Apr 19th, 12:00 AM

Proceedings, MSVSCC 2012

Old Dominion University, Department of Modeling, Simulation & Visualization Engineering

Old Dominion University, Virginia Modeling, Analysis & Simulation Center

OLD DOMINION UNIVERSITY
IDEA FUSION

Virginia Modeling, Analysis and Simulation Center

Modeling, Simulation & Visualization
2012 Student Capstone Conference

Final Proceedings

April 19, 2012
Old Dominion University

2012 Student Capstone Conference Final Proceedings Introduction

2012 marks the sixth year of the VMASC Capstone Conference for Modeling, Simulation and Gaming. This year our conference attracted a number of fine student written papers and presentations, resulting in 51 that were presented on April 19, 2012 at the conference.

The tracks that the papers were divided up into included the following:

- Science & Engineering
- Homeland Security/Military M&S
- Medical M&S
- Gaming & VR
- Training & Education
- Transportation
- Business & Industry

The following proceedings document is organized into chapters for each of the tracks, in the above order. At the beginning of each section is a front page piece giving the names of the papers and the authors.

Science and Engineering

VMASC Track Chair: Dr. Bridget Giles
MSVE Track Chair: Dr. Rick McKenzie

Highly Coupled and Minimally Cohesive Requirements
Author(s): John J. Johnson IV

A Review of Modeling Dynamic Changes Using Petri Nets
Author(s): Kristen Mark

Methodology for Valuing and Simulating a Network Protected by Homomorphic Encryption
Author(s): Geoffrey M. Hansen

Memes and Narremes in Narrative Networks
Author(s): David K. Wright

Determination of Optimum Fuel Injector Parameter Values for a Model Scramjet Combustor
Author(s): Dr. G. Selby, and A. Humadi

Theoretical Design and Application of a Robotic Hand
Author(s): Daniel Biller and Cheng Lin

Test Environment for Analyzing Event Based Simulation Executives
Author(s): Spencer Lane, and James F. Leathrum Jr.

Initialization of a Large-Scale Simulation Using Mined Data
Author(s): Olcay Sahin

Highly Coupled and Minimally Cohesive Requirements

Measuring Traceability and Design Space

John J. Johnson IV
Systems Thinking and Solutions
Broadlands, VA, USA
johnjohnson@ThinkSTsolutions.com

***Abstract*—Systems Engineers operate in the design space where design decisions determine solutions. Design decisions depend on clearly defined requirements. Systems Engineers are challenged with translating needs in the problem space into requirements for the design space. Design space decisions can constrain or expand the set of potential solutions for the requirements. Current methods for exploring design space are focused on optimizing the search for the best solutions.**

A characteristic of good systems engineering designs is traceable requirements. The common problem in managing design space and requirements traceability is the absence of objective measurements.

This paper presents a framework that measures available design space and requirements traceability based on coupling and cohesion concepts. The quantitative basis of the framework enables the modeling and simulation of design decision implications related to requirements traceability and available design space.

Keywords-coupling; cohesion; system of system; design space exploration; requirements traceability; modeling and simulation; formalism

I. INTRODUCTION

The motivation for this paper is solving a common problem associated with system design space and requirements traceability.

Systems are collections of hardware, software, people, facilities, and procedures that are organized to accomplish common objectives [1]. Sponsors or customers need systems to solve problems that they are particularly interested in. They express their needs in terms of natural language originating requirements that define the problem [2]. The system engineering process translates these requirements into formal technical specifications for solutions that meet the customer's needs.

An interesting dilemma emerges during the systems engineering process; ambiguity vs explicitness. Consider the systems engineering V-model in Figure 1 [3].

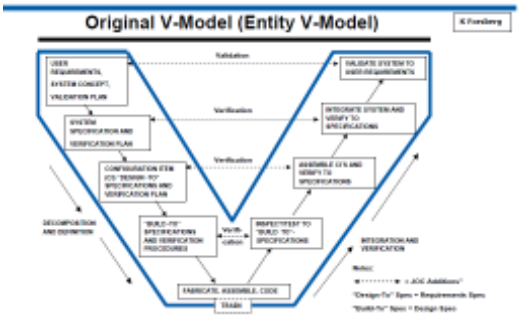


Figure 1: Systems Engineering V-Model

As requirements flow down the left side of the "V" from customers (i.e., users) through various levels of the systems engineering process, the volume and specificity of requirements increase and the solution becomes more defined. The dilemma is whether to specify more or fewer requirements.

In an analysis of systems defects for an electronic audio product, Lauesen & Vinter reported that 54% of the defects were related to requirements [4]. The report goes on to imply that more specific requirements and greater levels of abstraction might reduce ambiguity and consequently reduce requirement related system defects. This may in fact be true. However, more specific requirements will also reduce the design space domain and limit the range of available solutions. Specifying too many requirements can cause excessive cost, unrealistic expectations, schedule delays, and inhibit innovation [5][6].

Through out the development process, stakeholder requirements and outputs of the development process must stay in continuous alignment. Requirements traceability is intended to ensure that this objective is accomplished [7]. Requirements traceability can be defined as the ability to describe and follow the life of a requirement, in both a forward and backward directions and the degree to which there is a match between the system's design and its requirements [8][9] There are many motivations to incorporate effective requirements traceability into the system engineering process. These motivations include ensuring that systems do what they are required to do; improving project and change management; demonstrating compliance to outside stakeholders; and

facilitating the product's long-term maintenance [24]. Of particular importance, is the ability gained through requirements traceability to conduct impact assessment of requirements changes on the project plans, activities, and products produced by the system design process [10].

Modeling and simulating can help stakeholders gain a better understanding of their systems before they are actually constructed. A desirable attribute in the design of complex systems (i.e., system of systems) are highly traceable requirements with optimal design space at each stage of the design process. The common problem associated with design space and requirements traceability is the absence of quantitative definitions. Models can not simulate the traceability and design space characteristics of systems unless quantitative definitions are created.

This paper proposes advancing the state of the art by solving the requirements traceability and design space problem through formalism. The framework for the proposed solution is based on the concepts of coupling and cohesion. Coupling is a measure of the relationships between units while cohesion is a measure of relationships within units [11][12][13].

There are six major sections in this paper. Section I presented the background and problem statement and the advancements to the state of the art. Section II discusses works related to cohesion and coupling, Systems of Systems, design space, and requirements traceability. Section III presents formal definitions for the traceability and design space. Section IV is a presentation of the framework for applying cohesion concepts to measure design space followed by Section V which presents coupling concepts applied to measure traceability. Section VI is a brief discussion on the implications of the proposed framework on modeling and simulation.

II. Related Works

A. Coupling and Cohesion

Coupling and cohesion metrics lack formal and standardized definitions and thus for each metric there is more than one interpretation [11][12]. However, Briand and Morasca offer four common properties of cohesion and coupling metrics:

1) Non-negativity and Normalization Property: Measures are non-negative and independent of the system size.

2) Null Property: There are no members in the Cohesion or Coupling sets (i.e., the sets are empty) if there are no internal relationships between elements in the system.

3) Monotinicity Property: Additional internal relationships in the system cannot decrease cohesion.

4) Associative Property: Grouping two or more unrelated system elements cannot increase cohesion or coupling.

The Coupling of a subsystem characterizes its interdependence with other subsystems in the system. A subsystem's cohesion characterizes its intra dependencies [13][14]. The concept of cohesion assesses the tightness with which elements in related systems are connected [14]. There are two kinds of coupling, inbound and outbound [14]. Outbound coupling captures the relationship between elements inside a system to elements outside the system while inbound captures relations between elements from outside the system to elements inside the system.

It is generally held that, of two comparable systems, the one with lower coupling (i.e., higher isolation) and higher cohesion, is (or should be) of better quality, reliability and maintainability [14][15]. The property of high cohesion helps localize the impact of change to within a system. The ripple effect of change is less likely to propagate beyond the affected system of subsystem because of its reduced coupling with neighboring systems and subsystems [1]. Large-scale software-based systems warrant lengthy development cycles during which there is a constant evolution of stakeholder needs and technology specifications. It is imperative that in order to function satisfactorily over the life of the system, it must accommodate change. This desire for change tolerance tends to be focused on minimizing the impact of the change by isolating subsystems and reducing the opportunity for a ripple effect through out the system to other subsystems [1].

B. System of Systems (SoS)

A System of Systems is a "super system" comprised of other elements which themselves are independent complex operational systems and interact among themselves to achieve a common goal [3][16]. There are layers within the SoS consisting of smaller systems and even lower layers until eventually a level of elements is reached that cannot exist independently. Each independent element of a SoS achieves its designated goals whether or not it is attached to the "super system" or not.

C. Design Space

Requirements can also be considered constraints that limit the decision variables which determine the possible system solutions [17]. Designing systems is an iterative process involving many choices. Design Space Exploration (DSE) helps designers in discovering the optimal solutions among all possible combinations after mapping functional to architectural specifications [18]. The universe of all possible solutions is the design space [17]. DSE refers to the activity of discovering and evaluating design alternatives during system development [19]. Most DSE methods are focused on narrowing the design space or efficient search methods because there are too many potential solutions to consider [19][20][21]. However, reference [17] discusses expanding the design space in search of additional solutions through analogical reasoning, mutation and variable manipulation. Whether narrowing or expanding of the design space, current approaches do not offer a method for measurement or quantification.

D. Requirements Traceability

Stakeholders have the need to look across their SoS and determine impact of change at the enterprise level as well as at each subsystem level. There is also the need to justify the portfolio of systems in the enterprise to sponsoring stakeholders. For federal systems, an agency (i.e., enterprise)

is mandated by Congress to effectively and efficiently execute its missions and to justify its capital investments [22]. Justification is supported by linking systems acquired with capital investments to originating stakeholder requirements and legal mandates. This ability to link systems and requirements must be intentionally designed into the system.

There are numerous approaches that attempt to address requirements traceability. These approaches include cross referencing schemes, traceability matrices, and constraint networks, [24]. The approaches differ in the complexity of information they can trace, in the quantity of interconnections they can control, and in the extent to which they can maintain traceability when faced with ongoing changes in requirements. The common thread between requirements traceability approaches is the focus on reducing large volumes of information into manageable forms; enabling more accurate and complete traceability; and visualizing traceability [24]. As with Design Space, current approaches do not offer for measurement or quantification methods.

III. FORMALIZING DESIGN SPACE AND TRACEABILITY

A. *Design Space Formalism*

Systems are designed through an iterative process of decisions that systematically decompose ancestor requirements from the problem space into child requirements in the design space. Each decomposition is an abstraction of the ancestor requirements into increasing levels of detail so that the next phase of engineering can be completed. This decomposition process includes ancestor requirements that may have been implied but not specifically mentioned by stakeholders or those that might emerge as a result of the design process [1].

The decisions made during the design process create a set of possible design solutions called the design space [25]. The design space for determining a solution is constrained by the set of requirements abstracted from the ancestor requirements. All possible members of the universal requirements and solution sets may or may not be known, however, both sets are finite and countable. The relationship between the universal set of requirements and set of solutions is also surjective. A solution must have one or more requirements and a requirement must have one or more solutions. Otherwise, a requirement will be unsatisfied or a solution will be unnecessary.

The available design space is largest at the beginning of the design process [26]. At this point, very few if any design decisions have been made and all possible design choices are available. The greater the number of abstractions the more the design space is constrained, i.e., the smaller the design space becomes. The outcome of a declining design space is a declining set of available solutions. The process continues until a single solution has been defined.

Once the systems engineering process is complete, each systems in a SoS will have a unique solution. Each system's solution has a unique set of requirements. The set of sets for the SoS requirements are asymmetric. There may be common members in the requirement sets but no two sets are exactly the same. The relationship between the universal set of requirements and the universal set of solutions is inversely proportional.

Every solution cannot accommodate every requirement. Therefore, as requirements are specified through the decomposition process, potential solutions are eliminated. The decomposition process creates a proper subset of the universal solution set. This proper subset represents the solutions that have been eliminated by the requirement decomposition process. The compliment of this subset represents the available solutions (i.e., the design space).

As the number of requirement abstractions increase the number of eliminated solutions increases resulting in a diminishing design space.

B. *Traceability Formalism*

Consider a SoS containing systems A and B. The set of requirements for each system is an improper subset of the universal set of SoS requirements. Reusable requirements are defined by the intersection of sets A and B which represents the requirements that are common in both systems. The intersection between A and B represents the requirements traceability for the SoS. Set A ≠ Set B and the intersection of A and B is a non-empty set.

The traceability of the SoS has a proportional relationship with the size of the intersection between the improper sets of requirements in the SoS.

IV. COHESION AS A MEASURE OF DESIGN SPACE

Consider Figure 1 which represents one of three systems in an example SoS. It depicts the abstraction of system-1 as defined by a stakeholder. There are three branches of requirements and four levels of abstraction. The fourth level of abstraction occurs in branch B.

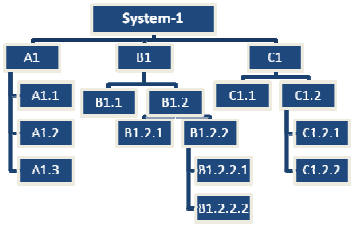


Figure 2: Node Tree for System-1

Each level of abstraction in the branches of the node tree equals one degree of separation from the original root requirement. For example, A1 in system-1 is one degree of separation from the root requirements System-1. A1.1 is two degrees of separation. The average degrees of separation for the system are the average degrees of separation of the branches in the node tree.

The cohesion of the system at any snap shot in time during system development, is the average degrees of separation at the snap shot, as a fraction of the average degrees of separation for the entire system upon completion (i.e., the final system).

Consider Figure 3 and Table 1. Figure 3 compares the second system in the SoS at the final design stage vs system-2 at a snap shot during an earlier stage of the design process. At the snap shot there are three branches in the node tree, three levels of abstraction, and nine requirements. The final design also has three branches. However, there are five levels of abstractions and thirteen requirements.

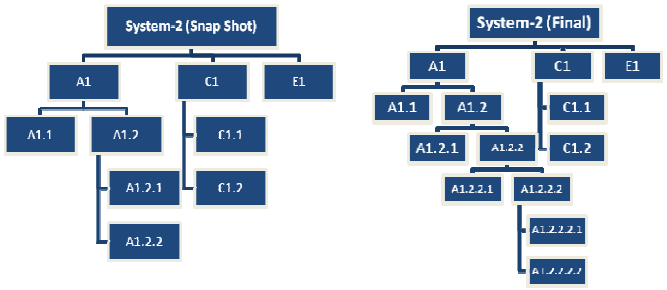


Figure 3: System-2 Snap Shot vs Final

Table 1: Cohesion Example

System-2 (Snap Shot)	Degrees of Separation	Requirements
Branch-A	3	5
Branch-B	0	0
Branch-C	2	3
Branch-D	0	0
Branch-E	1	1
Average	1.2	
Total		9

System-2 (Final)	Degrees of Separation	Requirements
Branch-A	5	9
Branch-B	0	0
Branch-C	2	3
Branch-D	0	0
Branch-E	1	1
Average	1.6	
Total		13

Equation 1: Cohesion = Average Snap Shot Degrees of Separation ÷ Average Final Degrees of Separation

System-2(Snap Shot) Cohesion = 1.2 ÷ 1.6 = 75%

The average degrees of separation for System-2 at the snap shot are 1.2. The average degrees of separation for the final design are 1.6. The cohesion for the system-2 snap shot is therefore 75%.

As degrees of separation from the root requirement increase more design choices are implicitly or explicitly made and that the available design space declines. Equation 2 and Figure 4 represent the relationship between available design space (ADS) and Cohesion. In this paper it has been assumed that the relationship is linear; however, it is quite possible that the relationship is of a higher order.

Equation 2: ADS = [-M1(1-Cohesion)]x100%

Where: M1 is the slope for the Traceability function

0< Cohesion <1

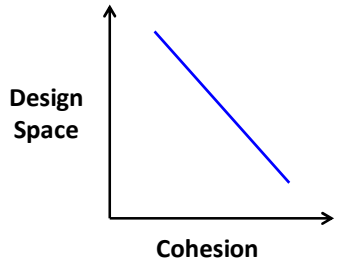


Figure 4: ADS as a Function of Cohesion

System-2 Snap Shot ADS = (1-.75) x100% = 25%

The interpretation of the cohesion and ADS statistics in the System-2 snap shot example is:

- 75% of the design space has been consumed by design activities completed as of the snap shot.
- Design activities remaining after the snap shot are limited to 25% of the design space.

V. COUPLING AS A MEASURE OF TRACEABILITY

Systems are coupled if methods or instances of variables in one system are used by the other. The basic idea underlying all coupling metrics is very simple: count how many intersystem interactions there are in the system [27]. An intersystem interaction is the occurrence of requirements in two or more systems that are members of the same SoS. These requirements shall be defined as reusable requirements.

Consider Figure 5 which represents two of the three systems in an example SoS. It depicts the traceability between System-1 and System-2. System-1 has four levels of abstraction, three branches and 16 requirements. System-2 has five levels of abstraction three branches and 13 requirements. Even though both systems have the same number of branches, only two of the branches are common in both systems (branch A and C). It is observed that for branch A, that only three of the possible nine requirements are common in both systems. Within in branch C there are three common requirements.

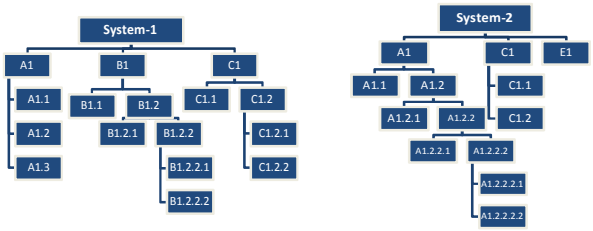


Figure 5: SoS Traceability

The coupling measurement for a SoS can be normalized by dividing the total reusable requirements by the total requirements in the SoS (see Equation 3, Figure 6 and Table 2).

Equation 3: Coupling =∑ reusable requirements ÷∑ system requirements

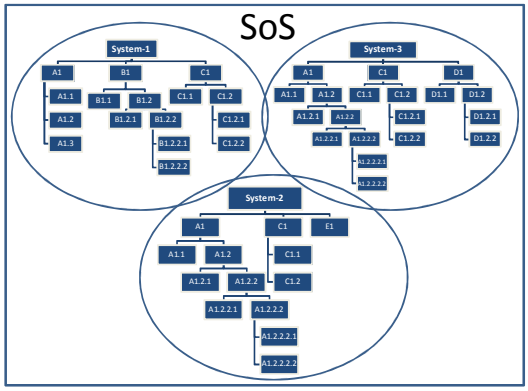


Figure 6: SoS Traceability

Figure 6 depicts the entire SoS. Total requirements vs reusable requirements for each branch of the node tree are summarized in Table 2.

Table 2: SoS Coupling

SoS		
	Total Requirements	Reusable Requirements
Branch-A	22	21
Branch-B	7	0
Branch-C	13	13
Branch-D	5	0
Branch-E	1	0
Total	48	34

SoS Coupling = 34÷ 48 = .71

Increasing reusable requirements relative to total system requirements increases the traceability in the SoS. Equation 3, Equation 4, and Figure 7 define the relationship between SoS traceability and Coupling. Again, the assumption has been made that the relationship is linear with the acknowledgement that the actual relationship could be of a higher order.

Equation 4: Traceability = M1(Coupling)x100%

Where: M1 is the slope for the ADS function

0≤ Coupling <1

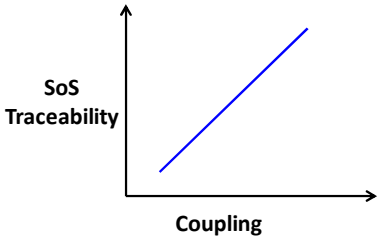


Figure 7: SoS Traceability as a Function of Coupling

The interpretation of the coupling and traceability statistics is 71% of the requirements in the SoS have interclass interactions and are reused in two or more subsystems of the SoS.

VI. System Modeling & Simulation

Models are approximations of real-world systems and simulations are dynamic representations of the behavior of those systems over temporal and special scales [28]. Modeling and simulation (M&S) creates virtual worlds where decision makers can explore the consequences of their decision without incurring the expense, risk, or time to do so in the real world. [29].

Understanding the dynamic behaviors and potential outcomes of a SoS is an appropriate application for M&S. Decision makers may wish to study the impact of their decisions on a variety of outcomes (i.e., dependent variables). However, dependent variables can only be studied through simulation if they are defined in terms of independent variables that are formally represented in the model.

Cohesion and coupling are independent variables based on decisions that impact reusable requirements and degrees of separation between requirements. Adding Cohesion and coupling as attributes of a SoS model allows decision makers to test drive the SoS and explore the "what if" implications of their decisions on design space and traceability.

VII. Conclusions

There are risks and benefits to high degrees of traceability in systems. Stakeholders will appreciate the ability to justify design decisions and trace investments to requirements. Analyst will appreciate the ability to quickly determine the impact of change on the system given fewer unique requirements and more reusable requirements. The coupling metric also creates the ability to objectively compare systems on the basis of their traceability traits. However, the risk of too much interdependency or coupling can open the door the ripple effects through the SoS when there are failures associate with a reusable requirement. In either case the ability to measure traceability in a SoS promotes better decision making and awareness of risk.

Innovation is generally accepted as a positive and desirable trait. Providing engineers with a generous design space appropriate for the level of design activity, enables and promotes innovation. When optimizing design space is one of the goals for the design process, an objective measurement is necessary to assess performance and drive improvement. The cohesion metric creates the ability to objectively compare design teams and programs on the basis of their management of design space throughout the development process.

The coupling and cohesion measurements proposed in this paper require additional study. The proposed cohesion measurement requires information about the final design of the system. A method for estimating the final design is necessary in order to make the cohesion measurement useful during the design process. Otherwise, the value of measuring cohesion is limited to analyzing historical events. Additional work is also needed to formally define coupling vs traceability and cohesion vs design space. It was assumed in this paper that the relationships are linear, however, they may in fact be non-linear.

Future work includes a pilot applying the proposed measurement framework presented in this paper will be conducted on an actual SoS to test the hypothesis that cohesion and coupling explain traceability and design space variance. The empirical data will help to develop a formal model of the relationships between the cohesion and coupling independent variables and the Design Space and Traceability dependent variables. The metrics will be revised and the concepts adjusted based on the results of the pilot.

ACKNOWLEDGMENTS

Andreas Tolk, Ph.D. is acknowledged for his technical contributions and general support during the development of this paper. Saikou Diallo, PhD is acknowledged for his review of the paper and insightful suggestions.

REFERENCES

[1] Buede, D. (2000). The Engineering Design of Systems. New York: John Wiley & Sons.

[2] Ravichandar, R., J. D. Arthur, S. A. Bohner, D. P. Tegarden, (2008) Improving change tolerance through Capabilities-based design: an empirical analysis *Journal of Software Maintenance and Evolution: Research and Practice*, vol. 20, issue 2, pp. 135 - 170, : John Wiley & Sons.

[3] Clark, J.O. (2008), System of systems engineering and family of systems engineering from a standards perspective. System of Systems Engineering, 2008. SoSE '08. IEEE International Conference on. IEEE Conference Publications.

[4] Lauesen, S., & Vinter, O. (2000). Preventing Requirement Defects. Proceedings of the Sixth International Workshop on Requirements. Stockholm.

[5] Little, T. (2005). Requirements: The More the Better?, Ask Magazine, Issue 14, National Aeronautics and Space Administration (NASA), http://askmagazine.nasa.gov/issues/14/features/ask14_requirements_littl e.html

[6] Knowles, I.; Malhotra, A.; Stadterman, T.J.; Munamarty, R , (1995), Framework for a dual-use reliability program standard, Reliability and Maintainability Symposium, 1995. Proceedings., Annual. Page(s): 102 – 105. IEEE Conference Publications

[7] Ramesh, B.; Jarke, M. (2001). Toward reference models for requirements traceability . Software Engineering, IEEE Transactions on, Volume: 27 , Issue: 1, Page(s): 58 – 93. IEEE Conference Publications.

[8] IEEE Standards Software Engineering, IEEE Standard Glossary of Software Engineering Terminology, IEEE Std. 610-1990, The Institute of Electrical and Electronics Engineers, 1999, ISBN 0-7381-1559-2.

[9] Gotel, O.C.Z.; Finkelstein, C.W. An analysis of the requirements traceability problem. Requirements Engineering, 1994., Proceedings of the First International Conference on. Page(s): 94 – 101. IEEE Conference Publications.

[10] CMMISM for Systems Engineering/Software Engineering, Version 1.02 (CMMISW/SW, V 1.02); CMMI Staged Representation, CMU/SEI-2000-TR-018, ESC-TR-2000-018; Continuous Representation, CMU/SEI-2000-TR-019, ESC-TR-2000-019; Product Development Team; Software Engineering Institute; November 2000.

[11] Husein, S., & Oxley, A. (2009). A Coupling and Cohesion Metrics Suite for Object-Oriented Software. International Conference on Computer Technology and Development (pp. 421 - 425). INSPEC.

[12] Marcus, A., & Poshyvanyk, D. (2005). The Conceptual Cohesion of Classes. Proceedings of the 21st IEEE International Conference on Software Maintenance (pp. 133- 142). IEEE.

[13] Allen, E., Khoshgoftaar, T., & Chen, Y. (2001). Measuring coupling and cohesion of software modules: an information-theory approach. *Software Metrics Symposium, 2001* (pp. 124-134). London: IEEE

[14] Briand, L.C.; Morasca, S.; Basili, V.R. (1996). Property-based software engineering measurement. Software Engineering, IEEE Transactions on, Volume: 22 , Issue: 1 IEEE Journals & Magazines.

[15] Moser, S., & Misic, V. (1997). Measuring class coupling and cohesion: a formal meta-model approach . Software Engineering Conference (pp. 31 - 40). INSPEC.

[16] Karcanias, N., & Hessami, A. (2010). Complexity and the notion of system of systems: Part (II): defining the notion of system of systems. World Automation Congress (WAC), (pp. 1-7). IEEE.

[17] Gero, J.; Kumar, K. (2004). Expanding Design Spaces Through New Design Variables. Design Studies ,Volume 14, Issue 2, Pages 98-224. Elsevier Ltd.

[18] Mura, M., Murillo, L., & Prevostini, M. (2008). Model-based Design Space Exploration for RTES with SysML and MARTE. Forum on Specification, Verification and Design Languages (pp. 203 - 208). INSPEC.

[19] Kang, E.; Jackson, E.; Schulte, W. (2011). An Approach for Effective Design Space Exploration. Foundations of Computer Software. Modeling, Development, and Verification of Adaptive Systems, Pages 33-54. Springer Dordrecht Heidelberg.

[20] Thompson, A.; Layzell, P.; Zebulum, R.S. (1999). Explorations in design space: unconventional electronics design through artificial evolution Evolutionary Computation, IEEE Transactions on, Volume: 3 , Issue: 3 Page(s): 167 – 196. IEEE Journals & Magazines.

[21] Schafer, B. C.; Wakabayashi, K. (2010). Design Space Exploration Acceleration Through Operation Clustering. Computer-Aided Design of Integrated Circuits and Systems, IEEE Transactions on, Volume: 29 , Issue: 1, Page(s): 153 – 157, IEEE Journals & Magazines.

[22] US Congress. (1980). Submissions of Plans to Congress. 14 USC Sect 663 . DC, USA: US Congress.

[23] Office of the Deputy Under Secretary of Defense for Acquisition and Technology, Systems and Software Engineering. Systems Engineering Guide for Systems of Systems, Version 1.0. Washington, DC: ODUSD(A&T)SSE, 2008.

[24] Panis, M.C. (2010). Successful Deployment of Requirements Traceability in a Commercial Engineering Organization...Really. Requirements Engineering Conference (RE), 2010 18th IEEE International, Page(s): 303 – 307. IEEE Conference Publications.

[25] Saxena, T.; Karsai, G. (2011). A Meta-Framework for Design Space Exploration. Engineering of Computer Based Systems (ECBS), 2011 18th IEEE International Conference and Workshops on. Page(s): 71 – 80. IEEE Conference Publications.

[26] Pimentel, A. (2005). A Case for Visualization-integrated System-level Design Space Exploration. Embedded Computer Systems: Architectures, Modeling, and Simulation. 5th international workshop, SAMOS 2005, Samos, Greece, July 18-20, 2005. Page(s): 455 – 469. Springer Dordrecht Heidelberg.

[27] Gui, G., & Scott, P. (2008). New Coupling and Cohesion Metrics for Evaluation of Software Component Reusability. The 9th International Conference for Young Computer Scientists (pp. 1181 - 1186). INSPEC.

[28] Sokolowski, J.; Banks, C. (2009). Principles of Modeling and Simulation: A Multidisciplinary Approach. John Wiley & Sons.

[29] Sterman, J. (2000). Business Dynamics, Systems Thinking for a Complexed World. Mcgraw-Hill

A Review of Modeling Dynamic Changes Using Petri Nets

Kristen Mark
Dept. of MSVE
Old Dominion University
Norfolk, VA

I. Introduction

There exists a large number of systems that are modeled or controlled using some version of a Petri Net (PN). Some systems are well modeled with a generic PN, whereas others have inspired users to research and design new types of PNs. Various dynamic systems have been studied over the years, and have inspired numerous developments relating to PNs.

A. Dynamic Systems

For the purposes of this report, the term *dynamic system* refers to a system that undergoes frequent dynamic changes. Furthermore, the term *dynamic change* refers to the "…problem of handling old cases in a new process, e.g., how to transfer cases to a new, i.e., improved version of the process" [1]. In essence, the dynamic system is a system that changes frequently over time.

Dynamic systems represent a significant portion of systems that may need to be modeled or controlled. When a PN is used as part of a controller, the reasons for keeping the PN up-to-date are obvious: if a controller of a large system is out of sync, it is not likely to work properly. This could cause a variety of problems ranging from a loss of the time used to update the PN to safety violations that could result in injuries or deaths.

Maintaining a model of a dynamic system is also important. While a model is a necessary and important tool for evaluation during pre-execution stages, the model is still useful while the implemented process is live. After the process has gone live, the model can still be used "…for what-if scenarios, for question answering, and for making pro-active suggestions" [2]. Models allow the user to study the system without needing to use the actual system. This can save time and money, as well as promote safety. If the model does not take into account the dynamic changes, the user is unable to reap all the benefits that an accurate model can provide.

C. History of Research in Dynamic Systems

The problem of modeling dynamic systems has come to the fore over the last couple of decades. The origin of the studies seems to have evolved around an article written by Clarence A. Ellis and Karem Keddara in 1993. This article, entitled *Dynamic Change within Workflow Systems*, introduced the problem of performing "…structural changes dynamically while the system is in execution" [3]. In short, the article presented the idea that while it was possible to flush a system before implementing a system change, it is not always practical. In cases where systems make frequent changes, or changes that are time-sensitive, this is especially true. Instead, Ellis and Keddara believed that the creation of a process allowing a system to be reconfigured without disrupting the model was ideal. This inspired a number of proposed solutions to the problem of implementing systems with dynamic changes.

D. Petri Nets and Dynamic Systems

Though there are many ways that a dynamic system can be modeled, PNs are often used. When modeling a system that is concurrent, PNs are often seen as the ideal tool [4]. When one or more events are occurring simultaneously, PNs help to keep the model both simple and precise. PNs are used to because they provide a graphical interpretation of complex systems. Their graphical nature means that PNs are easy to understand while the mathematical foundation keeps them unambiguous [5]. The nature of PNs make them an ideal modeling technology for dynamic systems.

II. Dynamic Changes in Petri Nets

There are various methods that are used for modeling dynamic systems, and more are being researched and introduced. Some recently proposed solutions include Reconfigurable Petri Nets (RPNs), Improved Net Rewriting Systems (INRSs), and Intelligent Token Petri Nets (ITPNs).

A. Reconfigurable Petri Net

The RPN is an extension of previous solution called a Net Rewriting System (NRS). The RPN was proposed by Marisa Llorens and Javier Oliver in 1999 and was further expanded upon in 2004 in [6]. They offer that the combination of NRSs and PNs in which the flow relationships can change at runtime would allow for better modeling of dynamic systems.

1. Basis of a RPN

A RPN relies heavily on the foundation of a NRS. NRS combines PNs and graph rewriting systems. The general idea is that the PN defines the system configuration while the graph rewriting rule defines a change in configuration. Each configuration has a transfer relation equation. This equation relates two unconnected graph pieces together. When the transfer relation conditions match up, the module in the original PN is switched

with the module provided by the graph rewriting rule [6].

2. Explanation of a RPN

In [6], Llorens and Oliver propose that the places and transitions in a PN should remain the same. However, they believe that flow relations can be changed or even removed in order to reach the desired outcome. In the original NRS, "..the application of a rewriting rule not only affects net structure but net marking,"[6]. Alternatively, in the RPN process the application of a net rewriting rule does not affect the net marking. And furthermore, the places and transitions remain the same while the flow relations are altered. In the case of the RPN, the net rewriting rule is applied in a way that is less disruptive to the underlying PN.

B. Improved Net Rewriting System

The INRS is and alternative to the RPN. It was introduced by Jun Li, Xianzhong Dai, and Zhengda Meng in [7]. These three men believed that the RPN introduced by Llorens and Oliver was useable, but not in the format it is presented in. By modifying the original NRS, Li, Dai, and Meng introduce a new model of PN.

In [7], Li, Dai, and Meng explained that the RPN does allow for dynamic changes in a PN. However, they believed that it also had at least one major flaw: the RPN solution did not allow PNs to keep important behavioral properties. These properties include liveness, reversibility, and boundedness. The inclusion of these properties is an integral part of deadlock prevention in PN. These properties also ensure that there are a finite set of states and that the PN behaves in a cyclical fashion. Li, Dai, and Meng felt that these were important properties and should be preserved.

As an alternative to the RPN, the INRS was introduced. The idea of an INRS is more closely related to the NRS than a RPN. An INRS works much the same way as an NRS, but instead of replacing a few pieces of a PN, the creators proposed various rules be used to replace portions of the original PN with entire subnets. For example, there may be a set subnet for the removal of one place from the original PN. The portion of the original PN that contains the place to be removed is replaced with a new subnet that does not contain the undesired place.

In this INRS, each subnet must be a valid PN. Because of this requirement, the properties of PN are not lost with a reconfiguration. Li, Dai, and Meng believed that this solution was a far better fit.

C. Intelligent Token Petri Net

The idea of a ITPN was introduced by NaiQi Wu and MengChu Zhou in [8]. They found that previous models of PNs were unsuitable for some highly dynamic environments. The specific example that Wu and Zhou offered is a reconfigurable Automated Manufacturing System (AMS). However, this solution could be applied to other dynamic environments just as easily.

The various PN that were used to control the AMS before the ITPN was proposed are very briefly explained in order to increase the understanding of the background of the ITPN. Because the ITPN is offered as a replacement for these PN, and does not rely too heavily on them, only a general overview is presented.

1. Basis of an ITPN

i. Process-Oriented Petri Nets: POPN – This model of PN divided places into two types: processes and resources. This allows the part flow and flow relations to be defined explicitly.

ii. Colored Petri Net (CPN) – This model of PN uses color coded tokens and places to differentiate between separate entities. By noting the color combination of a token in a place, the user can differentiate between processes.

iii. Resource-Oriented Petri Net (ROPN) – In this model of PN, each type of resource is modeled as one place, as opposed to having a separate place for each resource. Colors are used to describe processes in the PN, and color can change during execution.

iv. Object-Oriented Petri Net (OOPN) – OOPNs define a resource as a subnet of transitions and places as opposed to a single place. The color of a give token is defined by a central controller, and thus allows the color to change depending on the output of a subnet.

2. Explanation of an ITPN

In [8], Wu and Zhou explained why they found none of the previous PNs to be an ideal solution. POPNs modeled well, but could very easily grow to be too large to use easily. CPNs were introduced as a result. They did manage to make the model much more compact. However, the CPNs lost the ability to explicitly define part flow and the large number of combinations made CPNs very complicated. When ROPNs came into use, they reintroduced the explicit flow of parts and further shrunk the size of the PN.

Unfortunately, each of the given example required that the process sequence be defined before execution began, which was not ideal situation for a highly dynamic environment. So OOPN were introduced to dynamic systems. OOPNs introduced a lot of flexibility to the PN, as they are easily modifiable. However, because of their disjointed nature, many of the properties of PN, such as liveness, could no longer be studied.

As a result, Wu and Zhou introduced the idea of an ITPN in [8]. They felt that if a place acted as a server, and a token acted as a client, system changes could easily be modeled. This proposed PN has each token holds knowledge of which process it should undergo at any given time. The token can then be route itself to a particular place based on its needs. This seems to be an ideal solution because as places are add, changed, an removed, the tokens are still routed to the appropriate places as long as at least one exists.

For example, a system contains two resources that are listed as offering a service A. Tokens are routed to the resources, which are either

set as separate places or as a single place with a combined maximum capacity, as the resources are needed. Should one resource be removed, the tokens are no longer routed to that resource, but will continue to be routed to the remaining resource. The knowledge of the needed resource lets the tokens be routed to whichever resource is available at the given time. When a resource is added or removed, its availability changes, and so does the routing of the tokens.

III. Applications of Dynamic Petri Nets

There are numerous applications for PNs that are capable of dynamic changes. Whether the PN being used are for the modeling of a system or the control of a system, there are various places in which PN can be used. Two examples include Automated Manufacturing Systems (AMS) and nuclear power plant emergency management.

A. Automated Manufacturing System

Of the three main PN models discussed, two were created with an AMS in mind. PN are used both as models and controllers for various AMS environments. Li, Dai, and Meng found that RPNs were unsuitable for modeling a specific sort of AMS called a Reconfigurable Manufacturing System (RMS). As a result, they proposed the INRS in [7]. With the same goal in mind, Wu and Zhou proposed ITPNs as a way to model a general in [8] .

An AMS is, by nature, a highly dynamic system. An AMS will change for various reasons. Some of the many reasons include a system failure, a new process or machine, and the production of a new or different product. While an AMS is undergoing these changes, the controller must also change. Many AMS controllers are based on PNs. Should the controller not be able to change dynamically with the system, errors may occur. As earlier stated, the consequences of these errors may be benign or they may be much more serious.

B. Power Plant Emergency Management

A second example of PN being used to model dynamic processes was proposed by Madjid Tavana in [5]. Tavana felt that PNs were an ideal solution to the modeling of emergency management in a nuclear power plant. They may also be used to control emergency systems that are responsible for things like activating sprinklers, alerting the fire department, and initiating the evacuation of a building.

In some situations there is enough notice of an emergency that a management plan can be specifically formulated to handle it. In many situations, that is not the case. In these cases, there is usually a basic plan of action that is followed in order to keep people safe. While these plans are better than none, in actual emergencies, there can be a great deal of unexpected difficulties that may cause the emergency plans to go awry: such as broken equipment and blocked escape routes may

cause changes in the system. In these situations, it is especially important for the PNs to be able to react to dynamic changes.

IV. Conclusions

The PN is an important modeling tool. PNs have the ability to model concurrent changes while keeping the model both simple and precise. This makes PN useful for various applications, and has helped to continue to inspire new research into the topic of PNs.

The ability of a PN to react to dynamic changes in a system is an important one. Researchers have been proposing new solutions to this problem for many years, and continue to do so to this day. The introduction of the idea of a dynamic system was an important moment in the growth of PNs.

References

[1] H. Ehrig, W. Reisig, G. Rozenberg, and H. Weber, "Petri Net Technology for Communication-Based Systems," Berlin: Springer-Verlag, 2003.

[2] W. Reisig and G. Rozenberg, "Lectures on Petri Nets II: Applications," Berlin: Springer-Verlag, 1998.

[3] C. A. Ellis and K. Keddara, "Dynamic Change within Workflow Systems," 1993.

[4] P. A. Fishwick, "Simulation Model Design and Execution: Building Digital Worlds," Upper Saddle River, NJ: Prentice Hall, 1995.

[5] M. Tavana "Dynamic process modeling using Petri Nets with applications to nuclear power plant emergency management," International Journal of Simulation and Process Modeling, Vol. 4, No.2, 2008.

[6] M. Llorens and J. Oliver, "Structural Dynamic Changes in Concurrent Systems: Reconfigurable Petri Nets," IEEE Transaction on Computers, Vol. 53, No. 9, Sept. 2004.

[7] J. Li, X. Dai, and Z. Meng, "Automatic Reconfiguration of Petri Net Controllers for Reconfigurable Manufacturing Systems with an Improved Net Rewriting System-Based Approach," IEEE Transaction on Computers, Vol. 6 No. 1, Jan. 2009.

[8] N. Wu and M. Zhou, "Intelligent Token Petri Nets for modeling and control of reconfigurable automated manufacturing systems with dynamical changes," Transactions of the Institute of Measurement and Control, Vol. 33, No. 1, 2011.

Methodology for Valuing and Simulating a Network Protected by Homomorphic Encryption

Geoffrey M. Hansen, *Department of Systems Engineering, United States Military Academy*

***Abstract*— This paper analyzes two of the major security processes of a network that work in the cloud: layered homomorphic encryption and complete homomorphic encryption. This is done to test how well value-focused thinking and the System Decision-Making Process apply to the analysis of policies, software, and alternative security settings on a network.**

Two variations to the systems decision making process were introduced to more effectively value network security: the use of a correlation coefficient and the use of an adjust metric of risk. The new adjusted risk metric resulted in a much harsher metric that drastically lowers the value of the entire system for instances of weakness in as little of one area of security. The correlation coefficient proved to be a valuable tool in adding metrics without overpowering the areas of security that are more easily measured.

These metrics were intended to be simulated in a programmed virtual environment. However, a great deal of work is still required in this area before anything more than a demonstration of a network can be gained through simulation. The research in this area did prove that the best avenue of approach would be with the use of an agent-based simulation such as NetLogo.

These metrics, along with the simulation that was crafted, can give the user a better understanding of the impact of homomorphic encryption on their network set up. The biggest take away from the base case models examined in this test was that the two variations of homomorphic encryption that we analyze have similar incremental increases in total security of the systems that they were deployed on. However, the complete homomorphic encryption network's advances in security came with a significant cost to the availability of the data. The layered approach to homomorphic encryption proved to have similar advances in security while keeping data availability at a more acceptable level.

***Index Terms*—Valuing; Availability; Security; Information; Network; Homomorphic; Encryption**

I. INTRODUCTION

HERE are two major trends in computing that give this area Tof research a great deal of consequence. First, there is a major trend in consumer interface design and computing to seamlessly integrate all the new and different types of devices used in computer work. Through the use of the cloud, users will soon be capable of seamlessly accessing files from their computer, their phone, and their tablet. For instance, Windows 8 attempts to give the user an interface that can be used both on the user's computer and their tablet with easy access to cloud space provided by Microsoft [1]. Second, there is a growing trend of cyber crime and cyber warfare. The biggest challenge this threat presents is that there is a very difficult cost/benefit analysis when it comes to current security applications and policies.

For every avenue the network administrator opens for intended users, another channel becomes available for hackers to access the same information. The most secure networks in the world are deep within layers of physical security and completely cut off from connections to the outside world. The least secure have no protocols or security software and access everything and anything on the open-source internet. The former is a security expert's dream; the latter his nightmare. In the completely secure network, a hacker is provided absolutely no foothold to gain entry into the network. This level of security comes with a huge cost both in terms of dollars spent and in accessibility. In order to make the network more accessible, the network administrator must open more ports for the desired user in order to give the user access to their desired information. Opening more ports comes at the cost of more vulnerability in the system, i.e. more avenues that a hacker can come at the network. This trade-off gets even more complicated when additional aspects of network security, such as intrusion detection systems (IDS), intrusion prevention systems (IPS), and firewalls are added. This research attempts to close the gap in the analysis of the trade-offs of network security through the use of more tools for the system administrator.

II. RESEARCH REQUIREMENT ANALYSIS

A. Research Concentration

The purpose of this research is to propose new formulas to more effectively value a system's security. These changes to the process will be tested by valuing the added security benefits of homomorphic encryption, both in terms of complete homomorphic encryption and layered homomorphic encryption. The value will be returned in terms of the advantages given in security of information versus the availability of the information. This should provide the system

engineer a tool to decide if the added cost of a homomorphic encryption application and the necessary network protocols that would come with such an application would add enough value to make the application worth the costs.

B. *System Security Risks*

In general, System security can be compromised in one of three ways; through a confidentiality attack, an integrity attack, or an availability attack. Defending against these three types of attacks can often counteract each other.

Confidentiality attacks are most successful when the attacker forces the attacked portal to access files and sends them to the attacker's point of attack over an extended period of time, while keeping his/her overall activity under the network's detection threshold. These types of compromises to network security are most commonly seen as hardware compromises, portable drive compromises, and viruses that slave the attacked portal computer to send files back.

Integrity attacks are even harder to detect then confidentiality attacks due to the fact that little information is sent between the attacker and slaved portal through the attacked network. A famous example of this type of attack is the Stuxnet worm which was used to change critical temperature control systems in the Iranian nuclear facility, thereby crippling the advancements of the Iranian nuclear program [2]. Worms like Stuxnet are not the only forms of integrity attack. Viruses have also been known to conduct integrity attacks.

Availability attacks are by far the easiest to design. A successful availability attack can be made by forcing the network to perform nothing but tasks that are commonly performed within the network. This kind of compromise occurs when the slaved portion of the network forces enough tasks to the point of slowing down (or shutting down completely) the surrounding nodes. This causes a performance dip in the network, either by security protocol or excessive degradation [3]. These types of attacks can be performed with a simple virus, which slaves a single portal of attack, or a complex botnet program, which slaves multiple portals for a coordinated attack [4].

TABLE 1
REPORT ON MOST COMMONLY USED SECURITY METRICS [10]

Metric	% of Industry that Uses Metric
Viruses Detected in User Files	92.3%
Viruses Detected in E-mail Messages	82.3%
Invalid Logins (failed passwords)	84.6%
Intrusion Attempts	84.6%
Spam detected/filtered	76.9%
Unauthorized Website Access (content filtering)	69.2%
Invalid Logins (failed username)	69.2%
Viruses Detected on Websites	61.5%
Unauthorized Access Attempts (internal)	61.5%
Admin Violations (unauthorized changes to network)	61.5%
Intrusion Successes	53.8%
Unauthorized Information Disclosures	38.5%
Spam not detected (missed)	38.5%
Spam false positives	30.8%
Other	23.1%

C. *Currently Employed Security Metrics*

Nearly every organization, from small business to Fortune 500 Corporation, analyzes how vulnerable they are to security threats and what impact an attack would have on their organization. However, each organization has their own parameters to work with when designing their methods for risk identification and management. Once a given organization has identified their risk, they then have to make decisions as to what risks are worth preventing, given the tradeoffs associated with additional security. Great deals of companies even believe that, given these vast difference between themselves and other companies in industry and government, they have very little or nothing to learn from the industry of network security at large, hence they create their own policies from scratch. Typically, these companies are left analyzing the data they have access to in the raw form with little more than intuition and educated speculations [5]. Though each organization generally has their own mix of metrics used the metrics all generally have similar purposes. The Robert Francis Group, an IT research firm, performed a study on which metrics are most commonly used in industry. The results are presented in Table 1.[6].

Not all organizations have taken to such an incomplete and subjective style of risk management. Several organizations have developed their own processes that aim to give a more complete picture of a network's security. One such institution is the United States Government. The National Institute of Standards and Technology (NIST) have published a guideline for mitigating risk and establishing controls on risks that have passed their cost/benefit analysis. This process does produce a finite risk metric, but does so through subjective analysis. It combines the typical subjective approach based on given data to a scale based on subjective analysis. This can be dangerous, especially when the estimator has some kind of bias, either due to his skill level or his stake in the outcome [7].

D. *Cloud Computing*

As organizations struggle to successfully value their network's security, more and more corporate and private users are beginning to migrate towards cloud computing and cloud storage. This makes the problem of network security valuing even more important.

The concept of using the cloud in computing and storage is a relatively simple one. The cloud is made up of all available servers and resources. The user's computer uses the cloud much like if it were interacting with a single server. It does this by referencing what it sees as a location. In the years prior to cloud computing, network administrators would assign a specific, physical space on the hardware of a server. This is not the case when interacting with the cloud. The location is instead a pseudo-address. Instead of referencing a specific location, it references a location that has been carved from the cloud. This location can change from day to day without the user knowing or having to make any adjustments.

The power of the cloud is that it can leverage vastly more resources and can operate much more dynamically. The cloud can use algorithms to change the storage/computing locations

of data from location to location in order to maximize efficiency and effectiveness. Using a cloud configuration requires greater set up time and might require much more intense programming. However, cloud computing can save a great deal of money in hardware costs. It can also enable the user to perform all operations in the cloud, meaning that the system user can work on an entire file without ever downloading it to the user's local terminal [8].

E. Homomorphic Encryption

A homomorphic encryption algorithm is an encryption technique that allows the data to be changed or modified either in the encrypted form or in the plain text form while still producing the same result after an encryption or decryption. Basically any encryption algorithm that holds the following is considered homomorphic;

$$\text{Enc}(a \oplus b) = \text{Enc}(a) \oplus \text{Enc}(b) \quad (1)$$

Presently, encryption techniques that completely meet these criteria are still highly theoretical and experimental [9]. Current fully homomorphic encryption formulas are limited by their enormous computational costs. Some of the best completely homomorphic equations are still nearly a million times slower than traditional methods of performing computations on data, where the data must be fully decrypted prior to being analyzed [10].

If this technique was perfected, it could produce a level of security that would allow a system to access the cloud without the use of the systems security key. The system would only need the encryption equation to set up the queries necessary for its tasked processes. The terminal user would theoretically conduct queries and adjustments to data in the cloud without ever having to save a copy of the security key onto the system being used to analyze the encrypted data. This means that the network administrators can give access to their cloud without risking their encryption key or the raw data. This gives the network administrators further options when there is an analytical demand above its capabilities where it is not practical to expand, either due to a temporary increase in demand of analytical processing power or an insufficient demand to justify keeping the processing in house. The basic concept of this is that deciphered data can be queried by searching for deciphered lines of data that correlate with unencrypted data; the third party computation source never needs the unencrypted data, only the data which has been deciphered already.

Breakthroughs in homomorphic encryption have allowed less complete formulas to drastically decrease the amount of processing time needed to perform computations on data with only minor decreases in search complexity. For instance, with a pseudo-homomorphic formula that simply uses layered encryption and either division or addition, the computation for a basic query could be done with only two or three times the normal cost [11].

III. METHODOLOGY

A. Value-Focused Thinking

Value-focused thinking revolves around a stakeholder's values. Value is weighted to what we might care more or less about, generally based on a scale of 0-100. The idea behind value-focused thinking is to refrain from subjectively comparing alternatives when faced with a problem. This is extremely common in the area of network security due to the focus on quick reaction to risks and breaches in security. The premise behind value-focused thinking is that the decision maker can almost always maximize his gains from the alternatives presented if value is considered, usually through the combination of possible alternatives into a single candidate solution that produces the most value [12]. This process is applied to the metrics used in this paper in order to capture the most complete idea of the value of different alternatives in homomorphic encryption when applied to different organizations and network administrators' values.

B. The Systems Decision Making Process

Traditionally, a systems engineer would approach a new problem with an iterative process called the systems decision making process. This method is widely used in industry and academia, and is taught and used heavily at the United States Military Academy

First, a systems engineer would use graphical models to better understand the system at hand. This engineer would do this by defining the system, conducting integration definition (IDEF) models of the system, gathering the needs of the stakeholders, and producing a functional hierarchy. For network security analysis, this part of the systems decision making process is necessary.

The value provided to the user is then examined. The analyst uses his previously produced functional hierarchy to produce a qualitative value model which outlines the objectives of the system. It also assigns metrics to each objective, in order to produce areas of data collection. The metrics are then converted into a value of 0-100, based on the mathematical model that represent's the stakeholders' preferences. The stakeholder's preferences can be used to transform, through the use of global weights and summation, all the given metrics into a single value for the system. Each category of value is still considered for candidate solutions, as improvements in any of the functions can help maximize overall value of the system. Focusing on areas that provide sub-optimal value will help the systems engineering team create better alternatives.

The system engineer throughout this process can leverage sensitivity analysis and other mathematical approaches to aid in their decision-making. After value has been accounted for the alternatives are measured against costs. The alternatives that are more expensive without providing additional value are screened out. After this the stakeholder then uses his own costs versus value preferences to determine which alternative is best for his organization.

C. Adjusted Risk Metric

The U.S. Army calculates risk by multiplying the

probability of a particular event occurring by the severity of that event. The IT world takes a similar approach [13]. The newest and most complete approach to risk assessment contains four variables that are multiplied together to give a risk index commonly referred to as the business adjusted risk metric. The variables are; probability of a security breach occurring, vulnerability severity, severity of a policy violation, and the value of the assets lost [14]. The business adjusted risk metric is the result of a process familiar to anyone who has done a risk assessment within the military or the civilian world [15]. This paper will use quantitative, historical data in the creation of the variables in this metric. It has also been manipulated to return a value of 0-100 in order to fit with the rest of the value models used in this analysis. The multiplication of the variables causes any particularly low-value variable to severely lower the associated risk metric. This is logical, as a weakness in one area can negate the strengths of the network or system as a whole. The following are the variables used in the adjusted risk metric:

Value of Frequency of Security Breaches is simply a number generated by the value focused thinking It demonstrates, on a scale of 0-100, how frequently a given system is successfully and unsuccessfully attacked. An unsuccessful attack registers close to a value of 100, while any successful exploitation brings a value closer to zero.

Value of the Vulnerabilities Currently in the System is how much of the system is open to security threats. This could be a representation of how readily available information is. Metrics that would create this value are: percentage of users with authorized access to security software, percentage of highly-privileged terminated employees whose privileges where reviewed over a given time interval, and/or cycle time to deprovision users.

Value of the Current Policy Violations is a measure of the risks user make that are contrary to the policies set by network administrators. One of the most widely-used examples of this would be passwords. For instance, the metric could measure the percentage of system users whose passwords do not match the established policies or the percentage of users who keep the same password for multiple logins.

Value of the Assets Lost to Security Violations is where the value of network loss is placed. Since we have frequency, current potential holes in our system, and a measure of those holes it is important to include a variable that accounts for the effects of the actual losses to the network exploiter. Metrics in this category must account for the level of the information lost as well as the principle that, if security of a file is compromised on one terminal, all the similar files throughout the system are compromised; you do not need all copies of a particular file to exploit a given piece of information.

The completed model proposed looks like this:

Business Adjusted Risk Metric = (Value of Frequency of Security Breaches * Value of the Vulnerabilities Currently in the System * Value of Current Policy Violations * Value of the Assets Lost to Security Violations)/1,000,000 (2)

D. Correlation Coefficient

When measuring the security of a network there is a wide range of metrics and methods of measuring uniquely available to each network. The metrics available at any given network greatly vary based on the unique availability and structure constraints for any given network. This leads some categories of security and availability to contain a great deal more metrics, and without a method to equalize these categories would result in some areas to be artificially inflated. A set of metrics in any particular category can often overlap and highly correlate. The proposed method of eliminating this correlation is to determine the correlation between like processes and discounting one of the processes by that correlation.

Correlation between two groups of data, based on raw-score calculation, is determined by the following function: [16]

$$Correlation\ Coefficent(r) = \frac{N\sum XY-(\sum X)(\sum Y)}{\sqrt{[N\sum X^2-(\sum X)^2][N\sum Y^2-(\sum Y)^2]}} \text{ (Eq. 3)}$$

This produces a value for r to fall between -1 and 1 therefore it is necessary to take the absolute value of the coefficient before using it as a discount. It is also important to note that the increase in r is a result of increased correlation; therefore, in order to determine the amount that is not correlated you must subtracted r from 1. This produces the following equation that will give you the discount value for two correlated processes:

$$d = |1 - r| \text{ (Eq. 4)}$$

This equation can be used to discount the second of two correlated metrics after they have been valued.

E. Agent-based Simulation

There are two main types of simulation used in creating a virtual model of a complex computer network. These are discrete-event simulations and agent-based simulations. Both types of simulation allow the user several advantages over a study that would use the full scale models. However, an agent-based simulation such as NetLogo allows the user to replicate coding logic that most closely matches the coding used in real life networks [17]. For instance, a systems engineer can have a 'turtle' (NetLogo terminology for an entity in the virtual space) 'hatch' (NetLogo term for create out of a turtle) packets that have embedded variables that store information that a packet would carry. This means that routers will interact with packets with the same logic set that a router in a real network would interact and direct a packet with. This allows the simulator to create viruses, worms, and other malicious attacks on a network with similar coding logic. This gives the user as close to a real environment as possible to test network alternatives.

The coding centered on the entity rather than the process also allows for much more adaptive coding environment. In discrete-event simulation, processes are pre-mapped out; entities are only allowed to travel along a route/path, without the ability to interact with the route. This causes the programmer in the programmer to have a great deal more work to simulate a change in the route or process taken by the

entity. This is a serious issue when simulating the vulnerabilities of a network because the goal of just about any hacker is to interrupt the process of the network in some way.

IV. RESULTS

The results of both the simulated NetLogo model and systems value model are based on random generated results of data. They serve as an indicator of no real network. This is because no network that has access to this level of data would release their information for publication. However, both models are created to be robust enough that the reader can test them by simply adapting them to their confidential network data.

A. NETLOGO- Agent-based Simulation

Unfortunately, when dealing with sensitive or classified network configurations the set up of the network will drive the program itself into a classification too high for the general public to learn lessons from its development. However, there is a great deal of value in the system set up used in this paper as it demonstrates a generic network and can give the user a general idea of the effects of policies and programmed attacks on a network.

The model created in support of this paper was able to replicate the following functions in the NetLogo environment:

1. Clock function that displays the time elapsed in the network based on a division of the amount of ticks that have elapsed.
2. "Packet lost count" is based off the amount of packets that were killed before reaching their destination.
3. Display that shows each router's perception of possible directions to send a packet even though the current NetLogo code has not reached the point where it can model the refresh rate common in a real router.
4. "Connectivity" function that returns a percentage of the network that can be reached from any given node.
5. "Network loss" function which shows the effects of an attack on the value of the data that was compromised by the attacker.
6. Density function that shows how much redundancy there is in the network set-up.
7. Average time of packet movement through the network and average time of packet movement through the network after a change in the network. These will model the effects of changes in the network's structure by the user or by a simulated attack.
8. Set up functions that give the user an ability to change the basic arrangement of the network without hard coding.
9. Directed attack on a single computer.
10. Destruction of router-router and router-node links.

Currently a separate NetLogo program has also been programmed, but not merged, that can show traffic over nodes in a network.

B. Standard Base Case Model

Due to the fact that the simulation approach did not get to the point where simulations could have been run for statistically significant results a generic base-case scenario was simulated five hundred times with a normal distribution. The base-case was broken into two main categories: security and availability. These categories were broken into 81 and 23 security metrics, respectively, taken from Andrew Jaquith's book *Security Metrics*. The simulation returned the following results, given a stakeholder who values security with a global weight of 0.44 and availability with a global weight of 0.56:

TABLE 2
BASE CASE SIMULATION RESULTS

	Average	Standard Deviation
Security Value	46.29	4.40
Availability Value	49.65	10.67
Total Value	48.19	6.29

C. Complete Homomorphic Encryption

This type of homomorphic encryption is very theoretical in nature. It takes massive amounts of computational power to process the data. This means the speed of the network is almost completely disintegrated. This alone would make any network administrator factor out this option. However, the results of our simulation adjusted to the changes in security and availability returns the following results with the same global weights:

TABLE 3
COMPLETELY HOMOMORPHIC ENCRYPTION SIMULATION RESULTS

	Average	Standard Deviation
Security Value	53.08	4.19
Availability Value	32.38	4.12
Total Value	41.58	3.11

The bottom line of the results is that the minor increase in security is vastly overshadowed by the high decrease in availability to data. However, the sensitivity analysis to this model produces a much wider range of value for the system. A network administrator who values security above all else might come to a different conclusion based on the results.

D. Layered Homomorphic Encryption

This type of homomorphic encryption is much more limited in the type of searches and the way the user can interact with the data he is processing in the cloud. However, it has a much less significant impact on the speed of the system. The simulation of this network produced the following results:

TABLE 3
LAYERING HOMOMORPHIC ENCRYPTION SIMULATION RESULTS

	Average	Standard

		Deviation
Security Value	52.88	4.13
Availability Value	40.61	4.03
Total Value	52.17	2.94

As you can see, the system has almost half the decrease in availability of the data and a network manager might be much more reluctant to screen this alternative out due to reduced speed of the system, especially if he values security heavily.

V. Conclusion

The changes to the way raw data is collected in the systems approach to valuing the security and availability of data in a network, both in the computation of the Business Adjusted Risk metric and the correlation coefficient, proved to be very effective. The use of simulation proved to have a great deal of potential, but requires a much more work before it is capable of producing significant enough data to be used in decision making. Overall, the different types of homomorphic encryption only proved valuable in select cases.

VI. Future Work

The results section discussed the future work in the areas of research that were examined by this paper. However, there is a whole other field in network security algorithms that was not touched on by this paper, quantum computing, which is quickly approaching the point where it could complicate the majority of the work presented here. Though the focus of this paper was not on the encryption, quantum computing is quickly outpacing homomorphic encryption and thus needs to be studied accordingly.

Acknowledgment

Special thanks to Major L Nunn, Department of Systems Engineering, United States Military Academy for the advice and support. Also special thanks to Major S. Huddleston and his capstone cadets for the initial research and advice in agent-based simulation.

References

[1] *Newman, Jared. "Windows 8: What You Need to Know." PCWorld (2011) <http://www.pcworld.com/article/229285/windows_8_what_you_need_to_know.html>.*

[2] *Broad, William J., John Markoff and David E. Sanger. Israeli Test on Worm Called Crucial in Iran Nuclear Delay. 15 January 2011. 29 November 2011 <http://www.nytimes.com/2011/01/16/world/middleeast/16stuxnet.html?pagewanted=all>.*

[3] *Asman, Brian C., et al. Methodology for Analyzing the Compromise of a Deployed Tactical Network. SIEDS Manuscript Submission. West Point, NY: United States Military Academy Department of Systems Engineering , 2011.*

[4] *Kristoff, John. "Botnets." NANOG Presentation. Evanston, IL: Northwestern University.*

[5] *Hutton, Alex and Douglas Hubbard. The great IT risk measurement debate, part 2 Bill Brenner. 2 March 2011 http://www.csoonline.com/article/669869/the-great-it-risk-measurement-debate-part-2?page=2*

[6] *Group, Robert Frances. "Collecting Effective Security Metrics." http://www.csoonline.com/analyst/report2412.html. 2004.*

[7] *Stoneburner, Gary, Alice Goguen and Alexis Feringa. Risk Management Guide for Information Technology Systems. Special Report. Falls Church, VA: National Institute of Standards and Technology, 2002.*

[8] *Miller, Michael. Cloud Computing Web-Based Applications That Change the Way You Work and Collaborate Online. Que Publishing, 2009.*

[9] *Hara, Takahiro, Vladimir I Zadorozhny and Erik Buchmann. Wireless Sensor Network Technologies for the Information Explosion Era. Springer, 2010.*

[10] *Halevi, Shai, Yehuda Lindell and Benny Pinkas. "Secure Computation on the Web: Computing without Simultaneous Interaction ." Lecture Notes in Computer Science (2011): 132-150.*

[11] *Halevi, Shai, Yehuda Lindell and Benny Pinkas. "Secure Computation on the Web: Computing without Simultaneous Interaction ." Lecture Notes in Computer Science (2011): 132-150.*

[12] *Keeney, Ralph L. Value Focused Thinking A Path to Creative Decisionmaking. Harvard: Harvard University Press, 1992.*

[13] *The Army Composite Risk Management System employs a similar equation in their risk assessment. They multiply the probability of something happening by the severity of the incident to return their own risk metric. "Composite Risk Management FM No. 5-19 (100-14)." Field Manuel. Washington DC: Department of the Army, July 2006.*

[14] *Ravenel, J. Patrick. "Effective Operational Security Metrics." The ISSA Journal February 2006: 31-33.*

[15] *The business world commonly uses the business-adjusted risk formula (BAR) to demonstrate an index of 0-25 to give an indicator of the level of risk in a given system, the higher the number the higher the risk. Jaquith, Andrew. "The Security of Applications: Not All are Created Equal." Research Report. 2002.*

[16] *Spatz, Chris. Basic Statistics Tales of Distributions. Belmont: Wadsworth, Cengage Learning, 2010.*

[17] *Wilensky, U. (1999). NetLogo http://ccl.northwestern.edu/netlogo/ Center for Connected Learning and Computer-Based Modeling. Northwestern University, Evanston, IL. (accessed February 24, 2012)*

Memes and Narremes in Narrative Networks

David K. Wright[1]

[1] ODU, VMASC, Norfolk, VA, 23517

This paper explores variations on a model of affordance transmission and its connection to narrative networks. The affordance model is based on adopting behavior through memes and social cues, and draws from well-established sociological and psychological factors, such as attention, motivation, goals, and influence from social networks. The model tells *who* is mimicking a recently encountered, novel behavior; this aspect of affordance transmission improves on older metrics, such as *rate* of diffusion from system-dynamics model, and *where* transmission takes place from basic social network models. The three main advantages of the model over previous attempts to model affordance are unintentional learning, learning through winner-take-all attention games where attention cues issue from the environment and social networks, and learning within context [1].

There remain open questions, which is half the focus of this paper. Several psychological issues are explored, such as the interaction between and within attentional and motivational cues, and how these interactions can be modeled and used for decision-making and behavior adoption. Extensions to the model in the above paragraph involve replacing averages and static cognitive variables with stochastic random variables. This increase in variance requires more simulation runs, but the model must allow for outliers, if only for exploring the search space. The model currently conjectures affordance potential based on the model agent's present time; this aspect is extended in this paper to include the past and future, so that the agent can recombine old and predict new affordance memes in harmony with its goals. The last extension deals with multi-tasking; the agents can currently only perform one action at a time.

The second focus of the paper is the overlap of affordance transmission and narrative networks. Both involve the exposure to and acquisition of new information, and the decision to either dismiss or disseminate the information. The smallest unit of cultural information, the meme, is the thing that is transmitted; in narrative networks, the smallest unit of a narrative is loosely defined as a narreme. This paper will attempt to sharpen the definition of a narreme and use narremes to model narratives and their spread through a social network. The model is validated against popular narratives from ancient times up to the present.

[1] Nye, Benjamin. *Modeling Socially Transmitted Affordances: A Computational Model of Behavioral Adoption Tested Against Archival Data from the Stanford Prison Experiment*. 2011. Working paper.

Determination of Optimum Fuel Injector Parameter Values For a Model Scramjet Combustor

Department of Mechanical Engineering

Old Dominion University

Norfolk, Virginia

Dr. G. Selby

A.Humadi

ABSTRACT

A mathematical investigation has been conducted to study the effect of fuel injector angle, diameter and location on a supersonic hydrogen-air jet flame square cross section combustor. The mathematical model uses multi-species Navier-Stokes equations, K-epsilon and an axisymmetry turbulent model. Shock wave effects were examined by changing the fuel injector angle, diameter and location with special attention given toward the combustion efficiency, thrust, pressure losses and equivalence ratio of the scramjet combustor. It was determined that when the fuel was injected at 90^0 angle with diameter 0.2 cm and at distance 29 cm from the air inlet, the thrust increased to its maximum and the pressure losses decreased to its minimum with an equivalence ratio almost close to unity.

NOMENCLATURE

A_1	≡	inlet area
A_2	≡	outlet area
d_f	≡	Fuel Injector Diameter (cm)
L	≡	Fuel Injector Location (cm)
h_1	≡	first back-step height
h_2	≡	second back-step height
L_1	≡	first back-step location
L_2	≡	second back-step location
$\dot{m}$	≡	mass flow rate
m_a	≡	mass of air
m_f	≡	mass of fuel
m_{f_1}	≡	mass of fuel at the inlet
m_{f_2}	≡	mass of fuel at the outlet
p_1	≡	inlet pressure
p_2	≡	outlet pressure
p_s	≡	static pressure
p_t	≡	total pressure
RNG	≡	renormalized group
T	≡	thrust ≡ $(p_2 A_2 - p_1 A_1) + \dot{m} (V_2 - V_1)$
T_t	≡	total temperature
V_1	≡	inlet velocity
V_2	≡	outlet velocity

Greek symbols

ϵ	≡	epsilon

$\eta_c \equiv$ combustion efficiency $\equiv \frac{m_{f_1} - m_{f_2}}{m_{f_1}}$

$\Phi \equiv$ equivalence ratio $\equiv \frac{\left(\frac{m_f}{m_a}\right)_{actual}}{\left(\frac{m_f}{m_a}\right)_{theoritical}}$

$\Theta_1 \equiv$ First Back Step Angled (deg.)

$\Theta_2 \equiv$ Second Back Step Angle (deg.)

$\Pi \equiv$ Static pressure losses $\equiv P_1 - P_2$

$\alpha \equiv$ Fuel Injector Angle (deg.)

INTRODUCTION

In order to achieve desired engine performance, emissions, and fuel economy in a scramjet engine, the air-to-fuel ratio must be precisely controlled under all operating conditions. Increasing air residence time in a combustor will increase the air-fuel mixing efficiency, increasing fuel consumption, releasing more energy, and consequently, producing more power [1]. For a scramjet engine application, a fuel-injection system must be designed efficiently in order to obtain high mixing and combustion efficiencies with an equivalence ratio close to one. This is because the available time for fuel-air mixing and combustion is limited to times on the order of 1 ms, which makes the maintenance of supersonic combustion a difficult task [2].

The fuel-injection process provides not only sufficient fuel-air mixing but also decelerates the air flow in the combustor [3]. When the fuel is injected into a supersonic freestream, it creates an interaction shockwave (between air and fuel) as shown in Figure 1.

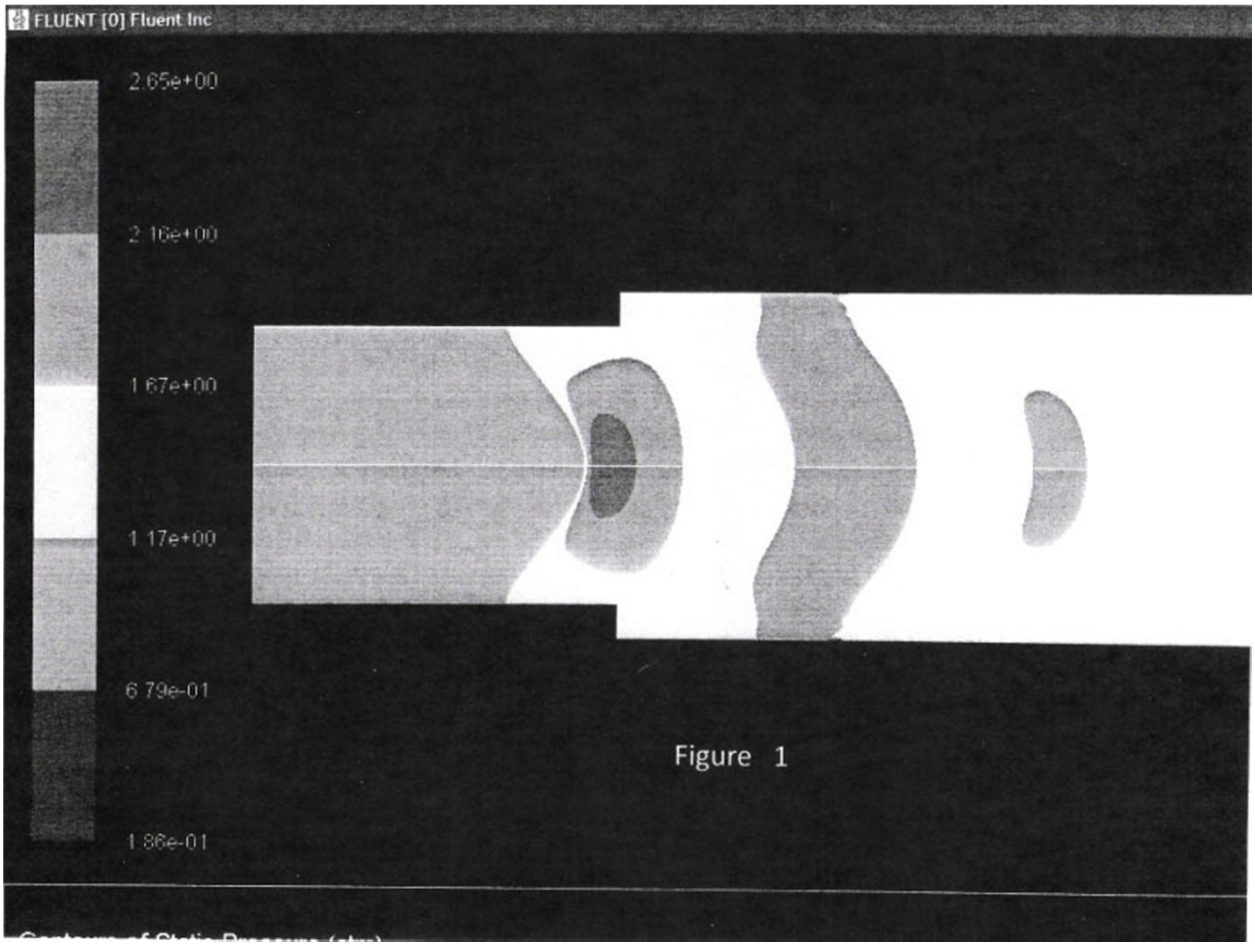


Figure 1

Several methods have been used in scramjet research to enhance fuel-air mixing and flame holding in order to achieve efficient combustion, including angled fuel injection into a supersonic air flow[4,5]. This method helps enhance the fuel-air mixing and reduces the total pressure losses, compared with normal injection, by generating a weaker bow shock. Moreover, transverse injection from the wall leads the fuel jet to interrupt the supersonic air flow and results in a bow shock in front of the injector [6]. The disadvantage of this method is increased pressure loss. Parallel air and fuel injection into the combustor was used by Jeong-Yeol Choi [7]. Raised (compression) and relieved (expansion) wall-mounted ramps were used by Abdel-Salam [8], fast fuel-air mixing was obtained by a relieved ramp.

The present work used normal fuel injection and utilizes the FLUENT CFD code to investigate the effect of variations of the fuel injector angle (α), diameter (d_f), and location (L) on the performance parameters, combustion efficiency (η_c), total pressure loss (Π), thrust (T), and equivalence ratio (Φ) in an axisymmetric scramjet engine.

MODEL SCRAMJET COMBUSTOR

An axisymmetric geometry was adopted in this investigation and appears with its dimensions in Figure 2.

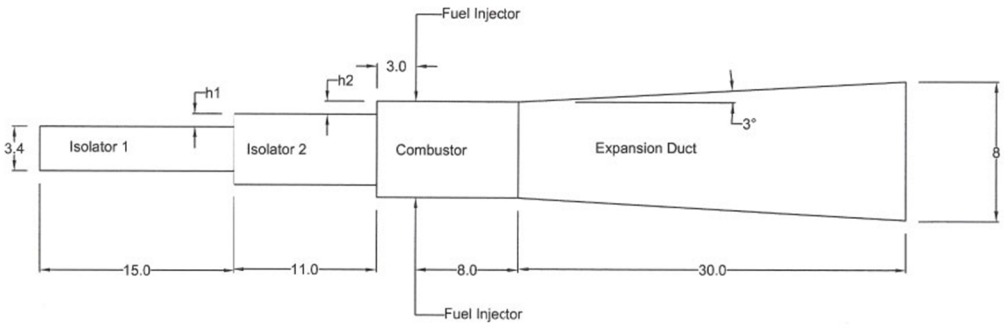


Figure 2 Dimensions in centimeter

It contains the following three sections:

(1) A constant cross-sectional area Isolator #1 with a height of 3.4 cm and length of 15 cm is connected to Isolator #2 which has a length of 11 cm and height of 3.9 cm. This provides upper and lower back steps with a height of 0.25 cm. At the outlet of isolator # 2 is the combustor which has a height of 4.9 cm and provides upper and lower back step heights of 0.5 cm. The two isolators are used in order to isolate the flow from its surroundings, stabilize the flow, and to

prevent interaction between the scramjet combustor and the inlet. The function of the two back-steps is to expand the flow and create shocks that reduce the flow velocity and maintain high mixing and combustion efficiencies [9]

(2) A combustor with a length of 11 cm which mixes and burns the fuel-air mixture. The fuel is injected normally into a high-speed air flow and has a strong influence on combustor performance. There are two 0.2-cm diameter fuel injectors located at the upper and lower surfaces of the combustor at X = 29 cm (measured from the air inlet). The injector's design must produce rapid mixing and combustion of the fuel-air mixture and accomplish minimization of the combustor's length and weight [10].

(3) The last section of the geometry is the square expansion duct which is 30 cm long with inlet dimensions of 4.9 cm x 4.9 cm. In order to prevent thermal choking (backfire), the duct divergence angle is 3^0 with the freestream direction (the outlet dimensions are 8 cm x 8 cm).

COMPUTATIONAL METHOD

A density-based solver (FLUENT) was used to solve the Navier-Stokes, continuity, energy, and species transport. These equations are simplified to algebraic equations and solved numerically to define the solution field within Fluent. Several iterations of the solution loop must be performed before a converged solution is obtained (typical iterations is in the 70,000 range). Two options exist for the density-based solver - implicit and explicit. The implicit solution approach is usually preferred to the explicit approach, which has a very strict limit on time-step size. The advantage of the implicit solver is that it performs much faster than the explicit solver and it is unconditionally stable with respect to time-step size. The solver uses a point implicit Gauss-Seidel/symmetric block (Gauss-Seidel/ILU) method to solve for variables.

To model the influence of turbulence, the K-ε model was used to characterize the turbulent flow. It includes two transport equations to represent the turbulent properties of the flow. The transport variables are turbulent kinetic energy (k), (which determines the energy in the turbulence) and the turbulent dissipation epsilon (ε) (determines the scale of the turbulence). The k- ε model has three users-defined options: Standard, RNG, and Realizable. The Realizable model was chosen, because it produces more accurate results for boundary layer flows than the Standard and RNG models.

The finite-rate/eddy-dissipation option was used in the species model because it is suitable for turbulent flow (high Reynolds number), fast chemical reactions as well as non-premixed air and fuel. The grid used to mesh the geometry in this study is composed of structured quadrilateral elements. Grid independence was examined to ensure that the present solutions are independent of the grid size. Grids with a number of cells between 40,000 and 80,000 were examined, and small variations in the values of pertinent variables were obtained. A grid with a mesh size of 70,000 cells was established as nominal. The Fluent computational algorithm used in this investigation was validated using the numerical and experimental results from [7], which is a study of a scramjet combustor. The boundary conditions for the air entering the isolator are: P_t =13.667atm , P_s = 1.12atm , and T_t = 1998K, In addition the fuel is injected at P_t = 2.12atm, P_s = 1.12atm and T_t = 294K, for the fuel injector diameter of 0.2 cm.

RESULTS and DISCUSSION

Two fuel injectors were tested in an axisymmetrical geometry. When the supersonic atmospheric air enters Isolator #2 and the combustor, it decelerates through a series of shocks that were created by the backsteps in order to reach the combustion chamber at an acceptable

level of uniform speed and to assure a stable combustion process. The sudden expansion of the flow at the first and second sets of backsteps creates shocks which can cause significant total pressure loss. In order to keep the combustion rate of the fuel constant, the pressure and temperature in the engine must also be constant [8]. The effect of variations to α, d_f, and L on η_c, Π, T, and Φ have been investigated as follows:

(1) $\boldsymbol{\alpha}$ was varied between 25^0 and 90^0 ,

(2) d_f was varied between 0.1 cm and 0.25 cm, and

(3) L was varied between 18 cm and 31 cm.

The following properties are kept constant for the present investigation:

1) Number of fuel injectors = 2,
2) $h_1 = 0.25$ cm, $h_2 = 0.5$ cm,
3) $\theta_1 = \theta_2 = 90^0$,
4) $L_1 = 15$ cm , $L_2 = 26$ cm,
5) Air inlet height = 3.4 cm,
6) Exhaust outlet height = 8.0 cm, and
7) An axisymmetry geometry (as defined within FLUENT) was adopted in this study.

EFFECT OF FUEL INJECTOR ANGLE ($\boldsymbol{\alpha}$) ON PERFORMANCE PARAMETERS

To determine the effect of variations to $\boldsymbol{\alpha}$ on η_c, Π , T and Φ, d_f and L were kept constant at 0.2 cm and 29 cm, respectively, while $\boldsymbol{\alpha}$ was varied between $25^{\boldsymbol{0}}$ and $90^{\boldsymbol{0}}$. The effect on η_c (as a function of X/d_f) for various values of $\boldsymbol{\alpha}$ ($25^{\boldsymbol{0}}$,$45^{\boldsymbol{0}}$,$65^{\boldsymbol{0}}$, and $90^{\boldsymbol{0}}$) is presented in Figure 3.

As α increases (angle measured from the X-axis clockwise), the strength of the interaction

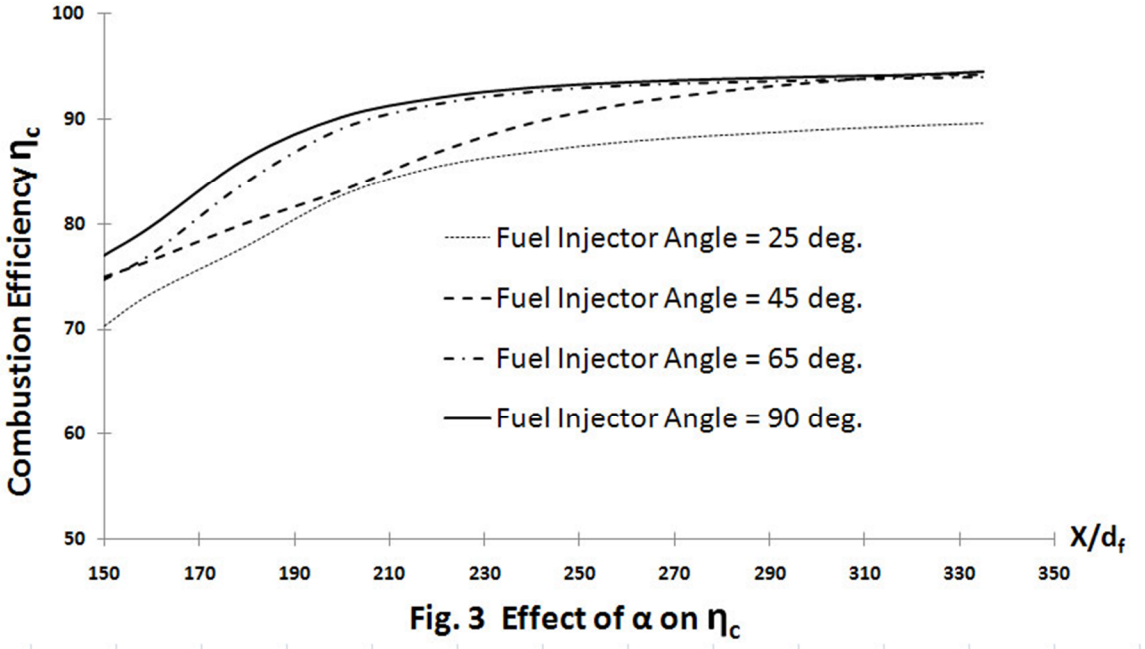


Fig. 3 Effect of α on η_c

between the fuel injector flow and the principal air stream increases, until the level of mixing reaches a maximum at $\boldsymbol{\alpha} = 90^{\mathbf{0}}$. The result is an increase in the strength of the resulting shockwave, a decrease in the average flow velocity, and an increase in the average pressure in the combustor. The mixing (air and fuel) efficiency increases at low air velocity, which in turn increases the extraction of the useful heat from the fuel. This is represented by the value of the equivalence ratio (Φ), as shown in Table1. If $\Phi > 1$, there is excess fuel in the fuel-oxidizer mixture and if $\Phi < 1$, there is excess air in the fuel-oxidizer mixture. This table also displays the relationship between α, η_c, Π, and T. As α approaches $90^{\mathbf{0}}$, T increases and Φ approaches unity, which is an indication of almost complete combustion (most of the fuel has reacted with the oxidizer). The smaller values of α produce Φ greater than unity, this is because of poor mixing between air with fuel and the weak shock resulted from the interaction between the fuel injector flow and the air stream flow.

Table 1. η_c, Π , T and Φ for various values of **α**

α (deg.)	η_c (%)	Π(atm)	T(N)	Φ
25	90.0	0.59	287	1.2
45	94.0	0.58	319	1.1
65	94.0	0.57	396	0.96
90	95.0	0.57	397	0.98

Figure 4 displays the pressure losses for different fuel injector angles. The minimum pressure losses occur when **α** = $90^{0.}$ This is due to the efficient shock wave that was created by the two

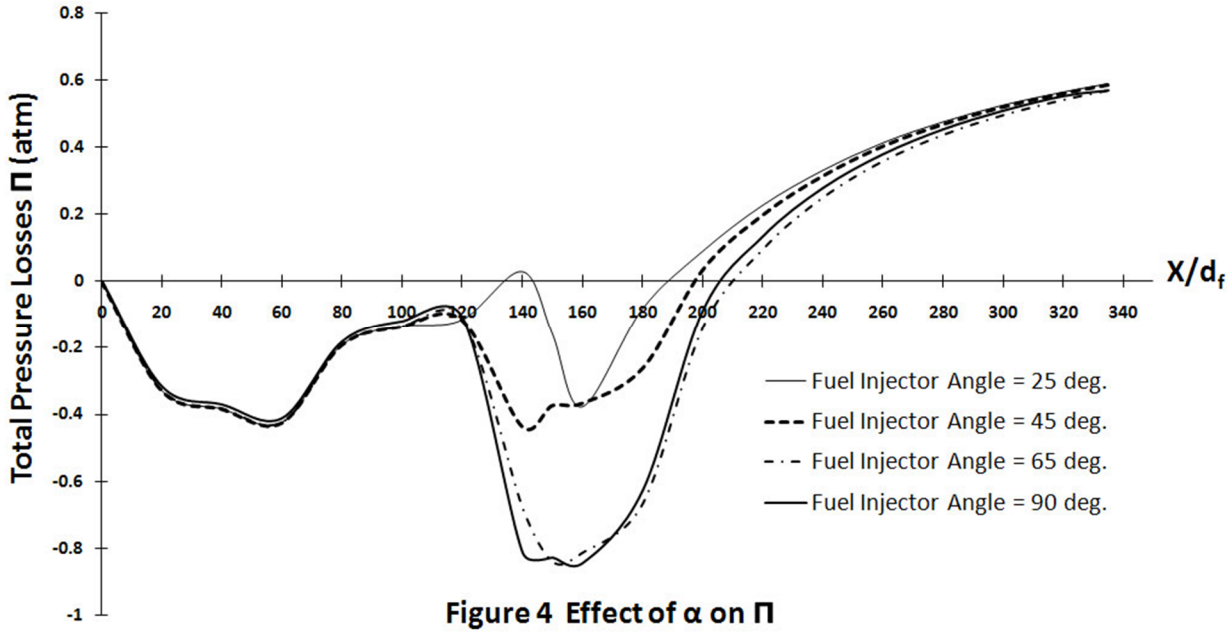


Figure 4 Effect of α on Π

backsteps and the interaction between the fuel injector flow and the freestream air flow, which causes freestream flow deceleration and leads to more desirable values of Φ and T. Fluent data and contour diagrams show that the interaction between injected fuel and the freestream air flow

creates maximum shock strength (with minimum mean velocity) when $\alpha = 90^0$ and minimum shock strength (with maximum mean velocity) when $\alpha = 25^0$. For the present steady-state adiabatic process, when supersonic air enters the 15 cm long constant cross section area Isolator 1, viscous effects cause the temperature to increase, and the flow properties to change along the duct.

For $0 \leq X/d_f \leq 60$, the pressure and the internal energy of the air flow increase due to an increase in temperature, which is caused by viscous effects. (For constant area compressible duct flow, with M>1, pressure increases in the freestream direction.)

For $60 \leq X/d_f \leq 120$, as the air flow reaches the first back step, the pressure decreases due to sudden enlargement of the duct (The sudden enlargement of the duct cross-sectional area creates expansion waves between the first and second backsteps that cause the pressure to decrease and velocity to increase).

For $120 \leq X/d_f \leq 140$, as the air reaches the second backstep, a shock wave is formed due to the second enlargement of the duct area, it cause a change in pressure and velocity of the air stream flow. The temperature and pressure of the air flow start to increase, due to the recirculating flow (behind the second backstep) which forces some of the hydrogen molecules to move upstream and react with air, except when $\alpha = 25^0$ the pressure continues to decrease because of the poor mixing efficiency, which leads to a poor η_c. (As the fuel injector angle increase, the mixing and combustion efficiency increase.)

For $140 \leq X/d_f \leq 160$, as the air reaches the fuel injectors, the interaction between the air stream flow and the fuel flow increases. As α increase the mixing and η_c increases, this causes the pressure, kinetic energy of the molecules and the Mach number of the mixture to increase

until it reaches maximum at $\boldsymbol{\alpha} = 90^0$.The minimum pressure was obtained when $\boldsymbol{\alpha} = 25^0$ because of the poor mixing efficiency. The inconsistent trend of the graphs in Π between the backstep and the inlet to the diverging section is due to the presence of shock waves and the flow interaction.

For $160 \leq X/d_f \leq 185$, as the highly energetic product mixture leaves the combustor and reaches the diverging section, the pressure begin to decrease due to the area enlargement of the exhaust duct, which cause the product mixture to expand.

For $185 \leq X/d_f \leq 335$, the pressure of the product mixture (for the four fuel injector angles) continues to decrease because of the diverging section, while the velocity reaches a maximum and becomes supersonic (M>2) at the outlet of the duct, (supersonic nozzle flow).

EFFECT OF FUEL INJECTOR DIAMETER($\mathbf{d_f}$)ON PERFORMANCE PARAMETERS

In order to investigate the effect of d_f on η_c , Π , T and Φ, L was kept constant at 29 cm (distance measured from air inlet), while d_f was varied between 0.1 cm and 0.25 cm. Based on results from a parametric study involving $\boldsymbol{\alpha}$, the most effective value of $\boldsymbol{\alpha}$ was chosen (90^0) and used in this investigation.

Figure 5 displays the effect of variations to d_f on η_c. The optimum η_c was obtained

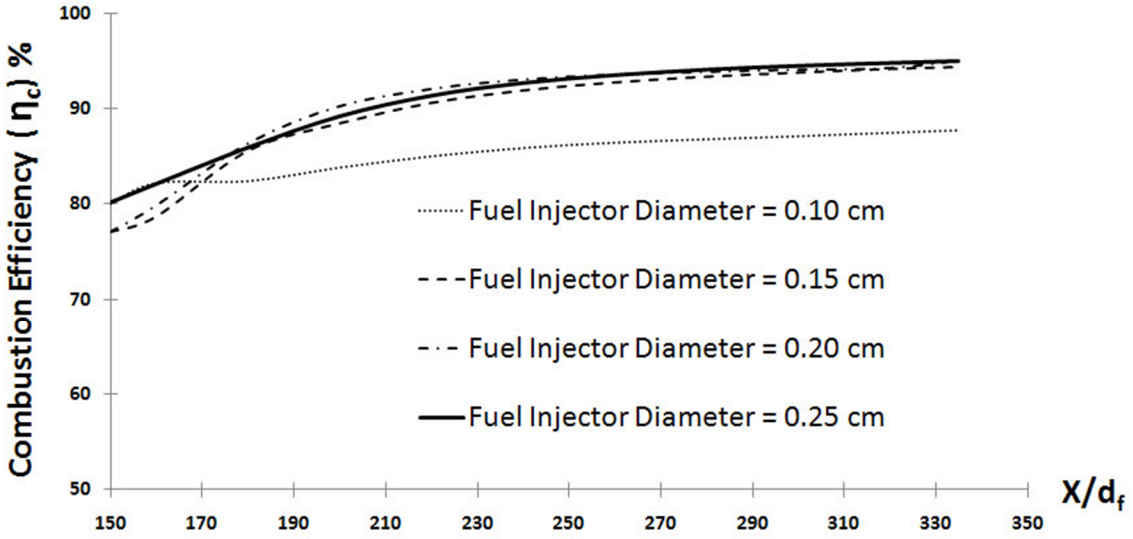


Fig. 5 Effect of d_f on η_c

when d_f = 0.2 cm, which is an indication of efficient mixing and combustion shown by the value of Φ being close to unity. The other values of fuel injector diameters (0.1, 0.15, 0.25) results in Φ greater than unity which is an indication of excess fuel in the product mixture. As the fuel injector diameter cross-section area decreases with constant mass flow rate, the fuel velocity increases, which decreases the mixing efficiency and leads to Φ greater than unity and a decrease in T, as shown in Table 2.

Table 2. η_c, Π , T and Φ for various values of d_f

d_f (cm)	η_c (%)	Π(atm)	T(N)	Φ
0.1	88	0.62	280	1.3
0.15	94	0.61	298	1.2
0.2	95	0.60	397	0.98
0.25	95	0.60	321	1.1

Figure 6 displays the effect of variations to d_f on Π. For $0 \leq X/d_f \leq 120$. Figure 6 shows similar trends as seen in Figure 4.

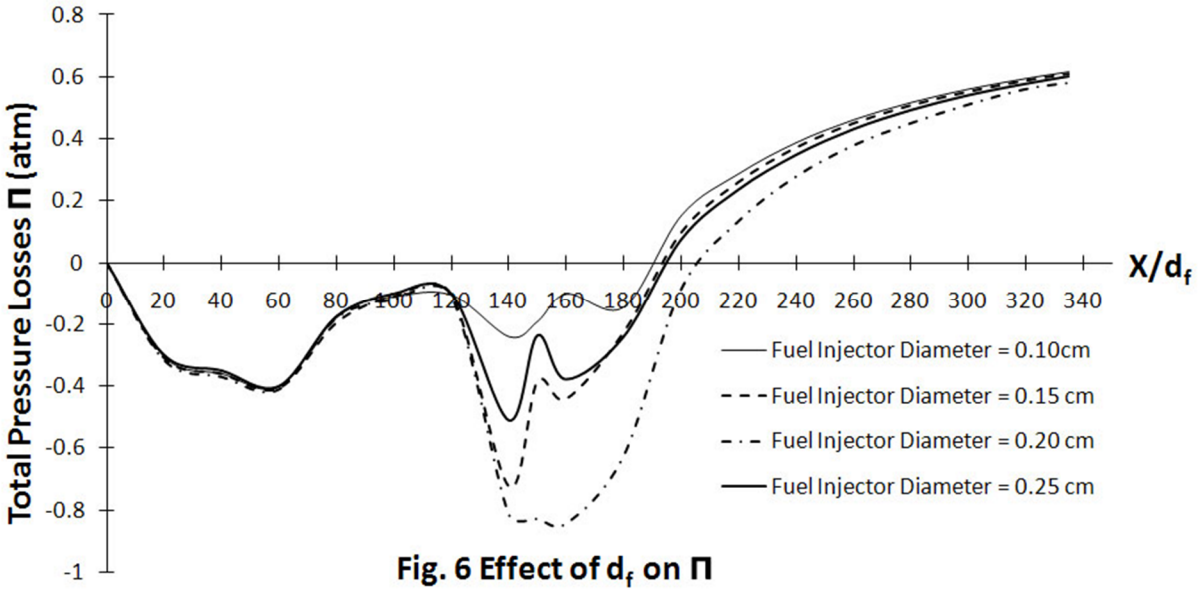


Fig. 6 Effect of d_f on Π

For $120 \leq X/d_f \leq 140$, as air reaches the second backstep, the temperature and pressure start to increase, because the vortices (behind the second backstep) force some of the hydrogen

molecules to move upstream and react with air. The maximum pressure was obtained when d_f = 0.2 cm (this is due to the efficient shock between the air flow and the fuel flow which leads to a better mixing efficiency), while the minimum pressure was obtained when d_f = 0.1 cm, this is because of a poor mixing efficiency (as the fuel injector diameter decrease with constant mass flow rate, its velocity increase and therefore the mixing efficiency decrease).

For $140 \leq X/d_f \leq 160$, as air enters the combustor, the pressure continues to increase for d_f = 0.2 cm, due to its efficient mixing and combustion efficiency. For d_f = 0.1 cm, the pressure decreases due to its high velocity, which leads to poor mixing efficiency. The inconsistent trend of the graphs in Π in the combustor area is due to the presence of shock waves and the changes in the fuel flow speed.

For $160 \leq X/d_f \leq 185$, the velocity increases and the pressure decreases as the flow leaves the combustor and approaches location X/d_f =185 (inlet to the expansion duct), except for d_f = 0.1 cm before it reaches location X/d_f =185 there is a minor increase in pressure and this due to a decrease in its velocity after a short distance from its injection. For $185 \leq X/d_f \leq 335$, the flow behavior is similar to that described in the previous section.

EFFECT OF FUEL INJECTOR LOCATION (L) ON PERFORMANCE PARAMETERS

In order to investigate the effect of L on η_c, Π, T and Φ, L was varied between 18 cm and 31 cm. Based on results from a parametric study involving α and d_f, the most effective values of α and d_f were found to be 90^0 and 0.2 cm respectively. These values were held constant during this phase of the study. Numerical calculations were performed in order to understand the correlation between the fuel injector location and combustion characteristics of the fuel

hydrogen. Fuel was injected in different location behind the first and second backstep, as is indicated in Table 3.

Table 3. η_c, Π , T and Φ for various values of L

L (cm)	η_c (%)	Π(atm)	T(N)	Φ
18	99	0.63	296	1.3
19	82	0.65	277	1.3
20	99	0.63	263	1.4
29	95	0.57	397	0.98
31	94	0.59	370	1.1

In the region between the first and second backstep (15 cm to 26 cm), the maximum T was obtained when L was at 18 cm from the air inlet. This is due the sudden expansion of the air flow behind the first backstep which create a shock waves that reduces the air flow velocity and resulted into a maximum T. However its Φ is greater than one which is an indication of incomplete combustion. The most efficient T with reasonable Φ were obtained when L was at 3 cm behind the second backstep (29 cm from the air inlet), and this is due the reduced air flow velocity by the shock waves which was created by the two backsteps.

Figure 7 display the effect of variation to L on the η_c, the maximum η_c was obtained when L was

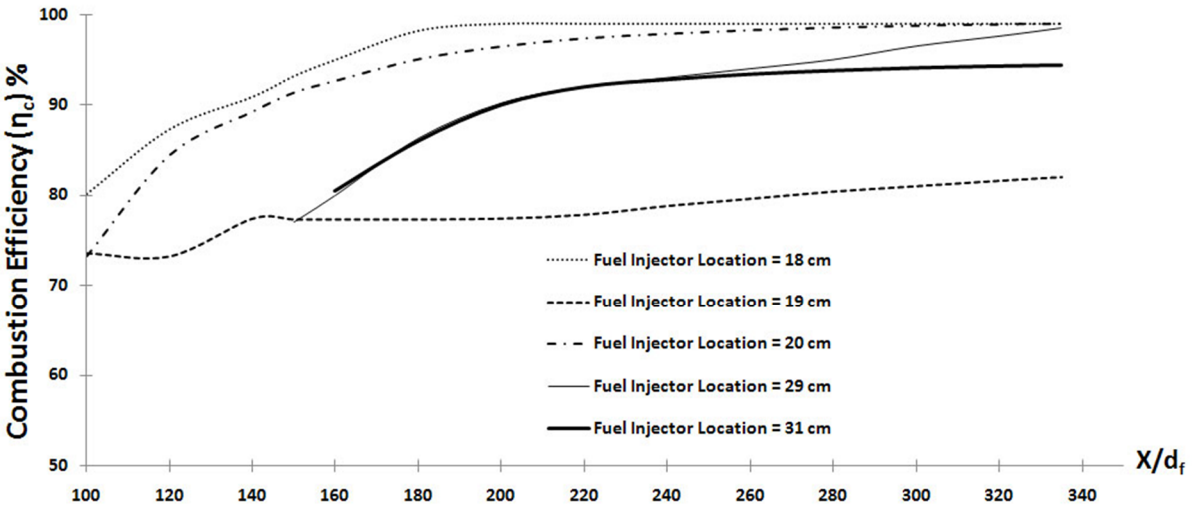


Fig. 7 Effect of L on η_c

at 18 cm and 29 cm from the air inlet. However the Φ is higher than 1 when L was at 18 cm which is an indication of excess fuel was found in the exhaust products, therefore more air was needed to complete the combustion. The optimum η_c was obtained when L was at 29 cm from the air inlet (3 cm behind the second backstep) as is indicated by the value of Φ being almost close to unity**.** This is due the reduced air flow velocity by the shock waves that were created by the sudden expansion of air between the first and second backstep, which leads to efficient mixing and combustion between air and fuel.

Figure 8 displays the effect of variations to L on Π. For $0 \leq X/d_f \leq 60$, as supersonic freestream air flow enters the 15-cm long constant cross sectional area duct, viscous effects cause the air flow temperature to increase, this lead to increase in the pressure of the air flow and a decrease in its velocity.

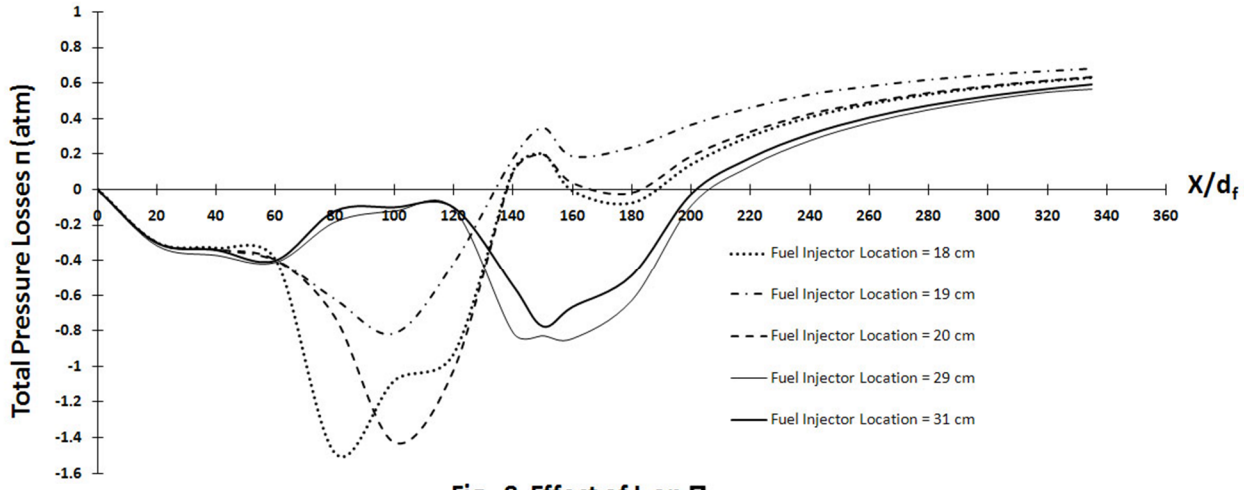


Fig. 8 Effect of L on Π

For $60 \leq X/d_f \leq 120$, between the first and second backstep, the fuel was injected at three different locations (18 cm, 19 cm, and 20 cm distance measured from the air inlet). The combustion of the fuel at the previous three locations**,** caused the pressure to increase, but the interference of the combustion products with the expansion waves (created by the sudden air flow expansion which caused by the first backstep) causes the pressure to decrease. The other two fuel injectors (located at distance 29 cm and 31 cm distance measured from the air inlet) show a decrease in the air flow pressure, this is due to the expansion wave created by the first backstep, as mentioned before.

For $120 \leq X/d_f \leq 160$, as the products of combustion for the first three locations of fuel injectors (18 cm,19 cm, and 20 cm) reaches the second backstep, the pressure continues to decrease, but there is small increase in pressure (probably due to the compression waves created by the second backstep). While the pressure increases for the other two locations of the fuel injectors (29 cm and 31 cm), due to the combustion process of fuel with air.

For $160 \leq X/d_f \leq 185$**,** as the products of combustion leave the combustor, the kinetic energy of the mixture molecules increased (this cause increase in the flow velocity), and the pressure begin to decrease as it approaches location $X/d_f = 185$**.**

For $185 \leq X/d_f \leq 335$, the pressure of the product mixture (for the five fuel injector locations) continues to decrease because of the diverging section angle increase, while the velocity reaches a maximum and becomes supersonic (M>2) at the outlet of the duct, (supersonic nozzle flow). The highest pressure losses with Φ greater than 1, was obtained when the fuel injectors were located behind the first backstep, and this is because one backstep was not efficient to reduce the air velocity before it reaches the fuel injector. The lowest pressure losses (with reasonable Φ and optimum T) were obtained when the fuel injector was at 29 cm and 31 cm, as shown in Table 3. This is because of the two backsteps that expand the air flow and creates the shock wave which reduces the air flow velocity before the air reaches the combustor, and lead to a maximum η_c. Therefore In order to have efficient mixing and maximum thrust, two backsteps are required before the combustion process begin.

CONCLUSION

Fuel injector effects were investigated numerically by changing fuel injector angle, diameter and location with particular emphasis on the combustion efficiency, total pressure losses, thrust, and equivalence ratio of a supersonic hydrogen-air jet flame in a model scramjet combustor. The main conclusions of the present study are as follows:

1. The strength of the shock wave increases as $\boldsymbol{\alpha}$ increases, until it reaches maximum at 90^0, it has strong effect on T and Φ. It causes a decrease in the average flow velocity, and an increase in the average Pressure. T increases significantly by 28 % when $\boldsymbol{\alpha} = 90^0$.

2. The fuel injector diameter has an effect on the mixing efficiency. As the fuel injector diameter decreases with constant mass flow rate, the fuel velocity increases, which decreases the mixing efficiency and leads to Φ greater than unity and a decrease in T as shown in Table 2. The efficient fuel injector diameter was found to be = 0.2 cm.
3. The location of the fuel injector behind the second backstep is more efficient than its location behind the first backstep, because of the air velocity was reduced by the shock waves that were created by the two backsteps.

REFERENCES

1. J.P. Drummond, G.S Diskin, and A.D. Cutler, “Fuel-Air Mixing and Combustion in Scramjets,” AIAA Paper, 2002-3878, 2002.
2. W. Xianyu, L. Xiaoshan, D. Meng, L. Weidong and W. Zhenguo, “Experimental Study on Effects of Fuel Injection on Scramjet Combustor Performance,” *Chinese Journal of Aeronautics*, vol. 20, no. 6, pp. 488-494, 2007.
3. Anonymous. “ZACM4921 HYPERSONICS, Literature Review, “Scramjet Fuel Injection Methods,” Internet: http://seit.unsw.adfa.edu.au/ojs/index.php/Hypersonics/article/view/15/6, June 4, 2007 [November 20, 2010].
4. Rodney D.W.Bowersox, “Turbulent Flow Structure Characterization of Angled Injection into a Supersonic Crossflow,” *Journal of Spacecraft and Rockets*, Vol. 34, No. 2, pp. 205-213, 1997.
5. E. Jeong, S. O’Byrne, I-S. Jeung and A.F.P. Houwing, “Investigation of Supersonic Combustion with Angled Injection in a Cavity-Based Combustor,” *Journal of Propulsion and Power*, vol. 24, no. 6, pp. 1258-1268, 2008.
6. K.Sundararaj and S. Dhandapani, “Numerical Simulation of Staged Transverse Injection of H2 Fuel in a Ducted Supersonic Air Stream with SST k-ω Turbulence Model,” *International Journal of Dynamics and Fluids*, vol. 2, no. 2, pp. 245-262, 2006.
7. J-H. Kim, H. Huh, Y. Yoon, I-S. Jeung, and Jeong-Yeol Choi, “Effects of Shock Waves on a Supersonic Hydrogen-Air Jet Flame in a Model Scramjet Combustor,” presented at the Second Asia-Pacific Conference on Combustion, Tainan, Taiwan, 1999.
8. T. M. Abdel-Salam, “Numerical Studies on Supersonic Mixing and Combustion Phenomena,” Ph.D. dissertation, Old Dominion University, Virginia, 2003.

9. Kuratani, N., Ikeda, Y., Nakajima, T., Tomioka, S., and Mitani, T., "Mixing Characteristics of Normal Injection into a Supersonic Backward-Facing Step Flow Measured with PIV," AIAA Paper 2002-237, 2002.

10. C.J. Steffen, Jr., "Fuel Injector Design Optimization for an Annular Scramjet Geometry," AIAA Paper, 2003-0651, 2003.

Theoretical Design and Application of a Robotic Hand

Author: Daniel Biller, Cheng Lin

Robotic Hand

I. Introduction

The robotic hand design was a fun and intriguing design throughout its process of being designed. Through Autodesk® Inventor® the hand was designed, visualized, and stimulated. Autodesk® Inventor® is a computer program used to design, visualize, and check for obstructions of objects [3]. The project required effort and persistence in attempting to design each part for the project. The main component in designing the project was that it had to be made with only currently available products in the market, and can be fabricated to the projects needs.

II. Problem/Motivation

The motivation behind the project was to further my expertise in designing and robotics, which led into designing a robotic hand. The robotic hand has to be relatively close to the size of a normal human hand, with the exception of only three fingers instead of five. The robotic hand consists of three fingers with three motors for each joint of the finger, just like a human finger. Although not indicated on the design due to software limitations, the cover on the finger is supposed to have rubber to cover the rest of the exposed areas of each joint. The motor used in the project is ordered from Robokits World, the 10RPM 12V DC Motor with Gearbox with alteration to the shaft length [1]. The torque provided by this motor is 12 kg*cm [1].

The purpose for designing the robotic hand is to be created so that it may be used for handling dangerous items in the workplace such as radioactive material, extremely delicate material with an extreme temperature difference from room temperature, and possibly a prosthetic with alterations. The use of the robotic hand will give assistance to daily human life.

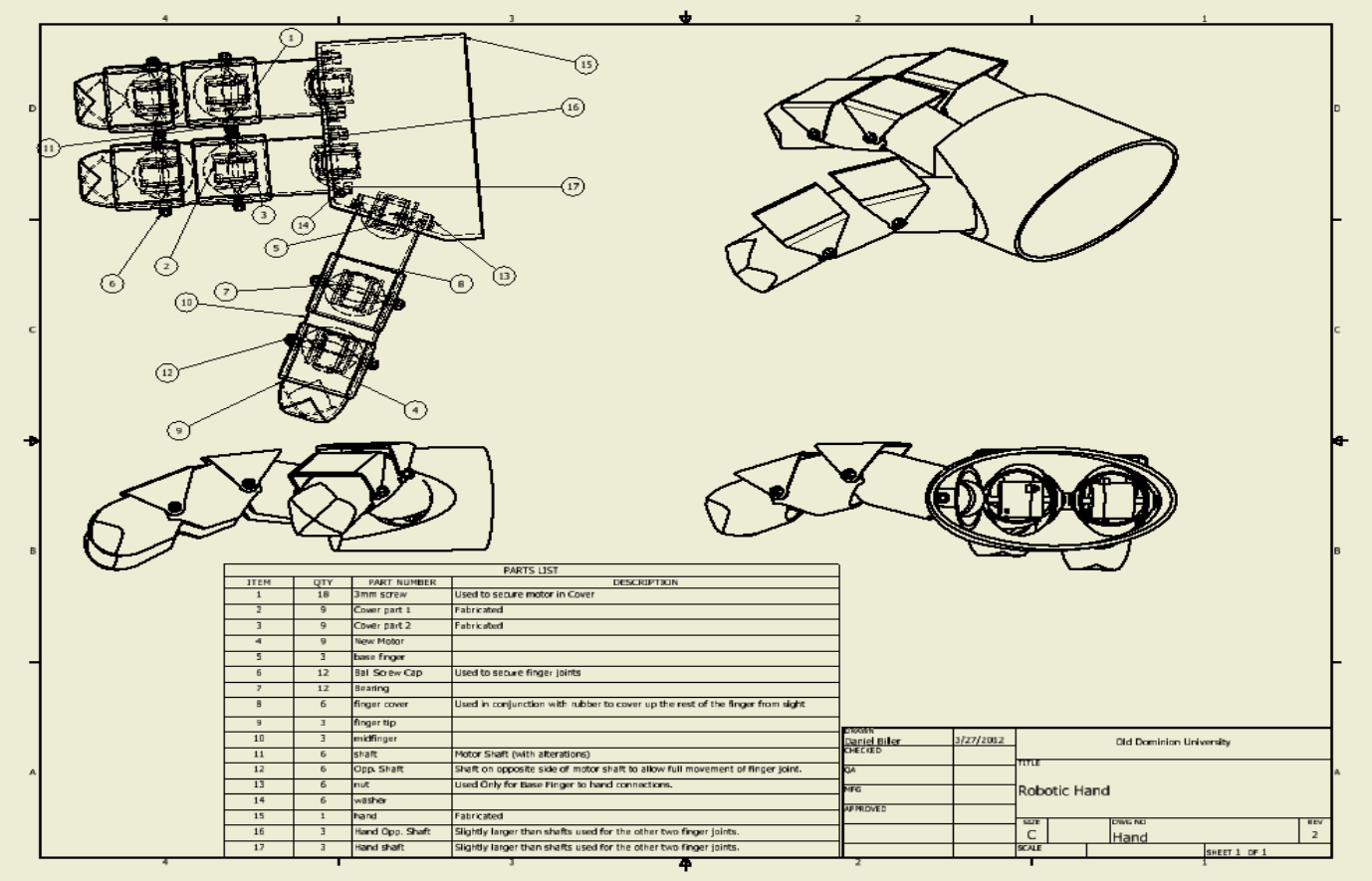
III. Method

The Autodesk® Inventor® was used to design and animate the robotic hand project. Each of the parts was made with cost and size in mind to make accessibility easy. The shaft on the motor and the spare shaft have screw threads at the end to constrain the rest of the hand and the motor in place. The cover for the motor, the finger parts, and the hand are all fabricated from plastic to ensure light weight and prevent possible metal shavings from the small enclosed areas so the enclosed areas are not clogged or damaged. A small bearing is used to translate the rotation of the motor to the corresponding finger part so that the finger may move.

The reason for using three fingers was to continue with the idea of a cost effective hand design. The primary purpose for the motor used is to stay within the cheaper materials while still using a motor that had enough torque and moved slow enough for manual controls. The cover for the motor is to allow another shaft to be equipped at the same position as the motor and cover

up the connection of the power to the motors. The finger cover shown would show rubber attached to it and the connecting parts of the finger, unfortunately due to software limitations, it is not shown. The purpose of the finger cover with rubber is to prevent the inner parts from being exposed and possibly prevented from moving if something got stuck in between the finger parts.

Although not indicated by the drawings, there is a pressure sensor, the P6000 Pressure Sensor, to be placed at the fingertip of each finger to ease off and stop movement to prevent crushing the item that is attempted to be held [2].



PARTS LIST

ITEM	QTY	PART NUMBER	DESCRIPTION
1	18	3mm screw	Used to secure motor in Cover
2	9	Cover part 1	Fabricated
3	9	Cover part 2	Fabricated
4	9	New Motor	
5	3	Base finger	
6	12	Ball Screw Cap	Used to secure finger joints
7	12	Bearing	
8	6	finger cover	Used in conjunction with rubber to cover up the rest of the finger from slight
9	3	finger tip	
10	3	midfinger	
11	6	shaft	Motor Shaft (with alterations)
12	6	Opp. Shaft	Shaft on opposite side of motor shaft to allow full movement of finger joint.
13	6	nut	Used Only for Base Finger to hand connections.
14	6	washer	
15	1	hand	Fabricated
16	3	Hand Opp. Shaft	Slightly larger than shafts used for the other two finger joints.
17	3	Hand shaft	Slightly larger than shafts used for the other two finger joints.

DRAWN: Daniel Biller 3/27/2012
CHECKED
QA
MFG
APPROVED

Old Dominion University

TITLE: Robotic Hand

SIZE: C | DWG NO: Hand | REV: 2

SCALE | SHEET 1 OF 1

Fig 1. Hand Assembled Shown views are top, front, and right side, and isometric views. Top view is used to show individual parts of the hand.

IV. Conclusion

The constrained area of the finger to move prevents a fist from truly forming, and only can move so far in without problems occurring. The motor that is currently used is slow enough for use so that the hand can be operated properly, although there may need more torque required to ensure true efficient movement in gripping items. Without expanding the size of the hand well beyond 1.5x the size of a human adult hand, which is about 10 cm for the hand and the finger length of 8 cm, the fingers cannot fully close using this design, but the design does allow rather large items to be grasped. Although this design has its flaws, it also has the greatest advantage of being lightweight, cheap, and rather easy to acquire parts.

V. Reference

1. 2008. 10RPM 12V DC Motor with Gearbox [Online]. http://robokitsworld.com/index.php?main_page=product_info&products_id=46
2. 2012. P6000 Miniature Pressure Sensor [Online]. http://www.kavlico.com/catalog/P6000.php?section=products&division=ind&category=miniature
3. Daniel T. Banach, Travis Jones, and Alan J. Kalameja, *Autodesk ® Inventor® 2011 Essentials Plus.* Clifton Park, New York: Delmar, 2011.

Test Environment for Analyzing Event Based Simulation Executives

Spencer Lane and James F. Leathrum, Jr.

***Abstract*—A project is currently underway to develop a test environment to test simulation executives. The system architecture is loosely based off of the Clemson Automated Testing System (CATS) for testing open software standards and involves automatic test generation. A Script Definition Language (SDL) was created to support automatic test generation. SDL Scripts can be created that, when executed by the test executive, simulate a wide range of application behaviors. This paper describes the architecture and design of this test environment.**

***Index Terms*—Automatic Test System, Discrete Event Simulation, Simulation Executive, Test Environment**

I. Introduction

One of the most important aspects of software engineering is testing. A program needs to be thoroughly tested and relatively bug free before it is shipped to potential customers. In the world of simulation, this process of verification and validation can be even more important as decisions will likely be made based on the results. In cases where the software to be tested was designed to meet a standard or set of requirements for which multiple implementations are envisioned, an automated software test environment facilitates the testing process. The test environment is developed to test the properties of the standard or requirements, ignoring implementation issues. This requires a cleanly defined interface to the software under test so that automated calls can be made and responses tracked.

Discrete event simulation software is typically driven by a structure called a simulation executive. The simulation executive maintains the simulation clock and a list of pending events which it executes in simulation time order. Simulation executives are paired with application software to generate the actual simulation. Each application is made up of multiple simulation objects which each contain a number of events. The application tells the simulation executive which events to schedule and when to schedule them.

Because of its role as the driving force behind simulation software, the simulation executive needs to be thoroughly and exhaustively tested using a wide variety of application behaviors. Testing with multiple applications is not always feasible due to the development costs of making new applications or modifying existing ones to work with the new simulation executive. A test environment however, could be designed that would simulate a wide range of application behaviors allowing for rapid, exhaustive testing of a new simulation executive. This test environment must support both functional and performance testing. In order to test functionality, the test environment would need to be able to generate representative calls to the simulation executive as well as accept responses which are then verified against expected responses. This paper presents the design of a system that would meet these needs.

II. Related Work

Test environments are generally found when multiple implementations of a given standard are expected, and creation of a test environment enables efficient accrediting of the implementations. Clemson University and the U.S. Navy partnered to create an automated testing system for use in testing operating systems. The Clemson Automated Testing System (CATS) was designed for testing systems throughout their lifecycle [1]. It is a requirements based testing system that uses what is essentially a script to define the requirements for a test and then automatically generate, run, execute, and analyze the results of that test [2]. These test specification files provide a way of documenting tests in a clear and consistent manner as well as providing a means of rapidly creating a large test suite. CATS provides a loose basis for the architecture discussed in this paper.

A similar test environment specific to the field of simulation was developed by the Mitre Corporation to test the Run Time Infrastructure (RTI) portion of the High Level Architecture (HLA). This system was designed to thoroughly

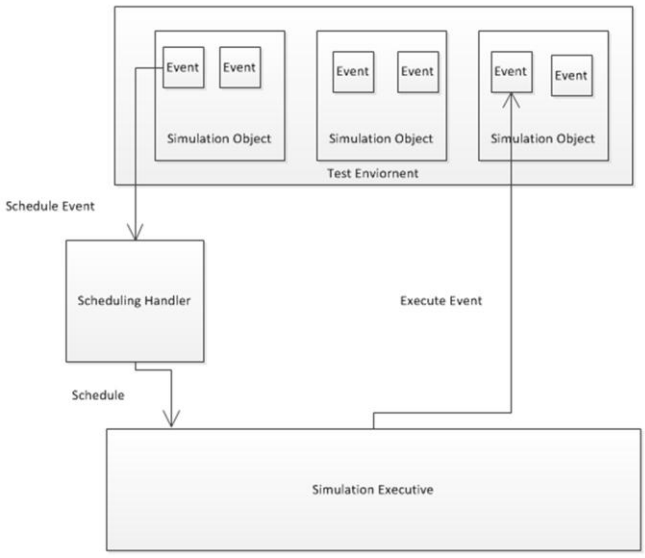


Fig. 1: Simulation Executive Test Environment System Architecture

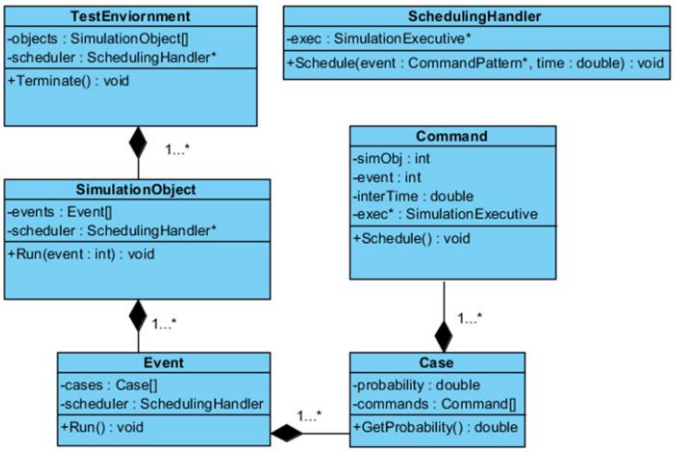


Fig. 2: Test Environment and Scheduling Handler Class Diagram

test and verify RTIs with little to no human interaction. The test controller in this environment executes multiple test scripts and verifies the output. One of the capabilities that sets this test system apart from others is that it is designed to automatically determine the sequence of events that lead to a failure state [3]. This ability to determine why a test failed aids in the development process by allowing for rapid bug finding.

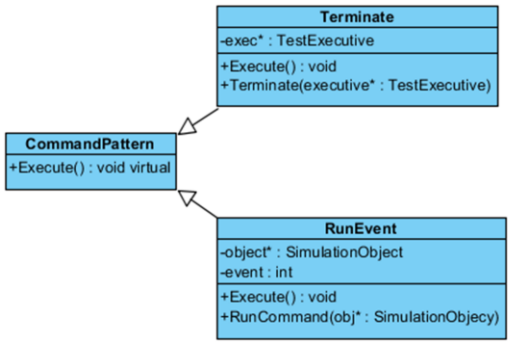


Fig. 3: Command Patterns Class Diagram

III. Test Environment Architecture

The test environment needs to be able to generate representative events, and call a scheduler method on the simulation executive that is under test. Because the test environment needs to work with a wide variety of simulation executives, an interface needs to be built between the test environment and the executive under test. This interface would have to be modified to fit with the application programming interface of the system under test.

The test environment itself needs to be able to instantiate multiple simulation objects, each with a set of potential events implemented by methods. The events are purely representational, containing only a list of other events

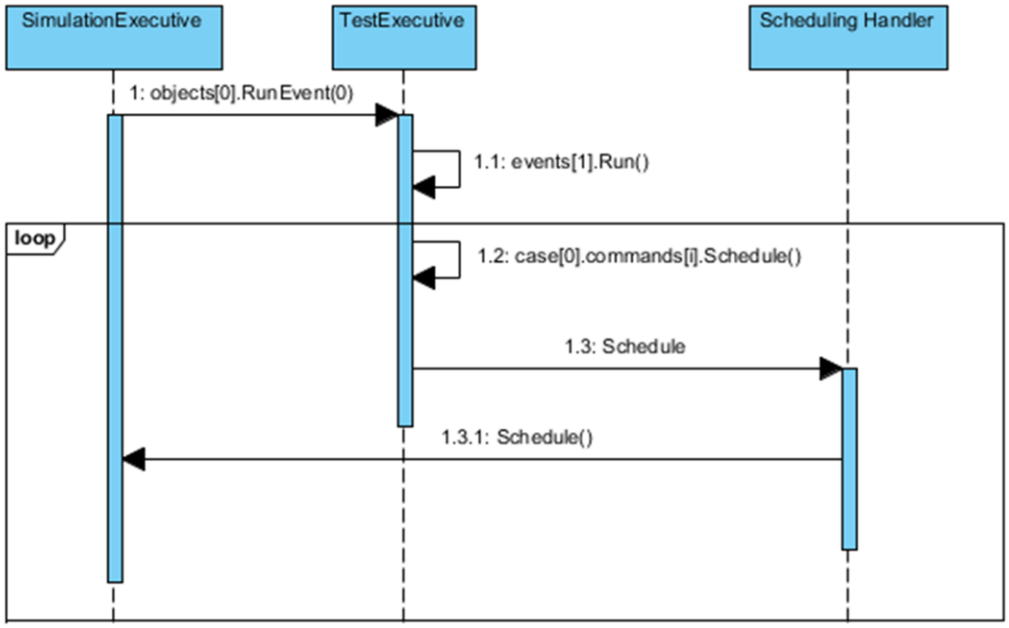


Fig. 4: Event Execution Sequence Diagram

to schedule. In order to introduce a stochastic process for generating the new events, each event has a set of cases associated with it. Each case has a probability and a list of events to schedule if that case is selected. This list stores a reference to the events to be executed as well as the time with which to schedule those events.

During execution of an event, that event schedules future events at a computed simulation time. The scheduled time is defined either deterministically through a script, or stochastically based on a provided distribution. An instance of the event for scheduling is created with an appropriate parameter list. The schedule method on the interface then puts the event in the right format for the simulation executive and causes that event to be scheduled. The simulation executive should then evoke the event at the correct simulation time. The execution of the event is handled by the simulation object that contains that event. At a high level, this leads to the system architecture seen in Figure 1. Figure 2 shows a UML representation of the test environment and scheduling handler.

In order to encapsulate events for later execution, a design pattern called a command pattern is used. Command patterns are data structures that contain all of the information necessary to make a later call to a function [4]. In this case there are two command patterns that are used to implement events: run event and terminate. The run event command pattern stores a reference to the simulation object that the event is on as well as a reference to the actual event to be scheduled. When the run event command pattern is executed, it calls the Run method on the specified simulation object with the parameter as the specified event. The terminate command pattern stores a reference to the test environment. When the terminate command pattern is executed, it calls the terminate method on the specified test environment. Figure 3 shows a UML representation of the Command Pattern base class as well as the specific commands for this system.

When a non-terminating event is executed, the Run method is called on the specific simulation object. This method then calls the Run method for the specified event. This method selects a case based on its probability and iterates through its list of commands calling the Schedule method for each. This method calls the Schedule method on the scheduling handler which then causes the event to be scheduled by the simulation executive. Figure 4 illustrates a sequence diagram for this process.

IV. Test Environment Design

The implementation of the test environment architecture involves a Script Definition Language (SDL) and the software implementation of the architecture. The SDL includes all of the pertinent data to simulate a wide range of application behaviors. The format of the SDL employed in this work is provided in Figure 5. Note that each case in the script has a probability associated with it. The sum of the probabilities of each case for a particular event must add to 1, ensuring that one of the cases is selected.

```
Number of Simulation Objects N (Int)
InitialEvents {
   SimulationObject# (Int) Event# (Int) Time (double)
   … … …
}
SimObject 0 {
    Number of Events I (Int)
    Event  0 {
        Number of Cases J (Int)
        Case 0 {
           Probability (double)
           Number of Commands (Int)
           SimulationObject# (Int) Event# (Int) Time (double)
           … … …
        }
        …
        Case J – 1 {
            Probability (double)
            Number of Commands (Int)
            SimulationObject# (Int) Event# (Int) Time (double)
            … … …
         }
     }
     …
     Event I -1 {
        Number of Cases K (Int)
        Case 0 {
           … … …
         }
    }
}
…
SimObject N – 1 {
    Number of Events L (Int)
    Event 0 {
    …
    }
    …
    Event L – 1 {
            terminate
    }
}
```

Fig. 5: Script Definition Language Outline

In the SDL, time can be defined in one of two ways. If the time is represented with just a double value, that value is deterministic. If the time is defined with a distribution name and then the parameters for that distribution, the time will be generated stochastically when the event is executed. The times in the Initial Events section represent the actual simulation time when these events will be executed. The times for each command represent the amount of time in the future that event will be scheduled. If an event is the terminating event, it will simply contain the word terminate instead of any cases.

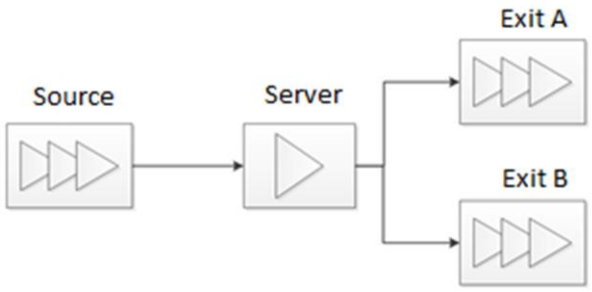


Fig. 6: Sample System

Scripts of this nature can be used to model a wide range of application behaviors. For instance, the system shown in Figure 6 consists of a single server with a single queue. Entities enter the system at a rate defined by an Exponential distribution with a mean of 0.5 minutes. Entities are served at a rate defined by a Normal Distribution with a mean of 2 minutes and a standard deviation of 0.2. Once an entity is served, it follows one of two paths out of the system. The first path, labeled A, takes 3 minutes to traverse. The second path, labeled B, takes 5 minutes to traverse. Path A is taken by 75% of the entities and Path B is taken by 25% of the entities. After following one of the paths, each entity leaves the system. This simple system can be simulated by using the script in Figure 7.

This script, when run through an interpreter, would create a new test environment with 2 simulation objects. Simulation object 1 represents the server and simulation object 2 represents the paths. On simulation object 1, event 0 represents the arrival process, event 1 represents the serving process and event 3 is the terminating event. On simulation object 2, event 0 represents leaving by path A and event 1 represents leaving by path B. This system is only an example. This language could be used to define a test environment that simulates the behavior of a wide variety of applications. Any discrete event application behavior could theoretically be modeled.

The SDL is used to specify a representative test that will be run on the simulation executive. An interpreter then reads this language and puts it into a format that is executable by the test environment.

```
2
Initial Events {
    0 0 0
    0 1 2
    0 2 480
}
Simulation Object 0 {
    3
    Event 0 {
        1
        Case 0 {
            1
            1
          0 0 Exponential 0.5
        }
    }
    Event 1 {
        2
        Case 0 {
            0.75
            2
          1 0 3
            0 1 Normal 2 0.2
        }
        Case 1 {
            0.25
            2
          1 0 5
            0 1 Normal 2 0.2
        }
    }
    Event 2 {
        terminate
    }
}
Simulation Object 1 {
    2
    Event 0 {
        0
    }
    Event 1{
        0
    }
}
```

Fig. 7: Sample SDL Script

V. Future Work

The test environment presented here is only a theoretical design. A full implementation is under development but is not yet complete. This implementation, when finalized, will contain all of the functionality defined in this paper.

In addition to this initial implementation, a way to track results will be designed and implemented. One way to do this would be to store a list of expected event executions as well as a list of actual calls. The Run method could be modified to include the capability to store the expected results and the Run Event method could be modified to include the ability to track actual results. The terminate method would then compare the two lists to determine whether the events were executed in the correct order at the correct times, identifying offending event executions.

Another requirement that is currently missing from this design is the ability to test performance. Obviously performance is an important aspect of software. If a program runs correctly but does so slowly, it might not be useful. A method of capturing and analyzing performance data based on performance requirements will be designed and implemented.

VI. Conclusions

When this test environment is fully implemented, it will be able to emulate a wide range of application behaviors allowing for thoroughly test the functionality of simulation executives. The resulting system creates a simulation of the expected behavior of simulation applications from the viewpoint of the simulation executive – effectively a simulation of a simulation. With the addition of the ability to test the performance of simulation executives, this test environment will be capable of aiding the development process of discrete event simulation executive software.

References

[1] J.F. Leathrum and K.A. Liburdy, "Automated testing of open software standards," *International Test Conference Proceedings*, 17-21 Oct 1993, pp.854-858,

[2] J.F. Leathrum and K.A. Liburdy, "A formal approach to requirements based testing in open systems standards," *Proceedings of the Second International Conference on Requirements Engineering*, 15-18 Apr 1996 pp.94-100,

[3] J. Tufarolo, J. Nielsen, S. Symington, R. Weatherly, A. Wilson, J. Ivers, T.C. Hyon, "Automated distributed system testing: designing an RTI verification system," Winter *Simulation Conference Proceedings* , 1999, vol.2, pp.1094-1102

[4] Christopher G. Lasater, *Design Patterns*, Jones & Bartlett Publishers, 2006

INITIALIZATION OF A LARGE-SCALE SIMULATION USING MINED DATA

Olcay Sahin,
Modeling, Simulation, and Visualization Engineering
Old Dominion University
Norfolk, Va *23529, USA*
osahi001@odu.edu

***Abstract*—This paper presents a method of semiautomatic initialization of a simulation through data mining. The method retrieves data from websites like Census, CIA, and Wikipedia and uses the data to initialize large-scale simulations. The method reduces time and human errors inputting the data. As a use case, the method is implemented to initialize a continuously running simulation environment.**

***Index Terms*— Initialization, data mining, large-scale simulation, continuously running simulation**

I. INTRODUCTION

Successful results from a large-scale simulation are contingent on an accurate initialization process because there are more components to initialize. If the initialization process runs unsuccessfully, because of incorrect inputs, re-initialization due to change of values, unforeseen runtime errors, or inconsistent outputs, the whole simulation needs to be reinitialized from the beginning. Ren, Buskens, & Gonzales, 2004 highligt the initialization problem in two points:

- It can take hours to initialize large-scale simulation which has hundreds or even thousands of components. This is important as naïve approaches to initialization require simulation components to initialize one at a time.
- Failures may occur during initialization. To restart the initialization process from the beginning is not desirable because reinitializing takes more time and errors may occur while initializing.

To deal with these problems Sheridan and Verplanck's (1978) proposes a taxonomy which has ten levels of automation, of particular interest is level five that says: "*which computer selects action and implements it if human approves.*" In other words, automation level five eliminates initialization problems since it makes initialization a semi-automatic process.

Initialization of a simulation is the procedure of setting the components to proper initial values (Ren, Buskens, & Gonzales, 2004; Schruben, Singh, & Tierney, 1983). This commonly accepted definition proposes the initialization occurs when data values are introduced as initial conditions of a simulation run.

There are different approaches to initialization. Kelton (1989) defined two types of initialization methods. Deterministic (fixed) initialization, where the simulation components are chosen as constant values, and stochastic (random) initialization when the simulation components are chosen from the steady-state probability distributions. Sandikci and Sabuncuoglu (2006) classify initialization into three groups - intelligent initialization methods, truncation (deletion) heuristics, and general studies. In the intelligent initialization transient is avoided by showing the system that conditions ready for run the simulation. In truncation heuristic the idea is deleting some observations from the beginning of the sequence. However general studies, Sandikci and Sabuncuoglu seems that as a combination of intelligent and truncation heuristic.

This paper is organized as follows:

- Section two provides a description of the proposed method.
- Section three provides how the method can be implemented using a use case
- Section four provides a summary
- Section five provides concluding remarks.

II. PROPOSED METHOD

Figure 1 provides a high level perspective on the method of initialization using data mining. This method is deterministic due to initial values are fixed for the components. This method is intelligent because the initial values are ready before running the simulation.

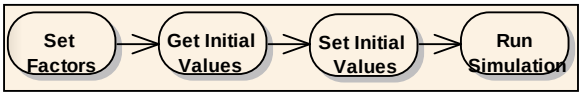


Figure 1 Semi-automatic Simulation Initialization

From figure 1, Set Factors select the factors that require initialization through data mining. Get Initial Values search and mine data values from the internet in order to provide initial conditions. These initial values are collected in one file in order to initialize factors (Set Initial Values). Lastly, the simulation is run using these initial values (Run Simulation). It is noted that the proposed method is not fully automatic because of the probability of obtaining inconsistent data during the mining process. The user must evaluate the consistency of the mined data values. Although, the proposed method facilitates the evaluation of inconsistencies through a double filtering process, it does not filter 100% of inconsistencies.

Figure 2 provide a higher level of detail on the data mining process (Get Initial Values) as it is the core of the process of initialization.

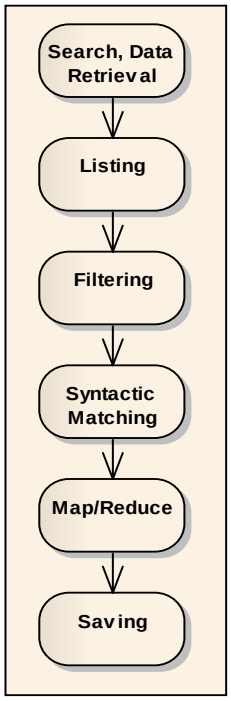


Figure 2 Get Initial Values Detail.

1. *Search, Data Retrieval:*

In order to conduct this step, search engines API are used. There are three major search engines application programming interface (API) on the web which are Google API, Yahoo API, and Bing API. They belong to Google®, Yahoo®, and Microsoft® respectively. Each API has pros and cons. Google is the largest search engine (based on the number of users) with 65% of people using their service on the web. However they don't allow getting the same amount of results from API as it does from the web user interface. There is a 100-query limit daily which is not useful for our method. Yahoo is the second biggest search engine on the web with 15% of users. Yahoo has API for developers and they allow a large number of requests, but they are not free. In third place, Microsoft's search engine takes 13% of users on the web among the other two search engines. Bing provides its API for free and full accessibility. In addition, Bing allows 50 results per request versus Google's 10 results per request. Bing's disadvantage is a limit of seven queries per second. The proposed method uses Bing API because of the free, unlimited number of search requests.

2. *Listing:*

Listing is done via a customized web crawler program that automatically retrieves the related links from the search results (Castillo, 2005). The crawler gets the links from Bing search results through a filtering process to remove undesirable content. The web crawler does not follow the links and download content of the found web pages. The downloading process is handled by other API which is called Alchemy.

3. *Filtering:*

Stored related links are sent to the filtering process for removing navigation links, advertisements, and other undesirable content. This process is handled by Alchemy API (Turner, Rouch, & Chuyko, 2012). There is another API called "TextWise" (TextWise, 2012) which creates same results as Alchemy. Alchemy was selected because it gave more detail filtering than TextWise. There are different filtering options Alchemy API provides such as:

- Named Entity Extraction
- Concept Tagging
- Keyword / Term Extraction
- Sentiment Analysis
- Relation Extraction
- Topic Categorization / Text Classification
- Automatic Language Identification
- Text Extraction / Web Page Cleaning
- Structured Data Extraction / Content Scraping
- Micro formats Parsing / Extraction
- RSS / ATOM Feed Detection.

In the proposed method text extraction/web page cleaning technique has been used for the filtering.

4. *Syntactic Matching:*

In the proposed method, syntactic structure matching is used to specify the intended meanings of a sentence based on the component words. The syntactic structure matching is programmed with a java library called regular expressions. The program is looking for any match in the cleared text documents that user enters as keyword. Program takes the whole sentence and loads to a file regardless of the meaning and place of the word.

5. *Hadoop Map/Reduce:*

The syntactically matched data is then funneled to the Hadoop server for map/reduce job for further filtering. Hadoop Map/Reduce is a software framework that allows processing of large amounts of data in parallel. Map/reduce gives best result in a short of time while delivering immediate results based on user's requests. In our case word frequency was requested for the related query.

6. *Saving:*

If the filtered matches are consistent for the user, he/she can select from three types of formats: extensible markup language (XML), javascript object notation (JSON), and text (TXT). Lastly, saved values are sent to the simulation.

Figure 3 shows the detailed search and syntactic activity diagram, from the user text inputting to the saving process. As seen on the figure, there is also an error handling process that warns the user if there are errors due to connectivity, no results or no matches, and unsupported web pages or formats messages.

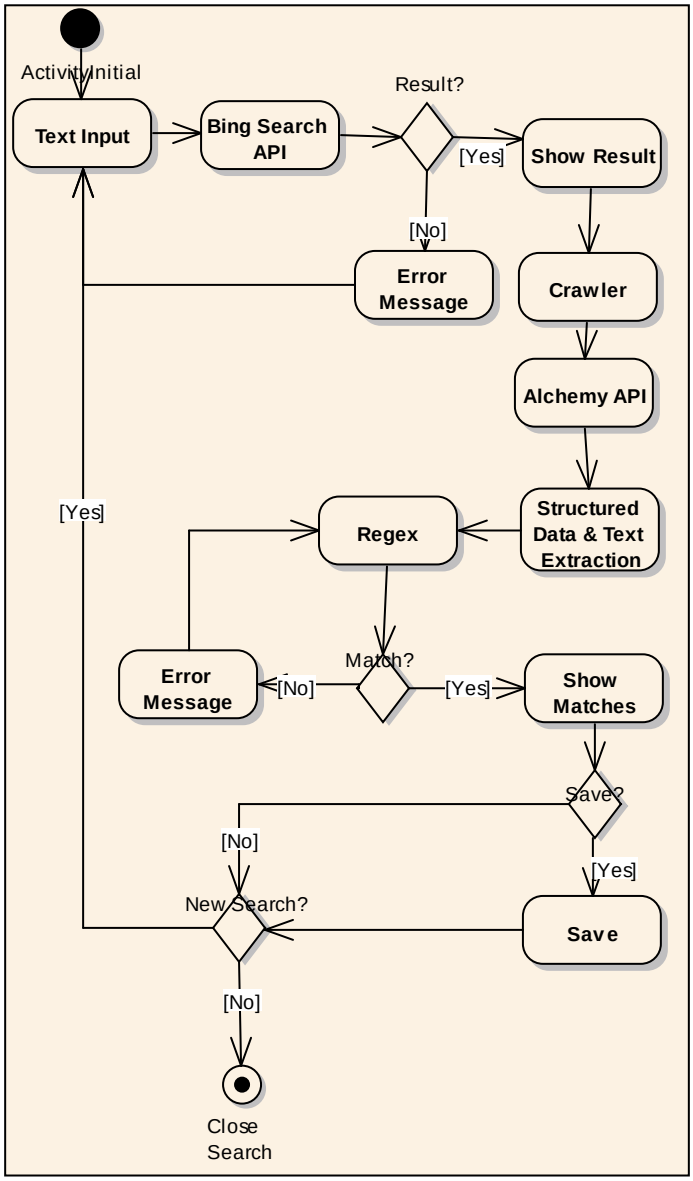


Figure 3. Activity Diagram for Method Implementation

III. IMPLEMENTED USE CASE

The semiautomatic initialization method is implemented in a continuously running simulation that seeks to support decision makers in a catastrophic event, in this case flooding due to sea level rise if the Hampton Roads region in southeastern Virginia.

Implementation of the use case of Sea Level Rise study was done using agent-based modeling and system dynamics. Relevant factors were captured from the Department of Homeland Security critical infrastructure resources and key resources taxonomy. The factors are:

1. Agriculture and Food
2. Banking and Finance

3. Chemical and Hazardous Materials Industry
4. Defense Industrial Base
5. Energy
6. Emergency Services
7. Information Technology
8. Telecommunications
9. Postal and Shipping
10. Healthcare and Public Health
11. Transportation
12. Water
13. National Monuments and Icons
14. Commercial Facilities
15. Government Facilities
16. Dams
17. Nuclear Facilities

Current mining and initialization do not cover all factors. They cover data values in the areas of demographics, water, energy, and transportation. The important aspect of these sub-factors is that they can be mined for initialized when simulating *any area*, be a city, a county, or a country as long as there is data available.

The sample of data mined for the U.S is shown in table 1. Table 2 shows data mined to initialize the simulation of the city of Suffolk and table 3 shows data mined for the city of Virginia Beach.

It is noted that these are search prototypes. Currently, the search is being expanded to address all the above mentioned factors in their totality.

USPopulation	301,461,533
USBirthRate	13.83
USDeathRate	8.38
USUnemploymentRate	9.6
USKwhPP	3.95E+12
USWaterAvgPerYear	575
Mean USSalary	62,363
Median Salary	81,537
Truck Freight/mil.ton	11712.41
Rail Freight/mil.ton	1978.584
Water Freight/mil.ton	1667.825
Air Freight/mil.ton	6.35
Pipeline Freight/mil.ton	3528.569
Ave.Food Per Month	277.50
US House Value	185,400

Table 1. US Mined Data

Suffolk Population	81,062	
USBirthRate	13.83	
USDeathRate	8.38	
USUnemploymentRate	9.6	
USKwhPP	3.95E+12	
USWaterAvgPerYear	575	
Suffolk Mean Salary	75,370	
Suffolk Median Salary	85,397	
Truck Freight/mil.ton	11712.41	
Rail Freight/mil.ton	1978.584	
Water Freight/mil.ton	1667.825	
Air Freight/mil.ton	6.35	
Pipeline Freight/mil.ton	3528.569	
Ave.Food Per Month	277.50	
Housing Value	243,900	
Latitude	36 43	N
Longitude	76 35	W

Table 2. City of Suffolk Mined Data

Virginia Beach Population	434,922	
USBirthRate	13.83	
USDeathRate	8.38	
USUnemploymentRate	9.6	
USKwhPP	3.95E+12	
USWaterAvgPerYear	575	
VirginiaBeach Mean Salary	72,029	
VirginiaBeach Median Salary	89,081	
Truck Freight/mil.ton	11712.41	
Rail Freight/mil.ton	1978.584	
Water Freight/mil.ton	1667.825	
Air Freight/mil.ton	6.35	
Pipeline Freight/mil.ton	3528.569	
Ave.Food Per Month	277.50	
Housing Value	268,600	
Latitude	36 51	N
Longitude	75 58	W

Table 3. City of Virginia Beach Mined Data

IV. SUMMARY

In order to get successful result from a simulation, data should be up to date and accurate. Especially challenging is data that comes from different sources and context, such as different web sites. The proposed method mines data from the internet and semi-automatically initializes a simulation. Mining data from the internet allows us to input current data that is available online. Semi-automatic initialization reduces the probability of error by reducing manual input. In addition, given that there are many variables to be initialized, the process reduces input time. Sources such as the US Census, the CIA, USDA, Wikipedia websites are among those used for mining and initializing simulations.

V. CONCLUSION

This paper proposes a method to semi-automatically initialize simulations using mined data from the internet. The method seeks to reduce inputting time, reduce error due to manual input, and use up to date information. In a continuously running simulation, we successfully tested the initialization of water, energy, and transportation factors in the prototype. All the initial data was successfully mined and initialized. For future work, a calibration process is proposed in order to test simulation outputs with data from the system.

VI. ACKNOWLEDGEMENT

The author is grateful to Dr. Saikou Diallo for his instructive feedback on the topic of the paper. The author is grateful to Dr. Jose Padilla for his continuous guidance, motivation and support which helped a great deal to the task of writing this paper.

VII. BIBLIOGRAPHY

Castillo, C. (2005). Effective web crawling. *SIGIR FORUM* , 55-56.

Kelton, W. D. (1989). Random Initialization Methods in Simulation. *IIE Transactions* , 355-367.

Ren, Y., Buskens, R., & Gonzales, O. (2004). Dependable initialization of large-scale distributed software. Dependable Systems and Networks, 2004 International Conference on, (pp. 335-344).

TextWise, L. (2012). *TextWise*. Retrieved March 29, 2012, from TextWise: http://textwise.com

Turner, E., Rouch, S., & Chuyko, A. (2012). *Alchemyapi*. Retrieved March 29, 2012, from AlchemyAPI: http://www.alchemyapi.com/

Sandıkçı, B., & Sabuncuoğlu, İ. (2006). Analysis of the behavior of the transient period in non-terminating simulations. European Journal of Operational Research , 252-267.

Homeland Security | Military Modeling and Simulation

VMASC Track Chair: Dr. Barry Ezell
MSVE Track Chair: Dr. Jim Leathrum

Viet Nam: 1961-1963 Using System Dynamics to Evaluate Alternative Policy Opportunities in South East Asia
Author(s): Joseph M. Bradley

Viruses to Missiles: Modeling Escalation in the Cyber Age
Author(s): Nicholas Reese

Adding Value to Dismounted C5ISR Systems Using the SDP
Author(s): Nicholas Coronato, Geoffrey Hansen, Bryan Musk, David Rattay, and Major Lawrence Nunn

System Dynamics Simulation and Agent-Based Modeling Tools for Analyzing Network Security
Author(s): Nicholas Coronato

Optimizing the Configuration of the XM-25 Into Dismounted Missions through Combat Modeling and Simulation
Author(s): Bryan Musk and Marc Wood

A Next-Generation Emergency Response System for First Responders using Retasking of Wireless Sensor Networks
Author(s): Syed R. Rizva, Chinmay Lokesh, Michele C. Weigle, and Stephan Olariu

Viet Nam: 1961-1963

Using system dynamics to evaluate alternative policy opportunities in South East Asia

Joseph M. Bradley
Old Dominion University: National Centers for System of Systems Engineering
Norfolk, Virginia, USA
jmbradle@odu.edu

***Abstract*—This paper examines the use of System Dynamics to first model the actual events in Viet Nam from John F. Kennedy's inauguration to his untimely death, and then propose alternate policies for both nations that would have likely altered the future of South Viet Nam, the United States and the region. The model selects Viet Cong population as the dependent variable, with a host of factors as independent variables drawn from a generic insurgency model developed by Sokolowski and Banks. Several different policy changes are examined for their effectiveness in improving the survivability of South Viet Nam. An examination of contemporaneous literature from near the end of this period indicates that a minority of analysts as well as government figures had been recommending the potential alternatives but their voices were not strong enough to alter the decisions that were actually made. Areas for further model development are discussed.**

Keywords-System of Systems, System Dynamics, Insurgency Modeling, and State Stability

I. INTRODUCTION

As the 1960's began, President-elect John F. Kennedy came into the office of the Presidency of the United States. He had already been well informed on the struggle in South East Asia and spoken and written publicly about the region. The change in administrations represented an opportunity to chart an alternative course from the one set by the Eisenhower administration. Yet, to a large extent, Kennedy and his administration chose or allowed South Vietnamese policies aligned with his predecessor and continued the arc that led the US and South Viet Nam to defeat. This paper examines the use of System Dynamics to first model the actual events in Viet Nam from John F. Kennedy's inauguration to his untimely death, and then propose alternate policies for both nations that might likely have altered the future of South Viet Nam, the United States and the region. An examination of the literature from near the end of this period indicates that a minority of analysts as well as government figures had been recommending similar alternatives but their voices were not strong enough to alter the decisions that were actually made.

II. DEVELOPING THE RESEARCH QUESTION, HISTORICAL BACKGROUND AND METHODOLOGY

A. The Research Question

This case study is tasked with developing an alternative case study and research question to a case previously developed by Sokolowski and Banks. In *Modeling and Simulation for Analyzing Global Events*, Sokolowski and Banks examine the history of Viet Nam in the period 1963 to 1965 with a System Dynamics model. Their primary interest was answering the research question stated as:

How can agent based modeling in a social network structure represent the behavior of President Johnson and Ho Chi Minh and the outcomes of the Vietnam conflict; and is this modeling useful in assessing alternative behaviors with changes in the agents and the structure of the social network?[1]

The authors have previously developed a generic insurgency model based on System Dynamics that has been applied to several insurgencies including Nigeria [2] and Columbia [3]. This generic model provides the incentive for the research question for this paper, stated as: How can governmental policy changes in the United States and South Viet Nam during the period 1961 to 1963 be measured, and then represented in a qualitatively developed and quantifiably supported model that can predict insurgency end strength and prescribe counterinsurgency strategy?

The goal of the research question was to provide insight into the use of System Dynamics to model large scale societal events, be able to evaluate the impact of proposed policy changes, and attempt to discern previously unexpected or emergent behavior between the factors and events. System Dynamics was chosen as it is a well-accepted technique for large scale problems, reflecting trends and responses to policies. An emerging body of literature provides guidance on the techniques to translate qualitative data to quantitative models.

A literature review of contemporaneous writings, primarily American, was conducted to ascertain the many factors in both societies that contributed to the perspectives of the leaders in

both countries, the policy actions set in place by those leaders and the outcomes that appear to be the conclusions of those policies and the then obscured strategies and tactics of the North Vietnamese.

B. The Historical Background

1) Post World War II

As World War II drew to a close, the Vietnamese resistance recognized the end of the Japanese occupation was near, but that the French colonial rule was likely to be restored. Seeking independence, the resistance leadership, personified by Ho Chi Minh, sought to be recognized as the legitimate government in Viet Nam. [4]

2) The French return and then Depart

Despite the efforts of Ho Chi Minh, France was restored to control of then French Indochina, which included Cambodia and Laos, and the Vietnamese struggle shifted to an insurgency against the French. After a long struggle, the Vietnamese succeeded in defeating the French at Dien Bien Phu. The withdrawal of the French and the subsequent internal struggles in Viet Nam led to negotiations in Geneva and the Accords that established a partition between the north and south. The Accords did not establish two separate countries, but called for a plebiscite to determine the leadership of the nation. [5]

3) The Divided Viet Nam

As a result of the partitioning called for in the 1954 Geneva Accords, and the failure to conduct a plebiscite in the north, there were large migrations of people seeking to relocate in the area of the country that they believed most aligned with their philosophy or would give them the greatest opportunities. Many fled the North based on their expectations of the repressive form of government to follow, while those in the South participated in the first democratic vote and elected Diem to lead them. However, not all communists left the South. A significant contingent remained, at the direction of Uncle Ho and the Politburo, to be in place when the time came to reunify the country under communist rule. [6] [7]

4) Eisenhower Administration Policies in Viet Nam

The Eisenhower policies were driven by the larger world view. The domino theory was the simplistic view of an ever expanding communist monolith. Eisenhower's support, first for the French and then for the Diem government, was to create a bulwark against the spread of communism. [8]

5) Diem Government Policies

The Diem government was structured around the historical mandarin philosophy. While the visible head was President Diem, his brother, Ngo Dinh Nhu, was the primary power and architect of the South Vietnamese government. Largely driven by their Catholic heritage, they were intolerant of the Buddhist majority, while at the same time focused on using their time in power to distribute the national wealth to personal allies, a practice often called corruption[9]. Reflecting their mandarin leanings and training, they rarely traveled outside the main cities, especially Saigon, and their circle of advisors included few from the 'people'[4] [10]. Further, the Diem regime seemed to ignore the security issues left from the partition. When partition occurred, most Vietnamese communists headed to the north, yet at least five thousand adherents remained behind, essentially sleepers in place for when the word would come to spring into action. Also stored were numbers of caches of weapons, similarly waiting in place for direction from the north. [7] The Diem government was also largely ineffective in providing security to loyal local government leaders. The Viet Cong assassination efforts were effective at removing good local leaders, as well as dissuading replacements from stepping forward. [6]

6) President Kennedy's Policies

President Kennedy came into office with some thoughts on how and why to proceed in Viet Nam. He requested a brief from his staff within weeks of his inauguration [11], and by May 1961 had publicly announced the support of the United States for the government of Viet Nam. [12] Over the course of his administration, various documents indicate that he understood the need to win the hearts and minds of the South Vietnamese populace, the increasing detail of his directions indicates some frustration by President Kennedy with the lack of response by the Diem government to his advice on religion, security, corruption and economic benefits for the greater populace. [13, 14] [15] At the same time, he increased American military support for the Vietnamese government from 500 advisors when he took office, to over 16,000 when he was assassinated.[16]

7) Chaos in Viet Nam and in America

Assassination of Presidents in both America and Viet Nam set the stage for the next phase in Viet Nam. The assassination of President Diem preceded that of President Kennedy by three weeks, however, the arc of the story was largely set and decisions in both America and Viet Nam to follow that arc only ensured the eventual outcome.[17] As discussed in the model analysis conclusions, the time to set the course for a free South Viet Nam had already passed.

C. Methodology

A highly simplified version of a generic insurgency model was adapted for this research. [2] The model selects the Viet Cong population as the dependent variable, with a host of factors as independent variables drawn from a generic insurgency model developed by Sokolowski and Banks. As shown in **Error! Reference source not found.**, the generic model has only two stocks, the Viet Cong population (VC) and the Susceptible Population (SP). The factors were also reduced from the generic models using only the propaganda, social network and government policies for this model. The literature review provided numerous examples of these factors allowing for a framework to map the qualitative descriptions to quantitative values. This also allowed for a reduction in the number of feedback loops, allowing more rapid development of the model. The literature also provided point estimates of the insurgent population in the period from Kennedy's inauguration to his assassination. For simplicity, the total South Viet Nam population was held constant over the period of analysis, and the Susceptible Population was simply the difference between the total and the Viet Cong population,

minus losses from the Viet Cong, due to elimination from the population. The literature also provided some qualitative estimates of the effectiveness of the ARVN and US forces in removing Viet Cong leadership in this period. This allowed a qualitative estimate of the Loss of Leadership factor for the baseline and alternative cases.

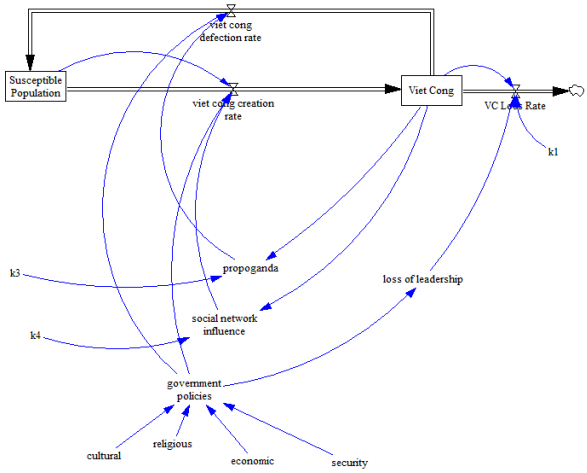


Figure 1: Simplified Insurgency Model

The Government Policy factors (Security, Economics, Religion and Cultural) were each modeled as an index. The range was 0.0 – 1.0, where 1.0 indicated a well-functioning, productive policy and 0.0 a complete failure. As will be discussed in more detail, three of those policies surface repeatedly in the literature: religion, economics and security. A subset of security, acting as a modifier to the policy, is loss of leadership. This was modeled as a multiplier to the effectiveness of Government Policy. The logic was driven by the concept that this was a specific role of Viet Nam's government, aided by US advisors that had the potential to accelerate the effectiveness of policy by removing Viet Cong leadership who acted against Government and the Susceptible Population. The factor of propaganda was assessed as significant in the Viet Cong defection rate, while the Social Network factor was more important to the Viet Cong creation rate. A summary of the factors and calibration data are shown in Table 1 Reference and alternative factor values, and Table 2 American Estimated Total and Viet Cong Faction Populations in Jan 1961, Jan 1962 and Nov 1963.

Table 1 Reference and alternative factor values

Factor	Baseline	Alternative
Cultural	0.85	NA
Religious	0.65	0.80
Economic	0.70	0.85
Security	0.60	0.80
Loss of Leadership	3.5	3.75

Table 2 American Estimated Total and Viet Cong Faction Populations in Jan 1961, Jan 1962 and Nov 1963

Population	Estimate (1000s) Jan 1961	Estimate (1000s) Jan 1962	Estimate (1000s) Nov 1963
Total South Viet Nam	14000	NA	NA
Viet Cong	5	16	70

III. Alternative Policies and Their Potential Outcomes

A. Economic – Impacts of a Free Press and Corruption Investigations

President Diem was not a supporter of the free press. He acted against the press when they challenged the corruption of his administration.[18] One alternative policy would have been for Diem to restore a free press. As a corollary, the free press would have been allowed to pursue reporting on corruption. These two combined effects would have likely improved the economic distribution of wealth, or at least allowed for people to believe that better economic conditions were imminent.

B. Religion – Could a Divider Have Been a Uniter?

As noted earlier, Viet Nam was historically Buddhist. The French occupation had given rise to a large Catholic community. During the partition, much of the Catholic population relocated to the South. The regime in the North was virulently anti-religious, a condition known to many in the South. The Diem government, however, repressed the Buddhists in the south. Given the experience of the new American President in overcoming his own religious minority status in a manner that did not fracture the country, an American policy that sought to create a larger tent in South Viet Nam might have allowed the two primary religious groups to work together to repel the Northern influence. [4] [10]

C. Security – The Impacts of Losing the Countryside

As noted earlier, the Viet Cong grabbed control of much of the countryside early in the campaign. They understood the benefits of appearing to be close to the people. Many of the farmers fled to the cities, well understanding the true nature of the Viet Cong. However, the government of South Viet Nam also retreated to the cities, falsely equating populace under control with progress in the war. One alternative policy would have been to contest the countryside with local leaders loyal to the South Viet Nam government who were protected from the assassination campaign run by the Viet Cong. [10]

The Viet Cong had also planted future agents in the South, simply by leaving behind adherents during the partition. The literature indicated that the South Viet Nam government was largely ineffective in identifying those adherents or their weapons caches. A more effective campaign could have been

designed to remove Viet Cong leaders. These actions would have the effective of raising Security, as well as the Loss of Leadership factor.[19] The Loss of Leadership factor was scaled on a 1 – 5 scale, and was estimated by this author to be the hardest factor to change, while it has the greatest payoff per percentage change. Also, unlike the other factors of Religion and Economy, where the government policy could be changed unilaterally, the security one is intertwined with Viet Cong actions. A sustained change in the factor implies that the Viet Cong would not have an effective counter strategy in place in a relatively short time.

IV. The Model - Alternative Policies and Their Potential Outcomes.

A. Model Results – Calibration

A highly simplified version of the generic insurgency model was adapted for this research. As shown in **Error! Reference source not found.**, the proposed model has only two stocks, the Viet Cong population (VC) and the Susceptible Population (SP). The factors were also reduced and only the propaganda (P), social network (SocNet) and government policies (GP) were used. The literature review provided numerous examples of these factors allowing for a framework to map the qualitative descriptions to quantitative values. The literature also provided point estimates of the insurgent population in the period from Kennedy's inauguration to his assassination, as well as some qualitative estimates of the effective of the ARVN and US forces in removing Viet Cong leadership in this period. A portrayal of the populations of Susceptible Population and Viet Cong are shown below in Figure 2: Baseline conditions for all factors.

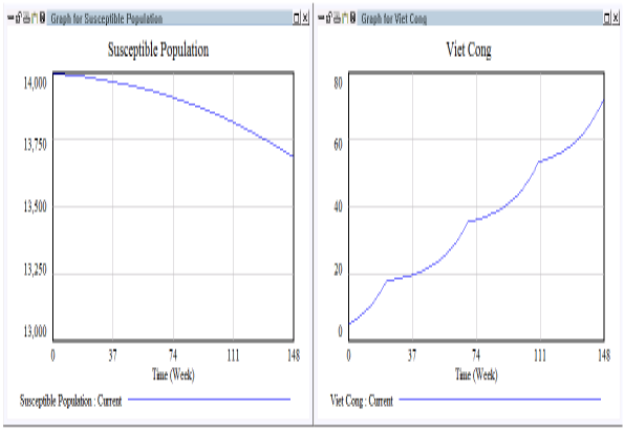


Figure 2: Baseline conditions for all factors

B. Economics – Allowing a Free Press and Corruption Investigations

As noted in 3.1, President Diem was not a supporter of the free press. In fact, he acted against the press when they challenged the corruption of his administration. One alternative US policy would have been to press Diem to restore a free press. The Economic factor was set at 0.70 for the baseline, and 0.85 for the alternative case. A portrayal of the populations of Susceptible Population and Viet Cong are show in Figure 3: Economic Factor is increased from 0.70 to 0.85.

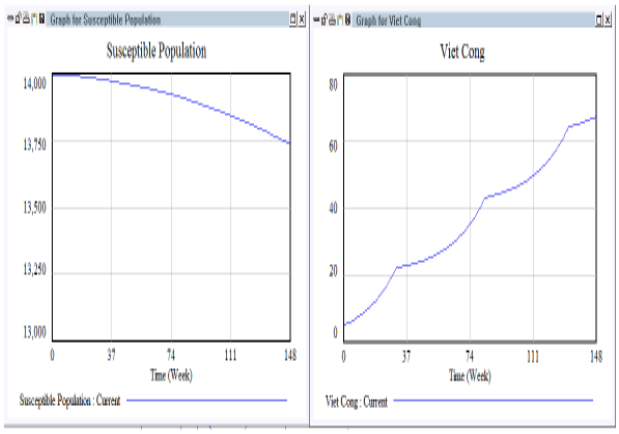


Figure 3: Economic Factor is increased from 0.70 to 0.85

C. Religion – Greater Tolerance for Buddhism

The religious factor (Rel) was initially set at 0.65. During the alternatives analysis, it was raised to 0.80, to represent a rapprochement between Diem and the leading Buddhists. A portrayal of the populations of Susceptible Population and Viet Cong are show in Figure 4: Religion Factor is increased from 0.65 to 0.80.

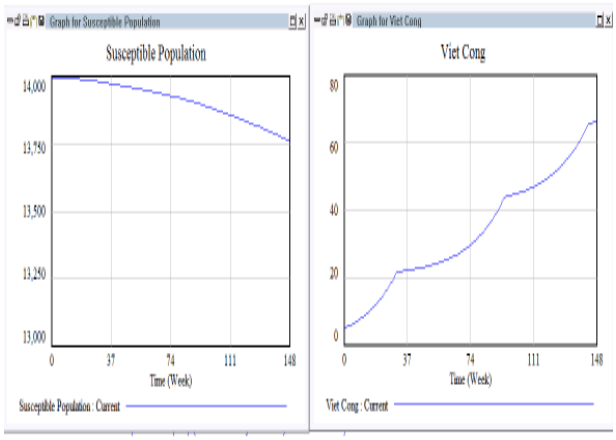


Figure 4: Religion Factor is increased from 0.65 to 0.80

D. Security – Keeping the Countryside and Targeting Viet Cong Leadership

During the alternative analysis, the Security factor was raised from 0.60 to 0.80, and the Loss of Leadership factor was increased from 3.50 to 3.75. A portrayal of the populations of Susceptible Population and Viet Cong are shown in Figure 5: Security Factor is raised from 0.60 to 0.80, and in Figure 6: Loss of Leadership is raised from 3.5 to 3.75.

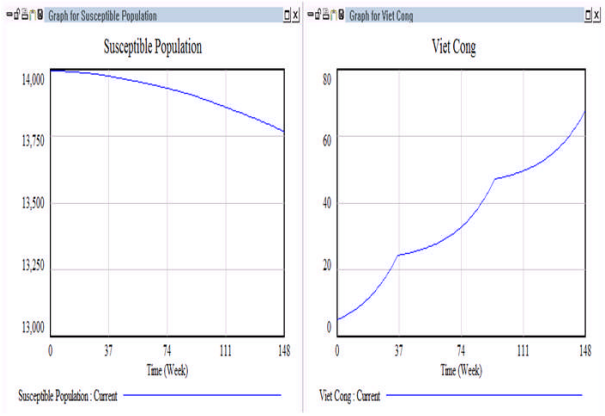


Figure 5: Security Factor is raised from 0.60 to 0.80

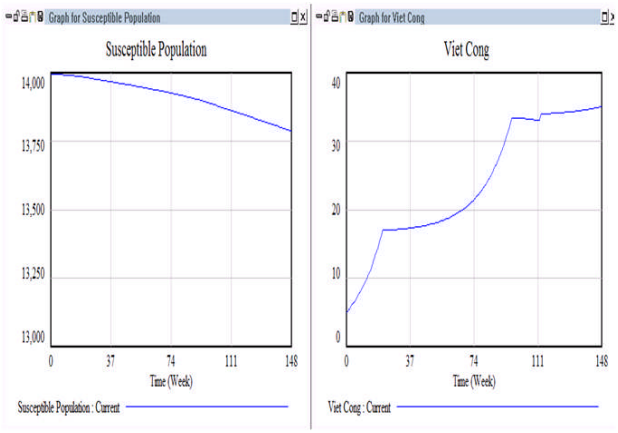


Figure 6: Loss of Leadership is raised from 3.5 to 3.75

E. Turning the Tide

The model indicates that, individually, the policy decisions would not have had significant impact on the growth of the Viet Cong. However, when applied as a package, the Viet Cong population gets capped just over one year after implementation of the policies, as shown in Figure 7: Impact of all factors being implemented. Additionally, the susceptible population losses are reduced by over 50%.

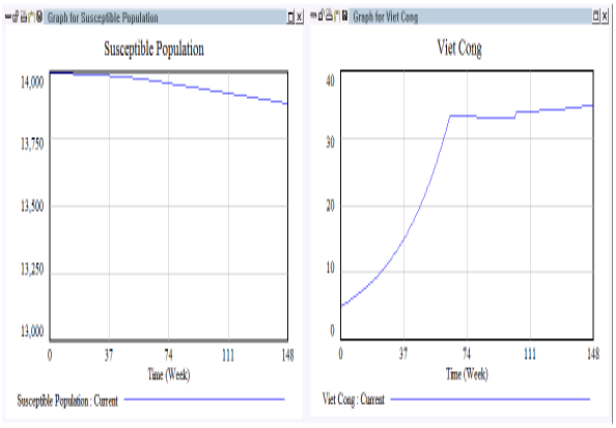


Figure 7: Impact of all factors being implemented

A summary of all estimated Viet Cong population for all the alternatives examined is contained in Table 3 Modeled Viet Cong population with postulated strategy change.

Table 3 Modeled Viet Cong population with postulated strategy change

Factor	Factor change	Estimated Viet Cong in Nov 1963
Baseline	NA	71.5
Religious	0.15 (23%)	66.3
Economic	0.15 (21%)	67.4
Security	.2 (33%)	67.6
Loss of Leadership	.25 (7%)	35.0
All factors	NA	34.8

In the final alternative, the tide is turned and Viet Cong population is controlled. It is noted that the model needs further research, testing and development before being suitable for leadership decision making. The Loss of Leadership factor is very dominating, and that should be further researched. Then, if the model were usable by decision makers, the challenge becomes converting those easy to write and model policy changes into action. Of note, there were many contemporaneous authors who had made similar suggestions to those offered here [4, 19, 20], yet the policies did not change. Even available National Security Action Memoranda indicate a growing awareness on the part of President Kennedy that President Diem was largely ignoring American advice, and that the situation on the ground was deteriorating rapidly [12-14]. Unfortunately for President Diem, and his brother, they were assassinated just 3 weeks before President Kennedy by many of the same leaders that they had repressed in late 1960 and restive military officers who saw the war being lost.

V. CONCLUSION

A System Dynamics model of the Kennedy administration period of the Viet Nam insurgency has been developed based on the interpretation of the policies in place or chosen by Presidents Diem and Kennedy. Alternative policy choices were proposed, modeled and assessed. While those policy proposals indicate a more likely chance of success for survival of the South Viet Nam regime, the real challenge would have been to implement those policy changes. Additionally, the model totally ignores the responses likely to be taken by the leaders in North Vietnam. It is clear from the literature review that those leaders were well versed in insurgency strategy and tactics, patient, and further had superb operational security. Their tactics were not always understood by the highest levels of American or South Vietnamese leadership, and it is highly conceivable that the NVA leaders could have developed effective counter strategies to those proposed in this paper.

Additionally, this model does not account for the significant infiltration from the North via the Ho Chi Minh trail. It has

assumed that all the Viet Cong are from converted Southerners. Yet in the period 1959 to 1963 (almost twice the period of the model), an estimated 50,000 persons were infiltrated to the South from the North.

The above elements would require a far more complex model, however, the concept of developing such a model would be highly useful in today's world of active insurgencies led by people equally capable as Ho Chi Minh, Nguyen Giap and the other members of the Northern leadership, in areas of the world that allow for reservoirs of resources outside the nation under threat.

In this context, the words of Andrew of St. Victor [21] come to mind:

How obscure the truth is, how deep it lies buried, how far from mortal sight it has plunged into the depths, how it will admit only a few, by how much work it is reached, how nearly no one ever succeeds, how it is dug out with difficulty and then only bit by bit.

This assignment has been a superb help in developing skills that enable this researcher's search for the truth.

REFERENCES

[1] J. A. Sokolowski and C. M. Banks, *Modeling and simulation for analyzing global events*. Hoboken, N.J.: Wiley, 2009.

[2] J. A. Sokolowski and C. M. Banks, "From empirical data to mathematical model: using population dynamics to characterize insurgencies," presented at the Proceedings of the 2007 summer computer simulation conference, San Diego, California, 2007.

[3] C. M. Banks and J. A. Sokolowski, "From War on Drugs to War against Terrorism: Modeling the evolution of Colombia's counter-insurgency," *Social Science Research*, vol. 38, pp. 146-154, 2009.

[4] E. G. Lansdale, "Viet Nam: Do We Understand Revolution?," *Foreign Affairs*, vol. 43, pp. 75-86, 1964.

[5] W. R. Fishel, "The National Liberation Front," *Vietnam Perspectives*, vol. 1, pp. 8-16, 1965.

[6] R. Thompson, "Squaring the Error," *Foreign Affairs*, vol. 46, pp. 442-453, 1968.

[7] J. M. Gates, "People's War in Vietnam," *The Journal of Military History*, vol. 54, pp. 325-344, 1990.

[8] G. A. Carver, "The Faceless Viet Cong," *Foreign Affairs*, vol. 44, pp. 347-372, 1966.

[9] G. A. Carver, "The Real Revolution in South Viet Nam," *Foreign Affairs*, vol. 43, pp. 387-408, 1965.

[10] G. H. Fox and C. A. Joiner, "Perceptions of the Vietnamese Public Administration System," *Administrative Science Quarterly*, vol. 8, pp. 443-481, 1964.

[11] J. F. Kennedy, "National Security Action Memorandum Number12 ." vol. The Papers of John F. Kennedy,, ed. Boston, Massachusetts John F. Kennedy Presidential Library and Museum, 1961.

[12] M. Bundy, "National Security Action Memorandum Number 65," vol. The Papers of John F. Kennedy, ed. Boston, Massachusetts: John F. Kennedy Presidential Library and Museum, 1961.

[13] M. Bundy, "National Security Action Memorandum Number 263," vol. The Papers of John F. Kennedy, ed. Boston, Massachusetts John F. Kennedy Presidential Library and Museum, 1963.

[14] M. Bundy, "National Security Action Memorandum Number 178," vol. The Papers of John F. Kennedy, ed. Boston, Massachusetts: John F. Kennedy Presidential Library and Museum, 1962.

[15] P. B. Davidson, *Vietnam at war : the history, 1946-1975*. Novato, Calif.: Presidio Press, 1988.

[16] R. Buzzanco, *Masters of war : military dissent and politics in the Vietnam era*. New York: Cambridge University Press, 1996.

[17] J. Garofano, "Tragedy or Choice in Vietnam? Learning to Think outside the Archival Box: A Review Essay," *International Security*, vol. 26, pp. 143-168, 2002.

[18] G. Lewy, *America in Vietnam*. New York: Oxford University Press, 1978.

[19] J. Farmer, "Counterinsurgency: Principles and Practices in Viet-Nam," DTIC Document1964.

[20] W. W. Rostow, "The Third Round," *Foreign Affairs*, vol. 42, pp. 1-10, 1963.

[21] M.-D. Chenu, *Nature, man, and society in the twelfth century; essays on new theological perspectives in the Latin West*. Chicago,: University of Chicago Press, 1968.

Appendix – Model Equations

$$\alpha(t) = GP * SocNet * SP(t)$$

$$\delta(t) = GP * VC(t) * P(t)$$

$$\mu(t) = integer(GP * k1 * LL * VC(t))$$

$$P(t) = k3 * VC(t)$$

$$SocNet(t) = k4 * VC(t)$$

$$GP = \frac{1}{Clt * Sec * Rel * Sec}$$

$$dSP/dt(t) = \alpha - \delta$$

$$dVC/dt(t) = \delta - \alpha - \mu$$

$$SP(t) = SP(0) + \int \frac{dSP(t)}{dt} * dt$$

$$VC(t) = VC(0) + \int \frac{dVC(t)}{dt} * dt$$

Where:

α(t) is the Viet Cong creation rate

δ(t) is the Viet Cong defection rate

μ(t) is the Viet Cong loss rate due to military action

P(t) is the propaganda factor

SocNet(t) is the social network influence factor

GP is the inverse of the product of the indexed policies for Cultural (Clt), Religious (Rel), Economic (Eco) and Security (Sec) spheres of government policy

LL is the Loss of Leadership (Viet Cong) factor

SP(t) is the Susceptible Population

VC(t) is the Viet Cong Population

Abstract

Paper Number 10
Title: Viruses to Missiles: Modeling Escalation in the Cyber Age

Author Name:

Nicholas Reese
Old Dominion University, College of Arts and Letters, Graduate Program in International Studies.

One of the emerging subjects of interest in International Relations is the opening of the Cyber Theater of operations. Cyber power as an instrument of state power is becoming a permanent part of world politics but it comes with may problems that have not yet been addressed. I have constructed an agent-based model to explore one implication of the use of cyber power.

Cyber power forces states to play a collective action game that includes an attribution problem as cyber provides anonymity to the belligerent. This model simulates such an environment and seeks to operationalize at what point a cyber war could tip into a kinetic event.

The model studies escalation by implementing a Game Theory-style collective action problem. States, patriotic hackers, and hackers are the three types of stationary agents connected by a network. The simulation will begin with the agents sending color-coded signals to one another representing cooperation and providing a positive payoff to each agent. The model generates a random defection, which represents a cyber attack and a negative payoff. Here, the agents have the option to choose cooperate or defect in return. After their payoff total reaches a critical threshold, the agents have the option to punish or escalate the conflict to a kinetic event.

This model will be an important step toward understanding relationship between cyber conflicts and traditional military conflicts. Understanding the tipping point will give policymakers a better understanding of the character of future conflicts that will certainly include a cyber dimension.

Adding Value to Dismounted C5ISR Systems Using the SDP

Nicholas Coronato, Geoffrey Hansen, Bryan Musk, David Rattay, MAJ Lawrence Nunn
United States Military Academy

***Abstract* – The purpose of this research is to examine the possibility of developing better command and control (C2) systems for dismounted combat leaders. We believe that, by utilizing value-focused thinking in the decision-making process, the U.S. military can acquire systems that provide maximum advantage to our Soldiers. By using a process known as the Systems Decision Process and keeping the needs and desires of the end-user in mind, we can find technologies that actually deliver value to dismounted leaders. We apply this concept to a case study that deals with eventually fielding commercial off-the-shelf (COTS) tablet devices in a deployed tactical environment. Input from experienced Soldiers helped generate a set of example alternatives that are feasible and could change the way leaders see and control the battlefield. Ideally, this approach could generate a myriad of solutions to current C2 problems with a higher chance of being embraced by the common Soldier on the ground.**

Index Terms - Computer simulation, Military communication, Technology management, Wireless communication.

I. Introduction

The current state of advanced Command, Control, Communications, Computers, Combat Systems, Intelligence, Surveillance, and Reconnaissance (C5ISR) projects might give the professional systems engineer some reason for concern. Over the past several decades, the Department of Defense has been seeking help from government agencies, research groups, and contractors to find a better way for dismounted leaders to communicate on the battlefield. Increased situational awareness is the primary focus of programs such as Land Warrior, Nett Warrior, and the Joint Battle Command-Platform. Some research groups – including the authors and the Virginia Modeling Analysis and Simulation Center – are looking into incorporating Coalition Battle Management Language in similar systems in order to provide a medium for digital correspondence between multinational units cooperating on a mission.

The challenge for future work in this area is to discover why the Army has yet to really field anything (other than prototypes), despite the years of research and millions of dollars spent. We believe that the problem is in the approach. Many programs focus on maximizing technological capability and appeal rather than optimizing the user experience and functionality. The Systems Decision Process focuses on delivering optimal value to the stakeholders throughout the entire system life cycle. In the case of a portable command and control system, it is the combat leader who has the most interest in the project; ultimately, he will be tasked to carry and interact with the devices over rugged terrain in addition to the rest of his essential gear. The war-fighter's wants and needs simply cannot be ignored. It is this lack of deliberate value-focused thinking that has thus far prevented successful integration of dismounted C5ISR systems into military operations.

In this essay, we provide an overview of the SDP as it relates to new C5ISR systems. After an analysis of the user requirements, we explain the methodology for the initial phases of the SDP, and conclude with the ways in which simulation can aid the systems engineer in solving this problem.

II. Background

A. Problem Analysis

The purpose for this research was to utilize the SDP as a methodology for adding value to current and future C5ISR systems as they are applied by dismounted combat leaders. The SDP is an approved, logically ordered process that keeps the stakeholders, clients, and end-users as a focus during each phase and throughout the system life cycle. Systems engineers first define the problem they face, then undergo a Solution Design phase, followed by Decision Making and Solution Implementation. Fig. 1 is the Systems Decision Process that was used by the design team in developing candidate prototype solutions.

This project does not aim at replacing already-established objectives in this area of research. However, we plan to refocus these goals through the use of stakeholder analysis in

order to bring the Soldier what he really needs. According to Drs. Parnell, Driscoll, and Henderson, "understanding who is affected by a system or solution to a decision problem provides the foundation for developing a complete definition of the problem" [1]. That foundation must be solid before a project can even take off.

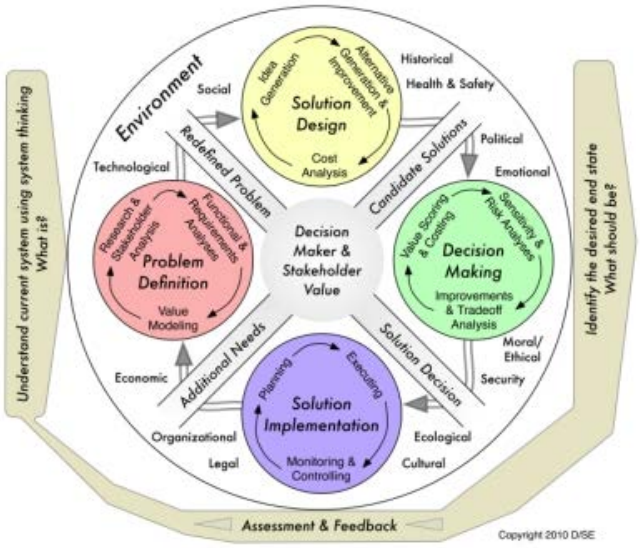


Fig. 1. The Systems Decision Process.

B. Past Work in Dismounted C5ISR Systems

The idea of using a computer-based combat system in dismounted operations has been around since the creation of the original Mission Need Statement for the Land Warrior system in 1993. This was followed by a system requirement document in 1994 [2]. The initial system focused primarily on creating a new weapon system that would enhance the war fighter's ability to target and engage the enemy [3].

Early Land Warrior systems focused almost entirely on reduced exposure firing positions. Early systems reduced the soldier's silhouette by nearly 75%, but the 18% reduction in accuracy proved too costly a price for fielding on the system. [4]. Later Land Warrior system would aim to give more tools to the soldier. However, extremely excessive learning times, up to twelve weeks to gain proficiency in some cases, and the small percentage of "thinking soldiers" who would actually use the full capabilities of the Land Warrior system once again provide to large of a barrier for fielding [5]. In order to maximize the value of dismounted C5ISR, systems must be created so that Soldiers *want* to use the devices; they must facilitate – not hinder – something Soldiers are doing. These failures should have signaled a need for more stakeholder input in future projects.

The Land Warrior-Stryker Interoperability (LW-SI) System was the last class of Land Warrior systems. LW-SI attempted to reduce the weight of the technology by refocusing the system from airborne to Stryker units. This allowed soldiers the ability to recharge batteries via their Stryker vehicle, thus reducing the overall weight of the system. This new version is also where we begin to see the focus on increasing situational awareness. Leaders of the Stryker battalion who tested the Land Warrior system in 2004 expressed an interest in the integration of a pre-loaded map, global positioning system, and text messaging [6]. However, this system still added too much weight to the soldier's already taxing payload. Due mostly to these shortcomings, the funding on the LW system was cut substantially for FY 2012 [7].

The system's budget cuts were also attributed to technological developments that allow for more effective and affordable alternatives [8]. One such alternative was leveraged in the newly created Nett Warrior program. This program primarily focused creating the ambitious types of situational awareness features that Land Warrior could not deliver, through the development of Army-specific platforms and interfaces [9]. The major advantage to this type of approach is the ability for the Army to make use of a wider range of developers. For instance, during the initial phase of the Nett Warrior program, one civilian software program was developed, called "SpeechTrans," that could potentially allow the Nett Warrior system to translate critical languages such as Arabic [10]. However, the Nett Warrior still weighed 8 pounds, compared to the ounces that a comparable Smartphone weighed at the time. Additionally, the commercially available Smartphones that were on the market at the end of the Nett Warrior program already had capabilities that outmatched the Nett Warrior System [11]. This called for a fresh look at battlefield C2 possibilities.

In 2011, the Army launched the Joint Battle Command Platform (JBC-P). The program has much of the same goals as the earlier Nett Warrior and Land Warrior Programs, but takes a different approach, and thus has already seen more approval from Soldiers. The JBC-P leverages a government-owned software framework, called Mobile/Handheld Computing Environment (CE), to allow anyone to create Android applications for the program [12]. Additionally, in March 2011, the Army launched their own version of an Android marketplace for Army created Applications [13]. This environment ensures that security risks – a serious concern with off-the-shelf commercial phones and public development – are mitigated. The program is still in the process of defining how the Smartphone will be implemented by the soldier, whether it will use commercial off-the-shelf (COTS) or government technologies, and what the interface of the system will look like [14]. Hopefully, JBC-P and similar programs will continue to focus more on stakeholder values as they work through the developmental phase. That is exactly the goal of this project team's research.

III. METHODOLOGY

Phase one of the SDP is Problem Definition, where the project team gains a thorough understanding of the problem

they will be attempting to solve. Through stakeholder analysis interviews, systems engineers identify the functions and objectives of the decision problem, as well as how to measure and weigh the value provided [15]. It is also important to predict and recognize external and environmental factors that will eventually play a role during the implementation phase. The primary outputs should be a set of concrete screening criteria for future solutions, and a list of consolidated values identified through the interviews.

Parnell, Driscoll, and Henderson consider interviews to be one of the best techniques for stakeholder analysis [16]. Despite the time-consuming nature of the interview process, they allow for the analyst to draw information from individual sources separately. There can be a lot of value in conducting a series of interviews, but it is key to focus on creating a stakeholder sample group that is diverse, unbiased, and experienced. Preparation, documentation, and thorough analysis are always encouraged.

A. Stakeholder Interviews

The authors approached the problem of dismounted C5ISR systems as a case study for applying the Systems Decision process. The following Stakeholder Analysis was performed through the use of interviews, as outlined above. A set of questions was generated prior to the interviews and used for each discussion conducted. This ensured that each of the stakeholders had a similar and fair interview.

During September 2011, separate initial stakeholder meetings were conducted with three US Army Majors. MAJ David Beskow, MAJ Benjamin Morales, and MAJ James Enos are instructors in the Department of Systems Engineering at the US Military Academy. They each have experience as Infantry officers, and were rated as top company commanders in their respective battalions. They discussed their values and concerns for a new piece of C5ISR technology and its implementation on the battlefield.

Sample questions for future interviews can be derived from these initial discussions. Some themes that were commonly discussed were:

- the placement of Smartphone devices and tablets on the Soldier's body;
- the durability of such a system in a combat environment;
- which members of the task organization would be equipped with devices;
- the learning curve for combat leaders to be comfortable with the system;
- integration of existing Army communication channels;
- replacement of the conventional Radio-Telephone Operator (RTO);
- and, perhaps most importantly, the ability of the end product to deliver increased situational awareness without becoming a distraction to the user.

Each officer agreed that there could be various designs for attaching Smartphones and tablet devices to a Soldier's combat kit. They emphasized, however, that the setup should not take away from a soldier's mobility or comfort. In addition, the device must be lighter than Land Warrior-type pieces of equipment. One possible arrangement, as proposed by MAJ Beskow, is a tablet that can flip down on the Soldier's chest to accompany a smaller phone device on the forearm. Obviously, this would not be appropriate for every rifleman, but would probably be a useful setup for squad or platoon leaders [17].

The officers interviewed suggested that they would not be interested in any gadget that will 1) subtract from their battlefield awareness or 2) require a significant amount of time to learn. These recommendations could steer the decision-making team away from technology similar to Land Warrior 1.0, which featured an eyepiece covering the Soldier's eye and providing constantly streaming data. This is where it is important to make the distinction between situational awareness – the goal – and battlefield awareness – the constraint. According to MAJ Enos, "the more user-friendly the system is, the less distracting the system becomes" [18]. Overall, the officers concluded that they would not like to train on a device that takes more than a few hours or a full day to learn.

The interviews shed light on the communications capabilities required of an effective C5ISR system. One important consideration is the ability to integrate a handheld device with the currently used communication systems: primarily, the individually carried MBITR radio and the ASIPS, carried by the RTO. If a system was to provide *more* capability, it should also incorporate the radio and voice communications already available in Army systems.

The amount of feedback provided by just three combat veterans provides an incredible amount of findings for the project team. Although some of the questions were designed to elicit an affirmative or negative response, the stakeholders were more excited to suggest other possibilities for the device. In this way, the project team can find new ideas that had not been considered prior to the analysis. For example, MAJ Beskow proposed what he called an "arms room concept," in which the individual leader has the freedom to choose which combination of Smartphone equipment to draw for each mission and what data to download in preparation [17].

The Virginia Modeling Analysis and Simulation Center (VMASC) is looking into similar systems, especially in terms of how they can be leveraged to provide more interoperability between multinational units in a coalition force. Stakeholder analysis with them in September 2011 yielded several new conclusions. The analysts at VMASC stressed the importance of having a device that is compatible for use with other NATO forces. Command and control among multinational elements is more difficult due to the language barrier as well as the differences in doctrine between cooperating units. VMASC also recommended

research into providing a simulation capability in real-time; with this function, the combat leader could input possible courses of action in the Smartphone device and simulate the effects of each. This would create a huge advantage for battlefield commanders if pursued and developed properly [19]. The recurring theme in the collection of stakeholder recommendations has been that the more capabilities the team integrates into the system – while remaining user-friendly – the more valuable it will become. In addition, the infantry officers have emphasized that the system must minimize the weight and bulk the soldiers have to carry, as well as the potential distraction factor. This initial research further highlights the concept that programs such as Land Warrior and Nett Warrior may have forgotten about the most important stakeholder: the Soldier.

B. Value Modeling

Conducting stakeholder analysis allows the systems engineering team to properly construct the Qualitative Value Model, which is the next key step in the SDP. The Qualitative Value Model is a holistic illustration of the system, which outlines the necessary functions, objectives, and value measures for the system to operate in accordance with the stakeholder's needs, wants, and desires [20]. It serves as the blueprint for the system to be constructed. Along with the problem statement, it is the most important component to come out of the problem definition phase.

The authors utilized their stakeholder analysis to outline basic capabilities required of a new system. Below is the Qualitative Value Model for this case study, which will add value to current and future C5ISR systems as they are applied by dismounted combat leaders.

As you can see, the Qualitative Value Model is a top-down hierarchical structure which outlines how the system should function and how its performance can be measured. At the top of Fig. 2, the main function, *Exchange Battlefield Information Rapidly*, outlines the overarching purpose and vision of the system. Below the main function are three sub-functions, which further outline how the system should properly operate. The three sub-functions created were *Provide Combat Durability*, and *Provide User Accessibility*. The sub-function *Provide Combat Durability* indicates that the system needs to be able to withstand the physical wear and tear of daily missions on a combat deployment. The second sub-function, *Provide User Accessibility*, outlines how the system should perform as a platform for combat leaders. These two sub-functions should be synchronized and provide further clarification as to how the system should function as a whole.

To ensure that each sub-function delivers maximum value, corresponding "objectives" are created for each function. The sub-function *Provide Combat Durability* has three objectives. The first is to Maximize Lifespan. The lifespan of the system is completely determined by its battery life. The system needs to be able to have a lifespan that can last for long term missions. The next objective is to *Maximize Resistance to Weather*. The system needs to be able to survive the various elements of weather that combat leaders will operate in. The value measures for this objective are the maximum and minimum temperature specifications of the system. The final objective for this sub-function is to *Maximize Durability*. During missions combat leaders will be diving to take cover from direct and indirect fire, and conducting other various tactical skills that result in direct contact with the surrounding environment.

The sub-function *Provide User Accessibility* has two corresponding objectives. The first objective is to *Maximize Comfort*. This pertains to the user interface of the application and design of the platform in relation to how combat leaders will equip it. The system as a platform cannot interfere with and weigh down a combat leader's ability to complete the mission. If the system is too bulky, combat leaders may decide that it is not even feasible to use for missions. The next objective is to *Minimize Learning Time*. The system needs to be very simple and intuitive. Ideally, no tutorial should be needed; Soldiers should be able to pick up the system and master it within a few hours. The more the software is compatible with the device, the easier learning will be for the user.

The Qualitative Value Model serves as the initial step in framing the architecture of the system. Based off this Qualitative Value Model, solution designs can be created. A properly constructed Qualitative Value Model will ensure a successful Solution Design phase.

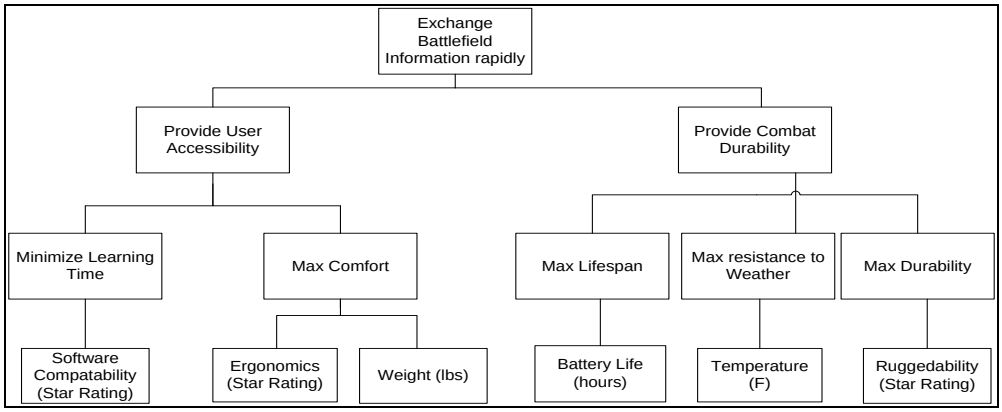


Fig. 2. Qualitative Value Model.

C. *Problem Statement*

After conducting the first phase of the SDP, the project team was able to develop a defined problem statement for this study: Provide dismounted leaders of coalition forces with a combat-ready C5ISR system that gives units digital interoperability, increased SA, and a more effective means of sharing information with cooperating elements.

IV. Solution Design & Scoring

A. *Alternative Generation*

The second phase of the SDP calls for potential ideas that might help to meet the problem statement. All of the stakeholder input is under consideration as the project team arrives at a set of feasible alternatives to present to the decision maker. In order to decide which alternative might be the best, or most valuable, the team must put each candidate through a solution scoring process.

For the case of our C5ISR system, the project team realized that there are two components to the most valuable system: the hardware configuration and the software interface to be used on the tablet. Value for the various physical devices under consideration is fairly easy to discern: each tablet or Smartphone has a specific weight, battery life, and temperature limitation as defined by the manufacturer. Scoring for the software can be slightly more difficult though; how can you know whether one digital interface is more suitable for combat operations than another? This is where modeling and simulation become very useful. A program called CogTool can be used to determine the best interface arrangement – what the user sees and interacts with on the screen – while combat simulators such as IWARS and ONESAF help provide insight into the command and control capabilities added by our Smartphone devices.

B. *CogTool*

Stakeholders expressed the need for a user-friendly system with a shallow learning curve. This implies the need for a touch-screen interface that is easily navigated and provides only the information that is necessary for the dismounted leader. The process of fully programming multiple interface designs and testing their effectiveness can be extremely time-consuming. CogTool offers its users a much simpler solution to this approach. When measuring the effectiveness of a software interface, processing time is what creates value. The programmer can model several different interfaces and the interactions between different screen views in CogTool. This allows us to test the time for a Soldier to observe the screen, process what he wants to do, and then interact with the device. Additional pauses can be added to simulate particular interactions that might require more time to think about. For instance, verifying a geographic location on a map may require more time than selecting from a drop-down menu, and CogTool helps us identify these differences in processing time. After testing various different interfaces, the project team can decide which provides most value to the end-user.

C. *Combat Simulations*

For the C5ISR system, the authors use the approved simulators IWARS and ONESAF as outputs for the system's performance. Both programs are high-resolution combat simulators that can simulate dismounted combat operations in a very detailed and visual manner. These simulators reveal how successful the technological application of Smartphones on the battlefield can be. Rather than conducting large scale field exercises, which consume resources, soldiers' time, and money, simulation can serve as the starting block for initial prototyping of the system. The use of simulation is appropriate because of its ability to provide rapid feedback and improvement towards the design of the system. Simulation also provides a better understanding of how improved situational awareness will affect decision making and the outcome of dismounted missions. More importantly,

simulation provides all of this information at a low cost and in a quick manner.

IWARS was used to simulate the benefit of C5ISR technology amongst dismounted units. The simulation consisted of a basic infantry squad patrolling in mountainous terrain similar to Afghanistan. Along their patrol route, an enemy ambush is in place, waiting for the unsuspecting friendly patrol to enter the kill zone. Having C5ISR technology, the dismounted patrol has instantaneous communication with a UAV which flies overhead. As a result, the dismounted friendly patrol gains a much higher level of situational awareness on the battlefield. Essentially, the friendly patrol on the ground can see what the UAV sees instantaneously. Having this advantage, the friendly patrol maneuvers properly to avoid the ambush and actually conducts a surprise hasty-raid against the enemy ambush element. The survivability of friendly units was high and most enemies were destroyed. Effectively, the dismounted friendly patrol used C5ISR technology to make faster and more informative decisions, saving lives on the battlefield. This observation from the simulation is the essence of what our C5ISR technology is intended to do. In the simulations without C5ISR, the friendly patrol was ambushed, and conducted an uphill assault against the ambush element resulting in a large loss of friendly units. Appendix I goes into further detail concerning the IWARS simulation and data that was collected.

Using ONESAF, the systems engineering team can communicate back and forth with dismounted units in the simulation using the application on any of our tablet devices. This provides quick and informative feedback to the team to better understand where the system's shortcomings are and how value can be optimized. Through this method, the user interface receives direct feedback from users and is improved rapidly. Using the App (user interface application) to communicate with the simulation can also serve as a train-up for new users of the system as well and possibly by a combat leader who doing reconnaissance or operational familiarization for a future mission.

Fig. 3 is a simple diagram outlining how the Smartphone system interacts with the simulators through the Coalition Battle Management Services (CBMS).

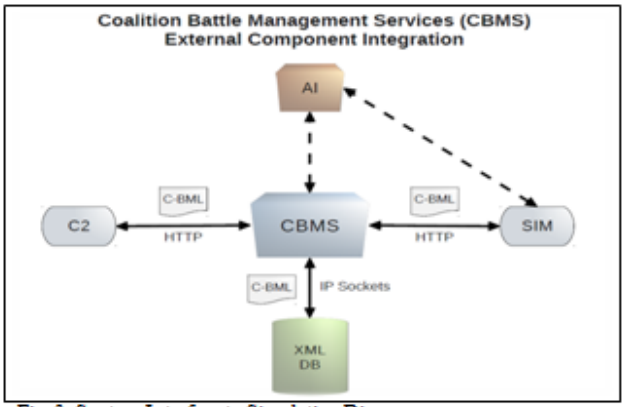


Fig. 3. System Interface to Simulation Diagram.

The App communicates with the simulation by sending and receiving XML files through CBMS. The AI, internal modeling of transmissions and situational awareness is modeled through the C2 node and the AI system. The AI system consists of the Message Transmission Service (MTS) and COMPOSER. MTS analyzes messages that will be sent amongst soldiers. The MTS then sends this data to COMPOSER. COMPOSER uses CES – SOCKET to communicate with MTS. COMPOSER models the transmission of communications then responds back to MTS with a modeled response of signal behavior (whether or not signal was sent, how long it took to transmit, etc.). MTS then sends the modeled response back into CBMS. CBMS then sends this information into the simulation, which helps make a much more realistic model of battlefield awareness and communication.

Situational Awareness Normalization and Dissemination Service (SANDS) models the awareness and artificial intelligence of each entity. Similar to MTS, SANDS models the entity's intelligence and awareness in relation to surrounding entities and the environment. This information is then sent into BCMS and into the simulation as well.

The project team can take a hypothetical combat situation that a dismounted leader might face, and process it through this simulation on our Smartphone device. After seeing how the entities in the simulation interact, we can draw conclusions on where our device can be improved before strapping it to live Soldiers for time-intensive testing and evaluation.

V. Conclusion

These first phases of the SDP are critical in developing candidate solutions for our C5ISR technology. The SDP's focus on delivering value to the client cannot be ignored in designing systems that may help save the lives of the war-fighter. Although the authors have provided examples of possible functions and objectives for a C5ISR system to be employed by dismounted commanders, we do not consider these initial outputs to be the only solutions. The main focus has been to show the importance of adding value through a process such as the SDP, and to offer ideas for more work in the endeavor to create better C2 systems.

VI. Future Work

Of course, the project is not completed with the conclusion of the Problem Definition phase. The team would apply the outputs (Candidate Solutions and testing criteria) to the next step: Decision Making. In order for the system to make it to successful implementation, it is crucial for the decision makers to remain focused on the end-user.

One area for further expansion involves a more diverse and comprehensive Stakeholder Analysis. The team should look into a wider selection of military personnel for idea generation, including Soldiers from multiple Army branches and varying pay-grades. Methodologies for obtaining

stakeholder input can be expanded to include surveys, focus groups, and research of published articles from credible sources.

It is the opinion of the authors that there should be a new approach to designing these systems. Land Warrior, Nett Warrior, and other programs previously looked at designing devices that would overload the user with battlefield information and could potentially cause a distraction during decisive engagements with the enemy. Future work should instead try to develop technologies that would help war-fighters make informed decisions on their unit's next actions, and deliver value to the American Soldier in the form of increased survivability.

REFERENCES

[1] G. Parnell, P. Driscoll, and D. Henderson. *Decision Making in Systems Engineering and Management*. New York: Wiley-Interscience, 2008. Print. 267.

[2] J. L. Dyer, "Annotated Bibliography of the Army Research Institute's Training Research Supporting the Land Warrior and Ground Soldier Systems: 1998-2009." Research Product. 2009-07, 1-2.

[3] Ibid. 7-8.

[4] J. L. Dyer, J. D. Salvetti, A. W. Vaughan, S. A. Beal P. Blankenbleckler, and M. Dlubac, "Reduced exposure firing with the Land Warrior system." ARI Research Report. 2005. vii.

[5] J. L. Dyer, G. W. Fober, R. Wampler, N. Blankenbeckler, M. Dlubac, and J. Centric, "Observations and assessments of Land Warrior training." Fort Benning, GA: Infantry Forces Research Unit, US Army Research Institute. 2000.

[6] J. L. Dyer, "Soldier survey results: Land Warrior-Rapid Fielding Initiative comparison." Special report to TRADOC Systems Manager-Soldier. 2004. 8.

[7] "Senate Committee on Armed Services Completes Markup of National Defense Authorization Act for Fiscal Year 2012." 18 June 2011. 12.

[8] M. Cox, "Army drops Land Warrior program." Army Times. 6 February 2007. 1-2.

[9] S. Ackerman, "Soldiers' Wearable Computers May Get an iPhone Brain." Wired.com. 14 April 2011. 1.

[10] S. Ackerman, "Psst, Military: there's Already a Universal Translator in the App Store." Wired.com. 19 April 2011. 1.

[11] S. Ackerman, "Army Hits Pause on 'Wearable Computer' Program." Wired.com. 28 July 2011. 1.

[12] C. Heininger, "Army develops Smartphone framework, applications for the front lines." Program Executive Office for Command, Control and Communications-Tactical. Army.mil. 18 April 2011.

[13] Army CIO/G-6, "Army Launches Apps Marketplace Prototype." Www.army.mil 19 March 2012.

[14] E. Montalbano, "Army Selects Android For Mobile Battlefield Network." InformationWeek. 22 April 2011.

[15] Parnell 267.

[16] Parnell 268.

[17] Beskow, David. Personal interview. West Point, NY. 13 Sept. 2011.

[18] Enos, James. Personal interview. West Point, NY. 13 Sept. 2011.

[19] Diallo, Saikou, and Andreas Tolk. Personal interview. Norfolk, VA. 19 Sept. 2011.

[20] Parnell 190.

AUTHOR INFORMATION

Nicholas Coronato Project Manager, Cadet Design Team, USMA Class of 2012, nicholas.coronato@us.army.mil

Geoffrey Hansen Lead Developer, Cadet Design Team, USMA Class of 2012, geoffrey.m.hansen@us.army.mil

Bryan Musk Systems Architect, Cadet Design Team, USMA Class of 2012, bryan.musk@us.army.mil

David Rattay Systems Analyst, Cadet Design Team, USMA Class of 2012, david.rattay@us.army.mil

MAJ Lawrence Nunn Faculty Advisor, Cadet Design Team, USMA Department of Systems Engineering, lawrence.nunn@usma.edu

System Dynamics Simulation and Agent-Based Modeling Tools for Analyzing Network Security

Nicholas Coronato
United States Military Academy

***Abstract* – This study will establish a baseline set of tools to model complexities of computer networks in terms of vulnerability and capability, both now and into the future. This report analyzes the advantages to a "combined-arms" strategy: agent-based simulation software coupled with a system dynamics model to show the effects of cyber policies and human behavior on our ability to keep information safe. As a case study, one can examine a typical computer network that belongs to a U.S. Army Brigade Combat Team during deployment, where security and functionality are both of high importance. It is also critical to look at how our approach can provide insight into policy-making for organizations such as the Department of Defense. While today's capabilities are a relevant baseline, it is important to introduce tools that can be applied to network systems in an even more significant time – the future.**

Index Terms – Computer simulation, Cyberspace, Information security, Military computing, Modeling.

I. Introduction

The advances made in the realm of cyber technology and network connectivity during the past decade alone indicate an increased focus on their importance in modern society. There is an impact on both the civilian and military environments. In *Cyber Silhouettes*, Timothy Thomas indicates that "just as cyber activities have introduced a change in the subjective nature of warfare in the civilian world, so-called information technologies have also changed the way militaries fight, maneuver, and organize" [1]. The scary truth of modern information warfare is that, regardless of how much time and money the United States spends on protecting our digital assets, *all* of our networks are vulnerable to attack. Any adversary – from a foreign state with bad intentions to a skilled hacker within our borders – could attempt to compromise the confidentiality, availability, or integrity of our information. The threats are very real and very numerous, yet there is an inherent difficulty in maintaining cyber security. We must first pursue a pro-active strategy focused on detection, as well as educating computer users on the effects of their actions on the network. For every measure taken to increase security in our networks, we lose critical functionality. The non-trivial challenge is to determine what decisions can be made now to improve our security outlook. How can we predict how our current policies will affect network behavior and security in the rapidly transforming cyber community?

II. Correspondence between Agent-Based Modeling and System Dynamics Simulation

A. Modeling and Simulation Approach

The purpose for this research was to design a methodology to facilitate the ability of an organization to monitor and predict the behavior of its computer networks over time. Additionally, it is a response to the call by Scholl [2] for research into the integration of system dynamics modeling and agent-based simulation. By utilizing both of these techniques in the context of computer networks, the advantages of a combined approach become clear, as well as some basic exploratory tools for analyzing network security. This study was also inspired by and builds upon a prior publication entitled "Methodology for Analyzing the Compromise of a Deployed Tactical Network" by Asman, Kim, Moschitto, Stauffer, and Huddleston [3].

Through modeling and simulation, we can test systems and policies to observe behavior when we have limited resources, time, or access to the actual system being modeled. While it is generally agreed upon that computer networks fall under the field of "complexity science," there has been some debate as to which modeling approach is most appropriate. Several studies have attempted to find the synergy between the fields of system dynamics (SD) modeling and agent-based simulation (ABS). These two non-linear modeling techniques emerged in systems engineering as separate approaches with little (or no) discussion on the benefits of collaboration between them [2].

After a brief beginners' description of the two techniques, there is an analysis of the potential advantages to applying each tool to the field of computer networks. It is important to note that this study does not reference hybrid simulation models, as the intent is to discover a way to integrate ABS and SD using two separate simulation software tools. The comparison considers each approach in general, the mathematical concepts behind them, and how easily results can be communicated and understood.

B. Agent-Based Simulation

A search for the definition of "agent-based modeling" revealed the following: A computer simulation technique that studies the interactions of a large number of entities known as "agents," which are units of programming that perceive situations and make decisions, usually of a specific or local nature. The objective is to observe global patterns that emerge from these myriad interactions [4].

At the core of agent-based modeling approaches is the assumption that "complexity arises from the interaction of individuals" [5]. The software used in these simulations typically represents an "agent" as an individual entity with set of rules that govern its behavior in the system, based on its perception of the surrounding environment. ABS is typically applied to problems in which we deal with systems at a more "bottom-up" or "physical" level – that is, less abstraction and more detail. However, this technique is useful because it can be implemented across all abstraction levels; the engineer can model objects of diverse nature and scale, from the individual data file in transit to the two offices passing email files [6]. This is not to say that a simple ABS model cannot provide useful information on dynamics of real-world systems and behavior patterns. It has become so popular because of its ability to portray systems at all levels of complexity.

The advantages of ABS as it relates to modeling computer networks are numerous. First, and perhaps most importantly, is the ability for agents' rules to be adaptive over time. Through the use of feedback and learning algorithms, an agent's schema can react to a changing environment [5]. This is key when considering systems such as complex computer networks, in which administrators deal with intelligent viruses and agent-based attacks more frequently every year. The fact that agents within the model can "think" for themselves is an element of ABS that is not necessarily possible in other approaches.

Various ABS modeling programs handle the concept of time differently, and this is another strong point. Some problems are better addressed with a discrete-event simulation, while others can be represented with continuous simulation. Finally, ABS is typically easier to maintain, as refinements can be made at the smallest level possible and result in local changes [6].

C. System Dynamics Simulation

SD is described by John Sterman, a leader in SD thinking, as "a method to enhance learning in complex systems… and design more effective policies" [6]. He is convinced that, by attempting to solve problems, people often naturally make them worse. A systems-thinking approach such as SD is a way to visualize and understand complex systems over time, especially with regard to changes and feedback. In simulation, it is an approach that focuses more on the structures of complex feedback systems, modeled by differential equations and their numerical solutions over time. It deals with aggregate details at the highest level of abstraction [7]. The most basic elements of the structural model are the causal loop diagram and stock-and-flow diagram.

With respect to modeling computer networks, SD can supplement the strengths of ABS. While it is unable to model behaviors of agents at the individual level, one of the most useful aspects of SD is the capability to predict a system's behavior patterns into the future, based on current conditions. As long as the parameters are properly defined and the system is correctly constructed in the model, software such as Vensim can provide quantitative data to the researcher through a continuous simulation that may help show behavior for years into the future. Of course, the structure of the system is assumed to remain constant throughout the simulation. Nonetheless, the primary advantage that SD provides to this study is an ability to understand aggregate behavior in a computer network and run a "flight-simulator" on potential security strategies aimed at influencing that behavior. The user of the SD tool can control certain variables in the model and identify effects on long-term behavior, and this will guide their policy decisions; thus, they test-drive their network policy before implementation. Sterman explains such a tool as providing "experiential as well as cognitive lessons that compress time and space so that we may experience the long-term consequences of our actions" [8].

D. A "Combined-Arms" Approach

With even a basic understanding of the differences between each approach, it is clear that there must be some synergy between SD and ABS. This paper explores the possibility of utilizing both SD and ABS to achieve predictive results that are more useful than either technique on its own can provide. Ever since the initial interest in multi-paradigm modeling, there has been a push to develop software that combines techniques, such as AnyLogic [9]. This type of software is not necessarily free or easy to use. For purposes of this study, I use two separate models of the same system

– NetLogo and VensimPLE – to compare the results and make informed decisions about security policies.

The average systems engineer may know little about overall network security, but a working knowledge of how users interact with a network in the real world allows for the creation of a fairly effective agent-based model. While ABS is powerful because it allows the user to construct a model without in-depth understanding of the global interdependencies [6], System Dynamics simulation fills in the gaps. As Nadine Schieritz puts it, "while system dynamics looks for [underlying] principles in system structure, agent-based modeling seeks them in agents' rules" [5].

III. Modeling the Network

A. *Defining the Network of Study*

The type of computer network that this study focuses on will be a standard deployed tactical network for a U.S. Army Brigade Combat Team (BCT), such as the one described by Asman's team [3]. In summary, this type of network is tasked with providing communication, information, situational awareness, command and control, and logistical support to a brigade (BDE) and its subordinate units. Fig. 1 below is borrowed from Asman in order to show the typical structure of a BCT tactical network.

Asman's group focused on modeling a "confidentiality" exploitation on a classified network. In this type of attack, an agent obtains access inside a network, locates targeted network files on servers or computers and copies those files [3]. Although not necessarily the most common or threatening cyber activity, it is fairly simple to understand and model.

For this study, the structure and typical activity of the BCT network is modeled as an ABS model with NetLogo and as an SD model with VensimPLE.

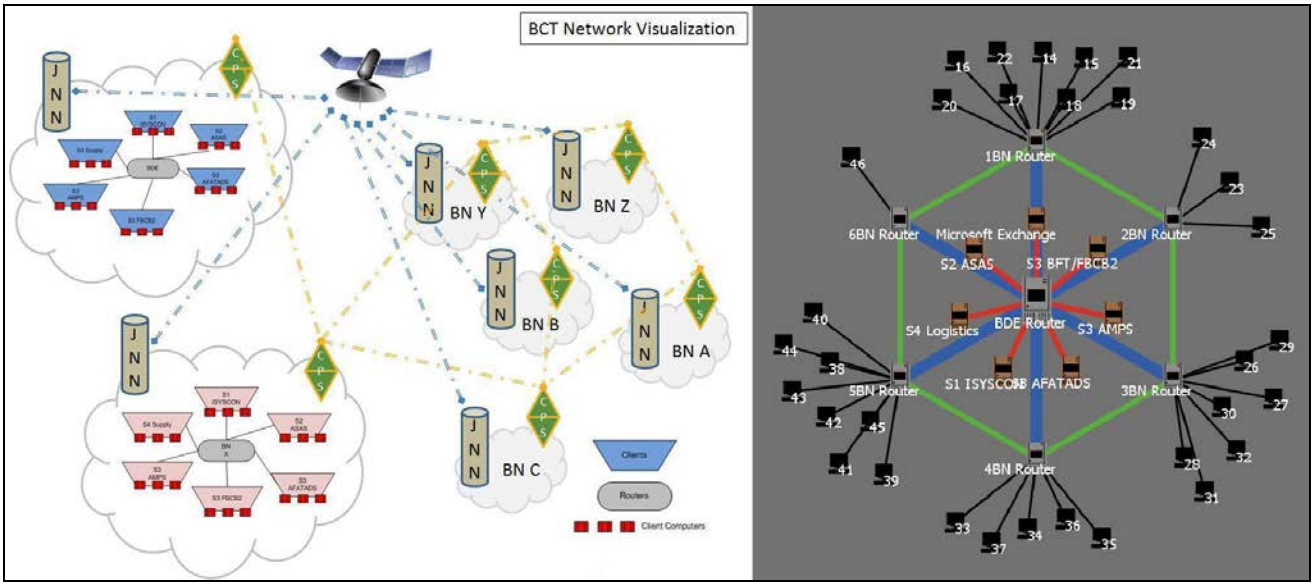


Fig. 1. BCT Network Visualization. "The figure above depicts the hypothetical topology and structure of a Brigade's (BDE) tactical network (left), and a view of the developed simulation model in a limited mesh hybrid configuration (right). The centrally located router is linked to both specific integrated system servers (ASAS, FBCB2, AMPS, etc.) and battalion (BN) routers. The battalion routers are also linked to adjacent battalion nodes. The width of the links denotes the bandwidth of that connection and the speed with which packets can travel. Blue links between routers represent a satellite link via Joint Node Networks (JNN) and green links represent a line-of-sight communications link between routers via Command Post Nodes (CPN)." [3].

B. *ABS Model*

NetLogo Software is available for free online and its use is not difficult for a fairly experienced coder. Asman's study created a very simplistic ABS model of a computer network using NetLogo. This initial model has been vastly expanded to more accurately portray the behavior of a real network. The structure remains the same in the new model, but agent rules make packet transfer more realistic. The new model, partially shown in Fig. 2, also allows for parameter adjustment during simulation, such as the ability to modify the rate at which files are being sent. The full view is located in Appendix I for reference. This ABS model shows packets (the smallest component of a file) being exchanged over a network with a limited hybrid configuration. Note that the location of each node in the network does not necessarily have geographic significance to a real-world computer network. However, this view highlights some of the advantages of ABS. Most importantly, individual packets can be identified by type (color) and location. If a network administrator were using this tool, they could watch as their network undergoes normal activity, and then interact with the model during simulation. For example, they can initiate a confidentiality attack on a particular computer from another node in the network. They can also destroy links in order to show the effects of isolating a particular set of nodes in an availability attack. All of the control that the user has in establishing the conditions for this model highlights the benefits of using ABS for the analysis of networks.

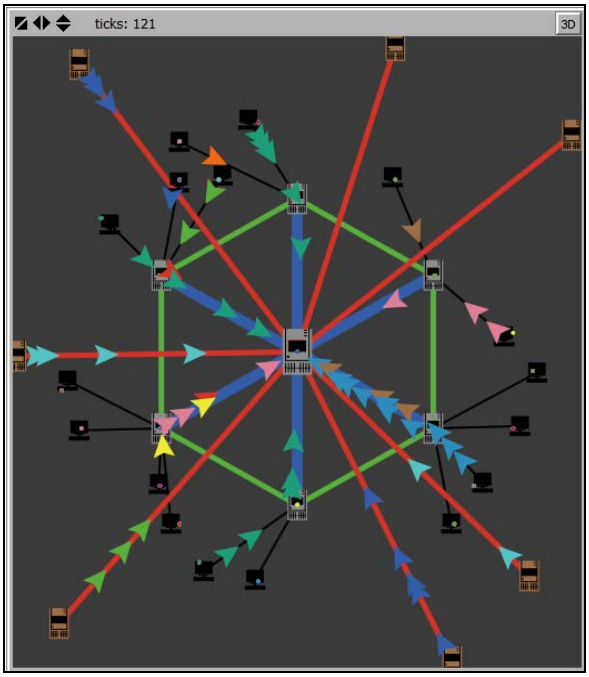


Fig. 2. Agent-Based Model of Network in NetLogo.

Additionally, there is a limited ability for feedback and data collection in this NetLogo model. The administrator can assign various weights to each type of file – as classified reports are more valued than standard email traffic – and the model will calculate how much "value" is lost during an attack. I will not explain more specifics for this model; Appendix I and Asman's paper can be referenced for full details.

C. *SD Model*

VensimPLE is the software that I used to develop a simple model for the same computer network. Although very basic and exploratory, this SD model was constructed with the intent of giving the administrator another tool for analyzing the BCT network. It is pictured in Fig. 3, and a larger view can be found in Appendix II. Note that, as opposed to NetLogo, a Vensim model cannot be manipulated during simulation, so the user must set the conditions prior to starting. In this model, the actual configuration of the network is unimportant and thus cannot be changed. It is most useful as a "flight simulator." My model attempts to do just that; it allows the user to play with the controls of his network and conduct multiple runs to identify strengths and weaknesses of certain policies. This is made simple with multiple slider bars for adjusting variables prior to a simulation run. Thus, although Vensim can be very difficult to master, this particular model is user-friendly for a network administrator with little experience in SD or this software.

The SD model is based on a few fundamental concepts. First, as Asman's team postulated, functionality of any computer network is inversely related to the level of security. As more protection is placed on a network, ease of use for a human interacting with the network is sacrificed. A familiar example would be the process of logging in to a computer; asking a user to input three different passwords, although effective in preventing successful hacks, impedes the productivity and satisfaction of that user. For that reason, a user may be willing to take shortcuts to make things easier; this is the second main concept at play in this SD model. As the functionality of the network drops, the desire for a human to comply with network security protocols also decreases, making an attack more likely. This balancing loop is essential in understanding the underlying structure of the SD model for a computer network. Simply put, more stringent security measures may actually increase the probability of an attack occurrence. The benefit of this SD model is that it allows the administrator to test drive his network for any length of time to see how different policies might play out. *What is the optimal security level for my network that operates at speed 'x' when we expect an attack of strength 'y'?* This is the type of question that this tool can help a network analyst answer.

With its advantages come some limitations. While the ABS model in NetLogo could assign value to individual files

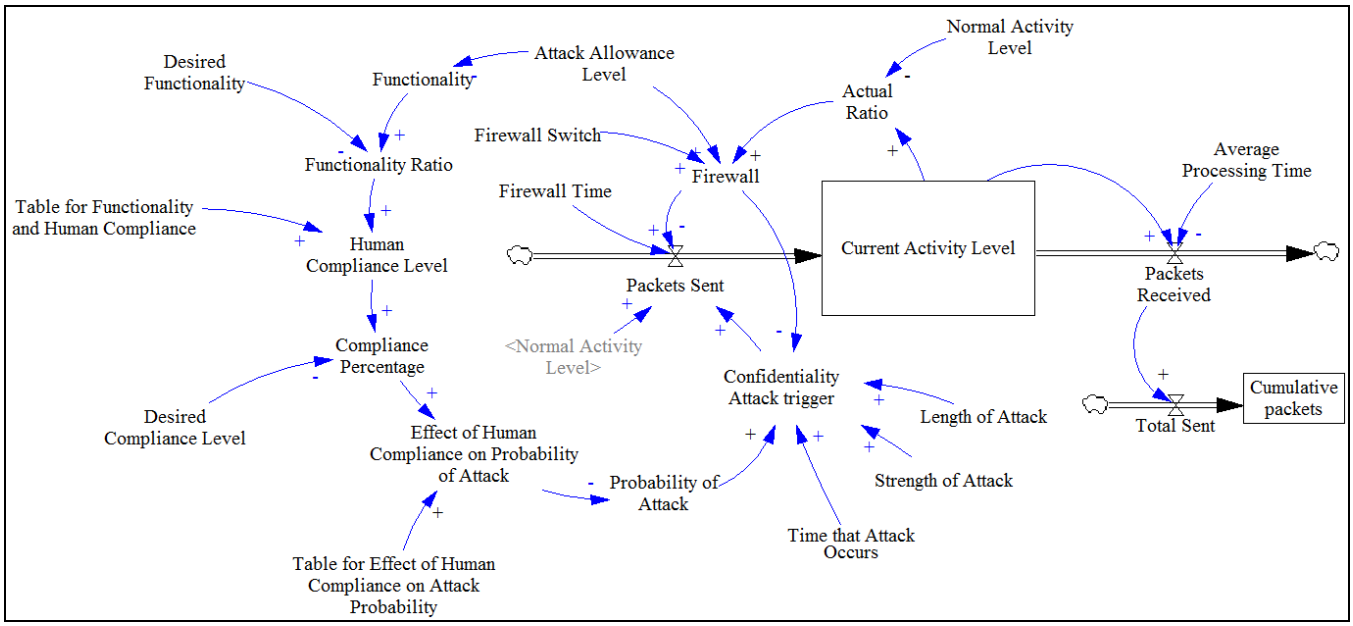


Fig. 3. System Dynamics Model of Computer Network Activity in VensimPLE.

lost during a confidentiality attack, the aggregate nature of an SD model makes that feature much more difficult to create. The Vensim model can give the user an idea of how network traffic, security, and functionality change over time, but it does not have the visual capabilities of NetLogo. Further work on an SD model is needed to provide an even more accurate representation of the BCT network under study.

IV. Observations

The results of this study are more qualitative than numerical. As the tools can be expanded and adapted to virtually any computer network, there is not much use to presenting specific output data for these experimental models. However, there are key observations from each simulation that would aid potential users.

The NetLogo model allows the user to set up the BCT network in multiple different arrangements, each of which has its advantages. By setting the levels for traffic of various file types in the network, the simulator can try to mimic the characteristics of their own network. Simulating for just a short amount of time allows for calculation of the amount of value that is lost from an availability or confidentiality attack. The ability to visualize the network while independent files are being passed can give an administrator some highly useful feedback on his system's weaknesses.

The Vensim model has the capability to show the user how well their network's security can defend against a future cyber attack. Essentially, the user defines the parameters of the network (normal activity level, normal security level, desired functionality ratio) and the specific severity and timing of attack. By using slider bars, the user can see, over time, how their firewall detects confidentiality attacks. If the packet traffic level gets too high on the network – such as the spike which would occur when large amounts of files are being downloaded – then the firewall detects a possible attack and shuts down that network activity. However, if the network administrator were to set the security level too low in order to maximize functionality, then the probability of an intrusion increases and an attack is eventually triggered. The most important observation that this SD tool provides is the insight into how functionality and human behavior interact to affect the overall network security.

V. Conclusion

SD delivers on many of the shortcomings of ABS, and vice-versa, as these two models indicate. In the example here, the ability of ABS to prescribe agent rules provides the ability to model packet traffic and nodes on a network more realistically. The complexity of the model is limited mainly by the programmer's creativity and understanding of network operations. The visual aspect of NetLogo and other ABS software is great for presentation and visual understanding of the interactions between agents – an area in which SD models pale in comparison. However, there being no endogenous feedback structure in NetLogo, we turn to Vensim.

If used effectively, SD can be used to not only visualize potential patterns of behavior in the future, but to identify ways to change that human behavior as it interacts with a system. Determining what behavior can be altered and

finding leverage points in the structure of the dynamic system are the first steps in designing effective policy changes. If a systems engineer were to analyze the SD model presented above, they would look for places in which human behavior has an influence and where a small change in one variable or structure brings large positive results to the overall system. In the Vensim model, the level of human compliance has a large role in whether or not an attack could occur; while computer users are abiding by security protocols, there is a better chance that an attack will be prevented. This is something that typical users probably do not fully understand with a naturally event-oriented world view.

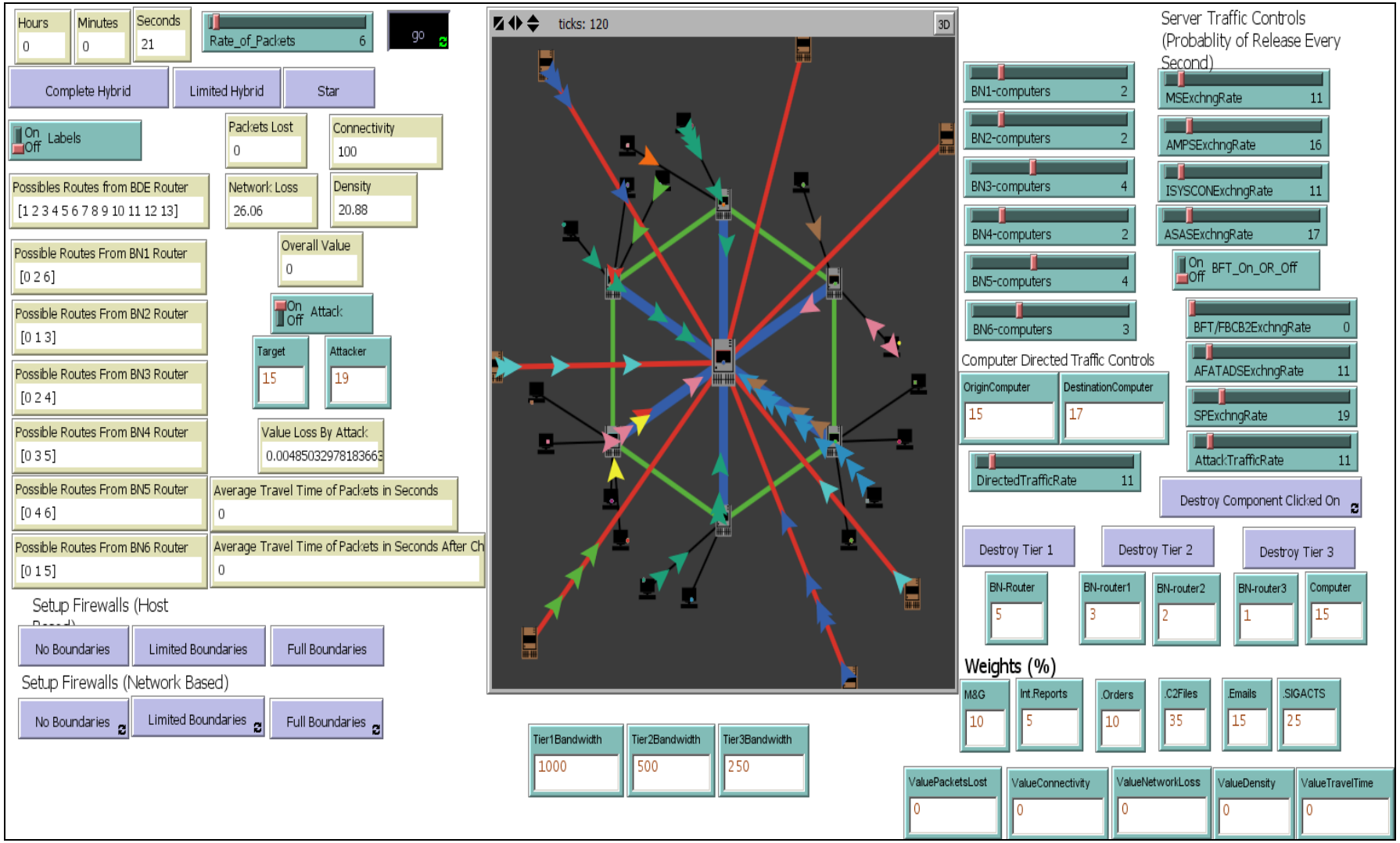
The network administrator could use SD analysis, along with an ABS model like the one presented, to show decision makers or other network users how they can combat cyber attacks in the future. Policy changes are bound to encounter resistance, but by showing the cause-and-effect relationships between human behavior and network performance, the network administrator may be able to change the world view. Once they understand the feedback structures at play, they may be more willing to comply with security policies in place. On the other hand, the problem may require a complete overhaul of the system's feedback structures. The ability to test-drive different policies and simulate attacks on the network can be incredibly valuable to those responsible for ensuring security and functionality for the Brigade Combat Team's network.

VI. Future Work

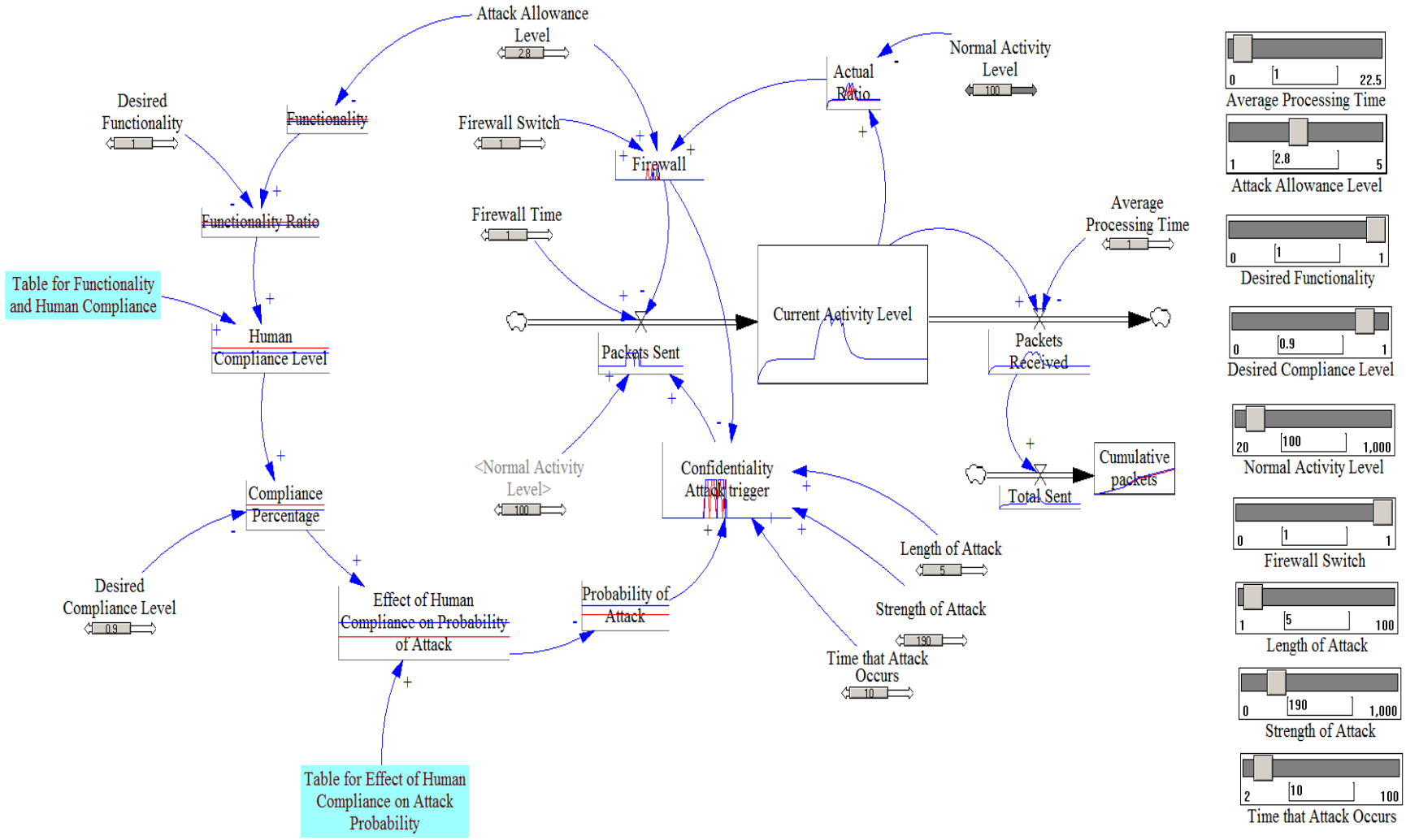
For purposes of simplicity, several potential areas of further work were left out of this exploratory study. Each of the two models can be expanded hugely and developed to more accurately represent a BCT network. Specifically, the SD model does not give the user the opportunity to experiment with randomly-initiated attacks. Also, bandwidth is a key component of functionality and activity level, but was not considered here. The relationships between functionality, human compliance, and probability of attack were not based on solid data, but a much more generalized interpretation of their interconnectedness. Further research would certainly provide a more accurate idea of how quickly the probability of successful attacks rises as network users take security shortcuts, for example. For both models, various types of firewalls could be considered, as there are key differences between host-based and network-based intrusion detection systems. As previously stated, these models only analyze confidentiality attacks on the network, but looking into simulation of integrity and availability attacks (as defined by Asman) presents another interesting challenge.

Future work is also necessary as networks become more complex. By continuing to develop a set of modeling tools such as these, we can incorporate new threats that arise and enhanced security measures to counter them. Intrusion detection systems, intrusion prevention systems, and proxy servers are relevant areas for future modeling. Homomorphic encryption is an example of a newer information security technique that could be included in future ABS and SD models. It may be impossible for modeling and simulation of computer networks to keep up with the ever-changing cyber world… but we must try.

APPENDIX I
FULL VIEW OF NETLOGO SIMULATION



APPENDIX II
FULL VIEW OF VENSIM MODEL



V. Acknowledgment

The author would like to thank several instructors for their guidance in research and model development: MAJ L. Nunn for constant guidance and support throughout the research and writing process; MAJ J. Enos for sparking an interest in System Dynamics, and enormous assistance in developing a working model; and MAJ S. Huddleston for inspiration from the beginning. Also, the author thanks G. Hansen, B. Asman, M. Kim, R. Moschitto, and C. Stauffer for all of the work put into the NetLogo model.

References

[1] T. L. Thomas, "Engaging Cyber or Information Ubiquity?" in *Cyber Silhouettes: Shadows Over Information Operations.* Fort Leavenworth, KS: Foreign Military Studies Office, 2005, ch. 1, p27.

[2] H. J. Scholl, "Agent-based and System Dynamics Modeling: A Call for Cross Study and Joint Research." *Proceedings of the 34th Hawaii International Conference on System Sciences – 2001*.

[3] B. C. Asman, M. H. Kim, R. A. Moschitto, J. C. Stauffer, S. H. Huddleston, "Methodology for Analyzing the Compromise of a Deployed Tactical Network," *Proceedings of the 2011 IEEE Systems and Information Engineering Design Symposium*.

[4] T. Marks-Tarlow, K. Clayton, and S. Guastello, "Terms & Definitions." Navigating The Complexity Space. Complexity Space Consulting, n.d. Web. 16 Mar. 2012. <http://thecomplexityspace.com/?page_id=19>.

[5] J. Sterman, "Learning in and about Complex Systems" in *Business Dynamics: Systems Thinking and Modeling for a Complex World*. Boston, MA: Irwin McGraw-Hill, 2000, ch. 1, p4.

[6] N. Schieritz, "Integrating System Dynamics and Agent-Based Modeling." *Proceedings of the 21st System Dynamics Conference*, July 20-24, 2003, New York City, NY.

[7] A. Borschchev and A. Filippov. "From System Dynamics and Discrete Event to Practical Agent Based Modeling: Reasons, Techniques, Tools." *The 22nd Conference of the System Dynamics Society*, July 25-29, 2004, Oxford, England.

[8] J. Sterman, "Teaching Takes Off: Flight Simulators for Management Education." Massachusetts Institute of Technology. N.p., n.d. Web. 21 Mar. 2012. <http://web.mit.edu/jsterman/www/SDG/beergame.html>.

[9] "AnyLogic Multi-Paradigm Simulator." PLM, PDM, Simulation, and Enterprise Architecture Consultants and Consulting Projects. Coensys, n.d. Web. 16 Mar. 2012. <http://www.coensys.com/anylogic.htm>.

Author Information

Nicholas Coronato, Cadet, USMA Class of 2012, nicholas.coronato@us.army.mil

Optimizing The Configuration Of The XM-25 Into Dismounted Missions Through Combat Modeling And Simulation

Bryan Musk and Marc Wood

***Abstract*—Combat Modeling and Simulation was used to determine the ideal configuration for the XM-25 weapon system at the platoon level for dismounted operations. The XM-25 weapon system is an area weapon system that incorporates airburst technology to precisely deliver hand grenade effects to counter defilade targets on the battlefield. Essentially, the shooter can decide the exact distance at which the round will detonate. The Infantry Warrior Simulation (IWARS), a high-resolution, agent-based simulator was selected to test an organic and weapon team configuration for three tactical scenarios. Although the results were not definitive, they suggest that there is not a one-size fits all configuration that will maximize the effectiveness of this combat multiplier. The common stakeholder view is that the decision to use the XM-25 in combat and subsequently the task organization within a platoon will depend upon the assigned mission.**

I. Introduction

Out in the rugged terrain of Afghanistan, an infantry squad is in contact with a team size element of Taliban insurgents. Knowing the terrain and its advantages like the back of their hand, the Taliban utilize large rocks and boulders for cover as they attempt to draw out coalition soldiers into an open and vulnerable area on the battlefield. There is no feasible or effective way for dismounted coalition soldiers to destroy enemy covered behind strong defilade positions unless they decide to maneuver to a better position which may expose them to hostile fire. The XM-25 counter defilade target engagement system is the solution to this growing problem. It is a handheld direct fire grenade launcher. The weapon system provides the shooter with air burst technology, which allows him to determine the exact point at which the round will detonate. This removes the advantage of cover on the battlefield for the enemy. Soon this weapon system will be integrated in infantry units Army wide; however, the growing concern amongst combat leaders during this integration is how the weapon system will be configured at the squad level. Should it be integrated into each fire team, directly participating in the main effort of dismounted missions, or will it be configured as a separate weapon team providing overall support for the main element (support by fire, over watch, etc)? Through the use of combat modeling, in particular, conducting simulations and output analysis in IWARS, an initial attempt towards discovering the ideal configuration of the XM-25 into dismounted missions can be conducted. The task organization and description of a conventional infantry rifle team and weapon team can be viewed in Appendix V (Figure 10 and Figure 11).

Combat modeling has long been used to evaluate various aspects of warfare. It can be traced backed to Frederick Lanchester during World War I. His equations provided the foundation for combat modeling; he used them to model various weapon systems, size of units, several different combat scenarios, and even line of sight. They have long been used as tools to provide enhanced decision making for combat leaders. Today, combat modeling has evolved with the fast paced digital world. Software such as IWARS (Infantry Warrior Simulation), OneSAF, and VBS2, are complex simulators which act as a conglomeration of models to produce an effective and realistic simulation of combat. For example, IWARS simulates ballistics, transmission of radio waves, personal fatigue of the soldier, the environment, and much more. IWARS simulates every aspect of a combat scenario to produce an overall realistic simulation, from which very detailed data can be collected. Because IWARS is able to simulate at such a detailed level it is known as a high resolution simulator. A high resolution simulator will provide the user the ability to tailor the simulation and collect data at a very detailed level. However, because a high-resolution simulation is so detailed, it cannot handle larger simulations (anything beyond platoon vs. a platoon sized engagements). Simulators such as OneSAF are aggregate simulators which can handle much larger and more complex scenarios, but data collection and the level of detail are much more limited. IWARS is the simulator of choice, due to its high resolution capabilities, intense data collection capacity, and specific mission detail required in the tactical scenarios.

The XM-25 weapon system is currently transitioning from its acquisition phase and entering the integration and implementation phase as units Army wide will be acquiring and using the weapon system [1]. Through the use of IWARS, a wide array of combat scenarios and different variations of the weapon's configuration can be simulated. The results may provide commanders with beneficial information regarding how they should configure the weapon system into their unit. Because the XM-25 weapon system is capable of accurate direct fire, but is too bulky for the carrier to have a powerful secondary weapon, the main concern for the weapon system is whether it should be integrated at the

Manuscript received March 23, 2012. This work was supported in part by the USMA Department of Systems Engineering, CSM (R) Wayne Batterson, MAJ John Hiltz, MAJ James Enos, MAJ William Parsons, MAJ Benjamin Morales, MAJ Joshua Gaspard, and MAJ Lawrence Nunn

Cadet Bryan Musk is with the United States Corps of Cadets, West Point, New York, 10997, USA (phone: 603-370-1725; e-mail: bryan.musk@usma.edu).

fire team level, similar to the grenadier, or in the weapon team attached in a similar fashion as a machine gun team. The optimal configuration of the XM-25 is the main focus for the simulations and analysis. Nine simulations, consisting of three different variations of the XM-25's configuration in a dismounted squad were simulated in three different scenarios. These configurations and scenarios were based from the recommendations from experienced infantry commanders. Data was then collected in order to recommend how the XM-25 should be utilized within a conventional infantry squad. The weapon is being procured and integrated into combat units, which requires a legitimate plan for how it can be tactically implemented. Pictures of the XM-25 weapon system and 25mm High Explosive Airburst Round can be viewed in Appendix I (Figure 2 and Figure 3).

An additional consideration towards the integration of the XM-25 gunner is his secondary weapon. A large overarching assumption is that secondary weapon for the XM-25 gunners will be the 9mm pistol. Due to the large size of the XM-25, carrying a rifle as a secondary weapon is infeasible. The data collected will include the survivability of friendly units and the destruction of enemy units. The accuracy and suppression radius of the XM-25's 25mm round will also be considered. Overall, this analysis illustrated how the use of combat modeling and systems engineering can be used as a real world application; in this instance for the recommendation for the XM-25's configuration into dismounted operations. More background information and history concerning the XM-25 can be found in Appendix IV.

II. Background

A. Fundamentals of Modeling and Simulation

Modeling and simulation is an effective tool for many large scale and resource intensive decisions. Often, there are numerous prototypes and variations for the optimal construction and implementation of any system. Simulation may save stakeholders and decision makers time and money. For example, a car company will conduct thousands of simulations on newly proposed vehicles before actually selecting and creating one. Simulation serves as a filtering tool to optimize the design and selection of proposed decisions and prototypes. The same principle holds true for the XM-25. Rather than spending numerous hours of field training and consuming unnecessary resources, the tactics techniques and procedures of the weapon system's configuration can first be studied in the context of simulation. This can scope how the XM-25 should be fitted into dismounted operations, which can then be further perfected in actual field training. Although the XM-25 has already been integrated into several real combat scenarios, it still has yet to be integrated across the entire Army. Simulation can serve as an aid to standardize the tactics, techniques, and procedures, of the weapon system's configuration and help units create scenarios to train with the weapon system before being deployed to combat with it.

B. IWARS and BRASS

IWARS has been selected to simulate the integration of the XM-25 because of its high resolution capabilities. It is an agent based simulator, which models the behaviors, awareness, and interaction amongst entities in the simulation. An agent based simulator focuses on modeling the individual entities to create a realistic environment and scenario for simulation [8]. IWARS also provides an intensive data output for analysis and is approved by DoD [9]. IWARS provides a multitude of terrain maps to run simulations in, along with a powerful a definitive complier known as Mission Builder, allowing the user to program the simulation for every single entity. Mission Builder utilizes a flow chart methodology to achieve simplicity as the user dictates the mission for each entity in the simulation. The user also has the ability to enter specific decision parameters which dictate to an entity when to start, end, or even restart a specific command. For example, the XM-25 gunner is given a command to switch to an M9. This command will only be executed when the entity's nearest threat is within 25 meters. Once the entity's nearest threat is outside of 25 meters, the entity will then switch back to using the XM-25. IWARS provides a clear cut Mission Builder interface which allows the user to ensure that each entity has a specified combination of commands to execute through the duration of the simulation. It also provides the user with a massive amount of data to select and analyze once the simulation is complete. IWARS is the proper simulator of choice for the XM-25's integration because of its capabilities to simulate command and control and the realistic behaviors for an XM-25 gunner.

The software BRASS (Batch Run Analysis And Simulation Studio) is the data output analysis program which runs in conjunction with IWARS. Upon completion of a simulation in IWARS, the user configures a corresponding output analysis file which specifies which data the user wishes to collect from the simulation. This data includes death, suppression level, range, fatigue, and numerous other criteria. Once the output analysis is complete, the user can instantly view the selected data from the single simulation that just ran. However, to gain a more accurate analysis of a simulation's output, the user should conduct multiple simulations and analyze the entire aggregation of data. After constructing the proper simulation and output analysis in IWARS, the user can upload these files into BRASS. The user may than select how many runs of the simulation for BRASS to analyze. BRASS will run the specified number of simulations and transfer the desired data, as prescribed by output analysis, into an Excel file. BRASS is a very powerful tool which can provide a momentous amount of data for a more conclusive analysis. All three scenarios and the three corresponding variations were uploaded into BRASS and run 20 times each. A screen shot of BRASS's interface with all the data uploaded can be viewed in Appendix II (Figure 4).

III. Methodology

A. Stakeholder Analysis

To gain a better understanding of how to properly simulate the integration of the XM-25 and what scenarios would be most realistic, stakeholder analysis was conducted. Various officers and NCOs (non-commissioned officers) who all have a background in infantry operations and even experience with the XM-25 offered their insight for the proper tactics, techniques, and procedures for the configuration of the XM-25. United States Army Majors Gaspard, Morales, Parsons, and Enos, all of which have experience as infantry platoon leaders and company commanders, but no experience with the XM-25, offered a command and control perspective towards benefits of various configurations of the weapon system. As a whole, they recommended that the XM-25 be integrated into the rifle squad (1 or 2 per squad), or insert them into the weapons team. Having the XM-25 gunners integrated into a rifle squad would mean that the XM-25 would serve a more active role (possibly in the main effort element) while conducting missions. The weapon system would be integrated with the squad as they conducted various missions to include urban ops, ambushes, hasty defenses etc. The main benefit for this configuration is that the weapon's technology would provide excellent direct fire on covered enemies as they approached the main objective. However, upon closing in on the objective, the weapon becomes ineffective once an enemy is within 25 meters. The XM-25 gunner is left with an M9 to defend himself and is essentially combat ineffective. The alternate configuration would be an XM-25 gunner integrated with a weapon team. Having the XM-25 integrated into the weapon team would mean that the XM-25 gunners would have a more specialized role in their missions. This would include a support by fire position, over watch position, or other designated roles similar to how the M240B is emplaced. However, this could lead to degradation in the squad leader's ability to employ the weapon system. Additionally, the weapon's technology may be limited as the gunners would likely remain in a stationary position, providing support by fire or over watch. All the officers agreed that the integration of the XM-25 should ultimately follow an arm's room concept, which means that the use and integration of the weapon system should be mission dependent. Additionally, they all confirmed that it should not replace the M203. The M203 is an invaluable part of each rifle squad due to its duality between the 40mm round and 5.56mm round. The M203 is a grenade launcher attached to the M4, allowing a soldier to combine the capabilities of an assault rifle with a grenade launcher.

The experienced infantry commanders also provided their input towards specific scenarios that would most realistically simulate the use of an XM-25 in modern squad level missions and illustrate specific tactical dimensions of the XM-25. They agreed that an urban environment is important to consider, largely because of the demand by coalition forces to conduct counter insurgency operations in urban environments. Within this scenario, the enemy would be hiding behind walls, fences, and windows in which a clear shot is nearly impossible. This is the ideal scenario to test the capabilities of the XM-25; however, the enemy units must be engaged before the distance of engagement closes within 25 meters. At this distance, the XM-25 is no longer combat effective and the gunner is left with an M9.

Another suggested scenario was a patrol in open terrain. In this scenario, the friendly patrol would react to enemy contact and be faced with the enemy firing from defilade positions. Enemy advantages include a drastic difference in elevation and natural elements of the environment which provide cover. This scenario tests the rapid deployment and responsiveness of the XM-25. A final scenario commonly suggested was an ambush. This scenario tests the command and control of the weapon system, along with its stopping power against a mobile enemy unit.

NCOs from 1/75 CAV and 1/427 IN, who used the XM-25 while deployed in Afghanistan also provided feedback for the integration of the weapon system. The NCOs also confirmed the notion that the XM-25 should be configured as 1 or 2 per squad, but also believed in an arms room concept as well. Having used the weapon in actual engagements, the NCOs believed it 100% effective when encountering enemies in defilade positions. Having the XM-25 as a separate attachment was also considered. They believed that 2 or 3 would be the ideal size for an attachment. Based on their experience, the XM-25 was a much better weapon system than the M203 in terms of accuracy and counter-mobility against the enemy. However, the blast radius of the M203 is much larger and more lethal than the XM-25. This group did not indicate a preference of replacing the M203 with the XM-25, rather, they endorsed the effectiveness of the XM-25 against the specific counter-defilade dismounted threat.

Stakeholder analysis served well in answering which data from the simulations would be most critical to analyze. A successful simulation is dependent on proper data analysis. The best measurements of performance in evaluating the tactics, techniques, and procedures of a new weapon's configuration into dismounted missions include survivability of friendly and enemy units and the effect (damage produced) of the round. These data measurements will serve to answer how successful the XM-25's configuration is. The best measurement of effectiveness is the accuracy of the round, in particular the range at which the XM-25 engaged enemies. This measurement unveils the specific range at which enemy units are suppressed, which provides friendly units with more freedom to maneuver. Several scenarios were created in IWARS from the recommended guidance received from the stakeholders. These scenarios include urban ops, ambush, and react to contact. Additionally, there were three different configurations (base, fire team, weapon team) of how the XM-25 was configured into the squad. Each configuration was run through each scenario. The same data measurements were analyzed for each scenario to discover which configuration consistently delivered the best results.

B. Model Scenarios and Variations

Scenario 1: Ambush

In this scenario, the friendly squad and attached weapon team conducted a linear ambush on a squad sized enemy patrol. This scenario takes place in a mountainous environment map of IWARS, which reflects Afghanistan. A 'kill zone' area feature was created for this scenario. Once the enemy patrol entered the kill zone, the fire team began to open fire, followed by the rest of the ambush element. Once fired upon, the enemy patrol reacted to contact and conducted a hasty assault on the ambush position. To realistically create the surprise element of the ambush, the enemy patrol is programmed to return fire and assault the ambush position only after being fired upon. This has to be done because it is nearly impossible to create a perfectly concealed terrain feature in IWARS for a squad to conduct a textbook ambush.

Scenario 2: React to Contact

For this scenario, the forces from the ambush scenario have switched roles. The enemy squad conducted a linear ambush on a friendly squad and weapon team patrolling in the Chorwon Valley (mountainous environment map). Once the friendly entities entered the kill zone, they were engaged. Upon receiving contact, the weapon team maneuvered to a support by fire position, while the rest of the squad returned fire and assaulted through the enemy ambush position. Once the squad passed through a designated phase line (imaginary line drawn through enemy ambush position), the weapon team then swept through the enemy ambush position to ensure that each enemy was killed.

Scenario 3: Urban Ops

In this scenario, an infantry squad conducted a hasty raid through an urban environment. Individual enemy entities were placed in various locations, hiding in alleys, windows, behind walls and on rooftops. As the squad entered the perimeter the urban environment, the weapons team maneuvered to the side of the urban environment and provided over watch as the main effort maneuvered through the streets and buildings. Once the squad passed through a designated phase line, the weapons team then swept through the urban environment.

Configuration 1: No XM-25

This variation serves as base model of squad level missions to gauge a basic squad's performance without the XM-25 weapon system. The task organization consists of a basic infantry squad (two rifle teams) along with a machine gun team.

Configuration 2: XM-25 Integrated With the Fire Team

In this variation, the XM-25 is integrated at the team level, similar to that of a grenadier. There is one XM-25 gunner placed in each team of the squad, replacing a conventional rifleman. The alternate weapon for the XM-25 gunner is an M9. By replacing the riflemen rather the grenadier, the simulation can still preserve and analyze the effect the M203 weapon system has in each scenario, which is a better representation of a realistic combat environment. Subtracting the M203 weapon system may drastically affect the squad's performance in each mission. As supported by stakeholder analysis, the M203 is an invaluable weapon system that dismounted operations often relies on and utilizes. Integrating the XM-25 at the team level requires the XM-25 gunners to take part in the main effort for the mission. As the main effort maneuvers throughout the mission, the fire team variation tests the capabilities of an XM-25 gunner to decisively engage the enemy while in a continuously maneuvering element. This variation allows the XM-25 gunners to contribute the advanced technology to the main element of the mission. This includes approaching the objective, which has the best vantage point towards the overall concentration of the enemy.

Configuration 3: XM-25 Integrated With the Weapon Team

For the final variation, the XM-25 is integrated with the attached weapon team. In the previous configurations, the weapon team consisted only of a M240B gunner and an assistant gunner. In this configuration the XM-25 gunner has an assistant gunner, armed with an M-4. An assistant gunner is common in a weapon team and provides fire support and extra ammunition for the gunner. The XM-25 gunner is playing more of a supportive role in this variation. This includes support by fire and over watch for the main effort during the different scenarios. This variation provides more stationary time for the XM-25 gunner to decisively engage targets and provide suppression on enemy units as the main effort maneuvers towards the objective. However, a support by fire role could possibly limit the number of decisive engagements in which the technology of the XM-25 is utilized.

C. Assumptions and Limitations

Although IWARS may be DoD approved and provide a relatively realistic simulation for analysis, it is still limited in its capacity to truly reflect a real combat scenario. For example, the entities in the simulation will only execute exactly what they are told. The entities lack the ability to make swift changes in their decision making, unless specified by parameters, which are inputted into the Mission Builder. However, there are only so many possible parameters the user could input. There will always be instances in each simulation in which an entity lacks an element of realism to make an obvious decision due to limited awareness and intelligence. Essentially, the entities lack the ability to think on their own. The biggest limitation along these lines is an entity's ability to care for a wounded comrade. During the simulation, the entities will continue their mission and disregard a fellow wounded soldier. Although this is a huge stretch from reality, it does not hinder the simulation from collecting valid data to analyze the XM-25's performance. It actually helps, by saving time and focusing the simulation towards the critical engagements from which data is collected. Additionally, the secondary weapon for the XM-25 gunners is the M9. This seems to be the most realistic option, but will also have a strong affect on the data, because the gunners will use the M9 for any enemies that are within 25 meters. The task organization for

each configuration of the XM-25's integration remained fairly constant. There were three variations: a base model (without the XM-25), fire team integration, and weapon team integration. Each variation consisted of a squad sized element conducting three different missions: Urban Ops, Ambush, and React to Contact. Each XM-25 gunner has an M9 as his alternate weapon. The integration for the XM-25 will vary for each simulation.

Several assumptions concerning the simulation were also made. In each simulation we must assume there is an open flow of communication and shared awareness amongst the entities. IWARS permits the entities from the same force to communicate with each other. Sharing known agents and other aspects of communications will provide a corresponding level of situational awareness amongst the entities from the same force in the simulation. Screenshots for each simulation can be viewed in Appendix III (Figure 5, Figure 6, and Figure 7).

IV. Results

After using BRASS to batch nine different types of simulations (three different configurations in three different scenarios), twenty times each, the results were thoroughly analyzed. Below, in Figure 8 displays the results concerning the average survivability of friendly (blue) and enemy (red) units in each simulation.

Base	
Ambush	
AVG Blue Killed	AVG Red Killed
2.25	12.6
React to Contact	
AVG Blue Killed	AVG Red Killed
4.6	11
Urban	
AVG Blue Killed	AVG Red Killed
3.7	6.15

Fire team	
Ambush	
AVG Blue Killed	AVG Red Killed
1.6	12.9
React to Contact	
AVG Blue Killed	AVG Red Killed
5.55	11
Urban	
AVG Blue Killed	AVG Red Killed
4.3	6.2

WPN SQD	
Ambush	
AVG Blue Killed	AVG Red Killed
1.5	12.9
React to Contact	
AVG Blue Killed	AVG Red Killed
4.15	11
Urban	
AVG Blue Killed	AVG Red Killed
4.4	6.3

Fig. 8 Average Survivability Data for each Simulation.

The data reflected the same recommendation from stakeholder analysis: the arm's room concept. The optimal configuration seemed to be based on the scenario. Simulations showed that the ideal configuration for the XM-25 is mission dependent.

Overall, the results for survivability of friendly and enemy units were not as conclusive as desired. The data delivered by a single configuration was not consistently the best. The weapon squad configuration seemed to have performed the slightly stronger than the other variations except for the base variation in the urban scenario. This is likely due to the close quarters of the urban scenario and the fact that the XM-25 gunner was left to use his M9. The weapon team variation's performance in the urban scenario for the survivability criteria is only marginally better than the other variations and by no means provides a definitive answer for how to best integrate the weapon system. The data concerning survivability strongly suggests a weapon team integration. This is probably due to the fact that the XM-25 gunner had an assistant gunner. The XM-25 is a fairly complex weapon system compared to the conventional weapons of the typical infantry squad and would be best placed in a weapons team where the user receives more training. This will allow for a better performance of the weapon system and ensures its maintenance is conducted properly. It seemed the only scenario in which the XM-25 system may hinder overall performance would be in a close quarter urban environment, which is likely due to the close range and lack of ability to effectively use the XM-25. The XM-25 gunners switched to an M9 as they maneuvered through the close quarters of the buildings which effectively created a combat ineffective soldier. This is a growing concern for the carrier of the XM-25. Because the weapon system is so large, the gunner cannot carry an assault rifle for close quarters. XM-25 would be great to take out enemies and clear rooms through the windows of buildings from afar, but in close quarters (less than 25 meter), it is ineffective. The survivability data seems to support the stakeholder analysis, which proposes that an arms room concept should be used when deciding how to integrate the weapon system. However, coalition units will not always be able to predict when they receive contact from the enemy. It is important for commanders to ensure the XM-25 weapon system is configured properly in the squad when contact is likely. The commander must use discretion when deciding to use the weapon system based on the nature of the mission. Some other important data points of consideration from results of the simulations include the suppression radius, average range at which enemies were hit by explosive munitions, and the maximum range at which enemies were hit by explosive munitions. These are the measurements of effectiveness, which provide data concerning how effective the XM-25 was at suppressing the enemy. The range data collected refers to the range at which an explosive munitions damaged an enemy entity. The suppression radius data refers to the radius from the point of explosion at which an entity may be damaged or temporarily hindered from completing any sort of action. Below, Figure 9 displays the results of the simulations concerning the average and maximum range (meters) and suppression radius of the respective explosive munitions; 40mm from the M203, and 25mm from the XM-25. The M203 weapon system's performance was compared to that of the XM-25 because it allowed for a more thorough assessment of explosive munitions. It creates a relative and measurable standard for the XM-25 to be compared to. The data concerning the range of the explosive munitions can appear somewhat misleading. Because the 40mm round from the M203 has such a high suppression radius, it also has a larger casualty producing radius. Thus demonstrates that a grenadier can cause damage to an enemy entity at a greater

Suppression Radius (meters)	
XM-25	30
M203	61

Base	
Ambush	
AVG RANGE M203	AVG RANGE XM-25
129.14	-
M203 MAX Range	XM-25 Max Range
197.68	-
React to Contact	
AVG RANGE M203	AVG RANGE XM-25
101.60	-
M203 MAX Range	XM-25 Max Range
111.04	-
Urban	
AVG RANGE M203	AVG RANGE XM-25
92.94	-
M203 MAX Range	XM-25 Max Range
153.59	-

Fire team	
Ambush	
AVG RANGE M203	AVG RANGE XM-25
147.21	157.87
M203 MAX Range	XM-25 Max Range
197.59	198.57
React to Contact	
AVG RANGE M203	AVG RANGE XM-25
80.95	0.00
M203 MAX Range	XM-25 Max Range
116.91	0.00
Urban	
AVG RANGE M203	AVG RANGE XM-25
103.74	79.03
M203 MAX Range	XM-25 Max Range
201.72	122.07

WPN SQD	
Ambush	
AVG RANGE M203	AVG RANGE XM-25
113.66	104.54
M203 MAX Range	XM-25 Max Range
195.79	196.29
React to Contact	
AVG RANGE M203	AVG RANGE XM-25
90.31	0.00
M203 MAX Range	XM-25 Max Range
118.12	0.00
Urban	
AVG RANGE M203	AVG RANGE XM-25
96.79	53.50
M203 MAX Range	XM-25 Max Range
203.44	83.35

Fig. 9 Suppression Radius and Range Data for Explosive Munitions. Average Survivability Data for each Simulation.

range even though it fired an inaccurate shot. Additionally, each scenario and variation placed the XM-25 gunners in different positions relative to the enemies, while the position of the grenadiers remained constant. It is noted in the react to contact phase that the XM-25 gunners were not able to fire a single shot. To remedy this, the path of maneuver was altered for the XM-25 gunners. However, the concentration of gun fire on the XM-25 gunner was too strong to make an effective change. This is because the enemy entities in IWARS considered the XM-25 gunner the highest threat and always targeted the XM-25 gunners first. In the urban scenario, the XM-25 had better range data while part of the fire team, this is likely due to the route of movement into the urban environment by the main effort. As they entered, they had much better line of sight towards rooftops at a greater range than the weapons team which established an over watch position by the side of the buildings. The ambush scenario seemed to perform fairly even for both the fire team and weapon team variations. The fire team did have a higher average range, but there were two XM-25 gunners in the fire team variation as opposed to only one in the weapon team variation. The max ranges for the XM-25 in both of these variations were nearly the same. The range data seems to support the same recommendation as the survivability data. The data further suggests that an arms room concept should applied by commanders. Missions involving close quarters will less likely need the XM-25, while open environment missions would greatly benefit from it.

V. Conclusion

The results clearly support the recommendations and input received from stakeholder analysis. Effectively, the configuration of the XM-25 into a dismounted squad or even the decision to use it is mission dependent. An urban scenario, with the high likelihood of enemy contact in less than 25 meters, is not favorable for an XM-25 gunner. However, an ambush or patrol in rolling terrain, full of cover for the enemy, is ideal for the XM-25 gunner. Should the XM-25 gunner have an AG at all times, the playing field for the weapon system will open up to many more scenarios. This is only initial research and by no means definitive enough to establish doctrine on how the XM-25 should be configured. However, it sheds light on the fact that commanders must understand the versatility of the modern battlefield and best equip their soldiers to successfully destroy the enemy. The best recommendation for the configuration of the XM-25 is completely mission dependent.

This analysis has displayed how combat modeling is an effective tool to evaluate weapon systems, before they are fully integrated in an actual combat environment. However, the limitations of the simulations and the software may sometimes prevent a realistic trend or notable occurrence from being observed. For example, in the react to contact scenario, the XM-25 gunners were immediately targeted by enemy entities largely due to the fact that the XM-25 gunner posed the largest threat. Because the enemy entities communicate a shared awareness of the highest threat, they quickly focus their efforts towards the same threat. Further research and work could be conducted to create ideal parameters for the decision making process for how entities decide to engage a target. IWARS allows the user to designate weighted values considering different criteria (range, threat, line of sight, fatigue, etc.) to consider when an entity decides which enemy to engage. A Design of Experiment could be conducted to determine the best combination of weighted values for a realistic and decisive engagement to occur. A design of experiment could also be conducted to discover the best secondary weapon for XM-25 gunners.

Even though the simulation may have its shortcomings and limitations, it still is a vivid display of how combat modeling and simulation saves time, money, and resources. The

integration of the XM-25 will likely be a very versatile process. Commanders must understand the weapon system and its capabilities. They must have properly trained soldiers who can understand how the decision to use the weapon on a mission plays a large role in the commander's intent. The numerous variations and scenarios in IWARS made this observation apparent. Simulations and analysis have illustrated how combat modeling and systems engineering can be applied to practical real world problems. IWARS provides the military the fortunate option to test numerous ideas and prototypes before experimenting with them on the battlefield, where much more then time and basic resources are on the line.

APPENDIX I

XM-25 Weapon System Photos

Fig. 2. XM-25 Weapon system and 25mm HEAB Round

Fig. 3. Soldier Aiming the XM-25 Weapon System

APPENDIX II
BRASS SOFTWARE

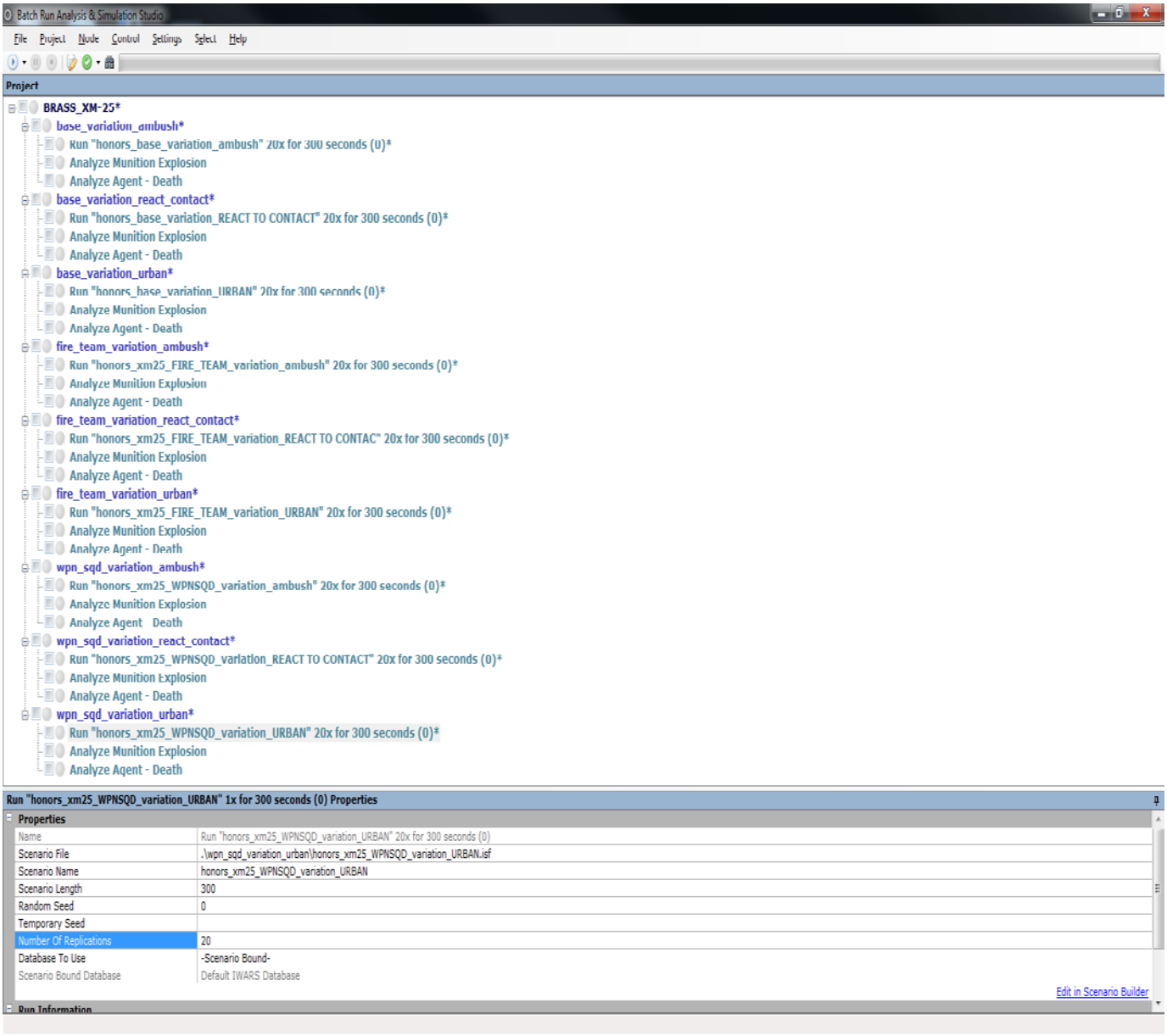


Fig. 4. BRASS Conducting Batch Runs

APPENDIX III
SIMULATION SCREEN SHOTS

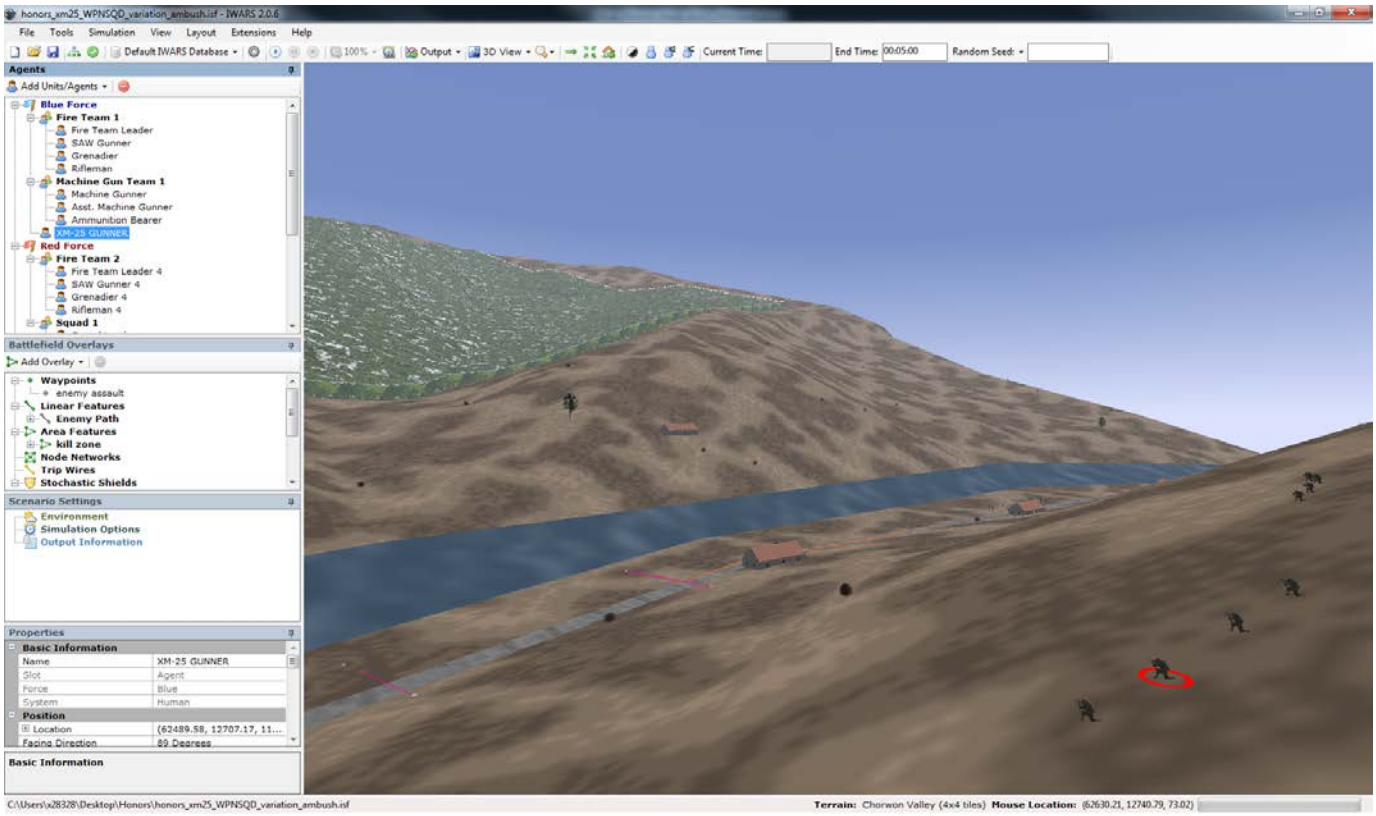


Fig. 5. A screenshot of the ambush scenario and weapon team variation. The friendly ambush position is located on the right side of the screenshot. They are waiting for the enemy patrol to enter the kill zone. The zone is identified by the magenta outlined box on the road. The enemy patrol is following the road. Once they are fired upon as they enter the kill zone, the enemy patrol will return fire and attempt to assault the ambush position. The weapon team is located closer towards the screen shot, and the fire team is the further element from the screen shot along the hill. The XM-25 gunner is identified by a red circle. In the react to contact scenario, the enemy entities and friendly entities will switch roles.

APPENDIX III CONTD.
SIMULATION SCREEN SHOTS

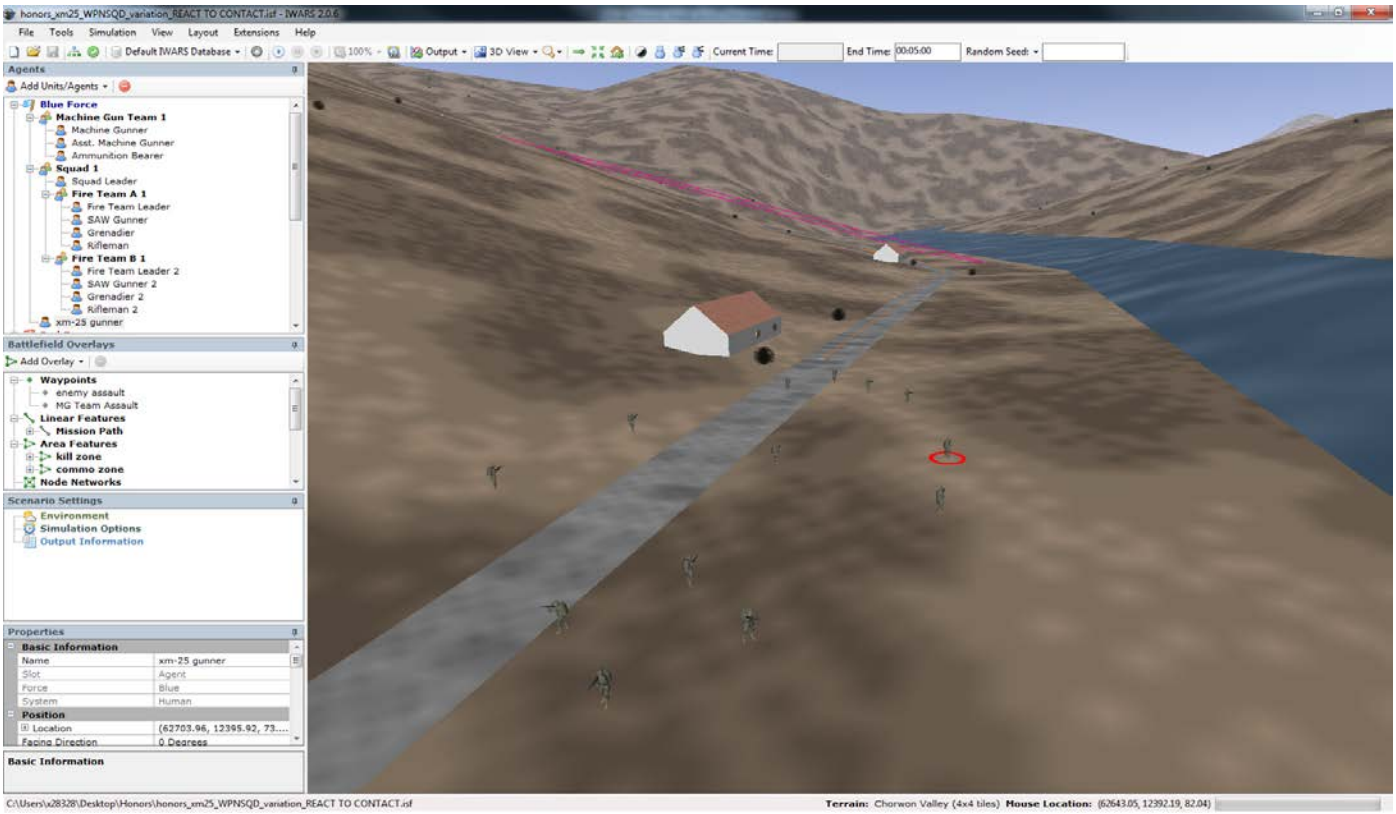


Fig. 6. A screen shot from the react to contact scenario. The friendly patrol is about to enter the kill zone and react to contact from an enemy ambush on the hill side. This is the weapon team variation. The XM-25 gunner is identified by a red circle. In the react to contact scenario, the enemy entities and friendly entities will switch roles.

APPENDIX III CONTD.
SIMULATION SCREEN SHOTS

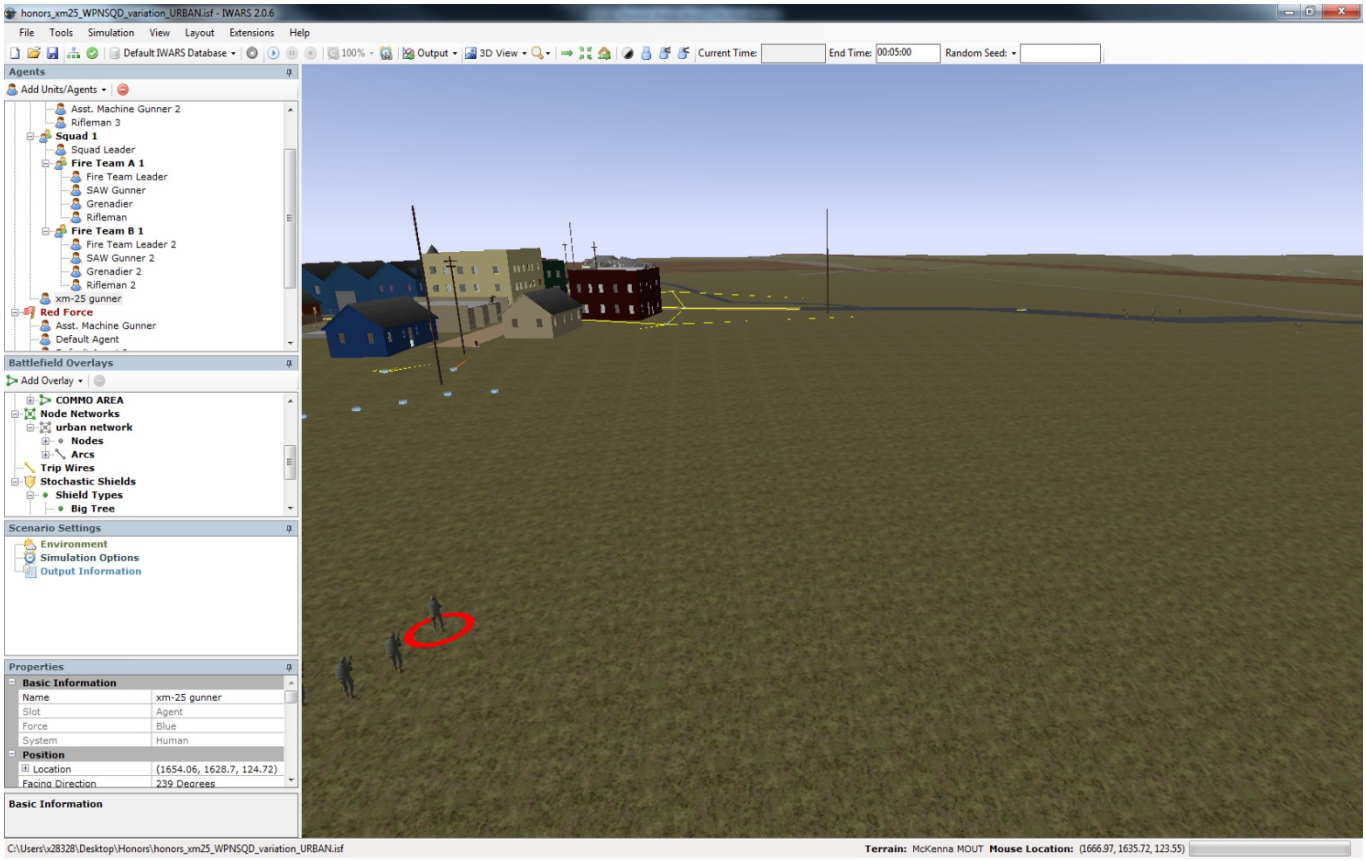


Fig. 7. A screenshot of the urban scenario and weapon team variation. The fire team is located towards the bottom left of the screen shot. They will approach the urban environment from the side and provide over watch. The fire team elements will approach the urban environment from the road. Once the fire team passes the designated phase line, the weapon team will sweep through the urban environment. Enemy entities are scattered throughout the buildings and streets.

APPENDIX IV
XM-25 INFORMATION AND HISTORY

The XM-25 is a counter defilade target engagement system. It is an advanced direct-fire grenade launcher weapon system. The XM-25 is a unique advanced weapon due to its target acquisition and fire control capabilities. It is semiautomatic, magazine fed, and fires a 25 mm round. The XM-25 is capable of shooting high explosive air burst rounds. Essentially, the 25mm round will detonate at a specified range rather than on contact. This is all possible due to the laser rangefinder system of the XM-25. The laser rangefinder acts as a cross-hair for the shooter. Upon acquiring his target, the shooter presses the rangefinder button and the distance (meters) will be displayed for the shooter. The shooter then has the option to add or subtract distance to the lased target to determine where the round will detonate. For example, if an insurgent is behind a rock 100 meters away, a XM-25 gunner would laze the rock. He would then be able to add 2 meters to that distance. The gunner can then aim just above the rock and fire. After traveling the desired distance (just above the rock and 2 meters beyond it) it will detonate. The resulting air-detonated blast from the round at the desired distance will destroy the enemy behind the rock. This drastically reduces the enemy's ability to use cover on the battlefield. "With the XM-25, Soldiers don't have to actually hit that vital area to dispatch the enemy, they only have to aim the launcher's air burst fragmentation warhead nearby. The warhead's blast is equivalent to a hand grenade" [1]. Ideally, the XM-25 is perfect for scenarios in which the enemy is hiding in a defilade position in which a clear shot is not possible. This is becoming more and more of an issue as coalition forces encounter the enemy in a wide variety of complex battlefields. The capabilities of the XM-25 are an incredible solution to this growing challenge.

The XM-25 was designed by Heckler & Koch and Alliant Tech systems [1]. By 2010, the weapon system began to be field tested in Afghanistan. Its performance and capabilities resonated very well with soldiers. 1/75 CAV utilized the XM-25 while deployed and quickly realized the benefits of the XM-25 airburst technology. SFC Shriver noted the weapons performance by saying, "There have been multiple engagements where the enemy knows, 'Hey, I'll hide behind this giant boulder. No matter what you hit me with you can't get through this boulder.' With the Counter Defilade Weapon System - it is invaluable. You can hit anybody within range, behind anything" [3]. The XM-25 has served as a valuable asset for infantry units in firefights. The tactics of using large cover to draw out coalition units into open space on the battlefield is eliminated through the XM-25's capabilities.

The XM-25 is continually compared with its counterpart, the M203 grenade launcher. The M203 is an indirect fire grenade launcher which shoots a 40mm round. It has a larger blast radius, but is far less accurate then the XM-25[3]. However, the M203 has dual capability because it is integrated with the M4. Below, Figure 1 displays a comparison between the accuracy of the XM-25, M4, and M203 [3]:

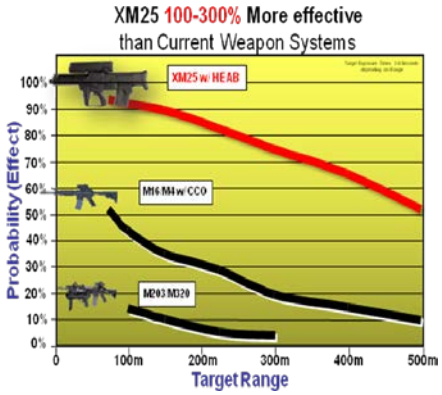


Fig. 1 Accuracy comparison between the XM-25, M4/M16, and M203

Figure 1 clearly illustrates how accuracy is the strongest quality of the XM-25. 1LT Featherman shared his experience in observing the accuracy of the XM-25 while in an engagement with the enemy, "The engagement we had was perfect. We had a 40mm [M203] right next to the XM-25, and even with a fresh XM-25 gunner, the difference in accuracy of the rounds was astounding. The 40mm shot high twice and low 3 times, never really getting a single shot on target. Whereas the XM-25 put the first 3 rounds right where they needed to be". The XM-25 is beginning to be utilized onto today's battlefield and the Army's leadership is quickly learning the capabilities and benefits of the weapon system.

APPENDIX V
TASK ORGANIZATION

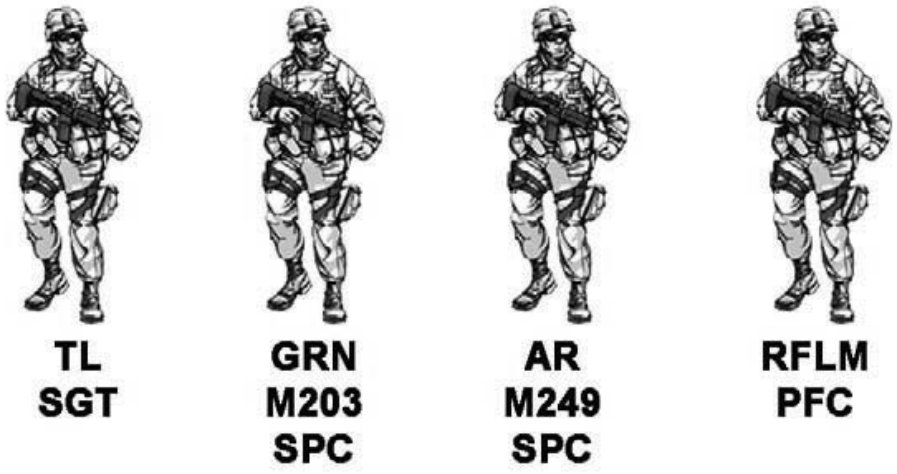


Fig. 10. A conventional infantry rifle team [10]. The rifle team consists of the Team Leader (TL) who carries the M4 (5.56mm assault rifle), the grenadier (GRN) who carries the M203 (a 40mm grenade launcher attached to the M4), the automatic rifleman (AR) who carries the M249 (a 5.56mm automatic machine gun), and the rifleman (RFLM) who carries the M4. An infantry squad consists of two rifle teams. The XM-25's configuration in the fire team would replace the rifleman with the XM-25.

Fig. 11. A conventional machine gun team attached to a squad [10]. The team consists of the machine gunner, who carries the M240B (a 7.76mm automatic machine gun) and an assistant gunner (AG) who provides extra ammunition, close quarter protection for the team, and carries the M4. The XM-25's configuration in the weapon team configuration would add an XM-25 gunner and an AG to the weapon team.

References

[1] Grant, Greg. “Army Sending Precision Grenade Launcher to Afghanistan.” Defense Tech. March, 1, 2012 <http://defensetech.org/2010/05/06/army-sending-precision-grenade-launcher-to-afghanistan/>

[2] Gaspard, Joshua. Personal Interview. West Point, NY. 15 Feb. 2011.

[3] Hiltz, John. Presentations and data collected on the XM-25 used intensely for initial research. West Point, NY. 15 Feb. 2012.

[4] Enos, James. Personal Interview. West Point, NY. 15 Feb. 2012. Parsons, William. Personal Interview. West Point, NY. 15 Feb. 2012.

[5] Morales, Benjamin. Personal Interview. West Point, NY. 15 Feb. 2012.

[6] Batterson, Wayne. Personal Interview and Assistance on setting up simulations. West Point, NY.

[7] Borshchev, Andrei. Filippov, Alexei. “System Dynamics and Discrete Event to Practical Agent Based Modeling: Reasons, Techniques, Tools”. XJ Technologies and St. Petersburg Technical University. March 1, 2012.

[8] IWARS Methodology Guide. Version 2.0.1. October 2009.

[9] FM 3-21.8. The Infantry Rifle Platoon and Squad. March 2007 Headquarters Department of the Army.

Author Information

Bryan Musk, Cadet, USMA Class of 2012, bryan.musk@usma.edu

Marc Wood, LTC, USMA, Instructor, marc.wood@usma.edu

A Next-Generation Emergency Response System for First Responders using Retasking of Wireless Sensor Networks

Syed R. Rizvi, Chinmay Lokesh, Michele C. Weigle and Stephan Olariu, *Old Dominion University*

***Abstract*— The objective of our research is to design a real-time information system to improve emergency-response functions by bringing together information to respond to a terrorist attack, natural disaster or other small or large-scale emergency. We call this system *ALERT: An Architecture for the Emergency Retasking of Wireless Sensor Networks*. The novel contribution of this research to the emergency response strategies is the seamless integration of various wireless sensor networks by *retasking* them with explicit missions involving a dynamically changing situation. Our thesis is that judicious re-tasking of independently-deployed sensor networks, working in tandem with an effective sensor capability integration scheme, will lead to improved emergency-response functions and a seamless return to normal conditions. Preliminary results have shown that retasking sensor networks for emergency response is a promising new paradigm that can not only promote a wider adoption of sensor network systems in support of guarding our national infrastructure and public safety, but can also provide invaluable help with disaster management and search-and-rescue operations.**

***Index Terms*— Emergency Response, Aggregation and Forwarding Node (AFN), Patrol, Search and Rescue (PSAR), and Wireless Sensor Networks (WSN).**

I. Introduction

On August 23, 2011 an earthquake struck the East Coast of the U.S., centered near Richmond, Virginia with a magnitude rating of 5.9. Tremors were felt all throughout the Mid-Atlantic and Northeast. On the previous day, Colorado was hit with a 5.3 magnitude earthquake, the state's largest in decades. A series of minor earthquakes hit Northern and Southern California during the same period, but the largest in this period was the one centered near Richmond. Though no injuries or damage were reported, several buildings were evacuated. Had the earthquake been a serious one, the injury, loss of life, and property damage associated with it would have been enormous similar to the tragic loss during and after the massive earthquake and tsunami that shook Japan in March 2011. Its impact became an international issue because it was the cause of release of radiation from the Fukushima Daiichi nuclear power station. An initial review of the Japanese response in four critical areas suggests important lessons for the whole world when evaluating national and international capacity to deal with catastrophes. These four critical areas are preparedness and response, communicating risk, international assistance, and critical infrastructure.

Fig 1. Rescue workers search for Japan earthquake victims in *Tamura, Iwate Prefecture* [8].

The objective of our research is to design a real-time information system to improve emergency-response functions by bringing together information to respond to a terrorist attack, natural disaster or other small or large-scale emergency. We call this system *ALERT: An Architecture for the Emergency Retasking of Wireless Sensor Networks*. An example of such an emergency response function is a search-and-rescue operation performed by first responders. Typically, Wireless Sensor Networks (WSNs) composed of a large number of nodes, with processing, sensing and radio communication capabilities, scattered throughout a certain geographical region, have been applied to many real-world problems. Remote monitoring applications have sensed animal behavior and habitat, structural integrity of bridges, volcanic activity, and forest fire danger, to name only a few successes.

Manuscript received April 13, 2012. This work was supported in part by the National Science Foundation under CNS 1116238.

Syed R. Rizvi is a PhD student at the Computer Science Department of Old Dominion University, Norfolk, VA 23529 USA (srizvi@cs.odu.edu).

Chinmay Lokesh is a MS student at the Computer Science Department of Old Dominion University, Norfolk, VA 23529 USA (clokesh@cs.odu.edu).

Michele C. Weigle is an assistant professor in Computer Science of Old Dominion University, Norfolk, VA 23529 USA (mweigle@cs.odu.edu).

Stephan Olariu is a tenured full professor in Computer Science of Old Dominion University, Norfolk, VA 23529 USA (olariu@cs.odu.edu).

These networks leveraged the relatively small form-factor of motes and their multi-hop wireless communication to provide dense sensing in difficult environments. Recently, there has been a surge of interest in sensors and sensor networks. We refer the reader to [2,3,4,6,9] for recent results and applications. In our system, the critical role of monitoring various parts of national infrastructure, from government buildings to power plants, to bridges, roads and tunnels is achieved through sensor network technology. The novel contribution of this research to the emergency response strategies is the seamless integration of various WSNs by retasking them with explicit missions involving a dynamically changing situation. This means that under normal conditions the sensor networks monitor the specific attributes for which they were deployed (e.g., air quality, temperature, pressure, noise levels). This is where heterogeneous sensor networks come into existence. The deployment of heterogeneous sensor networks in the real world is inevitable due to their specific objectives, and increase in reliability without significantly increasing the cost. However, should an emergency occur, the sensors in the affected area must be retasked and integrated into an emergency response system. Authorized personnel could task the sensor networks with explicit missions in support of mitigating the emergency at hand.

II. Related Work

Today's first responders rely primarily on hand-held high-frequency radios for communication among themselves. First responders often do not have access to hazardous materials information in a particular structure, let alone the real-time information about the areas of interest. We envision a waterproof, heat-resistant, wrist pouch device, wirelessly linked to heterogeneous wireless sensor networks deployed in the affected area, that would enable first responders to access just-in-time information that is relevant to their current mission.

The Sahana Free and Open Source Disaster Management System, one of the famous disaster management systems well known for its comprehensive approach to manage information in response to disasters, was conceived during the 2004 Sri Lanka tsunami. The system was developed to help manage the disaster and was deployed by the Sri Lankan government's Center of National Operations (CNO), which included the Center of Humanitarian Agencies (CHA). The main drawback of this system is that it does not address the instant access to the sensed information needed by the first responders [5]. There appears to be no electronic technology solution currently available to address their needs. In other words, there are no comprehensive solutions that involve emerging technologies like wireless sensor networks for role-specific, mission-specific information gathering and sharing for first responders. Additionally, there are no widely accepted design principles for retasking independently-deployed sensor networks and for integrating their capabilities.

Our work presents an important step towards adaptive and scalable computing architecture. Results have shown that retasking sensor networks for emergency response is a promising new paradigm that can not only promote a wider adoption of sensor network systems in support of guarding our national infrastructure and public safety, but can also provide invaluable help with disaster management and search-and-rescue operations. Our research will have a broad societal impact as sensor networks are expected to be integrated into the fabric of the society. Large geographical areas will be provided with integrated sensor networks that can provide invaluable help with disaster management. Our research can be readily extended in support of detecting trends and unanticipated events, the two key ingredients of an early-warning system.

III. Network Model

We anticipate that in the near future a multitude of stand-alone wireless sensor networks will be deployed in the same area of interest (AoI) by various infrastructure providers in support of application-specific missions. These sensor networks are populated by massive number of tiny, commodity sensors deployed for an aircraft or possibly embedded in the asphalt covering access roads, the surface of a bridge, etc. Due to the massive deployment, the exact location of individual sensors is a priori unknown. Alongside with the tiny sensors, more powerful devices referred to as aggregation and forwarding nodes (AFNs), are also deployed. The AFNs are

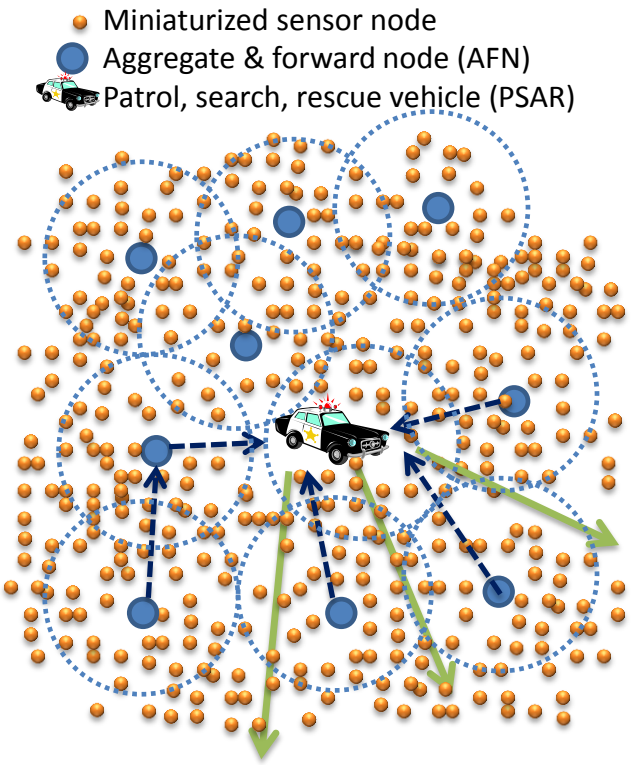


Fig 2. Illustrating sensors, AFNs and PSAR.

endowed with special radio interfaces for long distance communications, miniaturized GPS, and appropriate networking tools for data collection and aggregation. AFNs may be stationary or mobile. We assume that all sensors within radius *R* of an AFN are able to receive messages directly from the AFN and, likewise, that the AFN is able to receive messages directly from the sensors. The AFNs can organize the sensors in their immediate vicinity into a dynamic virtual infrastructure supportive of the overall mission of the network. Additionally, we assume that all AFNs within a certain radius of a patrol, search and rescue vehicle (PSAR) are able to receive messages directly from the PSAR and, likewise, that the PSAR is able to receive messages directly from the AFNs [7]. In Figure 2, a PSAR is communicating with the AFNs that are in its range which are in the front of the PSAR.

Each sensor may have sensory capabilities c_1, c_2, ..., c_k (such as temperature, humidity, or motion), each with its own technology-dependent resolution. The sensors also posses on-board resources (such as energy budget, CPU clock rate, memory size), which are distinct from their sensing capabilities. There are various techniques that allow the sensors to acquire a course-grain location awareness in terms of the sub-region of the AoI in which they reside. In turn, this allows the sensors to be addressed by the ID of their sub-region [1,2,3]. In addition, assuming that each sensing capability has an identifier describing its type, we assume a sensor naming scheme based on sub-region and sensing capabilities. The capability envelope (CE) of the sensor network system in the AoI defines the joint set of capabilities that the sensors, originally deployed in support of individual missions posses collectively, given that their capabilities are integrated. Since the various wireless sensor networks were deployed independently of each other by different infrastructure providers, and since their openers do not necessarily act in convert, the CE of the system is usually unknown.

In the normal mode of operation, each of the sensor networks attends to the specific mission in support of which it had been deployed. However, should an emergency occur (e.g. fire, chemical spill, hurricane, terrorist attack), the sensors in the co-located networks must participate in the high-priority tasks inherent to addressing the emergency situation. This may involve identifying survivor, directing first aid to the wounded, monitoring contamination levels, and so on. In this exceptional mode of operation, the corporate capabilities of the sensor network system should be made available to authorized in-situ users (e.g. first-responders) that may be in touch with a remote entity (e.g. mission control). The in-situ users have the authority to act as mobile BS, tasking the sensors directly.

IV. WORKING SCENARIO

To illustrate our technical discussion and the approaches proposed by ALERT, we refer to the following working

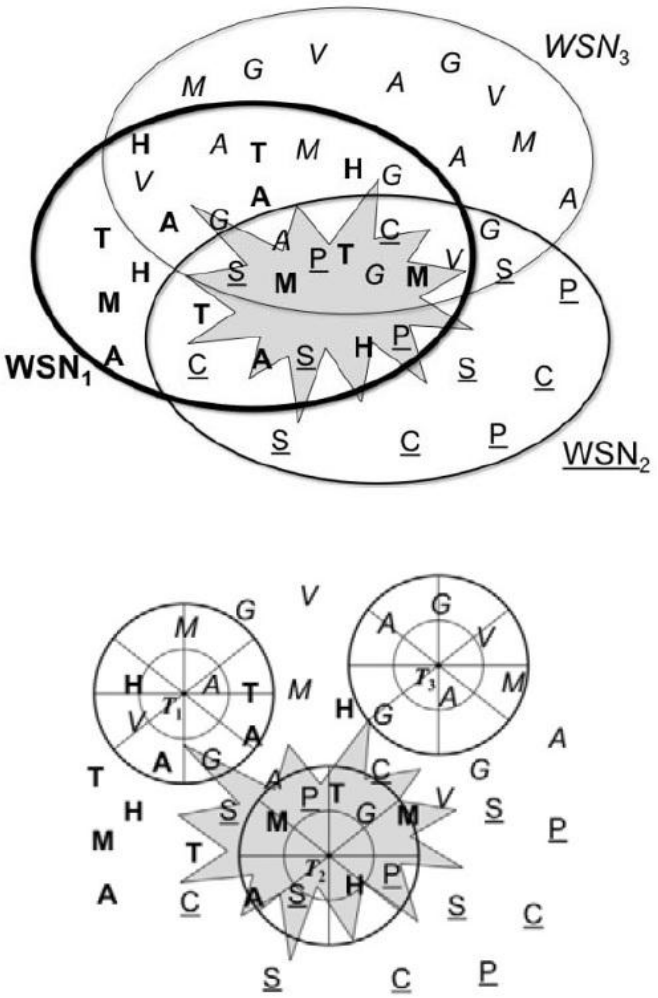


Fig 3. Three non-integrated *WSNs* (top). The same *WSNs*, with regions near first-responders *(T_i)* integrated by the virtual infrastructure (bottom).

scenario. As shown in Figure 3, assume that three wireless sensor networks WSN_1, WSN_2, and WSN_3 have been independently deployed over a common Area of Interest (AoI) in the vicinity of a chemical plant. To keep things simple, we assume that each sensor can sense only one attribute of the environment. WSN_1 deployed for environmental monitoring includes temperature (T), humidity (H), motion (M), and acoustic (A) sensors; WSN_2 deployed for pollution monitoring includes air pressure (P), smoke (S), and chemical (C) sensors, and WSN_3 deployed for traffic monitoring contains video (V), acoustic (A), motion (M), and GPS (G) sensors. Let us assume that there was an explosion at the chemical plant. As a result, the factory is on fire, and thick smoke, toxic gases and other hazardous chemicals were released over a large surrounding area. The first indications of problems are increased levels of smoke (S) and chemicals (C) detected by sensor network WSN_2. As a result, a helicopter was dispatched to the area to determine what has happened. On board the helicopter there are a number of first responders who, collectively, have the authority to mandate the transition of the ALERT-based sensor network system to the exceptional mode of operation. In the face of the disaster, the three networks must be retasked, morphing into short-lived mission-driven

sensor networks in support of the following high priority missions:

1. Assess the extent and potential effect of hazardous materials;
2. Identity and locate survivors;
3. Detect the wounded and assess their condition;
4. Guide rescue teams to the tapped and the wounded;
5. Detect fires, and identify directions of their spreading;

TABLE I
ILLUSTRATING CAPABILITY REASSIGNMENT TO EMERGENCY MISSIONS

	Types of Sensors								
Mission	T	H	V	A	G	M	C	P	S
1	X	X					X	X	X
2				X	X	X			
3			X	X					
4			X		X	X	X		
5	X	X						X	X
6			X				X		X

6. Detect damage power lines, leaks from gas lines and hazmat spills.

A glance at Table 1 confirms that the set of sensor capabilities required for performing missions 1-6 is feasible for the sensor network system in our scenario. In addition, Table 1 captures a possible reassignment of sensor capabilities in order to serve individual missions. For example, first of all, an assessment of the situation is needed (Mission 1). Mission 1 requires capabilities {T, H, C, P, S} which are included in set of capabilities provided by the entire network. Similarly, the task of detecting fire (Mission 5) is predicated on the fact in order for fire to be present four conditions must be met: the temperature must exceed a certain threshold, humidity must be below a certain threshold, barometric pressure must be low and there must be smoke. Notice that, originally, temperature sensors, humidity sensors, pressure sensors and smoke detectors were serving separate missions. In order to detect the presence of fire, they have to be grouped into a mission-driven network.

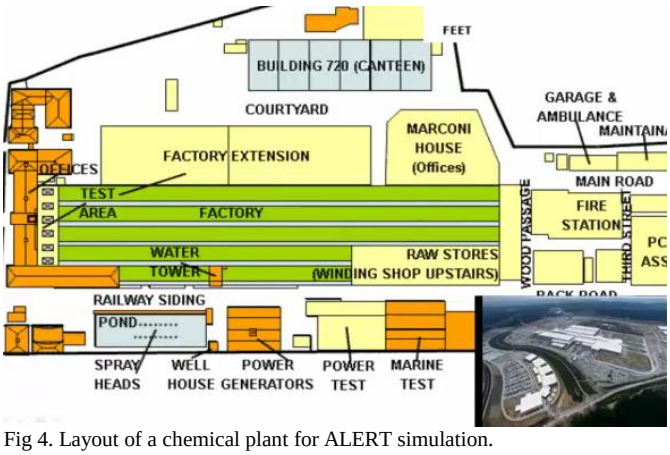


Fig 4. Layout of a chemical plant for ALERT simulation.

V. SIMULATION AND RESULTS

Figures 4-6 illustrate a Flash-based model for the above

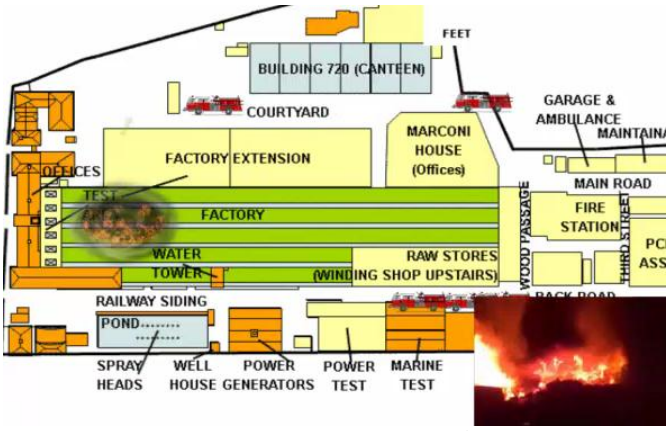


Fig 5. Simulating an explosion at the Test Area hall of the chemical plant.

mentioned scenario that is simulated using Flash MX ActionScript. Figure 4 shows the layout of a chemical plant where an explosion occurs (Figure 5). Consequently, the factory sets on fire, and thick smoke, toxic gases and other hazardous chemicals are released over a large surrounding area. Figure 6 shows in green arrows the establishment of evacuation paths in order to accomplish mission 4, i.e. to guide

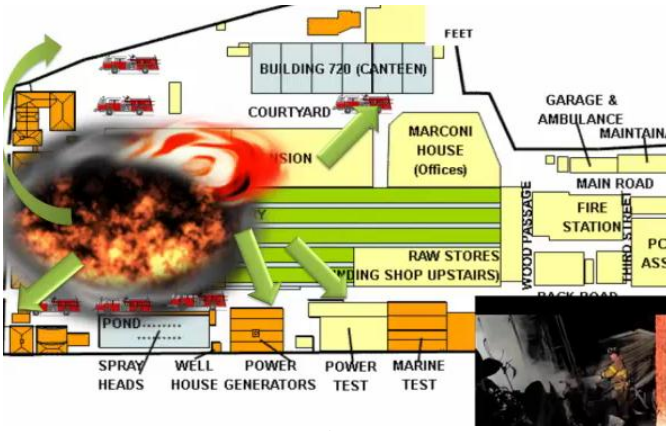


Fig 6. Simulating the establishment of evacuation path using ALERT.

rescue teams to the tapped and the wounded.

A small section of this scenario was implemented independently using the crossbow hardware consisting of IRIS motes (Figure 7), MTS310 sensor boards, MIB520

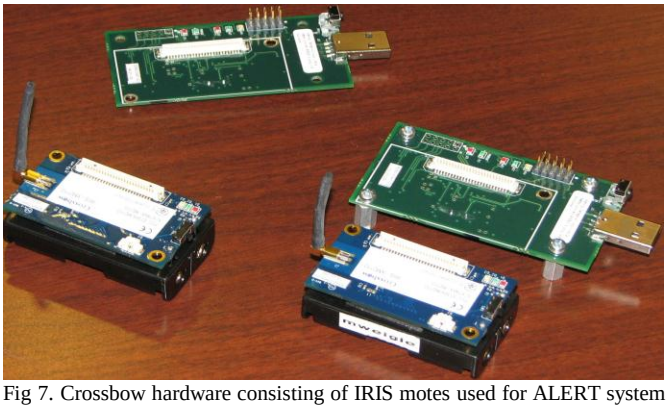


Fig 7. Crossbow hardware consisting of IRIS motes used for ALERT system validation.

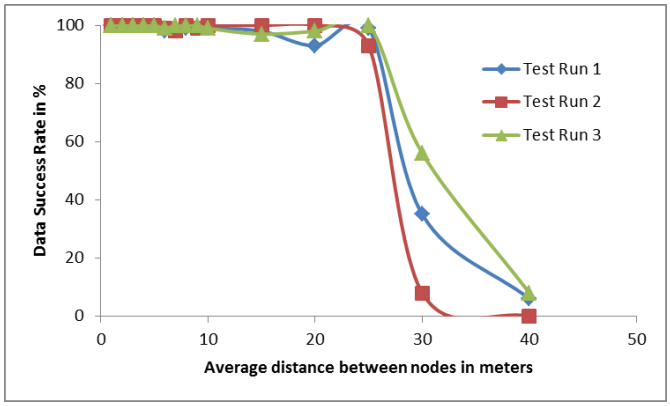


Fig 8. Data success rate verses average distance between wireless nodes.

programming boards using nesC programming on the TinyOS platform. Correctness of data obtained as a response of AFN's query was translated into Data Success Rate. The average distance between the sensor nodes is varied in order to determine the fall off and drop off levels. As can be observed in Figure 8, the fall off occurs at about 25 meters, and the drop off occurs at about 40 meters.

VI. Conclusion

The effectiveness of the ALERT system is dependent on the basic transmission capabilities and orientation of motes. At shorter ranges (<10m) transmission looks promising with high reliability (>95%) at all times. At larger ranges (>25m), quality of communications may be acceptable, but repeated tests have shown that reliability is very low. At medium ranges, from 10m to 25m, communication usually has consistent reliability. Our research can be readily extended in support of detecting trends and unanticipated events, the two key ingredients of an early-warning system.

Acknowledgment

We would like to thank Mat Kelly and Liang Chen of Intelligent Networking and Systems Research Group at the Computer Science Department of Old Dominion University.

References

[1] A. A. Abbasi and M. Younis, A survery on clustering algorithms for wireless sensor networks, *Computer Communications*, 30 (14-15), 2826-2841, 2007.

[2] Hady S. AbdelSalam, Stephan Olariu, "BEES: BioinspirEd backbonE Selection in Wireless Sensor Networks," *IEEE Transactions on Parallel and Distributed Systems*, vol. 23, no. 1, pp. 44-51, Jan. 2012, doi:10.1109/TPDS.2011.100.

[3] Hady S. AbdelSalam and Stephan Olariu, "Toward Efficient Task Management in Wireless Sensor Networks," *IEEE Transactions on Computers*, Vol. 60, No. 11, November 2011, pp. 1638-1651.

[4] Hady S. AbdelSalam, Stephan Olariu, and Syed R. Rizvi, Tiling-Based Localization Scheme for Sensor Networks Using a Single Beacon, Proceedings of the *IEEE Global Communications Conference (IEEE GLOBECOM 2008)* , New Orleans, LA, December, 2008.

[5] D Bitner, F Boon, M Careem, G Threadgold, "SahanaGIS: a Rapid Deployment GIS Framework for Disaster Management Humanitarian Relief Operations," Proceedings of the *Free and Open Source Software for Geospatial Conference*, Capetown 29 September - 4th October 2008.

[6] Manal AL-Bzoor, Laiali Almazaydeh, and Syed S. Rizvi. Hierarchical Coordination for Data Gathering (HCDG) in *Wireless Sensor Networks. International Journal of Computer Networks (IJCN)*, Vol. 5, Issue. 5, pp. 443 - 455, 2011.

[7] S. Olariu, M. Eltoweissy, and M. Younis. ANSWER: AutoNomouS net-Worked sEnsoR system. *Journal of Parallel and Distributed Computing*, 67(1):111–124, 2007.

[8] Mary Rose, "Japan Quake – Tsunami in Unforgettable Photos". Available: http://society.ezinemark.com/japan-quake-tsunami-in-unforgettable-photos-7736a56b91c8.html.

[9] Bryan Sarazin and Syed S. Rizvi. A Self-Deployment Obstacle Avoidance (SOA) Algorithm for Mobile Sensor Networks, *International Journal of Computer Science and Security (IJCSS)*, Vol. 4, Issue. 3, pp. 316-330, 2010.

Syed R. Rizvi is a PhD student at the Computer Science Department of Old Dominion University (ODU). Since 2002, Syed Rizvi has been working as a NASA Contractor developing advanced technologies, primarily for NASA Langley Research Center. Mr. Rizvi is the author of the book *Microcontroller Programming: An Introduction, ISBN 978-1439850770*, published by CRC Press. He received his M.S. in Computer Science from ODU and B.S. in Electrical Engineering. His research interests are related to embedded systems, vehicular ad-hoc networks, wireless networks, QoS provisioning, wireless multimedia, and web applications.

Chinmay Lokesh is a MS student at Computer Science Department of Old Dominion University. He is currently a Research Assistant at Intelligent Networking and Systems Research Group at the Computer Science Department of Old Dominion University. His research interests include sensor networks, vehicular networks, and cloud and distributed computing.

Michele C. Weigle is an Assistant Professor of Computer Science at Old Dominion University. She received her Ph.D. from the University of North Carolina at Chapel Hill in 2003. Her research interests include vehicular networks, wireless and mobile networks, network protocol evaluation, network simulation and modeling, and Internet congestion control. She is a member of ACM, ACM SIGCOMM, ACM SIGMOBILE, IEEE, and IEEE ComSoc.

Stephan Olariu is a tenured full professor in Computer Science at Old Dominion University, Norfolk, Virginia. He is a world-renowned technologist in the areas of parallel and distributed systems, parallel and distributed architectures and networks. He was invited and visited more than 120 universities and research institutes around the world lecturing on topics ranging from wireless networks and mobile computing, to biology-inspired algorithms and applications, to telemedicine, to wireless location systems, and vehicular networks. Professor Olariu is the Director of the Vehicular Networking Research Group at Old Dominion University. Prof. Olariu is an Associate Editor of Networks, International Journal of Foundations of Computer Science, and serves on the editorial board of Journal of Parallel and Distributed Computing and as Associate Editor of IEEE Transactions on Parallel and Distributed Systems

Medical Modeling and Simulation

VMASC Track Chair: Dr. Andi Parodi
MSVE Track Chair: Dr. Michel Audette

Dual Task Interference in Laparoscopic Psychomotor Performance: A Proposed Simulator Study
Author(s): Erik G. Prytz

A Framework for Therapy Planning in Robotic Deep Brain Stimulation
Author(s): Tansweer Rashid, Gregory S. Fischer, and Michel A. Audette

Cavitation based Ultrasonic Surgical Aspirator Modeling for Brain Tumor Annihilation Simulation
Author(s): F. Mahmud, A. Demuren, and M.A. Audette

Health Nexus: A Serious Game for Treatment and Prevention of Obesity and Diabetes
Author(s): Joseph C. Miller, and Yuzhong Shen

Simulation of Infectious Disease Spread in a Clinical Lab Setting: A Pilot Study
Author(s): Lydia Wigglesworth-Ballard, Lynn Wiles, and Holly Gaff

Simulation for Social Change: An ABM of Grassroots Organizing to Combat HIV/AIDs in South Africa
Author(s): Erika Frydenlund, and Savannah L. Eck

3D Meshless Enrichment Technique for Handling Discontinuities in Soft Tissue Models
Author(s): Rifat Aras, Yuzhong Shen, and Michel Audette

An Agent-Based Model of the Dynamics of a Tick-Borne Disease
Author(s): Daniel Drake Tillinghast, and Holly Gaff

A Continuous Simulation of Adoptive Immunotherapy Thresholds
Author(s): Tyrell Gardner

Finite Element Analysis of Deformation of Bars Used to Correct Pectus Excavatum in Children
Author(s): Rohit N. Nikam, Stephen B. Knisley, Frederic D. McKenzie, and Robert E. Kelly

Towards a Digital Cranial Nerve Model for Neurosurgery Simulation
Author(s): Sharmin Sultana, and Michel A. Audette

Towards a Framework for Poly-affine Initialized Elastic Registration of MR Spine Images for Detection of Abnormalities
Author(s): Rabia Haq, and Michel Audette

The Application of Interactive Behavior Change Technologies to Enhance Patient Education Among Adults Diagnosed with Diabetes
Author(s): Koren S. Goodman, Holly Gaff, Elizabeth F. Giles, and Gianluca De Leo

Dual Task Interference in Laparoscopic Psychomotor Performance: A Proposed Simulator Study

Erik G. Prytz

***Abstract*—Dual-task paradigms have been used extensively in research to measure workload. In such studies, the performance of a secondary task is used to determine the workload of a primary task. However, it is also known that such dual-task designs may lead to primary task decrements due to task interference. This paper outlines an experimental method for investigating if Fitts' law can be used to measure such dual-task interference. The proposed method includes using a custom-built laparoscopic simulator known as a box trainer.**

I. Introduction

Laparoscopic surgery carries general benefits compared to traditional open site surgery for multiple procedures [1]. However, it also imposes greater mental workload for the surgeon. Thus, some research has been carried out to determine accurate workload measures to be used in training new surgeons. One suggested method is a secondary task performed in conjunction with a primary, typically simulator based and training focused, laparoscopic task. It is not clear if such a secondary task would interfere with basic psychomotor performance of the primary task. To ensure that a secondary task only operates on higher cognitive levels, and does not interfere with the basic psychomotor performance of the primary task, a mathematical model could potentially be used to compare such psychomotor performance under dual-task and single-task conditions. Fitts' law [2] is such a mathematical model. It is widely used in human factors psychology and human-computer interaction. Fitts' law describes the relationship between target size, distance, and movement time for simple psychomotor tasks such as pointing. A simple pointing movement is of course a simplification of the more complex psychomotor skills needed for laparoscopic surgery. However, for the purpose of investigating if there is indeed interference between the secondary task and basic psychomotor tasks it provides a more straightforward measure and a reasonable first step.

The purpose of this paper is to outline an experiment using a simple laparoscopic surgery simulator to investigate potential dual-task interference using Fitts' law. If the law is capable of showing differences between single-task and dual-task conditions, it could potentially be a useful tool to investigate psychomotor dual-task interference.

Manuscript submitted March 8, 2012.

Erik G. Prytz is a PhD student in Human Factors Psychology at the Department of Psychology at Old Dominion University, Norfolk, VA 23529 USA (e-mail: epryt001@odu.edu).

II. Background

A. Basics of Laparoscopic Surgery

Laparoscopic surgery is also known as Minimally Invasive Surgery (MIS). Unlike regular surgery, where a large incision is made to open up the operating area to the surgeon, laparoscopic surgery uses long, thin tools that are inserted through smaller incisions. The surgeon views his or her tools within the body cavity by a camera inserted through a similar small incision. This type of surgery carries benefits compared to traditional surgery for a number of different procedures [1].

Unfortunately, these benefits come at a cost. Some of the disadvantages that have been observed with laparoscopic surgery are tools with poor ergonomic design [3], [4], uncomfortable working postures [3], and haptic feedback, while present, is reduced and transformed **Error! Reference source not found.**. Further, the surgeon is restricted to visual feedback from the video feed displayed on a monitor. The monitor may not be optimally placed and lined up with the surgical site, requiring the surgeon to look to either side, up, or down which may affect the surgical performance. Another important point is that the surgeon is operating in 3D, moving his or her instruments along the X-, Y-, and Z-axes, whereas the 2D visual feedback on the monitor deprives the surgeon of important visual depth cues such as binocular disparity. In addition, the camera position and alignment can also have an effect on the task performance [6].

These and other factors speak to the fact that laparoscopic surgery is associated with higher mental workload than traditional open surgery [7]. Some researchers have investigated ways of measuring mental workload using a secondary task paradigm during laparoscopic training [8]. The idea is that the secondary task would be a more sensitive measure of laparoscopic surgery skills than the traditional performance measures of speed and accuracy from laparoscopic training tasks alone.

B. Secondary Task as Workload Measure

Secondary task measures are based on the assumption that humans have limited mental resources. Essentially, a human does not have an infinite mental capacity for performing

multiple things at the same time. When a person is engaged in their primary task, there are only so many spare mental resources are left for other tasks. If a person is then asked to perform a second task that requires more mental resources than what are available, performance on that secondary task will vary depending on how much spare mental capacity the person can allocate to the task. As the person's proficiency in the primary task increases, mental resources are freed up and can be allocated to the secondary task, thus improving secondary task performance. Therefore, even though the primary task performance measure may not necessarily change as a practitioner improves, adding a secondary task that is sensitive to the spare mental resources may provide an estimate of skill acquisition. There are multiple theories on how mental resources are used and shared. One common theory is Multiple Resource Theory (MRT; [9]). According to MRT, there are multiple dimensions along which the mental resources are distributed. A secondary task would only be a sensitive workload measure if it matched the primary task dimensions. A primary task heavy in the visuo-spatial dimension, such as laparoscopy, would require a secondary task that draws upon the same resources, such as the task used by Stefanidis and his colleagues [8]. Another theory is the bottleneck theory [10] that assumes there is one single pool of mental resources. In some cases, one task will occupy the entire pool, and thus create a bottleneck that does not allow any other tasks to be performed in parallel.

One interesting question that arises then is what impact a secondary task would have on a laparoscopic primary task. Specifically, one important question is whether adding a secondary task that draws upon visuo-spatial processing resources would interfere not only with those mental resources (as predicted by MRT), but also would interfere with the very basic, fundamental level of laparoscopic performance, i.e. psychomotor performance (as predicted by bottleneck theory). One way to assess the potential impact of a secondary task on basic psychomotor task performance would be to use Fitts' Law.

C. *Fitts' Law*

Fitts' law is a mathematical model that describes the relationship between psychomotor movement time, target size, and distance [2]. It can be used to both describe and predict simple psychomotor tasks, such as pointing to a target, given the target size and pointing distance. It has been used to analyze or predict movement of fingers [2], time to press keys on a keyboard, moving a mouse cursor to an icon, and has been included in an ISO standard (9241) for ergonomic design of office visual display terminals [11]. In short, it is one of the most common models for pointing task device interaction [11]. A commonly used formulation is the Shannon [12] formula:

$$T = a + b\log_2\left(1 + \frac{D}{W}\right) \tag{1}$$

where T is the movement time, a and b are tool and task specific constants, D is the travel distance, and W is the width of the target. This formulation assumes that the person performing the task actually hits the target at the end of the movement. When using Fitts' law in an experimental setting, this may not always be the case, given the tendency of humans to engage in a speed-accuracy trade-off [13]. It is therefore suggested [14] that the following formulation be used:

$$ID = \log_2(1 + \frac{D}{W}) \tag{2}$$

where *ID* is the *index of difficulty*, measured in bits. Then, when the data have been collected, the standard deviation of end-point positions, σ, for each participant can be calculated and used to compute the effective target width, *We*:

$$We = 4.133\sigma \tag{3}$$

Given *We*, an *effective index of difficulty* can be calculated by:

$$IDe = \log_2\left(1 + \frac{D}{We}\right) \tag{4}$$

Equation (4) circumvents the assumption of perfect accuracy, and is suitable when experimental participants engage in a speed-accuracy trade-off, or are poorly motivated and thus have lower performance than expected.

Finally, going back to the Shannon formulation in Equation 1, the constants *a* (intercept) and *b* (slope) can be found by fitting the data to a least-squares linear regression, giving the movement time *T* by:

$$T = a + b \times IDe \tag{5}$$

Thus, Fitts' law can be used to study a subset of basic psychomotor skills used during laparoscopic surgery, i.e., moving an instrument to a certain point in space. The intercept *a* and slope *b* can both be determined and compared over time or between practitioners or conditions. Another suggested use of Fitts' law is as a throughput measure [14]. The throughput measure combines both the speed and accuracy, effectively combining also the intercept *a* and slope *b*, into one single dependent measure that can be used to compare experimental conditions. The throughput formula as suggested by [14] is:

$$TP = \frac{1}{y}\sum_{i=1}^{y}\left(\frac{1}{x}\sum_{j=1}^{x}\frac{IDe_{ij}}{MT_{ij}}\right) \tag{6}$$

where *y* is the number of participants and *x* the number of conditions. In addition to the *TP* measure, the average movement times and accuracy for each condition should also be reported to show how the conditions differ [14].

D. *Summary*

In the proposed experiment, a low-fidelity laparoscopic box trainer will be used with a secondary task, much like the one used by Stefanidis et al. [8], to investigate how dual-task

conditions affect the psychomotor performance of laparoscopic tasks as measured by Fitts' law. Specifically, this study would determine if there are any differences among the variables included in the Fitts' law model between a single-task and dual-task condition. The null hypothesis is that there is no dual-task interference as measured by the variables in the Fitts' law model. The alternative hypothesis is that Fitts' law can detect a difference, which would support the use of the law as a measure of dual-task interference.

III. Method

In this section the proposed methodology for investigating dual-task interference with Fitts' law using simulated laparoscopic tasks will be outlined. First, it is important to note that a laparoscopic box trainer simulator will be used. These types of simulators are typically low-fidelity with abstract tasks that primarily focus on maneuvering the tools and manipulating objects. The tasks typically used, and those outlined for this experiment, do not call for any medical judgment.

A. Participants

College undergraduate students will be recruited for this study. This reduces the likelihood of any participant having prior experience with laparoscopic tasks, which means that the participants will be novices.

B. Primary Task

The primary task should be a pointing task to satisfy the requirements of Fitts' law. The task will be a pointing task where the different targets are distributed along the X-, Y-, and Z-axes. The participants will be asked to point to a target using a laparoscopic instrument from a specified starting position as fast as they can. The participants will view the laparoscopic instrument and target on a monitor placed in front of them and fed with the video images from a camera inside the box trainer.

The physical size of the targets, i.e. the width *W*, will remain constant over distance. As the target positions are manipulated along the Z-axis, they will appear smaller or larger on the monitor.

C. Secondary Task

The secondary task will be a visuo-spatial task, as this will draw on the same mental resources need for the primary task [13]. The participants will be shown an image of four brightly colored balls in a 3D-tunnel. Their task will be to determine if the placement of the colored balls is identical to a standard image, or if one of balls has moved in depth (further away or closer toward them) or along the circumference of the tunnel. They will make a response using foot-pedals. If the image was the standard image, they will respond by pressing the right foot pedal. However, if one of the colored balls was in a different position, they will press the left foot pedal. In this experiment, the image will be shown for 500 ms at the same time as the participants are performing the primary task.

To prevent the participants from anticipating the image, and thus delaying the initiation of the pointing task until after the image has been displayed, the presentation of secondary task will be randomized so that it only appears on some trials. This will also allow a comparison between primary task performance alone and in combination with the secondary task using the throughput measure *TP*. To further reduce the risk of participants anticipating the image, the image will be delayed by a random interval of 200 to 400 ms in 100 ms increments. This will ensure that the participants have engaged in the ballistic movement of pointing to the target when the secondary task is shown. Finally, if the participants complete the primary task (i.e., hits the target) before the 500 ms display time has expired, the image will be removed from the screen at that time.

D. Apparatus

A low-fidelity laparoscopic simulator (box trainer) will be constructed for this experiment (see Fig. 1).

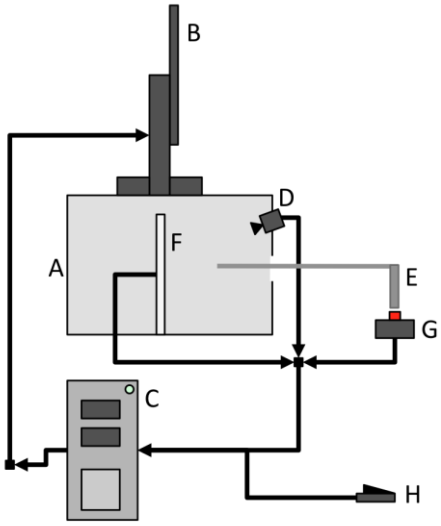


Fig. 1. The setup of the experimental simulator, viewed from the left side.

It will consist of a box (A) with a monitor (B) on top connected to a computer (C). The monitor will be a 15" LED-type monitor with a resolution of 1440x900 pixels at 110 PPI. The computer C is connected to a camera (D; Logitech HD C270h), with the video feed shown on monitor B at a resolution of 720p and refresh rate of 30 FPS. A hole in the front of the box will be used for the laparoscopic tool (E). Computer B will be connected to a drawing tablet (F; Wacom Bamboo Create) and button (G; U-HID Nano programmable circuit board). The drawing tablet F will be used to record responses. A stylus (not shown) will be attached to the end of the laparoscopic tool. As the participant points to a target shown on the drawing tablet, a custom program, developed as an Adobe Air runtime standalone application using Actionscript 3.0, on computer C will record where exactly on the tablet the participant pointed. This will allow for an accurate calculation of accuracy and effective target width *We*.

The participant will be required to press button G with the handle of the laparoscopic tool. The fixed location of the button (with the laparoscopic tool extending 11 cm into the box), will assure that the starting position is the same for all participants. This determines the specific distance *D* that the participant must move using the laparoscopic instrument to

reach the target. The same custom program running on computer C will also handle the secondary task, and overlay the secondary task image on top of the video feed from camera D. A set of foot pedals (H) is connected to computer C, where the same custom program will record the participant's responses to the secondary task.

In order to manipulate the distance *D*, the tablet will be moved depth-wise inside the box. There will be three distances, one at the far end of the box away from the participant (20 cm from the point of the laparoscopic tool in the starting position), one in the middle (14 cm from the end of the tool), and one toward the front close to the participant (8 cm from the end of the tool). Each distance will also vary the displayed width *W* of the targets, due to the physical target being close or further away from the camera.

E. Procedure

The participant will receive instructions for performing the primary and secondary task and then be allowed to practice the secondary task before starting the trials. Each trial will consist of the participant pressing a button (G in Fig. 1) with the handle of the laparoscopic instrument (E in Fig. 1). After a random interval of 1, 2, or 3 seconds, an auditory signal will sound. Upon hearing the signal, the participant will point to a specified target. If the trial includes a secondary task, an image will be shown on the monitor (B in Fig. 1) after a delay of 200, 300, or 400 ms. The image is displayed for 500 ms or until the participant hits the target, whichever comes first. Once the participant has hit the target, he or she responds to the secondary task using the foot pedals. The tablet is then moved to a different depth position in the box, and the participant repeats the trial. The trials will be arranged in blocks of 15. Each block will include one trial for each of the 5 targets at each of the 3 depths, and 6 trials will be without the secondary task, and 9 will be with the secondary task. The participants will go through a total 240 trials, or 16 blocks. The participants will be encouraged to move as fast as they can when pointing.

F. Independent and Dependent Variables

The independent variables for the primary task are the distance *D* and width of target *W*. The dependent variables are throughput *TP*, movement time *T*, and distance from target (accuracy). When using Fitts' law, it is important that the movement time does not include reaction time or dwell time [1]. Thus, there will be two separate measures of time for the primary task; *T1* from the onset of the auditory signal to the release of button G (that is, a reaction time from signal to initiation of movement), and *T2* from the release of the button to hitting the target (that is, a movement time from the onset of motion to the endpoint). *T2* will be used to calculate Fitts' law.

The independent variables for the secondary task are the presence or absence of a signal (colored ball out of position), and image delay time (200, 300, 400 ms). The dependent measures are accuracy (in percent correct detections), and response time. The response time, *T3*, is measured from the participant hitting the target to pressing the foot pedal.

IV. Projected Analysis

Using the independent and dependent measures listed above, the throughput for the different conditions can be calculated. These data will be analyzed in a repeated measures ANOVA to determine if there are any differences between performing the primary task with and without a secondary task. Further, the movement time *T2* can be fitt to Fitts' law using a least squares regression. As the custom program will record the exact point of contact for each trial and participant, the effective width *We* can also be calculated and compared across depths or secondary task presence.

V. Final Remarks

The experimental method proposed in this paper is designed to investigate whether Fitts' law can detect dual-task interference for a simulated laparoscopic task. It is a first step in looking at task interference for the basic psychomotor skills necessary for any type of laparoscopic surgery. The experiment will use a custom built simulator to capture the variables necessary to calculate the components of Fitts' law.

References

[1] Himal, H. S. (2002). Minimally invasive (laparoscopic) surgery. *Surgical endoscopy*, *16*(12), 1647-52.

[2] Fitts, P. M. (1954). The Information Capacity of the Human Motor System in Controlling the Amplitude of Movement. *Journal of Experimental Psychology*, *47*(6).

[3] Nguyen, N. T., Ho, H. S., Smith, W. D., Philipps, C., Lewis, C., De Vera, R. M., & Berguer, R. (2001). An ergonomic evaluation of surgeons' axial skeletal and upper extremity movements during laparoscopic and open surgery. *American journal of surgery*, *182*(6), 720-4.

[4] van Det, M. J., Meijerink, W. J., Hoff, C., Totté, E. R., Pierie, J. P. (2009). Optimal ergonomics for laparoscopic surgery in minimally invasive surgery suites: a review and guidelines. *Surgical endoscopy*, *23*(6), 1279-85.

[5] Bholat, O. S., Haluck, R. S., Murray, W. B., Gorman, P. J., & Krummel, T. M. (1999). Tactile feedback is present during minimally invasive surgery. *Journal of the American College of Surgeons*, *189(4)*, 349-55. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/10509459

[6] Delucia, P. R., & Griswold, J. A. (2011). Effects of camera arrangement on perceptual-motor performance in minimally invasive surgery. *Journal of experimental psychology. Applied*, *17*(3), 210-32.

[7] Berguer, R., Smith, W. D., & Chung, Y. H. (2001). Performing laparoscopic surgery is significantly more stressful for the surgeon than open surgery. *Surgical endoscopy*, *15*(10), 1204-7.

[8] Stefanidis, D., Scerbo, M. W., Sechrist, C., Mostafavi, A., & Heniford, B. T. (2008). Do novices display automaticity during simulator training? *American journal of surgery*, *195*(2), 210-3.

[9] Wickens, C. D. (1984). Processing resources in attention. In R. Parasuraman & D. R. Davies (Eds.), Varieties of attention (pp. 63-102). San Diego, CA: Academic Press

[10] Pashler, H. (1994). Dual-task interference in simple tasks: data and theory. *Psychological bulletin*, *116*(2), 220-44.

[11] Carroll, J. M., (2003) HCI Models, Theories, and Frameworks: Towards a Multidisciplinary Science Morgan Kaufmann, San Francisco, CA

[12] MacKenzie, S. I. (1992). Fitts' Law as a Research and Design Tool in Human-Computer Interaction. *Human-Computer Interaction*, *7*, 91-139.

[13] Wickens, C. & Hollands, J. (2000). *Engineering Psychology and Human Performance*. Prentice Hall: Upper Saddle River, NJ.

[14] Soukoreff, R. W., & MacKenzie, I. S. (2004). Towards a standard for pointing device evaluation: Perspectives on 27 years of Fitts' law research in HCI. *International Journal of Human-Computer Studies*, *61*, 751-789.

A Framework for Therapy Planning in Robotic Deep Brain Stimulation

Tanweer Rashid, Gregory S. Fischer, and Michel A. Audette

***Abstract*—Deep drain stimulation is a process where electrodes are inserted into the deep brain regions and electric fields are generated by these electrodes in the hopes of restoring brain functions. This procedure has been successfully applied in treating diseases like Parkinson's Disease, dystonia and even severe depression. This research effort aims to model the deep brain structures that can be used for robotic surgical planning. The brain models will be registered with MR data of the brain during the surgical procedure, and the robot system will utilize this data for accurate navigation.**

***Index Terms*—Surgery navigation, medical modeling and simulation, simplex mesh, deep brain stimulation, multi-surface model.**

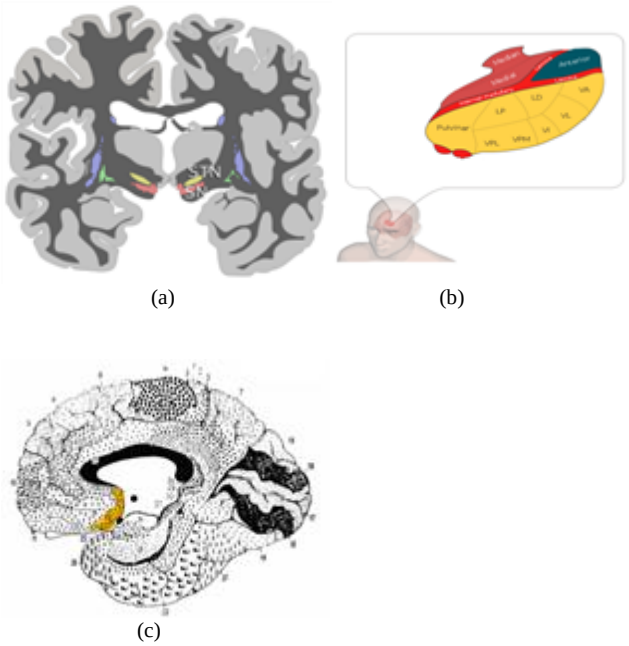
I. Introduction

Deep brain stimulation (DBS) involves surgically implanting tiny electrodes into the deep brain structures in order to treat neurological diseases. Electric pulses are subsequently sent into the brain via the electrodes to stimulate parts of the brain. This technique is used for treating various disorders like Parkinson's disease (PD), dystonia, chronic pain and severe depression. In the case of Parkinson's Disease, the disease for which DBS is most strongly indicated [1], the electrodes are implanted onto the subthalamic nucleus (STN), as illustrated in figure 1. Other targets include the anterior nucleus of the thalamus (ANT), which is emerging as the target for epilepsy, and the subgenual anterior cingulate (Cg25), which is targeted in treatment-resistant depression (TRD). In the future, our methods could also be applied to other types of neurological treatments, such as new drugs and genetic therapies.

II. Motivation

The motivation for this research arises from the need to have a reliable guidance system for deep brain electrode insertion. In particular, methods whose targeting relies only on preoperative information, and fail to integrate intraoperative image guidance, neglect the effect of brain shift. In a study by Miyagi [2], intraoperative brain shift statistics are reported for unilateral and bilateral targeting of the STN, with a mean displacement along Y of over 2mm, a standard deviation of over 3mm (fig. 3), and a maximum of 4mm. Similar statistics are reported by [3]. Given the size of the STN: on average, 5.9 mm in anteroposterior, 3.7 mm in mediolateral, and 5 mm in dorsoventral dimensions [4], brain shift statistics provide justification for the integration of a robotic approach and intraoperative imaging in DBS procedures: the target is reached more accurately and safely. The same argument can be made for the ANT and Cg25, where a procedure targeting the ANT in particular must avoid the ventricles.

The goal is to have a non-metallic surgical robotic implantation of electrodes into the brain. Furthermore, the patient and robot assembly will be placed within an MRI scanner to get real- time images of the patient's brain during surgery. Image registration will be used to map patient's brain MR image to a digital atlas of deep brain structures. Both the



(a) (b)

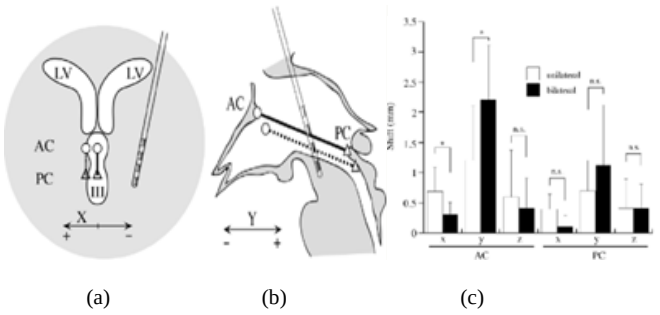
(c)

Fig. 1. Subcortical targets: (a) subthalamic nucleus (yellow, labeled STN); (b) ante-rior nucleus of thalamus (shown in blue); (c) subgenual anterior cingulate (yellow). Reproduced with permission from psychology.wikia.com and Wikimedia Commons.

Tanweer Rashid is a PhD student in the Modeling, Simulation and Visualization Engineering Department at Old Dominion University, Norfolk, VA 23508, USA. (e-mail: trash001@ odu.edu).

Dr. Gregory S. Fischer is Assistant Professor in the Department of Mechanical Engineering in Worcester Polytechnic Institute, Worcester, MA 01609, USA. (e-mail: gfischer@wpi.edu).

Dr. Michel A. Audette is Assistant Professor in the Modeling, Simulation and Visualization Engineering Department at Old Dominion University, Norfolk, VA 23508, USA. (e-mail: maudette@odu.edu).

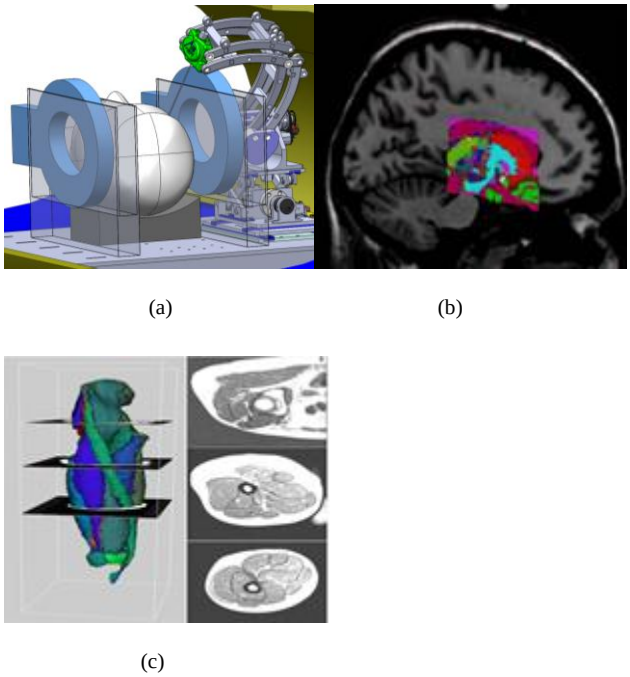


(a) (b) (c)

Fig. 2. DBS-induced brain shift along (a) X- and (b) Y; (c) mean shift statistics for unilateral and bilateral DBS, with peak along Y for bilateral. Reproduced with permission from [2].

robot and the atlas are depicted in figure 3. Since image registration is a time- consuming process, particularly for non-rigid transformations, one of our requirements is to make the registration process very fast. This will allow the estimation of the deformations in the patient's brain in near real-time, and make adjustments the robot accordingly.

To take deformations into consideration, the brain needs to be properly modeled. We make use of deformable simplex meshes, recently improved by Gilles [6] to support multi-surface modeling, as depicted in figure 3(c). Simplex models are connected meshes, characterized by each vertex having three neighboring vertices, which are topologically dual of triangulations. The justification for this multi-surface mesh approach is the computational efficiency of any non-rigid registration process built on such a sparse representation, given Gilles' precise control over resolution. Dr. Gilles has made this software available to this project, and our adaptation of these software classes to our requirements is currently underway.



(a) (b)

(c)

Fig. 3. Components of image-guided robotic deep brain therapy: (a) MR-compatible robot, (b) deep brain atlas, registered to patient data [5], (c) multi-surface simplex mesh, applied here to musculoskeletal modeling [6].

Fig. 4 shows the major stages of the project, and the anticipated time of completion of each stage. The first stage consists of building multi-simplex and multi-surface representations of the brain. This stage is anticipated to take one year to complete. In the second stage is a pre-op processing stage. The patient data will be registered to model data, and this is anticipated to be completed with two years. The third stage deals with the robotic surgical system. In this stage, the planning and awareness of the robotic system will be done and this will be completed within 2.5. The final stage is an intra-op process. The patient model will be registered to MR data, and this stage will be completed with 3.5 years. The total time from start to finish is an anticipated 3.5 years.

BIBLIOGRAPHY

[1] Deep Brain Stimulation. Perlmutter JS, Mink JW. 2006, Ann. Rev. Neuro., pp. Vol. 29: 229-257.

[2] Brain shift: an error factor during implantation of deep brain. Miyagi Y, Shima F, Sasaki T. Nov. 2007, J Neurosurg, , pp. 107:5, pp. 989-997.

[3] Assessment of Brain Shift Related to Deep Brain Stimulation Surgery. Khan MF, Mewes K, Gross RE, Skrinjar O. 2008, Stereotact Funct Neurosurg , pp. 86:44-53.

[4] Determining the position and size of the subthalamic nucleus based on magnetic resonance imaging results in patients with advanced Parkinson disease. Richter EO, Hoque T, Halliday W, Lozano AM, Saint-Cyr JA. 2004 Mar, J Neurosurg. , pp. 100(3):541-6.

[5] The creation of a brain atlas for image guided neurosurgery using serial histological data. Chakravarty MM, Bertrand G, Hodge CP, Sadikot AF, Collins DL. 2006, NeuroImage , pp. 30: 359 – 376.

[6] Musculoskeletal MRI segmentation using multi-resolution simplex meshes with medial representations. Gilles B, Magnenat-Thalmann N. 2010 Jun, Med Image Anal. , pp. 14(3):291-302 2010 .

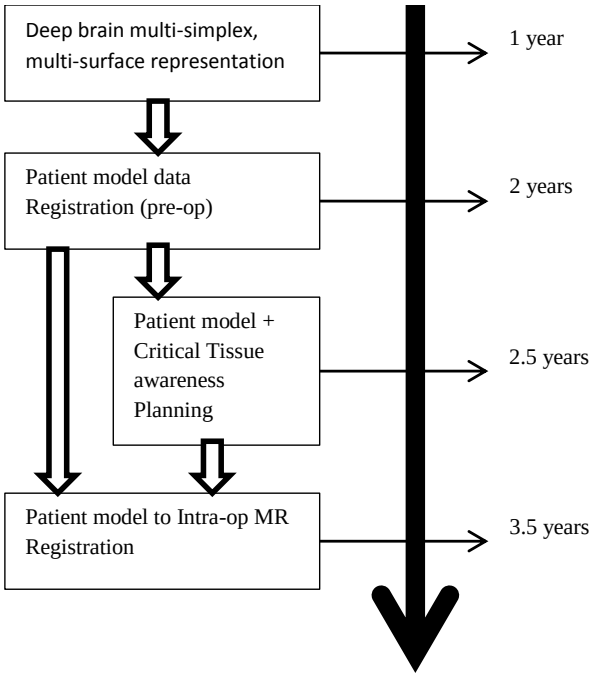


Fig. 4. Major stages of the project and anticipated timeline of completion of each stage.

Cavitation based Ultrasonic Surgical Aspirator Modeling for Brain Tumor Annihilation Simulation

F. Mahmud[1], A. Demuren[2], M.A. Audette[1]
[1] Dept. Modeling, Simulation and Visualization Engineering;
[2]Dept. Mechanical Engineering
Old Dominion University, Norfolk, VA, USA

***Abstract*— this research will present preliminary work in the development of an interactive simulator of the ultrasonic surgical aspirator. Our current progress includes the first phase of the project that will address non-interactive, predictive simulation of the process. At this level we are trying to model the bi-phasic, unsteady phenomena of cavitation models that are mesh independent in nature with the help of ANSYS FLUENT and GAMBIT. Future works will be to use the first phase as a gold standard for a somewhat simplified version, hardware-accelerated interactive implementation that will run on an open-source simulation platform, in conjunction with a 6 degree-of-freedom haptic device. We are also pursuing work on the characterization of tumor properties.**

***Index Terms*— Cavitation, Meshless Method, Ultrasonic Surgical Aspirator (USA).**

I. INTRODUCTION

Simulation of models are to make decisions on predictive or unpredictive nature of a system. Based on the complexity and type of a system, we need to identify some parameters to control or just to understand the response of the system. In the field of medical domain, surgery simulation has been shown to improve both surgical skill and patient outcome, especially in a context where surgical residencies hours are being shortened, in relation to traditional residencies. In this context, the traditional method of training whereby the inexperienced surgeon merely watches the senior surgeon operate, and only gradually assumes an active role, exhibits limitations and entails risks to the patient. Related to this, ultrasound based diagnostic is one of the most widely used diagnostic tools in medical/surgical domain. By applying ultrasonic sound characteristics to the Ultrasonic Surgical Aspirator (USA) in the modern armamentarium has profound consequences for practical neurosurgery simulation, given that the USA is the main tool used in debulking tumors. However, the aspirator can also cause small pockets of gas in body fluids or tissues to expand and contract/collapse in a phenomenon called cavitation. This tool uses cavitation to vaporize tumors in a tissue-selective manner, irrigates the area with water, which cools the vibrating tip of the tool as well as facilitates the third process: the aspiration of tumor tissue. While there are on-going efforts in neurosurgery simulation, including existing work on an interactive simulation of the USA (Jiang D, 2010), real-time performance often comes at the expense of the predictive aspect of the underlying model, in other words- the simulation lacks validity. Specifically, existing models entirely avoid modeling the cavitation process, concentrating on purely mechanical processes: suction and jackhammering. The cavitation process in brain tumor annihilation needs to further investigate for improving of the aspirator itself and at the same time to achieve more success rate in surgery. However, to faithfully account for the tissue-selective aspect of the USA, related to water content, the simulation must emphasize cavitation as well as model tumor properties, which correlate with major histological groupings. A USA simulator without cavitation cannot differentiate between a tumor with high water content and one that is fibrous or calcified, or even between tumor and healthy brain tissue, and is of limited clinical relevance. Therefore, our current effort is to characterize the bi-phasic phenomena of cavitation in unsteady nature to incorporate this with our future hardware-accelerated interactive implementation of a haptic device.

Manuscript received March 9, 2012.

[1]: F. Mahmud and M. Audette are with the Dept. Modeling, Simulation and Visualization Engineering, Old Dominion University, Norfolk, VA 23508 USA (phone: 757-683-3720; e-mail: fmahm001@odu.edu, maudette@odu.edu).

[2]: A. Demuren is with Dept. Mechanical Engineering, Old Dominion University, Norfolk, VA 23508 USA (e-mail: ademuren@odu.edu).

II. LITERATURE REVIEW

While dealing with cavitation phenomenon in our current study, we are more interested in a particular type of cavitation called “Acoustic cavitation” which has been researched by many researchers and thus has different understandings from different points of views based on particular circumstances. It is mainly the formation of vapor-filled bubbles in a liquid during the short periodic intervals of negative pressure, or tensile stress that accompany the passage of a sound wave. It might be bubbles or vapor cavities from acoustic inception, or the dynamics of such a behavior or even the effects of cavity dynamics on materials or bubbles. Some literatures identified and discussed their experimental findings on cavitation specially by discussing acoustic parameters (such as acoustic pressure and frequency), thermodynamic parameters (like ambient pressure and temperature) as well as the liquid parameters (like density, viscosity, surface tension, and gas saturation (Flynn, 1964) (L.D.Rozenberg, 1971)

(A.Prosperetti) (W.L.Nyborg, 1978) (Neppiras, 1980).

Strasberg (M.Strasberg, 1959) investigated the conditions influencing the onset of acoustically induced cavitation in tap water with special attention to the effect of air-filled cavitation nuclei. It was identified that an obvious reason for the complexity of ultrasonically induced cavitation is the brevity of the individual cyclic pressure reductions occurring at ultrasonic frequencies.

Holland et al. (E.Apfel, 1990) did experiment on the thresholds for transient cavitation produced by pulsed ultrasound in a controlled nuclei environment. The measurements of the transient cavitation thresholds were presented in water, in a fluid of higher viscosity, and in diluted whole blood.

Laborde et al. (J.-L.Laborde, 1998) investigated the acoustic cavitation bubbles forming place(s) and the intensity depending on the geometric and acoustic factors. They supported the fact that mathematical modeling might help to predict acoustic cavitation fields as long as we could identify cavitation place(s) and the intensity. As part of their research, they also presented linear acoustic equations and fluid dynamics equations to support their findings.

Suslick et al. (Kenneth S. Suslick, 1999) worked on acoustic cavitation to see its chemical consequences. They explored a diverse set of applications of ultrasound to enhance chemical reactivity. Ultimate results of their research include that acoustic cavitation is responsible for both sono-chemistry and sono-luminescence.

III. Procedures

The predictive modeling and simulation of the Cavitron Ultrasonic Surgical Aspirator (CUSA) for brain tumor annihilation will be developed with ANSYS FLUENT (ANSYS Fluent) via predictive computational fluid dynamics. Within ANSYS FLUENT, we are applying the Sauer and Scherr cavitation model (Sauer J, 2001), based on its applicability to unsteady cavitational flow. This is also the default cavitaion model in FLUENT. While the study of histopathological variability in tumor response to cavitation is still in its infancy, and will be pursued in our future research in conjunction with our tumor modeling effort, we argue that a basic numerical model of USA processes that emphasizes cavitation is a necessary step for producing a simulation faithful to tissue-specific neurosurgical ultrasonic aspiration. Moreover, an interactive simulator rigorously based on this predictive model could provide training for residents, that is broadly representative of their future caseload.

IV. Cavitation Model Formulation

The following assumptions are made in our standard bi-phasic cavitation model:

- The system under investigation must consist of a liquid and a vapor phase.
- Mass transfer between the liquid and vapor phase is assumed to take place. Both bubble formation (evaporation) and collapse (condensation) are taken into account in the cavitation model.
- The cavitation model is based on the Rayleigh-Plesset equation, describing the growth of a single vapor bubble in a liquid.

With the multiphase cavitation modeling approach, a basic two-phase cavitation model consists of using the standard viscous flow equations governing the transport of mixture (Mixture model) or phases (Eulerian multiphase), and a conventional turbulence model (k- ε model). In cavitation, the liquid-vapor mass transfer (evaporation and condensation) is governed by the vapor transport equation:

$$\frac{\partial(\alpha\rho_v)}{\partial t} + \nabla.(\alpha\rho_v \mathrm{Vv}) = \mathrm{R_e} - \mathrm{R_c} \ldots\ldots (1)$$

Here, the net mass source term is as follows:

$$R = \frac{\rho_v \rho_l d\alpha}{\rho \quad dt} \ldots\ldots (2)$$

Where, ρ_v = vapor density

ρ = mixture density (function of phase volume fraction and density)

ρ_l = liquid density

α = vapor volume fraction

Vv = vapor phase velocity

R_e, R_c = mass transfer source terms connected to the growth and collapse of the vapor bubbles respectively.

In Equation 1, the terms R_e and R_c account for the mass transfer between the vapor and liquid phases in cavitation. In ANSYS FLUENT, they are modeled based on the Rayleigh-Plesset equation describing the growth of a single vapor bubble in a liquid.

V. Results

Currently we are working to analyze the cavitation model by applying CFD techniques. Before running a mesh in ANSYS FLUENT with boundary conditions, the mesh should be created. In order to do that, we are using GAMBIT mesh generator (GAMBIT comes with ANYSYS product).

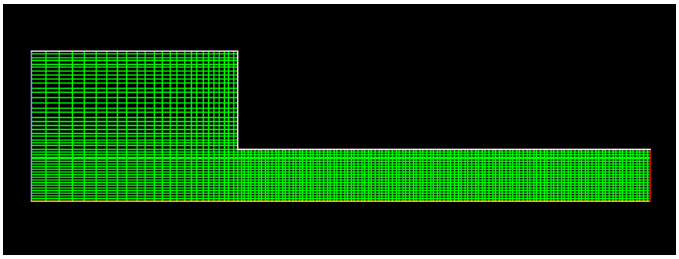


Figure 1: Experimental mesh grid to examine pressure-driven cavitation flow of water through a sharp-edged orifice.

The work is still in progress and further updated results will be available during the time of the conference. However, our current progress is to understanding the cavitation phenomenon in different experimental setup. We are considering a bi-phasic mixture model. Primary phase is liquid water whereas secondary phase is water vapor. The problem considers the cavitation caused by the flow separation after a sharp-edged orifice. The flow is pressure driven, with an inlet pressure of 5×10^5 Pa and an outlet pressure of 9.5×10^4 Pa. The orifice diameter is 4×10^{-3} m, and the geometrical parameters of the orifice are D/d = 2.88 and L/r = 8, where D, d, and L are the inlet diameter, orifice diameter, and orifice length respectively. The geometry and the mesh grid of the orifice are shown in Figure 1. Flow is from left to right through the orifice.

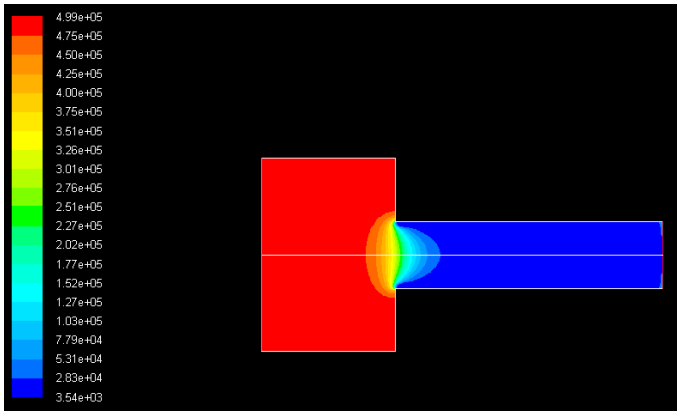


Figure 2: Mirrored view of contours of static pressure (mixer) of the experimental mesh.

In Figure 2, we can see the dramatic pressure drop at the flow restriction. Low static pressure is the major factor causing cavitation. Additionally, turbulence contributes to cavitation due to the effect of pressure fluctuation and turbulent diffusion, as will be shown in a plot that follows.

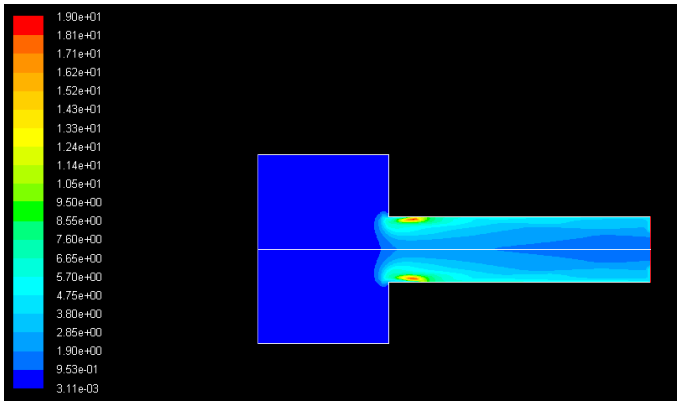


Figure 3: Mirrored view of contours of turbulent kinetic energy (mixer).

The high turbulent kinetic energy region near the neck of the orifice in Figure 3 coincides with the highest volume fraction of vapor in Figure 4. This indicates the correct prediction of a localized high phase change rate. The main flow then convects the vapor downstream.

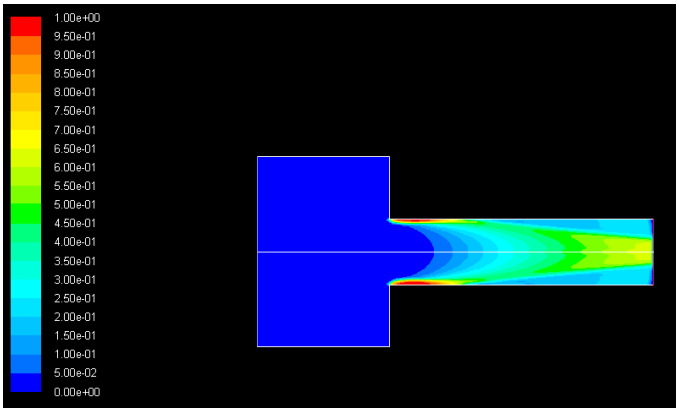


Figure 4: Mirrored view of contours of volume fraction (vapor).

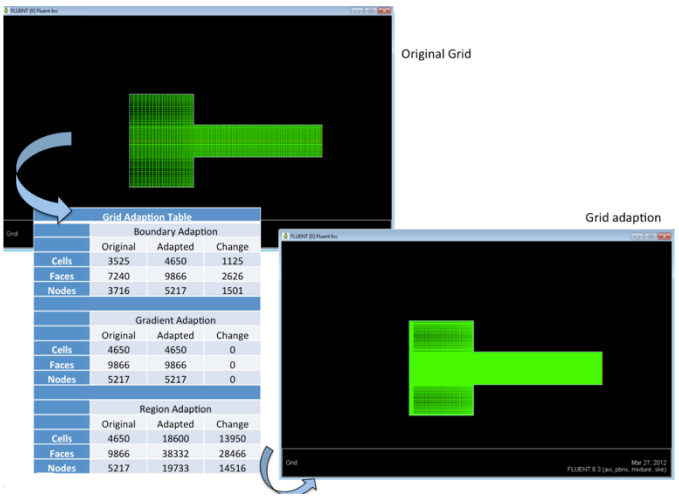


Figure 5: Mesh adaption.

Figure 5 shows how we adapted meshes in our analysis. Originally there were 3525 cells, 7240 faces and 3716 nodes in our analyzed geometry. By mesh adaption, we try to make the geometry more refined in various places to achieve mesh independent results. Therefore, in our case, we gained 13950 cells, 28466 faces and 14516 nodes to get a mesh independent result.

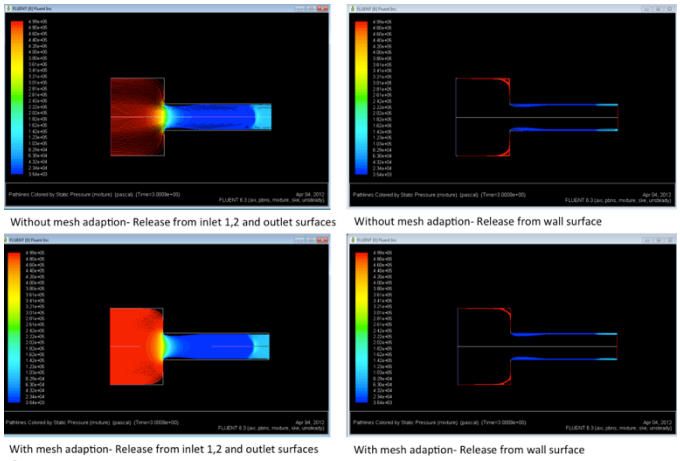


Figure 6: Pathlines for Static Pressure (Unsteady with adaptive time step).

In Figure 6, we can see the pathlines for static pressure along the walls and centerline. Pressure variations are also showed on the left of each window.

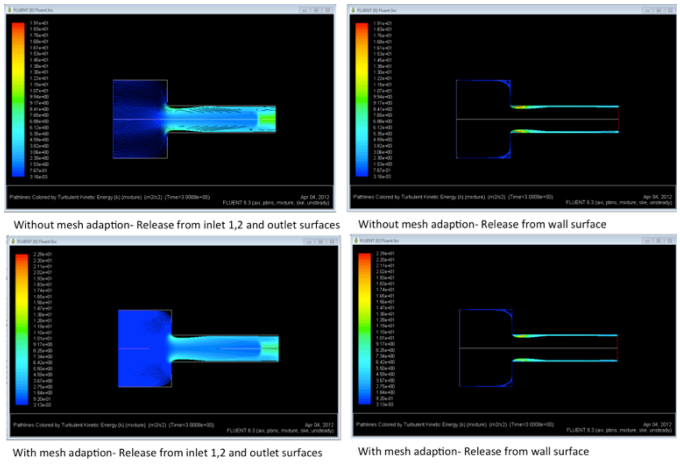


Figure 7: Pathlines for Turbulence Kinetic Energy (Unsteady with adaptive time step).

Figure 7 shows pathlines for turbulence K along the walls and centerline. To get more precise results, we also tried with different discretization methods to see the effects. Figure 8 shows static pressure contours for various discretization methods.

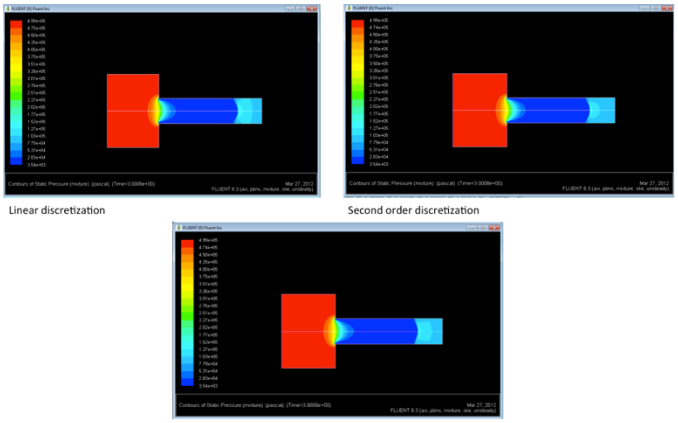


Figure 8: Effects of discretization on Static Pressure Contours (for unsteady case).

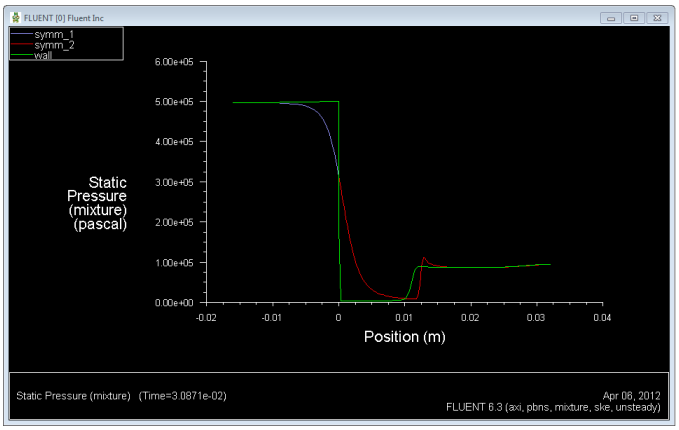


Figure 9: Static Pressure (P) with respect to Position (X).

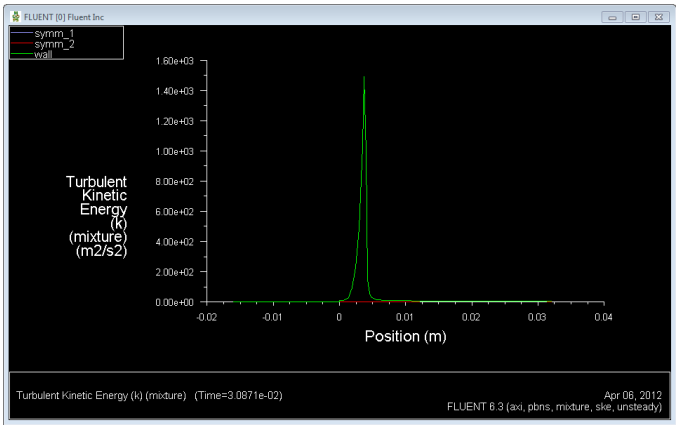


Figure 10: Turbulent Kinetic Energy (K) with respect to Position (X).

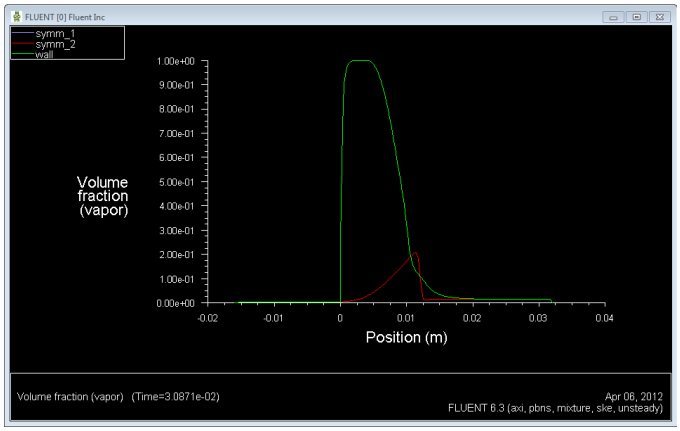


Figure 11: Vapor Fraction (V.F) with respect to Position (X).

Figures 9, 10 and 11 are the plots of static pressure, turbulence and vapor fraction with respect to position of the conduits we considered in our analysis. These plots clearly show where and at which points there are sudden changes of the parameters. The plots took both the walls and conduit centerline into consideration.

VI. Discussions

Current experimental approach is our first step to understand cavitation effects. For the multiphase mixture model, we normally specify conditions for the mixture (i.e., conditions that apply to all phases) and also conditions that are specific to the primary and secondary phases.

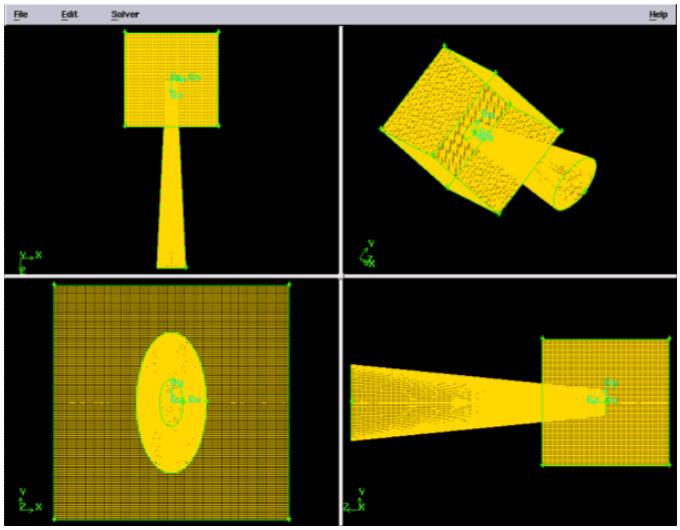


Figure 12: Future experimental mesh and geometry.

In our experiment, boundary conditions are only needed for the mixture and secondary phase of two boundaries: the pressure inlet (consisting of two boundary zones) and the pressure outlet. The pressure outlet is the downstream boundary, opposite the pressure inlets. To obviate dynamic re-meshing, our interactive cavitation model will be implemented with a meshless method, with a view to implementing the microstreaming, irrigation and suction with this technique in phase II. Meshless methods discretize partial differential equations, including continuum mechanics expressions, through shape functions with compact support defined on a local cloud of points (or nodes), rather than on an non-overlapping elements. These techniques have been applied to both solid mechanics and computational fluid dynamics (Katz AJ, 2009), but have not yet been used for modeling cavitation and the acoustic effect, to the best of our knowledge.

VII. Conclusion

The research is still in its initial stage to develop the control parameters for surgical aspirator modeling for brain tumor annihilation. Our current progress is to come with a solid base and theory for cavitation modeling.

References

[1] A.Prosperetti, M. a. (n.d.). Bubble Dynamics and cavitation. *Fluid mech.9* , 145-185.

[2] ANSYS Fluent. (n.d.). *ANSYS Fluent*. Retrieved from http://www.ansys.com/Products/Simulation+Technology/Fluid+Dynamics/ANSYS+Fluent

[3] E.Apfel, C. K. (1990). Thresholds for transient cavitation produced by pulsed ultrasound in a controlled nuclei environment. *Journal of Accoustic Society of America* , 2059-2069.

[4] Flynn, H. (1964). Physics of Acoustic Cavitation in Liquids. (W. Mason, Ed.) *Physical Acoustics , IB*, 57-172.

[5] J.-L.Laborde, C. J.-P. (1998). Acoustic cavitation field prediction at low and high frequency ultrasounds. *Ultrasonics , 36*, 581-587.

[6] Jiang D, C. N. (2010). Characterization of Suction and CUSA Interaction with Brain Tissue. *Proceedings of ISMBS*, (p. pp.11-19).

[7] Katz AJ. (2009). *Meshless Methods for Computational Fluid Dynamics,* . Stanford University, Ph.D. thesis, .

[8] Kenneth S. Suslick, Y. D. (1999). Acoustic cavitation and its chemical consequences. *Phil.Trans.R.Soc.Lond. , A*, 335-353.

[9] L.D.Rozenberg. (1971). High Intensity Ultrasonic Fields. New York: Plenum.

[10] M.Strasberg. (1959). Onset of Ultrasonic cavitation in Tap Water. *The Journal of the Acoustical Society of America , 31* (2), 163-176.

[11] Neppiras, E. A. (1980). Acoustic Cavitation. *Phys. Rep.* , 61-159.

[12] Sauer J, S. G. (2001). Unsteady cavitating flow—A new cavitation model based on a modified front capturing methodand bubble dynamics,. *Proceedings of 2000 ASME Fluid Engineering Summer Conference,*. Boston, MA,.

[13] W.L.Nyborg, W. a. (1978). Cavitation; Dynamics of Gas Bubbles; Applications. (F.J.Fry, Ed.) *Ultrasound: Its Applications in Medicine and Biology* , 77-159.

[14] Characterization of Suction and CUSA Interaction with Brain Tissue, Jiang D, Choudhury N, Mora V, Delorme S,. 2010. Proceedings of ISMBS. pp.11-19.
[15] Unsteady cavitating flow—A new cavitation model based on a modified front capturing method and bubble dynamics,. Sauer J, Schnerr GH,. Boston, MA, : s.n., 2001. Proceedings of 2000 ASME Fluid Engineering Summer Conference.
[16] http://ftp.rta.nato.int/public//PubFullText/RTO/EN/RTO-EN-AVT-143///EN-AVT-143-02.pdf

Health Nexus: A Serious Game for Treatment and Prevention of Obesity and Diabetes

Joseph C. Miller, III, and Yuzhong Shen

Abstract—**Serious games are video games that take advantage of games' intrinsic entertaining and challenging characteristics for serious purposes, such as education, training, advertisement, and political campaigns. This paper proposes to develop a serious game prototype for education and treatment of obesity and diabetes which will utilize the Microsoft Kinect, a motion sensor that allows the player to act as the game controller. The absence of physical game controllers allows the player to be fully immersed in the virtual environment presented by the games and leads to more realistic and engaging gaming experience. We intend to take advantage of the high desire to play video games and use that to advise people how to eat healthily and exercise correctly by utilizing Calorie and Gluten Level based health models to demonstrate effective eating and exercise habits to maintain so as to minimize the chances or mitigate the symptoms of having diabetes or obesity. This game was developed with the Microsoft Kinect SDK, XNA Game Studio, and .NET Framework in C# programming language.**

Index Terms—**diet and exercise, obesity and diabetes, serious games.**

I. Introduction

VIDEO games represent a form of entertainment that touches all ages, all demographics and a wide variety of interests. Video games statistics show that the bulk of gamers are between the ages of 18-49, with the number of older gamers increasing at 29% of all gamers in 2011. Video gaming is no longer a hobby limited to a certain age group or demographic, nor even to specific sexes. The same study by the Entertainment Software Association (ESA) shows that there are nearly an equal number of girls to boys gaming as of 2011, with 52% of gamers being male and 48% being female [1]. Video games appeal strongly to people across all demographics; an assertion that is verified by the sales numbers for the video games industry which has nearly tripled in the last ten years [1].

The passion for games can be deployed for more vital purposes such as education, training, and marketing. This project proposes to harness the popularity and availability of video games for the purpose of education and treatment of obesity and diabetes. To do so, a specific form of video game will be investigated, known as the serious game. In short, a serious game is a game that involves a specific topic and keeps the user entertained throughout the learning process. Game-based learning uses serious games with defined learning outcomes and objectives. The values of game-based learning have been recognized by organizations such as National Science Foundation and National Research Council [3-5]. NSF considers games as an important form of cyber-learning platform and technology [6]. It is this union of technology and education that the authors propose to utilize for the purpose of preventing and treating obesity and diabetes.

Obesity and diabetes are perhaps the two most prevalent diseases in the U.S. These two diseases present extraordinarily concerning health risks for multiple reasons. Not only are the effects of both obesity and diabetes quite serious, but being a victim of either ailment greatly increases the risks of additional health problems such as heart disease, cancer and kidney disease. Obesity and diabetes are serious health conditions and both have become far more prevalent, particularly in the United States over the past 20 years [7].

Obesity has been defined for adults by the Center for Disease Control and Prevention as “An adult who has a (Body Mass Index) of 30 or higher” [8]. The Body Mass Index, or BMI for short, is an accepted method of measuring body fat by using an individual’s weight and height, which is in turn used as a measure of obesity. To calculate BMI, there are two methods, one for the metric system and one for American measures [9].

These measures return a value detailing the amount of body fat of a person. If this value exceeds 30, then the person in question is obese and is subject to the symptoms and risks of obesity. The main causes of obesity, in short, involve consuming more energy than is released through physical activity. When one eats or drinks one takes in energy in the form of Calories, whereas when one performs physical activities one consumes Calories. If intake and consumption of Calories are in balance, then weight remains stable. If the balance tips in favor of one or the other, one will gain or lose weight depending on the nature of the imbalance. Though Caloric imbalance is the leading cause of obesity, environmental, genetic, diseases and drugs may also factor into obesity [10].

Diabetes is an equally troubling condition and has been diagnosed in approximately 11% of adults in the United States as of 2010 [11]. Diabetes comes in three main types: Type-1, Type-2, and gestational. Type-1 diabetes occurs when the

Manuscript received October 9, 2001.

body's immune system destroys pancreatic beta cells, which are the only cells in the body that produce insulin, the hormone that regulates blood glucose levels [12]. Type-1 diabetes is the most serious as sufferers of this infliction require regular insulin shots or pumps in order to survive, as their own body has destroyed its source of insulin. Type-2 diabetes occurs when body cells develop insulin resistance; insulin is produced but is not used correctly by the body's cells to break down sugar. Gestational diabetes only occurs in pregnant women and is a mild form of glucose intolerance. Type-2 diabetes is by far the most common, accounting for 90-95% of all reported cases of diabetes [12].

II. Game Design and Architecture

Obesity and diabetes have two treatments in common: diet and exercise. The Health Nexus game software design is hierarchy based and reflects these major applications. The architecture is shown in the following figure.

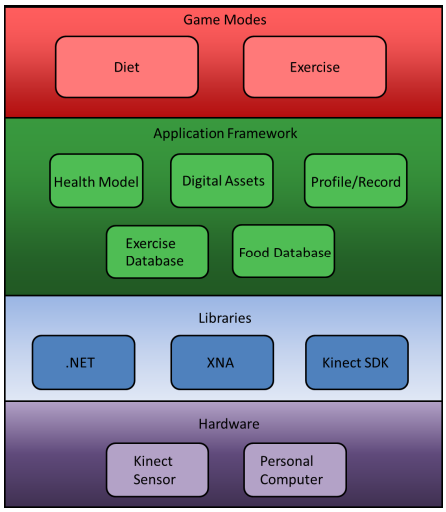


Figure 1: Health Nexus Software Architecture

Figure 1 demonstrates the component hierarchy of the health game, which spans four tiers: Hardware, Libraries, Application Framework and Game Modes.

The hardware components of the game are the interface between the player and the health game. These components are fairly straightforward, consisting of a personal computer and a Kinect motion sensor. The main control device for the game is the Kinect sensor which provides a controller-less interface between the player and the game via skeletal tracking. The personal computer is necessary for running the game as well as serving as an interface for the Kinect sensor.

The development of the framework described above requires a set of software libraries in order to design an appropriate interface between the components and the player. These software libraries provide high level development functionality that is conducive to the game development environment. The .NET Framework, XNA Game Studio and the Microsoft Kinect Software Development Kit (SDK) were chosen for the development of the serious health game. The .NET Framework was originally released by Microsoft in 2002 and runs on Microsoft Windows. The framework consists of two major components, a library of functions known as the Base Class Library that controls user interfacing, data access, network communications, etc. as well as a software environment known as the Common Language Runtime, or CLR. These two components in tandem allow the .NET Framework to operate across a multitude of languages, including C++, C#, and Visual Basic. The Base Class Library encapsulates many common functions, which is convenient for higher level design purposes, while the CLR provides a consistent execution environment which manages memory usage, exception handling and security for all programs developed using the .NET Framework.

Microsoft XNA Game Studio (XNA) is a game development library that was designed by Microsoft in 2006 as XNA Game Studio Express. XNA covers a wide range of functionality, from a library of useful game development tools, game asset pipeline management and a managed runtime environment for execution of developed games. XNA contains over 400 classes related to game design and content management and allows for cross platform development for Windows, Windows Phone and Xbox 360. The only supported languages for XNA at this time are C# and Visual Basic so the C# programming language will be utilized in conjunction with XNA.

The Microsoft Kinect SDK serves as a software interface between the Kinect motion sensor and the PC. This kit allows developers to access raw data returned from the Kinect sensor including depth data, color data, sound data and skeletal tracking data. The skeletal tracking data is of particular value as it is utilized in the game to provide animations of 3D avatars based upon player movement.

In order to realistically and accurately develop the game towards dieting and exercise, a solid software framework is required. This framework consists of a set of components designed to facilitate dieting and exercise by incorporating the aforementioned software libraries and hardware.

A. *Health Model*

First of the components is the health model. For the purposes of this project a health model will be generated concerning Caloric intake, sugar level intake and exercises, allowing players of the game to see how much weight is gained or lost and how sugar saturates their diet is based on a set of rules determined by the model.

The first approach is to plan and maintain an effective diet. An effective diet has a multitude of definitions, but for the purposes of this project we assume the simplest; an effective diet consists of consuming foods which provide the necessary nutrients for the body to function while keeping the total Calorie count as low as possible. Therefore the calculations of the desired daily Caloric intake are determined by a set of equations known as the Harris-Benedict (HB) [13]. This set of equations calculates the daily amount of Calories required in order for a person to maintain a current weight based upon current height, current weight, age and sex [14]. The HB equations also take into account the relative activity of the

person, such as daily exercise routines, sports, outdoor job work, etc. The HB equations for men and women are given as follows:

$$DEE = (66 + 13.7w + 5h - 6.8a) \times AM, \quad (1)$$

where *DEE* is the daily energy expenditure, *w* is weight in kilograms (kg), *h* is height in centimeters (cm), *a* is age in years (yr) and *AM* is a coefficient based on level of activity, discussed in Table 1.

$$DEE = (655 + 9.6w + 1.7h - 4.7a) \times AM, \quad (2)$$

where *w* is weight in kilograms (kg), *h* is height in centimeters (cm), *a* is age in years (yr) and *AM* is a coefficient based on level of activity, discussed in Table 1.

TABLE 1
ACTIVITY MEASURE COEFFICIENTS

Amount of Exercise	AM
Little to no exercise	1.2
Light exercise (1–3 days per week)	1.375
Moderate exercise (3–5 days per week)	1.55
Heavy exercise (6–7 days per week)	1.725
Very heavy exercise (twice per day, extra heavy workouts)	1.9

Table 1 details the activity measure coefficients based on the activity level of the person in question. As one may see, the coefficient increases proportionally to the amount of exercise and exertion taken in a given week. Since the Harris-Benedict equations calculate the amount of energy required to maintain your current weight, it is sensible to assume that the more Calories a body burns in a given period of time the higher Calorie consumption is required to off-balance the amount of Calories burned.

The HB equations are the most accurate measure of daily Caloric intake generated to date, surpassing other metrics such as the Owen, Mifflin–St. Jeor, and WHO/FAU/UNU equations in both accuracy and versatility, measuring accurate Caloric needs across both sexes and all Body Mass Index categories [15]. Therefore the HB equations will be utilized to calculate the daily Caloric needs for each individual player and weight loss or gain will be measured by the difference between the desired daily Caloric intake and the actual daily Caloric intake as determined by in game choices.

With the rules for Caloric intake determined it is time to describe a suitable method of displaying sugar intake. Unfortunately there is currently no way to measure blood sugar levels without actual blood testing, so we must investigate alternative methods of managing a patient's blood sugar. Therefore we intend to introduce a second method of monitoring sugar intake: glycemic load.

The glycemic load considers both the rate of digestion and the total amount of carbohydrates present in the food being eaten. Thus it is possible, using glycemic load, to eat a food that has a high GI but low total number of carbohydrates. This applies to many healthy foods that have a high GI such as pumpkin or parsnips which have a low GL [43]. The equation for calculating the GL of a food is shown in (3).

$$GL = \frac{(GI \times N_{carb})}{100}, \quad (3)$$

where N_{carb} is the number of carbohydrates per 100 grams of food.

The final consideration of the health model is the exercises. Exercises are excellent methods of consuming surplus Calories. However, the various methods of exercise do not consume Calories at the same rate: Calorie consumption is based on the duration and intensity of the exercise. The health model proposed for this game accounts for these differences in exercise by modifying the Activity Measure coefficients utilized by the Harris-Benedict equations shown in Table 1. The game will detect and record the exercises chosen by the player and the duration that the player exercises in order to select an accurate Activity Measure to calculate the player's Calorie consumption per day.

B. Digital Assets

The next component is the set of digital assets of the game, e.g. character models, location models, background art, music, sound effects, etc. These assets determine the theme and aesthetic value of the game, which in turn enhances playability and entertainment value; a necessity for both diet and exercise in terms of keeping the player's attention.

A prominent asset utilized in the Health Nexus game is the character avatar. The character avatar consists of a fully three dimensional skinned model of a human that is controlled by the player. Generating such a model required several steps: generating a tree-based skeleton hierarchy, skinning the hierarchy to humanoid geometry and texture mapping the geometry.

The first step is generating the skeleton hierarchy. As the Kinect beta SDK returns a set of 21 joint positions, so it is therefore sensible to make that set of joints into our skeletal structure set. The skeleton hierarchy generated is shown in Figure 2.

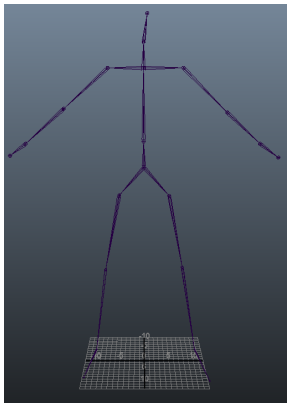


Figure 2: Skeleton Hierarchy

The second step in generating the avatar is skinning the skeleton to humanoid geometry. This is done by assigning each vertex in a mesh a set of influences. These influences are known as weights and determine the transformation effect that each joint in a hierarchy exerts on a given vertex in a mesh. When a joint in a skinned model is transformed from its original state, (e.g. rotation) each vertex that is influenced by the joint deforms based on the weight of the influence exerted.

The final step in generating a realistic 3D humanoid model is texture mapping. To produce a more realistic avatar, textures were applied to the surface of the mesh geometry to alter the physical appearance of the models, e.g. clothing textures, skin textures, textures for eye color, shoes, etc.

With the application of the above three steps two skinned and textured 3D avatars were generated. Players are allowed to select which avatar to use as a playable character in the game and the avatars will mirror any movements made by the players. The completed avatars are shown in Figure 3.

Figure 3: Male and Female 3D Avatars

C. *Remaining Software Components*

The remaining components consist of the profile and record, exercise database and food database. The profile and record keeps all the information pertaining to a particular player; it comprises a list of statistics and histories. This can be used both as a reference to the statistics of the player currently playing the game, in the case of the health model equations, for example, as well as to create a history of exercises and foods consumed. This is helpful from a dietary standpoint as the player may access all prior foods to see which combination of foods is more beneficial and also from an exercise standpoint as the player may reference how many calories were lost by performing exercises.

Finally we have the databases for exercises and foods. The former stores a series of gestures as interpreted by the Kinect, which at this time comprises of jumping jacks, running in place and leg lifts. This is an easily expandable component, as new exercises may easily be added into the database. The latter stores a list of foods and their nutritional facts, which are displayed to the player in the cafeteria when they are selecting their meals. This database is just as easily expandable as new foods may be appended as necessary.

III. Implementation and Results

The Health Nexus serious game consists of five major modes: Profile Selection, Cafeteria, Training Yard, Challenge and Results. These modes are the primary scenes that the player interacts with and are each critical to the gameplay mechanics.

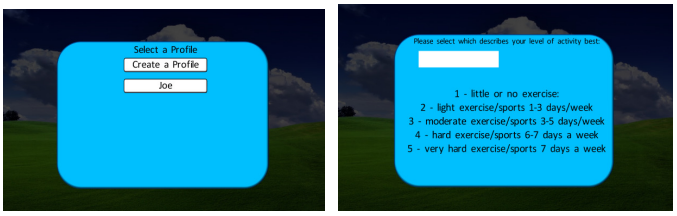


Figure 4: Profile Selection (left) and Creation (right)

The first game mode the player encounters is the profile mode. In the profile mode, players may choose a profile or create a new one. Each profile is unique to a player and stores the data about that player: name, age, weight, height, avatar, level of activity, cash and history. These variables comprise all relevant in game statistics for each player and are serialized to an XML file upon the exit of the game saving them for the next time the game is executed.

The cafeteria scene is where all diet related player interaction takes place. The player is transported into an outdoor cafeteria manned by the helpful Chef Sparty who gives the player a menu with which to choose their daily allotment of foods from. A screen capture of the cafeteria is shown in Figure 5.

Upon entrance to the cafeteria players are welcomed by Chef Sparty who presents them with the list of foods they may select on the left. Players may then select foods and check the nutritional values for one serving which are provided by Chef Sparty. The chef shows the player a current list of foods the player has selected along with the nutritional values of those foods including Calories, grams of fat and grams of sugar. For diabetics, the glycemic load will also be shown. Below the list of currently selected foods, an option to clear the diet and start over is given. Players may cancel or confirm food selections and finalize their food choices when finished picking a diet for the day. After diet finalization takes place, the list of food consumed and the current date are serialized to the player's XML file and the game proceeds to the exercise modes.

Figure 5: Cafeteria Mode

The exercise modes currently consist of three main exercises: leg lifts, jumping jacks and running in place. Players are required to choose at least one of these exercises and perform the exercise beyond a minimum length of time or number of repetitions. Players are awarded bonus Calorie loss and cash if they exceed the minimum exercise requirements or if they perform multiple exercises.

An important aspect of this mode is the detection of the exercises. Proper detection of the various exercises requires monitoring the movements of the player and identifying certain gestures. In the case of jumping jacks it was necessary to identify the location of both hands and feet and track them to determine if the positions were being executed in order before incrementing the number of jumping jacks performed. The game checks to see if the player's hands are within six inches of each other and higher than the player's head and if the feet are further apart than the player's knees; if so, the game recognizes that the player has assumed the motion position. To detect the rest position, the game checks to see if the feet are within six inches of each other and if the hands are held below the waistline. If the game detects the rest position after the motion position is identified, the count of jumping jacks is incremented. Running in place and leg lifts are detected in a similar manner. Screen captures for leg lifts and jumping jacks are shown as follows.

Figure 6: Jumping Jacks (left) and Leg Lifts (right)

After concluding their chosen exercises in the training yard, players are automatically entered into the challenge mode. The challenge encourages players to expend themselves by timing them during a short sprinting event. Players begin at a small track and must sprint 50 meters before ending the event. Players are rewarded larger amounts of cash for faster run times but players' max speed is dependent on current weight and dieting and exercise choices made in the prior modes. Poor choices such as overeating or minimal exercise result in a very low top speed whereas healthy choices result in a higher top speed.

Figure 7: Challenge Mode

Finally, after the player has progressed through the cafeteria, training yard and challenge modes, they may view the results of the day's exertions. Players are shown one of the overlays in Figure 8, depending on which gender character they selected.

Figure 8: Results Overlay

Figure 8 shows the two results overlays players may see. Each overlay lays out the current statistics of the player on the left side and the Calorie totals based on player statistics and decisions made throughout the prior three modes on the right. At the bottom of the overlay are three buttons which players may select. These buttons allow the player to see the results of their choices and exercises, in terms of weight gained or lost, over a period of a day, week or month. After viewing these results, the player may choose one of these periods of time to increment the game by. All choices made by the player are then applied each day for the duration of that time and the player is rewarded for his/her efforts.

IV. Conclusion and Future Work

This paper proposed a serious health game that helps to prevent and treat obesity and diabetes by promoting a healthy diet and regular exercise regimen. This game is driven by input from the Microsoft Kinect motion sensor which allows the player to control a skinned 3D avatar. The current edition of the game consists of multiple game modes including a cafeteria to select a diet, a training yard where the player may perform flexibility and aerobic exercises, a challenge mode to test the player's physical acumen and a results overlay which gives feedback to the player about the healthiness of his/her dietary and exercise choices.

Health Nexus in its current state is a prototype; it may be extended in several fashions. More exercises and foods could be added to the databases and new backgrounds and scenes could be added. The challenge mode could also be expanded to include several types of challenges such as cross country running and competition with computer controlled avatars. In game non-player characters such as the chef could be updated with predetermined animations to give the player the illusion of interaction with them. Finally, more 3D avatars could be generated and added into the game for the player to choose from, giving them a greater selection of characters and increasing the entertainment value of the game. All of these additional improvements could be added to the game in the form of optional upgrades purchasable from the in-game store which the player may purchase with the cash obtained through the game modes. All of these possible improvements would increase the educational and entertainment values of the game by providing players with a more interactive environment as well as a more realistic education on the relative value of various diets and exercises.

Once the final serious game is completed a validation study will be conducted to assess the effectiveness of the game in treating and preventing obesity and diabetes. This study will track user choices and effects in game and compare them to real choices the user makes outside of the game to determine how effective the game is at changing dietary and exercise lifestyles.

References

[1] Entertainment Software Association, "2011 Essential Facts About the Computer and Video Game Industry," 2011.

[2] M. Zyda, "From Visual Simulation to Virtual Reality to Games," *IEEE Computer,* vol. 38 pp. 25-32, 2005.

[3] The National Academies, "Improving Undergraduate Instruction in Science, Technology, Engineering and Mathematics," ed, 2003.

[4] Project Kaleidoscope, "Recommendations for Action in Support of Undergraduate Science, Technology, Engineering, and Mathematics," ed, 2002.

[5] American Association for the Advancement of Science, "Invention and Impact: Building Excellence in Undergraduate Science, Technology, Engineering and Mathematics Education," ed, 2005.

[6] NSF Task Force on Cyberlearning. (2008, Fostering Learning in the Networked World: The Cyberlearning Opportunity and Challenge.

[7] Centers for Disease Control and Prevention. (2010). *U.S. Obesity Trends*. Available: http://www.cdc.gov/obesity/data/trends.html

[8] Centers for Disease Control and Prevention. (2010). *Defining Overweight and Obesity*. Available: http://www.cdc.gov/obesity/defining.html

[9] Centers for Disease Control and Prevention. (2011). *About BMI for Adults*. Available: http://www.cdc.gov/healthyweight/assessing/bmi/adult_bmi/index.html

[10] Centers for Disease Control and Prevention. (2011). *Causes and Consequences*. Available: http://www.cdc.gov/obesity/causes/index.html

[11] Centers for Disease Control and Prevention. (2011). *National Diabetes Fact Sheet, 2011*. Available: http://www.cdc.gov/diabetes/pubs/pdf/ndfs_2011.pdf

[12] Centers for Disease Control and Prevention. (2011). *2011 National Diabetes Fact Sheet*. Available: http://www.cdc.gov/diabetes/pubs/general11.htm

[13] J. A. Harris and F. G. Benedict, "A biometric study of basal metabolism in man," in *Carnegie Institution of Washington publication, no. 279*, C. I. o. Washington, Ed., ed. Washington: Carnegie Institution of Washington 1919.

[14] C. L. Kien and F. Ugrasbul, "Prediction of daily energy expenditure during a feeding trial using measurements of resting energy expenditure, fat-free mass, or Harris-Benedict equations," *The American Journal of Clinical Nutrition,* vol. 80, pp. 876-880, 2004.

[15] R. E. Hasson, C. A. Howe, B. L. Jones, and P. S. Freedson, "Accuracy of four resting metabolic rate prediction equations: Effects of sex, body mass index, age, and race/ethnicity," *Journal of Science and Medicine in Sport,* vol. 14, pp. 344-351, 2011.

Simulation of Infectious Disease Spread in a Clinical Lab Setting: A Pilot Study

Lydia Wigglesworth-Ballard, Lynn Wiles, PhD, Holly Gaff, PhD

I. Introduction

Infection control in healthcare settings is essential for ensuring the safety of patients and those who work in these facilities. A high level of commitment towards adherence to best infection control practices is necessary to not only reduce, but also potentially prevent transmission of Healthcare-associated infections (HAIs). Healthy People 2020, managed by the U.S. Department of Health and Human Services (HHS), have set objectives to reduce and potentially eliminate catheter line-associated bloodstream infections (CLABSI) and invasive healthcare-associated methicillin-resistant Staphylococcus aureus (MRSA) infections [1]. This research focuses on the use of Simulated Infectious Diseases (SID), a powder-like substance to simulate infection during clinical procedures. SID becomes visible under ultraviolet light demonstrating how infections are spread or contained following simulated training sessions. Simpler substances present logistical issues for complex training environments beyond simple hand washing and may mechanically or cosmetically damage delicate medical equipment and simulators.

II. Problem

A Healthcare-associated Infection (HAI) is an infection that patients acquire in a healthcare setting, while being treated for another condition for which they were originally admitted [2]. Depending on the type of HAI, patients who become infected can have symptoms ranging from redness, swelling and pain at the sight of the infection to even more problematic, systemic conditions such as sepsis, which can lead to organ failure and even death [3]. This increased morbidity and mortality is an area of concern for the health and economic outcomes for both the patient and the healthcare facility. In addition, these infections are often resistant to at least one antibiotic, and so eliminating these would benefit society at large as well. The pathogens that cause these infections are not visible with the naked eye, therefore, creating challenges with the current methods used to train students and healthcare workers. With the current methods used for training, it is very difficult to assess individual compliance and the transfer in promoting individual commitment to infection control.

Lydia E. Wigglesworth-Ballard is a doctoral student in Health Services Research in the College of Community and Environmental Sciences, Old Dominion University, Norfolk, VA 23529 USA (email: lwigg005@odu.edu).

III. Pilot study

The author has conducted a pilot study to test SID in a clinical lab with students, study protocols, and design. The Old Dominion University Institutional Review Board (IRB) application was approved for this study.

The specific aims for this study was to: (1) incorporate SID into training scenarios using patient simulator mannequins to train nursing students in an interactive simulation environment by visually replicating spread of infectious bacteria while performing common clinical tasks in a virtual setting, (2) evaluate healthcare student's performance by measuring the magnitude and distance of SID spread from the original location photographically, and (3) compare outcomes of infection control knowledge, opinions, and behaviors after completing the training with SID versus traditional training.

IV. Methods

Participants were recruited through the Clinical Management Adult Health Nursing III: Critical Care course and were divided into two groups. SID was placed on specific mannequins to represent infected patients. Each group was instructed to perform the clinical tasks on simulated patient mannequins that were infected and not infected with SID. The participants were told that the purpose of this study is to test if the video camera equipment can be used for instructors to evaluate and grade students from outside of the skills lab. Scenario Group 1 was asked to perform a tracheotomy cleaning, wound dressing change, and vitals. They were not informed of the presence of SID until after the lab session was complete. Scenario Group 2 was asked to perform a tracheotomy cleaning, wound dressing change, and vitals. They were informed of the presence of SID immediately before the lab session began. Both groups were instructed to perform the clinical tasks for their patient while interacting with the patient as they would in the real world.

An infection control knowledge questionnaire was obtained from the Institute for Healthcare Improvement (IHI), a collaborative group that consist of the Center for Disease Control and Prevention (CDC), The Association for Professionals in Infection Control and Epidemiology (APIC), and the Society of Healthcare Epidemiology of America (SHEA). The participants were given the IHI questionnaire and they were debriefed immediately at the conclusion of the lab session regarding the true nature of the study. The participants also participated in a second debrief at the end of

the semester where they were able to discuss how they felt about the study, ask questions, and express any concerns. The lab sessions were videotaped and photographs were taken of the patient simulator mannequins and surrounding work area to document the distance and the magnitude of spread from its original location.

V. Results

Because of unforeseen circumstances the scenario for group 2 was changed immediately before the lab session. With this change in mannequins, the new mannequin was not set up for a tracheotomy cleaning , so resuscitation was performed instead. This made it difficult to compare the amount of spread between the two groups. The participants were observed having no interaction with the simulated patient mannequins while performing the clinical tasks, as instructed. Participants were also unexpectedly prompted and reminded about hand hygiene and infection control while performing the clinical tasks. The participants in Group 2, who were informed of the presence of SID prior to the lab session, were observed to have significantly higher vigilance of hand hygiene than Group 1. The distance and magnitude of spread were consistently different based on the type of clinical task performed.

VI. Lessons Learned

The original study design sought to obtain some measurable data to determine if there was a difference in spread between group 1, who had their lab session at the beginning of the semester course and was informed of SID after the lab session, and group 2, whose lab session was later during the semester course and was informed of SID immediately before the lab session. Rather than a distinct set of measurable data, this pilot study provided lessons learned that has been very valuable in the design for the main study.

A. Scenarios

It was difficult for the participants to obtain a real world perspective with the faculty and assistants interacting with the students during the lab session. The participants were also grouped as three to each mannequin. This provided an environment for them to interact with each other and distracted them from interacting with the patient. This also made it difficult to determine which participant contributed to the spread. The main study will include a more structured environment and observation through video only.

B. Measurement of Spread

During preliminary testing of SID, spread was defined as any relocation of SID from its original planted location to another location on the mannequin or surrounding clinical area, including the student (i.e. no spread or spread). The results of the pilot study showed that the type of spread varied with the type of clinical procedure. Data collected on actual distance and magnitude of spread will need to be compared to a baseline based on the type of clinical procedure.

VII. Acknowledgment

This work was conducted with the assistance and support of Dr. Holly Gaff and is gratefully appreciated. An appreciation and thank you also to Dr. Ginger Watson and Dr. Lynn Wiles for assistance with this research.

References

[1] U.S. Department of Health and Human Services. Office of Disease and Health Promotion. Healthy People 2020. Washington, DC. Available at http://www.healthypeople.gov/2020. Accessed February 6, 2012.

[2] The Centers for Disease Control and Prevention. Available at http://www.cdc.gov/hai. Accessed February 6, 2012.

[3] U.S. National Library of Medicine. Bethesda, MD. Available at http://www.ncbi.nlm.nih.gov/pubmedhealth/PMH0001687. Accessed February 6, 2012.

Simulating for Social Change: An ABM of Grassroots Organizing to Combat HIV/AIDs in South Africa

Erika Frydenlund and Savannah L. Eck

Abstract— **The HIV/AIDS pandemic has devastated traditional family and social network structures around the world. In the township of Khayelitsha, on the outskirts of Cape Town, South Africa, high HIV/AIDS infection rates and subsequent deaths have claimed a generation of parental-aged citizens, often leaving elderly women as sole guardians of their grandchildren. In addition to the financial strain of families built around an elderly woman head of household, misinformation about HIV/AIDS continues to threaten the well-being of their grandchildren's generation. Through grassroots efforts, the women of Grandmothers Against Poverty and AIDS (GAPA) reach out to residents of Khayelitsha to dispel myths, protect children from violence and infection, and break the taboos and social norms surrounding sexual education. This project utilizes a completed ethnographic-based master's thesis from the Graduate Program in International Studies at Old Dominion University to examine how agent-based modeling can advance traditional qualitative research. Using an agent-based model (ABM) based on GAPA, this project examines the efficacy of a grassroots organization to spread accurate information and education about HIV/AIDS through the community via the members' social networks. The researchers advocate for use of this tool to promote policy development that supports grassroots movements to improve community-based human security.**

Index Terms—**HIV/AIDs, agent-based modeling, grassroots activism, social networks, community outreach**

I. Introduction

THE purpose of this paper is two-fold. First, the authors intended to supplement existing interview-based data with quantitative pseudo-data in order to make a stronger argument for government support to grassroots organizations. In particular, the Grandmothers Against Poverty and AIDS of Khayelitsha Township in South Africa serve as a case study on how grassroots activists can reach out into their local community and begin to change belief systems and provide current information that helps stem the transmission of HIV/AIDS. Second, the authors intend through this work to serve as yet another example in the wider Modeling and Simulation community of the benefits of cross-disciplinary research that engages data collected in the field and extends it in ways that hold the potential to affect wider policy change. Rather than view this paper as a final project, the authors hope the reader will internalize its greater purpose of serving as an intermediary between the traditional binary of qualitative and quantitative research in academia.

Manuscript received March 9, 2012.

E. Frydenlund is a doctoral candidate in the Graduate Program in International Studies and Department of Women's Studies at Old Dominion University in Norfolk, VA 23529-0086 USA. E-mail: efryd002@odu.edu.

S. L. Eck is a master's student in the Graduate Program in International Studies and Department of Women's Studies at Old Dominion University in Norfolk, VA 23529-0086 USA. E-mail: seckx001@odu.edu.

II. Background

A. HIV/AIDS as a Human Security Problem

As many elder African women face continual socio-economic challenges, HIV/AIDS brings a new decimating force to communities and social relationships. In the wake of its aftermath, a generation is lost and children are left in the care of elder women. Women continually pick up the pieces as traditional family units crumble around them. As a direct result of HIV/AIDS, the most heavily hit populations are the very young and very old. The joint United Nations Program on HIV/AIDS (UNAIDS), along with the United Nations Children's Fund (UNICEF), estimates that more than 12.3 million, or 12% of African children, are orphans, while measures from South Africa and Uganda reveal that 40% of orphaned children are living with a grandparent [1]. In order to combat this disease and ensure the security and stability of Africa and the globalized world, a "broad approach in which the marginalization as well as their [women's] lack of provisions for sustaining their health are seen as dual, but interrelated, requirements that must both be addressed for the effective participation of women in the struggle to arrest the spread [2]." Women, suffering at the hands of HIV/AIDS, serve as caretakers, support systems and economic breadwinners, thus creating a gendered dynamic of the pandemic. They are therefore distinctly situated as vital agents of human security and social change.

Geographic and historical context is central to this study because the former apartheid system of governance in South Africa created a severely unequal social and economic environment conducive to the massive spread of HIV/AIDS. The World Health Organization (WHO) maintains that, "HIV infection thrives on poverty and marginalization. The epidemic is sustained by social disruptions; by historical inequalities of wealth, gender and race; and by migrant labor practices [3]." Additionally, "in late 2004 the South African Department of Health estimated that there were 6.28 million

infected South Africans, the largest national total anywhere [4].” According to Ostergard, “For South Africa, the security impact of the epidemic is already showing. From the civil service to the economy, South Africa’s political and economic security has been compromised from the strain that the epidemic has placed on resources and manpower [5].” Government involvement in stalling the spread of HIV/AIDS, particularly in townships, has remained appallingly low. One research notes, “…the massive debt remnants from apartheid, severe poverty, and a lack of social cohesion surpassed HIV/AIDS concerns on the national agenda. In retrospect, the idealism that swept the Mandela era did not live up to the goals of the National AIDS policy [6].”

From the larger context of post-apartheid South Africa and the continued spread of HIV/AIDS, a focus is required on the regional context under which GAPA grandmothers live, work and struggle. Established in the mid-1980s as a reaction to the forced residential segregation policies of apartheid, Khayelitsha, located 25 miles outside of Cape Town, served as a relocation site for those designated as “African” in the former classification system. Meaning “new home” in traditional Xhosa language, Khayelitsha boasts a population of over 300,000 individuals according to the 2001 census report [7]-[8]. Population estimates in South Africa are often considered low because of the inability to enumerate all individuals in a township due to population density and ongoing security concerns. In 2001, 50.8% of Khayelitsha’s population was unemployed and the majority of households (57.4%) lived in “shacks” in informal settlements; 64.6% of the total population resides in informal dwellings [8]. Rape and sexual violence are rampant in correlation with high unemployment levels [9].

By understanding both the geographic and historical context, the conditions under which HIV/AIDS is exacerbated and the reasons for grassroots campaigns such as GAPA to combat the disease is clear. The pressing dilemmas of HIV/AIDS, security and underdevelopment form the main components of analysis to explore social change and resistance through the lens of elder women activists and community leaders within Khayelitsha.

B. GAPA and Grassroots Organizing

Beginning as a self-help project aimed to assist grandmothers coping with the psychosocial aspects of poverty and AIDS in their families and communities, Kathleen Broderick, an occupational therapist from the University of Cape Town worked with ten grandmothers in home-based support projects. Following this project, many grandmother members collectively called for a larger implementation of the program to help elder women in the community of Khayelitsha.

Grandmother run and executed, GAPA serves as a model for the advancement of social, political and economic development on a community-based level. Founded in 2001 as an official non-profit organization, GAPA now boasts around 300 grandmother members and runs workshops focused on various topics including but not limited to HIV/AIDS education, gardening and business skills, will writing and bereavement support. GAPA members must be age fifty or above and be directly impacted by HIV/AIDS either themselves or through their families.[1]

Various elements of security, economic support and education play vital roles in the mission of GAPA. Certain insecurities create vulnerabilities specific to elder women, who struggle against the dual burden of serving as both the provider and caregiver. GAPA's model encompasses every element of human security as defined by the United Nations Development Programme and works to create stability and support in the midst of poverty and widespread HIV/AIDS.

Already, GAPA has shared its message in other regions of Africa and is working to replicate their organizational model, which has been so helpful for elder women in Khayelitsha Township. As the Madrid International Plan of Action for Ageing states, “rather than viewed homogeneously as vulnerable, dependent and passive recipients of resources, older persons should be mainstreamed into societal and developmental processes as continuing agents of change and contributors to and beneficiaries of development [10].” GAPA is equipped to provide elder women with a productive and applicable model for socio-economic development. As Sanam Anderlini states, “They are committed because it is their own lives: the political is deeply personal. They have no exit strategy. Time and again, women prove the invaluable contributions they can make, and their willingness to work for peace [11].”

C. The Problem

One of the core functions that GAPA serves in its community of Khayelitsha is that of HIV/AIDS-related education outreach. In particular, the Grandmothers strive to provide current, valid, factual information to community members struggling with the epidemic. Through this activist community engagement, GAPA members help to dispel social and cultural taboos surrounding public dialogue about sexual education. This is achieved both through one-on-one interactions with community members as well as monthly meetings, various topic-specific workshops and outreach to community leaders.

Focusing only on the educational outreach aspect of GAPA’s many community service projects, the Grandmothers can be viewed as “mavens” who act as “information brokers” within their community [12]. From qualitative interviews gathered from influential Grandmothers within GAPA arises the research question: how effective is a grassroots campaign based on word-of-mouth strategies in educating the public about HIV/AIDS? The aim of this research is to develop, based on previously collected ethnographic data, a quantitative model of GAPA’s education outreach efforts in order to enhance the argument that additional government funding for grassroots organizations can, in fact, alleviate the effects of HIV/AIDS in affected communities.

III. METHODOLOGY

This study focuses on an Agent-Based Model (ABM) inspired by ethnographic data collected in Khayelitsha

[1] For more information, see the GAPA website at: http://www.gapa.org.za/

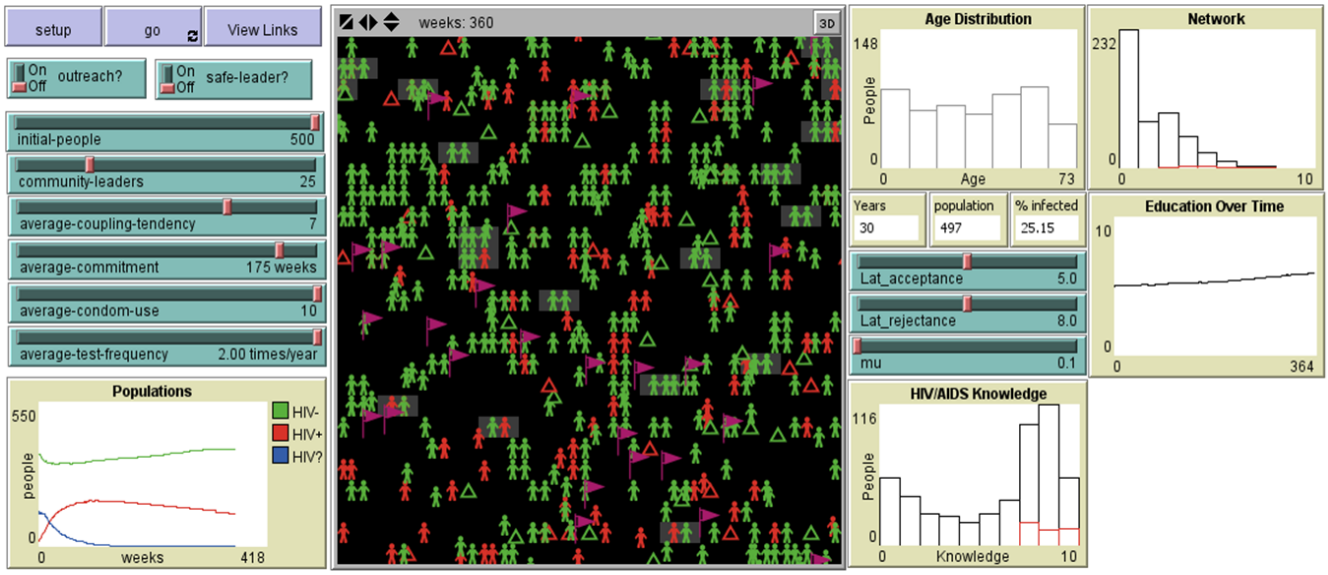


Figure 1: NetLogo Interface for GAPA ABM

Township between July 2010 and July 2011.[2] As mentioned above, collecting data in the insecure environment of a South African township presents challenges to traditional ethnographic field researchers. First, field research itself is a costly endeavor. Additionally, in the townships of South Africa, the social stigma of HIV/AIDS requires survey participants to risk social, economic and political costs thus making it difficult to conduct interviews on sensitive issues surrounding sexual behavior and belief systems.

When looking at the effects of grassroots community outreach, ethnographic field research is further complicated by the devastating effects HIV/AIDS has had on social networks. The disease causes disruptions in ties between friends and families as people fall ill or die, resulting in dynamic rewiring of these networks. A snapshot in time, captured by a field researcher, cannot by itself fully depict the rapidly shifting nature of these social ties and the effects these might have on grassroots efforts.

ABM, by considering the individual, autonomous and heterogeneous decisions of actors within the system provides the unique capability of incorporating field research into a dynamic system of social networks. Local interactions of agents within the model mimic communication between individuals in the real world. By growing the system from the "bottom up," ABM provides an experimental space in which the researcher can explore different combinations of "treatments" upon the population [13]. ABM also gives the researcher the ability to examine multitudes of nonlinear factors interacting in complex ways.

In the case of GAPA's community outreach efforts, ABM serves as a tool to help researchers quantitatively evaluate effective strategies that Grandmothers employ to educate neighbors and change social taboos surrounding public discussion of HIV/AIDS. The researcher can look at how information would spread if GAPA only used word-of-mouth tools versus targeting influential people to spread knowledge on their behalf. This information can then support interview data when advocating for increased government involvement and financial support for organizations such as GAPA.

IV. Model Development

The model was developed using NetLogo, a well-known open-source software for coding ABMs [14]. In addition to the obvious benefits of free software that can be readily shared between researchers, the developers of NetLogo encourage a community of model sharing that allows developers to build off of and learn from the ABM code of others.

A. Underlying Model Extensions

The GAPA ABM is an extension and hybridization of two smaller models. First, the agents in the model enacting the spread of HIV/AIDS are using procedures (algorithms) from an AIDS ABM written by NetLogo's founder [15]. As the focus of the research on GAPA is education outreach and its effect on HIV/AIDS, not HIV/AIDS directly, this previously written model was used to establish the epidemiological basis of the current model. Second, coding for the ways that GAPA spreads information through connectors in the population was inspired by another previously written model on "mavens" [16]. Finally, the ways in which GAPA agents are able to affect the belief system of others in the model is based on the mathematical design of Jager and Amblard [17]. This is not to suggest that the GAPA ABM coding is unoriginal, but rather to illustrate how the greater ABM community has become one of collaborative modeling efforts.

[2] For an intensive qualitative analysis of this data, see: Eck, Savannah. "Securing South Africa's Future: Grandmothers Against Poverty and AIDS as a Model for Social Development Change." Master's Thesis, Old Dominion University, 2012. Unpublished.

B. Model Assumptions

First and foremost, this model is not striving to provide answers to the spread of HIV/AIDS or in any way imply that grassroots outreach is the "cure." This model assumes that HIV/AIDS transmission is occurring at an alarming rate within Khayelitsha Township and that education-based outreach can have an effect on its transmission. The model is seeded with data specifically relevant to Khayelitsha where available because this is the current environment in which the Grandmothers operate. The model adopts Wilensky's HIV/AIDS transmission assumptions[3] as well as the assumptions related to attitude shifts adopted by Jager and Amblard [17]. The researchers assume that GAPA agents have the ability to shift the perceptions of others to varying degrees and in varying directions. Not all outreach attempts are successful in this way.

C. Factor Selection

As seen in the GAPA ABM interface (Figure 1), there are numerous variables interacting within the model. The most important factors to note for the purposes of this study involve the spread of information/ideals and the subsequent effect on risk factors for contracting or transmitting HIV/AIDS through the population. For simplicity, the researchers have dubbed the variable "education," but this encapsulates more than simply factual information about HIV/AIDS. While this type of information is provided by GAPA members to the community, the actual "educational" impact is at a deeper ideological level. This education variable is more of a "belief system" factor that incorporates both the amount of factual knowledge an agent has about HIV/AIDS as well as her own beliefs about its transmission and associated risk-related behaviors. This is an important distinction because GAPA does more than simply act as replacement-teachers, they become community mentors who evoke social change through the disruption of social norms and taboos surrounding public discussion of HIV/AIDS; this is their unique contribution to society.

When GAPA agents in the model shift the belief systems of other agents, an additional shift occurs in risk-taking behavior related to HIV/AIDS transmission. The three factors that a shift in belief affects are the average length they commit to a relationship, the propensity for condom use, and the frequency with which they get tested for HIV/AIDS each year. Of course, in reality, GAPA's outreach encourages other behavioral changes in the population such as increased nutrition and better understanding of treatment regimens, but for the illustration of GAPA's far-reaching effects, only these three factors were selected. Future models could incorporate more elaborate factors and possible interactions with changes in belief systems.

V. The Model

A. Lifecycle of Events

As described above, the majority of agents in the model are "people" with heterogeneous belief systems, ages, sex and HIV/AIDS related risk behaviors. These agents interact and transmit HIV/AIDS at a rate consistent with current estimates of Khayelitsha. A small number of female agents over the age of 50 (requirements of the organization) are members of GAPA. These agents attempt to change the belief systems of other agents through two types of interactions. First, GAPA agents approach individuals if they are nearby and share their own beliefs. Second, GAPA agents approach "community leaders," those known to have large social networks and represent educators and religious leaders, among other high prestige individuals in the community. Occasionally GAPA strategies will focus only on community leaders who are known to be "safe," or non-perpetrators of violence or intentional perpetuators of misinformation.

As laid out by Jager and Amblard, when a GAPA member approaches an agent, she attempts to shift his belief system. Each agent has a certain propensity for being convinced. If the agent finds the Grandmother's beliefs to be within a range of acceptance, the agent will shift his viewpoints toward hers and adjust his behavior to be less risky. If, on the other hand, the agent finds the Grandmother's beliefs to be unbelievable, he will instead shift in the opposite direction and adjust his behaviors to be more risky. The thresholds for acceptance and rejection of new beliefs as well as the weight placed on an agent's change are interactive in the user interface.

B. Verification and Validation

Several researchers participated in the concept development of this model resulting in pseudo-code that has been verified by field researchers of Khayelitsha as adequately reflecting social dynamics in that township. Additionally, the final coding for this model has been reviewed by an outside researcher for errors or biased programming. While still in its final prototype stages, the model accurately reflects birth, death and HIV/AIDS transmission rates in the Khayelitsha population based on currently known biometrics. The City of Cape Town, South Africa is conducting its 10-year census in 2012 and the results will be used to seed this model with even more accurate metrics. Based on what is currently known about the population, however, the model reflects future estimated population pyramids [18].

VI. Experiments

The BehaviorSpace feature of NetLogo was used to automate experiments and pseudo-data collection. Due to resource constraints, several factors were held constant while conducting these initial experiments on the model. Future experimentation will allow for increased variation of these factors. To reduce computational costs, the following factors were constrained:

- Population, fixed at 500
- Number of community leaders, fixed at 25 or 5% of the population
- Number of years, fixed at 30 into the future
- Weight placed on belief changes, fixed at 0.10
- Latitude of acceptance, fixed at 5
- Latitude of rejection, fixed at 8

Three experimental groups were established as follows: (1) GAPA agents only educate others on a one-on-one basis, (2)

[3] For details of this, see the documentation section of Wilensky's AIDS model [15].

GAPA educates one-on-one *and* reaches out to "safe" community leaders, and (3) GAPA educates one-on-one and reaches out to *all* community leaders. For each of these experimental groups, the risk behaviors associated with tendency for sexual contact, length of relationship commitment, condom use, and HIV/AIDS testing frequency were varied across their respective ranges. The average education of the population and HIV/AIDS infection rate was recorded for each run. Each combination of factors was repeated 10 times for a total of 9,000 runs.

VII. Results

In order to determine differences between the experimental groups, t-tests were conducted for each treatment combination. All groups were statistically significantly different from one another, with p-values extremely close to zero. Of particular interest, treatment group (3), where GAPA members educate one-on-one and access only "safe" leaders, the mean of the education variable was higher (5.5679 versus 5.1659) than the experiment group where all leaders are approached. This difference is also noted in the infection rate for those respective experimental groups (0.3486 versus 0.3874). Experimental group (1) experienced higher average education and lower infection rates in comparison with the other two groups.

The researchers anticipate that these unexpected results are actually derived from a coding anomaly which will be fixed in future versions and re-evaluated. Currently, when GAPA's access to community leaders in the model is turned off, community leaders are not influencing the behaviors of their networks. In the future, the networks should disseminate information whether or not being approached by GAPA Grandmothers which will create a natural negative feedback in the model.

VIII. Conclusion

The GAPA ABM is still in the late prototype development stage and there remain some coding irregularities to be pursued, as mentioned above. Despite this, the model's applicability to real-world policy advocacy is quite promising. In future versions of the model, the researchers intend to explore some of GAPA's other outreach programs such as health and nutrition workshops, community gardening and childcare as a means to deflect sexual assaults on neighborhood youth. GAPA provides many services to the community that ultimately contribute to increased human security within Khayelitsha, and with time, these factors will become part of a model that represents the full reach of grassroots activist efforts.

Acknowledgment

The authors would like to thank David C. Earnest for his assistance in both concept development and review of materials required in the development of this model. We would also like to thank Jennifer N. Fish for her help in model development and promoting this research in the greater academic community of Sociology.

References

[1] Jackson, Hall and Mhambi in Fuller, Linda K. *African Women's Unique Vulnerabilities to HIV/AIDs: Communication Perspectives and Promises*. New York: Palgrave Macmillan, 2008. Print. p. 8.

[2] Ankrah in Fuller p. 15.

[3] Shisana, Olive, Nompumelelo Zungu-Dirwayi and William Shisana. "AIDS : a threat to human security." in Chen, Lincoln C, Jennifer Leaning, and Vasant Narasimhan. *Global Health Challenges for Human Security*. Cambridge, Mass: Global Equity Initiative, 2003. Print. p.143.

[4] Fourie, Pieter. *The Political Management of HIV and AIDS in South Africa: One Burden Too Many?* Basingstoke England; New York: Palgrave Macmillan, 2006. Print. p.xvi.

[5] Price-Smith, Andrew, Matthew Tubin and Ostergard, Robert L. *HIV/AIDS, and the Threat to National and International Security.* Houndmills, Basingstoke, Hampshire; New York: Palgrave Macmillan, 2007. Print. p. 241.

[6] Fourie, Pieter. The Political Management of HIV and AIDS in South Africa: One Burden Too Many? Basingstoke England; New York: Palgrave Macmillan, 2006. Print. p. 100-101.

[7] "Baphumelele Caring for Community." Baphumelele Children's Home. January 4, 2011 <http://www.baphumelele.org.za/information/>.

[8] "A Population Profile of Khayelitsha: Socio- Economic Information from the 2001 Census." Census in Brief. Government of the City of Cape Town, South Africa. 2001. <http://www.capetown.gov.za/en/stats/CityReports/Pages/PopulationProPopul.aspx>

[9] Muula, Adamson S. "HIV Infection and AIDS Among Young Women in South Africa." Croatian Medical Journal. 2008 June; 49(3): 423–435. doi: 10.3325/cmj.2008.3.423.

[10] Ferreira. M. 2008. Training in Africa towards forging implementation and monitoring of the Madrid International Plan of Action on Ageing. BOLD (Journal of the International Institute of Ageing (United Nations-Malta)), 18(3): 2-6. p. 2.

[11] Anderlini, Sanam Naraghi. *Women Building Peace: What They Do, Why It Matters*. Boulder: Lynne Rienner Pub, 2007. Print. p. 232.

[12] Gladwell, Malcolm. *The tipping point: how little things can make a big difference*. Boston: Little, Brown. 2000. Print. p. 69.

[13] Epstein, Joshua M. *Generative Social Science: StudiesiIn Agent-Based Computational Modeling.* Princeton: Princeton University Press. 2006. Print. pp 5-7.

[14] Wilensky, U. *NetLogo*. Center for Connected Learning and Computer-Based Modeling, Northwestern University. Evanston, IL. 1999. Software. <http://ccl.northwestern.edu/netlogo/>.

[15] Wilensky, U. *NetLogo AIDS model*. Center for Connected Learning and Computer-Based Modeling, Northwestern University, Evanston, IL. 1997. <http://ccl.northwestern.edu/netlogo/models/AIDS>.

[16] Earnest, David C. *NetLogo Mavens model*. Old Dominion University, Norfolk, Virginia. 2011.

[17] Jager, Wander and Frederic Amblard. "Uniformity, Bipolarization and Pluriformity Captured as Generic Stylized Behavior with an Agent-Based Simulation Model of Attitude Change." *Computational & Mathematical Organization Theory*, 10, 2004. p. 295–303.

[18] U.S. Census Bureau. *The International Data Base (IDB): South Africa.* Census Bureau: International Programs Center for Demographic and Economic Studies. 2011.

3D Meshless Enrichment Technique for Handling Discontinuities in Soft Tissue Models

Rifat Aras, Yuzhong Shen, Michel Audette

I. INTRODUCTION

In anatomy, the soft tissue term refers to tissues that connect, support, or surround other structures and organs of the body. Connective tissue that includes tendons, ligaments, fascia, skin, fibrous tissues, and fat, and non-connective tissue that includes muscles, nerves, and blood vessels are types of soft tissue [1]. There has been numerous approaches towards accurate modeling of soft tissue. Among these are: *free-form deformations* [2], *mass-spring models*[3], and *finite element modeling* (FEM) [4-6].

In this paper, we are adopting a numerical method of another class called *meshless methods*. Unlike the FEM, meshless methods rely entirely on nodes to approximate the physical problem at hand. Furthermore, meshless methods do not require a background mesh to define the neighborhood relation of the nodes, instead they work on overlapping support regions of a point cloud as opposed to FEM's non-overlapping element based representation. FEM techniques have been widely used to model physical phenomena, where the domain of interest supports the continuum assumption. When the target problem includes extreme mesh distortions, moving discontinuities that do not align with the element edges such as propagating cracks, and advanced material transformations such as solid-liquid phase transition, the continuum assumption disappears and FEM techniques become impractical. To alleviate these limitations, various meshless methods have been developed [7].

II. PROBLEM / MOTIVATION

Several papers have been published on the application of meshless methods to soft tissue problems. The work of Guo and Qin [8] focuses on Element Free Galerkin (EFG) method with a Moving Least Square (MLS) based approximation scheme. The 3D object to be modeled is represented as an implicit surface, which is obtained through an octree-based distance field computation. The authors justify the use of implicit surfaces because of their properties like trivial inside / outside test and the ability to update the topology of the implicit surface easily without any ambiguity. Furthermore, the octree structure is used for hierarchical sampling of the meshless nodes in the volume, and its cells are used as integration cells, which makes this paper different than the others, regarding the spatial integration method (cell integration vs. nodal integration). Octree structure also helps keeping track of the neighborhood information for the meshless nodes. In their 2010 work, Zhu et al. [9] propose a unified particle-based simulation framework. The framework is capable of simulating laparoscopic surgery procedures and organ deformation, controlling the movement of surgical instruments and detecting collision between the instruments and the physical model, and computing haptic feedback. CT image data is used to extract both physical points and the surface of the organ for visualization. The elasticity of the soft tissue is modeled by Smoothed Particle Hydrodynamics (SPH), in

which a function $f(x)$ in the domain Ω is approximated as:

$$f^h(\boldsymbol{x}) = \sum_I w(\boldsymbol{x} - \boldsymbol{x}_I) f(\boldsymbol{x}_I) \Delta V_I \qquad (1)$$

where $w(\boldsymbol{x} - \boldsymbol{x}_I)$ is the weight function and ΔV_I is the volume quantity associated with the node I. In this work, the neighborhood and volume information for the node are precomputed, which can cause problems with complex interactions with the model such as cutting.

In their recent work, Hao and Huang [10] present their soft tissue deformation implementation that uses meshless method. This paper follows the work of Müller et al. [11]. The soft tissue that is modeled is assumed to be made of Hookean material. MLS approximation is used to define the shape functions associated with the individual nodes. Computation of the shape function is followed by the calculations of the deformation gradient, strain, stress, and internal and external force components. This work does not contain any discussion related to the validation of the soft tissue model and whether it is possible to extend the technique for more complex operations such as cutting of the tissue.

Many engineering problems involve some sort of discontinuity in the domain of interest. Discontinuities may be caused when the object is composed of different materials or when there is a crack / cut between the meshless nodes. Such discontinuities require special treatment in meshless methods. there are three main classes of techniques to treat discontinuity of the field variable (displacement). These techniques are: (1) modification of the weight function, (2) intrinsic enrichment of the basis of the approximation, and (3) the techniques based on extrinsic enrichment.

Visibility method is an example for the techniques that modify the weight function. In this method, the crack segment is treated as an opaque object and the influence of a node to another one is decided by drawing a line between the two, and testing whether the line intersects with the crack segment or not. Although being simple in nature, this method can lead to unwanted discontinuities along the lines that pass through the crack tips. Another disadvantage of this method is that it cannot be used to treat non-convex boundaries.

Element Free Galerkin (EFG) method is a meshless technique that uses intrinsic basis for approximating the field function. To model strong discontinuities (discontinuities involving the field variable), the intrinsic basis can be modified by using the information from the crack geometry. For a 2D problem domain, the linear basis

$$p^T(\boldsymbol{x}) = [1, x, y] \qquad (2)$$

can be extended to

$$\begin{aligned} p^T(\boldsymbol{x}) = [1, x, y, \\ \sqrt{r}\sin\left(\frac{\theta}{2}\right), \sqrt{r}\cos\left(\frac{\theta}{2}\right), \\ \sqrt{r}\sin\left(\frac{\theta}{2}\right)\sin(\theta), \sqrt{r}\cos\left(\frac{\theta}{2}\right)\sin(\theta)] \end{aligned} \qquad (3)$$

where r is the radial distance to the crack tip and θ is the incident angle to the crack [12]. A disadvantage of the intrinsic enrichment technique is the additional computational cost that comes from the increased size of the basis, which is not suitable for real-time operation requirements. This work focuses on handling discontinuities in meshless methods by using 3D enrichment technique. This paper describes the on-going work, through which we are extending to 3D the recently published technique of Barbieri et al. [13]. The strength of this technique comes from its ability to apply new enrichments due to new crack segments in a

multiplicative way, its low computation cost, and its ease of integration to an existing method.

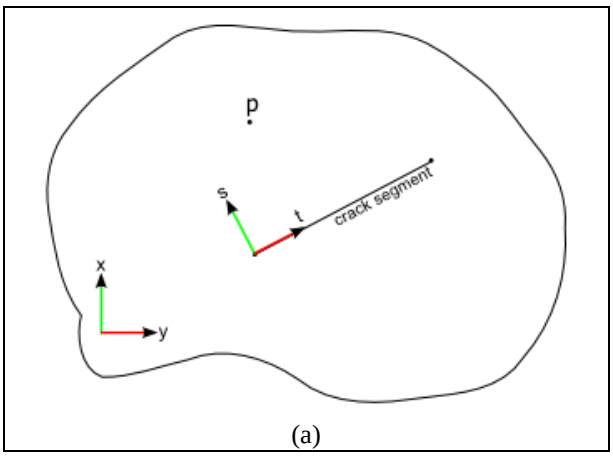
III. SIGNED DISTANCE FUNCTION BASED MESHLESS ENRICHMENT

Barbieri et al. [13] proposed an enrichment technique based on a signed distance function for handling discontinuities with multiple boundaries. Their method processes cracks as piecewise linear segments and calculates the signed distances of a meshless node to these segments that are later on multiplied with the shape function of the node.

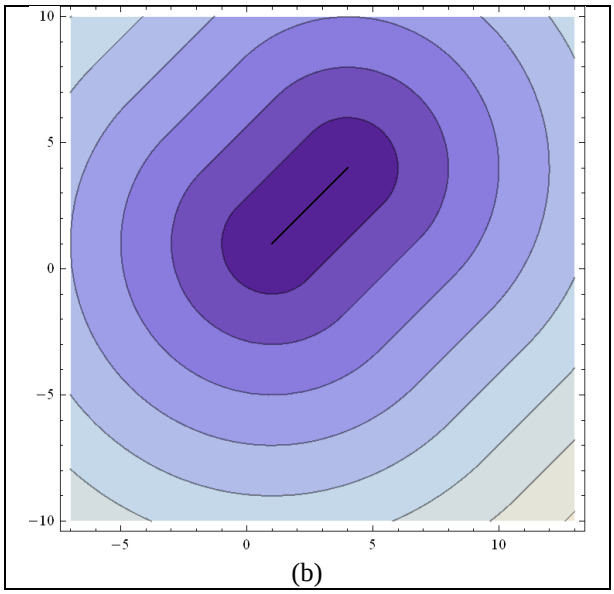
The distance function is computed in the local coordinate system of the crack piece (Figure 1). The 2D distance function for a given coordinate (x, y), can therefore be computed in terms of the local coordinates (s, t) as follows (Figure 1):

$$d_2(x, y) = \sqrt{d_1^+(t)^2 + s^2} \qquad (2)$$

where $d_1^+(t)$ is the positive part of the 1D signed distance function for a 1D segment (in the local coordinates). When we take the partial derivative of this distance function with respect to the normal coordinate axis s, we obtain a discontinuous function across the segment that is 1 on one side of the crack and -1 on the other side and varies smoothly around the crack (Figure 2).



(a)



(b)

Figure 1. (a) The local coordinate system of a crack segment. The coordinates of the meshless nodes and integration points are transformed from the world coordinate system (x,y) to the crack's local coordinate system (s,t). (b) The 2D distance function for a given crack line segment.

The proposed distance function can be extended to 3D. In our implementation, we realized the crack / cut surface as a triangle strip. Therefore, the 3D distance function is computed in the triangle's local coordinate system (Figure 3). For a triangle ABC, the planar signed distance of a point P with local coordinates (s, t, u) can be computed as:

$$sd_2^{ABC}(s, t, u) = Min\left(d_2^{AB}(s, t), d_2^{BC}(s, t), d_2^{CA}(s, t)\right) \qquad (3)$$

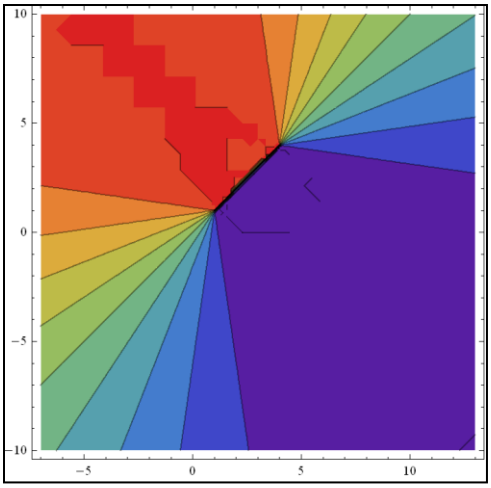


Figure 2. The partial derivative of the distance function with respect to the normal coordinate axis.

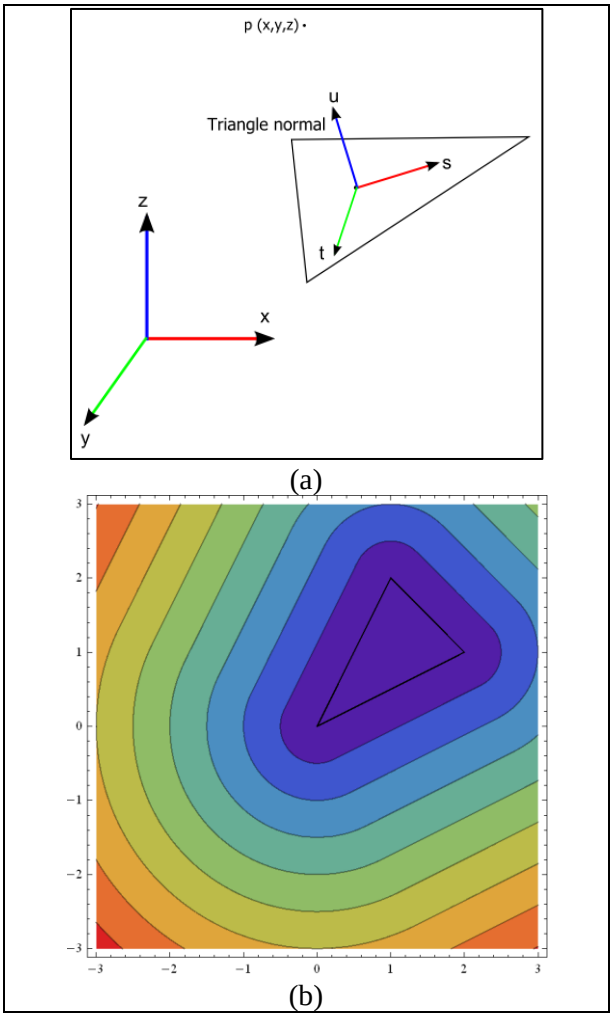


Figure 3. (a) Transformation from the world coordinate system (x,y,z) to the triangle's local coordinate system (s,t,u). (b) Visualization of the computed distance function using the local coordinates.

IV. RESULTS

To demonstrate the effect of the 3D enrichment technique on the shape functions of a meshless scheme, we have discretized a volume and computed the shape function for a meshless node whose support sphere intersects with a crack triangle. The visualization of the shape function shows that the 3D extension to the signed distance based enrichment works as expected (Figure 4). To show that the proposed enrichment technique is applicable to the soft tissue cutting context, we have incorporated it into a cut propagation setup, where a cut is represented as a piecewise linear segment. As the cut segment propagates, the enrichment technique ensured that the influence of the shape function of the node does not pass beyond the crack segment (Figure 5).

V. REFERENCES

[1] H. Gray, *Anatomy of the human body*: Lea & Febiger, 1918.

[2] T. W. Sederberg and S. R. Parry, "Free-form deformation of solid geometric models," *ACM Siggraph Computer Graphics*, vol. 20, pp. 151-160, 1986.

[3] K. Waters and D. Terzopoulos, "A physical model of facial tissue and muscle articulation," 1990, pp. 77-82.

[4] J. N. Reddy, *An Introduction to the Finite Element Method*, Third Edition ed.: McGraw-Hill, 2005.

[5] M. Bro-Nielsen, "Finite element modeling in surgery simulation," *Proceedings of the IEEE*, vol. 86, pp. 490-503, 1998.

[6] W. He-xiang, H. Ai-min, and L. Xue-mei, "Real-time cutting method for soft tissue based on TLED algorithm," 2010, pp. V3-393-V3-396.

[7] Y. Chen, J. D. Lee, and A. Eskandarian, *Meshless methods in solid mechanics*: Springer Verlag, 2006.

[8] X. H. Guo and H. Qin, "Meshless methods for physics-based modeling and simulation of deformable models,"

Science in China Series F: Information Sciences, vol. 52, pp. 401-417, 2009.

[9] B. Zhu, L. Gu, X. Peng, and Z. Zhou, "A point-based simulation framework for minimally invasive surgery," *Biomedical Simulation*, pp. 130-138, 2010.

[10] A. Hao and Z. Huang, "A physical based meshless method for soft tissue deforming," 2011, pp. 293-296.

[11] M. Müller, R. Keiser, A. Nealen, M. Pauly, M. Gross, and M. Alexa, "Point based animation of elastic, plastic and melting objects," 2004, pp. 141-151.

[12] V. P. Nguyen, T. Rabczuk, S. Bordas, and M. Duflot, "Meshless methods: a review and computer implementation aspects," *Mathematics and Computers in Simulation*, vol. 79, pp. 763-813, 2008.

[13] E. Barbieri, N. Petrinic, M. Meo, and V. L. Tagarielli, "A new weight-function enrichment in meshless methods for multiple cracks in linear elasticity," *International Journal for Numerical Methods in Engineering*, pp. n/a-n/a, 2011.

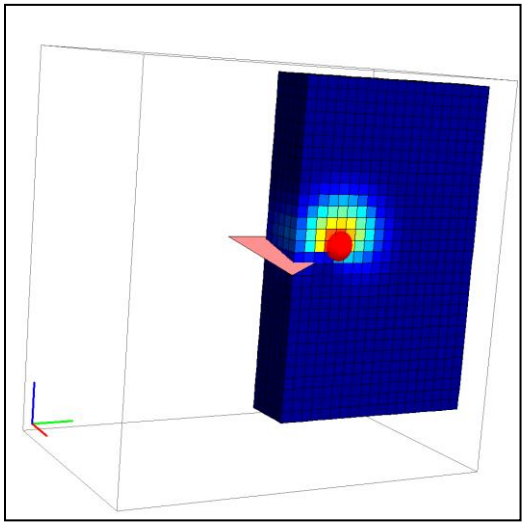


Figure 4. Visualization of the shape function for a meshless node (red sphere). Enrichment technique ensures that the meshless node does not influence the regions beyond the crack face.

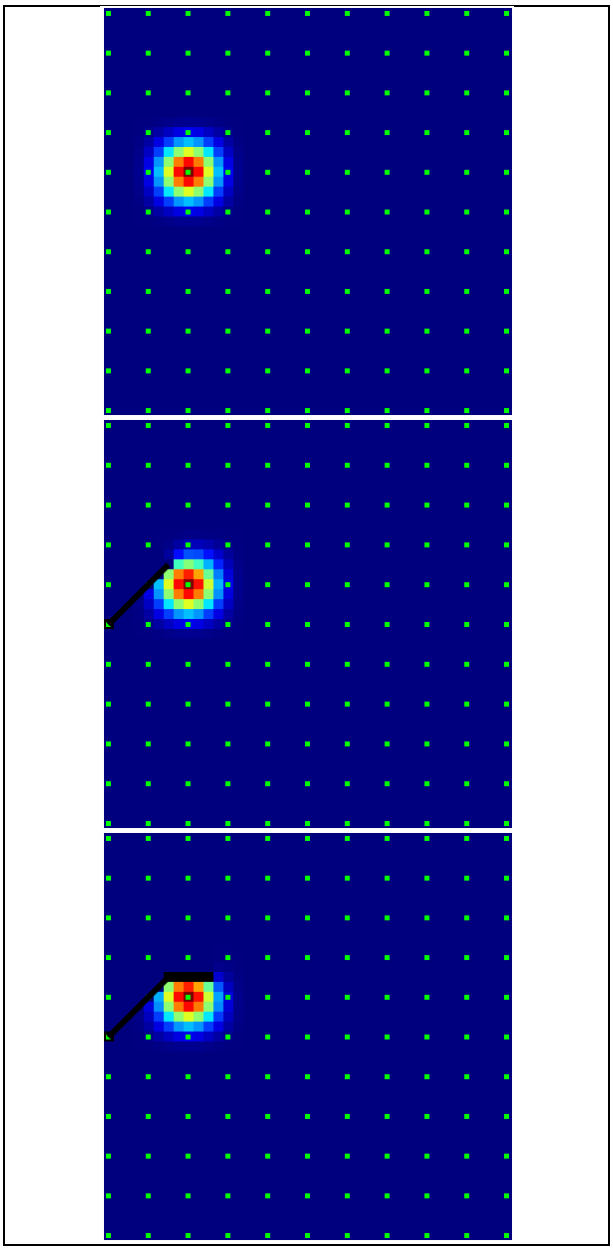


Figure 5. Propagation of a cut line in time and its effect on the shape function of a node with enrichment applied.

An Agent-based model of the dynamics of a tick-borne disease

Daniel Drake Tillinghast[1], Holly Gaff[1]

[1]Old Dominion University 5115 Hampton Boulevard Norfolk, VA 23529

An agent-based model was created to explore the transmission capability and resulting prevalence of a tick-borne disease in a current terrestrial system. The model was created in NetLogo and included a predator-prey model to explore the dynamics of the primary hosts and the tick-borne disease transmission. The simulation was run multiple times for a real-time equivalent of decades, using a variety of parameter sets to quantify the length of time the disease would remain in the environment. Results help identify the key criteria that substantially affect the predicted risk of the disease. Future renditions of the model will be able to provide warnings of seasons of severe contagion.

Index Terms— **Biological system modeling, Rocky Mountain Spotted Fever, Vector-borne disease, NetLogo**

I. Introduction

Rocky Mountain Spotted Fever (RMSF) was discovered in the Snake River Valley in Idaho, in 1896, by Major Marshall H. Wood [1]. Initially it was thought that the disease was contained in the Rocky Mountains of the United States, hence the name. However, in 1910, it was shown to have a presence in the country beyond its birthplace and its causative agent, the bacterium *Rickettsia ricketsii,* was isolated from ticks by Howard Taylor Ricketts [2]. RMSF has been a nationally notified disease in the United States since 1993.

Despite being one of the oldest known tick-borne diseases, RMSF has been overshadowed in the U.S. by a plethora of other diseases such as: Lyme disease, ehrlichiosis, anaplasmosis and babesiosis. Despite the publicity elicited by these other diseases, RMSF is still the most lethal tick-borne infection in the U.S with a 20% case fatality rate [3]. This is often resulting from delayed initiation of antimicrobial therapy because the characteristic rash may be absent early in the course of disease or because of overall masking of the disease underneath common influenza-like symptoms.

The disease has been studied for the past 20 years, but it is still unclear how to quantify the way RMSF behaves in the wild. Many studies show variable infection capability, duration, and fitness of the disease across different animal species [4], [5]. It is further complicated by the difficulty to track the primary carriers of this disease, the American dog tick (*Dermacentor variabilis*), Rocky Mountain wood tick (*Dermacentor andersoni*), and brown dog tick (*Rhipicephalus sanguineus)*. The creation of RMSF-SIM was inspired by TICK-SIM, an agent based model that was used as a tool to explore questions concerning Lyme disease. Following the same mindset, RMSF-SIM explores similar questions concerning RMSF. [6] The model attempts to encompass the life cycle of the American Dog Tick and track the progress of RMSF within a closed ecosystem, in an effort to track meaningful characteristics that the disease depends on in which to survive.

II. Methods

1) The Model

The model description follows the ODD (Overview, Design concepts and Details) protocol describing individual and agent-based models developed by Grimm and consists of seven elements [7]. The first three elements provide an overview, the fourth element explains general concepts underlying the model design, and the remaining three elements provide additional details.

2) Purpose

RMSF-SIM will help identify the key criteria that substantially affect the predicted risk of RMSF. Future renditions of the model will be able to provide warnings of seasons of severe contagion.

3) Purpose

In this agent-based model, three populations are considered: ticks, predator hosts and prey hosts. Each tick and each host along their respective trades are followed over successive time intervals. This model uses a generic version of each of these agents but the parameter values are based on those of the dog tick. The other two classes of agent are generic predator and prey animals. The dog tick differentiates between preferred hosts during its life cycles using the prey, an assumed smaller animal, for a blood meal in the first two life stages and the predator for the blood meal in third life stage.

The environment is set up as a 40X30 patches of forest to grassland ecosystem with non-wrapping boundaries. The system is divided between one-fourth forest and three-fourths grassland, with the forest serving as a home for all host agents and the grassland serving as the primary food source for the prey agents.

Each dog tick agent has a unique identification number, sex, location, life stage (egg, larva, nymph, adult), time in current stage, infection status (susceptible, infectious), current activity (resting, questing, feeding, or laying eggs for females) and hosts (list of all hosts used for blood meals). Similarly, host agents have a unique identification number, age, energy level, safety level, list of all ticks currently on the host, list of all ticks ever on the host and infectious status (susceptible, infectious).

4) Process Overview and Scheduling

The model follows the same steps of every parameter execution and resulting day to day cycle as shown in Figure 1. The mortality of ticks is based on the time of the year with the highest probabilities of death in the winter and summer. Unless otherwise stated within the table of parameters, all variables are kept static to ensure that the destabilization or success of the disease is due to inherent characteristics of itself and not the ecosystem being unable to supply due to its own inherent breakdown or unrealistic overabundance of a resource.

5) Process Overview and Scheduling

The forest portion of the system serves as a home location for all host agents and does not change with any experimental run. The grassland follows a variable change of food growth depending on the current time in the year and serves as the primary food source for all prey agents. The system is for all purposes closed from outside input once started.

The host agents in the simulation make up a predator-prey system that ticks interact within to find blood meals. The hosts have three main variables that govern their life cycles: age, energy, safety. Age is an overall time within system regardless of any variable. Energy is a currency that either host pays to interact within the system. All hosts must maintain a minimum energy level before a starvation factor effects the host's ability to continue within the system. Safety is a stress and survival characteristic of the particular host. While outside of the forest environment, a host agent builds up levels of stress until they have to navigate back to their forest home. If they cannot arrive within an acceptable time frame or they cannot achieve the fitness necessary to properly navigate, the agent is killed due to a "force of nature". All hosts are replenished in system following a breeding ability that is separately governed for either host. Everything within the system, except for ticks, emits a scent that allows each agent to navigate towards depending on its current goal whether it is a food source or their respective home.

The tick agents are the primary vector of the disease. The bacterium, *Rickettsia ricketsii*, is able to be vertically transmitted, from an infected female to the eggs she lays, for up to three generations. Every successive generation that is born carrying the disease has their respective egg laying capability cut to one-third of total capacity for three generations. Ticks can only sense hosts within their immediate vicinity and there is still only a small probability of actually successfully attaching to that host. The density dependence is only indirectly implemented by giving a maximum number of ticks per host. Tick agents have a diapause or hibernation ability if they were interrupted in the middle of a blood meal and were unable to secure another before the change of seasons allowing them to compete the following year if they are not killed otherwise.

6) Initialization

Each initialization is initiated in a divided 40X30 patch grid of ¼ forest and ¾ grassland with an initial 75 predators and 500 prey spread randomly across the grid. Additionally 15,250 ticks (10,000 eggs, 5000 larvae, 250 nymphs) are randomly spread across the grid. The simulations are assumed to begin on January 1, at which time adult stages will not be present. All other parameters are given in Table 1.

7) Simulation Experiments

Two experiments were conducted for this initial analysis. The first experiment ran the model 369 times (3 times for each variation parameter) The three variables that were changed were infection ability of the predator to infect the dog tick during the blood meal, the infection ability of the prey to infect the dog tick during the blood meal, as well as the initial starting infected population of ticks. Initial infected levels were chosen following current research on areas containing RMSF [8]. Counts of the number of infected predators, prey and dog ticks were recorded for each day until the extinction of the disease or the end of 20 years, whichever occurred first.

The second experiment varied the dog tick's ability to infect the predator during the blood meal, the dog tick's ability to infect the prey during the blood meal, as well as the initial infected population of dog ticks. Initial infected levels were chosen following current research on areas containing RMSF [8]. Accounts of the system were the same as the first experiment. All other parameters relevant to the disease were maintained at 50% to have a baseline of comparison.

The model was programmed using the programmable modeling environment, Netlogo. The software was authored by Uri Wilensky in 1999 and is freely available. (http://ccl.northwestern.edu/netlogo/).

III. Results

1) Experiment I

A selection of the minimum, median, and maximum variation of variables from the two experiments are shown in Figures 2, 3, and 4. In the first experiment, the disease was unable to maintain itself with any certainty within the system for the 20-year interval regardless of the infection ability of host to tick or quantity of the initial infected dog tick population. The disease either killed all dog ticks within the system or all carriers of the disease while the healthy dog ticks recovered within the system. By chance, a few of the runs were able to keep the disease in the system for the 20-year run. The disease lasted for an average 10-15 years. It can clearly be seen that the lower parameter values had fewer infected dog ticks while the highest parameter values had more infected dog ticks nearly the entire time.

2) Experiment II

The second experiment saw a slightly higher longer persistence time for the disease over the average but the disease was not able to last for a consistent duration within the system (Figures 5, 6, and 7). The lifecycle of the disease exhibited remarkably different characteristics than the previous experiment. While there was no common quantity of infection rate or infected population that allowed the disease to survive consistently, there was a similar result as in Experiment I showing that the majority of the ticks were infected with the highest parameter values.

IV. Conclusion

While the infection rates and initial infected populations impacted the proportion of the tick population infected while in the system, there was no discernible effect on the length of time RMSF remained in the system. The model supports current research on RMSF's inability to sustain itself within a closed environment. Additionally, the model can be used to explore and identify parameter values for specific areas that have RMSF cases. The model will be improved and streamlined to understand other characteristics of the disease within an ecosystem. The model will also be base lined so it can be manipulated to fit traits of other vector based diseases to correlate lab data to ecosystem lifestyles.

ACKNOWLEDGMENT

I would like acknowledge Darien Tyler Tillinghast, Heather Marie Raught, for their help in troubleshooting and resolving issues within the model as well as Darien's creation of the Visual Basic script to parse and categorize the hundreds of runs from the model. I would like to thank my mentors: Dr. Holly Gaff, Dr. David Gauthier, and Dr. Jin Wang for their leadership within the RUMS (Research for Undergraduate of Math and Science) scholarship as well as the Honor's College of Old Dominion University for their funding and support. Last I would like to thank Dr. Daniel Sonenshine and Robyn Nadolny for their biological expertise and input on various ticks and the RMSF disease.

REFERENCES

[1] V. Harden "Rocky Mountain Spotted Fever" Baltimore: The Johns Hopkins University Press, 1990

[2] H Ricketts "Some aspects of Rocky Mountain spotted fever as shown by recent investigations" Rev Infect Dis 199;13:1227–40. Reprinted from the Medical Record 1909;76:843–55.

[3] Center For Disease Control, "Nationally Notifiable Conditions (Annual lists of infectious conditions)" Internet: http://www.cdc.gov/osels/ph_surveillance/nndss/phs/infdis.htm, December 29, 2011 [March 6, 2012]

[4] S. Brown "Immunology of Acquired Resistance to Ticks" Parasitology Today Vol. 1 no. 6 1985.

[5] F. Bozeman, A. Shirai, J. Humphries, H. Fuller "Ecology of Rocky Mountain Spotted Fever II. Natural Infection of Wild Mammals and Birds in Virginia and Maryland" The American Society of Tropical Medicine and Hygiene vol 16 no. 1 1967

[6] H. Gaff, "Preliminary Analysis of an Agent-Based Model For a Tick-Borne Disease" Mathematical Biosciences and Engineering vol 8, no. 2, 2011

[7] V. Grimm, *et al* "A Standard Protocol for Describing Individual-based and Agent-based Models, Ecological Modeling vol. 198 pp. 115-26, 2006.

[8] Schriefer ME, Azad AF. "Ecology and natural history of Rickettsia rickettsia" in *Ecological dynamics of tick-borne zoonoses*. Sonenshine DE, Mather TN, Oxford: Oxford University Press;1994

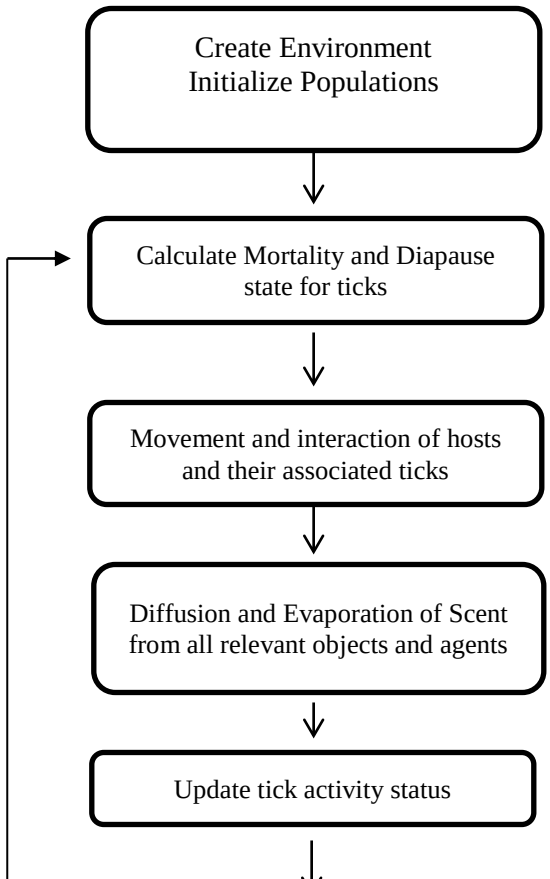


Fig 1. Flow Diagram for agent-based model

Table I. Parameter values. These parameters values are used for all runs of the model.

Parameter	Value
Initial Tick Population	15,250 (10,000 Eggs, 5000 Larvae, 250 Nymphs)
Initial Host Population	75 Predators, 500 Prey
Eggs per Female	3000 Healthy, 1800 Healthy at Generation 1 w/RMSF, 900 Healthy at Generation 2 w/RMSF, 0 at Generation 3 w/RMSF
Time from egg to questing larval stage	90 Days
Time from fed larval to questing nymphal stage	270 Days
Time from fed nymphal to questing adult stage	360 Days
Tick Mortality	0.1 in Jan, Feb, Mar, Jul, Oct, Nov, and Dec 0.01 in Apr May Jun Aug Sep
Maximum questing time	28 Days
Length of Blood Meal	3 Days
Prob. of successful attachment on predator	0.75
Prob. of successful attachment on prey	0.75
Maximum ticks per host	200
Prob. of predator infecting tick	0.1-0.9 in Experiment 1, 0.5 in Experiment 2
Prob. of prey infecting tick	0.1-0.9 in Experiment 1, 0.5 in Experiment2
Prob. of tick infecting predator	0.5 in Experiment 1, 0.1-0.9 in Experiment2
Initial infected population	6%, 17% 35%
Length of predator and prey infection	7 days

Minimum amount of energy before predator or prey begins to starve	20

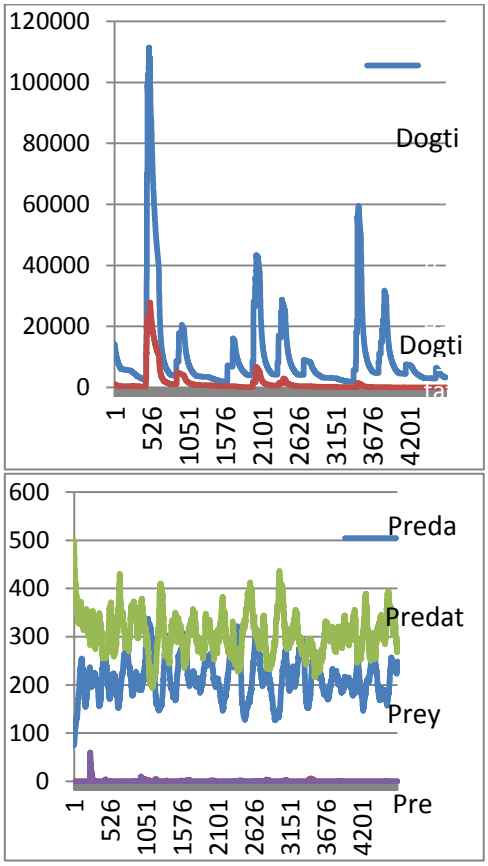


Figure 2: Minimum Infection Rate of Host Infect Tick
Prey Infect Tick: 10% Initial Infect Pop: 6%
Pred Infect Tick: 10% All Other Variables: 50%

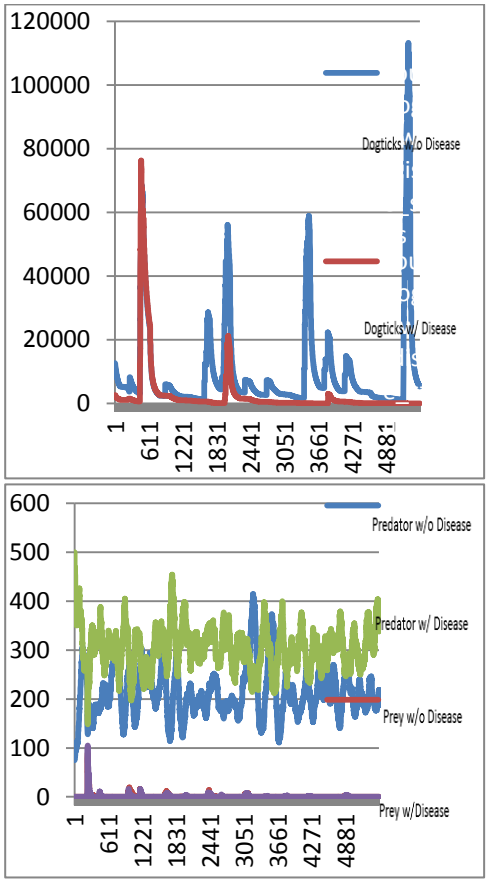


Figure 3: Median Infection Ability of Host Infect Tick
Prey Infect Tick: 50% Initial Infect Pop: 17%
Pred Infect Tick: 50% All Other Variables: 50%

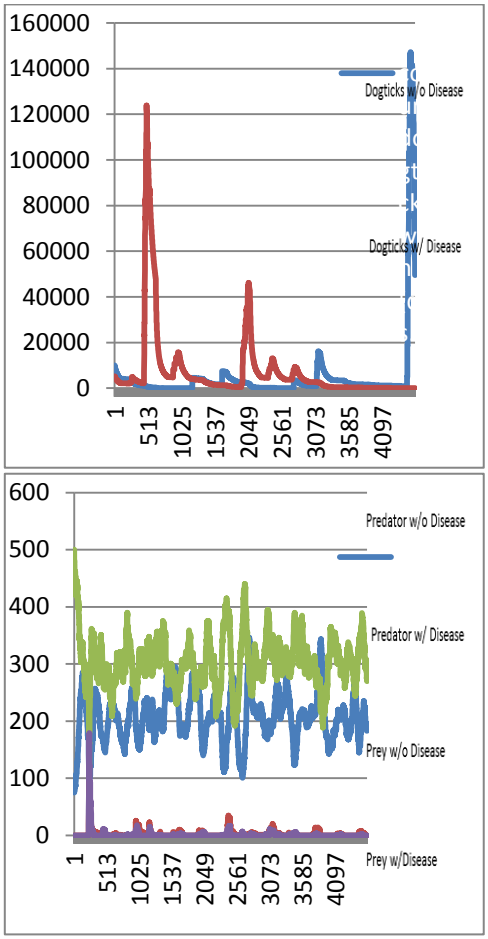


Figure 4: Maximum Infection Ability of Host Infect Tick
Prey Infect Tick: 90% Initial Infect Pop: 34%
Pred Infect Tick: 90% All Other Variables: 50%

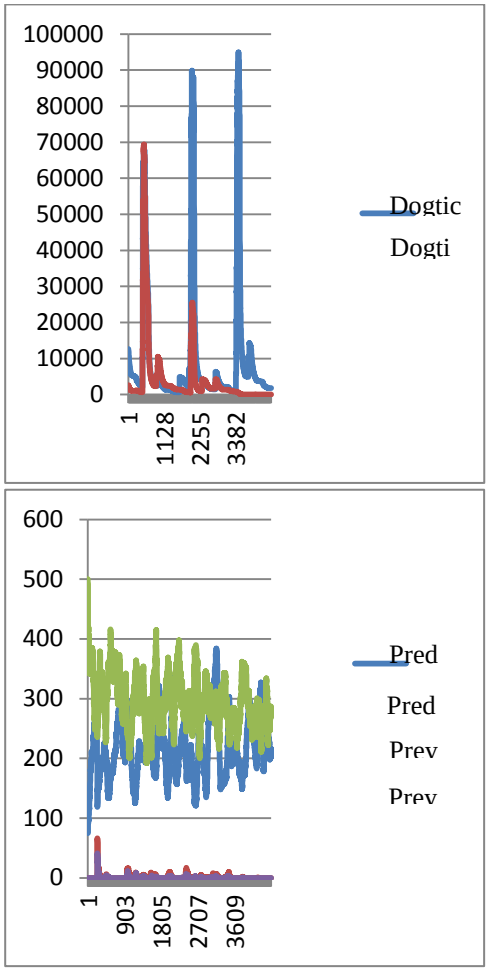


Figure 5: Minimum Infection Ability of Tick Infect Host
Tick Infect Prey: 10% Initial Infect Pop: 17%
Tick Infect Pred: 30% All Other Variables: 50%

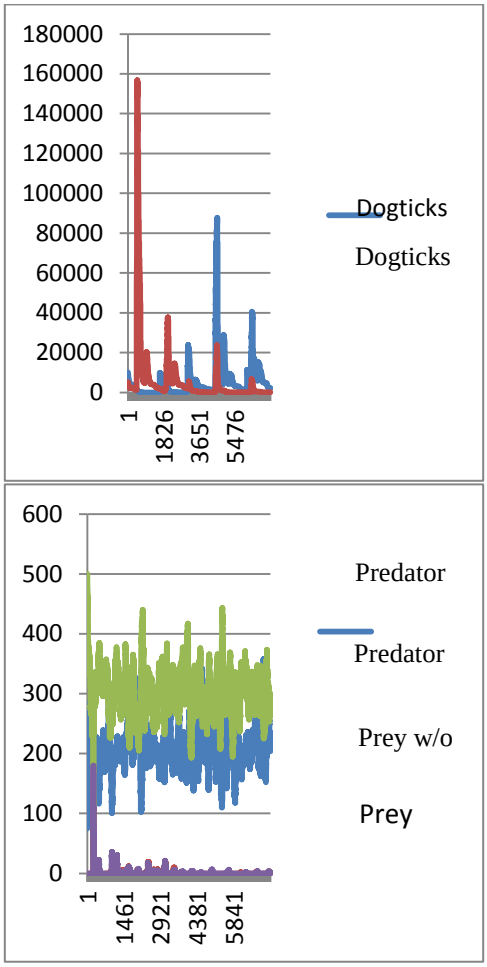


Figure 6: Median Infection Ability of Tick Infect Host
Tick Infect Prey: 50% Initial Infect Pop: 34%
Tick Infect Pred: 50% All Other Variables: 50%

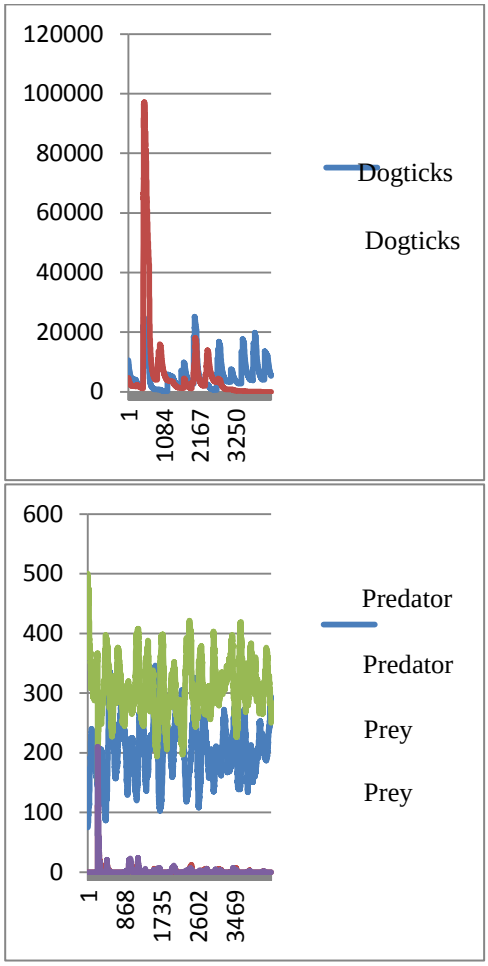


Figure 7: Maximum Infection Ability of Tick Infect Host
Tick Infect Prey: 90% Initial Infect Pop: 34%
Tick Infect Pred: 90% All Other Variables: 50%

A Continuous Simulation of Adoptive Immunotherapy Thresholds

Tyrell Gardner

***Abstract*---Adoptive immunotherapy is a new way to treat cancer which uses the patient's own immune system to eliminate the tumor. This is done by dosing the patient with large amounts of the cytokine interleukin 2. The immune system already produces interleukin 2 and with the added amount tumor elimination becomes possible. The problem is that interleukin 2 is toxic at high levels and patients have different thresholds for the treatment. If a patient's threshold is high enough, then cancer remission is possible. With Kirschner's differential equation model for immunotherapy a continuous simulation can be created to model different patient thresholds**.

***Index Terms*--- cytokine, effector, interleukin 2, continuous simulation**

I. INRODUCTION

Cancer involves the uncontrolled growth of the host's own cells [1]. Most treatments involve either chemotherapy or surgery; however, chemotherapy can cause mutations in non-tumor cells and surgery can lead to a vast array of obstacles for the host. Even with our great advances in technology, our two most common treatments for cancer are primitive at best. New avenues of treatment, that eliminate side effects and maximize treatment success, are always being pursued.

One of these new avenues is adoptive immunotherapy. Adoptive immunotherapy improves the host's immune system ex vivo and uses the boosted immune system to eradicate the tumor [2]. The immune system is usually boosted by introducing a large quantity of cytokine into the immune system. A cytokine is a type of hormone that helps regulate both natural and specific immunity. One of the most common cytokines used in adoptive immunotherapy is interleukin 2 (IL-2) [2]. The problem with using IL-2 to treat cancer is that IL-2 is very toxic at high levels [2]. When a patient being treated with IL-2 becomes too toxic the treatment must be ceased. Kirschner's current mathematical model used for adoptive immunotherapy does not take this cease in treatment into account. Instead, Kirschner's model, demonstrates the effect of one continuous stream of treatment [2].

Using Kirschner's original mathematical model as a base, a new mathematical model was created by Jackson et. al. which demonstrates a cease in treatment based on the patient's threshold for interleukin 2 [1]. The work presented in this paper develops a continuous simulation based on the new mathematical model created by Jackson to simulate the effects different IL-2 thresholds have on tumor eradication. By simulating adoptive immunotherapy with a cease in treatment, physicians will have better knowledge of how to treat specific patients based on their IL-2 threshold.

II. MODEL

The model employed in the simulation was taken directly from Kirschner's model (above it seems you are using Jackson's model, be clear which) describing the growth interactions between effector cells, tumor cells, and cytokines [2]. The effector cells modeled in this simulation were T-lymphocytes. T-lymphocytes are white blood cells that are crucial to cell mediated immunity. The tumor cells can be defined as any rapidly reproducing cell whose reproduction is uncontrolled. Cytokines are hormones that help control the effector cells, and the one that I focused on was interleukin 2. The change in these three populations was modeled using the three ordinary differential equations shown below.

$$\frac{dx}{dt} = cy - \mu_2 x + \frac{p_1 xz}{g_1 + z} + s_1 \quad (1)$$

$$\frac{dy}{dt} = r_2 y(1 - by) - \frac{axy}{g_2 + y} \quad (2)$$

$$\frac{dz}{dt} = \frac{p_2 xy}{g_3 + y} - \mu_3 z + s_2 \quad (3)$$

The first equation models the rate of change of the effector cell population. In this equation the first term is used to model the antigenicity of the tumor. Antigenicity can be defined as how different the tumor cells are from the host cells and is shown as the constant $\boldsymbol{c}$ in equation (1). A high antigenicity would indicate an increased number of effector cells. The second term models the lifespan of an effector cell with the average lifespan being $1/\boldsymbol{\mu_2}$. In the third term, the proliferation rate of the effector cells in

response to the cytokines is modeled. The constant $\mathbf{p_1}$ represents the effects cytokines produced by a specific T-lymphocyte have on that T-lymphocyte and other T-lymphocytes, and the constant $\mathbf{g_1}$ represents an immune saturation constant [2]. The final term $\mathbf{s_1}$ is an additional treatment term that allows for effector cells to be directly added into the population from an outside source.

The second equation models the rate of change of the tumor cell population. In the first term the growth of the tumor is modeled using a limiting growth approach with $\mathbf{r_2}$ representing the rate and $\boldsymbol{b}$ representing the growth limit. The second term takes into account the loss of tumor cells to effector cells with the constant **a** representing the strength of the immune response and $\mathbf{g_2}$ representing the saturation of the immune response.

The third equation models the rate of change for the cytokine population. The first term is the source term for the cytokines. This term depends on the interactions between effector cells and tumor cells at a constant rate $\mathbf{p_2}$ and takes into account the saturation of the immune system with constant $\mathbf{g_3}$. The second term is the loss term with an average loss of cytokines being $1/\boldsymbol{\mu_3}$ [2]. The third term $\mathbf{s_2}$ is a cytokine treatment term that represents an external introduction of cytokines into the system.

TABLE I
PARAMETER ESTIMATION FOR MODEL

Symbol	Value
c	$0 \leq c \leq 0.05$
μ_2	0.03
p_1	0.1245
g_1	2×10^7
g_2	1×10^5
r_2	0.18
b	1×10^{-9}
a	1
μ_3	10
p_2	5
g_3	1×10^3

The $\mathbf{s_2}$ term was the most important term in the simulation. According to Jackson, "the IL-2 treatment term should be changed to be dependent on both the population of effector cells and on time" [2]. This would indicate that the treatment term is similar to a switch which can be turned on and off based on both time and a patient's immune system threshold. For the first, second and third equation parameters were obtained from the models in references [3], [5], and [4] respectively. For Table 1 all units are in days^{-1} except $\mathbf{g_1}$, $\mathbf{g_2}$, $\mathbf{g_3}$, and **b** which are in microliters.

III. SIMULATION

Ptolemy II version 8.0.1 was employed to build a simulation based on the three differential equations previously discussed. Since each equation is only a first order differential equation, only three integrators are required for the entire simulation. Each integrator used a 4th order Runge Kutta numerical method for solving and was connected to a series of blocks. The blocks were connected in such a way that they represented the different terms previously discussed. These terms were then fed back into the integrator blocks. The information feed into the integrator represented the rate of change for each population and the information coming out of the integrator was the actual population size. Figure 1 shows the simulation diagrams for the three differential equations. The connector representing each term is clearly labeled in the figure and all parameter names were pulled directly from the model.

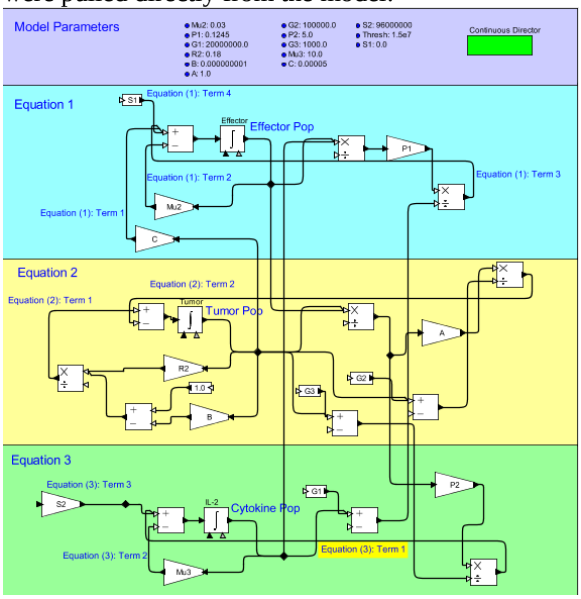


Fig. 1. Simulation Diagram of Adoptive Immunotherapy Treatment using IL-2 and Patient Threshold

In order to gather data, a graph output and a data output are attached directly after each integrator as seen in Figure 2. The graphing output was used to plot the population size with respect to time. This enables the monitoring of different population sizes as the simulation runs. The data output allowed the easy transfer of results to an Excel spreadsheet for analysis purposes.

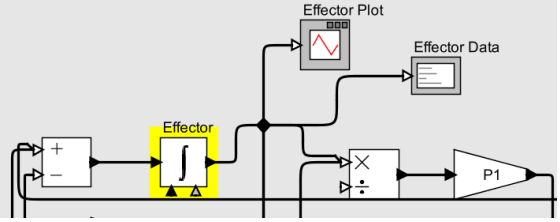


Fig. 2. Graphical and Data Output used in Ptolemy II

The most difficult piece of the simulation implementation was setting up the threshold for the IL-2 treatment. The tumor growth steady state of 3000 days was borrowed from Jackson's model [1]. According to Jackson, after 3000 days the tumor should be at maximum capacity allowing for a worst case scenario analysis of the treatment. After 3000 days the IL-2 treatment is supposed to start and continue until the immune threshold of the patient is reached. Figure 3 provides the block diagram of the Ptolemy implementation. As long as the simulation time was greater than 3000 days and the IL-2 population size was smaller than the patient threshold then the treatment continued. By implementing the switch it was very easy to change the patient threshold.

$$\frac{dz}{dt} = \frac{p_2 xy}{g_3 + y} - \mu_3 z + s_2 \qquad (3)$$

s_2=0 t < 3000
s_2=9.6x10^7 t ≥ 3000, z < threshold
s_2=0 t ≥ 3000, z ≥ threshold

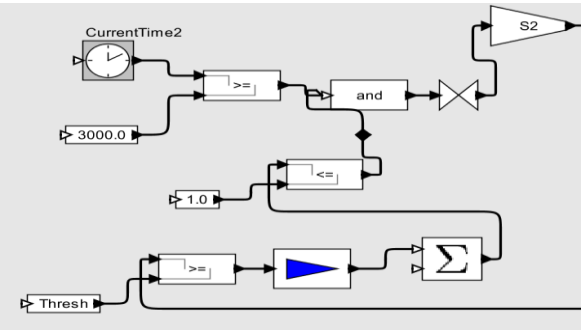


Fig. 3. Switch Implementation in Ptolemy for turning on treatment after 3000 days and turning off treatment once the threshold is reachedIL-2 Treatment

IV. RESULTS AND ANALYSIS

When the IL-2 treatment is started, the effector cells increase in number due to the surge in IL-2. This increase in effector cells can be seen in Figures 6, 7, and 8, and it is just the thing the body needs in order to eradicate the tumor. When the patient's threshold was set to 0.8 x 10^7 the increase in IL-2 was not enough to show any change in either the T-Lymphocyte growth or the tumor. Figure 6 shows a slight increase in IL-2 with no change in the effector

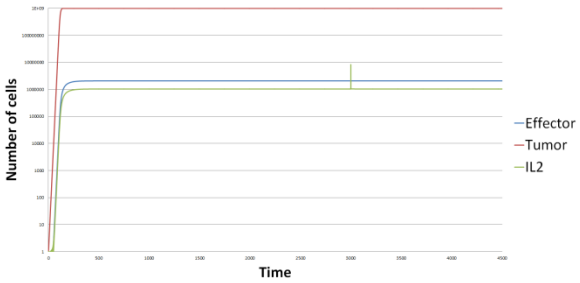


Fig. 6. Graph done in Excel displaying simulation results with an IL-2 Threshold of 0.8 x 10^7

cell population or the tumor cell population. The same was true when the threshold was increased to 1.0 x 10^7. Looking at Figure 7 shows that even with an increase of 0.2 x 10^7 there was no change in the

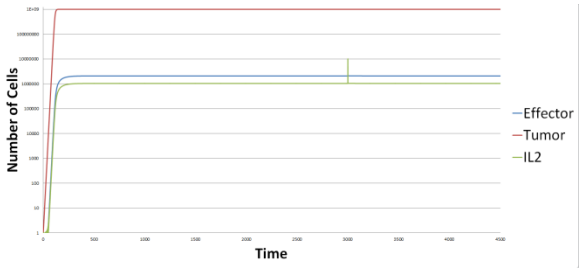


Fig. 7. Graph done in Excel displaying simulation results with an IL-2 Threshold of 1.0 x 10^7

effector cell population or tumor cell population. It was not until the threshold is raised to 1.5 x 10^7 that a change in the effector cell population and the tumor cell population was seen. Figure 8 shows the results when the threshold was raised to 1.5 x 10^7. It is quite easy to see the surge in

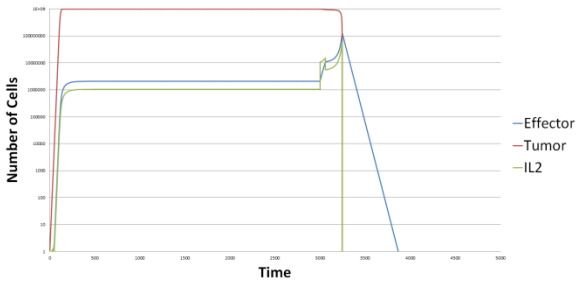


Fig. 8. Graph done in Excel displaying simulation results with an IL-2 Threshold of 1.5 x 10^7

IL-2 when the treatment is started at day 3000. This surge caused a subsequent surge in the T-Lymphocyte cells. As the IL-2 growth reached the threshold the treatment was ceased; however, the T-

Lymphocyte cells continued to grow and the tumor was eliminated. As soon as the tumor cells started to drop, the remaining IL-2 and T-Lymphocyte cells started to drop as well.

V. CONCLUSION

Continuous simulation was an excellent choice for modeling the threshold behavior of adoptive immunotherapy. The results for each threshold were easily reproduced in a matter of minutes. By combining the model created by Kirschner with continuous simulation a useful tool is created for physicians. A physician dealing with cancer patients for whom adoptive immunotherapy is an option will have a better idea of the success of the treatment before the treatment is even administered. Although it is not modeled here, this simulation could also provide the most likely time for a relapse in each patient. With these times the physician could re-administer treatment before the relapse.

Unfortunately there has not been any clinical data collected to compare with the results from this simulation. A series of clinical trials could be performed on groups of patients with different treatment thresholds. Another set of trials could be performed on the set of patients who go into remission. The time frame for relapse in these patients could then be recorded and compared with the time for relapse that could be predicted by our model. These are just a few of the ideas that lay ahead for improving this model of adoptive immunotherapy.

REFERENCES

[1] Jackson T, Usman A, Cunningham C. Application of the Mathematical Model of Tumor-Immune Interactions for IL-2 Adoptive Immunotherapy to Studies on Patients with Metastatic Melanoma or Renal Cell Cancer. *Biological Mathematics: Department of Mathematics, University of Michigan.* 1:1-10, *2004.*

[2] Kirschner D. Panetta JC. Modeling immunotherapy of the tumor-immune interaction. [Journal Article] *Journal of Mathematical Biology. 37(3):235-52, 1998.*

[3] R. J. DeBoer, Pauline Hogeweg, Hub F. J. Dullens, Roel A. DeWeger and Willem DenOtter. Macrophage T Lymphocyte interactions in the anti-tumor immune response: A mathematical model. *The Journal of Immunology, 134(4): 2748-2758, 1985.*

[4] S. A. Rosenberg and M. T. Lotze. Cancer immunotherapy using interleukin-2 and interleukin-2-activated lymphocytes. *Annual Review of Immunology, 4: 681-709, 1986*

[5] V. A. Kuznetsov, I. A. Makalkin, M. A. Taylor and A. S. Perelson. Nonlinear dynamics of immunogenic tumors: Parameter estimation and global bifurcation analysis. *Bull. Math. Biol., 56(2): 295-321, 1994*

Finite Element Analysis of Deformation of Bars Used to Correct Pectus Excavatum in Children

Rohit N. Nikam, Stephen B. Knisley, Frederic D. McKenzie, Robert E. Kelly

I. INTRODUCTION

Pectus Excavatum (PE) is the most common chest wall deformity. In PE patients, middle lower portion of the sternum is depressed producing concave appearance of the anterior part of the chest wall. Correction of the Pectus Excavatum problem is accomplished by using minimally invasive Nuss technique of chest remodeling developed by Dr. Donald Nuss in 1987. In this procedure, specially made Nuss bar is implanted in order to lift the sternum to its normal position [4][5].

This minimally invasive technique has been modified several times to improve safety and effectiveness of Nuss bar placement and removal. There are still few factors in the current Nuss treatment that need to be improved with the help of biomechanical analysis. One of the significant aspects of the Nuss procedure is the optimum bar shape prior to implantation. In PE, sternum force is an important factor for chest remodeling. This downward force between bar and sternum creates bending moment and can also change the end-to-end distance and shape of the bar. If ends of the Bar protrude too far, the Pectus Bar must be removed and further bent to produce the correct shape upon re implantation, i.e., hyper bending. The goal of this study is to conduct a mechanical analysis of the bar bending and the forces to predict an appropriate amount of hyper bending. If successful, this would eliminate the need to remove and implant the bar, thus shortening surgery time.

II. PROBLEM / MOTIVATION

Preoperative preparation for the Nuss Procedure involves measurement of chest wall from mid axial line to mid axial line. Then correct length of pectus bar is selected and formed to proper shape using the pectus blender in order to achieve maximum bar stability [4]. Sometimes, it is necessary to slightly exaggerate the curvature of the bar. Experience at Children's Hospital of the King's Daughters (CHKD) has also indicated that ideal shape of the bar, i.e., amount of bending, is not always correctly indicated by simply the desired shape of the patient's chest. The bar which is performed to match the desired shape of the chest will slightly straighten after implantation due to body forces. This can cause the ends of the bar to protrude out of the chest, thus requiring the surgeon to remove the bar and re-bend it to have additional bending (hyper bending). The goal of hyper bending is to produce the optimal bar shape after the bar is subjected to unbending by the sternum force. The amount of hyper bending needed is currently determined by the experience of the surgeon at the time of implantation. A Finite Element Analysis (FEA) of a Pectus Bar under different load conditions was carried out in an attempt to quantify the hyper bending needed.

A. Overall Objective:

To create a model of the bending and the body forces for a realistic bar in a patient being treated for the pectus excavatum and analyze the forces and bending in different realistic situations.

III. MATERIALS AND METHODS

A. Related Work and Data Analysis:

In the Nuss Procedure, Pectus bar should be in proper size and shape to sustain the sternum load. In order to conduct the realistic simulation of a pectus bar inside the patient's chest, different thoracic cage parameters were collected from the study conducted by A. R. Comeau [2]. A total of 25 pediatric rib cage cases for 1 to 18 year old patients were collected [2]. The parameters measured were lateral distance, anterior-posterior distance and sternum width. This data was analyzed by KaleidaGraph's least square polynomial curve fits to predict intermediate age parameters. Fonkalsrud and Reemtsen studied PE patients with different age groups to find out force required to elevate the sternum [3]. Data collected from their study was implemented in the FEA of a Pectus Bar for distinct age groups.

B. Estimation of Realistic Forces and their Locations:

From the experience at CHKD, computed tomography images, post-operative X-rays and information available in the literature, we developed our FEA model with the assumption that the Pectus Bar is loaded at center with a downward sternum force [1]. Sternum force with total load ranging from 44.48 N to 155.69 N was applied for different age groups [3][6]. Total force (F_{st}) was uniformly distributed over the sternum width. In order to maintain the equilibrium of the Pectus Bar, two supports (F_s) were provided at equal distance from the center as shown in Figure1.

Dr. Stephen B. Knisley,
Department of Mechanical and Aerospace Engineering,
Old Dominion University, 253, Kaufman Hall, Norfolk,
VA, 23529, USA
e-mail: sknisley@odu.edu

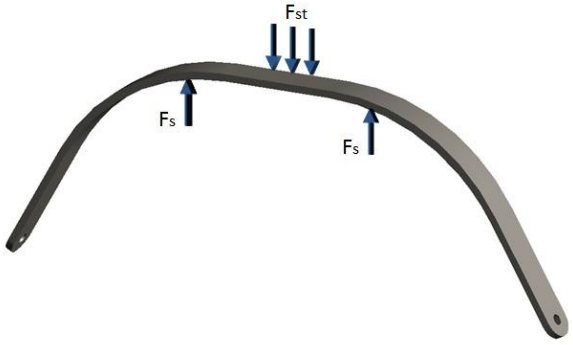


Figure1. Pectus Bar 3-D model with downward sternum force (F_{st}) and upward support forces (F_s).

C. Finite Element Analysis (FEA) of the Pectus Bar:

The dimensions of a Pectus Bar produced by Biomet Microfixation were measured and patient specific bar size data was collected from the Biomet Microfixation Pectus Excavatum Brochure. Considering all possible parameters, 3-D models of the Pectus Bar were created using Pro/Engineer Wildfire 5.0 software for 6,8,10,12,14 and 16 year old patients [2][7].

Further, this model was imported into MSC/PATRAN 2010 software for pre processing. High chromium stainless steel (316L Grade) was selected for the bar material properties with elastic modulus 193 GPa and Poisson's ratio 0.3. These values were collected from www.azom.com, article ID = 2382. Linear static analysis was selected for this study [8]. After completing the analysis setup, the PATRAN file was sent to NASTRAN 2010 for Finite Element Analysis (FEA). During post processing in PATRAN, displacements of the free end of the Pectus Bar in the horizontal direction were recorded.

D. Validation:

Finite Element Analysis results can be evaluated by conducting a bench test. In order to fulfill this requirement, an experimental test rig has been set up in the laboratory. In this study, bending and body forces acting on the Pectus Bar will be analyzed. These results will be compared with the FEA observations.

IV. RESULTS

FEA was conducted for 6 – 16 year old patients with the different sternum load conditions described above. The results of the horizontal displacement of free end of the Pectus Bar at different loading conditions are shown in Table 1.

Table 1
Pectus Bar free end displacements in horizontal direction

Sternum Force (N)	Bar displacement in horizontal direction (mm) related to patient age (years)					
	6 years	8 years	10 years	12 years	14 years	16 years
44.48	0.017	0.014	0.016	0.021	0.026	0.031
66.72	0.016	0.021	0.023	0.032	0.038	0.047
88.96	0.021	0.028	0.031	0.042	0.051	0.062
111.20	0.027	0.035	0.039	0.053	0.064	0.077
133.45	0.032	0.042	0.047	0.063	0.077	0.093
155.69	0.037	0.048	0.055	0.074	0.089	0.108

V. CONCLUSION/FINDINGS

According to Finite Element Analysis (FEA) results (Table1), Pectus Bar displacement in the horizontal plane is too small to account for the hyper bending required by surgeons. The FEA model could give smaller displacements compared to the displacements of surgical bars if the value for modulus in the model is higher than that in surgical bars. Material hardness can be increased by cold working, thermal factors or surface treatment, and is not determined solely by the material composition. Thus, the material properties of stainless steel based solely on composition may not represent the properties of bars used in surgery. Moreover, positions of the supporting points (F_s) should be validated to improve the accuracy of this study. Bench tests will be conducted in order to validate the Finite Element Analysis results including the supporting positions and the assumptions concerning the material properties of the stainless steel.

REFERENCES

[1] P. Y. Chang, Z.-Y. Hsu, D.-P. Chen, J.-Y. Lai, and C.-J. Wang, "Preliminary analysis of the forces on the thoracic cage of patients with pectus excavatum after the Nuss procedure," Clinical Biomechanics, vol. 23, pp. 881-885, 2008.

[2] A. R. Comeau, "Age-related Changes in geometric characteristics of the pediatric thoracic cage and comparison of thorax shape with a pediatric CPR manikin," Master of Science, Biomedical Engineering, Drexel University, 2010.

[3] E. W. Fonkalsrud and B. Reemtsen, "Force required to elevate the sternum of pectus excavatum patients," Journal of the American College of Surgeons, vol. 195, pp. 575-577, 2002.

[4] D. Nuss, R. E. Kelly Jr, D. P. Croitoru, and M. E. Katz, "A 10-year review of a minimally invasive technique for the correction of pectus excavatum," *Journal of Pediatric Surgery*, vol. 33, pp. 545-552, 1998.

[5] S. Uemura, Y. Nakagawa, A. Yoshida, and Y. Choda, "Experience in 100 cases with the Nuss procedure using a technique for stabilization of the pectus bar," Pediatric Surgery International, vol. 19, pp. 186-189, 2003.

[6] P. G. Weber, H. P. Huemmer, and B. Reingruber, "Forces to be overcome in correction of pectus excavatum," The Journal of Thoracic and Cardiovascular Surgery, vol. 132, pp. 1369-1373, 2006.

[7] Y. Wei, D. Sun, P. Liu, and Y. Gao, "Pectus Excavatum Nuss Orthopedic finite element simulation," in Biomedical Engineering and Informatics (BMEI), 2010 3rd International Conference on, 2010, pp. 1236-1239.

[8] Y.-b. Wei, Y.-k. Shi, H. Wang, and Y. Gao, "Simulation of Nuss Orthopedic for Pectus Excavatum," in Biomedical Engineering and Informatics, 2009. BMEI '09. 2nd International Conference on, 2009, pp. 1-4.

Towards a digital cranial nerve model for neurosurgery simulation

Sharmin Sultana, Michel A. Audette

Extended Abstract— Twelve cranial nerves emerge from the base of the brain, most from the brainstem, and innervate various parts of the head and neck. Cranial nerves are responsible for our senses of vision, hearing, smell, and taste, for our balance, and for motor functions related to facial expression, motor functions such as biting, chewing, and swallowing, and speech. There are 12 cranial nerves, on each of the left and right sides, connected to the central nervous system (CNS). These nerves can be divided into purely sensory (I, II, VIII), purely motor (II, IV, VI, XI, XII) and mixed (V, VII, IX, X).

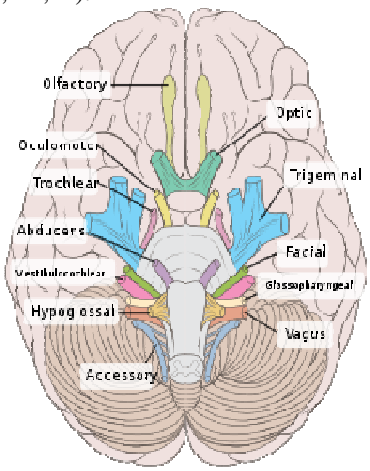


Figure 1 Cranial Nerves of Human Brain. Reproduced with permission from Bartleby.com

Cranial nerves are often at risk in neurosurgical procedures. A single surgical maneuver deleterious to a cranial nerve can have serious life-long consequences ranging from loss of eyesight or hearing to facial paralysis. Moreover, cranial nerves are often difficult to identify in routine clinical MR data, due to limitations in the imaging resolution as well as partial volume effects.

The development of a ***digital cranial nerve model*** that could be registered to a patient's anatomy has some applications in relation to these critical neurologic structures:

1. Preventing lesions to cranial nerves in ***neurosurgical simulation***, in a manner *consistent with neurosurgical approaches*. Such a model would appropriately penalize a deleterious gesture during simulation-based training, as well as lead to the development of new surgical techniques or to the choice of an alternative angle of approach during surgery,
based upon precise intra-operative localization of the cranial nerves.

2. Preventing lesions to cranial nerves by registering the cranial nerve anatomy to the patient's brain image, in ***model-guided neurosurgical avoidance of cranial nerves***. The model would ensure the optimal pre-operative planning around, and intra-operative avoidance of these critical tissues, particularly in skull base. Also, as high-res MR data eventually becomes a common component of surgical workflow, the identification of cranial nerves can build on this model registration, rather than an interactive segmentation.

3. ***Model-guided neurosurgical targeting of cranial nerves.*** The model would then serve to pre- and intra-operatively localize a specific nerve as a surgical target, for procedures such as nerve stimulation in epilepsy treatment or the decompression of a cranial nerve of a patient presenting neuralgia symptoms.

In our research we will construct ***a digital model of cranial nerves*** from high-res MR-CISS data, in conjunction with semi-automatic centerline extraction and expert neuro-radiological identification of points along each path. This model will encode the path and radius along the path of each nerve. The model will also resolve spatial overlaps between several nerves, or between nerves and blood vessels, or nerves and the cranium.

The current method for avoiding injuries to cranial nerves is intra-operative electrophysiological monitoring [4]. For cranial motor nerves, stimulodetection techniques are used, including electrical stimulation of nerve trunks and electromyographic recording of evoked motor responses [5]. These techniques can be used for monitoring the trigeminal, facial, glossopharyngeal, pneumogastric, spinal accessory, and hypoglossal nerve, in particular during surgical removal of tumors of the cerebellopontine angle or skull base. At the beginning of an operation, electrical stimulation is only used to identify the nerve structures. As removal of the tumor progresses, the goal is to verify that a surgical injury to the nerve is avoided by looking for the absence of any change regarding amplitude, morphology, and latency of motor responses. Our Cranial Nerve Model offers an intuitive and spatially precise depiction of the nerves to plan their localization pre-operatively and better anticipate where to use EP monitoring intra-operatively. From a surgical planning standpoint, our nerve model potentiates EP.

Much of the medical image analysis literature that deals with tubular structures has been applied to the identification of blood vessels. There are three basic approaches to vessel

extraction: centerline-based modeling, spatial filtering, and voxel labeling.

Centerline extraction is the basis of our planned approach to nerve modeling. It starts from an initial point near a nerve and then performs a multi-scale traversal of the nerve's centerline. Along the centerline, it subsequently estimates the radius of the structure [1].

we are currently developing synthetic angiographic data, in conjunction with quantitative validation, and the first step towards this development is the ability to give each voxel in a synthetic image an intensity value that varies based on whether it is contained in the tube model or not, and that ultimately we want to synthesize several tubes near each other. The points in image of figure 2 can ascertain containment in the tube. Red points pass the containment test within the tube while the green points fail the containment test.

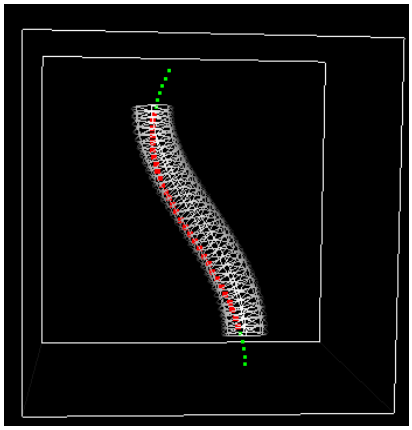


Figure 2 Implementation of a tube with seed points using VTK

To represent the outer surface of the nerve we will use deformable simplex surface which is a good approach to segment three dimensional objects based on the geometry of simplex mesh [2] [3]. We have started modeling cranial nerve with the help of Dr. Benjamin Gilles simplex Mesh implementation. Tube Tool kit (TubeTK) and/or Vascular Modeling Toolkit (VMTK) will be used to model the medial surface.

Flow Diagram :

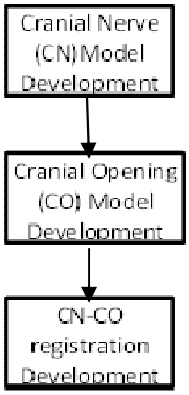


Figure 3 Flow Diagram of the project

Estimated Timeline:

Model	Year1		Year 2		Year 3		Year 4	
	Mo 1-6	Mo 7-12	Mo 1-6	Mo 7-12	Mo 1-6	Mo 7-12	Mo 1-6	Mo 7-12
CN	→	→	→	→				
CO			→	→	→	→	→	
CN-CO registration						→	→	→

References

[1] Stephen R. Aylward, "Initialization, Noise, Singularities, and Scale in Height Ridge Traversal for Tubular Object Centerline Extraction" IEEE Transaction on Medical Imaging, VOL. 21, NO. 2, Feb 2002.

[2] Benjamin Gilles, Nadia Magnenat-Thalmann , "Musculoskeletal MRI segmentation using multi-resolution simplex meshes with medial representations", Medical Image Analysis 14 (2010) 291–302.

[3] Herve Delingette, "General Object Reconstruction based on Simplex Meshes" , International Journal of Computer vision, 32, 111-142

[4] [4] Youssef AS, Downes AE. "Intraoperative neurophysiological monitoring in vestibular schwannoma surgery: advances and clinical implications." Neurosurg Focus. (2009 Oct): 27(4):E9.

[5] [5] Lefaucheur JP, Neves DO, Vial C. "Electrophysiological monitoring of cranial motor nerves (V, VII, IX, X, XI, (2009 Apr): 55(2):136-41.

Towards a framework for poly-affine initialized elastic registration of MR spine images for detection of abnormalities

Rabia Haq, Michel Audette
Old dominion University, Norfolk, VA 23508, USA

Recent advancements in computer science and modeling, simulation and visualization engineering have made it possible to visualize human anatomy. Biomedical image processing techniques have been developed to facilitate medical research, patient diagnosis and treatment planning for consolidation of patient image data. Biomedical images acquired through various modes, such as MRI or CT, and at various time periods need to be spatially aligned for comparison, which is known as image registration.

Image registration can be utilized for registration of MR images of the spine for isolation of tumors in the spine region. Tumors of the spine consist of tumors existing in the primary central nervous system as well as secondary lesions, metastases, of breast, lung, prostate and kidney tumors. Up to 10% of all cancer patients develop secondary spinal lesions (Jacobs et al., 2001), with an increased chance of metastases with increase in survival rate. Figure 1 shows an example of spinal metastasic tumor. These tumors of the spine need to be robustly identified and isolated to assist in surgery planning.

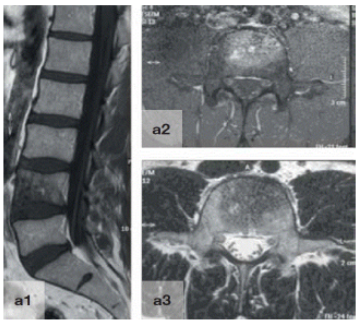


Fig. 1: (a1) MRI sagittal and (a2, a3) axial images of spinal metastasic tumor.

The proposed framework attempts to perform poly-affine initialized elastic registration of two MR images of the spine. Figure 2 depicts this approach for bringing two spine images into spatial alignment.

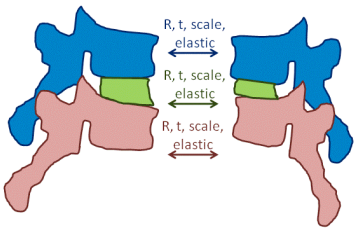


Fig. 2: poly-affine initialized multi-surface registration of the spine.

Registration of the spine is particularly challenging due to the arbitrary alignment of vertebrae in a particular image. Each vertebra in the spine has a topological consistent shape across individuals, which can be exploited to identify and isolate components, one for each vertebra, within the MR image. Multi-surface simplex meshes can be augmented to extract component surface information, as depicted in figure 3 taken from Gilles et al. (2010). These extracted components can be locally affinely registered to perform vertebra-to-vertebra registration between two images similar to polyaffine initialized elastic registration by Audette et al. (2008), and interpolate elastic deformations between these registered components to finalize the process.

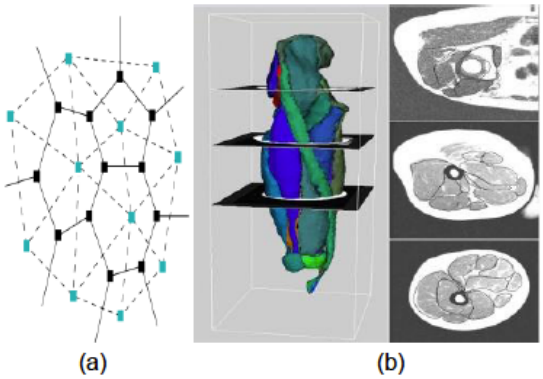


Fig. 3: Simplex model: (a) topology (in black) with three neighboring vertices and duality in triangulated mesh (in blue) (b) multi-surface simplex model of the thigh

This proposed MR spine image registration framework could be employed for development of an MRI-based 3D digital spine atlas to assist with automatic spine tumor segmentation analogous to automatic tumor segmentation of MR images of the brain (Prastawa et al., 2004).

REFERENCES:

Audette, M., Brooks, R., Funnell, R., Strauss, G., Arbel, T., 2008. Piecewise affine initialized spline-based patient-specific registration of a high-resolution ear model for surgical guidance. MICCAI Workshop on Image Guidance and Computer Assistance for Soft-Tissue Interventions.

Gilles, B., Magnenat-Thalmann, N., 2010. Musculoskeletal MRI segmentation using multi-resolution simplex meshes with medial representations. Medical Image Analysis 14(3) 291-302.

Jacobs, W.B., Perrin, R.G., 2001. Evaluation and Treatment of Spinal Metastases: An Overview. Neurosurgery Focus 11(6):e10.

Prastawa, M., Bulllitt, E., Ho, S., Gerig, G., 2004. A Brain Tumor Segmentation Framework Based on Outlier Detection. Medical Image Analysis 8(3) 275-283.

The Application of Interactive Behavior Change Technologies to Enhance Patient Education Among Adults Diagnosed with Diabetes

Koren S. Goodman, PhD Candidate, Holly Gaff, PhD, Elizabeth F. Giles, PhD, PT,
Gianluca De Leo, PhD, MBA

***Abstract*—The use of interactive behavior change technologies (IBCT) by individuals with type 2 diabetes may be a significant mechanism to provide educational and behavioral change messages. The application of these technological strategies promote overall compliance by minimizing traditionally imposed barriers that impact accessibility, complications experienced, and lack of patient education. This research will evaluate interactive behavior change technologies that aim to increase patient education, medication compliance, and quality disease management among adults diagnosed with diabetes when employing the Technology Acceptance Model as a theoretical framework.**

***Index Terms*—Technology acceptance, interactive behavior change technologies, type 2 diabetes, and behavioral change messages.**

I. Introduction

PROJECTED prevalence rates for diabetes are expected to increase through the year 2050 [1]. Direct and indirect costs associated with diabetes in the United States accounted for $174 billion in 2007 [2]. Approximately 25.8 million adults in the United States are affected by the chronic disease diabetes. It is important to thoroughly understand this disease to modify its deleterious effects, develop appropriate and effective treatment options, decrease its economic burden, and improve the quality of life among those diagnosed and those at risk [2]-[4].

Manuscript submitted March 23, 2012. This work was supported by Modeling and Simulation Funding from the Old Dominion University.

Koren S. Goodman is a doctoral candidate in the Health Services Research program in the College of Health Sciences, Old Dominion University, Norfolk, VA 23529 USA (email: kgood006@odu.edu).

Holly D. Gaff, PhD is a Virginia Modeling, Analysis and Simulation Center (VMASC) Affiliated Faculty Member and Assistant Professor in the Department of Biological Sciences in the College of Sciences, Old Dominion University, Norfolk, VA 23529 USA (email: hgaff@odu.edu).

Elizabeth F. Giles, PhD, PT is Senior Lecturer and Director of Clinical Education in the School of Physical Therapy, College of Health Sciences at Old Dominion University, Norfolk, VA 23529 (e-mail: egiles@odu.edu).

G. De Leo, PhD, MBA is an Assistant Professor with a joint appointment in the College of Health Sciences and the Virginia Modeling, Analysis and Simulation Center, Old Dominion University, Norfolk, VA 23529 USA. (email: gdeleo@odu.edu).

Diabetes is a group of diseases characterized by high levels of blood glucose, with multiple complications and premature mortality. Diabetes is associated with older age, obesity, family history of diabetes, history of gestational diabetes, impaired glucose metabolism, physical inactivity, and race and ethnicity [2]-[4]. The care of persons with diabetes is of growing importance to public health because of the newly diagnosed cases each year. Preventive care practices reduce progression among populations disproportionately affected by diabetes [5]-[7].

II. Problem and Motivation

The purpose of this study is to examine the effectiveness of an Interactive Behavior Change Technology in delivering diabetes education to adults, physician diagnosed with type 2 diabetes who are geographically and/or medically underserved. Interactive Behavior Change Technologies (IBCT) are "computer-based tools and systems, including hardware and software that can be used to address health behavior change" [8]. Telemedicine is a technology application used to enhance health promotion and the delivery of health care education [9]. Telemedicine is valuable for communities which are rural and underserved because of its ability to deliver diabetes education to improve clinical outcomes and reduce the risk of diabetes complications [10]-[12]. One of the least expensive forms of IBCT technologies includes telecommunication lines such as mobile phones or an automated telephone call center. Some of the more sophisticated IBCT applications utilize virtual reality [13].

III. Theoretical framework

Predictive models and theoretical frameworks will be explored to examine the effectiveness of an Interactive Behavior Change Technology in delivering diabetes education to adults, physician diagnosed with type 2 diabetes who are geographically and/or medically underserved. The Theory of Planned Behavior and Self Determination Theory are two models that have been investigated as theoretical frameworks for this study. A robust model, the Theory of Planned Behavior posits that the intent to perform a behavior is influenced by an individual's attitude, subjective norm, and perceived behavioral control. This theoretical framework

suggests that an individual's perceived behavioral control is a predictor of behavior [14]-[18]. Self Determination Theory examines the motivation for a behavior. This theory suggests the importance of examining an individual's environment to understand the intent of that behavior [19].

IV. Method

This study will employ a randomized, two-group, pretest-posttest design covering a period of 3 months. Potential participants will be recruited from the established patients of the Riverside Shore Diabetes Center (SDC) of the Riverside Shore Memorial Hospital [20]. Potential participants will include adults, aged 18-64 years, physician diagnosed with the chronic condition type 2 diabetes. Patients will be residents of the Eastern Shore, Virginia, living in Accomack or Northampton counties. All participants will have received a physician diagnosis of type 2 diabetes, having uncontrolled blood glucose levels, over a 3-month period. The HbA1C test measures the amount of glycated hemoglobin in the blood. Providers use this test to assess average blood glucose control over the past 2-3 months [7].

The intervention for this study will consist of the patient receiving a telephone call three times per week from an automated telephone call center following an initial visit with the Shore Diabetes Center. The call center will inquire about the patient's healthcare practices and provide educational messages on how to effectively manage their chronic condition. These educational messages will provide content for glucose monitoring, HbA1c, nutrition therapy, retinopathy screening and treatment, foot care, weight and waist management, physical activity, blood pressure, cholesterol, smoking habits, medication adherence, and the importance of measuring urinary protein [7]. Following the intervention, patients will return to the center for a follow-up assessment.

V. Summary

There is a growing burden of cost associated with the chronic disease diabetes. The onset of complications from diabetes can be delayed or prevented with a healthy lifestyle and behavioral change. The key to reducing the impact of risk factors is the ability to prevent as appropriate, screen, and finally, to treat.

Technology has made a significant impact on diabetes self-management for individuals living in rural and underserved communities over the past decade [21]. Patients enrolled in a diabetes management care program offering increased education about prescription therapy and lifestyle modifications were 16% less likely to have a hospitalization incidence [3],[22]. The delivery of educational messages utilizing technology is becoming essential to quality diabetes management and quality diabetes care. This is a cost effective disease management tool to prevent life-threatening complications experienced from diabetes. The use of technology to enhance diabetes education may overcome traditional barriers [8].

References

[1] J.P. Boyle, A.A. Honeycutt, K.M., Narayan, T.J. Hoerger, L.S. Geiss, H. Chen, and T.J. Thompson, "Projection of diabetes burden through 2050: Impact of changing demography and disease prevalence in the U.S.," *Diabetes Care*, vol. 24, no. 5, pp. 1936-1940, 2001.

[2] Centers for Disease Control and Prevention. (2010, September 5). [Online]. Available at http://www.cdc.gov

[3] A. Greisinger, R. Balkrishnan, R. Shenolikar, O. Wehmanen, S. Muhammad, and P. Champion, "Diabetes care management participation in a primary care setting and subsequent hospitalization risk," *Disease Management*, vol. 7, no. 4, pp. 325-333, 2004.

[4] World Health Organization. (2010, September 5). [Online]. Available at http://www.www.who.int/en/

[5] D. Clancy, D. Cope, K. Magruder, P. Huang, K. Salter, and A. Fields, "Evaluating group visits in an uninsured or inadequately insured patient population," *The Diabetes Educator*, vol. 29, no. 4, pp. 292-302, 2003.

[6] M. Safran and F. Vinicor, "The war against diabetes-How will we know if we are winning?" *Diabetes Care*, vol. 22, no. 3, pp. 508-516, 1999.

[7] American Diabetes Association. (2010, September 5). [Online]. Available at http://www.diabetes.org

[8] R.E. Glasgow, S.S. Bulls, J.D. Piette, and J. Steiner, "Interactive behavior change technology: A partial solution to the competing demands of primary care," *American Journal of Preventive Medicine*, vol. 27, no. 2S, pp. 80-87, 2004.

[9] J.C. Lin, "Applying telecommunication technology to health-care delivery," *Engineering in Medicine and Biology Magazine, IEEE*, vol. 44, no. 1, pp. 28-31, 1999.

[10] P. Bray, M. Roupe, S. Young, J. Harrell, D. Cummings, and L. Wetstone, "Feasibility and effectiveness of system redesign for diabetes care management in rural areas: The eastern North Carolina experience," *The Diabetes Educator*, vol. 31, no. 5, pp. 712-718, 2005.

[11] J. East, P. Krishnamurth, B. Freed, and G. Nosovitski, "Impact of diabetes electronic management system on patient care in a community clinic," *Journal of Medical Quality*, vol. 18, no. 4, pp. 150-154, 2003.

[12] T. Gary, M. Turner, L. Bone, H. Yeh, N. Wang, F. Hill-Briggs, D. Levine, N. Power, M. Hill, C. Saudek, M. McGuire, and F. Brancati, "A randomized controlled trial of the effects of nurse case manager and community health worker team interventions in urban African-Americans with type 2 diabetes," *Controlled Clinical Trials*, vol. 25, no. 1, pp. 53-66, 2004.

[13] K. Patrick, "Information technology and the future of preventive medicine potential, pitfalls, and policy," *American Journal of Preventive Medicine*, vol. 19, no. 2, pp. 132-135, 2000.

[14] K. Glanz, B.K. Rimer, and F.M. Lewis (Eds)., "Health behavior and health education: Theory, research, and practice," 3rd ed., San Francisco: Jossey-Bass, 2002.

[15] Icek Ajzen. (2010, October 12). [Online]. Available at http://www.people.umass.edu/aizen/index.html

[16] D.O. Omondi, M.K. Walingo, G.M. Mbagaya, and L.O.A. Othuon, "Understanding physical activity behavior of type 2 diabetics using theory of planned behavior and structural equation modeling," *International Journal of Social Sciences*, vol. 5, no. 3, pp. 160-167, 2010.

[17] C.L. Blue, "Does the theory of planned behavior identify diabetes-related cognitions for intention to be physically active and eat a healthy diet?" Public Health Nursing, vol. 24, no. 2, pp. 141-150, 2007.

[18] A. Kagee and M. van der Merwe, "Predicting treatment adherence among patients attending primary health care clinics: The utility of the theory of planned behavior," *South African Journal of Psychology*, vol. 36. no. 4., pp. 699-714, 2006.

[19] R. Ryan and E. Deci, "Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being," *American Psychologist*, vol. 55, no. 1, pp. 68-78, 2000.

[20] Riverside Shore Memorial Hospital. (2010, December 1). [Online]. Available at http://shorehealthservices.org/diabetes.html

[21] T. O'Brien and S.A. Denham, "Diabetes care and education in rural regions," *The Diabetes Educator*, vol. 34 no. 2, pp. 334-347, 2008.

J. Pope, "Implementing EHRs requires a shift in thinking," *Health Management Technology*, vol. 27, no. 6, pp. 24-26, 2006.

Gaming and Virtual Reality

VMASC Track Chair: Mr. Hector Garcia
MSVE Track Chair: Dr. Yuzhong Shen

The Effects of Game Design Principles on Learning Outcomes in Serious Games
Author(s): Michael W. Martin, and Yuzhong Shen

Flood Watch: A Mobile App for Flood Reporting
Author(s): Gary Lawson, and Yuzhong Shen

Design of Video Overlay and Camera Pan/Tilt for UAV
Author(s): Rahul Shambhuni and Yiannis Papelis

Emerging Trends in Geospatial Information Visualization for Decision Making in Regard of Chronic Disease Management
Author(s): Ange-Lionel, and Rafael Diaz

The Effects of Game Design Principles on Learning Outcomes in Serious Games

Michael W. Martin, Yuzhong Shen

***Abstract*—This paper details the administration and results of an experiment conducted to assess the impact of three video game design concepts upon learning outcomes. The principles tested include game aesthetics, player choice, and player competition. The experiment participants were asked to play a serious game over the course of a week, and the learning outcome was measured by comparing their pre-test and post-test scores. The results of a one-tailed t-test indicated, with a p-value of 0.043, that there was a statistically significant effect of the aesthetic presentation of the game upon the learning outcome. There was no indication of a significant effect by the player choice or player competition conditions, but the results from these experiment groups do illuminate some interesting potential interactions between the conditions and learning, as well as possible future lines of experimental inquiry.**

***Index Terms*—Serious Games, Game Design Principles, Learning Outcomes, Learning**

I. Introduction

Video games are an ascendant aspect of modern culture. In the mid 2000's, the commercial software entertainment industry eclipsed Hollywood in terms of revenue [1], and in the 2008, sales of video games surpassed the global sales of DVDs [2]. The ubiquity of smart phones and the growing popularity of online communities such as Facebook™ have served as the vectors by which video games have infiltrated many aspects of our daily lives.

For the past decade and a half, there has been a commensurate interest in the application of these video games towards productive ends. This emerging field is known as "serious games". In some ways, serious games are the aspiring successor to the defunct field of "edutainment" [3]. Whereas edutainment may have been seen as the worst of game development in the 1980s and 1990s, serious games intend to capitalize on the current explosive growth and acceptance of video games.

There are several of educational scholars and notable video game developers who have offered a wealth of insight into the potential and perceived benefits of serious games [4-8]. Much of this work is founded upon informative personal and professional anecdotal experiences in the video game development or educational fields. However, there is a shortage of empirical data to support these suppositions. There are many presumptions regarding the possible merits of the field, but there is not a large body of experimental data from which to verify that the merits exist, and do indeed result from serious game usage. Further, while the theorized or anecdotally observed results may be well detailed, there is little discussion how to actually create serious games and engender the described benefits. This research seeks to address this shortfall by conducting quantitative experiments to empirically test the efficacy of fundamental game design principles in eliciting learning outcomes.

This work builds upon the theoretical foundation presented previously by the authors [9-11]. The experiment detailed in this paper was carried out using the Element Solitaire© serious game, which was developed by the authors. This game teaches students the placement of elements on the periodic table of elements. Its learning content is simple declarative knowledge (rote memorization), and the game is intended for use by novice learners in the chemistry field. The game is conceptually based on a combination of the game of solitaire with the periodic table of elements.

II. Experiment Design

This experiment was designed to test the following three hypotheses:

> *1. A serious game with added aesthetic features, to include graphics and music, will result in higher learning outcomes than an identical one without music and graphics.*
>
> *2. A serious game with meaningful choices will result in higher learning outcomes than an identical game without meaningful choices.*
>
> *3. A serious game with competition will result in higher learning outcomes than an identical one without competition.*

These hypotheses, in turn, were directly translated into experimental conditions. The first condition was the presence and quality of aesthetic effects within the game. For Element Solitaire©, this includes enhanced graphical presentation, as

Michael W. Martin is a PhD Candidate in the Modeling, Simulation and Visualization Engineering Department at Old Dominion University, Norfolk, VA, 23529 USA (e-mail: mmart081@odu.edu).

Yuzhong Shen is an Assistant Professor in the Department of Modeling, Simulation, and Visualization Engineering at Old Dominion University, Norfolk, VA, 23529 USA (E-mail: yshen@odu.edu).

well as additional aesthetic effects such as sound effects, music, card animation, and "sparkle" particle effects. Figures 1, 2, 3, and 4 provide comparisons of the graphical differences in this condition. However, it is important to emphasize that the difference in aesthetic presentation went beyond static graphics, and included not only animation, but also sound effect and music. Care was taken, however, to ensure that the same learning content is presented in each condition.

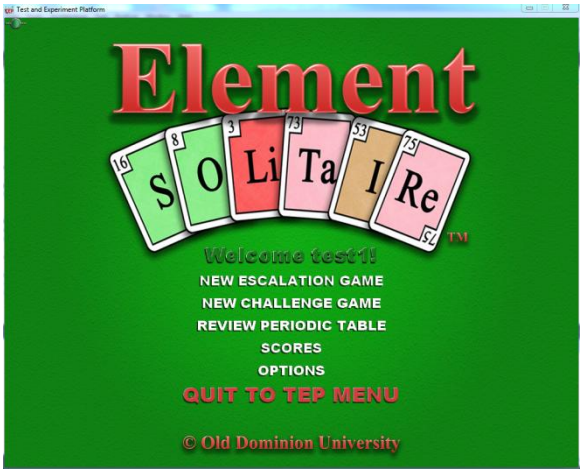


Fig. 1. High Aesthetic Presentation. This figure shows the title screen from the high aesthetic presentation of the game.

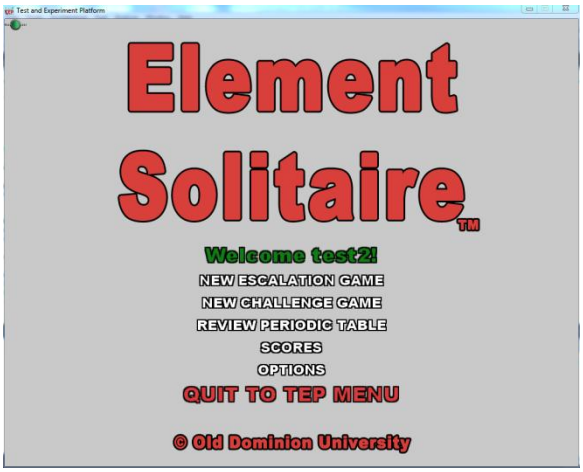


Fig. 2. Low Aesthetic Presentation. This figure shows the title screen from the low aesthetic presentation of the game.

The second condition is the presence of specific choices within the game. These choices are made of two specific options available to the player – the "hint" and the "skip". The learner has a budget of ten "skips" they can use to delay having to place an element which they might be unsure of. The "hint" gives the learner the option to "purchase" hints for a small score penalty. These hints give the player information needed to correctly place an element, but diminish the score which they receive when they ultimately place the element correctly.

The third condition is the presence of a network enabled score board. The network scoreboard allowed the player to see what is commonly referred to in the entertainment software industry as a "leaderboard". This leaderboard listed the names and scores of the top ten players. This is slightly different from a more direct scoreboard, as each player may only have one entry on the board, marking their highest score. Players without this "competition" condition are only shown their own scores, in a traditional scoreboard fashion. Figures 5 and 6 show the difference in the scoreboards between the two condition levels. Note that for privacy purposes, the usernames in these pictures have been intentionally blurred out.

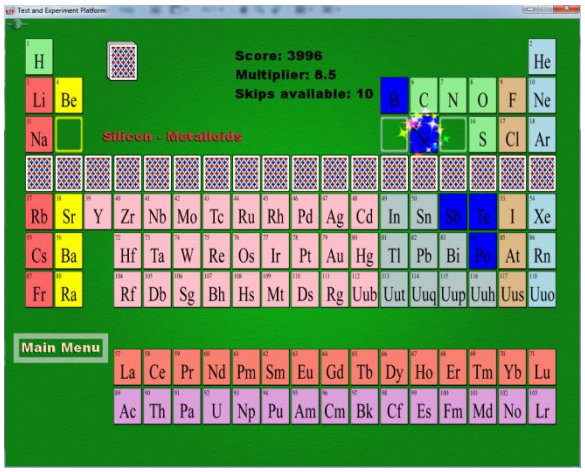


Fig. 3. High Aesthetic Game. This figure shows the game in play with the high aesthetic condition. The sparkle effect is visible around the newly placed Silicon (Si) element.

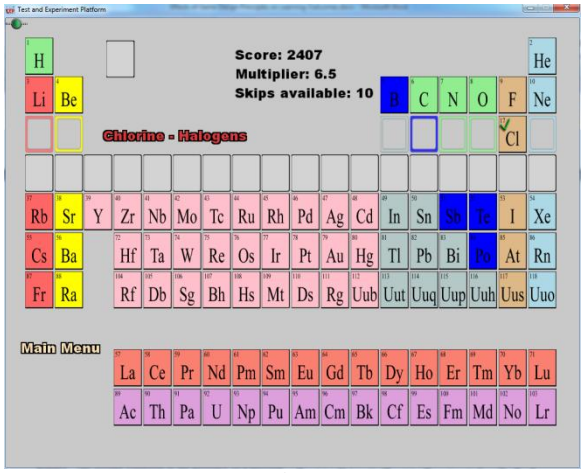


Fig. 4. Low Aesthetic Game. This figure shows the game in play with the low aesthetic condition.

The first experimental treatment group was a baseline treatment, in which all three conditions were present. The game featured the full aesthetic presentation, the full breadth of available gameplay choices, and the networked leaderboard. This version of Element Solitaire© was considered to be the standard version of the game, and this treatment groups served as the baseline from which to compare the performance of the other experimental treatments.

These remaining experimental treatments were reductive iterations of this baseline treatment. The second treatment retained the "choices" and "score board" conditions, and omitted the "aesthetics" condition. The third treatment

retained the "choices" and "aesthetics" conditions, but omitted the "score board" condition. The fourth and final treatment omitted the "choices" condition, while retaining the "score board" and "aesthetics" conditions. The control group effectively omitted all conditions by not being exposed to the game. These "control group" participants were only asked to take the pre- and post-test. The Group-to-Condition interaction is shown in Figure 7.

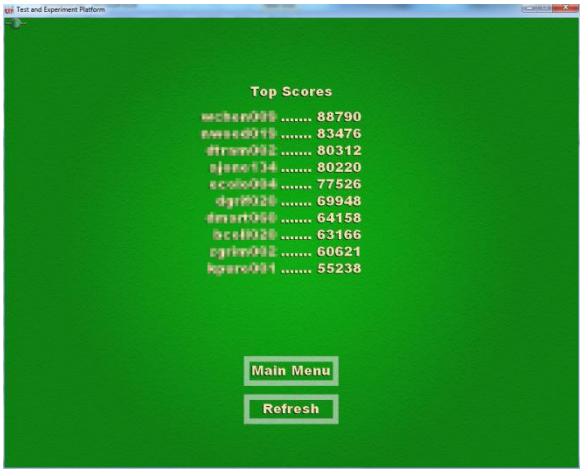


Fig. 5. Competition Scoreboard. This figure shows the score board from a high competition condition participant. The scoreboard shows a leaderboard displaying the single top score of each of the ten highest scoring players. The participants' usernames are blurred to protect their identities.

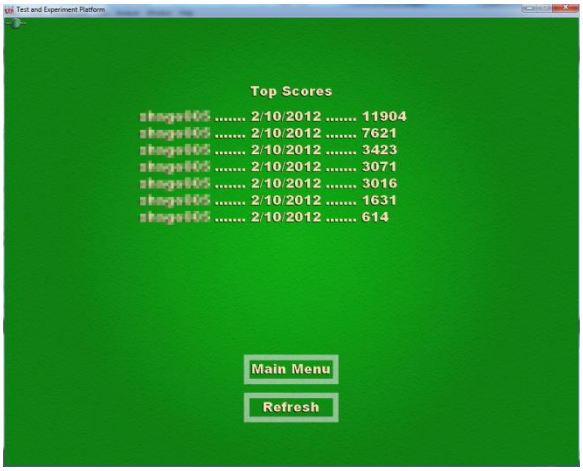


Fig. 6. No Competition condition. This figure shows the scoreboard from a "No Competition" condition participant. The participant's username is blurred to protect their identity.

Group to Condition Interactions

Group \ Condition	Aesthetics	Choice	Competition (Score Board)
Control Group	-	-	-
Baseline	X	X	X
Diminished Aesthetics	-	X	X
Diminished Choices	X	-	X
No Competition	X	X	-

- denotes absence of condition
X denotes presence of condition

Fig. 7. Group to Condition Interactions. This chart demonstrates how the three hypothesis conditions are translated into five experimental groups.

All treatment groups took a pre-test, were asked to play Element Solitaire© four times (excluding the Control Group) and then were asked to take the post-test as a means of assessing their learning outcomes. The learning outcome is defined, for the purposes of this experiment, as the difference between the pre- and post-tests scores. The post-test was administered one week after the pre-test, and participants had that week during which to play the game. The dependent variable for the experiment is the difference in the participants' pre- and post-test scores.

III. Experiment Execution

A. Participants

The research participants were drawn from the Old Dominion University Chemistry Department. Specifically, the participants were all students in the CHEM 123 Foundations of Chemistry course. This course is described as a core requirement for science and engineering majors, and is intended to prepare students for subsequent studies in molecular science. The students taking this course are primarily freshmen or sophomores in their undergraduate studies, and are typically between 18 and 23 years old. This particular demographic was chosen to increase the probability that the participants would be novices with regards to the learning content covered by the serious game interventions, but that they would still have a sufficient threshold level of interest in the subject matter.

The invitation to participate in the program was given to two sections of this course, each taught by different instructors. The two classes had a total of 533 students. From these potential participants, 235 registered to participate in the experiment. 172 of the participants actually completed the experiment. It should be reiterated that the experiment was relatively intensive in terms of required effort. In order to complete the experiment, the participants had to take a pre-test on the periodic table of elements, play the game a minimum of four times, and then take the post-test. The estimated combined total time commitment is estimated at approximately one hour, over the course of one week.

The participation goal was to have 40 participants in each experimental group. The actual results fell slightly short of this goal. However, the number of samples in each treatment still provides a sufficient number of measures for statistical analysis. Fig. 8 shows the number of participants that completed the experiment, broken down by experimental group.

Completed Samples Per Treatment	
Total	172
Control	38
Baseline	32
Diminished Aesthetics	29
Diminished Choices	36
No Competition	37

Fig. 8. Total number of Completed Participants. This chart shows the total number of participants who completed the experiment, broken down by experimental group.

B. *Experiment Administration*

For the execution of the experiment, the research participant population was given a digital document (in PDF format) which provided a brief overview of the experiment, anticipated time requirements, and a link from which they could download the experiment program.

The experiment program, entitled Testing and Experiment Platform (TEP), was designed to encapsulate the experimental design and administration into a desktop application which the participants could use at their own convenience. Providing this testing platform alleviated some of the practical barriers to student participation in a relatively time consuming, voluntary experiment.

When a participant first used the TEP program, they were prompted to enter a unique username. The participant was then given a pre-test, in which they were asked to enter in the symbol names for as many of the 118 chemical elements as they can remember. Once this was complete, the pre-test score and username were stored on a network server.

After a participant was registered, the network server assigned the participant to a treatment group. Treatment assignments were incremented with each participant, ensuring that each treatment group never varied in population by more than one.

Based on the treatment assigned by the server, the TEP client then provided the participants with access to the variant of the Element Solitaire© serious game intervention appropriate to their treatment group. The end result was intended to make the treatment assignments opaque to the participants. Whenever the participants played the serious game, detailed records of their gameplay sessions were automatically reported to the server. These records include such information as a date-time stamp for when a game was started, what kind of game was played, how long the participant played the game, what score they received, whether they completed the game, what elements they missed, and how many hints or skips they used.

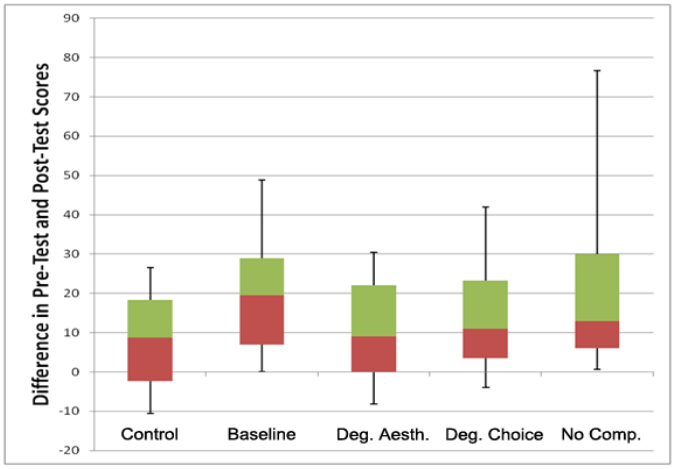


Fig. 9. Boxplot for the Measures of Learning Outcome, grouped by Treatment.

IV. Results

As described in Section III, 172 of the original 235 participants completed the experiment. While the original apportionment of participants to experimental groups was tightly controlled, obviously the apportionment of completed participants varied. Figure 9 shows the boxplots of the differences between each participant's pre-test and post-test score, as grouped by treatment. For these boxplots, the red box represents the region of scores between the 25% and 50% percentile. The green box represents the region between the 50% and 75% percentile, and the whiskers represent the respective 10^{th} and 90^{th} percentiles.

Initial inspection of the boxplots indicates that the baseline treatment group had a relatively high positive difference in the pre-test and post-test scores in comparison to all other treatment groups, with the exception of the No Competition group. Even with the No Competition group, the median value in the Baseline group is well below the median value for the Baseline group, as evidenced in Figure 10. However, the mean for the No Competition group was actually higher than the mean for the Baseline group.

	Control	*Baseline*	*Deg. Aesth.*	*Deg. Choice*	*No Comp.*
Mean	7.92	20.09	8.90	16.67	23.08
Median	6.50	19.5	9	11	13
St. Dev.	23.94	21.40	26.96	25.33	30.14
Count	38	32	29	36	37

Fig. 10. Descriptive Statistics for the Differences in Pre-Test and Post-Test for each experimental group.

T-tests and Resultant P-Values

Test	Baseline – Deg. Aesth.	Baseline – Deg. Choice	Baseline – No Comp.	Baseline – Control	Deg. Aesth. – Control
P-Value	0.0426	0.2749	0.3176	0.016	0.4394
Result	Significant – Reject H_o	Not Significant – Cannot Reject H_o	Not Significant – Cannot Reject H_o	Significant – Reject H_o	Not Significant – Cannot Reject H_o

$$H_0: The\ difference\ between\ treatments = 0$$
$$H_A: The\ difference\ between\ treatments > 0$$

Critical α-value of 0.05

Fig. 11. T-test and P-value Results for Mean Comparisons Between Experimental Groups. The first row identifies the means being compared, the second row cites the resultant p-value, and the third row describes the interetation.

In order to assess the statistical significance of the results, the authors selected a t-test [12-15]. The t-test is more tolerant of smaller sample sizes than the more common and more discriminating z-test. One implication of this test selection is that the results are more conservative, in favor of the null hypothesis, which creates a greater burden of proof. Figure 11 shows the t-test results for five comparisons of the treatment means. The means being compared are listed in the top row. The p-values indicate the probability of finding the given observations if the null hypotheses are true. The authors have selected a critical value (α) of 0.05 as the maximum p-value acceptable for rejection of a null hypothesis.

These tests indicate that there was a statistically significant difference between the learning outcomes of the Baseline and the Degraded Aesthetics experimental groups. In fact, the negative impact of the degraded aesthetics was so substantial

that there was no statistical significance between the measured learning outcome of the Degraded Graphics experimental group and the Control group, who did not play the game at all.

The tests do not, however, indicate a statistically significant difference between the Baseline group and the Degraded Choice experimental group or No Competition experimental group. In the case of the No Competition group, cursory examination of the descriptive statistics for each group in Figure 10, as well as their respective boxplots in Figure 9, suggests that in some measures, the No Competition group might actually have perform better than the baseline group.

As a means of comparison, the test of the difference between then Baseline group and the Control group is included, and it does indeed, indicate that the Baseline Group has a statistically significantly higher learning outcome than those who did not play the game.

V. Discussion

A. Aesthetic Presentation

The results of the t-test comparing the Baseline Group and the Degraded Aesthetic Experimental Group do support the hypothesis that aesthetic presentation has an impact on learning outcomes. While the authors expected some indication of this outcome, the scope of the effect on the learning outcome, to the point where those playing with the degraded graphics did not perform statistically better than those who did not play the game, was surprising.

This surprise was amplified even more when the authors examined the number of average games played by each group. The mean for the number of games which the Baseline group played and completed was 5.21 and 4.15, respectively (not all played games were completed). In comparison, the mean of played and completed games for the Degraded Aesthetic group was 6.23 and 4.98.

Caution should be taken when comparing these numbers as the experiment design created a significant confounding effect upon the number of games any participant played. Specifically, the participants were told they had to play a minimum of four games. One could imagine scenarios in which this experiment imperative could either artificially inflate or depress the number of games which a participant might play otherwise. However, keeping that in mind, both experimental groups were subject to this same effect.

Bearing this caveat in mind, the difference still suggests that graphics not only improved the learning effect, but possibly made the learning more effective: those with better graphics improved much more with fewer games.

One possible explanation for this result might be a variation on the psychological "halo" effect [16]. The high quality aesthetic presentation might create a positive bias in the mind of the user, prompting them to engage in a more effective and efficient manner with the learning material. In contrast, the low quality aesthetic presentation may cause the user to dismiss the potential learning out of hand.

Unfortunately for serious game developers, the practical corollary to this principle is that enhanced graphics tend to cost more and take more time to develop. The entertainment software industry has been struggling with the burden of ballooning game development budgets, attributable, in large part, to the increased cost of creating high quality assets [2, 17-19]. The results of this research seem to indicate that serious game developers might have to be concerned with the same challenge of delivering high quality aesthetic presentations to their users as faced by the entertainment industry.

B. Choice

The statistical analysis of the Degraded Choice experimental group indicates that the treatment had not effect upon the learning outcome. In trying got understand this outcome, the authors examined the records of the games played.

In the course of the experiment, 1067 separate games were recorded. 789 of these games were played by participants who were not in the Degraded Choice treatment group. In those 789 games, only 3 total "hints" were used. In contrast, 349 "skips" were used.

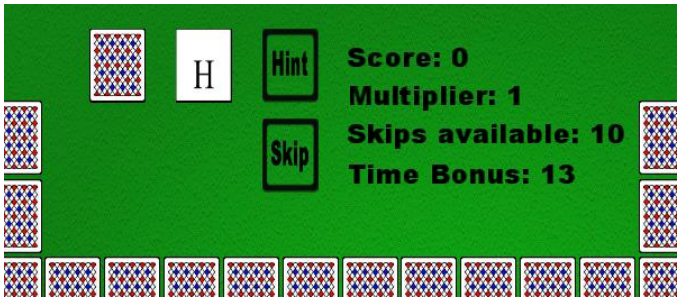


Fig. 12. The Placement of the Hint and Skip Controls.

Given the proximity and similarity of the "hint" user interface to the "skip" user interface, as showing in Figure 12, it seems unlikely that the participants were unaware of the presence of the hint button. More likely, it seems that the Hint button was never considered a viable or rewarding option to the player.

Game Developer D. Schubert provides insight into the role of player choice in creating entertainment game, and provides guidelines for creating "meaningful choices"[20]. Within this context, it seems likely that the relatively low penalty for failing to place an element (points are deducted) may make the option of getting a hint an unviable and non-meaningful choice.

A possible reformulation of the current game might be to end the game when the player misplaces a specified number of elements. This version of player "death" could incentivize the use of the hint button, and induce the player to consider more carefully their available options. It may be informative to compare this alter game structure which places more emphasis on player choice with the existing, more forgiving game structure to further explore the role of choice in learning.

C. Competition

As with Choice, the analysis revealed no statistically significant difference between the No Competition and Baseline experimental groups. As discussed in Section IV and demonstrated in Figures 9 and 10, however, the No Competition experiment group exhibits a much wider variance

in the performance than any of the other experimental groups, and its mean value is actually higher than the mean of the Baseline Group.

Possible explanations for this result can be found in educational research on the subject of competition. Educational literature discusses proposed pre-requisites for creating a learning environment conducive to constructive competition, as opposed to destructive competition [21-23]. Constructive competition can enhance learner motivation and performance, while destructive competition can have the opposite effect. Without strict guidance and cues on the nature of the competition, the affective result upon the individual learner depends, in large part, upon how they internalize the competition. Competitive personalities may be motivated by the competitive nature of the game, while non-competitive personalities may become de-motivated. The two divergent responses to the same stimuli could create a potentially bi-modal response.

This may account for the wide variation seen in the No Competition experimental group measurements. While some participants might have perceived the leaderboard as a destructive competitive artifact, the No Competition experimental group was only shown a history of their own scores. In effect, they were only competing with themselves, which, in some cases, may have been perceived as more constructive.

D. Validity and Experiment Integrity

During the administration of the experiment, it became apparent that efforts to create opaque walls between the treatment groups were not entirely effective. Within a few days of beginning the experiment, the authors received e-mails from concerned students asking why they didn't get to "play the game". Quick examination of the server database revealed that all of these students were in the control group. The authors did not receive any e-mails from participants asking about the differences between the experimental groups, but it would seem likely that the control group participants were not unique in being exposed to the interventions assigned to other treatment groups.

The underlying intent for obscuring the different treatment groups from the participants was to maintain the integrity of the treatment groups in order to reduce the potential for treatment diffusion and compensatory rivalry effects between the respective subjects. While the effects of compensatory rivalry may be harder to judge, the high investment requirement in participating in the experiment, or even playing just one game, makes the likelihood of treatment diffusion effects very low. Even if a participant were to find out about the other treatments, actually using these other treatments, due to the setup of the TEP program, would involve a significant amount of effort. While it feasibly could have happened, it does not seem likely to have happened, at least not with sufficient frequency to have affected the outcomes.

Lastly, there is also the specter that the convenience of the TEP program also created the potential for participants to cheat while using the program by referencing some external aid. Looking at the pre-test data, there are some anomalous results at the upper end of the spectrum, as shown in Figure 13. There is no evidence that these scores are the result of cheating, but regardless, they represent likely outliers in the distribution. It is presumable that someone who scores that highly on the pre-test may not fit the desired participant target demographic descriptor of a "novice" in the target learning content.

In spite of these potentially outlier values, t-tests on the experimental data with those participants culled out do not produce dramatically different results. The p-value for the difference between the Baseline and the Degraded Aesthetics group drops slightly to 0.016. The difference between the Baseline and No Choice groups does drop to 0.11, but it is still not below the critical value of 0.05. And lastly, the p-value for the difference between the Baseline and No Competition group increases slightly to 0.35.

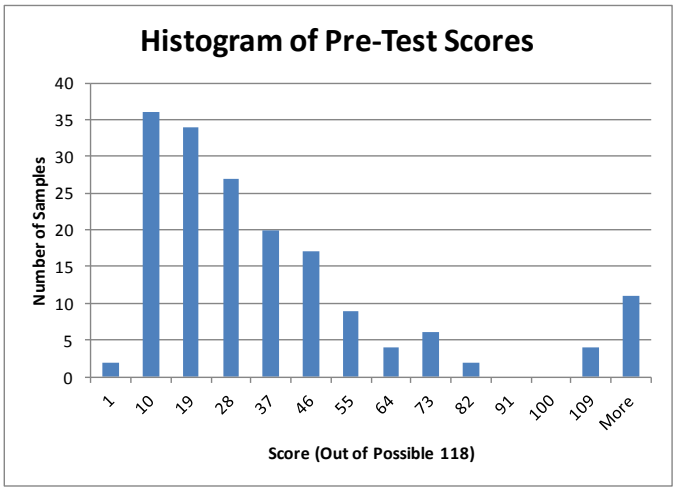


Fig. 13. Histogram of Pre-Test Scores. Note the tail on the right side, consisting of 15 participants who scored higher than 100.

VI. Conclusions

The results of this experiment support the hypothesis that the aesthetic presentation of a serious game can have a significant effect upon learning outcome it produces. A degraded aesthetic presentation may indeed completely nullify the learning outcomes of playing a serious game.

The experiment does not indicate a significant effect from the choices provided in Element Solitaire©. However, the usage reports indicate that the participants saw little need to use the available options. This indicates a design shortfall in the game itself, and also highlights some consideration for future versions of the game.

Lastly, there was no evidence to support the hypothesis that competition increases learning outcomes. However, the responses indicate some interesting considerations for future iterations of Element Solitaire©, specifically with regards to how competition is presented to the learner.

References

[1] M. Yi, "They Got Game," in *San Francisco Chronicle*, ed. San Francisco: Hearst Newspapers, 2004.

[2] N. V. Zelfden. (2009, 6 March). *What's Killing the Video-Game Business?* Available: http://www.slate.com/articles/technology/gaming/2009/02/whats_killing_the_videogame_business.single.html
[3] M. Zyda, "From Visual Simulation to Virtual Reality to Games," *Computer*, p. 8, September 2005 2005.
[4] M. Prensky, *Digital Game-Based Learning*. New York: McGraw-Hill, 2001.
[5] R. Koster, *A Theory of Fun for Game Design*. Scottsdale, AZ: Paraglyph Press, 2005.
[6] J. P. Gee, *What Video Games Have to Teach Us About Learning and Literacy*, Rev. and updated ed. New York: Palgrave Macmillan, 2007.
[7] D. Michael, *Serious Games: Games That Educate, Train and Inform*. Boston, Mass.: Thomson Course Technology, 2006.
[8] B. E. Shelton and D. A. Wiley, Eds., *The Design and Use of Simulation Computer Games in Education* (Modeling and Simulations for Learning and Instruction. Rotterdam, The Netherlands: Sense Publishers, 2007, p.^pp. Pages.
[9] M. Martin and Y. Shen, "Defining and Leveraging Game Characteristics for Serious Games," presented at the MODSIM Conference & Expo 2010, Virginia Beach, VA, 2010.
[10] M. Martin and Y. Shen, "Defining Learning Space in a Serious Game in Terms of Operative and Resultant Actions," presented at the Virginia Modeling, Analysis and Simulation Center 2011 Student Capstone Conference, Sufolk, VA, 2010.
[11] M. W. Martin and Y. Shen, "Differentiating Between Serious Games and Computer Aided Instruction," presented at the Virginia Modeling, Analysis and Simulation Center 2010 Student Capstone Conference, Suffolk, VA, 2010.
[12] J. L. Devore, *Probability and Statistics for Engineering and the Sciences*, 6th ed. Belmont, CA: Thomson-Brooks/Cole, 2004.
[13] D. C. Montgomery and G. C. Runger, *Applied Statistics and Probability for Engineers*, 3rd ed. New York: Wiley, 2003.
[14] D. J. Rumsey, *Statistics for Dummies*. Hoboken, N.J.: Wiley, 2003.
[15] D. J. Rumsey, *Statistics II for Dummies*. Hoboken, NJ ;: Wiley Pub., 2009.
[16] R. E. Nisbett and T. D. Wilson, "The Halo Effect: Evidence for Unconscious Alteration of Judgments," *Journal of Personality and Social Psychology*, vol. 35, pp. 250-256, 1977.
[17] (2009, 6 March). *Interview: Krome's Robert Walsh*. Available: http://www.develop-online.net/features/484/Interview-Kromes-Robert-Walsh
[18] Y. Takatsuki. (2007, 6 March). *Cost headache for game developers*. Available: http://news.bbc.co.uk/2/hi/business/7151961.stm
[19] K. Graft. (2010, 6 March). *EA: Budgets for Games Going In Reverse Direction*. Available: http://www.gamasutra.com/view/news/30075/EA_Budgets_For_Games_Going_In_Reverse_Direction.php
[20] D. Schubert, "Designing Choice," *Game Developer Magazine*, p. 2, 2008.
[21] P. Williams and S. Sheridan, "Conditions for Collaborative Learning and Constructive Competition in School," *Educational Research*, vol. 52, pp. 335-350, 2010.
[22] S. Sheridan and P. Williams, "Developing Individual Goals, Shared Goals, and the Goals of Others: Dimensions of Constructive Competition in Learning Contexts," *Scandinavian Journal of Educational Research*, vol. 55, pp. 145-164, 2011.
[23] J. C. Burguillo, "Using Game Theory and Competition-Based Learning to Stimulate Student Motivation and Performance," *Computers & Education*, vol. 55, pp. 566-575, 2010.

Flood Watch: A Mobile App for Flood Reporting

Gary Lawson and Yuzhong Shen
Department of Modeling, Simulation, and Visualization Engineering
Old Dominion University
Norfolk, Virginia
glaws003@odu.edu and yshen@odu.edu

***Abstract*—Annually severe weather claims thousands of lives and causes billions of dollars in damages and flooding is the most costly. Many services exist to warn users of imminent severe weather including television and radio broadcasts as well as updates via the Internet. However, there may be situations where these services are unavailable, or may be too broad to capture the conditions felt by each city or county. To answer these problems, a prototype Android smartphone application, Flood Watch, has been developed. It allows users to report and browse flooding in their local area.**

Flood Watch was derived from an open source platform, Ushahidi; however, Ushahidi does not allow for reports to be queried by location or date which severely limits its functionality. Additionally, reports submitted via mobile phones must be approved manually by an administrator before they are publicly available. This paper addresses these problems and discusses the solutions implemented in the Flood Watch prototype.

Keywords-Flood, Android, Mobile, Alert, App, Ushahidi, Flood Watch

I. INTRODUCTION

Flood Watch is a prototype app developed for Android mobile devices that alerts users of flooding in their local area. The app is based on an open source implementation provided by Ushahidi. This paper expands on the original platform and discusses the contributions made to Flood Watch. This section will provide a brief background on flooding in the United States and why mobile phone apps are an appropriate avenue for a service such as flood reporting. Additionally this section will provide an introduction to the prototype app.

In 2010, severe weather cost the United States approximately $10 billion in damages alone with 490 fatalities and over 2,000 injured [1]. The highest ranking severe weather event for 2010 was flooding accounting for over $5 billion dollars in damages and 103 deaths [1]. Flooding is responsible for approximately 94 fatalities per year [2]. Currently the National Weather Service, managed by the National Oceanic and Atmospheric Administration, is responsible for issuing warnings and alerts through local television and radio frequencies. Additionally, information regarding weather alerts and warnings may be found on the web [3]. Alerts and warnings would be beneficial to mobile users. Although there are weather apps that update users of imminent condition, an app that informs users of current local conditions would be beneficial. This requires residents to own a mobile device with broadband access.

As of 2011, there were over 6 billion mobile-cellular subscriptions; accounting for approximately 87% of the global population [4]. Of these 6 billion mobile subscriptions, approximately 1.2 billion include active mobile-broadband subscriptions [4]. These figures are expected to increase throughout the following years. Therefore, it is reasonable to assume that most citizens may be alerted by a product, such as the one developed and described in this paper, in the event of flooding.

Flood Watch is a mobile app developed for Android mobile devices. It is associated with a web based server responsible for collecting, archiving, and providing detailed reports to the mobile app. Flood Watch heavily relies on users to upload report information whenever flooding is encountered as well as provide pictures of the occurrence. The app takes advantage of Google Maps to display the uploaded reports using icons that mark the physical locations of the uploaded reports. Additionally, users may view the available reports in a list format sorted chronologically with the most recent updates at the top.

Flood Watch was developed in response to the high flood waters recorded in the Hampton Roads area during Hurricane Irene in August 2011. In Norfolk, VA, there were reported waters levels of 7.54ft during the surge [5]. Similarly, Hurricane Isabel of 2003 pushed water levels to 7.89ft, and Hurley, the nor'easter of 2009 pushed 7.75ft [5]. The Hampton Roads area could benefit greatly from a service such as Flood Watch. However, Hampton Roads is not the only area affected by severe flooding. For this reason, Flood Watch was developed with the intent to provide user reported information throughout the United States. If Flood Watch is widely used and accepted, it could be expanded to other countries as well.

The subsequent sections are outlined as follows: Section II provides an overview of the Ushahidi platform. Section III describes the limitations of the Ushahidi platform and app and addresses them by describing the implemented solutions. Section IV outlines potential improvements and future work and Section V is the conclusion to this document.

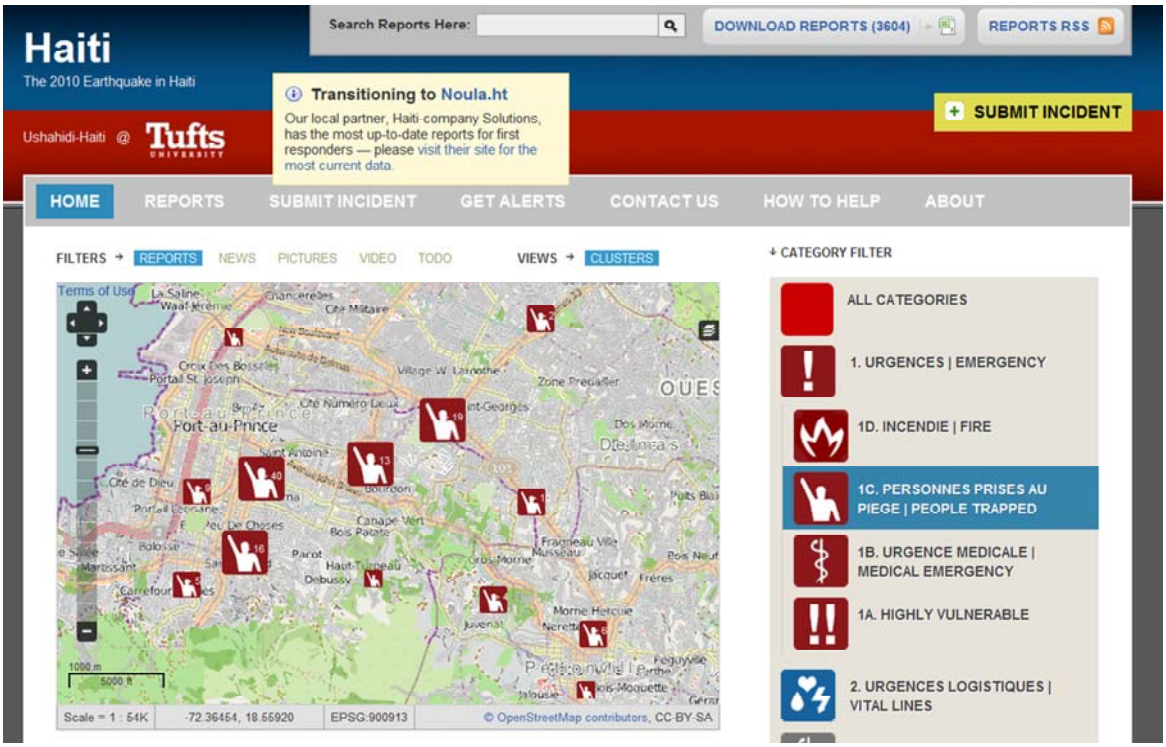


Figure 1: Ushahidi Web Platform

II. USHAHIDI

Ushahidi is a nonprofit organization based in Africa who has developed free, open source software primarily used for information collection, visualization, and interactive mapping [6]. Its original development was to map accounts of violence occurring in Kenya in 2008 via the web or mobile app [6]. The Ushahidi platform is a web based tool that allows users to submit reports to plot specific occurrences of events. These events may be any type of information that takes place in physical locations; for instance Figure 1 provides an example deployment of Ushahidi used to track the earthquake in Haiti of 2010 [7]. The current view shows the areas and count of people trapped in various areas of Haiti. Additionally there is an Ushahidi mobile app available for Android and iOS that provides users with the same functionality as the web platform [8].

The Haiti web page as shown in Figure 1 relied on reports from many different sources; however, most came from individuals who were affected by the earthquake. Likewise, many other deployments similar to the one shown above heavily rely on users submitting inquiries about events related to the deployment. Flood Watch, very similar to these other Ushahidi deployments, heavily relies on inquiries provided by users experiencing flooding. For instance, news reporters generally take footage of flooded streets throughout hurricanes and other large storms; it would be advantageous for them to also provide short reports with photos of the same footage to forewarn users of the flooding occurring with the Flood Watch app.

A. Ushahidi Web Platform

The default platform may be found on the Crowdmap website [9]. Here, users may create an account and create or follow deployments of the platform. A deployment is defined as an instance of the Ushahidi platform with a particular purpose; for instance, Flood Watch itself is a deployment as it is an instance of the platform with the purpose of alerting users of flooding.

Figure 1 provides an example of the default layout of the Ushahidi Platform [9]. The main page provides the user with a map courtesy of Google Maps and a category filter specific to the deployments topic. The user may navigate within the map pane to see markers representing the locations of reports that have been approved. Selecting a particular marker on the map provides the user with a few details about the report including the title and any picture associated with it, as well as the option to view the full report in detail. Selecting this option, or by selecting the "Reports" tab will bring up a list of recent reports submitted to the deployment that have been approved by the administrator of the deployment.

Throughout this document, there have been several references to a report that is submitted to the Ushahidi Platform and then available to the user. These reports are

used to identify events, and therefore must contain some specific information. Each report must have a title, date/time, latitude/longitude, formal location name, the category the information applies to and a description. Additionally there are a few optional fields; a report may have the name and email address of the user whom submitted the report as well as an image file.

It was also mentioned that reports available on the main page of the deployment must be approved. By default any report submitted either on the web or by mobile device is not approved. Approval of reports is primarily handled by the administrator of the deployment; however, there is an option that allows all reports submitted via the web page to be automatically approved. This excludes reports generated and submitted via a mobile device.

B. Ushahidi Mobile App

Ushahidi has also provided open source support for Android and iOS mobile devices. This research focuses on support for Android devices but may be expanded to include iOS support at a later time. Figure 2 below provides an example of the mobile app displaying reports using Google Maps [10]. Figure 3 below provides an example of the "Add Report" screen with the Google Maps input option [10]. Similar to the web platform, the mobile app also allows users to submit reports to the deployment as well as view approved reports retrieved from the web deployment server.

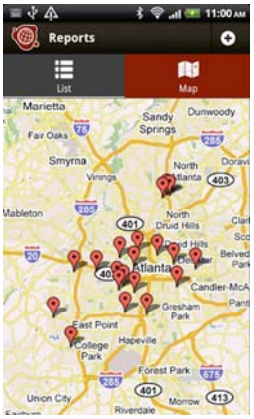


Figure 2: Reports shown using Google Maps

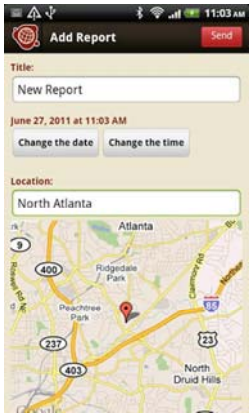


Figure 3: "Add Report" screen

III. CONTRIBUTIONS

This section outlines the limitations of the original Ushahidi web platform and mobile app. As these limitations are addressed, the implemented solutions in the current installment of the Flood Watch app will be described. However, before the limitations may be described, the expectations of the Flood Watch app must be set forth.

The information collected by Flood Watch is flooding event data. Users are encouraged, wherever the occurrence may arise, to provide short reports detailing the flooding that has been encountered. To keep the reports as short as possible, much of the report overhead previously required has been stripped away. For Flood Watch, users are only required to input the location of the flooding using the Google Maps input field and to choose a category that generally describes the severity of the flooding encountered.

Users may add additional information such as a more detailed description of the flooding; for example a description of hazards around the area. Finally, the user may submit an image, either from the phones gallery or by taking a new picture via the camera and appending it to the report. The user will not be required to title their report or input a formal location name, as these items will either be automated.

For the most basic implementation of the Flood Watch app, there are two major limitations to the original platform/mobile app: report querying based on location and age.

A. Location Based Report Query

When information has been collected, it must be then distributed to the users, and more specifically, the users who will benefit from the available information. As was mentioned in Section II, any report that is submitted via the mobile device must be approved by the administrator. However, the Flood Watch app will require that these reports are made available immediately to all users who will benefit from them. It is not beneficial for users in California to receive flood warnings from users in Virginia. There must be some type of filtering involved to prohibit users from receiving alerts outside the local area.

To compensate for the approval requirement, changes to the Ushahidi web platform must be implemented. This requires a dedicated server that will support the modified Ushahidi web platform. The web platform uses PHP to handle server requests and an SQL database to store all of the report information. When a request is sent from the mobile app to the server, an SQL query request is developed and passed to the database that returns a list of reports.

By default, the query simply requests for all of the most recently approved reports, up to a specific limit. The result is then a list of the most recent and approved reports with a maximum count governed by the limit. To retrieve all reports, the filter for approved reports had to be removed. However, a much larger problem remains. The reports being returned must be meaningful to the recipient. This is handled by filtering the query by location that may be implemented one of two ways: search provided latitude and longitude coordinates and a radius, or by a formal location name. Both have been implemented, and the choice of how to search is offered to the user in the preferences. A description of each implementation follows.

The default location query will be based on location name as it is easier to implement and requires zero calculations. As mentioned previously, most of the extraneous data from the reports generated on the mobile app was removed or

automated, for instance location name. Location name was one of the automated fields, as it plays an important role in the location search. When the app is loaded, and periodically while the app is active, the users' location is determined using either the mobile network or GPS provider, whichever is available. If a location cannot be found, the default location "Norfolk, VA" is used as the default position for the Flood Watch deployment is in Norfolk.

When the users' location may be found, the latitude/longitude coordinates are parsed using a Geocoder to determine the nearest city. This information is saved and used for reports generated by the user as well as for location querying. When reports are generated, the location name is automatically filled with city information in the following format: "City, State (Abbreviated) Zip Code"; for example, "Norfolk, VA 23505". Likewise, when querying reports, a search using location name passes the same information as a filter to the server. All reports that have matching location names are then returned.

When the user wishes to query based on a radius about their current position, similar information is required. By default, and unaltered, the reports generated contain latitude and longitude coordinates. When querying based on latitude/longitude position and a radius, this information must be sent to the server as a filter. The radius unit is set to miles by default but may be converted to kilometers, and the actual value may be updated in the preferences menu. Similar to querying by location name, the most recent latitude/longitude information is used as a filter.

Once this information is retrieved by the server it must be tested against all available reports. To determine the distance between two latitude/longitude coordinates, the Haversine Formula [11] is used as shown below:

$$a = sin^2\left(\frac{\Delta lat}{2}\right) + cos(lat_1) \cdot cos(lat_2) \cdot sin^2\left(\frac{\Delta lon}{2}\right) \quad (1)$$

$$c = 2 \cdot atan2\left(\sqrt{a}, \sqrt{(1-a)}\right) \quad (2)$$

$$distance = R_{EARTH} \cdot c \quad (3)$$

Here, '*a*' represents the square of half the chord length between the two points, and '*c*' represents the angular distance in radians [11]. With the resulting distance, it can be determined whether or not the report's location is within the boundary set by the radius. One final limitation must be addressed for the most basic implementation of the Flood Watch app, which is querying by report age.

B. Age Based Report Query

Filtering must be implemented for cases where a flood report has expired. Generally, flooding doesn't last for very long, a few hours to a couple days except in more extreme cases; however, in these extreme cases, this app would not be as beneficial due to mandatory evacuations. Therefore, at the discretion of the user, a filtering system must be implemented to compensate for expired reports and remove them from the alert system. When the user is requesting the most recent reports from the server, in addition to location information, the user is also providing the age preference for a report. The age is specified as a unit of days with a minimum value of 1.

When the server receives the age limit, all available reports are first compared to determine whether or not they meet the age requirement. Because the reports are in chronological order, once a report has failed the test, no other reports must be tested and a list is then generated of all appropriate reports. To test the reports, the date of the report is converted to a number ranging from 1-365 that represents the day in the year. Likewise, the current date is converted to day in the year format and the age is then subtracted. Finally, the reports date and the determined maximum age date are compared. If the report is newer or resulting on the same day as the maximum age date, it is kept. Special cases have been included to compensate for ages spanning to a previous year.

Once a list containing all appropriately aged reports has been composed, the location based query may begin. Especially in cases where the user prefers to perform the latitude/longitude test, the smaller report list will assist in reducing the number of individual tests that must be conducted. After all queries have been implemented, an XML file is created and returned to the mobile app that contains the requested report information.

C. Results

This section provides explanations and screen captures of several of the main features of the prototype Flood Watch app including the Flood Watch web platform and mobile app.

Figure 4 provides an example of the Flood Watch app running on the HTC Evo3D Android smartphone. The screen shown is the splash screen where the ODU logo and Ushahidi logo are displayed. This screen is very similar to the default Ushahidi splash screen as to provide credit to Ushahidi for providing the base functionality required for Flood Watch.

Figure 5 provides an extended screen capture of the "Add Report" screen present in the current Flood Watch app. The user may flick or slide the screen up or down to see this entire page. Additionally, an on screen keyboard comes up when the user clicks into a field such as Description. As mentioned previously, the user is required to select a category from the list of available categories. The user may also include a brief description of the occurrence as well as a photo that may be taken using the camera or referenced from one previously taken. If the user would like to edit the current position to be more accurate, the map is fully interactive. The user may zoom in/out and tap on the map to move the location of the marker. When the position is updated, the latitude/longitude fields automatically update. Likewise, the user may directly enter latitude/longitude

coordinates. When the user has completed the report, they may send it off for other users to see.

Figure 4: Splash Screen from physical device

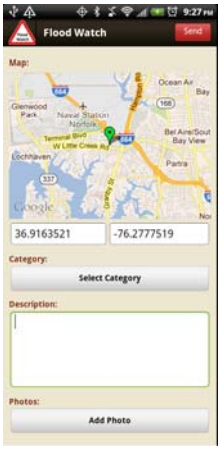


Figure 5: Extended "Add Report" screen

Figure 6 provides a screen capture of the "View Reports" page in list view. Although there are only 2 reports shown in the current screen capture, the list may contain as many as 500 unique reports. The total number of reports is user dependent, but the default length is 20 reports. Here the user is provided with a small version of the photo provided in the report (if any) as well as some information about the report. Current the date/time, city, and verification status are provided; however in future versions of the app, the verification status will be replaced by the severity of the flooding. Report verification is not currently supported by the Flood Watch app because there is not currently a method for verifying reports. It is assumed that the users submitting reports are providing correct information.

Figure 7 provides a screen capture of the "View Reports" page in map view. Here, reports are marked by icons hovering over the reports specific location. Users may zoom in/out of the map to view a much broader area of the map as well as tap on icons representing reports to see more detailed information about the report in question.

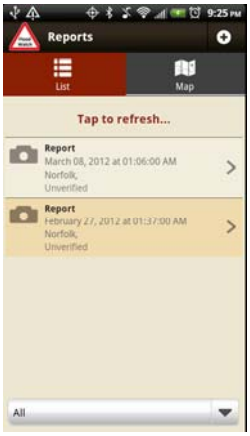


Figure 6: "View Reports" List View

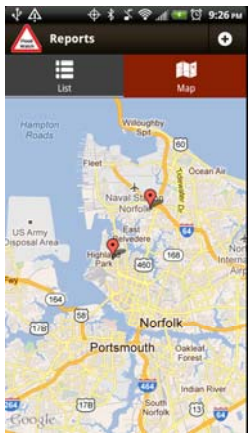


Figure 7: "View Reports" Map View

Figure 8 is the main page for the Flood Watch web platform. Here users may also view the most recent reporting of flooding provided by submitted reports via the mobile app or web. Additionally, users may also provide flooding reports via the web page using the "Submit Report" button. From the web, users may filter reports present on the map by category, a feature that currently is not available on the mobile device but may later be included.

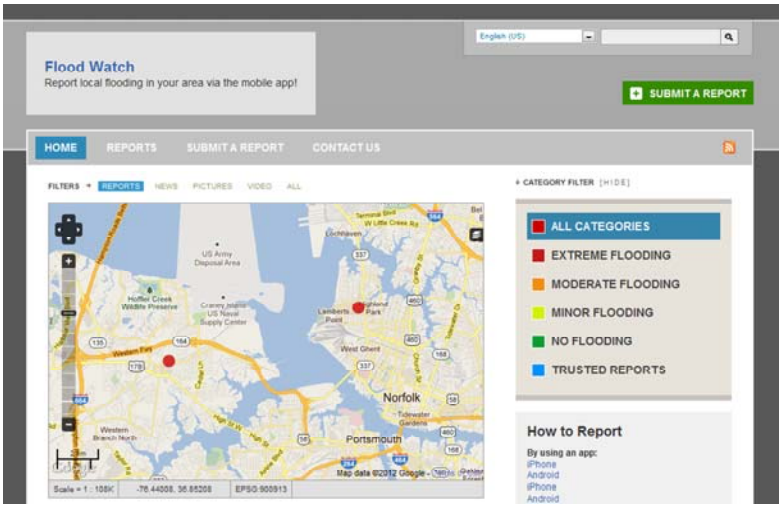


Figure 8: Flood Watch Web Platform

IV. Future Work

This section outlines some potential improvements that could be implemented to make the Flood Watch application stronger, as well as potential future work. The current implementation requires heavy use of the application by trustworthy users; that is, users who are willing to report flooding and only report flooding that exists. This app would have to heavily rely on police officers and news crews updating flooding information. While it's possible for common users to also provide input for recent flooding activity, most users would be expected to stay indoors. The app could benefit from retrieving weather alerts and updates from the National Weather Service. These general alerts, in addition to contributions made by the Flood Watch community would greatly benefit the common user by providing up-to-date flooding information. The automated alerts provided by the National Weather service could be parsed from the RSS feed. When relevant alert information is found, the user gets a notification icon on their screen alerting them to check the app.

Another improvement this app would benefit from is additional options for users with the location query. Currently, the users' location is automatically found before requesting reports. Including an option to input a location, either from selecting it from Google Maps, entering a latitude/longitude coordinate, or entering the city and state of the location they wish to search would be beneficial. This option would most commonly be used when users' are planning to travel and would like to have an idea of the conditions in a specific area.

Finally, some future work could be dedicated to the current user interface. Currently the app is the same as provided by Ushahidi; however, with the additions to querying options, it would be beneficial to offer users shortcuts to updating the query preferences as well as updating locations (if entered manually by the user as suggested above). Also, the color scheme currently used doesn't compliment the water theme present from the very nature of the app.

V. Conclusions

To conclude, Flood Watch is a mobile app for Android devices developed from the Ushahidi web and mobile platforms. It has been developed to provide users with another method in which they may receive flooding alerts and to better protect them against damages to their vehicles and to prevent a loss of life. The original platform has been modified to include certain functionalities not present in the original Ushahidi platforms including location queries and age queries that reduce the number of irrelevant reports the user must consider when being alerted of flooding in the area. Flood Watch, especially with some of the suggested improvements, is a valuable app for the Android Market.

References

[1] National Weather Service. (2010, March 1). *Summary of Natural Hazard Statistics for 2010 in the United States*. Available: http://www.nws.noaa.gov/om/hazstats/sum10.pdf

[2] National Weather Service. (2012, March 1). *Hydrologic Information Center - Flood Loss Data*. Available: http://www.nws.noaa.gov/hic/flood_stats/recent_individual_deaths.shtml

[3] National Weather Service. (2012, March 1). *National Weather Service Alert System*. Available: http://alerts.weather.gov/

[4] ITU World Telecommunications and ICT Indicators Database. (2011, March 1). *ICT Facts and Figures*. Available: http://www.itu.int/ITU-D/ict/facts/2011/material/ICTFactsFigures2011.pdf

[5] D. Forster. (2011, Storm surge from Hurricane Irene falls short of predictions. *PilotOnline*. Available: http://hamptonroads.com/2011/08/storm-surge-now-worry-irene-hampton-roads

[6] Ushahidi. (2012, March 1). *Ushahidi*. Available: http://ushahidi.com/

[7] Ushahidi. (2010). *The 2010 Earthquake in Haiti*. Available: http://haiti.ushahidi.com/

[8] Github. (2012, March, 1). *Ushahidi (Team Ushahidi)*. Available: https://github.com/ushahidi

[9] Ushahidi. (2012, March, 1). *Crowdmap*. Available: https://crowdmap.com/

[10] Ushahidi, "Ushahidi Screenshots of Android Market," ed, 2012.

[11] C. Veness. (2011, March 1). *Calculate distance, bearing and more between Latitude/Longitude points*. Available: http://www.movable-type.co.uk/scripts/latlong.html

Final Paper

Design of Video Overlay and Camera Pan/Tilt for UAV

Rahul Shambhuni
Modeling, Simulation & Visualization Engineering
Old Dominion University
Norfolk, United States

Yiannis Papelis, Ph.D., Menion Croll
Virginia Modeling, Analysis and Simulation Center
Old Dominion University
Norfolk, United States

Abstract: The idea of this paper is to develop a video overlay for the on board camera output on the base station display screen and work out for positioning on board camera, so that it always points to a given target location according to commands it received. This UAV designing is used as live model which is used in the Live-Virtual - Constructive (LVC) simulation of a command control system(C2).

I. INTRODUCTION

UAV's are capable of taking up volatile missions were especially human intervention is dangerous and expensive e.g. wide range reconnaissance, disaster relief support , military operations etc[2][3][4]. UAV's flies autonomously depending upon the navigation instruments such as Global Positioning System(GPS) and Inertial Measurements Units(IMU). Information of Absolute position of UAV can be determined by satellite using GPS and when IMU's are used the data is acquired from accelerometers and gyroscopes.

Using the outputs from the (GPS) and (IMU) camera needs to continuously point to a given target location. This output from this camera is wirelessly transmitted to base station ,the transmitted information is to be displayed on the monitor screen . These commands are given using microcontrollers. There is a special microcontroller chip called "propeller backpack" for displaying video along with overlay textual information in synchronous with incoming video. For commanding camera using servo motors two PWM pins of PIC 32 microcontroller is used. The video overlay information helps us in having a graphics overlay information such as Altitude, Battery life , tilting angles such as roll, pitch and yaw all help us controlling the mission to way point smoothly with more information at our view. Controlling of camera helps as to always keep track of our target irrespective of our heading direction.

II. MOTIVATION

Small fixed wing UAV's, rotary wing and ducted-fan designs are developed and demonstrated[5] in different scientific fields. But motivation for this paper has come from developing and testing a live-virtual-constructive simulation using a quad copter

live model to test it's result in the real world and using the information in virtual environment and develop a constructive simulation based on live and virtual situations. This project helps in fulfilling the gap between the virtual and real world using quad copter . In here the quad copter has been assigned a mission that it need to be complete it autonomously and return back to the base station, with video information of the enemy location.

III. METHODS

The below figure gives the overview of system architecture that is used in the quad copter. The outputs from the High Level CPU (PIC) and camera are transmitted to the video overlay CPU which processes the input signals and display the result on the monitor. Using the signals received pan/tilt of the camera is controlled by Pulse Width Modulation(PWM) signals. The GPS from serial port is read via High Level CPU and they are inputted to video overlay CPU. Camera output is also given into video overlay CPU . The GPS readings are inputted into PIC microcontroller and wireless receiver inputs are given to Propeller chip.

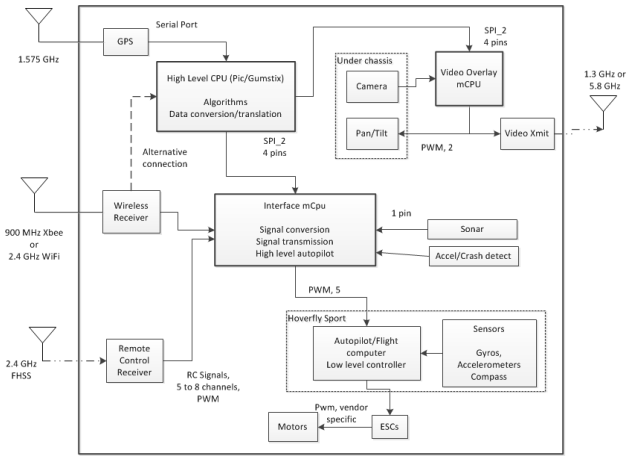


Figure 1:Quad copter System Architecture

In order to have a video overlay we used a video overlay CPU which is called propeller backpack(#28327) which has a 32pin propeller P8X32A micro controller. The programming for this CPU is done on a platform known as Propeller Tool using a SPIN Language specially designed for propeller CPU[1]. The output of camera is connected to video in of backpack and the video out to display. The display screen position is specified by X and Y coordinates of its upper left-hand corner. The maximum practical window size of the display that we are using is 40 columns by 13 rows. The commands that are used to display graphics on screen using the propeller CPU are DEFWIN(winno, cols, rows),where winno is window number from 1-63 ,which defines the window , APNDSP(disp,winno,x,y) will append window winno(0-63) to the end of display(1-63) at position (x,y) and finally USEWIN(winno) to make changes to current window. The characters that were used from the backpack CPU to have some graphics are binary arrays from $00 to $FF having symbols varying from alphabets, numbers and special characters.

In order to position camera continuously to target location using on board servo motors vector mathematical calculations are initially carried out. The final formulas derived are given below:

$$\alpha = \cos^{-1} \frac{(a-x)*(X-x) + (b-y)*(Y-y) + (c-z)*(Z-z)}{\mathrm{Sqrt}((a-x)2+(b-y)2+(c-z)2)\mathrm{Sqrt}((X-x)2+(Y-y)2+(Z-z)2)}$$

where α(radians) is the angle that camera should rotate in order to point to target location, P(x,y,z) is the location of quad copter , R(X,Y,Z) is the target location, Q(a,b,c) is point where camera is initially pointing .Similarly calculations are carried out considering the hovering state of the quad copter. Let that changing angle be β, then cos $\beta = \frac{Sqrt(X-x)2+(Y-y)2+(Z-z)2}{Sqrt(X-x)2+(Y-y)2+(z)2}$.

These mathematical calculations are to be given as inputs to UNO 32 chip kit which is programmed in

special chip kit MPIDE environment to produce PWM pulses to rotate servo to the required angles derived above.

IV. EXPERIMENTS AND RESULTS

The experiments and results all were done in the lab. For the video overlay using propeller chip, the display screen was divided into two portions, top left hand corner, and center of screen. The PIC microcontroller is programmed such as it is made to operate at 80MHz frequency by setting it to clock mode. Two objects have to be imported into program 'basic_sio.spin', and 'Parallax Serial Terminal.spin'. 'basic_sio.spin' sets the input output pins for use in propeller chip and 'Parallax Serial Terminal.spin' to serial out the commands to backpack video overlay chip. The top left hand corner was used to display Altitude and battery life of the Quad Copter. The center of screen shows us the angle the Quad Copter get tilted up or down in two dimensional space with respect to horizontal x- axis. The graphical display of tilt is always maintained in synchronous with the Quad Copter tilt. The direction of tilt is divide into segments of 3 degrees i.e., the angle from 30 to 33 degrees is shown as a 30 degree angle and the next angle from 34 degrees to 36 degrees is shown as 34 degree line. If the tilt is in anti clockwise direction with respect to horizontal x-axis it is taken as positive angle and if it is clock wise direction it is taken as negative angle. The clock speed of Propeller Chip is set such that the display on screen gets continuously updated without any lag. This information gets dynamically updated on the screen by reading the values from the PIC and Propeller Chip.

Figure 2: Video Overlay On Incoming video

For Camera positioning using servo motors, two servos are used, one for yaw and other for pitch. The servos used are Hextronik HXT 500 5GR Servo, it's operating voltage is 3V to 6V. Used PWM pins 6 and 9 on the PIC to rotate the camera mounted on the servo. Each angle input is given to the servo and a time of 15ms is given to it to rotate to given angle. Two servo objects have been created in PIC to control two servos and there are attached to pins 6 and 9 , and they are rotated according to logic that 'camera should be always trying to point to target location', by always keeping track of it's current location and heading using GPS. The camera is mounted on front of ground vehicle so all the trigonometric calculations are done considering this point.

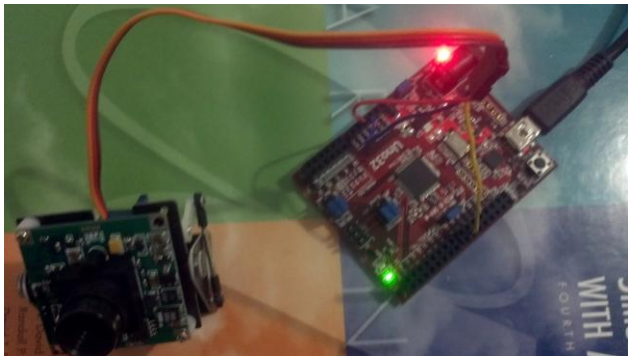


Figure 3:Controlling Servo motor using PIC

A experiment was performed on ground vehicle, outdoors, by giving it target mission in Northing and Easting values using Universal Transverse

Mercator (UTM) grid by keeping track of it's current heading and current location through PIC 32 micro controller. The steering values to motor wheels are given into Propeller Chip, whose calculations are similar to that of camera angle calculations, and according to these values it tracks target autonomously. Ground vehicle was given a single point and it generates a 4 way points on a circle of given radius and it successfully tracked them one after the other autonomously.

V. CONCLUSION

A wirelessly transmitted video can be captured on the display screen and text overlay can be made on any position of the screen in synchronous with the video using a single chip and no external circuit is required. The work on servo motor with aim for continuously looking towards a target even if the robot is headed to different point can be achieved which have very broad range of applications in military domain. This autonomous capability of UAV with vision camera controlled according to the mission commands and which communicates back to ground station are very helpful in LVC environment. Further work can be done on the video overlay by graphically replicating the compass on the incoming video to track the robot. The work on camera pan/tilt can also be applied in video games or graphical simulations where we keep track of user interacting with simulation or game.

VI. REFERENCES

[1] *Propeller Manual*, Version 1.2, Parallax Inc., Rocklin, CA, 2011.

[2] Office of the Secretary of Defense, USA, "Unmanned Aerial Vehicle (UAV) Roadmap 2005-2030," [R). August 2005.

[3] P. Campoy, J. F. Correa, I. Mondrag6n, C. Martinez, M. Olivares, L. Mejias, J. Artieda, "Computer vision onboard UAVs for civilian tasks," Journal of Intelligent and Robotic Systems. 2009, vol.54(1-3), pp.l05-135.

[4] Yu-chi LIU , Qiong-hai DAI,"A survey of Computer Vision Applied in Aerial Robotics Vehicle," International Conference on Optics, Photonics and Energy Engineering,2010.

[5] Samad, T, Bay, J.S, Godbole, D, "Network Centric Systems for Military Operations in Urban Terrain: The Role of UAVs", Proceedings of IEEE.2007,Vol.95,pp.92-107.

EMERGING TRENDS IN GEOSPATIAL INFORMATION VISUALIZATION FOR DECISION MAKING IN REGARD OF CHRONIC DESEASE MANAGEMENT

Ange-Lionel TOBA, Old Dominion University-Frank Batten College of Engineering & Technology, Department of Modeling, Simulation and Visualization Norfolk, VA
dtoba001@odu.du

Rafael Diaz, PhD, Research Professor, Old Dominion University, Old Dominion University-Virginia Modeling Analysis & Simulation Center Suffolk, VA
rdiaz@odu.edu

***Keywords*—Chronic conditions, System Dynamics, GIS, QALY, Information Visualization**

I. INTRODUCTION

The healthcare system in the US is subject to an enormous pressure, especially, from current and future demands of chronically-ill patients. Premiums are quite high and large amounts of money are being injected in this sector, with 75 percent allocated only to patients with chronic conditions [1]. In order to optimize the sustainability of the healthcare system, research in cost-utility analyses ought to be performed. 7 out of 10 of Americans die each year from chronic diseases [2]. A better management of this situation should be done in order to bring more efficiency and reduce the impacts of chronic diseases on people and the ambulatory system. *Diaz et al.* [3] provided a System Dynamics approach, with the objective of studying and analyzing the power of the interventions, not only from an economic point of view but also from a health condition standpoint.

II. PROBLEM/MOTIVATION

The focus of this paper is to present the process of integrating two simulation models that allows an analysis of the variation/improvement of the Quality Adjusted Life Year (QALY) of the patients under a pre- and a post-intervention condition. , and One approach to study this issue is location-based intervention, through GIS mapping visualization. The type of population, divided into group sex-based, color-based, age-based, as proposed by *Diaz et al.* [4], is used for this purpose. These authors analyzed the relationship between a given cohort of the US population and its ability to accede to ambulatory service is studied. The paper suggests using ArcGIS to incorporate location factors to focalize a potential intervention action for a given intervention. This study intends to provide an alternative to the problem of chronic disease management, by defining where to start the mitigation and response plan. The simulation and visualization of the output over time allows having

a better picture and a more accurate sense of the problem [5].

III. METHODS

This research extends the work of *Diaz et al.* [3] that use a simulation approach based on System Dynamics (SD) to measure and analyze the effects of generic interventions on the QALY and other public health metrics. The model presented in this research combines two simulation models that are designed in VENSIM, namely, (A) the categorization of the population, as of race, age and sex grouped, that was also used by *Diaz et al.* [4], and (B) the determination of the aspects related to the improvement of QALY and public health [3]. The association of these models helps in finding which population and location is more susceptible to the action of a given intervention that seek to improve the health status of a targeted population. A population cluster selected in (A) is considered as the effective fraction of population targeted in (B), so their particular health status, effect on the healthcare budget and system in a larger context can be examined. The simulations are executed for three regions in HR. Results are stored in a GIS-readable file. Each category is then separately analyzed. The simulation is executed after receiving an input from SABLE, which serves as an interface allowing entering the specifications and specific parameters of the case to study. The final input to be used in this case is Information Visualization (InfoVis), as it offers interactive and visual representation of data (previously obtained) to get a superior view of what the information being visualized means. InfoVis is a capable mechanism to communicate the findings to decision makers, as they can observe and understand how things unfold and make smart decision.

IV. CONCLUSION

The objective of using InfoVis is to improve information foraging, pinpoint trends, eventual anomalies, and enable the understanding of the observed data. In this research, the ability to accurately convey the scope of the situations is critical. This technique enables the user to evaluate the evolution of the health care status for certain population, and thus, user decision-making is improved. The model displays and lays out the inferences provided by the results, analyzing the overall healthcare system characteristics given a specific population chosen. Option is given to the user to compare scenario and visualize the differences in order to locate more accurately the underlying issue. Therefore, users may have graph outputs that assist in the decision-making allocation process as well as a visual geographical representation from GIS. This provides explanation about the shape and dynamics of the behaviors found.

REFERENCES

[1] Kung HC, Hoyert DL, Xu JQ, Murphy SL. *Deaths: Final data for 2005*. National Vital Statistics Reports, April 24th, 2008*; Vol 56 Number 10.* Available Online from: http://www.cdc.gov/nchs/data/nvsr/nvsr56/nvsr56_10.pdf

[2] (2009) Centers for Chronic Disease and Prevention-CDC: *The Power of Prevention Chronic disease, the public health challenge of the 21st century*. [Online]. Available: http://www.cdc.gov/chronicdisease/pdf/2009-Power-of-Prevention.pdf

[3] Diaz Rafael, Behr Joshua, Tulpule Mandar. "*Simulation-based optimization for improving chronic healthcare management interventions*". The 23rd European Modeling & Simulation Symposium. September 12-14, 2011. Rome, Italy.

[4] Mandar Tulpule, Rafael Diaz, "*Towards an Appropriate Population Model for the United States: A System Dynamics Approach"* Capstone, Old Dominion University. *Norfolk, VA Extended Abstract*

[5] Card, S. and Mackinlay, J. and Shneiderman, Feb 1999, *Readings in Information Visualization: Using Vision to Think*, p. 7.

.

Training and Education

VMASC Track Chair: Mr. Sol Sherfey
MSVE Track Chair: Dr. Roland Mielke

Recognition of Emotional Responses Using a FaceReader System
Author(s): Enilda J. Romero, and Dr. Ginger S. Watson

Cognitive Load in a Simulation for Mooring a Coast Guard Cutter
Author(s): Jennifer R. Morrison, Don G. Robinson, and Michael M. Martin

A Visual Aid as a Cue for the Detection of Critical Signals in Simulated Maternal-Fetal Heart Rate Patterns
Author(s): Rebecca A. Kennedy, Brittany L. Anderson-Montoya, Mark W. Scerbo, and Lee A. Belfore II

Mathematics in Biological Context: the use of Modeling, Analysis and Simulation to Enhance STEM education
Author(s): J. Eggleston, H. Gaff, and G.S. Watson

Recognition of Emotional Responses Using a FaceReader System

Enilda J. Romero and Dr. Ginger S. Watson

***Abstract*—The focus of this investigation was to determine the emotional responses of the participants as they interacted with an animated pedagogical agent in a web-based interactive environment using the FaceReader system. This inquiry sought to investigate whether the emotional behaviors of the animated agent would affect the emotional responses of the participants.**

***Index Terms*—Animated Agents, Emotional Responses, Interactive Environment.**

I. Introduction

THE FaceReader is modeling software used to detect facial expression information from the faces of the users. According to Den Uyl & van Kuilenburg (2005) the FaceReader has an accuracy of 89%. The system is based on Ekman and Friesen's Theory of the Facial Action Coding System (FACS), which states that the basic emotions correspond with specific facial models (Zaman & Shrimpton-Smith, 2006). The Face Reader system has been used as a reliable measuring tool in human-computer interaction studies (Terzis et al., 2010). However, there have been very few approaches for the purpose of affect recognition with regards to learning (Terzis et al., 2010). It is possible that the human ability to read emotions from one's facial expressions is the basis of facial affect processing that can lead to expanding interfaces with emotional communication and, in turn, to obtaining a more flexible, adaptable, and natural interaction between the learner and the elements of the interface (Den Uyl & van Kuilenburg, 2005). For the purpose of the research detailed in this abstract, the ability to read emotions from someone's faces would result in a more adaptable interaction between the learner and the animated agent.

Manuscript received March 9, 2012.

E. J. Romero is with the Instructional Design & Technology Program, STEM Education and Professional Studies Department, Old Dominion University, Norfolk, VA 23529 USA (e-mail: eromero@odu.edu).

G. S. Watson is with the Instructional Design & Technology Program, STEM Education and Professional Studies Department, Old Dominion University, Norfolk, VA 23529 USA. (e-mail: gswatson@odu.edu).

II. Methodology

A. Focus and Research Question

The focus of this investigation was to determine the emotional responses of the participants as they interacted with an animated pedagogical agent in a web-based interactive environment. This inquiry sought to investigate whether the emotional behaviors of the animated agent would affect the emotional responses of the participants. For example, would the participant express anger based on the angry emotional behaviors of the animated agent?

B. Participants

The participants were 66 paid volunteers who were undergraduate and graduate students in a public university in southeastern United States.

C. Equipment

The system consisted of the Noldus FaceReader software and a webcam that collected video feed while the participants interacted with the web-based interactive environment.

D. Treatments

This study used a between-subject design. Participants were randomly assigned to one of three experimental treatments: interactive environment without an agent, interactive environment with a non-expressive animated pedagogical agent, and interactive environment with an emotionally expressive animated pedagogical agent. The 66 participants were equally distributed among the three treatments (n = 22).

Data records from 13 participants were eliminated from statistical analysis. Calibration issues of the FaceReader software prevented the system from collecting adequate data from these participants. After the elimination of the 13 data records, the number of participants per treatment was as follows: 18 participants in the interactive environment without an agent condition, 18 in the interactive environment with a non-expressive animated pedagogical agent condition, and 17 in the interactive environment with an emotionally expressive animated pedagogical agent condition.

III. Results

To evaluate the relationship between the three interactive environments and the emotional responses of the participants, several ANOVA were conducted. The dependent variable was the emotional response data collected from each participant through the face reader text file. The statistical analysis of the mean and min response for each emotion indicated that there was no significant difference in the emotional responses between the interactive environments. However, the statistical analysis of the maximum response for each emotion indicated that there was a significant difference in the sad emotional response between the interactive environments, $F(2, 50) = 3.43$, $p > .05$. Specifically, participants in the interactive environment with an emotionally expressive animated pedagogical agent ($M = 0.4842$, $SD = 0.2610$) experienced more sadness than participants assigned to the interactive environment without an agent ($M = 0.2721$, $SD = 0.2438$).

In order to evaluate the emotional responses in the most neutral screens compared to the most emotional screens in the three interactive environments, separate ANOVA were conducted. The statistical analysis of the mean, min, and maximum response in the emotional screens and the mean and maximum response of the neutral screens indicated that there was no significant difference in the emotional responses between the interactive environments. On the other hand, the statistical analysis of the min response in the neutral screens indicated that there was a significant difference in the scared emotional response between the interactive environments, $F(2, 46) = 3.55$, $p > .05$.

TABLE I
MEAN AND STANDARD DEVIATION FOR EMOTIONAL RESPONSE DATA OF THE THREE INTERACTIVE ENVIRONMENTS

	Treatments					
	Without an agent		With an animated pedagogical agent		With an emotionally expressive animated pedagogical agent	
	M	**SD**	**M**	**SD**	**M**	**SD**
Sad - Max (all screens)	0.2721	0.2438	0.4317	0.2476	0.4842	0.2610
Scared - Min (neutral screens)	0.0106	0.0198	0.0014	0.0048	0.0007	0.0013

IV. Discussion

The present study examined how emotionally expressive animated pedagogical agents in a web-based interactive environment influenced learners' emotional response. As expected and as suggested by Heckman and Wobbrock (2000), learners interacting with the emotionally expressive agent displayed more facially expressed sadness and fearfulness than learners without an agent. In these treatments, there was a strong connection between the learners and the agent. This connection could be due, in part, to the level of credibility and believability of the animated agent. It is possible that the learners were able to identify themselves with the animated agent presented and therefore, were able to create an emotional connection to the agent.

In conclusion, the present study presented relevant results on the use of animated pedagogical agents in a web-based interactive environment. It also demonstrated that the use of physiological measures such as the emotional FaceReader system can produce more specific insights on the perceptual experiences that occur as learners interact with an animated agent.

References

[1] Den Uyl, M. J., & van Kuilenburg, H. (2005) The facereader: Online facial expression recognition. In *Proceedings of MB 2005*, 589-590.

[2] Zaman, B., & Shrimpton-Smith, T. (2006) The FaceReader: Measuring instant fun of use. *NordiCHI 2006*, 457-460.

[3] Terzis, V., Moridis, C.N., & Economides, A.A. (2010). Measuring instant emotions during a self-assessment test: The use of FaceReader. *Proceedings of Measuring Behavior 2010 7th International Conference on Methods and Techniques in Behavioral Research.* Eindhoven, The Netherlands.

[4] Heckman, C.E., & Wobbrock, J.O. (2000). Put your best face forward: Anthropomorphic agents, e-commerce, e-commerce consumers, and the law. *Proceedings 4th International Conference on Autonomous Agents*. Utrecht, Netherlands.

Cognitive Load in a Simulation for Mooring a Coast Guard Cutter

Jennifer R. Morrison, Don G. Robison, and Michael M. Martin

***Abstract*—Mooring a cutter for the Coast Guard is a complex, multivariate task. The current training process includes classroom instruction followed by limited use of a single, expensive, and remote simulator. The purpose of this case study is to document the design and development of an instructional simulation that will bridge the gap between classroom instruction and actual practice while managing the limits of working memory.**

***Index Terms*—Cognitive load, instructional simulation, design case study**

I. Introduction

The challenge of bringing a ship safely in to a pier is essentially a physics problem with multiple variables. The ship possesses physical properties (e.g., mass or hull resistance to water) that interact with force vectors (e.g., wind force, current force, engine thrust, and rudder angle). There are other distracters operating during the actual task performance, but successfully mooring a ship boils down to effective management of the physics problem.

A. The Current Training Process

The current training process for novice shiphandlers begins with classroom instruction that addresses the principles of ship handling. These principles include the importance of preparing beforehand for a mooring, managing engine and rudder vectors, managing wind and water current vectors, and also managing line handling. In addition to the classroom instruction, both a two dimensional and full bridge simulator are used. The full bridge simulator is optimized for training deck watch officers in cruising scenarios but it is not optimized to facilitate mooring training. The two dimensional simulator is effective, but it only allows for a few users at a time and is not accessible outside of scheduled class times.

Manuscript received March 9, 2012.

J. R. Morrison is a PhD student in the Instructional Design and Technology Program in the Darden College of Education, Old Dominion University, Norfolk, VA 23529 USA (email: jmorr065@odu.edu)

D. G. Robison is a PhD student in the Instructional Design and Technology Program in the Darden College of Education, Old Dominion University, Norfolk, VA 23529 USA (email:drobi036@odu.edu)

M. M. Martin is a PhD student in the Simulation and Gaming Program in the Batten College of Engineering, Old Dominion University, Norfolk, VA 23529 USA (email:mmart081@odu.edu)

Following the classroom training and limited time in the two dimensional simulator, cadets practice mooring using 50 foot long training boats. The ship-handling training process continues aboard the student's first cutter where junior ship handlers practice mooring the ship in what are affectionately called "dock-crashing drills." Cadets currently report to their first cutters with a good foundation of background knowledge, concepts and principles, but have limited opportunities to apply what they have learned. There exists a need to provide learners an intermediate step between classroom instruction and time spent practicing within the simulator and on actual ships.

II. Instructional Simulation

The purpose of this design case is to describe the process of creating an instructional simulation to bridge the gap between learning about the facts, concepts, and principles involved in mooring a Coast Guard cutter and the actual practice of mooring a Coast Guard cutter. Since mooring a ship is a complex task that requires extensive practice, learners could benefit by having unlimited access to a practice simulation. The design solution to this issue was to a PC based instructional simulation that allows learners to download and run the simulation on their personal computers so that they may use the simulation any time they desire, thus eliminating the access limitation of the simulator.

Because of the complex, multivariate nature of this task performance, the dominant design challenge was managing cognitive load. Cognitive load theory is concerned with the limits of working memory and the resulting instructional message design factors that influence learning. Cognitive load theory is comprised of three components: intrinsic cognitive load, extraneous cognitive load, and germane cognitive load [1]. Intrinsic cognitive load (ICL) describes the nature of the instructional material, specifically the structure of the to be learned content. When individual elements must be simultaneously held in working memory for a learner to acquire the content and achieve learning goals, the material is said to have high element interactivity and a high level of intrinsic cognitive load. For example, learning the letters of the alphabet has low element interactivity as the individual items may be learned in isolation. In the case of learning how to moor a ship, however, the learner must hold numerous elements in working memory to successfully moor, such as environmental hazards, ship characteristics, and port attributes.

The second component of cognitive load theory, extraneous cognitive load (ECL), is concerned with the design of

instructional materials or the learning activities presented to the learner [2]. Strategies for reducing ECL, such as incorporating worked examples in mathematics instruction, integrating separate sources of information into a single source, and presenting information in dual modalities, have been the subject of research [2-4].

The third component of cognitive load theory is germane cognitive load (GCL). This final component refers to the remaining working memory capacity for schema development [1]. When intrinsic cognitive load is high due to high element interactivity, and ECL is high due to poor message design, learners may not possess enough working memory capacity for schema development (i.e., germane cognitive load). An important task for the design of instructional materials is to account for ICL, effectively structure and present content to reduce ECL, which allows for enough GCL for the learner to achieve the intended instructional goals.

The following is a description of the process of designing supplementary instruction and an accessible web-based simulation for mooring practice while incorporating instructional design principles for managing cognitive load. The instructional design process follows the steps described by Morrison, Ross, Kalman, and Kemp [5].

III. The Design Process

A. The Design Team

The concept for this simulation was birthed during a class on instructional simulations. The design and development team consisted of three Ph.D. students at Old Dominion University, led by the class instructor, who has extensive experience in instructional modeling and simulation. The design and development team included:

- An Instructional Design Ph.D. student with a research interest in cognitive load and the use of representations in instructional materials.
- An Instructional Design Ph.D. student with over 20 years of practical experience in instructional design. He is a retired Coast Guard officer and served as the project SME, as an instructional designer, as graphics lead, and as Flash and website developer.
- A Simulation and Gaming Ph.D. student with skills in game development, specifically, programming in XNA. He served as the technical lead and developer.

B. Procedure

In keeping with the overall nature of the project, the development procedure utilized an informal, yet highly iterative management process. In many regards, the project followed an informal application of recent project development philosophies, such as Agile [6] or Scrum [7]. With rare exceptions, the project members met weekly to discuss the current project status, near term goals, and any alterations to ongoing previous goals. In effect, the project team created informal "sprint" cycles to further the project's development. The process proved effective at accommodating the shifting priorities of work required by external demands upon the project members, while at the same time maintaining a clear focus on iterative improvements towards defined near and long term goals.

To facilitate project collaboration, the team explored a number of software file sharing solutions. It became quickly apparent that simple file sharing using e-mail attachments would become cumbersome and difficult to maintain. Before the course project was complete, the team explored using cloud based shared storage to share and preserve critical project files. This approach took the form of Microsoft's SkyDrive™. Shortly after the course was over, persistent challenges with accessibility and ease of use with SkyDrive™ led the team to try a cloud-based file synchronization software solution called DropBox™. DropBox offers the same core features as SkyDrive™, but has the added notable addition that the files being shared are automatically synchronized to each team members' local computer storage. Once the software was installed and set up, the files updates and shared occur in near real-time, with no effort required on the part of either the sender or receiver. This solution greatly increased the efficiency by which the project team could share important work and synchronize their efforts.

C. Problem Identification

While cadets receive classroom instruction, they do not have sufficient opportunity to apply those principles in dynamic or realistic contexts. They needed an opportunity to apply the principles they learned between the classroom experience and actually mooring a cutter. The problem is amplified during a cadet's first actual mooring experience. The cadet is forced to perform at a level that far exceeds any prior experience, and the ship's Commanding Officer must trust a multiple-million dollar vessel to a novice. A simulation with a presentation that reduces intrinsic cognitive load and would be available to cadets on their laptops appeared to be an optimal solution to the problem, offering a learning experience that bridges the gap between the classroom and actual ships.

D. Learner Analysis and Prerequisite Skill Identification

The design process began with a learner analysis to identify entry-level skills and prerequisite knowledge for the target audience. Coast Guard Academy cadets are high-achieving students who have demonstrated conspicuous academic success prior to obtaining an appointment to the Academy. However, most of the cadets, regardless of their background, will have had little experience with ships. The design solution assumes that the cadet users of the simulation will have had the basic instruction in ship-handling described earlier. They are familiar with the facts, concepts, and principles involved in mooring Coast Guard cutters, but have had little opportunity for practice and lack skill in the application of this knowledge.

E. Task Analysis

The team used subject matter experts (SME's) to identify performance-related variables and important principles. During the initial stages of design, one of the team members, a retired Coast Guard officer, served as the SME. In analyzing the mooring process, the design team identified three distinct categories of simulation variables: environmental factors, ship

characteristics, and port attributes. These categories shaped the instructional objectives and guided the simulation design.

Environmental factors account for variables such as the speed and direction of both wind and current, which have a prominent impact on the mooring process. Accounting for these factors constitutes a major instructional objective. This simulation allows the user to train on a variety of ships, each with their own movement and performance characteristics, such as speed, acceleration and deceleration, and turning radii, which in turn interact with the navigational variables. Finally, the port attributes constitute another important set of variables. These attributes include pier characteristics, navigational hazards, water depths, and navigational aids. Table 1 provides an overview of the relevant variables identified in the task analysis.

TABLE I
SIMULATION VARIABLES

Ship Variables	
	Sailing characteristics of the cutter (tactical data)
	Speed controls (Engine thrust)
	Directional controls (Rudder effects)
Environmental Variables	
	Effects of water current
	Effects of wind speed
Port Variables	
	Navigation hazards
	Aids to navigation
	Water depths
	Pier characteristics
	Wind and current tell-tales

Relevant ship-handling variables for mooring.

After task-specific relevant variables were identified, SME's were queried regarding relevant principles for ship-handling and cues for determining what actions are indicated in different situations. Principles for the following concepts and practices were developed for inclusion in supplementary instructional materials:

- Relative impact of wind and water current
- Approaching the pier with off-dock forces
- Approaching the pier with on-dock forces
- Twisting ship
- The effects of a short forward thrust on one engine
- The effects of a short reverse thrust on one engine
- The effects of using full rudder with engine bursts
- Identifying on-scene indicators of wind and current effects

A primary objective was developed for the simulation based on information gleaned from the task analysis: Given various natural mooring conditions, the learner will employ the appropriate commands to properly moor a Coast Guard cutter.

IV. INITIAL DESIGN

A. Instructional Materials

The initial purpose for the design of the instructional materials was to provide novice shiphandlers optional, supplementary content to reinforce concepts and principles acquired during classroom instruction. The instructional design team identified the following important principles: wind and current effects, controllable and uncontrollable forces, momentum alongside, and the angle of approach. The principle refresher section could be accessed from the home screen and learners were presented with either static graphics or animations paired with text describing the specific principle. For example, animations were designed to demonstrate the differing effects of wind and current on a ship and reinforced content presented in the text. The initial design of the instructional materials did not incorporate a level of interactivity for learners as the content was merely intended to provide a refresher on important principles.

B. Instructional Simulation

One of the key early design decisions was to manage intrinsic cognitive load by limiting task performance to a part-task trainer [8]. The simulation as designed presents learners with the cognitive challenge of vector management, and eliminated variables such as managing shipping contacts or managing line handling. The learner is focused only on the difficult task of vector management. Additionally, intrinsic cognitive load was reduced by the introduction of a simplified three step paradigm for the task, making it easier for novice ship-handlers to assimilate each step.

The simulation provides more difficult scenarios to advanced learners based on the learner's self-description. The novice simulation presents a simplified environment [1]. Extraneous materials have been removed, and the wind and current challenges are modest. Novices also follow a different pathway into the actual simulation. First, novices are presented with a model of the performance that has been simplified into setting up for mooring, the approach, and maneuvering alongside the pier. This chunking of tasks should help the novice learner who is developing a paradigm for this performance [9].

More advanced ship-handlers may choose a path that takes them directly to the simulation. To reinforce the chunked paradigm, aspects of the three step paradigm presented to novices are included as part of the simulation. For example, if an advanced ship-handler elects to go directly to the simulation, the first step is a port-brief that forecasts wind and current at the time of mooring. This forces simulation users to perform important elements of the 'setting up' step.

The design allows users starting a simulation to select difficulty level. The user may choose between "easy," " moderate," "difficult," and "custom" mooring challenges. The more advanced the selection, the wider the possible range of wind or water current force. The "custom" selection allows users to enter the direction and force of both wind and water current. In this way, the design allows novices to experience a simplified simulation with appropriate challenge, while allowing advanced ship-handlers rapid access to more challenging simulation scenarios.

C. Simulation Fidelity

The initial design of the simulation was heavily influenced by Alessi's [10] discussion of moderating simulation fidelity to achieve the optimal effect upon the learner. The target learner

for VectorMoor is a novice ship handler, who was a small amount of formal education on the learning content, but will not have had much, if any, practical education. For novice learners, Alessi recommends reducing the level of fidelity [10].

VectorMoor moderates it's fidelity along all four of the fidelity aspects described by Alessi: the underlying model, presentations, user actions, and system feedback. The underlying model is not a full feature hyrdodymanics simulation, but rather a simple rules-based simulation, using simple second-order algebraic equations to model the movement of the ship. The model is designed to reinforce specific learning goals, rather to strictly represent hydrodynamic interactions. The design also chooses to present the ship to the user in the two dimensional, overhear, third person perspective, instead of a more realistic first person perspective. Again, this is intended to allow the user to focus specifically on the learning objectives, such as visualizing the vector effects of forces on the ship. The user's actions are similarly simplified, allowing the user to focus more on what he wants the ship to do, rather than how to do it. Lastly, the feedback system provides instantaneous feedback to the user, letting them know, in a realistic manner, about nearby hazards, and providing negative feedback in the event of a crash. The design also calls for the implementation of a "trackline" notification system, which would provide the player with constant feedback on their performance with regards to navigational planning, but this part has not yet been implemented in the program.

VectorMoor also seeks to manage the cognitive load it creates by using the concept of signaling, as described by Mayer and Moreno [11]. In the simulation, VectorMoor incorporates visual cues to focus the user's attention on key environmental information. This process, called signaling by Mayer and Moreno [11], provides information to the user on how to process the material presented in the simulation. VectorMoor uses a combination of diegetic and non-diegetic cues to helps users reduce processing of extraneous material, and instead focus on the important information. The user is cued to the wind direction via wind sock which are present in different areas of the simulation world. Similarly, waves upon the water cue the user to the current conditions. Lastly, the simulation provides a non-diegetic highlighted square around the designated birthing area to focus the user on their ultimate mooring goal.

D. Feedback

Feedback on a shiphandler's performance during the simulation was initially designed as a text message to the user stating whether they completed a successful or unsuccessful mooring.

V. Formative Evaluation

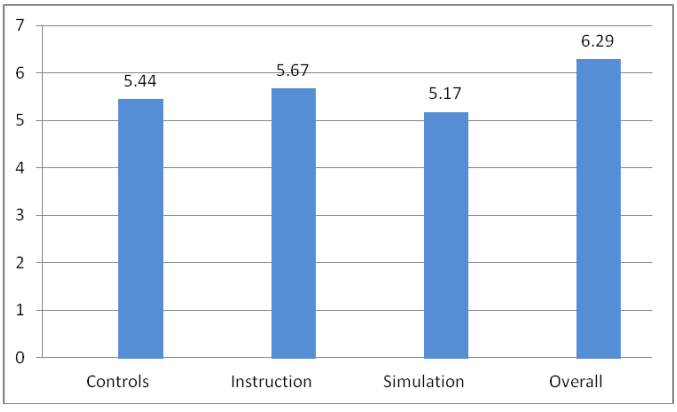
The initial phase of formative evaluation involved an expert review by an instructional designer on the instructional simulation. The first suggested revisions included adjusting the mode of feedback within the simulation, which occurred when the learner came close to environmental hazards or unsuccessfully moored the ship. Rather than visual text feedback, auditory feedback was included that more accurately reflected the form of feedback shiphandlers would experience on a real ship. For example, when the shiphandler maneuvered the ship too close to an environmental hazard, a recorded warning from the navigation team would play. Additionally, if the shiphandler crashed the ship or maneuvered the ship at a speed greater than 1 knot next to the pier, the general quarters alarm would play and the screen shook. The second suggested revision involved the fidelity of objects represented. Specifically, as the user approached a bridge to pass under, it was unclear where pylons were positioned although one knew they existed. This was a fairly minor change but led to revisions to several objects represented to improve fidelity within the simulation.

The second phase of formative evaluation involved pilot testing with expert shiphandlers and a ship captain. A think-aloud protocol with individual users, where their thoughts and comments were verbalized and documented, was conducted as they interacted with the simulation. Feedback included fidelity of the simulation, fidelity of the objects represented, and adjustment to the target birth area. Shiphandlers commented that the ship was turning slower than it would in real life and was also traveling faster than what was indicated on screen. Second, it was suggested that the shoal water should be indicated on the screen. In real life, shiphandlers would rely on a navigation team to aid in avoiding shoal water. Within the simulation, this support should be provided as there is not a navigation team to assist. Finally, the shiphandlers suggested that the target birth area be larger since positioning of a real ship would vary based on environmental conditions.

A small group evaluation on the instructional content was also conducted as a think-aloud. The few revisions suggested on the instructional material included adjusting verbage to the technical terms used by shiphandlers, indicating wind and current conditions during all animations, and slight changes regarding the fidelity of the animations.

Shiphandlers were also asked to complete a usability survey, which consisted of 16 questions on a 7 point Likert-type scale (0 ="strongly disagree"; 7 = "strongly agree"). Questions pertained to the controls, instructional content, simulation, and general questions.

TABLE II
USABILITY SURVEY RESULTS



In general, shiphandlers reported they felt the simulation and instructional content was a very valuable tool to bridge the gap between classroom instruction and time spent with the full simulator and practice on an actual ship. Additionally, one participant stated that the use of the VectorMoor simulation could make time in the simulator more valuable since novice learners would have more experience applying classroom instruction to the shiphandling procedure. The shiphandlers commented that they enjoyed the challenge presented and were motivated to continue use until they successfully moored the ship in the simulation.

VI. Revisions

Several revisions were made to the supplementary instructional materials and the simulation based on results from the formative evaluation.

A. Instructional Materials

The decision was made to further develop the instructional materials as stand-alone pieces rather than supplementary to classroom instruction. Since the design of the novice simulation presented the mooring process in three steps, the instructional content was merged with the novice simulation. Principles were divided based on the applicability to one of the three mooring stages and were presented as "just in time" information [12]. The RUL-EG instructional approach was incorporated for each principle, as described by Morrison, Ross, Kalman, and Kemp [5]. Specifically, a demonstration of the principle was presented, followed by an opportunity for active practice. Each principle was first presented as a narrated animation, allowing for offloading from the visual to the auditory channel and reduced extraneous cognitive load [13]. The incorporation of narrated animations rather than text paired with animations or static graphics also reduced any potential split attention effects caused by learners having to integrate disparate sources of information [4]. The practice activities also allowed the learner to practice relevant principles within each stage of the mooring process and should facilitate assimilation of steps [9].

B. Simulation

During the extended development of the simulation following the completion of the course, the simulation underwent several major revisions. The most notable of these changes was the inclusion of realistic terrain. Figure 1 shows the initial state of simulation. Terrain in this version consisted of simple rectangular geometry, which bore little resemblance to actual locations that the user might train to moor at. Figure 2 shows the current state of the simulation. Most notably, this version of the simulation includes realistic terrain modeled on the waterways surrounding the United States Coast Guard Academy. Additionally, a number of smaller aesthetic changes have been undertaken to make the presentation of user interface more consistent with the diegetic narrative of the simulation. For example, the presentation of the ship's heading and speed have been altered to resemble displays which might be found on a ship.

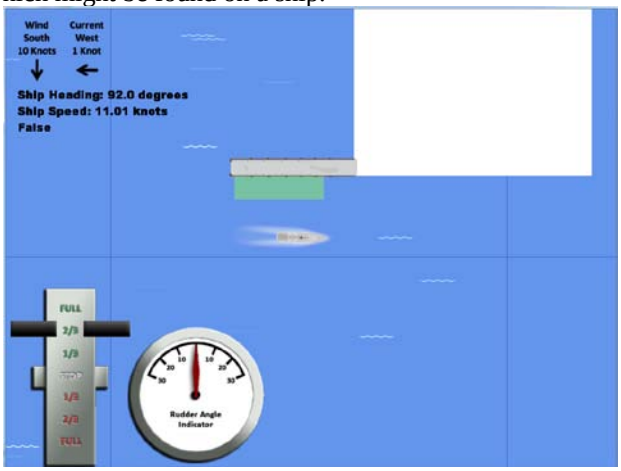


Fig. 1. Initial simulation design

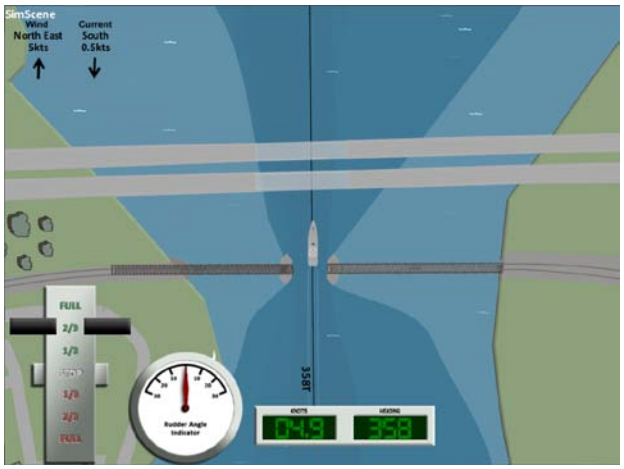


Fig. 2. Current simulation design

Revisions have also been undertaken to increase the instructional efficacy of the simulation, specifically with regards to the use of scaffolding to help manage the cognitive load upon the user [12].A primary example of this is inclusion of the trackline to help novice users orient themselves as they approach the pier. The trackline is shown in figure 2 as a black line in the water, with a heading reading next to it (358T). The guide is something which a shiphandler would have available on a navigational chart, but which, of course, would not be actually present in the water. This scaffolding is available for novice users, and can be turned off for more experienced users. Another example of scaffolding is the inclusion of a mooring approach selection screen, as shown in figure 3. This screen prompts the user to plan out their approach to the birthing area based on the present environmental conditions, and gives the user feedback to help them adjust as needed.

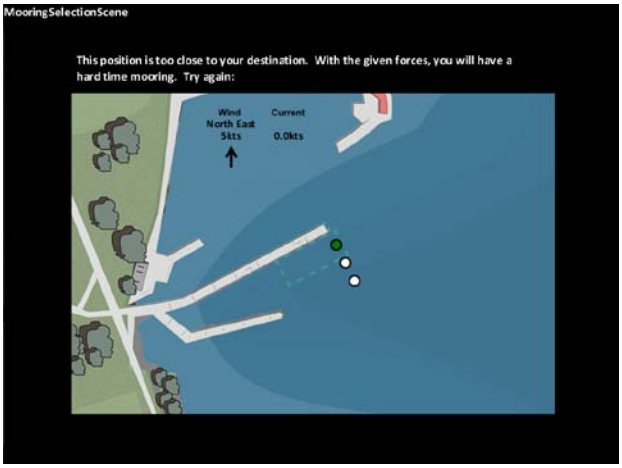


Fig. 3. Mooring approach selection screen

Lastly, there have been a number of other additions to the simulation designed to increase the utility of the simulation, and in an effort, as described in the section on Development to make the simulation a complete instructional package. These additions include the ability to customize the "difficulty" level of the mooring by altering the wind and current in order to tailor the challenge level for each individual user, as well as the instruction and practice on using the simulation controls.

All of these revisions and incremental improvements are undertaken to reach the design goal for the simulation. In turn, this design goal itself is undergoing a constant review and iteration based on feedback from subject matter experts and quality assurance checks by the project team members. The agile design philosophy employed by the project team allows these sorts of revision to be easily integrated in to the development process, while ensuring that the project is continually progressing towards its defined goals.

Time trials were also conducted to verify that the speed readout for the cutter was accurate. A straight speed trial course was laid out from the bridge clearance north exactly .5 miles. The simulation cutter was run at a speed of either 5 or 7 knots according to the readout and then timed with a stopwatch accurate to the hundredth of a second. In our initial trials, the simulation cutter actually operated at about six times the posted speed. Modifications to the model were made to correct this discrepancy.

VII. Future Work

Future work on the instructional simulation includes the development of additional harbors and cutters. A summative evaluation process, which will explore the effectiveness and efficiency of the instructional materials and simulation, will also occur. Finally, the simulation may be used in future research examining visual aesthetics on motivation and performance, as well as factors affecting the fidelity of the environment on performance.

References

[1] J. Sweller, *Instructional Design in Technical Areas*. Camberwell, Victoria: The Australian Council for Educational Research Ltd., 1999.

[2] J. Sweller, *et al.*, *Cognitive Load Theory*. New York, NY: Springer, 2011.

[3] J. Sweller, *et al.*, "Cognitive architecture and instructional design," *Educational Psychology Review*, vol. 10, pp. 251-296, 1998.

[4] R. E. Mayer, *Multi-media Learning*, 2nd ed. Cambridge, MA: Cambridge University Press, 2009.

[5] G. R. Morrison, *et al.*, *Designing Effective Instruction*, 6th ed. Hoboken, NJ: John Wiley & Sons, Inc., 2011.

[6] P. DeGrace and L. H. Stahl, *Wicked problems, righteous solutions: A catalogue of modern software engineering paradigms*. Englewood Cliffs, NJ: Yourdon Press, 1990.

[7] C. Larman, *Agile and iterative development: A manager's guide*. Boston: Addison-Wesley, 2004.

[8] J. J. G. van Merrienboer and J. Sweller, "Cognitive load theory and complex learning: Recent developments and future directions," *Educational Psychology Review*, vol. 17, pp. 147-177, 2005.

[9] E. Pollock, *et al.*, "Assimilating complex information," *Learning and Instruction*, vol. 12, pp. 61-86, 2002.

[10] S. Alessi, "Fidelity in the design of instructional simulations.," *Journal of Computer-based Instruction*, vol. 15, pp. 40-47, 1988.

[11] R. E. Mayer and R. Moreno, "Nine ways to reduce cognitive load in multimedia learning," *Educational Psychologist*, vol. 38, pp. 43-52, Winter2003 2003.

[12] J. J. G. van Merrienboer, *et al.*, "Taking the load off a learner's mind: Instructional design for complex learning," *Educational Psychologist*, vol. 38, pp. 5-13, Winter2003 2003.

[13] S. Y. Mousavi, *et al.*, "Reducing cognitive load by mixing auditory and visual presentation modes," *Journal of Educational Psychology*, vol. 87, pp. 319-334, 1999.

A Visual Aid as a Cue for the Detection of Critical Signals in Simulated Maternal-Fetal Heart Rate Patterns

Rebecca A. Kennedy, Brittany L. Anderson-Montoya, Mark W. Scerbo, Lee A. Belfore II
Old Dominion University
Alfred Z. Abuhamad, Suneet P. Chauhan, Stephen S. Davis
Eastern Virginia Medical School

***Abstract*—The present study used static images from simulated maternal-fetal heart tracings to determine whether a visual aid would serve as a cue to improve detection performance of critical signals embedded in the fetal heart rate. Twenty-one undergraduates twice viewed 80 images, with and without the presence of a visual aid. The visual aid was a turquoise crosshair cursor that was overlaid on the images in a predetermined location meant to facilitate detection. Proportions of correct responses were analyzed according to four different signal-to-noise (S/N) ratios and five different delays (in sec) between the peak of the contraction and the onset of the signal in fetal heart rate. The results indicated that lower S/N ratios and shorter onset delays made signals more difficult to identify. However, overall levels of accuracy were significantly higher for conditions in which the visual aid was present. These findings provide initial evidence that utilizing visual aids may enhance the ability of clinicians to detect critical signals in maternal-fetal heart rate patterns.**

***Index Terms*—maternal-fetal heart rate, visual aid, cueing, signal detection**

I. Introduction

A. Maternal-Fetal Heart Rate

During labor, the wellbeing of the fetus is primarily assessed using electronic fetal heart monitoring [1]. An elastic belt with transducers is placed around the woman's abdomen [2]. Contractions (measured in mmHG), and fetal heart rate (measured in beats per minute; bpm) are presented either via a paper strip or a computer screen [1, 3]. Clinicians assess fetal heart rate for critical patterns that may indicate fetal distress and necessitate intervention.

Unlike many other physiological monitoring systems currently used in healthcare, fetal heart rate activity is predominantly presented in analog format. This analog presentation introduces a source of subjectivity into the interpretation of tracings [2, 3]. Indeed, a number of studies have found disagreement among individuals over interpretation of fetal heart rate signals, and even disagreement when the same individual interprets the same tracing at two different points in time [1, 4, 5].

In 2008, the National Institute of Child Health and Human Development (NICHD) proposed a three-tiered system for interpreting fetal-heart rate patterns [6]. Patterns falling under the Category I designation are considered normal and reassuring, whereas Category II patterns are considered indeterminate, and Category III patterns are considered abnormal. Category III patterns require quick evaluation of threat to fetus wellbeing.

One such Category III pattern that is often indicative for the need for intervention is a late deceleration, or late decel [3, 6]. A late decel is a decrease in the fetal-heart rate from baseline that occurs after the peak of a contraction. This pattern is very similar to an early decel, which is a decrease in fetal-heart rate that is coincident with the contraction, often making it difficult to identify. Whereas late decels are considered an ominous pattern, early decels are considered a Category I, reassuring pattern. It is therefore imperative that clinicians correctly distinguish between late and early decels.

Research findings have suggested that distinguishing between early and late decels is problematic. Figueras, Albela, Bonino, Palacio, Barrau, et al. [7] found lack of agreement concerning the number of late decels present in fetal heart rate tracings. Todros, Plazzotta, Biolcati, and Lombardo [8] found that individuals were able to detect decelerations in fetal heart rate, but there was disagreement about whether the decels were early or late.

Recently, Anderson and colleagues assessed how well individuals could differentiate between early and late decels under different levels of fetal heart rate variability and with the late decels occurring at differing time frames. In the first study, Anderson, Scerbo, Belfore, and Abuhamad [9] asked untrained undergraduate students to view static images of decels and respond as to whether they thought the decel was early or late. The images were generated using a maternal-fetal heart rate simulator created by Belfore, Scerbo, and Anderson [10]. The results of the study indicated that the participants had trouble

differentiating between the two types of decels, especially at the 4- and 12-sec delays.

In a second study, Anderson, Scerbo, Belfore, Abuhamad, and Davis [11] replicated the static images study with trained labor and delivery providers. The results mirrored those of the untrained undergraduates. That is, even trained labor and delivery clinicians had difficulty distinguishing between early and late decels. Further, Anderson and colleagues [12] have also examined how well individuals could detect decels under different signal-to-noise ratios (absent, minimal, moderate, and marked) and found that as the signal-to-noise (S/N) ratio decreased, individuals made fewer correct detections.

The results from previous studies suggest the need for a visual inspection aid that can be used to highlight critical patterns in fetal heart rate signals.

B. Cueing

Research has demonstrated that cues can be an effective method for improving the detection of targets and signals. Cues have been successfully implemented in a wide variety of applications, including industrial inspection tasks [13] and vigilance tasks [14]. Cues may function to direct attention toward targets, particularly when they are reliable [15]. When attention is appropriately focused, accuracy and response times may therefore increase [16].

In one study, Chaney and Teel [17] compared the effectiveness of training, visual aids (a type of cue), and both training and visual aids on performance on an inspection task. The results showed that training alone resulted in a 32% increase in defect detections, visual aids alone resulted in a 42% increase, and use of both resulted in a 71% increase. The findings suggest that visual aids effectively improve performance on visual detection tasks. Further, incorporating visual aids into a training program may be a particularly useful method for improving performance.

Because accurate maternal-fetal heart rate monitoring depends on successful detection of critical signals, we reasoned that employing a visual aid to serve as a cue would facilitate detection among observers. In accordance with this reasoning, a visual aid was created that highlighted the location of late decel embedded within the simulated maternal-fetal heart rate tracings.

C. Present Study

The purpose of the present study was to serve as an initial investigation of how a visual aid may affect observers' abilities to detect critical fetal heart rate signals. Specifically, the present study examined how well individuals could differentiate between early and late decels embedded within simulated maternal-fetal heart rate tracings under four different S/N ratios (representing absent, minimal, moderate, and marked variabilities) and varying delayed signal onsets (4, 8, 12, and 16 sec). Further, the study examined whether this ability to detect signals would be moderated by the use of a visual aid.

Based on the findings of Anderson et al. [9, 11, 12], it was expected that decreases in S/N ratio would result in lower detection performance, and late decels occurring at shorter onset delays would also result in fewer correct detections. Further, these effects were expected to be moderated by the presence of a visual aid serving as a cue. Specifically, it was predicted that the visual aid would improve detection performance primarily at lower S/N ratios and shorter decel onset intervals.

II. Method

A. Participants

Twenty-one undergraduate students (14 male; 7 female) from Old Dominion University participated in this study. They ranged in age from 18 to 24 (M = 19.33) and all had normal or corrected-to-normal vision. The ODU internal review board approved the study.

B. Maternal-Fetal Heart Rate Display

The static images were created on a maternal-fetal heart rate simulator [10]. Each image was 18 x 15.5 cm and consisted of a white background with two red grids. The grid at the top of the display showed the simulated fetal heart rate tracing. The bottom grid showed maternal contractions. Every image contained one contraction and one fetal heart rate decel, either early or late. The maternal and fetal tracings were displayed in blue.

Images were created to represent four different S/N ratios. The four S/N ratios were generated to approximate the four categories designated by the NICHD [6]: absent (S/N ratio = 10:1), minimal (S/N ratio = 5:1), moderate (S/N ratio = 2:1), and marked (S/N ratio = 1:1).

For images created for the visual aid condition, a turquoise, semitransparent crosshair cursor was overlaid on the image. The cursor was placed so that the horizontal bar was positioned at the baseline (i.e., mean) of the fetal heart rate beat-to-beat variability, and the vertical bar was positioned at the peak of the contraction. Examples of an early decel without and with the visual aid overlaid are displayed in Figs. 1 and 2, respectively.

A. Procedure

Participants twice viewed 80 images of simulated decels embedded in fetal-heart rate tracings. Each image was presented on a computer screen for 3 sec, immediately after which they used a computer mouse to select either "early decel" or "late decel." Prior to performing the experimental session, participants viewed examples of early and late decels and instructors gave descriptions of each. Further, they viewed practice images and indicated the presence of early or late decels. Participants were given feedback about their responses during practice.

The experimental session was performed in two blocks (with and without visual aids), the order of which was counter-balanced. For the early decels, the nadir was coincident with the contraction peak (0-sec delay). For the late decels, the nadir from the contraction peak was delayed in 4-sec increments. The set contained 16 early and 64 late decel

images (sixteen each with 4-, 8-, 12-, and 16-sec delays). Within each set of 16 images, 4 depicted absent, minimal, moderate, and marked heart rate variability as defined by the NICHD in 2008.

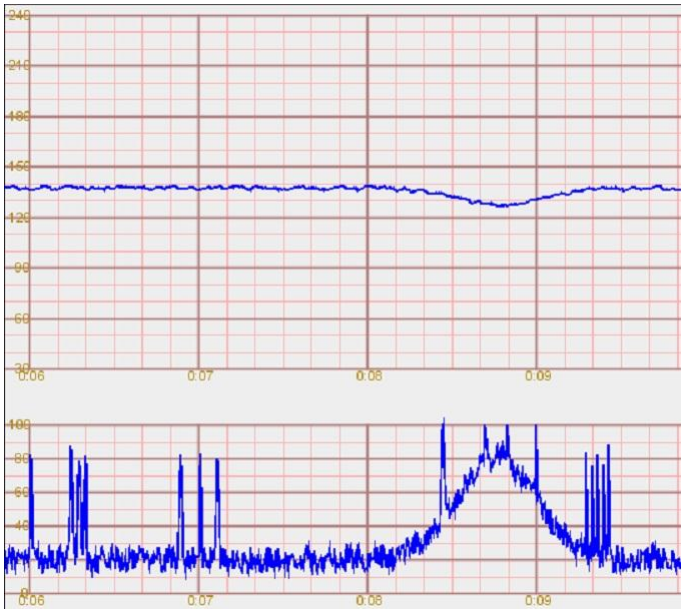


Fig. 1. Example of an early decel with minimal variability.

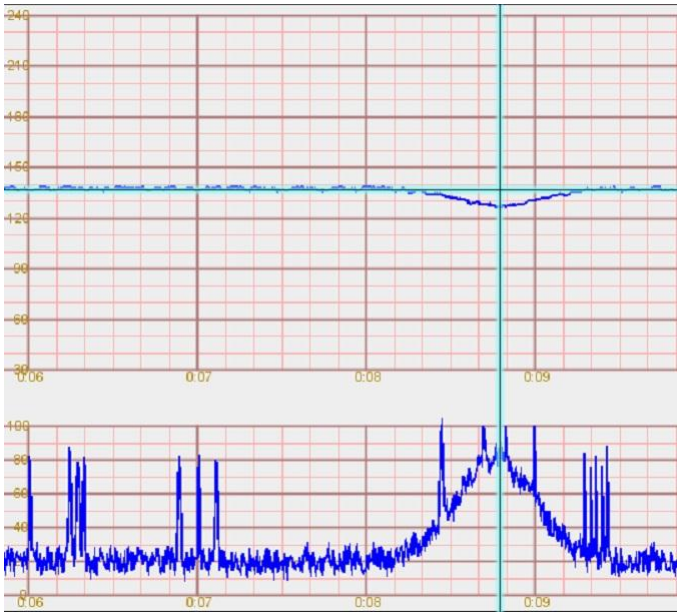


Fig. 2. Example of an early decel with minimal variability and a visual aid.

III. Results

The data were analyzed using a 2 (aid presence) x 2 (block order) x 4 (variability) x 5 (signal delay) mixed factorial analysis of variance (ANOVA). Aid presence, signal delay, and S/N variability were repeated measures, and order of aid-absent and aid-present blocks was a between-subjects factor. The dependent measure was proportion of correct responses; that is, the proportion of correct identifications (whether late decel or early decel) out of the total number of trials for which that signal was present.

There was a significant main effect of visual aid presence on the overall proportion of correct responses, $F(1, 19) = 25.23$, $p < .05$, partial $\eta^2 = .570$. Correct responses were significantly higher when the aid was present ($M = .74$, $SE = .02$) than when the aid was absent ($M = .66$, $SE = .02$).

For signals in which there was a delayed onset (i.e., late decels), correct responses increased as signal onset delay increased. The means for each delay interval are displayed in Table 1. A Bonferroni Sidak test revealed that the means were all significantly different from one another except between 0- and 12-sec, 0- and 16-sec, and 8- and 12-sec delays.

As variability increased (S/N decreased) among absent, minimal, moderate, and marked categories, correct responses also decreased. Table 2 displays the correct responses for each variability category. A Bonferroni Sidak test indicated that the means for each category differed significantly from one another except for the difference between the absent and minimal categories.

Last, there was a significant Delay x Variability interaction, $F(12, 228) = 13.25$, $p < .05$, partial $\eta^2 = .411$. This interaction is displayed in Fig. 3. A test of simple effects showed a significant difference in correct responses between the 4-sec and 12-sec delays and between 4-sec and 16-sec delays in the minimal variability condition, $F(4, 107) = 32.70$, $p < .05$. A Bonferroni post hoc test revealed higher correct responses in the 12-sec ($M = .94$, $SE = .10$) and 16-sec ($M = .98$, $SE = .08$) delay conditions than in the 4-sec ($M = .33$, $SE = .29$) delay condition. No other effects were significant.

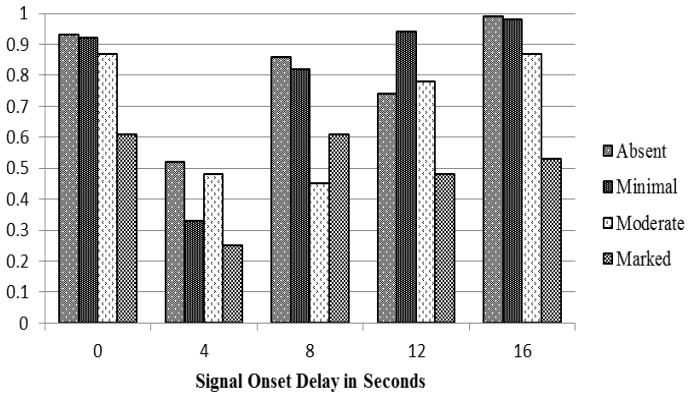


Fig. 3. Mean proportion of correct detections for each S/N ration and signal onset delay.

IV. Discussion

The purpose of the study was to determine whether cueing through visual aids would assist individuals in differentiating between early and late decels in simulated fetal heart rate tracings. Proportion of correct responses was examined across four different S/N ratios (absent, minimal, moderate, and marked) and with the late decel onsets occurring at incremental 4-sec intervals (4-, 8-, 12-, and 16-sec) from the peak of the contraction.

As predicted, the presence of visual aids was associated with overall higher proportions of correct responses. These results provide initial evidence that an auxiliary visual aid may serve as a cue to assist individuals with directing attention and detecting critical events in maternal-fetal heart rate tracings.

Results concerning the S/N ratios and signal onset delays were consistent with previous findings by Anderson and colleagues (2008, 2010, 2011) suggesting that observers have difficulty discriminating between early and late decels. Consistent with Anderson et al. (2008), the proportion of correct responses decreased with decreases in the S/N ratio. Thus, as the variability category levels progress from minimal to marked, detecting decels of the same magnitude becomes more difficult. Additionally, signal onset delay affected detection rates. Late decels were fairly easy to distinguish from early decels at longer onset delay (12 and 16 sec), but performance was near chance levels at short onset delays (4 and 8 sec). These results are also consistent with those reported by Anderson et al. (2010, 2011).

Finally, it was predicted that the visual aid would improve detection performance primarily for the lower S/N ratios. This hypothesis was partially supported. The Delay x Variability interaction revealed that the benefit of the visual aids was limited primarily to the minimal variability category (S/N 5:1). In minimal variability, the visual aids improved performance at the 8-sec relative to the 4-sec delay interval to near optimal levels. This finding was not observed under the lower S/N ratios (moderate and marked variabilities). It is possible that under these conditions the beat-to-beat variability masks the presence of decels to a degree that even the presence of visual cannot overcome.

V. Conclusion

The results of the present study provide initial evidence that an auxiliary visual aid was useful in helping novices to distinguish between early and late decels in simulated maternal-fetal heart rate static images. The clinical implications of these findings are that a visual aid may improve clinicians' performance in detecting critical fetal heart rate patterns. However, although the use of the visual aid would be helpful, it would not provide a complete solution for critical signal detection in this context. Decels occurring under moderate and marked levels of fetal heart rate variability, particularly with short onsets, may continue to prove challenging to observers. Further research is needed to examine how performance is affected when the user manipulates the visual aid in real time using a simulated dynamic tracing.

References

[1] Freeman, R. K. (2002). Problems with intrapartum fetal heart rate monitoring interpretation and patient management. *Obstetrics and Gynecology, 100*, 813-826.

[2] Sweha, A., Hacker, T. W., & Nuovo, J. (1999). Interpretation of the electronic fetal heart rate during labor. *American Family Physician, 59*, 2487-2500.

[3] Menihan, C.A., & Zottoli, E. K. (2001). *Electronic Fetal Monitoring: Concepts and Applications*. Philadelphia, PA: Lippincott.

[4] Blackwell, S. C., Grobman, W. A., Antoniewicz, L., Hutchinson, M., & Bannerman, C. G. (2011). Interobserver and intraobserver reliability of the NICHD 3-Tier fetal heart rate interpretation system. *American Journal of Obstetrics and Gynecology, 205,* e1-5.

[5] Chauhan, S. P., Klauser, C. K., Woodring, T. C., Sanderson, M., Magann, E. F., & Morrison, J. C. (2008). Intrapartum nonreassuring fetal heart rate tracing and prediction of adverse outcomes: Interobserver variability. *American Journal of Obstetrics and Gynecology, 199*, 623.e1-623.e5.

[6] Macones, G.A., Hankins, G.D., Spong, C.Y., Hauth, J., & Moore, T. (2008). The 2008 National Institute of Child Health and Human Development workshop report on electronic fetal monitoring: Update on definitions, interpretation, and research guidelines. *Obstetrics and Gynecology, 112*, 661-666.

[7] Figueras, F., Albela, S., Bonino, M. P., Palacio, M., Barrau, E., Hernandez, S., et al. (2005). Visual analysis of antepartum fetal heart rate tracings: Inter- and intra-observer agreement and impact of knowledge of neonatal outcome. *Journal of Perinatal Medicine, 33*, 241-245.

[8] Todros, T., Preve, C. U., Plazzotta, C., Biolcati, M., & Lombardo, P. (1996). Fetal heart rate tracings: Observers versus computer assessment. *European Journal of Obstetrics and Gynecology and Reproductive Biology, 68*, 83-86.

[9] Anderson, B. L., Scerbo, M. W., Belfore, L. A., & Abuhamad. A. (2010). When is a deceleration perceived as a late deceleration*? Simulation in Healthcare, 4*, 311.

[10] Belfore II, L. A., Scerbo, M. W., & Anderson, B. L. (2007, September). *A fetal heart rate monitor simulator.* Paper presented at the meeting of ModSim World, Virginia Beach, VA.

[11] Anderson, B. L., Scerbo, M. W., Belfore II, L. A., Abuhamad, A. Z., & Davis, S. S. (2011). *Can clinicians tell when a decel is a late decel?* Poster presented at the 2011 International Meeting on Simulation in Healthcare.

[12] Anderson, B. L., Scerbo, M. W. Belfore, L. A., & Abuhamad, A. (2008). Detecting critical patterns in maternal-fetal heart rate signals. In *Proceedings of the Human Factors and Ergonomics Society 52nd Annual Meeting* (pp. 1478-1482). Santa Monica, CA: Human Factors and Ergonomics Society.

[13] Harris, D.H., & Chaney, F.B. (1969). *Human factors in quality assurance.* New York: John Wiley & Sons. W. D. Doyle, "Magnetization reversal in films with biaxial anisotropy," in *1987 Proc. INTERMAG Conf.*, pp. 2.2-1–2.2-6.

[14] Gunn, D. V., Warm, J. S., Nelson, W. T., Bolia, R. S., Schumsky, D. A., & Corcoran, K. J. (2005). Target acquisition with UAVs: vigilance displays and advanced cuing interfaces. *Human Factors, 47*, 488-97.

[15] Posner, M.I., Nissen, M.J. & Ogden, W.C. (1978). Attended and unattended processing modes: The role of set for spatial location. In J.H.I. Pick & E. Saltzman (Eds.), *Modes of perceiving and processing information* (pp.137 – 157). Hillsdale, NJ: Erlbaum.

[16] Wickens, C. D. & Hollands, J. G. (2000). *Engineering psychology and human performance. (3rd. ed.).* Upper Saddle River, NJ: Prentice Hall.

[17] Chaney, F.B., & Teel, K.S. (1967). Improving inspector performance through training and visual aids. *Journal of Applied Psychology, 51*, 311 – 315.

Mathematics in a biological context: the use of modeling, analysis and simulation to enhance STEM education

J. Eggleston, H. Gaff, and G. S. Watson

Abstract—In mathematics, student cognition of calculations may be enhanced by a direct application and understanding of pertinent biological concepts. With this in mind, we sought to test this instructional strategy by creating a one-week college level mathematics curriculum using the biogeographical modeling context of the Theory of Island Biogeography.
.

***Index Terms*—Biodiversity, Biogeography, Mathematical model, Graph theory.**

I. Introduction

The demands of an instructor to meet the challenges in the classroom are immense. From fostering individual student cognition while meeting all required standards for learning, to keeping pace with ever-changing technological advances, teachers face a never-ending demand for their time and energy. The demands for STEM faculty are even greater given there is a societal, and often institutional, push for integrated and higher-order learning coupled with the desire for learner-centered pedagogically methods. One way to meet these demands for learning, while minimizing workload for individual teachers, is to use formal instructional design methods to create and test effective instructional materials for use by many. This also allows the collective creation of materials to be shared by all. However, the experience in a classroom both as a student and an instructor, provides a varied measure of success and failure for instructors to appropriately design, develop, and integrate effective instructional materials. In order to meet these challenges, funding agencies, professional associations, and departments have organized efforts to bring together experts in content, cognition, and instructional technologies to create instruction that is innovative, learner-centered, and strategically focused on integrating content and scientific methods appropriate for STEM. This paper describes the design and development of one such unit of instruction.

Manuscript received March 23, 2012. This work was supported in part by a Modeling and Simulation Graduate Research Fellowship, Office of Research, Old Dominion University and the National Science Foundation.

J. Eggleston, H. Gaff, and G.S. Watson are affiliated with the Virginia Modeling Analysis and Simulation Center and Old Dominion University, Norfolk, VA, 23529. (757-773-3734; email: jeggl001@odu.edu).

II. Background

Island biogeography is the study of the factors affecting species diversity of natural communities. As our natural landscape becomes more and more fragmented, it resembles a series of islands. The breaking up of landscape can degrade the ecosystem, reducing the ability of it to sustain itself [1]. This continued fragmentation could then result in the collapse and final disintegration of a given ecosystem, causing a decline in biodiversity [2]; a good measure of the health of the planet system is that of biodiversity [3]. The more species and genetic diversity, the more stable the ecosystem [4]. Continuing research supports that this loss of species can have definite consequences for an ecosystem [5]. As research continues to focus on the sustainability of the planet, we need this diversity to withstand stresses from changing weather patterns, disease and human population growth [6].

The Theory of Island Biogeography, the attempt to predict the number of species on an "island," was solidified in the 1960's [7]), and is still very valid today. It is important both to understand the definition and principles governing species [8], species richness, island biogeography and sustainability, as well as to be able to present the research data in an understandable form. It is vital to be able to ascertain the impact an introduction of new individuals from an adjacent habitat plays and how those "islands" are key to the continuation, or persistence of a population [9].

A fundamental way to evaluate and represent the relationship between biodiversity or in this case species richness, and the area of a given habitat is to graph the data, which is known as a species area curve [10]. With respect to our theory of island biogeography, or habitat fragmentation, this graphical representation of species richness versus habitat area, not only allows for the comparison of habitats of varying size, but more clearly illustrates existing relationships. Most importantly for habitat managers and ecologists are how species richness varies with the size of the habitat. If habitat size is replaced with the distance from the mainland to an island, or a smaller fragment to a larger central habitat, using a species curve will show these relationships.

Species richness and habitat size and/or distance data are analyzed by graphical representations such as the previously mentioned species area curves, as well as by using biological indices. A graph of the basic data for a species-area relationship will appear as a curve with data points concentrated or condensed at one end. Using a log-log plot

with this same data will scale the data. To scale data is to place the data in proportion to each other in such a manner that existing relationships between our species richness and either area or distance are easier to see. Ecological indices, such as Simpson's and Shannon-Wiener, are similar to those extensively used in economics to assess the distribution of wealth in a region. Indices allow us to reduce down and compare habitats and the populations of species in terms of the probability that we will encounter a given species in a given location.

III. METHODS

Our instructional design process and resulting documentation included the structured analysis of the instructional problem, characteristics of the learner, content, instructional objectives, content sequencing, instructional strategies, and elements of message design, development of instruction, and evaluation instruments and methods [11]. Specifically, we used a prescribed project template from Mathematics for Planetary Earth 2013 (MPE 2013) [12] and current standards for effective instructional design [11]. Our module included instructional materials to guide both the student and the instructor through the biological concepts and specific mathematics. The materials included pre-session readings and evaluations, lecture materials, in-class work, homework, as well as additional resources for the student and instructor, from literature and online references, to a glossary.

Our materials were developed using a variety of available datasets to explore methods of calculating biodiversity and measuring landscape, as well as the relationship between those. These points were then used to teach logarithms by estimating slopes and intercepts from a log-log plot of the number of species in a given location, against a variety of metrics, including island size and distance from mainland. Finally, the plots were used to simulate and model the level of fragmentation that would push the system to a given level of species loss. The mathematics and analyses were used as a framework for a final discussion of the biological concept.

IV. FUTURE WORK

Our continuing work includes the summative, formative, and confirmative evaluations of our field test of this module at ODU, with IRB approval, in the mathematics department, Spring and Summer of 2012. The formative analysis will consist of expert review and small group implementation [13]. Experts, consisting of at least one content expert from biology-math and one from instructional design and technology, will review all material commenting on accuracy of content, perceived effectiveness of instructional strategies, and appropriateness of assessment methods. The instruction will be revised and specific recommendations will be implemented prior to testing with students. The small group assessments will involve presentation of the revised materials with a single class and will rely on pre-test and post-test comparisons of content knowledge, a survey of learner perceptions of the instruction and their own learning, structured observation scales recorded by researchers observing the instruction, and interviews with a convenience sample of learners. Scores will be used to establish reliability for the measures and collective findings will be used to modify the instruction as needed. Upon final analysis, we expect our results to reflect an enhanced understanding and willingness to learn the mathematical concepts. Additionally, with an online release enhanced with a variety of models and simulations as a part of MPE 2013, we expect those results to reflect a greater conceptual understanding of the material, but not necessarily a greater application.

ACKNOWLEDGMENT

J. Eggleston and H. Gaff thank the Center for Discrete Mathematics and Theoretical Computer Science for the opportunity to participate in the Mathematics for Planetary Earth 2013 workshop.

REFERENCES

[1] Chapin, F. S., M. S. Torn, and M. Tateno. "Principles of ecosystem sustainability," *The American Naturalist*, vol.148, pp. 1016-1037, 1996.

[2] Ibáñez, I., J. S. Clark, M. C. Dietze, K. Feeley, M. Hersh, S. LaDeau, A. McBride, N. E. Welch, and M. S. Wolosin. "Predicting biodiversity change: outside the climate envelope, beyond the species-area curve," *Ecology*, vol. 87, pp. 1896-1906, 2006.

[3] McGrady-Steed, J., P. M. Harris, and P. J. Morin. "Biodiversity regulates ecosystem predictability," *Nature*, vol. 390, pp. 162-165, 1997.

[4] Butchart, S. H. M., M. Walpole, B. Collen, A. van Strien, J. P. W. Scharlemann, R. E. A. Almond, J. E. M. Baillie, B. Bomhard, C. Brown, J. Bruno, K. E. Carpenter, G. M. Carr, J. Chanson, A. M. Chenery, J. Csirke, N. C. Davidson, F. Dentener, M. Foster, A. Galli, J. N. Galloway, P. Genovesi, R. D. Gregory, M. Hockings, V. Kapos, J.-F. Lamarque, F. Leverington, J. Loh, M. A. McGeoch, L. McRae, A. Minasyan, M. H. Morcillo, T. E. E. Oldfield, D. Pauly, S. Quader, C. Revenga, J. R. Sauer, B. Skolnik, D. Spear, D. Stanwell-Smith, S. N. Stuart, A. Symes, M. Tierney, T. D. Tyrrell, J.-C. Vié, and R. Watson. "Global Biodiversity: Indicators of Recent Declines," *Science*, vol. 328, pp. 1164-1168, 2010.

[5] Worm, B., E. B. Barbier, N. Beaumont, J. E. Duffy, C. Folke, B. S. Halpern, J. B. C. Jackson, H. K. Lotze, F. Micheli, S. R. Palumbi, E. Sala, K. A. Selkoe, J. J. Stachowicz, and R. Watson. "Impacts of Biodiversity Loss on Ocean Ecosystem Services," *Science*, vol. 314, pp. 787-790, 2006.

[6] Heller, N. E. and E. S. Zavaleta. "Biodiversity management in the face of climate change: A review of 22 years of recommendations," *Biological Conservation*, vol. 142, pp. 14-32, 2009.

[7] MacArthur, R. H. and E. O. Wilson. "An equilibrium theory of insular zoogeography," *Evolution*, vol. 17, pp. 373-387, 1963.

[8] Mayr, E. "A local flora and the biological species concept," *American Journal of Botany* , vol. 79, pp. 222-238, 1992.

[9] Quinn, J. F. and S. P. Harrison. "Effects of habitat fragmentation and isolation on species richness: evidence from biogeographic patterns," *Oecologia*, vol. 75, pp. 132-140, 1988.

[10] Cain, S. A. "The Species-Area Curve," *American Midland Naturalist*, vol. 19, pp. 573-581, 1938

[11] Morrison, G.R., S.M. Ross, J.E. Kemp, and H. Kalman, *Designing Effective Instruction, 6th ed.* Hoboken, NJ: Wiley, 2011, pp.

[12] Proceedings of the 2011 Center for Discrete Mathematics & Theoretical Computer Sciences Workshop "Mathematics for Planetary Earth Sustainability Curriculum", Rutgers University, Piscataway, NJ, October 20-22, 2011.

[13] Tessmer, M. (1993). *Planningg and Conducting Formative Evaluations: Improving the Quality of Educaiton and Training*. Abingdon, Oxon: Routledge.

Transportation

VMASC Track Chair: Dr. Mike Robinson
MSVE Track Chair: Dr. ManWo Ng

Modeling Inland Multimodal Transportation using the Generic Shortest Path Problem
Author(s): Umama Ahmed and Yishu Zheng

Command and Control Quad-Copter and Ground Vehicle: Autonomous Task Execution
Author(s): Christopher Lynch, and Yiannis Papelis

Using Vehicle-to-Vehicle Communications to Clear Paths for Emergency Vehicles
Author(s): Craig A. Jordan, and Mecit Cetin

Evaluation of Traffic Assignment with Road Capacity Uncertainty
Author(s): Linmin Pei

Autonomous Mission Execution for Quad-Copter
Author(s): Hamdi Kavak, and Yiannis Papelis

Demand Responsive Signal Control Strategy for Over-Saturated Signalized Intersection
Author(s): Rahul M. Rajbhara, and Mecit Cetin

Modeling Inland Multimodal Transportation using the Generic Shortest Path Problem

Umama Ahmed
Department of Modeling, Simulation and Visualization Engineering
Old Dominion University
uahme001@odu.edu

Yishu Zheng
Department of Modeling, Simulation and Visualization Engineering
Old Dominion University
yzhen003@odu.edu

Abstract - Efficient and reliable freight transportation is a fundamental requirement for a nation's economic growth. The cost of freight transportation directly affects the price of products, which in turn impacts the consumers demand and spending, and thus affects the economy of a country. Therefore selection of an appropriate transportation mode for goods movement is critical. The transportation mode choice decision is made considering the objectives: transfer cost minimization, time reduction and increased level of service reliability. In this paper we modify the generic single source shortest path problem in order to solve an inland multimodal transportation network problem.

Keywords - Intermodal transportation, freight transportation, networks, shortest path problem.

I. Introduction

Transportation is described as a service that provides movement of goods and people via different modes [1]. The movement and exchange of goods via freight transportation permits the development of other social and economic activities to take place. Freight transportation is an important and complex domain since it involves several players and different levels of decision. Thanks to the standardized containers, for the first time the intermodal transportation of goods became commercially feasible.

Intermodal freight transportation is defined as the movement of loads from origin to destination by means of at least two different types of transportation modes: truck, rail, air, ship, and pipeline [1]. The transfer operations are therefore performed at intermodal terminals in which two or more transportation means connect their transportation way. Often, transportation is provided by different carriers; therefore, question arises when companies have to decide what transportation service mix to use in order to reach the optimal desired results: cost minimization, time reduction, level of service reliability are often objectives that companies are looking for while search for transportation services.

In this paper, we present an inland intermodal freight transportation network problem. We utilize the basic concepts of solving a shortest path problem, in order to adjust the formulation into the problem. The objective of the model is to determine the best combination of transportation modes to be utilized to move the freight from an origin to its destination. For exemplify the problem, a simple network is created to better address the problem.

We first present a brief literature review of the problem. Then, we introduce the problem description and motivation. Later on, we present the adjusted formulation, with a simple network for better exemplification. Finally, we describe future work and research direction in the conclusion.

II. Motivation

Different transportation modes have different cost and time functions depending on the origin-destination distance. Road transportation by truck has a lower cost function for short distances, but its cost function grows faster than the rail and maritime cost functions as distance increases. For long distance freight transportation, rail and ocean shipping is highly chosen because of its low transportation rate comparing to other modes; however, these modes are less time-effective than air and road transport. While air transportation is the most time-effective mode for moving high value products or perishables, it is not cost effective for large bulk nonperishable products due to high cost function. Pipeline is the cheapest transportation mode; nevertheless, it is restricted to move only fluids in bulk over fixed routes under specific conditions [1]. For inland

transport of bulk products, truck and rail transport is most suitable because of wide network coverage.

Freight transportation has to adapt to the changing economic environments, operational practices, and social trends. Recent rise in the price of oil is significantly raising the cost of transportation in all modes of freight transportation. Even though the increase in oil prices has strongly impacted the freight transportation cost, these cost fluctuations differs based on transportation modes. A study done by U.S Department of Maritime Transportation indicated that the projected price of diesel fuel price for 2020 over the year 2002 will cause truck line-haul costs to increase 62 percent ($0.87 per mile) over 2002. On the other hand, for the same period, rail line-haul costs will experience an increase of 50 percent ($0.15 per mile) [2]. Therefore, to adapt the high rise in oil prices it is necessary to integrate the services of higher price suppliers like trucking and lower cost suppliers such as rail for inland transportation of bulk goods. Integration of the two different modes (intermodal transportation) will have the benefit of road transport's relative time effectiveness over rail and rail transport's cost effectiveness over truck.

In this paper, we work on selecting the best inland transportation mix for goods to be moved from origin to destination in order to minimize the transportation cost with given time constraint.

A simple transportation network with inland transportation modes will be considered in order to present our model formulation. As result, the outcome will show the best mode choice based on the objective and constraints. Two possible scenarios will result from our analysis:

1. Goods transferred from origin to destination by truck only.
2. Goods transferred from origin to destination by integration of truck and rail transportation.

III. Literature review

The shortest path problem is a network problem of finding the shortest way between a set of nodes, such that the total weight (w) of the path is minimized [2]. Given a directed network $G = (V, E, w)$ defined by a set V of n number of nodes, and a set E of m number of directed links in G, each arc $(i,j)\ \epsilon\ E$ has an associated cost or weight $w_{ij}\ \epsilon\ E$. The objective is to find for each non-source node $i\ \epsilon\ V$ a shortest directed path from the source node s to the sink node v. This problem is also translated as sending one unit of flow with the lowest cost w_{ij} from node s to each of the nodes in $V - \{s\}$ in an uncapacitated network. The problem formulation from this point of view follows the linear programming formulation displayed below:

$$min. \sum_{(i,j)\epsilon\ E} w_{i,j} x_{i,j} \quad (1)$$

$$subject\ to:$$

$$\sum_{j:(i,j)\ \epsilon\ E} x_{ij} - \sum_{j:(j,i)\ \epsilon\ E} x_{ji} = \begin{cases} n-1 \ \ for\ i = s & (2) \\ -1 \ \ for\ all\ i\ \epsilon\ V - \{s\} & (3) \end{cases}$$

$$x_{i,j} \geq 0 \ \ for\ all\ (i,j)\epsilon\ E \quad (4)$$

Equation (1) describes the objective, to minimize the total weight (e.g. cost, time, etc.) of the chosen path. Therefore, if the route is chosen, x_{ij} will have value one, otherwise zero. For the constraints, we look at the number of paths exiting and entering a node, equation (2) states that for $i = s$, paths have to leave source node. Equation (3) states that for any other node, the number of links entering the node must be exactly one larger than the number of links leaving the node. Equation (4) defines x_{ij} cannot handle negative movements.

Operation research practitioners have investigated variations of the shortest path problem, and they are widely used in different fields such as communications, plant and facility layout, scheduling, robotics, transportation, etc. [3]-[6].

IV. Model Formulation

We will base our formulation in the generic single source shortest path problem, which help to find a path between n nodes with minimum distance, time or cost from a source to a destination [7]. The classical shortest path problem does not differentiate transportation modes. For our model (Figure 1), considering this simple three nodes network, we will exemplify a shortest path problem with two transportation modes. Assumptions are:

1. Two types of transportation modes are available: highway, and railway linking the specific nodes.

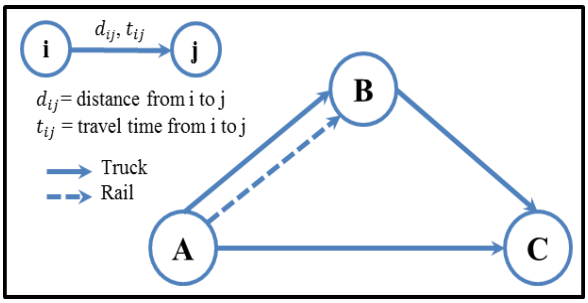


Figure 1.

2. All transportation services are reserved in advance, so there is no order and booking time. Transportation services are available at their schedule time.
3. Road and rail modal options are interchangeable for the origins and destinations.
4. The model does not consider any reliability issue related to freight transportation.
5. Since transportation by rail has discrete networks, they are only accessible via specific terminals.

We will transport the freight from origin A to destination C within provided transportation time in hours T. B is an intermediate node that has transshipment terminal for the transfer of goods. A and B is connected via both highway and railway; while B and C is only connected via highway. Goods may be transported from A to C in either of the following two transportation mix:

1. A to B by rail and then B to C by truck.
2. A to C by truck.

We consider c as the user cost function for the particular mode, t_{ij} as the transportation time (hours) incurred by transportation mode traveling from node i to node j, and d_{ij} as the distance incurred by traveling from node i to node j. Therefore, we reformulate the generic shortest path problem, and our objective function becomes:

$$min.\ x_{AB_h}d_{AB_h}c_h + x_{AB_r}d_{AB_r}c_r + x_{BC_h}d_{BC_h}c_h + x_{AC_h}d_{AC_h}c_h \quad (5)$$

s.t.

$$x_{AB_h} + x_{AB_r} \leq 1 \quad (6)$$

$$-(x_{AB_h} + x_{AB_r}) + x_{BC_h} = 0 \quad (7)$$

$$x_{AB_h} + x_{AB_r} + x_{AC_t} \leq 1 \quad (8)$$

$$x_{AB_h} + x_{BC_h} = 1 \quad (9)$$

$$x_{AB_h}t_{AB_h} + x_{AB_r}t_{AB_r} + x_{BC_h}t_{BC_h} + x_{AC_h}t_{AC_h} \leq T \quad (10)$$

$$x_h = \begin{cases} 1, & if\ transported\ by\ highway \\ 0, & otherwise \end{cases}$$

$$x_r = \begin{cases} 1, & if\ transported\ by\ railway \\ 0, & otherwise \end{cases} \quad (11)$$

Equation (5) presents the objective - to minimize the total transportation cost. Constraint (6) restricts to choose only one transportation mode from node A to B. Constraint (7) states that if route A to B is chosen, we must choose route B to C. Constraint (8) restricts to choose only one route. Equation (9) sets the network equilibrium constraint. Constraint (10) sets the total transportation time limit and it is possible that with the given time constrains no feasible solution is found. If that scenario exists, then the time constraint will be relaxed and provided transportation time will be increased. Equations (11) restrict our variables to be integer values from 0 to 1.

V. Methodology

In order to validate the formulation, we have solved this simple network using Microsoft Excel, Solver Add-Ins. For calculating the total transportation cost, we have considered the user cost for rail and truck transportation to be 2.7 and 5.0 cents per ton-mile [8]. The user cost is based on 2002 market rate, which we have considered in our calculation due to the unavailability of current value of users cost for rail and truck transportation. For calculating time to reach from origin to destination, a speed of 65 mph for truck and 50 mph for rail is considered [9]. Since for safety issues, truck drivers are required to take rest therefore for every 10 hours of operation we have considered 8 hours of break in operation during calculation of truck transportation time from origin to destination. As an example, for traveling at 65 mph on average, the required time for trucks to travel 1200 miles is around 18 hours. We have considered 8 hours of break during driving this distance and thus the required time to transport 1200 miles by truck is 26 hours (10 hours of driving+8 hours of break+8 hours of driving) as a total. Similarly, for railroad transportation, trains may stop in between terminals to load/unload other goods and therefore we have considered 4 hours of stop time for

every 10 hours of operation. Also an intermodal transfer time of 3 hours has been considered for transferring goods from rail to truck. As an example, traveling at 50 mph, the required time for train to travel 1200 miles is about 24 hours. However considering 4 hours of stop time for every 10 hours of operation and including 3 hours of transfer time in the transshipment terminal, the actual required time to travel the 1200 miles distance by train is found to be 35 hours (10 hours of operation+ 4 hours of stop+10 hours of operation+4 hours of stop+4 hours of operation+3 hours of transfer). From point A to B, we have considered the railway and highway route length to be same. We have ignored the variability of route length based on mode of transportation for calculation simplicity.

We analyzed the problem with two different time constraints. Results are presented on Table 2. The results indicated that when time constraint is wide, the multimodal transportation is the best solution for transporting goods. However, with narrow time constraint, a direct route from origin to destination by truck transportation is the best solution.

Varying the origin-destination distance, we ran our model several times. It becomes evident that when distance is less than 500 mile, for transferring goods, truck transportation is the optimized solution over multimodal transportation with both wide and narrow time constraints.

Route	AB_h*	AB_r*	BC_h*	AC_h*
Distance (miles)	1200	1200	350	1400

Table 1. Route length assumed for the simple origin-destination network

		Cost ($ /ton)	Time (hrs.)
	AB_r-BC_h	49.9	40
Route	AB_h-BC_h	77.5	40
	AC_h	70	37
Time constraint		48 hours	36 hours
Optimized solution		Route AB_r-BC_h	Route AC_h

Table 2. Results based on cost and time of different modes

*AB_r = A to B transported by rail
AB_h = A to B transported by truck
BC_h = B to C transported by truck
AC_h = A to C transported by truck

For our future work, we would like to analyze our model with a larger and more complex network. Furthermore, to have a more realistic model, we will also take into consideration the railway and highway route length variability.

VI. Conclusion

We have presented an inland intermodal transportation network problem to be solved based on the generic single source shortest path problem. We utilized a simple three nodes model to presents the way we can modify and reformulate the generic shortest path problem into an intermodal transportation network problem. More studies and analysis are needed to perform in order to model large networks with more nodes and links. With more transportation modes, larger network implies larger the number of constraints. Moreover, this approach might also be applied into larger scale networks including other transportation modes (maritime and air transportation). Furthermore, considerations regarding freight transportation scheduling, transportation mode capacity, pricing discrimination strategies are also topics that might need to add into this formulation in order to have a more accurate model.

References

[1] Talley, W. Port Economics. New York: Routledge, 2009.

[2] "Impact of High Oil Prices on Freight Transportation: Modal Shift Potential in Five Corridors", Transportation Economics & Management Systems, Inc., October 2008

[3] Zhu, X., Wilhelm, W. "Three-stage approaches for optimizing some variations of the resource constrained shortest-path sub-problem in a column generation context." European Journal of Operation Research, vol. 183, Issue 2, PP 564-577, Dec. 2007.

[4] Macharis, C., Pekin, E., Rietveld., "Location Analysis Model for Belgian Intermodal Terminals: toward an integration of the modal choices variables". Procedia – Social and Behavioral Sciences, vol. 20, pp 79-89, Sep. 2011.

[5] Bontekoning, Y.M., Macharis C., Trip J.J., "Is a new applied transportation research field emerging? – A review of intermodal rail-truck freight transport literature", Transportation Research Part A: Policy and Practice, vol. 38, Issue 1. pp 1-34, Jan. 2004.

[6] Grainic, T.G., Kim, K.H., "Chapter 8. Intermodal Transportation". Handbook in OR and MS., Eisevier B.V. vol. 14, pp 467 -529, 2007

[7] Ahuja, R.K., Magnanti, T.L., Orlin, J.B. Network Flows: Theory, Algorithms and Applications. Prentice-Hall, Englewood Cliffs, NJ, 1993.

[8] Brown, T. R., Hatch, A. B., "The Value of Rail Intermodal to the U.S. Economy", September 19, 2002.

[9] Eric Anderson, "Faster trains have a price, CSX says" Internet: http://www.timesunion.com/news/article/Faster-trains-have-a-price-CSX-says-562235.php, June 24, 2010 [March 7, 2012].

Command and Control Quad-Copter and Ground Vehicle: Autonomous Task Execution

Christopher Lynch
Modeling, Simulation & Visualization Engineering
Old Dominion University
Norfolk, United States

Yiannis Papelis, Ph.D., Menion Croll
Virginia Modeling, Analysis and Simulation Center
Old Dominion University
Norfolk, United States

Abstract - the purpose of this project is to design an autonomous movement system for a quad-copter (QC) and a remote controlled (RC) ground vehicle (GV) using Global Positioning System (GPS) coordinates. The QC and GV are specifically focused for live-virtual-constructive simulation of a command and control (C2) system.

I. INTRODUCTION

A QC and RC have been designed to receive commands from an external C2 system. This allows for the QC and the GV to be used for experimentation and to carry out missions. These missions are intended to be carried out by either of these vehicles autonomously. A number of factors must be considered in order to generate autonomous movement for the GV and even more considerations must be taken to generate autonomous flight for the QC. The main goal for both the QC and the GV is to constantly update their positions in order to move autonomously towards their intended target(s). A GPS device has been selected to track the position of the vehicles as they move.

The most important factor for the QC is that it must be able to constantly update its altitude, pitch, and roll so that it can remain airborne. The QC is currently intended to carry out the following tasks:

i) Hover at a given location and altitude
ii) Follow a series of waypoints
iii) Land autonomously

All of the initial components used for creating the autonomous movement were first tested on a RC GV. These tests were done in an effort to test the hardware and software without exposing the QC to the risk of being damaged due to an unexpected malfunction. The QC and GV can both be switched to RC mode and navigated manually if necessary.

II. PROBLEM / MOTIVATION

The idea for this project came from the growing need to create simulation environments that combine live, virtual, and constructive (LVC) simulations. An unmanned aerial vehicle (UAV) was selected as the device to use as a bridge between these simulation types. With this is mind, a QC was chosen as the UAV to use for this purpose. Adding autonomous flight capability to the QC allows for new components to be integrated and tested within the LVC environment. QCs are already designed for use in surveillance and search and rescue

operations [1]. The idea for using a GV arose from the desire to test the general software capabilities before putting the vehicle in the air.

III. METHODS

The overall system architecture for the QC is displayed in Figure 1. Propeller Board 2 is responsible for controlling the autonomous flight. This board requires several inputs in order to effectively control the QC. The necessary inputs include position data from the GPS, the tasks that it needs to carry out, and information from any other sensors contained on the vehicles.

The Propeller board constantly communicates with the PIC microcontroller to obtain the GPS data on position and altitude. The Propeller board obtains its current task list from Propeller Board 1 which receives the overall mission from the control center. The Propeller board then takes its current position based on the GPS data and calculates the necessary changes to pitch, roll, and thrust needed to move the QC towards the desired location using PID control loops.

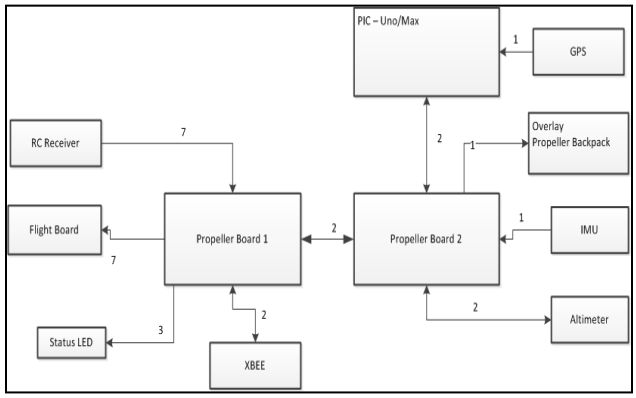


Figure 1: Quad-Copter Autonomy System Architecture

The setup of the GV is almost identical to the QC except that only one Propeller board is used, there is no flight board, and all of the inputs feed into this single Propeller board. The Propeller board was selected because it is designed for use with high-speed processing in embedded systems. The Propeller board uses eight processors, referred to as cogs, to perform simultaneous independent or cooperative tasks. This helps to speed up the responsiveness of the QC in correcting its position and executing functions [2].

The Propeller board handles all of the tasks carried out by the QC. The hover task is the most simple of the desired tasks. Ideally the QC only needs to check and correct its altitude while hovering. However, the QC also tends to drift out of position due to wind and its own residual motion and it needs to correct for this displacement as needed. The magnitude of the pitch and roll adjustments are based on the total distance that the QC has drifted from its target location.

The QC follows a series of waypoints by pointing the heading of the QC towards the desired waypoint. The QC is designed to move in a direct line from one waypoint to the next in the order that the waypoints were listed from the control center. The QC's speed is determined based on its distance from the target location. The QC is designed to slow down (reduce pitch) as it approaches its target in order to prevent it from overshooting its target.

The GV travels by adjusting its velocity and heading. The PIC completes some of the calculations for determining the needed speed and steering values and then transfers them to the Propeller. The Propeller takes these inputs and uses the values to adjust the motor. The GPS device is the key component in obtaining the information that is necessary to make these calculations because it provides the vehicles with their position information.

The main information that is retrieved from the GPS is the GPRMC sentence. This sentence contains all of the main information that is needed from the GPS. The fix status is the first item that is looked for

from the GPS. This tells the GV whether or not it has connected to three or more satellites yet. If the fix status is 'A' then the GPS is receiving accurate readings from the satellites. Next, the current latitude and longitude values in decimal degrees are obtained. The compass directions are also obtained with the latitude and longitude readings and are North/South in relation to the equator and East/West in relation to the Prime Meridian.

The next item in the sentence contains the current velocity of the GPS. This value is given as speed over ground in knots which requires further conversion on the PIC. The final value that is retrieved from this GPS sentence is the true course information [3]. True course means that the course is an angle between 0 and 360 degrees with north set as 0 degrees. The angle moves clockwise with east at 90°, south at 180°, and west at 270° [4]. This coordinate axis setup is displayed in figure 2.

Several steps are required for determining the desired steering value. First, the current heading of the vehicle, the current latitude and longitude positions of the vehicle, and the latitude and longitude coordinates of the target must be obtained from the PIC. The latitude and longitude values are then converted into the Universal Transverse Mercator (UTM) coordinate system. The UTM system projects the cylindrical map coordinates from the Earth into relatively undistorted, flat zones. These zones are broken into precise measurements in meters to within one meter [5]. This allows for the vehicles to track their movement in centimeters while they carry out their tasks.

Next, the heading of the vehicle (in degrees from north) needs to be compared against the heading that it needs to be taking to get to its target location. To do this, the current position of the vehicle is set to the (0, 0) location of the coordinate axis as seen in Figure 2. Then the angle in degrees from north of the target location can be calculated. This value is then subtracted from the current heading to produce the magnitude of the angle at which the vehicle needs to turn. The quadrant that the target location falls in relative to the vehicle determines whether the vehicle will turn left or right depending on which route will be shorter.

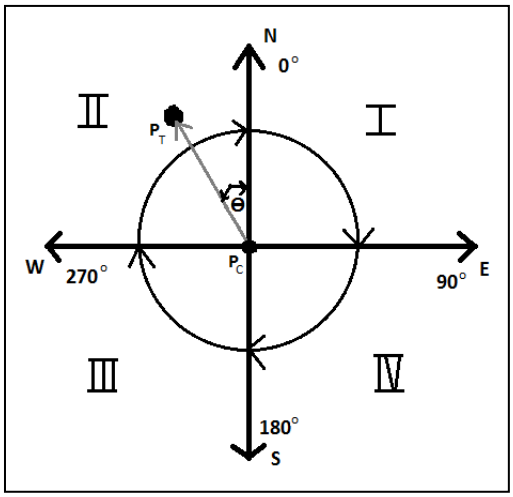


Figure 2: Angle Calculations from Coordinate System

A number of safety features have also been implemented in order to protect the vehicles while in autonomous mode. For instance, warnings have been setup on the QC to let it know if it is running low on battery. This tells the QC to immediately return home or to conduct an emergency landing if it does not have enough power to return home. In addition, a leash has been connected to the GV so that it can be stopped manually from driving into another object. It is possible that the GV stops receiving information from the Propeller board if the connection becomes loose while it is driving autonomously. In this case the leash also prevents the vehicle from having to be chased and recovered during testing.

IV. EXPERIMENTS

All of the current experimentation has been conducted on the RC GV. Two main experiments have been conducted with the GV. The first

experiment was simply to get the GV to move autonomously from any starting location to a specified target location. This required that a target position be specified using UTM coordinates that would be received on the PIC as part of the mission parameters. This mission is sent from the external C2 system and processed on the PIC. Initially, the GV was setup with static turn values for going either right or left. During the test the GV successfully turned right and left as expected. However, the vehicle did not travel in a very straight line to get to its target location.

The GV was then adjusted to turn based on the magnitude of the angle between its current heading and its desired heading. The test was conducted again and the travel path of the vehicle smoothed out considerably. In both of these cases, the GV was started from several different locations and in all cases it found its target without overshooting its mark.

The second experiment conducted with the GV was traveling along a series of waypoints to conduct a survey mission. This mission is setup to send a series of four or five waypoints to the vehicle that cause it to circle a target in an effort to survey the area. The GV successfully traveled from one point to the next while taking the shortest path between each point. Again the GV was started from several different starting positions and it successfully found all of its target waypoints. In one case, the GV overshot its target but it successfully turned around, found the correct point, and continued moving to the remaining points.

V. RESULTS

The GV successfully traveled to within a 2 meter radius of each waypoint. As soon as it detected that it had reached the current waypoint it adjusted course to move towards the next waypoint. One of the experiments is shown in Figures 3 and 4. Both of these figures were generated using Google Maps to plot latitude and longitude points. Figure 3 shows the location that is to be surveyed (in yellow) and the 5 waypoints that the vehicle needs to travel to in order to successfully survey the area around that location. Figure 3 was generated before conducting the experimental run. Figure 4 shows the path that the GV took in reaching all of the waypoints. It can be seen in Figure 4 that the GV alters course the most when it has arrived at a waypoint and needs to turn significantly to head to the next waypoint.

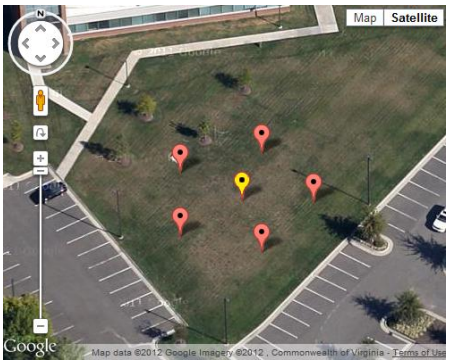


Figure 3: Waypoint Setup for Mission

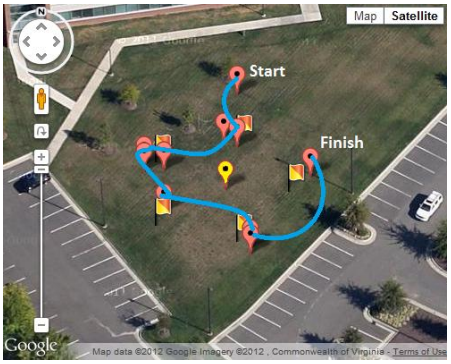


Figure 4: GV's Execution of Mission

VI. CONCLUSION / FINDINGS

The experimentation with the GV has shown that the autonomous task execution using GPS positioning works in a two-dimensional space. The autonomous task execution should be ready for testing on the QC soon. The only change that needs to be made within the autonomy function is that

altitude calculations will need to be computed as well. The GV has already shown that it can carry out a mission received from an external C2 system.

REFERENCES

[1] T. Luukkonen, "Modelling and control of a quadcopter," Aalto University, Finland, Aug. 2011.

[2] *Propeller Manual*, Version 1.2, Parallax Inc., Rocklin, CA, 2011.

[3] Baddeley, G. (2011, May 24). Glenn Baddeley - GPS - NMEA sentence information. Retrieved from http://home.mira.net/~gnb/gps/nmea.html

[4] U.S. Department of Transportation. (2008). *Pilot's Handbook of Aeronautical Knowledge* (FAA-H-8083-25A). Oklahoma City, OK: Federal Aviation Administration.

[5] Riesterer, J. (2008, April 7). UTM – Universal Transverse Mercator Geographic Coordinate System. In *Geospatial Training and Analysis Cooperative: Introduction to Topographic Maps*. Retrieved from http://geology.isu.edu/geostac/Field_Exercise/topomaps/utm.htm

Using Vehicle-to-Vehicle Communications to Clear Paths for Emergency Vehicles

Craig A. Jordan, Mecit Cetin

***Abstract*— This paper presents a new strategy to enable an emergency response vehicle (EV) to navigate through congestion at signalized intersections more efficiently. The proposed strategy involves communicating control messages to vehicles to change their behavior so that EVs can proceed through congested intersections as quickly as possible. To achieve this, the vehicle queue in one lane is split at a critical point to allow the EV to proceed at its desired speed while minimizing the disruption to the rest of the traffic. The proposed method makes use of the kinematic wave theory (i.e., shockwave theory) to determine the critical point in the vehicle queue. The proposed method is simulated in a microscopic traffic simulator for validation and evaluation. The results show that this strategy can shorten the trip time for EVs.**

I. Introduction

In 2009, the US Department of Transportation developed a five-year strategic research plan with the purpose of improving safety, mobility, and the environment. The foundation for the plan is vehicle-to-vehicle (V2V) and vehicle-to-infrastructure (V2I) communication, which uses a wireless communication system between vehicles and infrastructure to provide a connection for the transmission of traffic flow information. The idea is that vehicle information such as speed and location is communicated between vehicles and the infrastructure to create a smarter transportation network. Areas of research have included improving on-ramp merging at freeways, intelligent signal design, and queue length estimation on surface roads and freeways.

One area that has seen an increase in research is in individual emergency response vehicles (EVs). The idea is to equip an EV with a V2V/V2I communication system so that it can transmit its location, route, and final destination to vehicles in its path and to the infrastructure to improve travel time. V2V/V2I equipped vehicles in the immediate area of the EV and in the EV's path would receive a message indicating that an EV is coming and to move to the right and stop. This would allow for the safe passing of the EV. In addition, the infrastructure would be notified of the EV's path and location so that signal lights could be modified to provide green time for the EV's direction as it approaches an intersection. The infrastructure would transmit the location, path, and destination of the EV to any vehicle that might be impacted. This would allow for drivers to make better route choices. Implementation of this concept is currently underway in Europe. ERITCO, Europe's intelligent transportation system, has developed the Rescue system (http://www.ertico.com/assets/download/GST/RESCUE.pdf) which allows vehicles to be outfitted with a communication device and visual display that is designed to alert drivers of the approach of an EV. The system will also allow the transmission of data between non-emergency vehicles equipped with the system devices to alert drivers of the location of the EV at the emergency scene.

The focus of this paper is to develop efficient strategies to allow EVs to travel congested roadway segments that have signalized intersections. In particular, the paper deals with two-lane roads where vehicles queue at a signalized intersection. It is assumed that the EV is part of the queue, and the objective is to 1) manage the queued traffic such that the EV clears the intersection as quickly as possible, 2) control downstream signalized intersections in the EVs path to provide an unobstructed path for the EV, and 3) provide minimal impact to surround traffic. The EV in this context can be an ambulance or a police car that may be notified to respond to an emergency.

To develop a solution to the basic problem defined above, a strategy is developed where traffic on one of the lanes is stopped at a critical location to allow the EV to change lanes so that it can travel at its desired speed which is higher than the speed of the normal traffic. This critical location is found by making use of the Kinematic Wave Model which is a linear model to describe traffic flow dynamics, i.e., the widely known model of Lighthill-Whitham-Richards [1-3], or simply LWR. This model arises from the conservation of vehicles principle and a fundamental diagram that relates flow to density, and is well known to reproduce important traffic flow features (e.g., shockwaves). The original LWR model has been extended in numerous studies to represent additional complexities [4-6].

In this paper, the information exchanged between vehicles is not explicitly modeled. It is assumed that all vehicles can receive the alert messages sent by the EV and comply with the

Manuscript received March 9, 2012. *Research supported by the Transportation Research Institute (TRI) and Virginia Modeling Analysis and Simulation Center (VMASC) at Old Dominion University.

C. A. Jordan is with Virginia Modeling Analysis and Simulation Center at Old Dominion University, 1030 University Blvd, Suffolk, VA 23435, USA (e-mail: cjord023@odu.edu).

M. Cetin is with the Department of Civil and Environmental Engineering at Old Dominion University, Norfolk, VA 23529, USA (phone: 757-683-2700; fax: 757-683-5354; e-mail: mcetin@odu.edu).

given instructions. The main contribution of this paper is to demonstrate how the LWR model can be utilized to make predictions about the evolution of the traffic over time and space and how such information can be used to alter the vehicle trajectories to optimize the travel time of the EV under congested conditions.

II. Problem Formulation

A. Shockwaves at a Signalized Intersection

As a traffic light alternates between green and red phases it creates discontinuities or shockwaves in the traffic stream. The LWR theory is particularly suitable to predict these shockwaves since the boundary conditions are well defined (e.g., the back ward moving shockwaves starts at the stop bar when signal phase changes). Furthermore, the queue discharging process at signalized intersections has been shown to be quite stable [7], which allows for reliable shockwave speed predictions. Fig. 1 shows a typical shockwave diagram for the formation and dissipation of a queue at a signalized intersection. The backward moving shockwaves start at t_R and t_0, which represent the back of the queue and the front of the queue, respectively. The speed of the shockwave for the discharging queue, w, is assumed to be known. In addition, the free-flow speeds (or desired speeds) of regular vehicles and the EV are also assumed to be known.

B. Clearing Path for the EV at One Intersection

Fig. 1 shows the shockwaves that define the boundaries of the queue, as explained previously, and the trajectory for the EV. The EV joins the back of the queue at some arbitrary time t_0 at which time the traffic light is turned to green. Obviously, the light can be turned to green earlier to shorten the delay incurred by the EV if the arrival of the EV had been known but this is not the focus of this paper and has been investigated before [8-9]. As previously discussed, this paper is focused on developing a strategy for a condition where the EV is already in a queue at some arbitrary time. Since the EV is already in a queue and there are two lanes, the EV needs to wait until the queue discharge shockwave reaches its location, or time t_2 in Fig. 1.

The proposed strategy involves stopping vehicles in the adjacent lane from the EV at a downstream point so that the lane clears and the EV can switch lanes to travel at a higher speed. Initially, the EV is in lane 1, and at a particular time (t_3) and location (x_L), it switches to lane 2. The EV sends a message to the vehicle at location x_L in lane 2 to stop until the EV completes the lane change. The distance x_L is found such that the preceding vehicle on lane 2 clears the intersection by the time the EV reaches the intersection. The two trajectories meet at the intersection at time t_4. As indicated in Fig. 1, the speed of the EV is initially at u (the desired speed of the rest of the traffic) until time t_3. After time t_3, the EV travels at its own desired speed, v. Assuming the speeds of the EV and the preceding vehicle in the queue and the position of the EV from the intersection are known, a formulation can be developed based on the shockwaves to determine which vehicle in lane 2 needs to stop for a short duration to make way for the EV.

The following variables are used in the formulation:

w: the shockwave speed for the discharging flow at the signalized intersection
v: desired speed of the EV
u: desired speed for regular vehicles

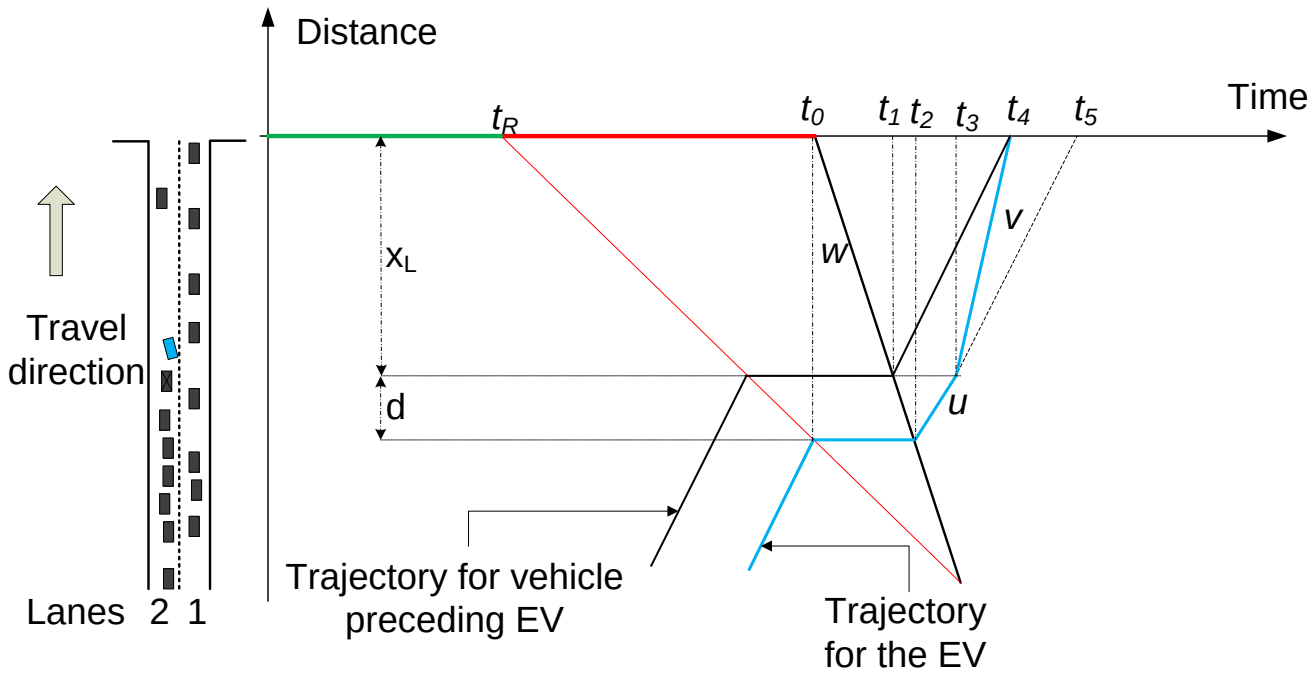


Figure 1. Shockwave diagram for a single queue at a traffic light and the trajectories of the EV and preceding vehicle.

x_L: the critical distance from the intersection to the point where queue needs to be split
d: distance from the EV (when in the queue) to the queue split point
D: distance from the EV (when in the queue) to the intersection

Assuming t_0 is zero, the departure time from the queue for the preceding vehicle can be found as follows.

$$t_1 = \frac{x_L}{w} \tag{1}$$

Likewise, t_2, departure time from the queue for the EV is

$$t_2 = \frac{x_L + d}{w}. \tag{2}$$

Furthermore, the time when the EV changes lanes and starts traveling at its own desired speed is found as,

$$t_3 = t_2 + \frac{d}{u} = \frac{x_L + d}{w} + \frac{d}{u} \quad . \tag{3}$$

Both the EV and the preceding vehicle reach the intersection at the same time t_4. For the preceding vehicle,

$$t_4 = t_1 + \frac{x_L}{u} = \frac{x_L}{w} + \frac{x_L}{u} \tag{4}$$

And, for the EV

$$t_4 = t_3 + \frac{x_L}{v} = \frac{x_L + d}{w} + \frac{d}{u} + \frac{x_L}{v} \tag{5}$$

Replacing d with "D-x_L" in (4) and (5) and solving them simultaneously for x_L results in the following relationship.

If the queue on lane 2 is split at location x_L then, the EV will be able to travel over distance x_L at its desired speed v.

$$x_L = D\frac{w^{-1} + u^{-1}}{w^{-1} + 2u^{-1} - v^{-1}} \tag{6}$$

This will results in time saving that is equal to the difference $t_5 - t_4$.

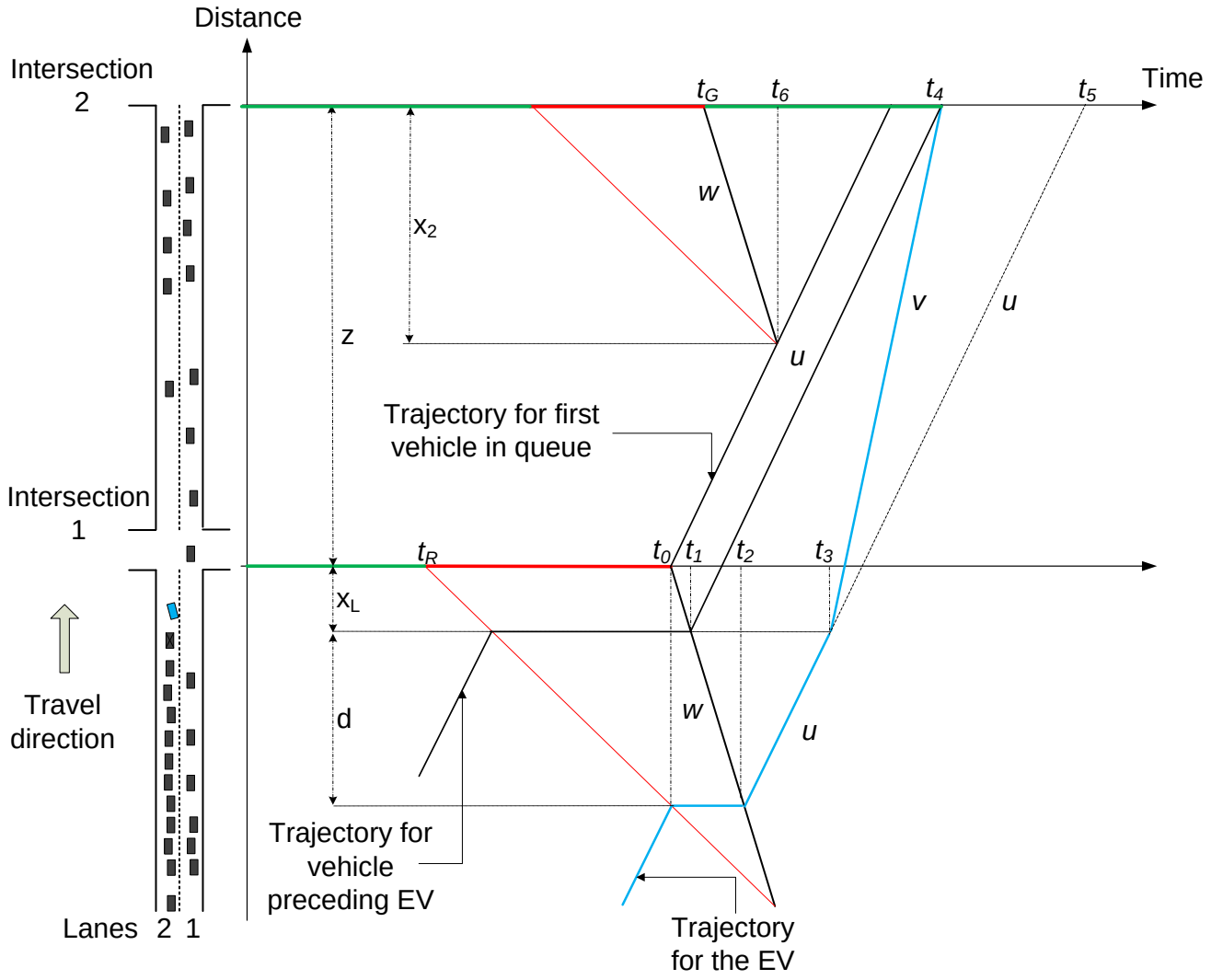


Figure 2. Shockwave diagram for a single queue at a traffic light and the trajectories of the EV and preceding vehicle to downstream intersection.

C. Clearing Path for the EV for Two Intersections

In this scenario, the destination of the EV is located at a point beyond two intersections. The shockwave corresponding to this scenario is shown in Fig. 2. Similar to Fig. 1, the trajectories of the EV and the preceding vehicle are shown. The idea is to find the critical point to stop traffic in the adjacent lane so the EV can change lanes to travel at its desired speed, *v*, through the first intersection and travel unimpeded through the second intersection. The formulation for calculating this critical point is similar to the previous scenario with the inclusion of a new variable *z*, (the distance between intersections).

Assuming t_0 is zero, the departure time from the queue for the preceding vehicle can be found as follows.

$$t_1 = z + \frac{x_L}{w} \tag{7}$$

The departure time from the queue for the EV is

$$t_2 = z + \frac{x_L + d}{w} \tag{8}$$

When the EV changes lanes and starts traveling at its desired speed can be found as,

$$t_3 = t_2 + \frac{d}{u} = z + \frac{x_L + d}{w} + \frac{d}{u} \tag{9}$$

The time at which the preceding vehicle of the EV and the EV reach the second intersection at the same time t_4 is calculated with the following two equations, respectively.

$$t_4 = t_1 + \frac{x_L + z}{u} = z + \frac{x_L}{w} + \frac{x_L + z}{u} \tag{10}$$

$$t_4 = t_3 + \frac{x_L + z}{v} = z + \frac{x_L + d}{w} + \frac{d}{u} + \frac{x_L + z}{v} \tag{11}$$

Replacing d with "D-x_L-z" in (10) and (11) and solving them for x_L provides the following equation.

$$x_L = D\frac{w^{-1} + u^{-1}}{w^{-1} + 2u^{-1} - v^{-1}} - z \tag{12}$$

If the queue on lane 2 is split at location x_L then the EV will be able to travel over distance $x_L + z$ at its desired speed v. This will result in time savings that equals the difference $t_5 - t_4$.

D. Calculation for Signal Time Changing

To clear potential vehicle queues at the downstream intersection, a formulation was performed to determine when the downstream traffic signal should turn green. This was determined by using the LWR method to specify the time at which the second intersection turns green in reference to when the EV enters the back of the queue (time t_0). Fig. 2 shows the shockwaves that define the boundaries of the queue and the trajectory of the EV. The formulation for calculating this critical time uses a new variable x_2 (the length of the queue at the downstream intersection).

The time the first vehicle departing from the first intersection reaches the back of the queue at the downstream intersection can be found as,

$$t_6 = \frac{z - x_2}{u} \tag{13}$$

The time it takes for the vehicle queue at the downstream intersection to discharge can be calculated with,

$$t_6 = t_G - \frac{x_2}{w} \tag{14}$$

where t_G is the time the signal turns green.

Solving (13) and (14) simultaneously for t_G provides the following equation.

$$t_G = \frac{z - x_2}{u} - \frac{x_2}{w} \tag{15}$$

Time t_G is relative to time t_0 and therefore the downstream intersection should turn green t_G seconds after time t_0.

III. Microscopic Simulation Modeling

A. Model Development

In order to evaluate the formulations developed in the previous section, a microscopic traffic simulation model was developed for the single intersection scenario discussed in Section II-B and for the two intersection scenario discussed in Section II-C. For this purpose, VISSIM, a widely known traffic simulator, was used for model development and analysis. VISSIM is a traffic micro-simulation tool and was chosen because of it flexibility in building complex traffic control logic and its ability to control certain aspects of the simulation with outside programming languages through the component object model (COM). The COM provides access for the developer to observe the modeled system after each simulation time step and make changes as necessary. Information for every road segment, vehicle, and traffic signal is available along with attributes assigned to these objects.

VISSIM allows the user to specify different vehicle types to be simulated on the roadway network simultaneously. For this study, two vehicle types were created to model different types of driving behaviors needed to simulate the traffic patterns in the presence of an EV. The use of the different vehicles types allowed for specific information to be sent to and received from vehicles during the model run. The two vehicles types used were "EV" and "NonEV." The information exchange between vehicles was not modeled explicitly but the control logic assumes that such exchange takes place without any communication delays or failures.

The base model consisted of a simple straight two-lane roadway approximately 2 kilometers long. The roadway

Figure 3. The EV entering the back of queue

Figure 4. The EV moves from lane 1 to lane 2

contained two intersections and no other obstructions. Vehicles loaded onto the network traveled the length of the roadway and stopped at the signalized intersections to form vehicle queues. VISSIM uses stochastic assignments of numerous parameters in the simulation based on distributions rather than exact values. The desired speed distribution is one of these parameters. The user can specify the shape of the distribution curve and the minimum and maximum values. For this study the range of the speed distribution was reduced so the only possible desired speeds were set to a small range. This was performed so that the model runs could be compared with the LWR model. The LWR assumes constant speeds and does not account for acceleration and deceleration behavior. For the base model, the desired speed distribution for "NonEV" vehicles was coded with a uniform distribution between the values of 50 km/h and 50.1 km/h. Traffic entered the network with varying flow rates between 800 and 1575 vehicles per hour per lane to create traffic congestion along the roadway and at the signalized intersections. The desired speed for the EV was set at 50 km/h prior to arriving at the vehicle queue at the first intersection. It was then changed to 80 km/h when it stopped at the back of the queue.

To make the model as simple as possible, only passenger cars were included in the vehicle composition. This made the evaluation of the data easier as it removes the effects of other vehicle types, such as heavy goods vehicles, might have on travel time.

VISSIM uses the Wiedemann 74 and Wiedemann 99 models to model car following behaviors. These models use the idea that vehicles can be in one of four driving states. These include free driving, approaching, following, and breaking. The model parameters can be modified by the user to adjust the specific driving behavior that is desired. In the study model, the driving behavior of the vehicles was set to the default values. Future work will include modifications of these values to more accurately represent the behavior of EV and other vehicles. In addition to car following, VISSIM also has parameters for lane changing. The parameters can be modified by the user to adjust the conditions for which a vehicle changes lanes. The parameter changes can affect the acceptable gap size for the lane-changing vehicle. In the model, the lane changing parameters were also kept at the default settings.

The model was run 50 times for each scenario to provide a large sample size of data. Each model run had a different random seed so that the vehicles that entered the network did so differently than the other model runs. In addition, the EV was added to the network at random times during the simulation so that its position in the vehicle queue at the first intersection would vary from between each model run.

The methodology used to control the vehicles on the network for the scenario in Section II-B is as follows. The EVs were added to the network at random times during the model run. After an EV entered the network, it traveled the length of the roadway and joined the back of the queue at the first signalized intersection. Fig. 3 provides a picture of a model run where the EV (white vehicle) enters the back of the queue. At this point, a message was sent alerting the vehicles in the adjacent lane that an EV is present. Specific vehicles were instructed to stop which were located upstream of distance "x_L" from the intersection, calculated based on (6). The vehicles that were instructed to stop are shown in Fig. 3 in red. In addition, the traffic signal was instructed to turn green. As the EV passed the queue of stopped vehicles in lane 2, it changed lanes to have an unobstructed path to the intersection. Fig. 4 provides a picture of a model run where the EV has changed lanes and is proceeding to the intersection. As soon as the EV changed lanes, the stopped vehicles were instructed to proceed at their original desired speed.

The methodology used to control the vehicles on the network for the scenario in Section II-C is similar to Section II-B and is as follows. The EVs were added to the network at random times during the model run. After an EV entered the network, it traveled the length of the roadway and joined the back of the queue at the first signalized intersection. At this point, a message was sent alerting the vehicles in the adjacent lane that an EV is present. Specific vehicles were instructed to stop which were located upstream of distance "x_L" from the first intersection, calculated based on (12). In addition, the traffic signal was instructed to turn green at the first intersection. The second intersection, downstream of the first, was instructed to turn green t_G seconds after the EV entered the back of the queue, calculated based on (15). As the EV passed the queue of stopped vehicles in lane 2 of the first intersection, it changed lanes to have an unobstructed path through the intersection. As soon as the EV changed lanes, the stopped vehicles were instructed to proceed at their

original desired speed. The preemptive changing of the downstream traffic signal at the second intersection allowed for vehicles queued to be discharged just prior to the arrival of the discharging vehicles from the first intersection. This allowed for the EV to travel through the first intersection and to the second intersection at its desired speed.

B. Simulation Results

After each run for the scenarios described in Section II-B and Section II-C was completed, vehicle speeds, locations, and the time at which vehicles passed certain locations were recorded. In particular, the time instant when the EV joined the back of the queue (time t_0), the time at which the EV departed from the queue (time t_2), the time at which the EV changed lanes (time t_3), and the time at which the EV entered the first intersection for the scenario in Section II-B (time t_4) and the second intersection for the scenario in Section II-C (time t_4) were recorded. Model runs were completed for both with and without the implementation of the strategy presented to compare the EV travel times and determine the time savings.

Table I and Table II provide samples of the collected data and comparison between the runs for the scenario in Section II-B and Section II-C, respectively. The first column shows the distance D in meters (D - z for the scenario in Section II-C) at which point the EV joined the back of the queue and sent a message to those vehicles in lane 2 that need to stop. The time points when the EV reaches a critical point are also tabulated. Fig. 1 can be consulted to relate these points to the vehicle trajectory and the formulation. As it can be observed, the proposed strategy improves the travel time of the EV – compare values under columns t_4 and t_5. The time savings are indicated in the last two columns. The percent saving is found by dividing the time savings by the travel time ($_(t_5 - t_4) /_(t_5 - t_2)$). It can also be observed that the percent saving increases as D increases.

Fig. 5 summarizes the results for the 50 runs for the scenario in Section II-B. It shows the percent reduction in travel time for the EV (the same information reported in the last column of Table I) versus the distance D. The data indicates that the time savings varies between 4% and 36% depending on the distance from the first intersection (D). It was noticed that as D increases beyond 600 meters, the increase in percentage savings begins to level off.

Fig. 6 summarizes the results for the 50 runs for the scenario in Section II-C. It shows the percent reduction in travel time for the EV (the same information reported in the last column of Table II) versus the distance D - z. The data indicates that the time savings varies between 2% and 34% depending on the distance from the first intersection (D - z). It was noticed that as D – z increases beyond 20 meters, the increase in percentage savings begins to level off.

Theoretically, the percent reduction in travel times should not change as D varies. Based on the used desired speeds, the theoretical percent reduction in travel times should be around 34%. However, as mentioned above, the LWR formulation

TABLE I. SAMPLE DATA FROM THE VISSIM SIMULATIONS SCENARIO SECTION II-B

Distance From First Intersection (m)	**Critical time points when the proposed strategy is implemented**				**Without any strategy**		**Savings**	
D	t_0	t_2	t_3	t_4	t_0	t_5	Time	Percent
153	02:06	02:32	02:35	02:42	02:06	02:43	00:01	8%
274	02:37	03:28	03:34	03:45	02:37	03:48	00:03	15%
416	03:15	04:33	04:37	04:55	03:15	05:01	00:07	24%
548	03:20	04:59	05:05	05:24	03:20	05:35	00:11	31%
972	04:36	07:40	07:51	08:25	04:36	08:45	00:21	31%

Note: Times are in minute : second format.

TABLE II. SAMPLE DATA FROM THE VISSIM SIMULATIONS SCENARIO SECTION II-C

Distance From First Intersection (m)	**Critical time points when the proposed strategy is implemented**				**Without any strategy**		**Savings**	
D - z	t_0	t_2	t_3	t_4	t_0	t_5	Time	Percent
10	04:39	04:43	04:46	05:39	04:39	05:49	00:10	14%
16	04:41	04:46	04:51	05:38	04:41	05:52	00:14	21%
35	05:00	05:08	05:14	06:00	05:00	06:16	00:16	24%
42	05:01	05:09	05:16	05:58	05:01	06:18	00:20	30%
67	05:15	05:28	05:36	06:15	05:15	06:39	00:24	33%

Note: Times are in minute : second format.

does not account for acceleration and deceleration behavior and assumes that vehicles can change speed instantaneously. When the distance is short the acceleration and deceleration can impact the travel times more substantially. However, as the D increases the savings in travel times approach the theoretical value of 34% (see Fig. 5 and Fig. 6).

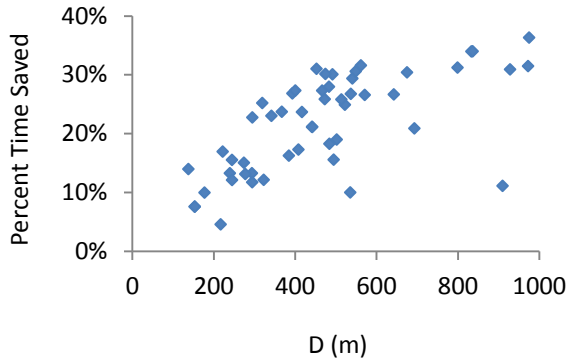


Figure 5. Percent improvement in travel time for the EV at varying distances for one intersection

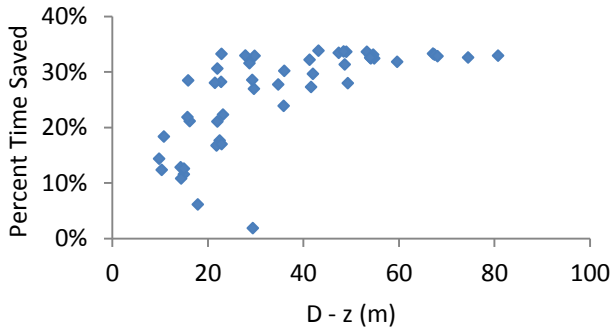


Figure 6. Percent improvement in travel time for the EV at varying distances for two intersections

IV. Conclusion

This paper presents a new strategy to enable emergency response vehicles to traverse a congested link and through two intersections as quickly as possible. The application of the concept is illustrated for two scenarios. The first scenario investigates one intersection on a two-lane road where the vehicular queue on one of the traffic lanes is split at a critical point so that an EV can proceed at its desired speed while the disruption to the rest of the traffic is minimized. The second scenario investigates two intersections on a two-lane road where the vehicular queue on one of the traffic lanes is split at a critical point and the downstream traffic signal is turned green at a critical point to allow an EV to proceed at its desired speed through both intersections. The formulations are developed based on the kinematic wave model for traffic flow to predict the queuing behavior at signalized intersections. The proposed method was simulated in VISSIM for validation and evaluation. The results show that this strategy can shorten the trip times for EVs significantly.

The research work presented here will be extended to model other types of strategies to improve the travel time of emergency response vehicles. These will then be tested in VISSIM on larger networks to evaluate benefits on more realistic travel trajectories.

References

[1] M. J. Lighthill and G. B. Whitham, "On Kinematic Waves. I. Flood Movement in Long Rivers," *Proceedings of the Royal Society of London. Series A. Mathematical and Physical Sciences*, vol. 229, pp. 281-316, May 10, 1955 1955.

[2] M. J. Lighthill and G. B. Whitham, "On Kinematic Waves. II. A Theory of Traffic Flow on Long Crowded Roads," *Proceedings of the Royal Society of London. Series A. Mathematical and Physical Sciences*, vol. 229, pp. 317-345, May 10, 1955 1955.

[3] P. I. Richards, "Shock Waves on the Highway," *Operations Research*, vol. 4, pp. 42-51, 1956.

[4] J. van Lint, *et al.*, "Fastlane: New Multiclass First-Order Traffic Flow Model," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2088, pp. 177-187, 2008.

[5] H. Zhang and W. Jin, "Kinematic Wave Traffic Flow Model for Mixed Traffic," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 1802, pp. 197-204, 2002.

[6] G. C. K. Wong and S. C. Wong, "A multi-class traffic flow model – an extension of LWR model with heterogeneous drivers," *Transportation Research Part A: Policy and Practice*, vol. 36, pp. 827-841, 2002.

[7] H. X. Liu, *et al.*, "Real-time queue length estimation for congested signalized intersections," *Transportation Research Part C: Emerging Technologies*, vol. 17, pp. 412-427, 2009.

[8] E. Oliveira and N. Duarte, "Making way for emergency vehicles [traffic congestion control]," in *European Simulation and Modelling Conference 2005. ESM 2005, 24-26 Oct. 2005*, Ostend, Belgium, 2005, pp. 128-35.

[9] G. Unibaso, J. Del Ser, S. Gil-Lopez, and B. Molinete, "A Novel CAM-based Traffic Light Preemption Algorithm for Efficient Guidance of Emergency Vehicles," in *2010 13th International IEEE Conference on Intelligent Transportation Systems (ITSC 2010), 19-22 Sept. 2010*, Piscataway, NJ, USA, 2010, pp. 74-9.

Evaluation of Traffic Assignment with Road Capacity Uncertainty

Linmin Pei

***Abstract*— this paper explores the impact of road capacity uncertainty on the evaluation of traffic assignment. It also shows the total system travel time of traffic system at the expected road capacity value is less than the expected total system travel time of the system in terms of Jensen's Inequality. An example has been provided to demonstrate the impact of road capacity uncertainty to the traffic assignment.**

***Keywords*— Traffic Assignment, Road Capacity Uncertainty, Jensen's Inequality**

I. Introduction

The Bureau of Public Road function (BPR function) is widely used to build traffic system models [1]. It contains link flow, link parameters, free flow travel time and road capacity. The total system travel time of the traffic assignment is impacted by the BPR function and the O-D demand. Waller et al discuss the impact of demand uncertainty to the evaluation of the traffic assignment [2].The total system travel time of the traffic system also is affected by BPR function's change that maybe results from road capacity. In practice, road capacity is not a fixed value for a link in transportation network. It depends on some other factors, such as the traffic situation, road situation, and weather etc. With the road capacity uncertainty, the BPR function is also uncertain. Generally, BPR function is nonlinear and convex. According to Jensen's Inequality [3], the average of the function (the BPR equation) is greater than the function of the average (of the links considered). This paper presents the total system travel time of traffic assignment with road capacity uncertainty based on the total system travel time. We analyze a simple transportation network with varying road capacity to measure the impact on traffic assignment.

II. Problem

Traffic assignment models are used to mimic the decisions that people make. The validity of the model is the key to properly estimate the network system. One of the factors that impact the model is road capacity. In reality, road capacity is not constant, but rather a random variable because it is affected by traffic situation, road situation, and weather, etc. The evaluation of the traffic assignment is affected by the uncertainty of BPR function because of the road capacity uncertainty.

III. Mathodology

For a network, the Dijkstra's algorithm is used to get the shortest path from a source node to sink node, and keeping the link flow of the network updated in term of Frank-Wolfe algorithm until each route of an O-D pair has same travel time, i.e., the system reach users' equilibrium. To test the impact of road capacity uncertainty of the system, a number of road capacities of a link will be applied into the system. The applied system performance (the expected value of total system travel time) of the traffic assignment can be evaluated and the system performance (total system travel time at the expected value of road capacity) also can be obtained after applying the expected value of road capacity into the system. The relationship of the two values is found after making a comparison between them.

The commonly used function to calculate travel time on a traffic network is known as Bureau of Public Roads function [1]:

$$t_a = t_a^0 \left[1 + \alpha \left(\frac{x_a}{C_a^{'}} \right)^{\beta} \right] \quad (1)$$

Where:

t_a and x_a are the travel time and flow on link a respectively,

t_a^0 is free travel time of link a

$C_a^{'}$ is road capacity of link a

α and β are model parameters.

To evaluate the total system travel time of the traffic network, the mean of road capacity is calculated by using:

$$E(C_a) = \frac{1}{n}\sum_{i=1}^{n} c_a^i \quad (2)$$

And the mean of the total system travel time of the traffic assignment (E (T)) is equal to the average total system travel time based on road capacity:

$$E(T) = \frac{1}{n}\sum_{i=1}^{n} T(i) \quad (3)$$

Where:

E (C_a) is the mean of road capacity.

C_a^i is the specific value of road capacity.

T is total system travel time.

E(T) is the mean of total system travel time.

T(i) is the total system travel time at the specific road capacity (C_a^i).

IV. EXPERIMENT

We have implemented a program in C++ based on Dijkstra's algorithm and Deterministic User Equilibrium. By using the program, the system reaches user equilibrium that each route for an OD pair has same cost. A small network will be applied into the program to test the impact of road capacity uncertainty. The network has four nodes and five links.

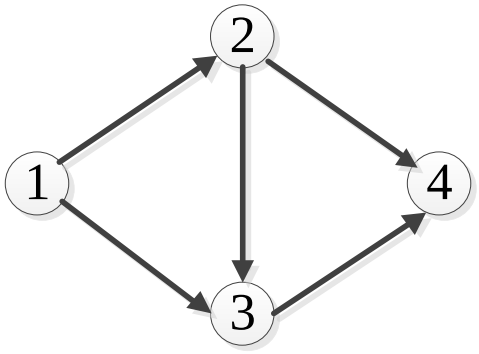


Fig. 1 Network of the traffic assignment

For this transportation network, assuming the network data is listed in Table 1.1 and the OD demands are shown in Table 1.2:

Table 1.1 Network data of the traffic assignment

Link No.	Head	Tail	t_a^0	C_a	α	β
1	1	2	3	c_a^*	0.15	4
2	1	3	1	3	1	3
3	2	3	1	2	0.15	2
4	2	4	1	2	0.15	2
5	3	4	2	2	1	4

Where:

t_a^0= free flow travel time (minutes).

C_a= road capacity (vehicles/hour).

α, β = parameters (typically0.15 and 4, respectively).

The O-D matrix data of the network is shown in Table 2.

Table 1.2 O-D matrix data of the network

Node	1	2	3	4
1	0	0	20	10
2	0	0	10	5
3	0	0	0	5
4	0	0	0	0

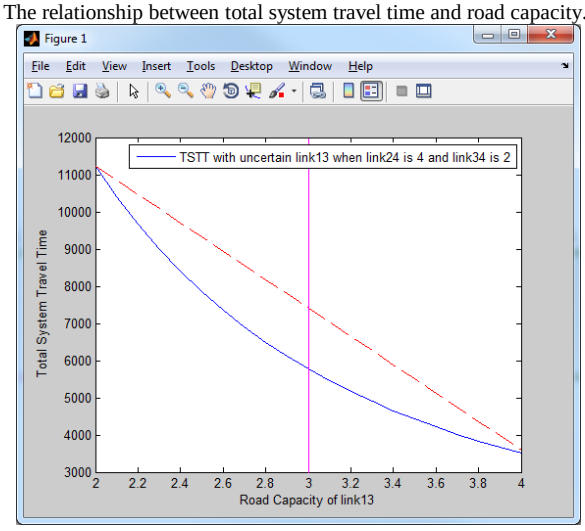
V. RESULTS

All the network data and O-D matrix data are applied into the program with different road capacity(c_a^*) of link12, the total system travel time is calculated and the result is listed into the Table 1.3.

Table 1.3 TSTT for particular road capacity (iteration = 2000)

Iteration=2000		
link24	4	
link34	2	
C12	2	11228
	2.1	10403.6
	2.2	9665.07
	2.3	8998.33
	2.4	8397.07
	2.5	7853.71
	2.6	7360.1
	2.7	6908.81
	2.8	6498.96
	2.9	6124.32
	3	5776.86
	3.1	5462.19
	3.2	5171.63
	3.3	4902.92
	3.4	4656.41
	3.5	4426.58
	3.6	4215.16
	3.7	4018.42

Fig. 2 The relationship between total system travel time and road capacity.



According to the result shown above, the mean of the road capacity (E (C_a^i)) is equal to 3.15.

And the mean of the total system travel time of the traffic assignment (E (T)) is equal to 6430.661.

The total system travel time of the traffic assignment at the expected value of road capacity is estimated by applying the mean of road capacity into the network. The total system travel time at the expected road capacity of link12 is equal to:

$$T\left(E\left(C_a^i\right)\right) = T(3.15) = 5960.49$$

After comparing to the mean of the total system travel time of the traffic assignment (E (T)) and The total system travel time at the expected road capacity (T (E (C_a^i))), the result is obvious:

$$T\left(E\left(C_a^i\right)\right) < E(T)$$

From the relationship shown above, it finds that the total system travel time of the traffic assignment at the mean of road capacity is less than the mean of the total system travel time of the network, and it shows that the consistency of Jensen's Inequality in this situation. Therefore, for a transportation network, using a fixed road capacity to evaluate the total system travel time underestimates the real total system travel time.

VI. Conclusion

For a traffic assignment, the road capacity uncertainty results in the uncertainty of the BPR function, and it impacts the total system travel time of the network. It has been shown using the expected value of road capacity in networks evaluated with traffic assignment underestimate evaluation of the system for the traffic assignment. The performance of the network is underestimated by using fixed road capacity to evaluate the system.

References

[1] Y. Sheffi, Urban Transportation Networks: Equilibrium Analysis with Mathematical Programming Methods, Englewood Cliffs, NJ.: Prentice-Hall, 1985.

[2] S. T. Waller, J. L. Schofer and A. K. Ziliaskopoulos, "Evaluation with Traffic Assignment Under Demand Uncertainty," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2001, no. 1771, pp. 69-74, 2001.

[3] M. D. Perlman, "Jensen's Inequality for a Convex Vector-Valued Function on an Infinite-Dimensional Space," *Journal of Multivariate Analysis*, vol. 4, no. 1, pp. 52-65, 1974.

Linmin Pei, a Ph.D. student in Department of Modeling, Simulation & Visualization Engineering at Old Dominion University.

Autonomous Mission Execution for Quad-Copter

Hamdi Kavak
Modeling, Simulation & Visualization Engineering
Old Dominion University
Norfolk, United States

Yiannis Papelis, Ph.D.
Virginia Modeling, Analysis and Simulation Center
Old Dominion University
Suffolk, United States

Abstract – this paper presents autonomous mission execution system developed for Quad-Copter. Quad-Copter is a part of Command and Control system that manages live-virtual-constructive simulation environment. For test case, patrol mission is implemented and tested on ground vehicle.

I. INTRODUCTION

Unmanned Aerial Vehicle (UAV) is a type of flying machine that performs various tasks autonomously, semi-autonomously or by remote controller. There are various sizes and designs in UAVs. UAVs may vary from couple of inches (robot quadrotors built in University of Pennsylvania [1]) to tens of feet long (Eitan by Israel Aerospace Industries). Size and design of the UAV is dependent on its production purpose.

Quad-Copter (QC) is an aerial vehicle that can be moved by four rotors. Rotors are placed symmetrically on top of cross shaped body. QC can move in any direction by running rotors with various combinations.

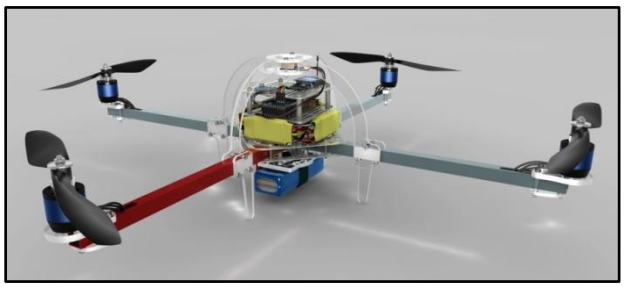


Figure 1: ArduCopter open-source Quad-Copter solution

The purpose of this study is designing a system for QC that is capable of executing predefined missions autonomously.

II. METHODS

A. Positioning the Vehicle

The vehicle's position is defined in three dimensions: Latitude, longitude and altitude.

Longitudes (meridians) are imaginary lines which are drawn on surface of earth from North Pole to South Pole. Longitudes divide the earth into 360 equal slices.

Latitudes (parallels) are imaginary circles on the surface of the Earth. Latitudes are parallel to the equator (0 degree latitude).

Altitude is the vertical distance between ground and object.

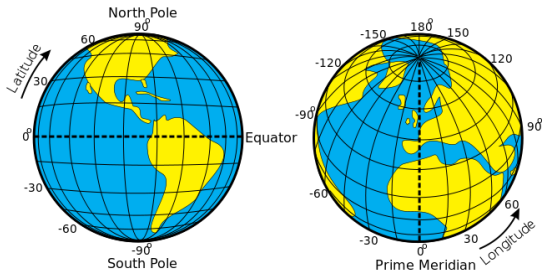


Figure 2[1]: Latitude and longitudes on earth

[1] Image Source: http://en.wikipedia.org/wiki/File:Longitude_(PSF).png

Vehicle's current position is constantly provided by Global Positioning System (GPS) receiver which is connected to *High Performance CPU*.

Latitude and longitude of a location is unique, but it needs to be converted into known units for easier calculation. 2-dimensional Cartesian coordinate system *UTM* is more suitable solution for this application [2]. In this system earth is divided into 60 zones. Each *UTM* zone is segmented into 20 latitude bands. Each point in the zone expressed as (x, y) coordinates (A.K.A easting, northing). We used centimeter unit for easting and northing coordinates.

B. Communication with the Vehicle

In order to continuously perform autonomous missions, a communication system is needed. One of the command and control system components called "Ground Station" is the communication layer that talks wirelessly with the vehicle. Messages can be sent in both ways. Sample message flow between ground station and vehicle is illustrated in Figure 3.

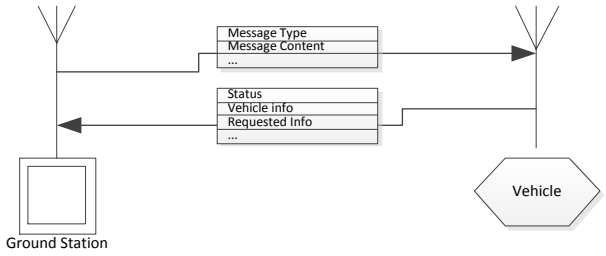


Figure 3: Communication between ground station and vehicle

The mission transmission procedure can be broken into the following steps:

1) The new mission is encoded in a predefined format and then sent by ground station.
2) *Interface mCPU* reads new messages through *Wireless Communication Module*.
3) *Interface mCPU* board places the mission into a specific memory location.
4) *High Performance CPU* receives the mission from memory location and parses it.
5) Mission is added into the first-in first-out (FIFO) mission queue.

Vehicle's components and communication between these components are shown in Figure 4.

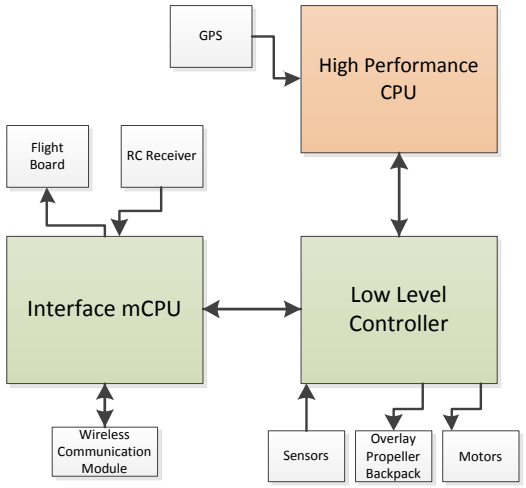


Figure 4: Vehicle components

C. Mission Creation and Execution

High level mission execution is handled by *High Performance CPU*. Once a new mission is detected, it is first set up with given parameters. Then waypoints are generated depending on parameters and mission type. Waypoints are defined as easting, northing points that vehicle should follow to accomplish the mission.

Once all waypoints are created, first waypoint is set as target point. Autopilot is notified for current target point in order to make vehicle move there. Physical movements are handled by autopilot system that is located in *Low Level Controller*.

On completion of visiting all waypoints, the mission is marked as accomplished. Autopilot is set free until new mission is received. Lastly, report is generated and sent to ground station. Simplified mission creation and execution flow is illustrated in Figure 5. This flow is applied in experiment part.

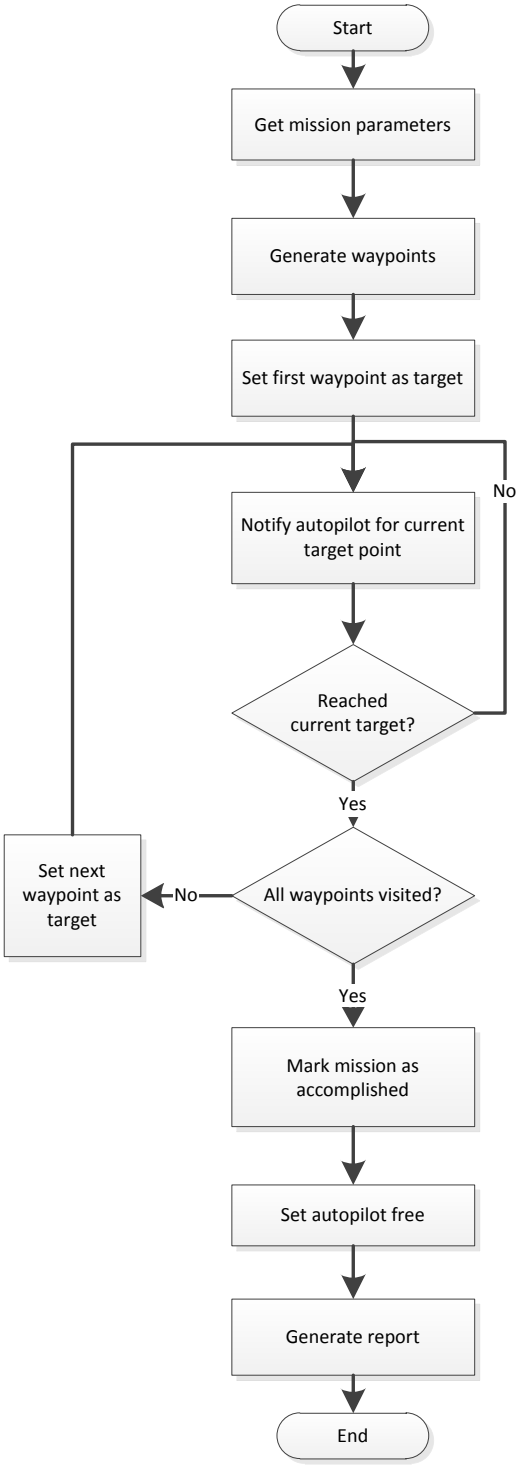


Figure 5: Mission execution flow

III. EXPERIMENT: *Patrol Mission*

The QC is one of the live components in live-virtual-constructive (LVC) simulation environment. For test purposes, ground vehicle (GV) is regarded to be more convenient than CQ. Because, having an accident in QC experiment might be more costly if you compare it with ground vehicle.

For test case *patrol* mission is implemented on ground vehicle. Patrol mission simply generates waypoints around the target point. Vehicle should visit all waypoints to accomplish the mission.

Patrol mission has five parameters. These are; *target easting*, *target northing*, *radius*, *number of waypoints* and *error margin*. Waypoint generation is the first step after parameter initialization (see Figure 5). Waypoint generation is carried out in following approach:

angle = 2 π / number of way points*

for *i = 1 to number of way points*
*$waypoint_{i,\ easting}$ = target easting + cos(i*angle)*radious*
*$waypoint_{i,\ northing}$ = target northing + cos(i*angle)*radious*
end

Subsequently, first waypoint was set as target point. Then autopilot notified until vehicle reaches the current target point. It is not always possible to move the vehicle to the exact target point. Error margin is defined to overcome this issue. If the vehicle's position is in the target range then it is considered to be reached to the target point. Following code shows how target range is calculated.

min_easting = target easting point – error margin
max_easting = target easting point + error margin
min_northing = target northing point – error margin
max_northing = target northing point + error margin

if *min_easting < current easting < max_easting* ***AND***
min_northing < current northing < max_northing ***THEN***
"GV reached the target point range"
ELSE
"Not in target point range. Keep moving"

In this patrol mission experiment, we used following parameters (in centimeter).

- Target easting: 37358881
- Target northing: 408126461
- Radius: 1000
- Number of waypoints: 5
- Error margin: 200

Target point and generated waypoints are illustrated in **Figure 6**. If a circle is drawn around the target point with 10 meter radius, these five points will be located on the circle. The distance between adjacent waypoint is equal.

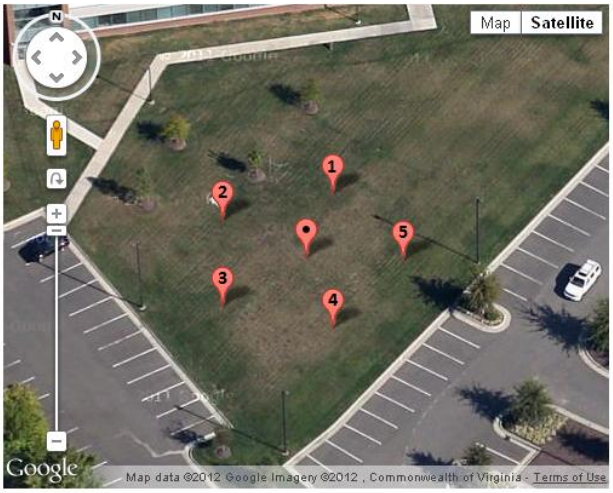


Figure 6[2]: Target point at center and generated waypoints around. (This area is the backside of Virginia Modeling Analysis and Simulation Center.)

The path taken by the ground vehicle is shown in Figure 7. You can see that vehicle does not exactly visiting the waypoints. It is passing close to the waypoint.

Figure 7[2]: The path taken by the vehicle in order to accomplish mission.

IV. CONCLUSION

This paper has presented an autonomous mission execution system. Presented system is successfully applied to LVC component called ground vehicle. New mission implementations will be applied for future study. Later, this system will be transferred to the Quad Copter.

For this experiment, clear view of the field is required. It is possible to overcome this drawback by adapting IR sensors.

REFERENCES

[1] Michael, N., Mellinger, D., Lindsey, Q., and Kumar, V., "The GRASP multiple micro UAV testbed," IEEE Robotics and Automation Magazine, 2010.

[2] NGA Coordinate Systems Analysis Team (CSAT), "A simplified definition and explanation of UTM and related systems", 2006.

[2] Image was generated by using Google Maps © API.

Demand responsive signal control strategy for over-saturated signalized intersection

Rahul M. Rajbhara*, Mecit Cetin**

Abstract— **Most of the existing vehicle detection systems, namely inductive loops, video cameras etc. are incapable of measuring the queue length information at signalized intersections in real-time. Some modern detection technologies however currently provide this capability. Probe vehicles equipped with wireless communications, particularly under the Vehicle Infrastructure Integration (VII) vision[1,2], is one good example of a modern detection technology. This paper presents a demand responsive traffic control strategy for an over-saturated signalized intersection. The control strategy incorporates real-time queue length information (derived from use of probe vehicles) into the signal control logic, based on which, maximum green times for all the phases are computed in each cycle. Use of variable maximum green times from cycle to cycle, helps in efficient allocation of the intersection capacity when traffic demand fluctuates. The proposed methodology is implemented for a single isolated typical 4 legged intersection and then evaluated in a microscopic traffic simulation environment (VISSIM). Three different methods are proposed and evaluated based on few important measures of effectiveness, such as the average delays, average queue size and travel time. Various numerical experiments are conducted where traffic demand is increased and decreased to replicate the demand fluctuations on field.**

Index Terms— **Delay, maximum green, queue length, signalized intersection, traffic signal control, VISSIM**

Acronyms: **QL - queue length, G_{max} - maximum green**

I. Introduction

Urban traffic congestion has been one of the major issues in transportation for several years. Congestion occurs on fixed capacity road networks when traffic grows beyond about 90% of capacity [3]. It can be classified into two main types namely, recurrent congestion and non-recurrent congestion. The former is caused by peak-period/rush hour traffic, while the later results from seven primary sources namely, bottlenecks, incidents, work zones, severe weather conditions, poor signal timing, special events and fluctuations in normal traffic. The three most prominent strategies known to Transportation Engineers, to deal with non-recurrent congestion are increasing road capacity; efficient traffic operation; and travel demand management (TDM). With the continuing socio-economic growth in metropolitan cities across the U.S, most of the existing transportation facilities are being used to their full capacities. Moreover, the scope of any further development of infrastructure is becoming unfeasible, due to the limitations of space and cost. Many travel demand management policies, like flexible work hours have been deployed in the busiest metropolitan areas of the country; nonetheless, they were found to have no substantial impact on the system. Moreover, efficient traffic operation which focuses on optimal control of traffic signals is perhaps the most effective method of relieving congestion in larger metropolitan areas. Intersections are the vital nodal points in transportation and efficient control of their traffic signals has a greater influence on the performance of entire network.

Thus, many research efforts over the past few decades have been directed towards the development of new strategies that can relieve the traffic engineer from the burden of data collection, and, at the same time, provide an improved level of traffic performance under over-saturated traffic conditions. This eventually resulted in the inception of traffic responsive signal control. Traffic Responsive Signal Control Strategy was initially developed in the form of first-generation control by UTCS (Urban Traffic Control System) of U.S Department of Transportation in the early 1970's [6]. The objective of this research project was testing and development of three generations of control strategies. In the first-generation control (1-GC), optimization was done offline, but implemented online. In the second-generation control (2-GC), both optimization and implementation was done online; however, to avoid transition disturbances, new timing plans were not implemented more often than every 10 minutes. The improvement of 2-GC strategy led to the third-generation control (3-GC), which can implement and evaluate a fully responsive, online traffic control system. In this strategy, the timing plans could be revised more frequently (3 – 5 minutes). Although, all three control strategies were designed to provide an increasing degree of traffic responsiveness, they could not fulfill the objective, due to an error in the procedure of the measurement-prediction cycle [5]. The systems responded to hypothetical conditions rather than actual traffic conditions, because of this error. Another setback of these strategies, addressed by Kreer[6], is the computation time needed to match the estimated traffic condition to the prevailing traffic condition for which the timing plan is assigned. Jiann-Shiou Yang [13] developed traffic responsive control to optimize intersection split times after special events at the City of

Manuscript received March, 30 2012. *R. Rajbhara, a graduate student at the Department of Civil and Environmental Engineering, Old Dominion University, Norfolk, VA – 23529 (e-mail: rrajb001@odu.edu)
**M. Cetin (corresponding author), Assistant Professor is with the Civil & Environmental Engineering Department, and the Transportation Research Institute, Old Dominion University, Norfolk, VA 23529 USA. (E-mail: mcetin@odu.edu).

Duluth Entertainment Convention Centre (DECC) area in Duluth, Minnesota. The strategy provided effective traffic signal timing (over a 30-minute time period) for high volumes of traffic resulting from DECC special events, such as hockey games, concerts, graduation ceremonies, conventions, etc. Optimization of splits was carried out in Synchro, which in reduced total delays. This approach however did not test the peak and off-peak conditions for robustness.

Recently, Comert et.al (2008) proposed new signal control logic, for a simple isolated intersection (two one-way streets) that utilizes the real-time queue length information. Unlike the typical actuated signal control, wherein the maximum green times are pre-determined and pre-set; their new control logic implemented the use of a simple formula, to estimate maximum green times. Comert.et.al calibrated a new equation to estimate the maximum green time for a phase based on the queue length information of the intersection. For the simple intersection with two one-way streets used in their research study, the maximum green for the major and minor street was set to be a function of queue length. The efficiency of this new control logic was then evaluated against the typical actuated signal control by conducting numerous experiments. Both the signal control methods were optimized for three different demand profiles (increasing or decreasing the flow-rates by a fixed percentage) over one-hour period. This study however was confined to a simple 2 phase intersection, to avoid the intricacy of a 3 or more phase signalized intersection.

This paper is intended to overcome the drawbacks of the former study (by Comert et.al) on incorporating the queue length information in actuated signal control. The major contributions of this research study over the previous research attempt are a) control logic is developed for a typical 8 phase signalized intersection, as opposed to a 2 phase intersection with two one-way streets; b) more accurate implementation of the control logic in VISSIM; c) testing new demand scenarios; d) revised formula for computing the maximum green (G_{max}) for each phase.

II. Problem Statement

Pre-timed (fixed-timed) and actuated signal control represents the two basic types of traffic control systems. Pre-timed control system is by far the most widely implemented method, where the signal variables like cycle length; phase splits; offsets etc. are predetermined based on historical volume data. Actuated signal control is a more advanced signal system which makes use of the vehicle detection technology (inductive loops, video cameras) to either extend the green while in green interval or call the controller while in red interval, and reacts to the prevailing traffic, by changing green times, splits, cycle length etc. Actuated control however works within the constraints of a maximum and a minimum green beyond which it begins to act as a typical fixed time signal control. Thus, due to the limited flexibility in their control logic, pre-timed and actuated control systems cannot fully accommodate the time varying traffic demands. There is a strong need of implementation of advanced signal control systems which can change the phase lengths in real-time to maximize the throughput and minimize delays at intersections. This type of control system can be highly demand responsive, and can efficiently respond to non-recurrent traffic congestion[5].

III. Methodology

Two models are created and tested for this research study, namely, the typical actuated method and the proposed demand-responsive traffic control. The following components are crucial in the model development.

A. *Phase sequencing logic (phase skipping).*

The phase sequence logic for a fully-actuated control is implemented in the logic and is illustrated in Figure1. Stage 1, which comprises of phases 1 and 5 is the start-up stage. The likelihood of the next stage is assessed using the detector readings. For instance, at a time step *t*, if the current stage is 4 and the detectors linked to phases 3 and 7 are both occupied, then the next stage will certainly be stage 5. If only phase 3 detector is occupied, then the anticipated next stage is stage 7. Stages 6 and 7 (or 2 and 3) are non-concurrent stages, meaning that the traffic controller can process only one of the two stages during a cycle. The following stage is the last stage in a cycle (and barrier 2) i.e. stage 8, irrespective of the detector readings. The control is then transferred to barrier 1. At the end of stage 8, the logic is returned to stage 1, 2, 3 or 4. This sequence is repeated every cycle until end of simulation.

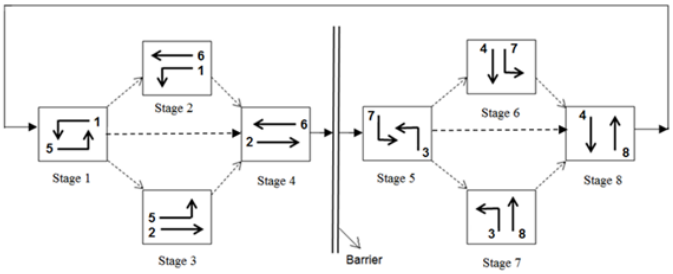


Figure 1 Phase sequencing logic

B. *Queue length measurement technique:*

In this study, probe vehicles are used to get the real-time queue lengths on various approaches of the intersection. Real-time information collected from probe vehicles (for instance, its position in the network, speed etc.), is used to determine the queue length on each link/approach of the intersection. The queue information of each link/phase during each cycle is stored and used for computing the maximum green time. Previous cycle (k) queue length information is extracted for the active phases, as the real-time queues cannot be measured for such phases. Various techniques can be implemented for updating the queue length information. For instance, for equation 1 the intersection queues (all approaches) are updated right before a phase is served (assigned green interval), and the respective Gmax is thereby computed. However, for equations 2 and 3, the queue length information is updated one phase at a

time. In this scenario, the queue length of the upcoming phase is updated just before the phase state is expected to transition from red to green.

C. *Maximum Green estimation*

Unlike a fixed maximum green in a typical actuated signal control, the G_{max} in the proposed methodology can be adaptively determined based on the queue length data. The queues are observed for all the phases that are in red interval during the current cycle (k+1). All the queues once observed are then archived for future use. The G_{max} for each phase is computed right before the green interval begins. In order to develop an equation which determines the maximum green for all phases, different Gmax estimation techniques with varying parameters are tested and evaluated using Average delay (per vehicle per second), and Maximum queue (in feet) as the measures of effectiveness (MOEs). Few equations that are tested and analyzed are illustrated below:

Equation 1,

$$Gmax_i^{k+1} = \max\left\{LB_i, \min\left(UB_i, \beta\left[\frac{QL_i^{k+1}}{QL_{p1}^{k}+QL_{p2}^{k}+\sum_{\substack{j=1 \\ j\neq p1,p2}}^{8} QL_j^{k+1}}\right] QL_i^{k+1}\right)\right\}$$

Equation 2,

$$Gmax_j^{k+1} = \max\left\{LB, \min\left(g_{TOT} * \left[\frac{QL_j^{k+1}}{\sum_{l=1}^{4}\max(QL_l^{t}, QL_{l+4}^{t})}\right]\right)\right\}$$

$$g_{TOT} = [C_{max} - L]$$

Equation 3,

$$Gmax_j^{k+1} = \max\left\{LB, \min\left(g_{TOT} * \left[\frac{QL_j^{k+1}}{\sum_{l=1}^{4}\max(QL_l^{k}, QL_{l+4}^{k})}\right]\right)\right\}$$

where, i = 1, 2, 3……..,8; j = 2, 4, 6 or 8
l = 1, 2, 3 or 4.
QL_{P1} = queue length for active phase in ring 1
QL_{P2} = queue length for active phase in ring 2
LB = the lower bound on green time (ideally the minimum green) for a phase (in seconds);
UB = the maximum allowable green time for a phase (in seconds);
β = the model parameter for converting the queue length (in number of vehicles) to time (in seconds);
QL_j^{k+1} = Queue length for the upcoming phase, j or i in cycle k+1;
QL_{l+4}^{t} = queue length of phase (l + 4), at time t;
g_{TOT} = the total effective green time;
L = the total lost time per cycle;
C_{max} = maximum allowable cycle length = 3 minutes

Equation 1 estimates the maximum green time for upcoming phase i, and is derived from the former study by Comert et. al. (2008). In this equation, the square bracket term represents the measure of fairness. In other words, the maximum green for a phase i, is determined by considering the overall demand at the intersection, relative to the queue at phase, i. For instance, if the queue length of phase i is 20, and the sum of QL is 100, then, that fraction (20/100) multiplied by the QL_i and a model parameter, β will give an estimate of the green time to be assigned to phase i. Equations 2 and 3 utilize only the queue lengths of the critical approaches. In other words, among the two con-current phases (2 and 6; 1 and 5 etc.) the higher QL (critical phase) is used as a measure of demand. The preliminary purpose of using the maximum Queue of the two phases is to avoid over-estimation of demand. One might notice that equations 2 and 3 are alike in many aspects except for the differing queue length observations. The queue lengths used in equation 2 are the most recently observed queues for phases 1 to 8 (at time, t). The time, t merely symbolizes the time instant of the most recent update of queue length for a phase, and is certainly different for different phases. Furthermore, equation 3 incorporates all the previous cycle queues (i.e. for phase 1 to 8) only. The g_{TOT} is introduced in equations 2 and 3 to capture the tolerance index for the inactive phases. The maximum time a vehicle on a movement, spends in the red interval, due to another movement with prolonged green (higher demand) can be justified by the use of tolerance index.

IV. STUDY SITE

The study site is an isolated 4-legged signalized intersection. It is a typical 8-phase intersection with 32 conflict points. The picture below (Figure 2) illustrates the geometry of the study intersection in VISSIM 5.3. Due to the flexibility and ease of using an arbitrary intersection, no real/existing intersection is considered for the case study.

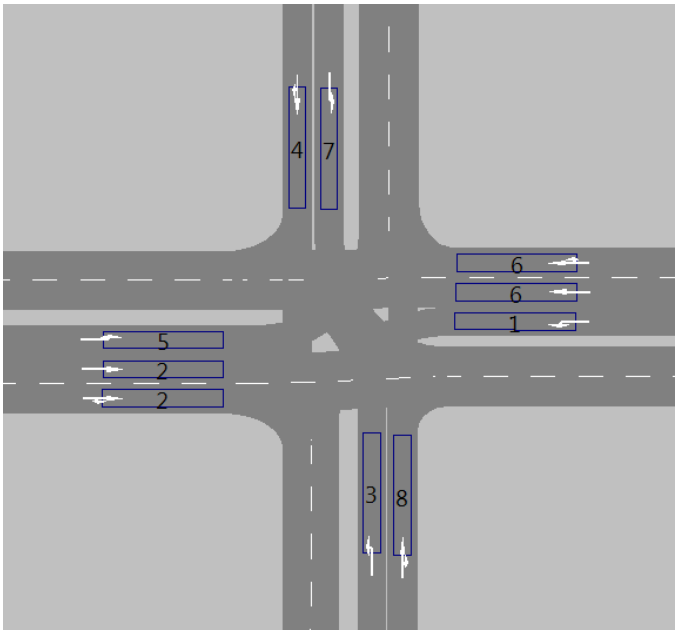


Figure 2 Intersection layout in VISSIM

V. EXPERIMENTS

A. Demand scenarios

Three different demand scenarios are set up to evaluate the robustness of the proposed models. The volume or hourly demand (vehicles per hour) on opposing approaches (major/minor street) is considered symmetrical. The only major difference between the three demand scenarios lies in the induced fluctuations at t = 900 seconds. For instance, the first demand profile has a steady influx of volume at NB and SB approaches, while it is uneven for EB and WB movements.

Time (in min)	Demand Profile #1 (vph)			
	NB	SB	EB	WB
0 - 5	700	700	1250	1250
6 - 10	500	500	1050	1050
11 - 15	600	600	1100	1100
16 - 25	550	550	2200	2200

Table 1 Demand scenario 1

Time (in min)	Demand Profile #2 (vph)			
	NB	SB	EB	WB
0 - 5	750	750	1100	1100
6 - 10	630	630	1150	1150
11 - 15	700	700	1300	1300
16 – 25	1300	1300	2100	2100
Time (in min)	Demand Profile #3 (vph)			
	NB	SB	EB	WB
0 - 5	500	500	1400	1400
6 – 10	350	350	1200	1200
11 – 15	400	400	1000	1000
16 – 25	1150	1150	2050	2050

Table 2 Demand scenarios 2 and 3

The demand surge apparently lasts for 10 minutes (16 – 25 minutes). It thereby causes a disruption in the efficient traffic signal operation at the intersection. Consequently, this unexpected demand surge leaves behind the residual queues for the first few cycles, until the adaptive G_{max} adjusts to meet the demand.

B. Results

Methods 1 and 2 are integrated in the signal control logic in COM interface and analyzed or tested under the three stipulated demand profiles to assess their robustness, using various measures of effectiveness. Minimum of three simulation runs (with varying random seeds) for each method under each demand profile are performed and the results are reported.

REFERENCES

[1] U.S. Department of Transportation, ITS Joint Program Office, VII Architecture and Functional Requirements, Version 1.1, July 2005.

[2] SmartSensor Design Catalog, Wavetronix LLC, 2007

[3] G. Comert; M. Cetin. *"Incorporating Queue Length Measurements into Actuated Signal Control: Evaluation of Efficiency Benefits at an Intersection"* Transportation Research Board, 88th Annual Meeting (2008).

[4] C.S Papacostas; P.D. Prevedouros, "*Transportation Engineering & Planning"*, 3rd Edition

[5] S. Hossain; L. Kattan; A. Radmanesh. *"Responsive Signal Control for Non-recurrent Traffic Congestion on an Arterial"*.

[6] Gartner, N. H. Demand-Responsive Traffic Signal Control Research. In Transportation Research, Vol. 19A, Issue 5/6, 1985, pp. 369-373.

Business and Industry

VMASC Track Chair: Dr. Andy Collins and Dr. Dean Chatfield
MSVE Track Chair: Ms. Yishu Zheng

Savings and the Real Interest Rate: An Empirical Analysis
Author(s): Mark C. Ingraham

Returns and the Bullwhip Effect
Author(s): Alan M. Pritchard and Dean C. Chatfield

A Conceptual Model of New Knowledge-based Venture Formation: Understanding the Roles of Networks in Entrepreneurship
Author(s): Xiaotian Wang, Andrew Collins, and Michael Provance

Imperfect Information and Heterogeneous Agents: Exploring Asset Bubbles and Crashes
Author(s): A.J. Costa and Dr. David Earnest

Algorithmic Trading Strategies
Author(s): Joseph Vesley and David C. Earnest

Exploration and Evaluation of Enterprise Innovation
Author(s): Mariusz A. Balaban, and Patrick T. Hester

Savings and the Real Interest Rate: An Empirical Analysis (January 2011)

Mark C. Ingraham, Old Dominion University, Department of Modeling & Simulation

Abstract— **Several studies have found a positive empirical correlation between savings and the real interest rate in an economy. At the same time, other studies have found a positive correlation between savings and inflation over the history of the American economy. There is a potential contradiction between the two findings because inflation will tend to reduce the real interest rate if the nominal interest rate remains the same. This paper runs multiple regressions on inflation, the real interest rate, and the savings rate. A possible explanation for these findings is ascertained. It is found that savings correlates with the nominal interest rate and the nominal interest rate increases with inflation. Bond prices may also play a role.**

***Index Terms*—Inflation, savings, real interest rate**

I. INTRODUCTION

The relation between the real interest rate, inflation, and the savings rate requires further study. Previous efforts have found a positive correlation between savings and real interest rates across countries, except in the United States (US) (De Serres et al. 2003, Heer et al. 2006, Tease & Jorgen 1991). Over the postwar period, studies have found a positive correlation between savings and inflation in the US. Because higher inflation will reduce real interest rates, all else equal, these data present a potential contradiction. (Roubini 2004)

It is possible that the US Federal Reserve Bank responds to higher expected inflation by increasing nominal interest rates to the point where real interest rates increase faster than inflation. This could mean that increased inflation in the US reduces the savings rate, *ceteris paribus*, but the effect is not apparent due to the nominal interest rate increasing in response to inflation. This paper aims to address part of this question by examining the relationship between savings and the real interest rate in the US.

Heer et al. (2006) examined the savings-inflation paradox in the US. The authors came to the conclusion that the paradox could be a statistical artifact from price index adjustments to income series. The authors also noted that the savings rate behavior in the US varied according to Federal Reserve chairmanship regimes. Specifically, there was a positive relationship during the Volcker Era, and a negative relation during the Greenspan and the 1965 to 1978 Eras. This indicates a possible role for monetary policy in affecting the savings rate. Monetary policy influences the real interest rate by setting the nominal interest rate. This paper seeks to determine whether the positive real interest vs. inflation rate relationship holds in the US as in other countries, and will further study nominal interest rate effects on the savings rate.

The primary purpose of this paper is to determine if real interest rates increases still cause savings rate increases in the US. The research hypothesis is as follows: There is a positive correlation between the monthly US savings and real interest rates from January 1970 to December 2010, significant at the $p = 0.05$ level.

The study will examine two periods in US monetary history. The Federal Reserve adopted a monetary aggregate approach to controlling the money supply in 1970. As mentioned by Orphanides (2007), inflation targeting or Taylor rule approaches have more accurately reflected Federal Reserve interest rate decisions since the early 1980s. Specifically, Fed policy since 1993 has reflected an inflation target in the range of two to three percent. Unlike some other central banks, however, the Federal Reserve has never adopted an explicit inflation target. The periods 1970 to 1992 and 1993 to 2010 are each included in the analysis.

A. Definitions

It is necessary to establish the definitions of technical terminology used in this paper. Inflation is an increase in the price level of goods and services in the economy. In the US, the Bureau of Labor Services (BLS) publishes an index of inflation called the Consumer Price Index (CPI). Two CPIs are calculated monthly. The CPI-W represents a special measure of inflation used for union wage negotiations. The CPI-U is a price index for a basket of goods representing purchases of urban consumers. This study, like many other econometric analyses of the US economy, makes use of the CPI-U. (Heer)

The nominal interest rate is the interest rate charged to banks in a country that desire to borrow from the national central bank. In the US, the nominal interest rate is most commonly referenced as the Federal Reserve's federal fund rate. (Romer & Romer) This paper uses monthly federal funds rate data to establish the nominal interest rate data set.

The real interest rate is the nominal interest rate minus the current rate of inflation. It is the return that banks must pay back loans borrowed at the nominal interest rate when the rate of inflation is taken into account. The real interest rate is important as it affects borrowing and savings behavior across the economy. Borrowing and savings behaviors may, in turn, affect the level of consumer and commercial demand in the economy, which affects inflation.

Manuscript received March 28, 2012.

II. Method

Multiple linear regressions is calculated to determine the relations between the savings rate, the real interest rate and the rate of inflation using monthly data for the US from January 1970 to December 2010. The source is the Bureau of Labor Statistics CPI-U data.

The inflation rate is provided by the CPI-U data published by the US Government. The nominal interest rate data will be provided by the monthly Federal Funds Rate provided by the Federal Reserve Bank of St. Louis. This data was found to be more comprehensive than the Federal Funds Rate data provided on the Federal Reserve Board (FRB) website. The FRB website only contains monthly Federal Funds Rate data for the period following 1990, whereas the St. Louis data covers the entire period under review. The real interest rate is calculated by subtracting the CPI inflation rate from the nominal interest rate for each point in time. The savings rate data will be provided by the US Department of Commerce Bureau of Economic Analysis monthly personal savings rate data. This data does not include information on US Government budget deficits or surpluses and is only relevant to private sector savings behavior.

The individual regressions are run and analyzed separately. The correlation between the savings rate and real interest rate will be subject to a hypothesis test. The array for each variable (the CPI-U, savings rate and real interest rate) will include 492 observations. The assumption is that the underlying economic dynamics of the US economy have been sufficiently similar over the time period under study that the same model can be applied to understand the entire period. Previous studies have taken an similar approach. For instance, analysis of the US inflation and savings rate relationship has been examined using regression analysis. (Federal Reserve Bank of St. Louis, 2007, Kandor et al., 1995, Heer et al., 2006) Another limitation of this approach is that any correlations detected do not necessarily indicate causal relationships. Additional theoretical analysis is required for interpretation.

A Student T test hypothesis test will be run to determine if there is a statistically significant positive correlation between the savings rate and the real interest rate at a $p = 0.05$ level of significance. The null hypothesis is that there is no significant relation, while the alternative hypothesis will be that such a relation exists. The linest() and ttest() functions in Microsoft Excel are used to determine the linear regressions and statistical significance of correlations, respectively. Linest() are applied according to the $y = mx + b$ method (the third input into the linest() function will be set to "1" or true, to allow b to vary from zero). Multiple linear regressions are run between the savings rate, the real interest rate, and inflation as measured by the CPI-U using the same method. Inflation and interest rates are treated as X values in examining the behavior of savings, the Y value. The proposed determinants of the savings rate, or the real/nominal interest and inflation rates, are each be plotted against the savings rate with a linear trend line and explained variance value.

Keeping with previous studies, such as in Heer (2006), this paper uses regression analysis to study the relations between macroeconomic variables despite imperfect normal distributions.

A. Data

Full US data from Government sources on inflation, savings rates and interest rates are only available for the recent postwar era. The year 1970 is selected as a starting year for this study on monetary explanations for the savings rate. A significant break with earlier practice occurred in 1970 as the FRB shifted from a money market policy to a monetary aggregate strategy. The Bureau of Labor Services consumer price index for all urban consumers (CPI-U) year-over-year change are used for the inflation rate. The CPI has faced criticism in that it does not accurately depict a basket of goods purchased by a typical consumer. (Heer et al. 2006) This measurement issue may affect the results, but constructing a more accurate price index is beyond the scope of this study.

B. Empirical results

The average of the savings rate values is 0.067*(171),* while the median is slightly lower, at 0.067. More significantly, however, when the savings rate is analyzed in terms of the two periods under study, i.e., 1970 to 1992 and 1993 to 2010, Arena output reports that the first period follows a lognormal distribution, while the second period follows a beta distribution.

Tables 1 displays descriptive statistics for the entire period of data.. Based on changes in FRB policies, the entire period of data was divided into two periods, Periods 1 and 2. Descriptive statistics for Periods 1 and 2 are presented in Table 2. Linear regression summary parameters for these three data partitions are presented in Tables 3 (Entire Period) and 4 (Periods 1 and 2).

The results of the statistical tests, all of which produced statistically significant results, are shown in Tables 3 and 4. A positive relation between savings and real interest rates is found. For the entire period, the relation between savings and inflation is positive, while the inflation and real interest rate relation is negative.

The real interest rate to savings relation is negative in the pre-1993 period but becomes positive in the second period. Inflation and the nominal rate both seem to be inversely related to savings post-1993. Post-1993 data with each bar graph representing an increment of one (1) percentage point of personal income saved. A normal distribution does not form a good fit to the data due to the small number of midrange savings rate values of approximately 4% that are observed during the post 1993 period. The reasons for this dual kurtosis distribution of savings rates, with peaks at 3 and 5%, is a possible avenue for future research. The observed results may relate to the brief interruption of the long-run decline in the US savings rate which occurred during the late 1990s dotcom bubble, without which the savings rate may have transitioned gradually downward through 4% rather than experiencing a large decrease after the bursting of the bubble. If correct, this interpretation can be generalized to explain irregularities in otherwise representative or normalized data which result from disruption of long-term trends due to temporary macroeconomic instabilities.

In all other regressions, both periods have the same sign as the aggregate relations.

The positive inflation and savings correlation is a confirmation of the results found in previously mentioned

studies. The relative strength of the correlations found is not necessarily indicative of their causal importance in the medium to long run. Even in the short run, expectations of future macroeconomic changes can influence behavior as much as or more than present conditions.

III. Discussion

A. Results

The relation between savings and the real interest rate is weak, at 0.15. There is a stronger relationship between the nominal interest rate and the savings rate, at 0.48. There is even stronger relation between the savings and inflation rates, which is found to have a value of 0.58.

The R^2 values for the savings against real interest rate regression indicate that 1.6% of the variance in the savings rate is explained by variance in the real interest rate. There are clearly other factors involved in determining the savings rate. The inflation rate accounts for approximately 39% of the variance in the savings rate, while the nominal interest rate accounts for approximately 37%.

Linear regression yields a value of 0.89 applies to the nominal interest rate versus inflation. This indicates that nominal interest rate increases tend to coincide with inflation rate increases, even in the short term time period reflected in monthly data.

B. Period 1: 1970 to 1992

In this period, there is a 0.24 slope on the inflation to savings regression, somewhat weaker than the result from aggregate data. The nominal rate also appears to have somewhat less effect on savings. The real interest rate appears to be negatively related to savings, although the relation is weak, at less than 0.08. The null hypothesis of a non-positive relation between savings and real rates would not warrant rejection if tested for this period.

C. Period II: 1993 to 2010

In this period, inflation has a negative relation with savings. It is worth noting the somewhat lower inflation rates during the 1990s and 2000s shown in Table 2. Thus, the observed positive savings/inflation relationship may be due to period effects, with unrelated factors causing a fall in savings and inflation over time in the US, rather than a direct relationship.

The nominal rate remains closely tied to inflation, but now becomes negatively correlated with savings. The relation, at 0.11, is not particularly strong. The relation between savings and the nominal rate is somewhat stronger in aggregate than in either period. Again, the aggregate results may not reflect causality, but rather coincidences of long term trends in the economy. Nominal rates fell somewhat after the early 1980s recession as "stagflation" ended. Savings fell as well, but this may have been due to other factors, such as demographics.

There is a positive but very weak relation between savings and the real interest rate in this period. The relation remains highly statistically significant, but has a value of less than 0.01. Although the hypothesis would be confirmed for this period, it appears that the observed aggregate relation is due primarily to long term economic trends rather than mono- or unidirectional causality.

D. Interpretation

An explanation of the observed macroeconomic relationships is put forward. The FRB increases the nominal interest rate in response to inflation, and the expectation of a series of gradual nominal interest rate increases may be a more important factor in determining the savings rate than the inflation rate is, at least in the short run. Economic agents have incentive to reduce borrowing and increase lending as the interest rate increases, as bond contracts can cause lock-in of higher nominal interest rates. These incentives remain if increased inflation is expected to temporarily reduce the real interest rate agents can expect to earn on savings.

Alternatively, bond price speculation may mediate the observed relationship, with agents preferring to purchase bonds ahead of expected nominal rate increases and sell them afterwards as higher nominal rates increase bond prices. Bond divestitures will decrease the rate of borrowing in the economy and raise the savings rate.

IV. Conclusion

A positive correlation between the savings rate and real interest rate exists. This paper fails to reject the initial hypothesis in aggregate for the period under study. However, the real interest rate does have a negative correlation with the savings rate in the pre-1993 period. The relation between these two variables is complicated by the positive inflation/savings correlation and negative inflation/real interest correlation. A possible explanation for the observed statistical relationships is that the savings rate responds primarily to the nominal interest rate, which the Federal Reserve chooses to increase when attempting to moderate high inflation. Another possible explanation is advanced regarding speculative motives for bond purchases and sales.

An increase in the real interest rate may increase the savings rate if the nominal interest rate increases or remains unchanged. Of these two proposed determinants of savings, correlations suggest that the nominal interest rate has a stronger effect on savings than the real interest rate, at least in the short run.

Additional research is needed to determine causal relationships between the variables under study. A Granger causality test or Generalized Methods of Moments analysis could be performed to determine to what degree inflation increases cause nominal rate increases. Additional insight could be gained from deriving theoretical models where the nominal interest rate is treated as a control variable and its effects adjusted to remain constant. Such a model could be applied to determine if the actual relation between inflation and savings is negative, *ceteris paribus*. Finally, analysis of time-series and causal relationships and correlations can help to answer questions about Fed statements' and inflation and interest rate expectations' influence on present savings behavior.

Appendix

TABLE I
Descriptive Statistics for Entire Period, 1970 through 2010

Variable	Mean	Std. Dev.	Minimum	Maximum
Inflation	0.04	0.03030*(6)*	-0.02	0.15

Nominal	0.06	0.03628*(9)*	0.00	0.19
Real	0.02	0.02531*(5)*	-0.05	0.10
Savings	0.07	0.02835*(9)*	0.01	0.15
Obs.	492			

TABLE II
DESCRIPTIVE STATISTICS FOR PERIODS 1 AND 2

Variable	**Mean**	**Std. Dev.**	**Min.**	**Max.**
Period 1: 1970 to 1992				
Inflation	0.06	0.03	0.01	0.15
Nominal	0.08	0.03	0.03	0.19
Real	0.02	0.03	-0.05	0.10
Savings	0.09	0.02	0.03	0.15
Obs.	276			
Period 2: 1993 to 2010				
Inflation	0.02	0.01	-0.02	0.06
Nominal	0.04	0.02	0.00	0.07
Real	0.01	0.02	-0.04	0.04
Savings	0.04	0.01	0.01	0.08
Obs.	216			

TABLE III
LINEAR REGRESSION FOR ENTIRE PERIOD, 1970 TO 2010

Slope of best fit line		**P value from T test**
Inflation (x value) versus y value		
Nominal	0.868031	< 0.00001
Real	-0.13197	< 0.00001
Savings	0.582556	< 0.00001
Nominal (x value) versus y value		
Real	0.394594	< 0.00001
Savings	0.477196	0.002904539
Real (x value) versus y value		
Savings	0.145681	< 0.00001

TABLE IV
LINEAR REGRESSION FOR PERIODS 1 AND 2

Slope of best fit line		**P value**
Period 1: 1970 to 1992		
Inflation versus:		
Nominal	0.638542957	< 0.00001
Real	-0.361457043	< 0.00001
Savings	0.242715633	< 0.00001
Nominal versus:		
Real	0.426166797	< 0.00001
Savings	0.165148016	0.001821331
Real versus:		
Savings	-0.070534378	< 0.00001
Period 2: 1993 to 2010		
Inflation versus:		
Nominal	0.709927763	< 0.00001
Real	-0.290072237	< 0.00001
Savings	-0.339592979	< 0.00001
Nominal versus:		
Real	0.773164138	< 0.00001
Savings	-0.107153241	0.002296481
Real versus:		
Savings	0.001563075	< 0.00001

REFERENCES

[[1] Bureau of Labor Statistics. "Consumer Price Index (CPI-U)." U.S. Department Of Labor Bureau of Labor Statistics," Web. 10 Oct 2011. <ftp://ftp.bls.gov/pub/special.requests/cpi/cpiai.txt>.

[[2] De Serres, Alain, and Florian Pelgrin. "The decline in private saving rates in the 1990s in OECD countries: how much can be explained by non-wealth determinants?" OECD Economic Studies . 36.I (2003): Print.

[[3] Federal Reserve Bank of St. Louis. US. A Guide to the National Income and Product Accounts of the United States (NIPA) personal savings rate. 2011. Print. <http://research.stlouisfed.org/fred2/data/PSAVERT.txt>.

[[4] Federal Reserve Bank of St. Louis. "Monthly federal funds rate." Archival Federal Reserve Economic Data. Web. 11 Nov 2011. <http://research.stlouisfed.org/fred2/series/FEDFUNDS?cid=47 >.

[[5] Heer, Burkhard, and Bernd Suessmuth. "The savings-inflation puzzle." CESIFO working papers. 1645. (2006): Print.

[[6] Kandor, Shmuel, Aahron Ofer, and Odeg Sarig. "Real interest rates and inflation: an ex-ante empirical analysis." Rodney L. White Center for Financial Research. (1995): Print.

[[7] Orthanides, Athanasios. "Active stabilization policy and inflation". Board of Governors of the Federal Reserve System. (2000).

[[8] Orthanides, Athanasios. "Taylor rules." Board of Governors of the Federal Reserve System. (2007).

[[9] Romer, Christina and Romer, David. "The NBER Monetary Economics Program." National Bureau of Economic Research. 10 Oct 2011. <http://www.nber.org/programs/me/me.html>.

[[10] Roubini, Nouriel. "Understanding the World Macroeconomy, Handout for Chapter 5: Output and Real Interest Rates." Output and real interest rates, (2004). 10 Oct 2011. <http://pages.stern.nyu.edu/~nroubini/NOTES/HAND5.HTM>.

[[11] Tease, Warren, and Jorgen Elmeskov. "Real interest rate trends: the influence of saving. Investment and other factors." OECD Economic Studies. 17. (1991). Print.

[[12] United States. Federal Reserve Board: open market operations archive, intended federal funds rate. <http://www.federalreserve.gov/monetarypolicy/openmarket_archive.htm>.

Returns and the Bullwhip Effect (April 2012)

Alan M. Pritchard and Dean C. Chatfield, *Member, VMASC*

I. Motivation

We investigate the issue of returns and their impact on the phenomenon known as the "Bullwhip Effect" (BWE) in a serial supply chain. The Bullwhip Effect is the amplification of demand variance as orders move upstream through the supply chain. The concept of the Bullwhip Effect is not a new one. Extensive evidence of the BWE has been found in both the operations research and the economics literature, and we are interested in whether the assumption of returnability significantly impacts the accuracy of the traditional bullwhip effect models where returns are not allowed.

Successfully quantifying the forces that lead to the BWE is important for developing strategies to eliminate its detrimental impacts. Many factors have been identified that contribute to the BWE. Forrester contributed the amplification found in his simulations to the structure, policies, and interactions within the supply chain [1]. In the now popular classroom game known as the Beer Distribution Game, Sterman showed that "misperceptions of feedback" by human decision makers gave rise to the BWE [2]. Metters [3] tried to identify a lower bound on the monetary implications of the BWE for a periodic, time-varying, stochastic demand industry with capacitated production. Overall, Metters concluded that reducing the BWE can improve profitability by 15-30 percentage points.

The implications of negative orders, however, have largely been ignored in the literature up to this point. Most researchers in the operations research and economics literature allow negative orders for analytical tractability, without exploring the implications of doing so. Chen et al. [4], Kahn [5], and Lee et al. [6] allow negative orders in their models and assume this *excess inventory is returned without cost*.

Manuscript received March 7, 2012. This work was supported in part by the Virginia Modeling, Analysis and Simulation Center and by Old Dominion University, Norfolk, Virginia, USA.

Alan M. Pritchard is a M.A. student in the Economics department in the College of Business at Old Dominion University, Norfolk, Virginia 23529 (e-mail: aprit003@odu.edu).

Dean C. Chatfield is an Assistant Professor in the Decision Sciences Department in the College of Business at Old Dominion University (phone: 757-683-4880; e-mail: dchatfie@odu.edu).

In their 2000 paper, Chen et al. acknowledge that the main drawback of their model is the assumption that excess inventory is returned without cost. They acknowledge that this assumption may not be realistic and perform some basic simulations to show that this assumption does not drastically impact the variance of the order quantity. While the demand variance amplification results from their simple two-stage model are very close for return permitted and return restricted scenarios, what is not shown is the impact further up the supply chain. The five-stage model we employ clearly shows significant differences in variance amplification at the upper echelons.

Others, such as Lee et al, assume the mean of the customer demand is significantly higher than the standard deviation of the customer demand so that the probability of a negative order is negligible. In our analysis, we find that this point is valid at the customer level, but as orders move up the supply chain the demand standard deviation will raise and the probability of negative orders will increase substantially.

II. Methods

We build a sophisticated agent/discrete-event hybrid simulation model of a five stage supply chain with a similar structure to the one described in Chen and Samroengraja [7]: a serial formation of customer, retailer, wholesaler, distributor, and factory levels, or nodes k = 0, 1, 2, 3, and 4. We utilize the NetLogo agent programming environment version 5.0 to construct our model, and we add the necessary discrete-event functionality by including additional code directly within our model, such that the final hybrid model is a self-contained NetLogo model not relying on external software or libraries.

The sequence of events in our model follows the routine set up by Sterman in the Beer Game, where each node places orders with the next upper node and this node fills those orders. The customer does not fill orders and the factory places orders with an outside supplier. Our model represents returns and return processing explicitly. We employ an adaptive (R, S) inventory policy with review period R =1. We utilize the standard protection period of $L + R$ time units and employ a normal approximation, as is the convention, when determining the order-up-to point (S).

In order to verify our model, we compare our simulation results to the analytical results of Chen et al, Chatfield et al. [8], and Dejonckheere et al. [9]. We verify by making adjustments to our model in order to fit the assumptions of the previous literature, and then we compare the amplification ratios of the different levels of the supply chain. It should be noted that the purpose of this research is to relax some of the assumptions of the previous literature and not merely replicate the existing work.

We create an experiment utilizing the following four factors, which have been shown to have statistically significant effects in previous studies of the BWE:

1) **Returns permission.** (Yes/No). If returns are permitted, then orders from the customer as well as replenishment orders placed by the stocking points may have a negative size. When returns are not permitted, the order size is truncated at 0.0
2) **Information Sharing.** (Yes/No). In cases where information sharing is present, all nodes have access to the customer demand information, whereas when no information sharing is allowed, each node only has access to the demand stream (orders) coming in from its downstream partner.
3) **Customer demand variance.** (c.v. = 0.1, 0.2, 0.4). Customer demand is normally distributed with a mean of 50 units. We utilize three levels of demand variance, which correspond to coefficient of variation (c.v.) levels of 0.1, 0.2, and 0.4, with c.v. defined as the standard deviation/mean. This results in customer demand distributions of N(50, 52), N(50, 102), and N(50, 202) using the notation N(μ, σ2) for the normal distribution.
4) **Lead time variance.** (c.v. = 0.0, 0.25, 0.50). Lead-time is gamma distributed with a mean of 4 time units. We utilize three levels of lead-time variance, which correspond to coefficient of variation levels of 0.0, 0.25, and 0.50. This equates to lead-time distributions of L=4 (constant), G(4, 12), and G(4, 22) using the notation G(μ, σ2) for the gamma distribution.

III. Results

While many studies have justified the allowance of negative orders by stating that the percentage of negative orders is insubstantial with a low demand coefficient of variation, we find evidence that clearly negates this assumption. This assumption may be valid for a two stage system consisting of only a customer and retailer; however, when multiple levels of the supply chain are modeled, the percentage of negative orders becomes substantial. This can be seen even with a very low customer demand coefficient of variation by observing that the percent of negative orders increases as you move up the supply chain. Increasing the demand coefficient of variation amplifies this effect even further. We find that with a moderate coefficient of variation, there are even a substantial percentage of negative orders at the retailer level. With a demand standard deviation of 20, we observed a 35% probability that an order placed by a factory will be a negative order.

We find that within scenarios utilizing the same customer demand c.v., the higher the lead time variance, the larger the effect returns permission will have on the BWE. The opposite is also true; within scenarios utilizing the same lead time variance, the higher the customer demand c.v., the larger the effect returns permission will have. Both of these effects are consistent with and without information sharing, but are especially evident without information sharing.

We acknowledge that there are instances where a supply chain's return policy is unchanging, however, in other instances, returns permission may be partially determined by the firm. In our research, we found that this decision can make a substantial difference in the BWE experienced by the supply chain. Allowing returns will amplify the BWE, even when information sharing is employed, so it is up to the manager to decide whether the customer service benefit of allowing returns from the downstream member of the supply chain is enough to compensate for the additional costs that these returns create.

IV. References

[1] Forrester, J.W., 1958. Industrial Dynamics – A Major Breakthrough for Decision Makers. *Harvard Business Review* 36 (4), 37-66.

[2] Sterman, J.D., 1989. Modeling Managerial Behavior: Misperceptions of Feedback in a Dynamic Decision Making Experiments. *Management Science* 35 (3), 321-339.

[3] Metters, R., 1997. Quantifying the Bullwhip Effect in Supply Chains. *Journal of Operations Management* 15 (2), 89-100.

[4] Chen, F., Drezner, Z., Ryan, J.K., Simchi-Levi, D., 2000. Quantifying the Bullwhip Effect in a Simple Supply Chain: The Impact of forecasting, Lead times, and Information. *Management Science* 46 (3), 436-443.

[5] Kahn, J. A., 1987. Inventories and the Volatility of Production. *American Economic Review* 77(4), 667-79.

[6] Lee, H., Padmanabhan, P., Whang, S., 1997. Information Distortion in a Supply Chain: The Bullwhip Effect. *Management Science* 43 (4), 546-558.

[7] Chen, F., Samroengraja, R., 2000. The Stationary Beer Game. *Production and Operations Management* 9 (1), 19-30.

[8] Chatfield, D.C., Kim, J., Harrison, T.P., Hayya, J.C., 2004. The bullwhip effect- impact of stochastic lead time, information quality, and information sharing: a simulation study. *Production and Operations Management* 13 (4), 340-353.

[9] Dejonckheere, J., Disney, S.M., Lambrecht, M.R., Towill, D.R., 2004. The Impact of Information Enrichment on the Bullwhip Effect in Supply Chains: A Control Theoretic Approach. *European Journal of Operations Research* 153 (3), 727-750.

A Conceptual Model of New Knowledge-based Venture Formation: Understanding the Roles of Networks in Entrepreneurship

Xiaotian Wang
Modeling, Simulation and Visualization Engineering Department
Old Dominion University
xwang009@odu.edu

Andrew Collins, Ph.D.
Virginia Modeling, Analysis and Simulation Center
Old Dominion University
ajcollin@odu.edu

Michael Provance, Ph.D.
College of Business and Public Administration
Old Dominion University
mprovanc@odu.edu

Abstract: In recent years, there is a burgeoning literature on the knowledge-based entrepreneurship, and researchers have pointed out two key mechanisms that lead to sustainable entrepreneurial activities: knowledge spill-over stemmed from agglomeration, and network-based new venture formation (i.e., social embeddedness model). While much has been discussed in comparing these two mechanisms from many theoretical angles, none has examined them through the lenses of the effectiveness of different network structures. Knowledge appropriation, either in the agglomeration model or in the social embeddedness model, does not take place in a vacuum of networks. Rather, the processes and results of knowledge appropriation are fundamentally shaped by the structures of business and social networks. In this paper, the authors explore the causal roles of network structure in affecting knowledge appropriation by using agent-based simulation methodology. Furthermore, the author intends to elucidate what kind of knowledge formation network is more effective in fostering new ventures and, accordingly, promoting the sustainable economic growth in regions.

1.0 INTRODCUTION

Joseph Schumpeter once described Entrepreneurship as “the gale of creative destruction” to take over in a totally or partially way of inferior innovations across both markets and industries, meanwhile, to create brand-new products including new business models [1]. In the light of this, entrepreneurship can be best understood as the strong propulsion of the dynamism of industries and long-lasting economic growth. As a consequence, the sources of sustainable entrepreneurship have long been one of the key questions of economics. A cursory review of the existing literature reveals that there are two principal theories in explaining the formation of new venture, that is, agglomeration theory and entrepreneurial action [2]-[5].

As for the camp of agglomeration theory, scholars treat clusters or agglomeration

economies as the central explanatory variable of sustainable entrepreneurship. Specifically, agglomeration or cluster is commonly referred to the phenomenon of "spatial clustering," or the clusters of firms within certain regions. In its essence, agglomeration suggests that the existence of these clusters can greatly foster the formation of new firms by providing both internal and external knowledge appropriation.

More specifically, it is argued that clustering promotes the generation of entrepreneurial knowledge networks, usually expressed as innovation systems in regions, through entrepreneurial knowledge carried by founders of incumbent firms to the new ventures or through knowledge exchanged in the connection between parent firms and spinouts [2], [6], [7]. As a corollary, knowledge networks, facilitated by clustering, can strongly affect the formation of new ventures by stimulating or spreading new or existing entrepreneurial knowledge, providing more effective and complex networks of information dissemination.

Despite the prevalence of the agglomeration theory, entrepreneurial action theory is also widely used to explain the formation of new ventures. Rather than emphasizing such environmental variables as clustering, it stresses the importance of the role of entrepreneurs who own entrepreneurial knowledge and strategic visions.

As elucidated by Alvarez et al. [8], the logics of entrepreneurial action theory can be exampled as that "*if opportunities exist as objective phenomena, then the task of ambitious entrepreneurs is to discover these opportunities – using whatever data collection techniques exist – and then exploit them – using whatever strategies are required – all as quickly as possible, before another entrepreneur discovers and exploits the opportunity*." In other words, entrepreneurship can be best theorized as the results of some entrepreneurs who have the access to the priority of the opportunity, that is, entrepreneurs who obtain the ownership of some new entrepreneurial knowledge. After imported into knowledge formation networks, the new knowledge owned by the entrepreneurs becomes the starting points for the networks.

Although the above-mentioned two theories highlight different mechanisms of knowledge-based venture formation, these mechanisms are not mutually exclusive. Rather, the two kinds of formation mechanisms coexist and even complement each other in the real world venture formation. This also dictates researchers to elaborate and clarify the relationships between these different mechanisms. Carayannis et al. [2], for example, argue that "*while the two camps agree upon the importance of regional innovation systems to nurture the creation and formation of new ventures … they differ in the ways in which knowledge is acquired and used to generate the entrepreneurial activity*."

Although much has been written in comparing these two theories, few, if any, have employed a network perspective. What if we treat each formation mechanism as a distinct network of knowledge diffusion? What if we simulate the network-oriented models of venture formation by using agent-based modeling? More fundamentally, can ABMS tell us whether different structures of knowledge networks are the key to the emergence of new ventures? If so, can ABMS tells us what kinds of knowledge network are more likely to foster the formation of new ventures under a particular regional setting? After examining the underlying assumptions of the two theories and ABMS methods, the authors find that answers are probably "Yes".

In the second parts of this paper, the authors use ABMS methods to simulate and compare agglomeration theory and entrepreneurial action theory by applying village (or clique) network and opinion leader network respectively. More

discussions about of relationship between the networks and theories will be delivered in 2.0.

Besides providing us another way to examine agglomeration theory and entrepreneurial action theory, there are multiple advantages to adopt a network-oriented approach. Provance et al. [5] argue in their working paper that agglomeration emphasizes the regions grow because of strong cluster that foster knowledge spillover locally. However, with the advancement of modern technology, knowledge acquisition and transmission mode can be ever-changing. Therefore, entrepreneurial formation mechanism is characterized with great complexity. Without a scrutiny of how knowledge network works, it is impossible to answer which type of knowledge formation network in the known information/knowledge transformation networks is the most effective and fast mode embedded in the regarding region ambience?

Meanwhile, the ambiguous concept of 'knowledge spillover' can no longer satisfy our needs to gain a concrete understanding about what kinds of knowledge network suits a particular regional environment. Our fundamental purpose is to find a best knowledge formation network that can foster a sustainable regional economic growth, and a network-oriented approach seems to be the only viable way to solve this problem.

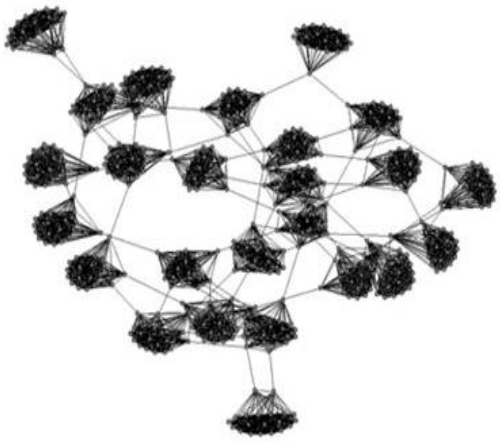
2.0 A NETWORK-ORIENTED APPROCH

In this part, the authors present an attempt to apply a network-orient approach to understand the formation of new ventures. Specifically, two ideal knowledge networks—Opinion-leader network and Village (Clique) Network—are introduced to model entrepreneurial action theory and agglomeration theory respectively.

Figure1 Opinion-Leader Network

Figure 1 is Opinion-leader network (OLN), which is used to model entrepreneurial action theory. Just as described by entrepreneurial action theory, in OLN most firms have few knowledge connections, while a few—the leading firms run by entrepreneurs of strategic visions (i.e., the opinion leaders)—have many. From a technical perspective, we can use a single parameter determines both the number of opinion leaders and the number of connections each has.

A real world version of this is the tablet industry. After the emergence of touch-pad, huge customer demand turns out to be the most important reason for the emergence of new type of companies. However, only few of them have plenty of knowledge networks, and the market is dominated by a small number of opinion-leader companies with unmatched access to knowledge connections (such as Apple, Amazon, and Google). The results turn out to be most "mass" firms are influenced by a few "leader" firms, producing the similar products with the same function and competing for lower price.



Village (Clique) Network

Figure2 Village Network

Figure 2 presents Village (clique) network (VCN), which is used to model the agglomeration theory here. Similar to the agglomeration theory, in VCN the firms are tightly clustered. Specifically, within each "clique" of VCN, every firm links to each other, and the follows of knowledge between them are largely similar, and only the rare firms that spans multiple cliques.

The real world examples can be Silicon Valley and Hollyhood. The advancement of a specific technique attracts more professional employees; accordingly, foster a venture cluster/village centric on one particular technique. Village Network process can be describe as: a small cluster with the same type of knowledge will attract large amounts of professional people and a huge cluster with the same type of knowledge.

It should be noted that these two networks are the "ideal type" networks, and thus neither comprehensive nor conclusive. Further studies are called for to improve our understanding about the ways in which knowledge networks affect venture formation.

3.0 Methodology

Agent-based modeling (ABM) will be the modeling paradigm used for this study.

Agent-based modeling is a powerful simulation technique that obtains "*growing acceptance and enthusiasm in various fields of social science in recent years"* [10], specifically, in political science, economics and finance et al. In ABM, system is described as a collection of agents which are making decisions automatically (implemented through programming) based on environment variation. Each individual agent, behaved as the human facing the issues in the real world, address the regarding factual situations and make decisions, provided simulated environment consisted of executable rules within the system [9].

The fundamental and a most important reason of applying ABM in social science is the character of generating "generative science" [11]. It captures emergent phenomena which are resulted from the interactions of individual entities. *Eric Bonabeau [9*] explains this in the informal language, ABM is a problem solving machine for answering the question "what could happen in the future, given simulated economic environment?", and this is the first reason that using this methodology in entrepreneurship.

Another reason is explained from the technical perspective. Simulation tool (Repast Simphony) of ABM is able to help us by handling multi-agents coexisting in multiple dimensions, which is a sparkling point for this study. In the process of simulation, agents coexist in two spaces: simulated space and 'real world space'. They interact with each other in the same space and also in the different space, to some certain extent, the simulated results make sense both in theoretical and in 'real world' angles.

The last but not the least reason is that it offers flexibility of programming required by

complex system study as Repast Simphony uses a JAVA development interface.

In the finalizing stage, the model will be developed and validated using the assistance of a Subject Matter Expert (SME) from the field of Entrepreneurship.

4.0 Conclusion

The expectation of this piece of study is obtaining a simulation result of one of two networks more positively affect the new venture formation. In fact, by knowing the structure of network are likely to help to focus data collection, conserving limited resources. Consequently, validation is going to be practical and digging further is also in expectation. Many drawbacks also exist in the initial stage, such as, more expertise needed. Actually, business theory is a very critical issue. Without the expertise of this field, more uncertainty will be brought in the study.

References

[1] Schumpeter, Joseph A. "*Capitalism, Socialism and Democracy*", 1942

[2] Elias G. Carayannis, Mike Provance, Nathaniel Givens, “Knowledge Arbitrage, Serendipity, and Knowledge Formality: Their Effects on Sustainable Entrepreneurial Activity in Regions”, IEEE Transactions on Engineering Management, vol.58, no.3, August, 2011

[3] E.G. Carayannis, “Knowledge-driven creative destruction, or leveraging knowledge for competitive advantage: Strategic knowledge arbitrage and serendipity as real options drivers triggered by co-opetition, co-evolution, and co-specialization,” Industry & Higher Education, vol. 22, no. 6, pp. 343-353, 2007.

[4]Mike Provance, Try it together: Three Essays on The Roles of Knowledge, Entrepreneurial Action, and Institutions in The Formation of New Ventures. Dissertation, 2010.

[5] Mike Provance, Andrew Collins, Elias G. Carayannis, “An Agent-Based Model of New Venture Creation: Conceptual Design for Simulating Entrepreneurship”, 2012

[6] E. G. Carayannis, “Firm evolution dynamics: Towards sustainable entrepreneurship and robust competitiveness in the knowledge economy and society,” Int. J. Innovat. Reg. Develop., vol. 1, no. 3, pp. 235–254, 2009.

[7] S. Klepper and S. Sleeper, “Entry by spinoffs,” Manage. Sci., vol. 51, no. 8, pp. 1291–1306, 2005.

[8] Sharon A. Alvarez and Jay B. Barney, “Discovery and Creation: Alternative Theories of Entrepreneurial Action,” Strategic Entrepreneurship Journal, 1: 11–26, 2007

[9] Eric Bonabeau, “Agent-based modeling: Methods and techniques for simulating human systems”, Proceedings of the National Academy of Sciences of the United States of America, vol.99, no. Suppl 3, 2002

[10] Steven C. Bankes, “Agent-based modeling: A revolution?”, Proceedings of the National Academy of Sciences of the United States of America, vol.99, no. Suppl 3, 2002

[11] Michael J. North and Charles M. Macal, Managing Business Complexity: Discovering Strategic Solutions with Agent-Based Modeling and Simulation , USA, Oxford University Press, 2007

Imperfect Information and Heterogeneous Agents: Exploring Asset Bubbles and Crashes

A.J. Costa and Dr. David Earnest
Graduate Program in International Studies
Old Dominion University
Norfolk, Virginia, U.S.A.

***Abstract*—Asset bubbles are difficult to explain through conventional economic methods. Agent-based modeling can be utilized for better exploration of this event. Heterogeneous agents with varied levels of risk tolerance and access to information engage in herd behavior to create bubbles and crashes.**

Keywords-bubble; finance; agent-based modeling.

I. INTRODUCTION

Conventional economics provides no place for a stock bubble. Some economists deny that asset bubbles even occur. Traditionally, the study of economics takes place in a rigid structure where humans are assumed to be instantly, constantly and homogeneously rational. Perfect information is taken as a given. Price, says Hayek, is the collective knowledge of all people participating in the economy and is thus a signal of true value [1]. There is an equilibrium price for supply and demand of any good, and a deviation from that equilibrium will be temporary, for the market will return to equilibrium.

However, asset bubbles do occur, as we have seen from the dot-com bubble of the early 2000's and the real estate bubble of 2008. The more severe crashes can bring with them the economic devastation we have seen since 2008 and the ramifications echo across the globe. Given the assumptions of traditional economics, how could this happen? This extended abstract will briefly review alternative theories on the dynamics of asset bubbles. An agent-based model will be proposed that can study this financial event using heterogeneous agents with differing levels of risk tolerance and separate information sets.

II. LITERATURE

Alternative theories and approaches to asset bubbles do exist in the literature. Hyman P. Minsky identifies three factors that lead to financial instability: rising debt/income ratios, increased asset and real estate values, and declining liquidity. As these three factors increase, the economy moves from a state of stability to instability. If enough sectors of the economy are affected, a general fall in asset prices will occur and result in a deep economic depression [2].

Paul Krugman blames "irrational exuberance" for the appearance of asset bubbles. As traders see the price of an asset rise, they assume it will go up forever and decide to invest in it while they have the chance to make profits from it. As more and more traders buy, the price continues to rise, its expected value rising above its real value (this is the "bubble"). However, the belief it will rise forever is faulty. Eventually it pops when some traders come to this realization and decide to sell out of that market. The herd behavior strikes again as more and more traders sell, prompting a crash in value [3].

This herd behavior that creates and pops asset bubbles is also evident in something known as the Period of Financial Distress. At the price peak, a few investors begin to sell their holdings of the asset and the expected value declines slowly, before the rest of the herd catches on and sells quickly, almost all at once. The Period of Financial Distress is the time between the peak and the crash [4]. Here we begin to see that people are not homogeneous, but rather have different information sets and attitudes toward risk. Individuals have differing levels of risk tolerance [5]. Some people have access to different information than others. Many studies make note of two types of traders, which could be known as "information traders" and "noise traders." "Information traders" are better-informed, choosing to make decisions based on fundamental information about a company and the market, whereas "noise traders" use the market price signal as their information. Information is not perfect, but indeed rather complex due to the effects of noise in the system, making it difficult to determine the real price of an asset [6].

III. METHOD

Asset bubbles involve heterogeneous traders with differing levels of risk tolerance and access to information, herd behavior, and feedback patterns. Computational finance has made great strides in modeling these bubbles. Agent-based modeling can be used in this case to better understand how bubbles occur.

The model we use draws from Stefan Thurner's 2011 model of financial markets [7]. It will be implemented in NetLogo on a 33 by 33 grid. Investment funds plus two types of investor with different access to price information buy and sell assets in the marketplace. Investor aggressiveness is variable. The following discusses agent types and procedures of the model.

A. Agents

- Assets are an agent in this model. Each has its own price.

- Funds own cash and assets and seek to maximize their wealth.
- Each fund has its own bank which may lend to it. Bank agents have no connections to one another.
- Investors trade based on information. They can see the hidden price of an asset and determine if it is over- or under-priced before buying it.
- Noise traders assume the demand function is the true price of an asset and base their buying decision on it.

B. Procedures

At each time step, the following procedures are carried out:

- Noise traders calculate their demand.
- Asset prices are calculated using the demand of noise traders, the demand of funds, and the number of assets in the model.
- Based on a series of equations, investors check the net asset value of the funds, checking it against past performance, and either add or subtract money from the funds.
- The funds calculate their payoff.
- Banks calculate the amount of leverage they are willing to provide funds based on past market volatility.
- If a fund's wealth has fallen below its initial wealth (it goes bankrupt), it is removed from the simulation.

C. Variable Parameters

The following parameters are variable and can be altered at simulation setup:

- The number of assets, funds, investors and noise traders.
- The aggressiveness of traders.
- The maximum amount of leverage banks are allowed.
- Price volatility, the stochastic element.
- The extent to which banks are monitored or regulated.

IV. CONCLUSION

My hope is that multiple runs of this model will yield greater insight to the conditions that create asset bubbles and crashes. History has shown these bubbles and crashes occur periodically with similar traits. Heterogeneous agents with differing levels of risk tolerance and information sets depart from conventional economic methods but may provide a more realistic understanding of the conditions leading to these events.

REFERENCES

[1] F.A. Hayek, "The use of knowledge in society," in The American Economic Review no. 35 (September 1945), pp. 519-530.

[2] Hyman P. Minsky, "Longer waves in financial relations: Financial factors in the more severe depressions," in American Economic Review no. 54 (May 1964), pp. 324-335.

[3] Paul Krugman, The Return of Depression Economics and the Crisis of 2008. New York: W.W. Norton & Company, Inc., 2009.

[4] Mauro Gallegati, Antonio Palestrini and J. Barkley Rosser, "The Period of Financial Distress in financial markets: Interacting heterogeneous agents and financial constraints," in Macroeconomic Dynamics no. 15 (February 2011), pp. 60-79.

[5] Carl Chiarella and Xue-Zhong He, "Heterogeneous beliefs, risk, and learning in a simple asset pricing model," in Computational Economics no. 19 (February 2002), pp. 95-132.

[6] Fischer Black, "Noise," in The Journal of Finance no. 41 (July 1986), pp. 529-543.

[7] Stefan Thurner, "Systemic financial risk: agent based models to understand the leverage cycle on national scales and its consequences," OECD/IFP Project on "Future Global Shocks."

Algorithmic Trading Strategies

Joseph Vesley and David C. Earnest, Ph.D

***Abstract*--Today, algorithmic traders—computers executing buy and sell orders without human intervention—conduct high-frequency, high-equity trades on most major stock markets[1]. This study simulates algorithmic trading to assess which algorithms produce the highest return to investors. Is it possible to create an automated program that can choose investments that perform at a higher rate than the financial market itself? Quantitative methods have been used (by institutional investors) to derive formulae that provide mathematical decision criteria on investments.**

***Index Terms*--Agent-Based Modeling, Applied Statistics, Artificial Intelligence, Computational Finance**

I. INTRODUCTION

Automated trading—that is, the sale or purchase of equities by computers without human intervention—today encompasses a substantial portion of trade in the New York Stock Exchange and other leading stock markets around the world. This interdisciplinary project combines insights from the fields of investment finance, artificial intelligence, and modeling and simulation to emulate the most common algorithmic trading strategies. Wall Street currently uses these computer programs to buy and sell stock based upon mathematical decision criteria[2]. Computer systems have a natural breadth and speed that cannot be matched by human traders, affording them a competitive advantage in a marketplace where the volume of trade and small margins can make the difference between profitability and insolvency.

Manuscript submitted April 3rd, 2012
Joseph Vesley is with Old Dominion University, 5115 Hampton Boulevard Norfolk, VA 23529 (804-832-4725; jvesl001@odu.edu)
David C. Earnest, Ph.D, is with the Graduate Program in International Studies, Old Dominion University, 5115 Hampton Boulevard Norfolk, VA 23529 (dearnest@odu.edu)

A human trader may have a more intuitive sense of the market, and thus may have a better grasp of one individual trade than a computer could, but a computer can conduct thousands of trades simultaneously without tiring. A computer can also perform rigorous calculations on data sets much more easily than a human trader ever could[3]. Using statistical analysis, simple rules are used (that are automated and highly replicable) to achieve scales of performance that are beyond human capacities to achieve. Algorithms are used to predict the mathematical attributes of a stock: probabilistic elements (such as confidence intervals and bell curves) that are completely separate from investing fundamentals. Even though stock values are supposedly caused by identifiable economic and industrial phenomena, some would argue that there are patterns to be recognized in the data itself. Using artificial intelligence to purchase stocks is contrary to classic investment theory, which analyzes nebulous concepts like productivity, demand, and motivation. These concepts are currently being adapted into algorithmic models (for example, to predict the market effects of news headlines) but that research is in its infancy and is highly secretive. The purely mathematic approach is also contrary to most conventional academic theory on statistics: the 'efficient markets' hypothesis'[4] states that all information that can possibly be known about a stock is directly included into a stock's valuation. That means that the markets cannot be beat.

II. ENVIRONMENT AND IMPLEMENTATION

Utilizing agent-based modeling (on the *Netlogo* platform[5]), artificial 'broker' agents are fed historical Standard and Poor's financial data[6] in real time. Brokers obtain the values that the stocks reach, and then internally process them. Brokers will make decisions based upon the location of the values they process. Broker agents will be able to purchase stock, hold it, or sell it. To simplify the simulation, we assume broker agents cannot purchase multiple shares of the same stock, although one can easily

implement this in a future simulation. It is assumed that, for the purposes of modeling, equity calculations would detract from fundamentally exploring automated choice algorithms. Giving weight to a investment choice (with equity) will be a critical step in modeling the potential profit of investment decisions, as any miniscule profit can be grown by multiple orders of magnitude with the use of leverage. Another simplification is that different brokers will own simultaneously shares of the same stocks, but their ownership will not affect the stock's price (as it could in real life). Brokers will not communicate with one another, and their actions will have no impact on the strategies ofother brokers. Broker agents will start with a small amount of base capital ($1,000 simulated dollars), with which they will purchase stocks. The program will record differences in prices at which the stock are bought and sold, monitoring the broker's performance by adding or subtracting from the broker's capital. Brokers will want to make as much money as possible while the model is running. Different broker agents will utilize different strategies to capitalize upon the movement trends of the 'market'. They will want to avoid being locked into a stock that is moving against their strategy. We will examine how much randomness exists between the movement of our 'stock' agents and our 'broker' agents. If they fare well (that is, they add to their base capital), it will be revealed that the model is somewhat deterministic. It is fairly obvious that functional improvements could be gained with added layers of complexity (such as equity, intermarket rises, and brokers learning from one another), although for the purposes of this model those will not be included. We simply intend to model the effectiveness of algorithmic trading as an idea.

To illustrate one approach to algorithmic trading, we implement a strategy known as statistical arbitrage.[7] Using a basic confidence interval, a 'band' is created around the stock's price (graphically speaking). When a stock's prices falls below this average range, the broker agent purchases the stock because it is likely that it will rise to be contained within this range again. When the stock rises above this range, the broker sells the stock, as this high price is expected to once again fall back within the mean range. The initial choice of which stock to buy will depend on how far below the mean a stock falls right after the initial confidence interval is established. The size of the confidence interval (that is, the number of standard deviations used as a threshold) is randomized across brokers for the purposes of experimentation and strategy optimization. This demonstrates an array of stock choices, considering the variation of parameters between agents. The strategy is considered fundamentally contrarian because the stock may be bought when it is falling, and sold when it is rising.

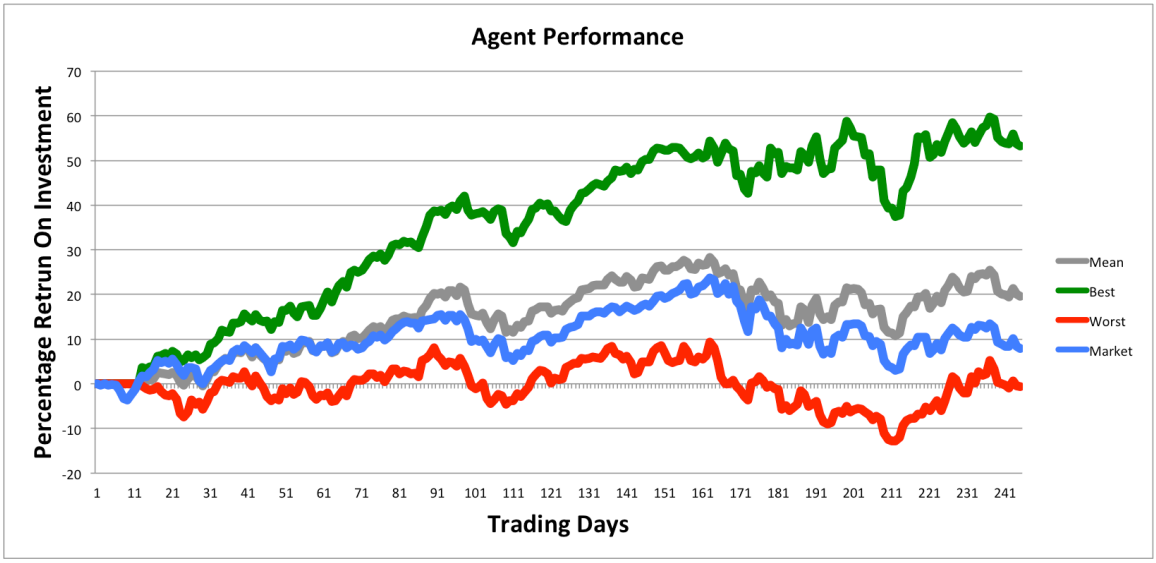
III.RESULTS

Netlogo utilizes a tool called *BehaviorSpace* for parameter sweeping. In our model, we used this tool to run the simulation of statistical arbitrage for ten iterations with the maximum number of stocks set at 10, 20, 50, 100 and 200. The number of broker agents was held constant at 20.

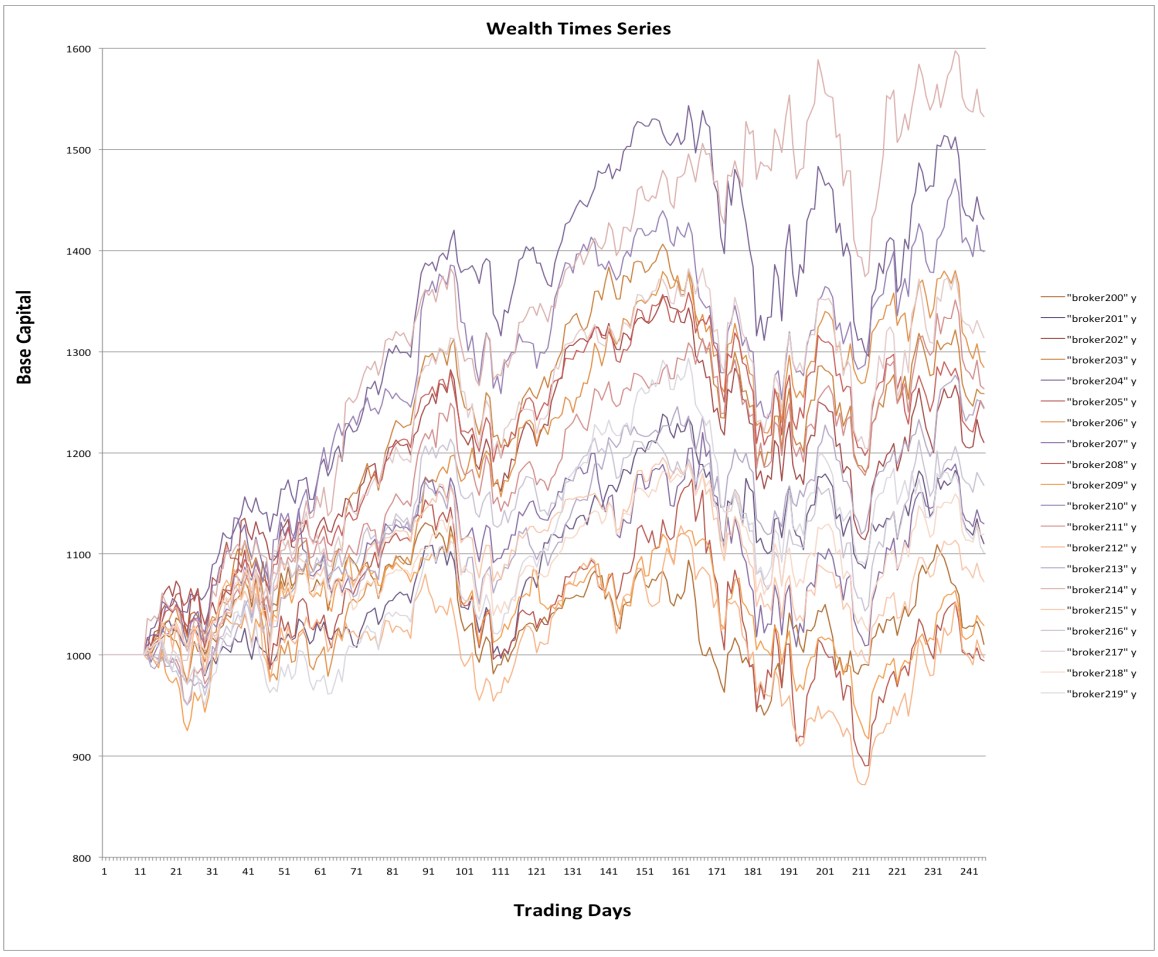
Maximum Number of Stocks Available	*Average Number of Brokers out of 20 whose Investment Choices 'Beat the Market'*
10	7.93
20	7.49
50	11.16
100	11.05
200	12.93

Initial results indicate that algorithmic trading strategies can outperform an index of S&P 500 equities with greater likelihood when given a wider range of stocks. These figures illustrate one relevant factor that influences the success of algorithmic strategies: the size of available portfolio choices. In markets characterized by limited time-series data, brokers tend to miss optimal buy- or sell windows. Conversely, in markets characterized by few portfolio choices (i.e. fewer available stocks) automated brokers have fewer opportunities to identify price movements that yield profits. This highlights an interesting and counterintuitive result: automated

A. Sample Return on Investment for Simulated Brokers, 200 stock choices maximum.



B. Agent's growth of initial starting capital, 200 stock choices maximum.



brokers require some volatility in price movements to yield profits that outperform the market index. With more frequent data points and a wide variety of stocks to pick from, the computer is limited only by its own processing power in its ability to capitalize upon this volatility.

IV. OTHER POTENTIAL TRADING ALGORITHMS

While the above results suggest algorithmic strategies may outperform market indices, they are not exhaustive. There are as many algorithms and methods for determining investment decisions as there are statistical methods and mathematical derivations. This list is intended to be the summation of the most practical frameworks/assumptions that could be adopted in the implementation of an algorithmic trader. This document specifically refers to the algorithmic trader simulation currently under construction in the Netlogo programming environment. The undertaking of predicting probability across an infinite number of potential environments results in unyielding and exponentially growing complexity.

A. Legend/Interpretation

B= DECISION TO PURCHASE A STOCK (BOOLEAN)

BC= BASE CAPITAL (AMOUNT OF MONEY BEING HELD)

C: CONFIDENCE INTERVAL (STANDARD DEVIATION PARAMETERS WILL BE EXPERIMENTED WITH)

E= EQUITY (NUMBER OF SHARES PURCHASED)

I= DENOTES PRESENCE OF CONSISTENT ITERATIONS OF MOVEMENT (ABSTRACT VARIABLE)

P= DENOTES PRICE MOVEMENT OVER TIME (ABSTRACT VARIABLE)

S= DECISION TO SELL A STOCK (BOOLEAN)

V= VALUE (STOCK'S CURRENT PRICE)

i, ii= DENOTES PRESENCE OF SUB-ARRAY; STOCKS THAT ARE RELATED (PAIRS)

B. Peak and Valley Recognition

This strategy is based upon simply utilizing the natural speed and breadth of computational systems; the powers of massively parallel decision-making. The broker invests in all possible stocks simultaneously. If there is a limit to the amount of stocks that can be bought (because of initial starting capital), stocks will be bought at random until capital runs out. Each stock will be held until it begins to fall. The broker will automatically purchase the cheapest stock that has recently risen- as soon as their excess capital allows it.

$$B \; if \; V^{(P+1)} = BC$$
$$S \; if \; V^{(P-1)}$$

C. Equity Decision/Random Walker

The equity (number of the *same* stocks purchased) of each buy/sell decision can be altered in this strategy. We ignore equity in other strategies to avoid distortion/misrepresentation of a strategy's guiding principles. In an active, real-world model these strategies could be combined for greater effectiveness, but this added layer of complexity would require analysis of its own and would thus detract from the examination of algorithmic trading fundamentals. When equity actually *is* included in the model, the strategy will focus upon that principle exclusively. This strategy will be a highly 'simulated' strategy, simply to illustrate a mathematic dynamic. In our Netlogo model, equity will be modeled on the broker side by using a multiplier.

The strategy is based upon the assumption that a stock's rise or fall is truly random; in any given simulated trading day it has a 50% chance of either rising or falling. If any homogenous movement exists, it automatically creates a higher probability for the movement's mirror effect to happen at the following data point. Thus, a stock that continuously rises will experience with each iteration a higher chance of a fall. The way to capitalize upon such movement is to give weight to buy/sell decisions. In the context of the stock market, this will be to purchase multiple shares based upon this 50/50 likelihood. This means that our broker agent will seek out and purchase stocks that have moved the most consistently downwards. The number of negative movements required before purchasing will vary across brokers. The multiplier for the number of stocks to purchase will grow with each iteration of downward movement. This multiplier could potentially be alterable for experimentation. The stock will then be purchased and held at that

level of equity. The stock will be sold when it begins to fall again, which will also reset the multiplier for that stock.

$E = I * V$
$B_E \; if \; V^{(P-1)=I}$
$S \; if \; V^{(P-1)}$

D. Pairs Trading/ Correlation

Correlation is the presence of simultaneous movement. It is easy to recognize that certain sectors of the economy will naturally move in tandem. This can be measured at varying levels of micro-andmacro-level movement. There is significant speculation in the market investigating spurious versus non-spurious correlation.

If $V_i^{(P+1)}$
while $V_{ii}^{(P+1)}$
then
BV_I
while $V_{ii}^{(P+1)}$

and likewise:

$SV_I \; if$
$V_{ii}^{(P-1)}$

This is a highly simplified strategy that would simply insert instances of simultaneous movement into the algorithmic trading program's memory. However, if correlation is blindly asserted without parameters, it is assumed that spuriousness will soon overwhelm the system, leading to many unprofitable investment decisions. It is for this reason that differing strengths of correlation will have to be utilized and experimented with so as to find optimal ranges of correlation strength. There is a vast array of methods for determining the correlation of pair, group, or population. Utilizing the varying formulae (for covariance, regression, etc…) will be a key component of the actual investment research process, but for the purposes of this simulation 'pairs trading' has not yet been explicitly modeled simply because the sheer size and complexity of the calculations would quickly overwhelm Netlogo, a platform that at best runs on a multicore personal computer. For example, even using the simple algorithm described above would require Netlogo to establish this relationship for every pair contained in the sample of stocks. If this were expanded in an attempt to measure the performance of the sample as it related to the population as a whole, entire matrices would have to be created for each individual stock's relationship to all of its peers. One can extrapolate how this might be too demanding for a single-computer simulation; research in this arena would be performed more easily in a wider-scale research context. These correlations would be better suited to a cluster or super-computing environment, which would allow for the thorough exploration of the differing levels of analysis (index, sector, or perhaps international correlation).

E. Constant Wavelet Transformation

The idea behind this strategy is that a 'spline of best fit' is to be made along the stock's time-series distribution of prices[8]. A stock would be bought or sold at the inflection point of our theoretical spline. This strategy would thus be able to take into account yet another important dynamic stock pricing: the relative speed/rate change in a stock's price. Utilizing this 'time' element could provide a crucial tool in price prediction: bubbles are more likely to be indicated by excessively fast growth, whereas slow and consistent growth would indicate a more dependable investment (with a longer time horizon).

IV. CONCLUSION

A. Implications

Trading on such simple rules may only provide minuscule profits on each individual trade, but as they are performed at such a large scale and with great speed (multiple orders of magnitude larger than a human could achieve) a consistent and potentially large profit could be realized cumulatively, across a multitude of trades. While our prototype model demonstrates the feasibility of our approach, it lacks the speed and scale to test competing trading algorithms; to incorporate real-time market data; and to simulate their effect on asset prices, such as the possibility of 'bubbles' and market crashes. With the assistance of more powerful computing to help us solve the problems of calculation space and time-scales, we can achieve the next three main objectives of the project:

1. *The algorithmic trading strategies to be utilized.*

The above strategies can be optimized endlessly, via the application of advanced simulation techniques such as genetic algorithms[6]—a form of artificial intelligence. The only upward limit on the sophistication of a strategy is the researcher's understanding of complex statistical methodologies.

2. *The optimum values to be used as parameters within these strategies.*

We can automate our "brokers" to experiment with values used in a strategy, such as the size of a confidence interval beyond which to execute trades, and then use these results to find what kind of behavior produced the most profitable results. For example, which strategy is more profitable—frequent low-margin trades or infrequent large-return trades? Our simulation can answer such questions by experimentally varying trader strategies, including the frequency of their trades and their "discount rates" (the value they attribute to expected future gains from ownership of an equity). This will be a major part of the research, as more time given towards such a study will doubtlessly produce insights that can be capitalized upon in further implementations. This is done by running simulations repeatedly with different values, creating a function for the program to monitor its performance as measured by a trader's profits.

3. *The various methods to interface such a program with a real-time market.*

Can the algorithms we develop produce profitable returns in the real market? To answer such a question, we need to develop the interface between the simulation and real-time market data. This interface is potentially "two-way": the software needs to read data from real-world markets, and then to send information to the market to execute real trades. The latter raises a host of regulatory, legal, financial and ethical concerns that are beyond the scope of this study. Yet these concerns have not preempted other firms from engaging in the exploitation of these principles[9]. The ability of the software to interact with real-world data is both a feasible and a necessary step toward understanding how automated traders interface with equity markets. Trade execution is a fundamental and complex portion of the algorithmic trading process[10]. This is our most complex technical challenge. However, it is not an insurmountable one. Research will be conducted to discover the most effective methods of interfacing with web-based market data, but it should not be too difficult to translate our program into a programing language that is more capable of such interactions. The simulation is currently written in the Netlogo programming environment/language, which is specifically designed for research into complex social behavior.

B. Outcomes

Ideally, we would implement a system upon a supercomputer (or cluster of computers) that made simulated trades utilizing these strategies in real time. Such computational resources would permit us to conduct ongoing optimization research, as the simulated traders implemented new strategies and simulated analysis produced the most advantageous parameters to be utilized. Even research on static data requires a high level of personal computing power; conducting actual trades would require even more. With computers experimenting with values and running simulations upon the data constantly, algorithmic traders could "learn", with the program able to update its own code based upon successful experimental results (effectively learning as a human would). Over time, the study could create "smart" automated brokers that use evolutionary computation to learn to beat the market. Agents theoretically could connect to an online brokerage/clearing house to make real-world trades, and then create some kind of consistent income. This process will begin by building models inductively, testing for effectiveness Features and functionalities can be added over time. Baseline algorithms will be created within a field (high-performance investment strategy) of upwardly spiraling complexity. The potential for expansion is truly vast.

REFERENCES

[1]"Black Box Traders Are on the March." *Telegraph.co.uk*. The Telegraph, 27 Aug. 2006. Web. 24 Feb. 2012. <http://www.telegraph.co.uk/finance/2946240/Black-box-traders-are-on-the-march.html>.
[2]"Ahead of the Tape." *The Economist*. The Economist Newspaper, 21 June 2007. Web. 24 Feb. 2012. <http://www.economist.com/node/9370718?story_id=9370718>.

[3]Duhigg, Charles. "Artificial Intelligence Applied Heavily to Picking Stocks - Business - International Herald Tribune." *The New York Times*. The New York Times. Web. 24 Feb. 2012. <http://www.nytimes.com/2006/11/23/business/worldbusiness/23iht-trading.3647885.html>.
[4] Fama, Eugene. "The Behavior of Stock-Market Prices." *The Journal of Business* Jan 38.1 (1965): 34-105. JSTOR. Web. 3 Apr. 2012. <http://www.jstor.org/stable/pdfplus/2350752.pdf?acceptTC=true>.
[5] http://ccl.northwestern.edu/netlogo/
[6] "Historical Data for S&P 500 Stocks." Trademan. Web. 4 Apr. 2012. <http://pages.swcp.com/stocks/index.html>.
[7] Avellanda, Marco, and Jeong-Hyun Lee. "Statistical Arbitrage in the U.S. Equities Market." NYU, 11 July 2008. Web. 30 Mar. 2012. <http://www.math.nyu.edu/faculty/avellane/AvellanedaLeeStatArb071108.pdf>.
[8] Palshikar, Girish K. "Simple Algorithms for Peak Detection in Time-Series." Tata Research Development and Design Centre (TRDDC). Web. 30 Mar. 2012. <http://www.tcs-trddc.com/trddc_website/pdf/SRL/Palshikar_SAPDTS_2009.pdf>.
[6]Holland, John H. 1992. “Genetic Algorithms.” Scientific American 267 (July): 66-72.
[9]"Moving Markets." *The Economist*. The Economist Newspaper, 02 Feb. 2006. Web. 24 Feb. 2012. <http://www.economist.com/node/5475381?story_id=E1_VQSVPRT>.
[10]Nuti, Giuseppe, Philip Treleaven, and Chaiyakorn Yingsaeree. "Algorithmic Trading." *IEEE Computer Society*. UK Centre in Financial Computing. Nov. 2011. Web. 24 Feb. 2012.

Exploration and Evaluation of Enterprise Innovation

Mariusz A. Balaban
ODU
mbala004@odu.edu

Patrick T. Hester, Ph.D.
ODU
pthester@odu.edu

***Abstract*— This paper explores the topic of enterprise innovation for evaluation, insight, and experimentation. The authors proposes a conceptual model of enterprise innovation, to assist in understanding enterprise innovation and as a backbone for evaluation of enterprise innovation capabilities. The proposed conceptual model is further explored using the Enterprise AID methodology [1] to evaluate enterprise innovation. Additionally, this paper provides a sample model as a test-bed for initial investigation of the benefits and limitations of the multi-method M&S approach in satisfying the need for simulating enterprise innovation.**

***Index Terms*—innovation, capability, performance, effectiveness**

I. Introduction

DURING the 20th century, and the beginning of the 21st century many theories, views, and perspectives on innovation have emerged[2-10]. These have spanned through many streams of research not limited to evolutionary theories, organizational theories, as well as economic, socio-political, cognitive, behavioral, and systemic views. Often, those different dimensions and domains of analysis overlap. Innovation can be classified based on novelty level [11] as incremental or radical; based on type as product, process, marketing, or organizational (business) [12]; based on degree of novelty diffusion as new to the firm, new to a market, and new to the world [12]. Theoretical models and their evolution at different levels of analysis have been developed, and for the purpose of this study can be summarized as follows: at a project level [13], at a firm level [14], and at a regional level [15]. The inherent complexity of innovation makes validation or further development of proposed theories difficult. Innovative enterprises are usually involved in Research and Development (R&D) activities, developing technological capabilities, and in significant marketing competencies [16]. Innovation can be linked to performance, evolution, growth, change, and sustainability of an enterprise [12].

Recently, open innovation paradigm gained its momentum. It sees knowledge inflow and outflow as a mechanism accelerating internal innovation, and allows for expanding new markets by external use of innovations [9, 17, 18]. For example buying a license to a technology or a whole company that owns desired technology has become a common approach. Similarly, selling Intellectual Property (IP) may be a good practice for getting extra money for not otherwise used knowledge. Spillovers are especially popular realization of open innovation, because they are relatively inexpensive or free, and often easy to obtain from universities, trade fairs, public and private research institutes, market partners, conferences, competitors, trade, and patent literature. Similarly customers, especially lead users, can provide valuable feedback leading to innovation [11].

One factor that distinguishes enterprises is their ability to innovate. On the one hand, decomposition of different types of innovation and understanding their effects on enterprise performance are problematic; on the other hand, it seems likely that current innovation capability of an enterprise defines its future position in the market, its value, and a perception of the enterprise as seen by its stakeholders. Measuring innovation can be a key indicator of the enterprise value and future prospects. It can also point to internal problems of the enterprise before they emerge. Anthony, Johnson, Sinfield, & Altman warn that too few innovation measures might not capture innovation trends [19]. Similarly, too many metrics can be confusing. Alegre, Lapiedra, & Chiva proposed two general measures to evaluate innovation: (1) efficacy, as the degree of success of innovation; and (2) efficiency, as the degree of difficulty in showing how hard it is to achieve goals [20]. Proposed by Essmann and Du Preez, the Innovation Capability Maturity Model (ICMM) claims to assess organizational innovation capabilities "reliant on the organization's employees relaying the internal situation via the questionnaire" [21]. The authors agrees with the general direction of this research, but the limitations of the proposed approach should be pointed out: metrics may lack objectivity due to using employees for assessment, and metrics used for comparing innovation capabilities based on enterprise goals cannot be used to compare between enterprises with different goals. In addition, it is impossible to use this approach to capture effects of innovation dynamism and to conduct dynamic experimentations. Enterprises come in different sizes, shapes, and forms, which makes comparison between them difficult. Teece, Pisano, & Shuen point to the dynamic capabilities of a firm as a determinant of its strategy [22]. Lack of a common quantitative metric of enterprise innovation that could be used to measure future prospects of an enterprise prevents easy comparison of enterprises. In light of the above, it seems appropriate to ask the question of how to measure innovation of different types of enterprises.

II. Key Terms

Before beginning a substantial exploration of innovation, it is necessary to define several key terms central to this discussion as follows:

- *Innovation* in this research is defined based on definitions of Meyers and Hester [23], and OECD [12]. The definition is tailored to the scope of enterprise level as follows: innovation is the exploitation, implementation of, or value creation from: new ideas generating a product, a process, a new marketing method, or a new organizational method in business practices, workplace organization, external relations, or their combinations. The terms "innovation" and "innovation capabilities" are used interchangeably throughout this research.
- The definition of *enterprise* used in this work follows a common dictionary description as undertakings, especially one of some scope, difficulty, complication, and risk [24].
- *A Multi-method M&S* is an approach that pertains to a combination of at least two M&S paradigms, for example: Discrete Event Simulation (DES), System Dynamics (SD), Agent Based Modeling (ABM), and any additional methods, e.g. Bayesian Networks (BN).

The following three definitions central to enterprise performance derived from the work of Hester and Meyers [1] will be used throughout this work:

- *Critical Operational Issues (COI)* are defined as "stakeholder needs identified and that must be satisfied for problem resolution; the emergent essentials of capability without which the posited problem solutions must be judged as unacceptable on functional grounds" [1].
- *Measures of Effectiveness (MOE)* are defined as "standards derived by stakeholders from stakeholder-identified COI, independent of solutions proposed for COI resolution but representing emergent properties that induce rank ordering on solutions proposed to satisfy critical needs" [1].
- *Measures of Performance (MOP)* are defined as "evaluations of systems' intrinsic functions by which can be judged, using MOEs, those systems' capabilities to ultimately resolve COI from which the MOE derive" [1].

III. M&S of an Enterprise as a Toolbox for Modeling Enterprise Innovation

The multi-method M&S approach allows for modeling of a system that comprises many levels of abstraction, allowing for representation of a system's elements in their more natural way of functioning. Enterprise modeling using a combination of M&S paradigms is a relatively new approach [25, 26]. Until recently, approaches to the multi-method were limited and rather difficult. Often, it was easier to adapt one paradigm and use it for elements that would have been modeled easily with a different paradigm, despite possible problems related to accuracy, simulation inefficiency, and usually higher modeling effort. Because of the problems mentioned, realization of the power of the multi-method approach emerged rather slowly [27, 28]. Some practitioners tried to build a custom communication interface (often using a third party database for exchange of data) between different Commercial of the Shelf Simulation Packages (COTSSP) [26], and some using High Level Architecture (HLA) [29]. Both groups analyzed potential benefits of using both DES and SD paradigms to simulate manufacturing enterprise systems. Rabelo, Helal, Jones, and Min [26] showed the benefits of integrating an operational level with various levels of management to capture long-term effects of decisions at various levels in an enterprise. Venkateswaran and Son described hybrid Hierarchical Production Planning (HPP) architecture, consisting of an aggregate planning level and a detailed scheduling level. The other domain where the multi-method approach has been applied is healthcare. The applicability of a hybrid simulation in healthcare was examined with a positive outcome by Chahal & Eldabi [30], and further analyzed based on two case studies by Brailsford, Desai, & Viana who see progress in combining DES and SD, but also concede that M&S practitioners are "a long way from genuinely combining these approaches" [31].

The most relevant study found by the authors with a connotation of modeling enterprise innovation using M&S was conducted by Bodner and Rouse [32]. This study focused on R&D value creation and valuation methods, and employed a single DES paradigm. The model theory was based on a "push" approach [33], which itself cannot represent enterprise innovation characteristics [14], but addresses a very important aspect: funding. A model of an innovative firm that used the SD method was proposed by Garcia [34]. Innovation diffusion processes have been modeled using SD [35, 36] and ABM [37-39]. An interesting model is proposed by the vendor of modeling software AnyLogic, which used an ABM paradigm in their "Portfolio model" that shows significant potential for further improvement and combination with other paradigms such as DES, SD and even BN. When some functionality does not exist in AnyLogic's library, one can add functionality by looking for or developing an appropriate Java code. This powerful Object Oriented Programming centric M&S software was used by Glazner in his dissertation [25]. Glazner used a combined DES/SD/AMB approach to model enterprise architecture. Several other elements pertaining to an enterprise were modeled using a single or multi-method M&S.

Models of an enterprise can require high levels of complexity and an appropriate level of detail to capture adequately the problem space related to innovation. A model with too much detail can be hard to understand and analyze. Finding the right balance in abstracting non-essential elements and focusing on important ones is an important rule to obtain a valid, efficient model. On the one hand, modeling of enterprise innovation by using combined endogenous and exogenous activities using multi-method M&S seems more difficult and less tangible than the referenced enterprise models. On the other hand, the multi-method simulation combining DES, SD, and ABM makes a powerful set of methods for modeling enterprise innovation, and it is advocated by the authors.

IV. The Need for Enterprise Innovation Evaluation

Assessments of the long-term effectiveness of enterprises are marked with uncertainties and risks. Evaluation of an enterprise should encompass a perception of stakeholders on its potential for profitability, sustainability, or ability for fulfillment of its mission. Evaluation of innovation of an enterprise would allow an individual to determine a long-term prospect for stakeholders, and would provide the ability to draw comparison between enterprises. Unfortunately, questions concerned with how much innovation capabilities affect the value of an enterprise are difficult to answer. For instance, there may be a case when one hypothetical enterprise "A" can be very innovative, but currently its performance measures are less than impressive because of heavy investments in a new product, which is still in the development phase. Hypothetical enterprise "B" became a very stale innovator; it employed restructuring that minimized spending on R&D and training. Even if its product is still very profitable, when the enterprise "A" starts diffusion of the new "killer" product, the enterprise "B" will lose their competitive advantage, which then will be reflected in diminishing profitability, and inability to sustain on its market. The ability to compare both enterprises in terms of their innovation, and the ability to look at the dynamic processes of change inside and outside of both enterprises would allow for improved investment decisions.

Another case where measuring and modeling of innovation capabilities can be beneficial is in internal enterprise decision-making processes. Oversight of a large enterprise involves decision making at strategic, tactical, and organization levels. Each of these levels can have significant effects on enterprise innovation, and then usually delayed effects on enterprise performance, and MOEs. Decision makers often lack tools that would allow for dynamic experimentation with an enterprise using factors that are harnessed to innovation and its activities, e.g. approaches to funding, HR policy, resources allocation, knowledge policy, organization alignments, R&D and product development processes, and marketing strategies. Not all elements can, or should, be addressed using M&S, but elements with a connotation of innovation emergence and diffusion are hard to address using only questionnaires, and static, analytic methods, therefore employing M&S seems more appropriate.

Challenging aspects of this research include linking innovation to its influencing factors, and integration of different views on innovation. Inter-relationships of these factors should be investigated in order to model enterprise innovation. Many scientists from different fields of science have described innovation using theoretical models [2-10], but these theories are difficult to validate because of their complexity and difficulty in capturing and analyzing the dynamic aspects of innovation. Complex dynamics make a closed-form analytical description of innovation mathematically challenging. In addition, empirical testing has a limitation regarding addressing "what if" questions. Based on the discussion above, there is a need for a methodology and a useful test-bed in the form of a simulation model, which allows for the conducting of experimentations pertaining to enterprise innovation.

The first step in filling the aforementioned gap was creation of a high-level conceptual model of enterprise innovation with a goal of its broad applicability to most, if not all, enterprises [40]. Hence, during a comprehensive literature review, the authors of this work searched for factors common to all enterprises, with specific regard to those relevant for capturing the complexity of innovation. Based on the initial conceptual model, hypothetical logical relationships of the main factors were proposed for better understanding of dependencies and complexity of the model.

This research looks at the multi-disciplinary methodology of modeling enterprise innovation by incorporating perspectives from several domains such as strategic management, organizational studies, economics, and engineering. The authors believe that this is necessary in order to build an enterprise innovation model that considers a broad range of viewpoints on innovation and factors included in the model. The multi-method M&S approach seems especially valuable in representing enterprise innovation, because it provides flexibility in experimenting with different system views and configurations. It can provide an experimental setting in the form of a simulation model of an enterprise and its surrounding environment, with representation of internal and external influences related to innovation. Additionally, a multi-layer approach can be necessary to model different views on innovation, with each layer describing a view catalyzed by a related concept or a theoretical model drawn from different domain, and represented in the most effective way using an appropriate modeling paradigm.

V. A Proposed Conceptual Model of Enterprise Innovation

The initial goal of this research was selection of main, high-level factors that could describe enterprise innovation. A detailed discussion of the conceptual model, its factors, and logical relationships of those factors are described in un unpublished work [40]. Presented below conceptual model of enterprise innovation can be implemented using different techniques. The very brief description of the conceptual model follows next.

Different factors of innovation and their arrangements were proposed in the literature [7, 41-44]. The main factors proposed by the authors and the basic structure of the enterprise innovation model are shown in Figure 1. Innovation capabilities are conceptualized as a central part of the model, and serves as a major link between inputs and outputs. They can be represented as a four-item tuple:

$$I_c = (P, S, M, O), where:$$

I_c is innovation capabilities, P is product innovation capabilities, S is process innovation capabilities, M is

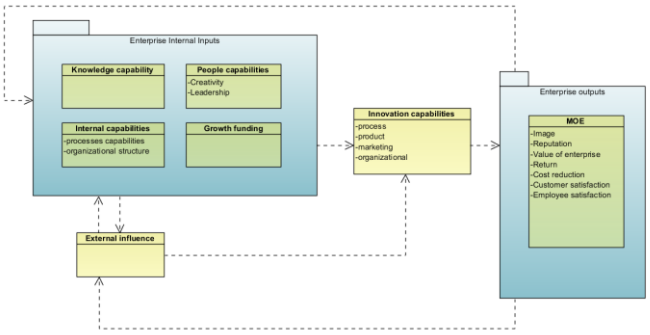


Fig. 1. Enterprise Innovation Model

marketing innovation capabilities, and *O* is organizational innovation capabilities.

On the input side, innovation capabilities are related to enterprise capabilities expressed by sum of its current state and internal change caused by the internal and the external factors:

$$I_c = C + \dot{C}\, where:$$

C is initial enterprise capabilities, and $\dot{C}$ is a change of enterprise capabilities that is possible due to enterprise innovations. Because all inputs can affect I_c we could say that:

$$I_c = (K, R, A, F, E) = E_c = C + \dot{C}\, where:$$

K is knowledge capabilities, *R* is people capabilities, *A* is internal view capabilities, *F* is funding capabilities, *E* is external influence, and E_c represents enterprise capabilities. Dependencies created between internal and external factors cause the enterprise change ($\dot{C}$), spurred by, for example, new demands posed by internal stakeholders or influenced by changes in external environment, for example, customers' needs or competition. The general idea of the model is a representation of innovation capabilities so that the output explains input using the feedback effect. The set of MOEs is usually different for different enterprises, so the proposed elements that are presented in Figure 1 are only used for better depiction of the idea. The input factors proposed in the conceptual model should be further analyzed and modeled at an appropriate level of abstraction. This stage will require identification of model parameters, particular to the enterprise under the study, and used to control of the high-level input factors, affecting enterprise innovation, and ultimately its performance.

VI. Evaluation of Enterprise Innovation

The Enterprise AID methodology proposed by Hester and Meyers [22], combined with the conceptual model proposed by the authors, were used to assess enterprise innovation capabilities. The Enterprise AID method can provide a backbone for the M&S phase, and a structure allowing for quantitative evaluation of enterprise innovation. The Enterprise AID methodology was extended to accommodate the needs of this work. Figure 2 depicts the proposed extension. For detailed explanation of the Enterprise AID process, please refer to [1].

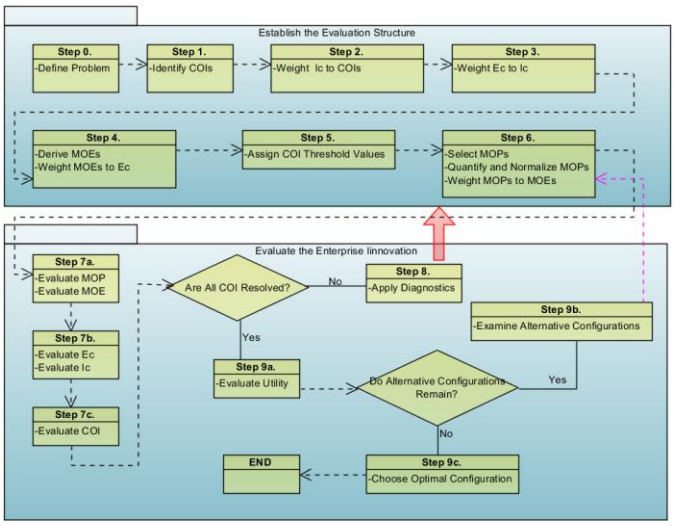


Fig. 2 Extended AID Application Flowchart based on [1]

Step 0 formulates the problem, which in this case, can relate to ensuring an appropriate level of innovation capabilities to accomplish mission objectives of an enterprise. Step 1 identifies COIs, all of which must be satisfied to resolve the problem. Every enterprise is different, with emphasis on the different type of innovation capabilities, hence in step 2 SMEs independently assign weights between COIs and the four types of innovation capabilities that enterprise can generate using, for example, the five-level relevance scale ranging from the insignificant level to the essential level. Obtained average weights are then converted, so their weighted sum equals to one. In step 3, SMEs independently assign weights between the proposed conceptual model main inputs, representing enterprise capabilities (E_c), and the four types of innovation capabilities, in the same manner as in step 2. In step 4, MOEs are derived such that the enterprise capabilities can be defensibly assessed. After reviewing all of the E_c-MOE correspondences, SMEs independently assign weights between all pairs in the same manner as in step 2. Next, in step 5 SMEs assign threshold values to COIs, which are recalculated for the "margin of safety", and will be used to determine the resolution of COIs. In step 6, MOPs are selected, quantified, and normalized. Next, in the same step SMEs assign weights between all the MOE-MOP pairs, in the same manner as in step 2. Evaluation of MOPs, MOEs, E_c, I_c, and COIs are conducted in steps 7a, 7b and 7c, which reveals the resolved or not resolved status of COIs. Extended AID evaluation structure model, with only one COI and sample MOEs and MOPs, is shown in Figure 3. Step 8 allows learning about possible causes of the unresolved COIs, ranking them in the order of significance, which points toward potential responses. Step 8 in the case of innovation evaluation could mean something like: What innovation capabilities require

critical attention, and how to improve them in the current situation? or; What additional innovation capabilities would allow to achieve the mission objectives? Steps 9a, 9b, and 9c evaluate the possible scenarios for different values of MOPs, evaluating utility values that allow for selection of the best enterprise configuration.

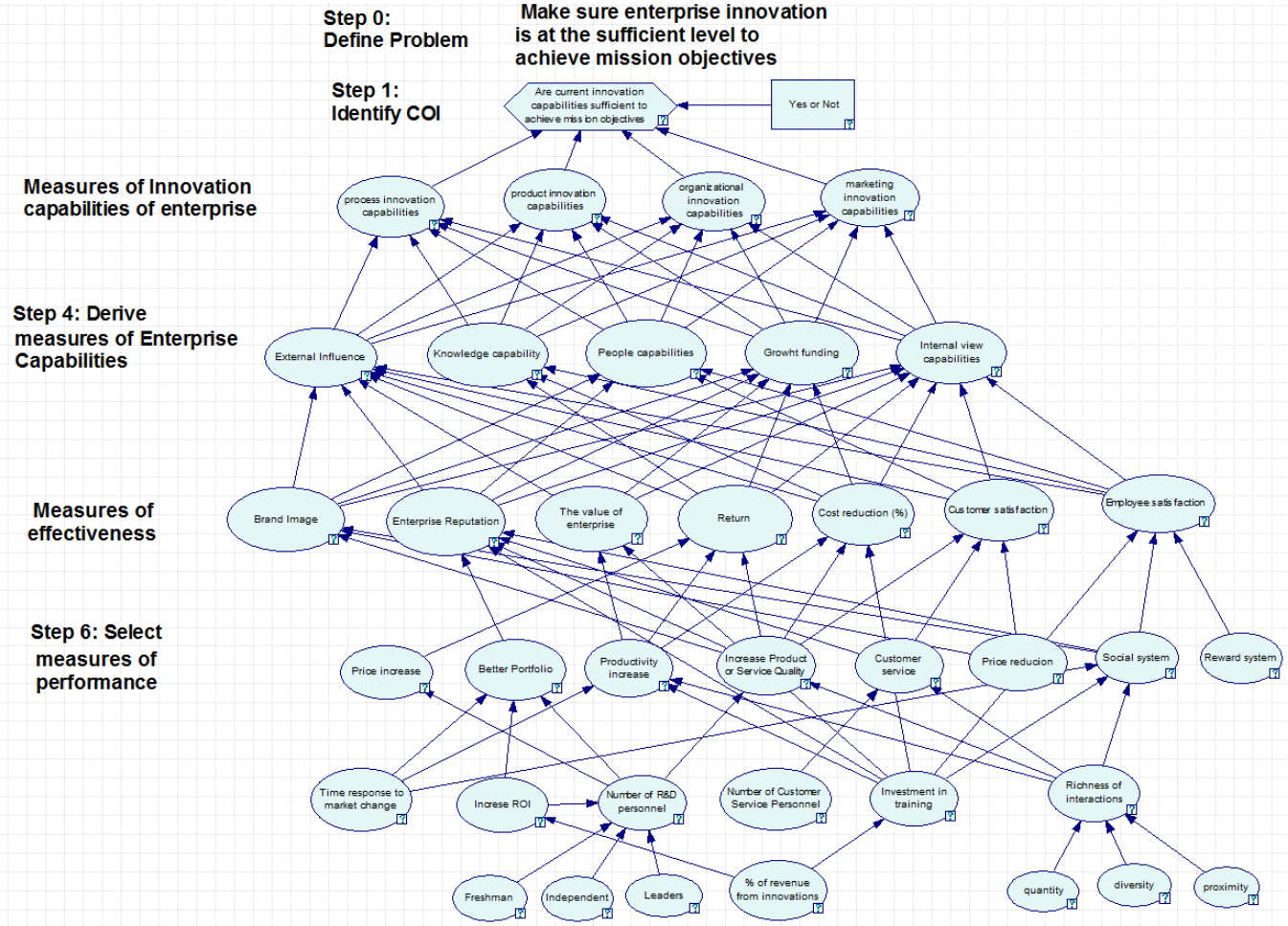


Fig. 3 Extended AID structure model for innovation enterprise evaluation

The Enterprise AID methodology was extended by adding two layers: the innovation capabilities layer and the enterprise capabilities layer. All the factors present in these two layers consist of the core of the model that should not change. Adjusting the weighted vales of these two layers allows for adjustment of the model based on different problems and consequently different COIs. In the case of comparing two or more enterprises based on some common problem or perspective, the model structure, and weights between COI, I_c, E_c, and (derived) MOEs layers should be the same in order to derive a meaningful comparison. In summary, the Enterprise AID methodology was extended based on the conceptual model of enterprise innovation. It is believed to be possible that proposed approach can be used to evaluate the enterprise with this revised model to compare enterprise innovation capabilities based on a given problem or perspective.

VII. Simulation Model Example

The purpose the simulation model was the evaluation of the usefulness of the multi-method M&S in representing enterprise innovation, versus using a single modeling paradigm. The developed model is not based on a real system, but the main structure of the model is common for modeling project portfolio and R&D processes [32]. The focus of this section is to address benefits, and problems of using the multi-method M&S to represent enterprise innovation.

DES seems useful when looking at enterprise innovation from the processing perspective. Usually, an enterprise has to manage many innovation projects, evaluate them, and properly allocate resources. DES allows for easier representation and analysis of projects throughout their life cycle compared to SD or ABM because, it has predefined blocks with appropriate functionality. Also, the changing logic of model in DES is faster than the other two paradigms. The effort to build a

queuing system using ABM is significant. Development of ABM normally involves coding most needed elements. When a change of logic is required, which often happens in real problems, more time would be consumed to adjust modeled elements in ABM than it would take in the case of using DES. For this reason, the project layer of the sample model was represented using DES as shown in Figure 4. Projects are represented as entities going through their life cycle. Generated projects go through initial funnel, where based on their properties e.g. estimated ROI, and estimated development time, a certain percentage are

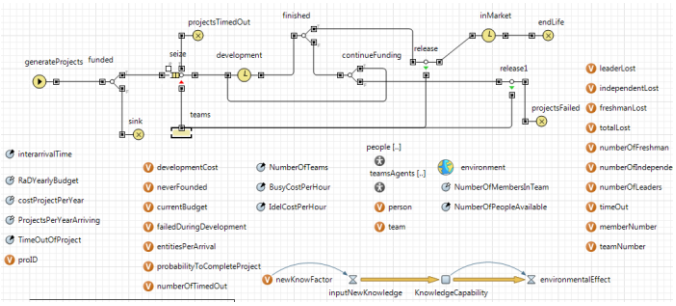


Fig. 4. Example of project process using DES

discarded. Those initially accepted for consideration await funding approval on a yearly basis, based on budget availability. Only a certain number of projects get to the development phase, and the rest of them wait until the next year's funding arrives. If the project waits for too long, it is discarded. Next is the development phase. A project is assigned a team, and then based on a yearly evaluation it can be completed, it can continue its development phase, or it can be rejected due to, e.g. a lack of funding. Finally, a project goes into the "market", or other environment for which it was developed.

From the financial perspectives of evaluation of an enterprise, this model may be used to evaluate and improve decisions about resource allocation, funding and hence the cost of the project development, versus the expected return. The limitations of using DES are apparent when trying to model among others, team characteristics, individuals, and market characteristics. The aforementioned characteristics are important for modeling people's capabilities, knowledge capabilities, and external influences of enterprise related to open innovation paradigm, as shown in Figure 1. Omissions of the factors that truly matter can cause an inaccurate representation of enterprise innovation. On the other hand, the current state of the art in representing human traits via DES is not sufficient, and more work is needed. The ABM method seems more appropriate in specifying individual behavior. A global view of the environment in which the enterprise operates can be most easily represented using SD approach, but using ABM should also be considered to represent individual behaviors of enterprises and effects of enterprises' interactions, which underline open innovation effects. In this sample model, additional layer representing the global view of enterprise's environment was not modeled. Similarly, dynamic traits of behavior of individuals (within an agent; as an ABM) can also be represented using SD, DES, or Bayesian Network.

Teams were modeled as two separate instances, one as a DES resource, and the second as an agent. A DES team is mostly used because of the ease of modeling queuing of projects based on available human resources. Behavior of the individual teams are defined as agents with internal logic based on the project dynamics model provided by Garcia [34], and used to acknowledge project completion, and estimate work remaining. The two layers can interact with each other and trigger events. Additionally, each of the agent-type teams consists of a collection of the individually behaving team members. Once the project leaves the development phase, the DES resource becomes idle, and both the agent-type team and individual agents are released as well. This dual-existence-layer method allows for the representation of the same element from different perspectives, capturing behaviors serving different purposes in the model. The dual, or potentially more, perspectives can be useful in representing additional functionality of DES's competing for a resource, or entity's complex decision-making. For example, a person standing in a store line "falls into" a queuing model specification, but the decision of balking, for example, can be based on complex behavior, and would require the ABM approach to better represent triggering that event. Figures 5 and 6 show the proposed internal representation of a team and an individual member, respectively.

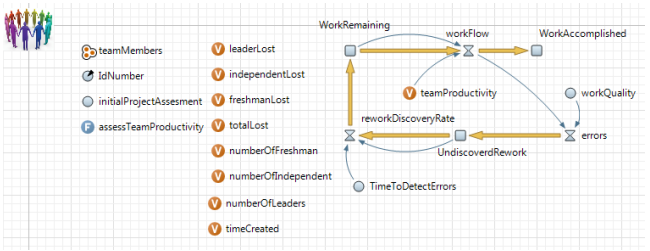


Fig. 5. An agent behavior of the project team

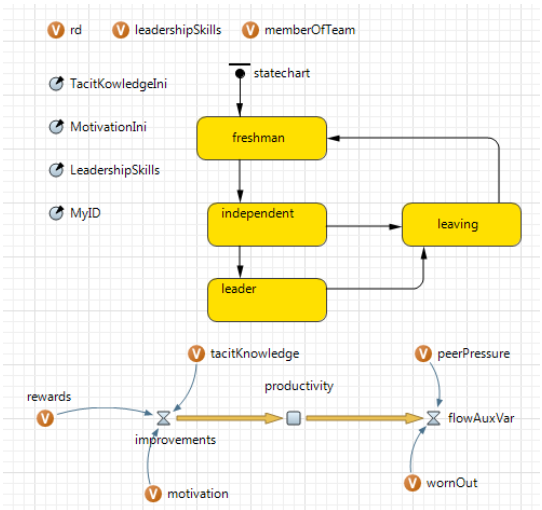


Fig. 6. An agent behavior of the individual team member

Computing power limitations can become a problem when using the multi-method M&S. When the behavior of agents is complex, and their interaction rules cause a significant number of events, then the execution time can be a problem.

Additionally, the validation difficulty of such a complex simulation model should be further discussed. Mixing of paradigms poses many challenges and questions that are still unanswered. Especially, communication between different paradigms with a clear understanding of accuracy risks needs to be addressed. Validation of simulation model of enterprise innovation using multi-method M&S should be examined based on the context of a study, addressing validation of components separately, and then simulation model as a whole. Because this model is in its early stage, it is difficult to define validation requirements. Different configurations and manners in which to assemble models are possible by using the given four paradigms. This should spur analysis of requirements for validity and credibility of the model based on each new configuration. In addition, validation of agents' behavior as humans seems especially challenging.

Simulation methods based on a single modeling paradigm often prove inadequate to represent effectively the complexity at different levels of a system (strategic, tactical, and operational). Because of large differences in scope, and time phases included in modeling of innovation within an enterprise lifecycle, the multi-method simulation combining DES, SD, and ABM makes a powerful set of methods for modeling enterprise innovation, and it is advocated by the authors.

VIII. CONCLUSION

In this paper, the authors proposed a conceptual model of enterprise innovation, to aid in understanding enterprise innovation. The extended Enterprise AID method was proposed to evaluate enterprise innovation. In addition, an initial investigation of the benefits and limitations of the multi-method M&S approach in satisfying the need for simulating enterprise innovation indicated its usefulness.

MARIUSZ A. BALABAN received his two Masters in Mechanical Engineering from Lublin Technical University, Poland. He is currently pursuing a Ph.D. in Modeling and Simulation at the Old Dominion University, where he also works as a research assistant. His research interests are R&D, innovation, governance, evaluation methods, and topics related to Modeling and Simulation, especially, the multi-method (hybrid) M&S approach. Mr. Balaban is a member of SCS, ACM SIG, and GK International.

Dr. PATRICK T. HESTER is an Assistant Professor of Engineering Management and Systems Engineering at Old Dominion University. He received a Ph.D. in Risk and Reliability Engineering (2007) at Vanderbilt University and a B.S.in Naval Architecture and Marine Engineering (2001) from Webb Institute of Naval Architecture.
Prior to joining the faculty at Old Dominion University, he was a Graduate Student Researcher in the Security Systems Analysis Department at Sandia National Laboratories, where he worked on developing a methodology for safeguard resource allocation to defend critical infrastructures against a multiple adversary threat. Prior to that, he was a Project Engineer at National Steel and Shipbuilding Company in charge of Modeling and Simulation experiments to test shipboard logistics and cargo load-out capabilities. His research interests include multi-objective decision making under uncertainty, probabilistic and non-probabilistic uncertainty analysis, critical infrastructure protection, and decision making using modeling and simulation.

REFERENCES

[1] P. Hester and T. Meyers, "Enterprise AID: A performance measurement system for enterprise assessment, improvement and design (NCSOSE-TR-2012-001)." Old Dominion University, NorfolkJanuary 23, 2012 2012.

[2] T. J. Hargrave and A. H. Van de Ven, "A collective action model of institutional innovation," *The Academy of Management Review* vol. 31, pp. 864-888, 2006.

[3] Z. Jian and Z. Jianqiu, "An Open Innovation Model for Business Innovation of Chinese Telecom Operators," in *Management and Service Science, 2009. MASS '09. International Conference on*, 2009, pp. 1-4.

[4] S. Roy, *et al.*, "Innovation Generation in Supply Chain Relationships: A Conceptual Model and Research Propositions," *Journal of the Academy of Marketing Science*, vol. 32, pp. 61-79, January 1, 2004 2004.

[5] J. Schmookler, *Invention and economic growth*: Harvard University Press Cambridge, MA, 1966.

[6] J. A. Schumpeter, *et al.*, *The theory of economic development: an inquiry into profits, capital, credit, interest, and the business cycle*: Harvard University Press Cambridge, MA, 1934.

[7] M. L. Shyu, *et al.*, "A Conceptual Model of Organizational Innovation: An Empirical Study on Universities of Technology in Taiwan," in *Management of Innovation and Technology, 2006 IEEE International Conference on*, 2006, pp. 186-190.

[8] R. M. Solow, "A contribution to the theory of economic growth," *The Quarterly Journal of Economics*, vol. 70, p. 65, 1956.

[9] L. Wang, *et al.*, "Developing a conceptual framework for business model innovation in the context of open innovation," in *Digital Ecosystems and Technologies, 2009. DEST '09. 3rd IEEE International Conference on*, 2009, pp. 453-458.

[10] Y. Wu, *et al.*, "Constructing Knowledge Innovation Models of Communication Management in Research Organization," in *Management and Service Science, 2009. MASS '09. International Conference on*, 2009, pp. 1-4.

[11] Harvard Business School Press, *Managing creativity and innovation*. Boston, Massachusetts: Harvard Business Press, 2003.

[12] OECD, *Oslo Manual: Guidelines for collecting and interpreting innovation data (2005)*: OECD / European Communities 2005.

[13] M. P. e. Cunha and J. F. S. Gomes, "Order and Disorder in Product Innovation Models," *Creativity and Innovation Management*, vol. 12, pp. 174-187, 2003.

[14] M. Hobday, "Firm-level innovation models: Perspectives on research in developed and developing countries," *Technology Analysis and Strategic Management*, vol. 17, pp. 121-146, June 2005 2005.

[15] F. Moulaert and F. Sekia, "Debates and Surveys," *Regional Studies*, vol. 37, pp. 289-302, 2003.

[16] J. R. Baldwin and P. Hanel, *Innovation and knowledge creation in an open economy: Canadian industry and international implications*: Cambridge Univ Pr, 2003.

[17] B. Calida and P. Hester, "Unraveling future research: an analysis of emergent literature in open innovation," *Annals of Innovation & Entrepreneurship*, vol. 1, 2010.

[18] H. Chesbrough, *et al.*, *Open innovation: Researching a new paradigm*: OUP Oxford, 2008.

[19] S. D. Anthony, *et al.*, *The innovator's guide to growth*. Boston, Massachusetts: Harvard Business Press, 2008.

[20] J. Alegre, *et al.*, "A measurement scale for product innovation performance," *European Journal of Innovation Management*, vol. 9, pp. 333-346, 2006.

[21] H. Essmann and N. Du Preez, "An Innovation Capability Maturity Model–Development and initial application," *World Academy of Science, Engineering and Technology*, vol. 53, pp. 435-446, 2009.

[22] D. J. Teece, *et al.*, "Dynamic capabilities and strategic management: organizing for innovation and growth," *Strategic Management Journal*, vol. 18:7, pp. 509-533, Aug., 1997 1997.

[23] T. Meyers and P. T. Hester, "Toward the what and how of measuring R&D system effectiveness," presented at the 7th European Conference on Management, Leadership and Governance, Nice, France, 2011.

[24] F. Mish, "Merriam-Webster's Collegiate Dictionary," ed: Springfield: Merriam-Webster, Inc., 2009.

[25] C. Glazner, "Understanding Enterprise Behavior Using Hybrid Simulation of Enterprise Architecture," MIT, 2009.
[26] L. Rabelo, *et al.*, "Enterprise simulation: a hybrid system approach," *International Journal of Computer Integrated Manufacturing*, vol. 18, pp. 498-508, 2005.
[27] S. Brailsford and N. Hilton, "A comparison of discrete event simulation and system dynamics for modelling health care systems," 2001.
[28] L. Rabelo, *et al.*, "New manufacturing modeling methodology: a hybrid approach to manufacturing enterprise simulation," 2003, pp. 1125-1133.
[29] J. Venkateswaran and Y. J. Son, "Hybrid system dynamic—discrete event simulation-based architecture for hierarchical production planning," *International Journal of Production Research*, vol. 43, pp. 4397-4429, 2005.
[30] K. Chahal and T. Eldabi, "Applicability of hybrid simulation to different modes of governance in UK healthcare," 2008, pp. 1469-1477.
[31] K. Chahal, "A generic framework for hybrid simulation in healthcare," 2010.
[32] D. A. Bodner and W. B. Rouse, "Understanding R&D Value Creation with Organizational Simulation," 2006.
[33] R. G. Cooper, "Stage-gate systems: a new tool for managing new products," *Business Horizons*, vol. 33, pp. 44-54, May - June 1990 1990.
[34] J. M. Garcia, *Theory and practical exercises of system dynamics*: MIT Sloan School of Management, 2006.
[35] F. H. Maier, "New product diffusion models in innovation management—a system dynamics perspective," *System Dynamics Review*, vol. 14, pp. 285-308, Winter 1998 1998.
[36] P. M. Milling, "Modeling innovation processes for decision support and management simulation," *System Dynamics Review*, vol. 12, pp. 211-234, 1996.
[37] S. A. Delre, *et al.*, "Targeting and timing promotional activities: An agent-based model for the takeoff of new products," *Journal of Business Research*, vol. 60, pp. 826-835, 2007.
[38] T. Houxing, "Simulation analysis of effect of knowledge diffusion on innovation ability of organizations," in *E -Business and E -Government (ICEE), 2011 International Conference on*, 2011, pp. 1-5.
[39] J. Li, *et al.*, "Research on Agent-Based Simulation Method for Innovation System," in *Information Management, Innovation Management and Industrial Engineering, 2008. ICIII '08. International Conference on*, 2008, pp. 431-434.
[40] M. A. Balaban, "Title," unpublished|.
[41] M. Augier and D. J. Teece, "Understanding complex organization: the role of know-how, internal structure, and human behavior in the evolution of capabilities," *Industrial and Corporate Change*, vol. 15, pp. 395-416, April 2006 2006.
[42] P. Merrill, *Innovation Generation: Creating an Innovation Process and an Innovative Culture*. Milwaukee, Wisconsin: Asq Quality Press, 2008.
[43] D. J. Teece and S. G. Winter, "The limits of neoclassical theory in management education," *The American Economic Review*, vol. 74, pp. 116-121, 1984.
[44] J. Tidd, *et al.*, *Managing innovation: integrating technological, market, and organizational change*. New York: John Wiley & Sons, 1997.