2005

Enhancing Syndromic Surveillance through Autonomic Health Grids

Hina Arora
Arizona State University, hina.arora@asu.edu

T. S. Raghu
Arizona State University, raghu.santanam@asu.edu

Ajay Vinze
Arizona State University, Ajay.Vinze@asu.edu

Peter Brittenham
IBM Software Group, peterbr@us.ibm.com

Enhancing Syndromic Surveillance through Autonomic Health Grids

Hina Arora
Department of Information Systems &
Center for Advancing Business through
Information Technology
Box 874606, W. P. Carey School of Business
Arizona State University
Tempe, AZ 85287-4606
hina.arora@asu.edu

T. S. Raghu
Department of Information Systems &
Center for Advancing Business through
Information Technology
Box 874606, W. P. Carey School of Business
Arizona State University
Tempe, AZ 85287-4606
raghu.santanam@asu.edu

Ajay Vinze
Department of Information Systems &
Center for Advancing Business through
Information Technology
Box 874606, W. P. Carey School of Business
Arizona State University
Tempe, AZ 85287-4606
ajay.vinze@asu.edu

Peter Brittenham
STSM, Autonomic Computing Architecture
IBM Software Group
peterbr@us.ibm.com

ABSTRACT

The Centers for Disease Control defines *syndromic surveillance* as, "an investigational approach where health department staff, assisted by automated data acquisition and generation of statistical alerts, monitor disease indicators in real-time or near real-time to detect outbreaks of disease earlier than would otherwise be possible with traditional public health methods" (CDC, 2004). While syndromic surveillance has traditionally been used in the context of detecting natural outbreaks, it is increasingly being used to develop systems to detect bioterrorism outbreaks. Timely response to a bioterrorism event requires accurate information exchange between clinicians and public health officials. This entails building highly complex surveillance systems that provide access to heterogeneous/distributed medical data, computational resources and collaborative services, for real-time decision making in a highly reliable and secure environment. In this paper we propose enhancing syndromic surveillance through grid and autonomic computing augmentations, and present our approach to a proof of concept modeling and simulation environment.

Keywords

syndromic surveillance, grid computing, autonomic computing.

1. INTRODUCTION

Bioterrorism attacks are low-probability, high-impact events where time is of the essence. The effectiveness of intervention depends on how quickly an epidemic can be detected, how well it can be characterized, and how rapidly a response is initiated (Buckeridge, Graham, O'Connor, Choy, Tu and Musen, 2002). In the case of an Anthrax outbreak, while pre-diagnostic data maybe available one day after exposure, diagnostic data may not be available until day 4, and death may occur as early as day 6 (Teich, Wagner, Mackenzie and Schafer, 2002). Since Anthrax is most treatable early on, timely detection and control of an outbreak is critical.

Surveillance systems need to continuously monitor different sources of pre-diagnostic data, analyze this data for the possibility of an outbreak, and should an outbreak be detected, plan a coordinated response. These tasks are non-trivial. Pre-diagnostic (and diagnostic) data is typically heterogeneous, distributed and voluminous. Analysis of the data involves computationally intensive statistical modeling and simulation of spatio-temporal syndromic data. Detection has to be low on false-positives and false-negatives. Both the analysis and detection tasks are knowledge intensive and place considerable cognitive burden on decision-makers. Of prime importance is also the maintenance of data privacy, confidentiality and security.

This research-in-progress focuses on developing system artifacts to reduce the cognitive burden on decision-makers involved in surveillance activities. An Autonomic Health Grid system is proposed as a means of enhancing syndromic surveillance, and a modeling and simulation environment that will be used to evaluate implications of performance characteristics on surveillance effectiveness. The paper is organized as follows: Section 2 discusses some of the challenges faced by syndromic surveillance. Section 3 briefly discusses the autonomic grid computing paradigm as a possible enhancement. Section 4 details the proposed Autonomic Health Grid and suggests a conceptual model. Section 5 describes our approach to simulation and modeling in this domain context.

2. SYNDROMIC SURVEILLANCE

This section provides a brief summary of techniques and challenges faced in syndromic surveillance. Figure 1 draws on earlier work (Buckeridge et al., 2002; CDC, 2004) to show the process flow in a surveillance system.

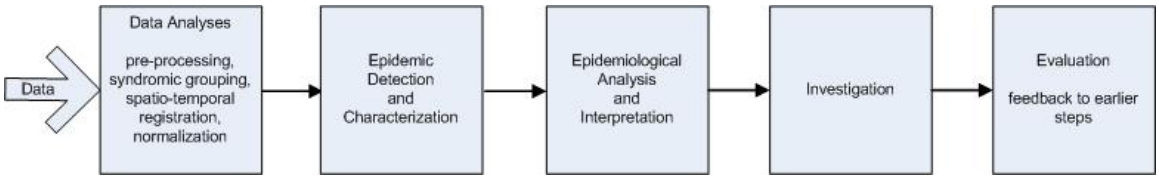

Figure 1. Process Flow

CDC has identified 5 "category-A" bioterrorism agents: inhalation anthrax, tularemia, pneumonic plague, botulism and smallpox. They have varying incubation periods, prodromes (early symptoms), and post-prodrome disease progression (Buehler, Berkelman, Hartley and Peters, 2003). Since syndromic surveillance relies on detection of outbreaks through early symptom identification, 11 syndrome categories related to category-A agents have been identified (e.g., botulism-like, hemorrhagic illness, neurological, rash, fever, and severe illness or death). These syndrome categories manifest themselves in different kinds of data depending on the time elapsed from the moment of exposure. Figure 2 draws on earlier work (Buckeridge et al., 2002; Reis and Mandl, 2003) to depict this time dependency of data sources. It is also essential to capture the geographic spread of the outbreak and the demographics of the affected population. While many of the data sources are currently non-standard, there have been numerous efforts at data standardization in recent years (such as ICD-9-CM, LOINC, UMLS, SNOMED, HL-7).

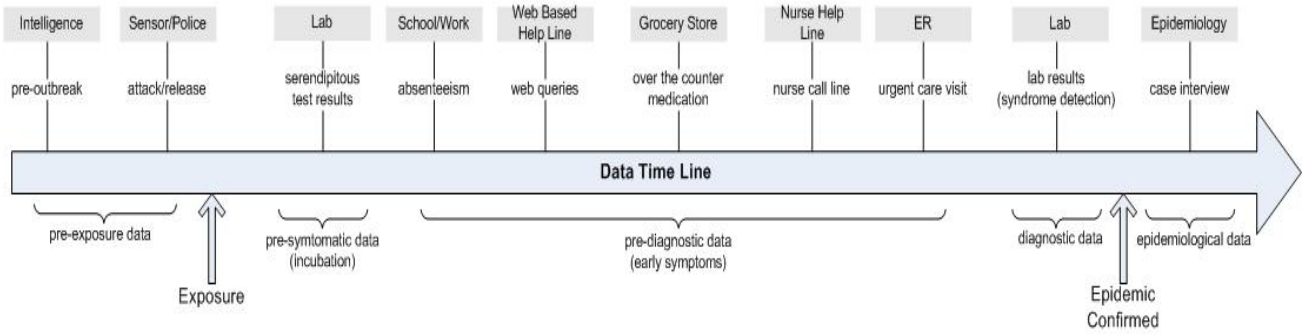

Figure 2. Data Flow

Data analyses involve pre-processing (filtering and transformation), syndromic grouping (one of 11 syndrome categories), spatio-temporal registration and normalization (accounting for expected variation in data). Tasks such as syndromic grouping are knowledge-intensive and require the application of domain-specific knowledge. Pure statistical analyses would not suffice in this case. A few systems have therefore used an ontological knowledge based approach to syndromic grouping (Buckeridge et al., 2002).

The goal of outbreak detection is to distinguish an abnormal pattern from a normal one. Syndromic surveillance approaches focus on detecting time and/or space clustering of outbreaks, i.e., a higher than expected number or rate of events than what is considered "normal" (Moore, Cooper, Tsui and Wagner, 2002). However, much like syndromic grouping, outbreak detection is an equally knowledge-intensive process that requires syndromic and epidemiological domain-specific knowledge. Thus the development of an ontological knowledge based approach would be valuable.

If an outbreak is detected, the next step involves epidemiological analysis, interpretation and investigation. Investigation involves the rules, procedures, and tools that support decision-making in response to a system signal, including adequate staffing with trained epidemiologists who can review, explore, and interpret the data in a timely manner (CDC, 2004).

Finally, the surveillance system needs to be evaluated across the following criteria: (1) Sensitivity (the likelihood of a positive result when testing a sample containing a bio-threat agent); (2) Specificity (the likelihood of a negative result when testing a sample that does not contain a bio-threat agent) (Bravata et al., 2002); (3) Timeliness (lapse of time from exposure to the initiation of a public health intervention) (4) Flexibility (adaptability to changing needs and risk thresholds); (5) Usefulness (demonstrated value relevant to public health); (6) Cost (Sosin, 2003).

3. AUTONOMIC GRID COMPUTING

Grid computing has been defined as "flexible, secure, coordinated resource sharing among dynamic collections of individuals, institutions, and resources" (Foster, Kesselman and Tuecke, 2001). We direct the reader to (Foster et al., 2001; GLOBUS) for a more complete overview of the grid architecture and available middleware. The grid architecture can be used to address the data and computational concerns described above (HealthGrid). However, by bringing together distributed, heterogeneous, dynamic systems to serve its purpose, the grid is inherently a complex environment to manage (Agarwal et al., 2003).

Autonomic computing has been suggested as a new paradigm to deal with the ever increasing complexity in today's systems. Autonomic Grids are currently being researched and at least two middleware systems have been suggested: OptimalGrid (Deen, Lehman and Kaufman, 2003) and AutoMate (Agarwal et al., 2003). We provide a brief summary of autonomic computing here, and direct the reader to (Kephart, Chess, 2003; White, Hanson, Whalley, Chess and Kephart, 2004) for a more complete overview. Autonomic systems are a collection of managed elements (such as CPU, database, business), that **m**onitor their environment, **a**nalyze the data, **p**lan a course of action and then **e**xecute it (known as the MAPE model). Elements can interact only through their specified interfaces. These systems are required to be self-managing (self-configuring, self-optimizing, self-healing and self-protecting) and capable of establishing and maintaining relationships with other autonomic elements. Everything in an autonomic system is done in accordance with high-level policies (representation in a standard external form of desired behaviors or constraints on behavior) set forth by human administrators. The MAPE model also assumes the existence of a common knowledge base that is shared among the four components of this model. In this context, an ontological knowledge base would allow for inferential capabilities and flexibility (Stojanovic et al., 2004).

4. AUTONOMIC HEALTH GRIDS

In an attempt to enhance syndrome surveillance systems with the autonomic grid architecture, a design of the Autonomic Health Grid as a self-managing system that is based on the MAPE model and uses the grid architecture is described in this section. Figure 3 is a conceptualization of our framework (for one level of hierarchy).

We consider every data source as a policy-driven autonomic managed element. Hence, hospitals, pharmacies, schools etc will each be autonomic elements that monitor-analyze-plan-execute. Each autonomic element monitors its own data. Analysis may be limited to preprocessing the data for the central surveillance system, and perhaps advising it of possible discrepancies. The central surveillance system then gathers all this data and associated advisories from various autonomic elements and analyzes the sum-total of data for an outbreak. If an outbreak is suspected, the central system might then ask for additional data from the autonomic elements. Having established an outbreak, the central system will then draw out a plan of action and execution that will then tap into the planning and execution phases of the autonomic elements' individual MAPE

cycles. Hence, the central system will end up controlling and changing the system states of the autonomic elements. Each element will also continuously learn from failures and update its knowledge base.

This approach has several advantages: (1) Autonomous - each autonomic element can manage its own policy-driven goals. (2) Scalable – each autonomic element can in turn automate other processes within it until the entire system is fully autonomic, allowing for a natural evolution of a fully autonomic system. (3) Self-configuring – the system will use a goal-driven self-assembly approach (White et al., 2004) based on a central registry. Before each element joins the system, it is given a high-level description of its goal (possibly by the central system, such as "monitor this demographic for this data"), and how to contact the registry. If the element is ready to commit to these goals, it registers itself, and starts collaborating with the central system via the MAPE cycle. The element may also choose to leave the registry subsequently. (4) Self-healing – if an autonomic element should fail, or choose to leave the registry, the surveillance system will then look for other data sources in that geographical area via the registry. It may also adapt its analysis to accommodate the failure. (5) Self-optimizing (in terms of performance, efficiency, cost) - an autonomic system will likely improve performance and efficiency by streamlining the processes within each autonomic element in order to become self-managing, and autonomic grids will bring down costs in terms of virtualizing resources and managing these complex systems. (6) Self-protecting (in terms of data privacy, confidentiality, protection) - by making each data source an anonymous, autonomic element that defines its own interfaces with other elements in the system, the data source is in complete control of its data.

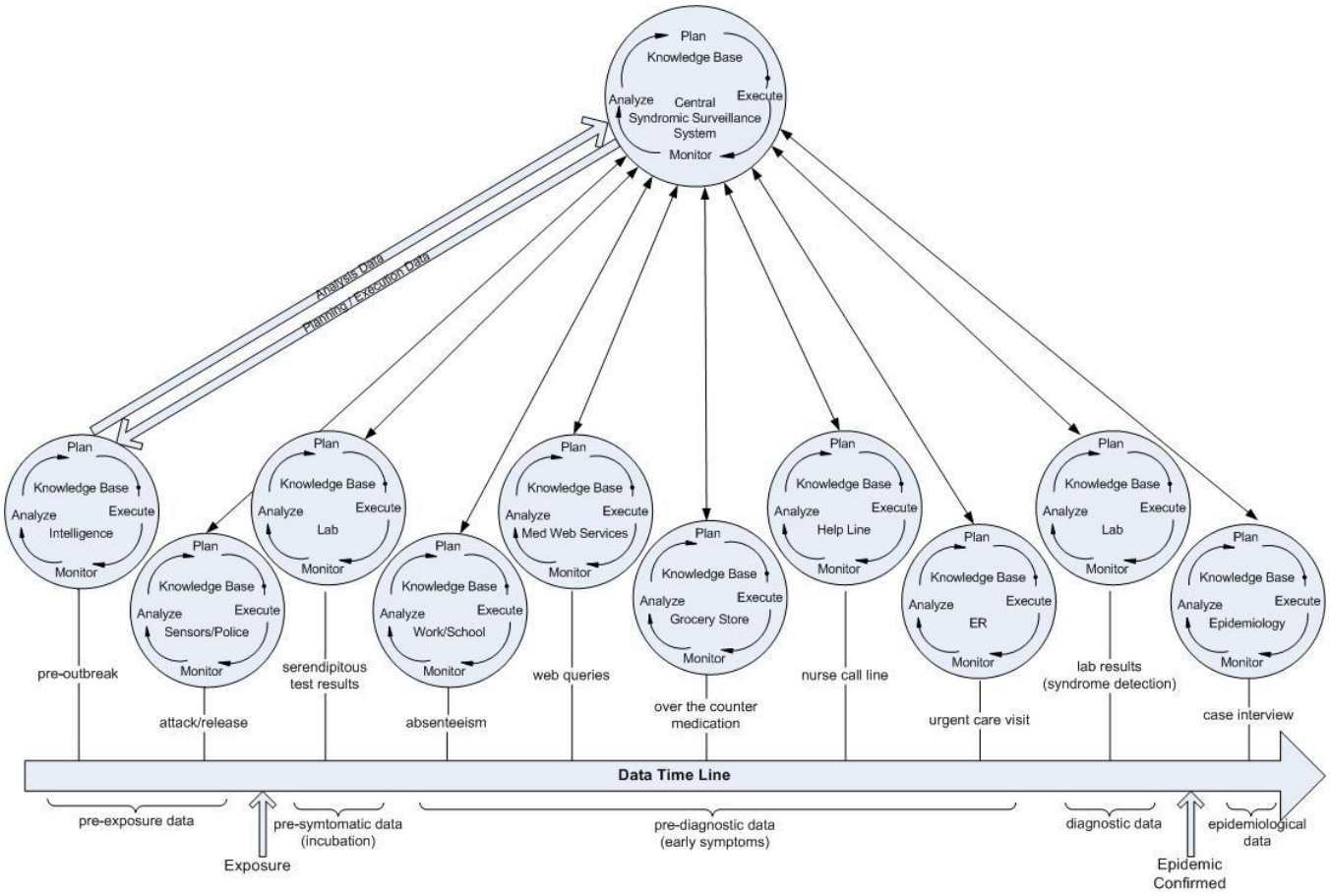

Figure 3. Autonomic Surveillance System

5. FUTURE WORK AND CONCLUSION

In extending this effort, a proof of concept simulation and modeling environment is being developed. In the planned setup, autonomic elements will be viewed as agents, and the autonomic surveillance system as a multi-agent system. This will be simulated (Carley et al., 2003; O'Toole, Mair and Inglesby, 2002; Goel, Belardo and Iwan, 2004) using the agent simulation toolkit, Repast (Collier, 2002). Repast has numerous packages, notably, the *analysis* package for data collection, the *engine* package for setting up, manipulating and driving the simulation, the *network* package to build network simulations, and the *space* package to simulate spatial relationships. Input data will be modeled based on real longitudinal data at our disposal. This data will then be perturbed (at various data points) to simulate a bio-aerosol inhalation anthrax attack based on inhalation anthrax syndrome characteristics as described in (Teich et al., 2002; Anthrax Factsheet). Through this period, the autonomic elements will continue "monitoring" their data, analyzing it, and reporting discrepancies based on a simple threshold mechanism. The central surveillance element will gather this data and run it through a spatio-temporal detection algorithm. Based on the detection of an outbreak, it will then plan and execute an action that changes the state of the autonomic elements. We will use this environment to study the following. Does having a dynamic registry mechanism afford flexibility and fault tolerance in terms of node failures? How does a policy-driven mechanism compare to micro-managing the elements? How can we build the knowledge bases of each of these autonomic elements? How can we make them adaptable to change? Can we show that syndromic and epidemiological ontologies will increase the specificity and sensitivity of our system? Does this architecture alleviate the cognitive dissonance among distributed decision makers?

This paper summarized the challenges faced by syndromic surveillance and how autonomic grids can be used to enhance these systems. A conceptual model was proposed. Finally, a simulation and modeling environment was described to test the proof of concept.

ACKNOWLEDGMENTS

This research has been partially funded by IBM T. J. Watson Research Center through an IBM Faculty Research Award.

REFERENCES

1. Agarwal, M., Bhat, V., Liu, H., Matossian, V., Putty, V., Schmidt, C., Zhang, G., Zhen, L., Parashar, M., Khargharia, B. and Hariri, S. (2003) AutoMate: Enabling Autonomic Applications on the Grid, *Proceedings of the Autonomic Computing Workshop, 5th Annual International Workshop on Active Middleware Services, 48-57.*
2. Anthrax Fact Sheet at http://www.bt.cdc.gov/agent/anthrax.
3. Bravata, D. M., McDonald, K., Owens, D. K., Buckeridge, D., Haberland, C., Rydzak, C., Schleinitz, M., Smith, W. M., Szeto, H., Wilkening, D., Musen, M., Duncan, B. W., Nouri, B., Dangiolo, M. B., Liu, H., Shofer, S., Graham, J. and Davies, S. (2002) Bioterrorism Preparedness and Response: Use of Information Technologies and Decision Support Systems, *Agency for Healthcare Research and Quality*, AHRQ Publication No. 02-E028.
4. Buckeridge, D. L., Graham, J. K., O'Connor, M. J., Choy, M. K., Tu, S. W. and Musen, M. A. (2002) Knowledge-based bioterrorism surveillance, *Proceedings AMIA Symposium*, 76-80.
5. Buehler, J. W., Berkelman, R. L., Hartley, D. M. and Peters, C. J. (2003) Syndromic surveillance and bioterrorism-related epidemics, *Emerging Infectious Diseases*, 9, 10, 1197-1204.
6. Carley, K. M., Fridsma, D., Casman, E., Altman, N., Chang, J., Kaminski, B., Nave, D. and Yahja, A. (2003) BioWar: Scalable Multi-Agent Social and Epidemiological Simulation of Bioterrorism Events, *NAACSOS conference proceedings* available at http://www.casos.ece.cmu.edu/terrorism/carley_et_al_2003_biowar.pdf.
7. CDC, (2004) Framework for evaluating public health surveillance systems for early detection of outbreaks: recommendations from the CDC Working Group, *MMWR*, 53, RR-5, 1-13.
8. Collier, N. (2002) Repast: An extensible framework for agent simulation, http://www.econ.iastate.edu/tesfatsi/RepastTutorial.Collier.pdf.
9. Deen, G., Lehman, T. and Kaufman, J. (2003) The Almaden OptimalGrid Project, *Active Middleware Services*, 14-21.
10. Foster, I., Kesselman, C. and Tuecke, S. (2001) The Anatomy of the Grid: Enabling Scalable Virtual Organizations, *International Journal of Supercomputer Applications*, 15, 3, 200-222.
11. GLOBUS at http://www.globus.org.
12. Goel, S., Belardo, S. and Iwan, L. (2004) A resilient network that can operate under duress: to support communication between government agencies during crisis situations, *Proceedings of the 37th Annual HICSS*, 123 - 133

13. HealthGrid at http://whitepaper.healthgrid.org.
14. Kephart, J. O. and Chess, D. M. (2003) The Vision of Autonomic Computing, *IEEE Computer*, 36, 1, 41-50.
15. Moore, A., Cooper, G., Tsui, R. and Wagner, M. (2002) Summary of biosurveillance-relevant technologies, *Internet Report*, http://www-cgi.cs.cmu.edu/~awm/biosurv-methods.pdf.
16. O'Toole, T., Mair, M. and Inglesby, T. V. (2002) Shining Light on 'Dark Winter, *Clinical and Infectious Diseases*, 34 , 972-983.
17. Reis, B. Y. and Mandl, K. D. (2003) Integrating syndromic surveillance data across multiple locations: effects on outbreak detection performance, *Proceedings AMIA Symposium*, 549-553.
18. REPAST at http://repast.sourceforge.net.
19. Sosin, D. M. (2003) Draft framework for evaluating syndromic surveillance systems, *Journal of Urban Health*, 80, 2, 8-13.
20. Stojanovic, L., Schneider, J., Maedche, A., Libischer, S., Studer, R., Lumpp, T., Abecker, A., Breiter, G. and Dinger, J. (2004) The Role of Ontologies in Autonomic Computing Systems, *IBM Systems Journal*, 43, 3, 598-616.
21. Syndrome Definitions at http://www.bt.cdc.gov/surveillance/syndromedef/pdf/syndromedefinitions.pdf.
22. Teich, J. M., Wagner, M. M., Mackenzie, C. F. and Schafer, K. O. (2002) The informatics response in disaster, terrorism, and war, *Journal of the American Medical Informatics Association*, 9, 2, 97-104.
23. White, S. R., Hanson, J. E., Whalley, I., Chess, D. M. and Kephart, J. O. (2004) An Architectural Approach to Autonomic Computing, *International Conference on Autonomic Computing*, 2-9