Available online at www.sciencedirect.com

ScienceDirect

Procedia Manufacturing 3 (2015) 5088 – 5094

Procedia MANUFACTURING

6th International Conference on Applied Human Factors and Ergonomics (AHFE 2015) and the Affiliated Conferences, AHFE 2015

Log Analysis of Cyber Security Training Exercises

Robert G. Abbott*,Jonathan McClain, Benjamin Anderson, Kevin Nauer, Austin Silva and Chris Forsythe

Sandia National Laboratories, Albuquerque, NM, USA

Abstract

Cyber security is a pervasive issue that impacts public and private organizations. While several published accounts describe the task demands of cyber security analysts, it is only recently that research has begun to investigate the cognitive and performance factors that distinguish novice from expert cyber security analysts. Research in this area is motivated by the need to understand how to better structure the education and training of cyber security professionals, a desire to identify selection factors that are predictive of professional success in cyber security and questions related to the development of software tools to augment human performance of cyber security tasks. However, a common hurdle faced by researchers involves gaining access to cyber security professionals for data collection activities, whether controlled experiments or semi-naturalistic observations. An often readily available and potentially valuable source of data may be found in the records generated through cyber security training exercises. These events frequently entail semi-realistic challenges that may be modeled on real-world occurrences, and occur outside normal operational settings, freeing participants from the sensitivities regarding information disclosure within operational environments. This paper describes an infrastructure tailored for the collection of human performance data within the context of cyber security training exercises. Techniques are described for mining the resulting data logs for relevant human performance variables. The results provide insights that go beyond current descriptive accounts of the cognitive processes and demands associated with cyber security job performance, providing quantitative characterizations of the activities undertaken in solving problems within this domain.

Peer-review under responsibility of AHFE Conference

*Keywords:*Cyber security; Computer security; Human performance; log analysis; Activity recognition

* Corresponding author. Tel.: +1-505-284-6765; fax: +1-505- 844-4728.
*E-mail address:*rgabbot@sandia.gov

Peer-review under responsibility of AHFE Conference
doi:10.1016/j.promfg.2015.07.523

1. Introduction

Cyber security professionals have become essential within organizations providing defense against various criminal, adversarial and malicious threats. However, the available pool of qualified personnel is insufficient to meet current demands. Furthermore, as the reliance upon information technologies continues to grow, the demand for cyber professionals will also grow. There has been substantial research, as well as monetary expenditures for commercial products, focused on software solutions to augment the performance of cyber security professionals. However, it is difficult to imagine closing the gap between the availability of cyber professionals and the demand for their services primarily through technology. The human remains a vital, inescapable element in the cyber defense of organizations[1]. The human component in cyber defense was illustrated in research assessing the utility of intrusion detection software[2].Intrusion detection software monitors network activity and generates alerts in response to suspicious patterns of activity. It was found that human operators enhanced the overall effectiveness of these products by increasing the proportion of alarms corresponding to legitimate threats that were detected, without decreasing the likelihood that attacks were detected. It has been recognized that there is a critical interplay between technology solutions and human operators[3].

A second mechanism for improving cyber defense is through the education and training of cyber professionals. Recent research has focused on identifying the knowledge and skills that underlie the progression from novice to competent to expert to elite cyber defenders. Analysis has been reported characterizing the tasks, decisions, workflow and demands associated with cyber security operations[4][5][6]. Paul and Whitley[7] described the process followed by cyber professionals in assessing and responding to alerts concerning suspicious network activity and the questions that arise at different steps in the process, with consideration of the domains of knowledge prompting specific questions. Two distinct forms of knowledge believed to be essential to cyber operations have been described[8]. First, there is knowledge of networking and security. Second, there is situated knowledge reflected in an understanding of what is normal for a given information network. It was noted that the former lends itself to transferring from one organization to another, but the latter places cyber professionals at a disadvantage when they move from one organization to another[9].

A common challenge faced by researchers in conducting research involving cyber security professionals has been access to the individuals within the operational settings that they work. This is partly attributable to the heavy workload typical of cyber operations, but also a product of sensitivities regarding capabilities and vulnerabilities. An alternative to studying actual operations can be found with cyber security training exercises[10][6]. These environments may be instrumented to provide detailed data concerning the activities of participants, use of software tools and success in accomplishing exercise objectives. This creates the opportunity for observational research. For example,[11]described the importance of team communication, structure and leadership in effective performance within the context of competitive cyber exercises. Other research has sought to characterize performance factors contributing to performance. For example, it was identified that individuals that integrated the use of specialized cyber security software tools with the use of generalized software tools (e.g., Microsoft Excel, Cygwin) performed better than those who more exclusively utilized the specialized tools[12]. Similarly, it was found that participants whose training emphasized adversary tactics and techniques surpassed the performance of participants with training that emphasized the features and functions of cyber security software tools[13].

With competitive cyber security exercises, doubt exists regarding the appropriate measures of performance[13]with this doubt a product of uncertainty regarding the appropriate metrics for assessing cyber security skills in general[1]. The current paper identifies measures that are attainable within the context of a competitive cyber security exercise. This assessment is based on the Tracer FIRE platform. Tracer FIRE was developed by the United States Department of Energy as a training environment that provides operational personnel an opportunity to exercise their skills within a semi-realistic environment. Research undertaken at Sandia National Laboratory has focused on instrumenting this environment to provide a range of measures regarding human-machine transactions and performance.

Through the instrumentation of training environments, opportunities exist for collecting real-time data concerning participant performance[14]. Such data may provide the input to automated student performance assessment. It has been demonstrated that superior training outcomes may be achieved by supplementing human instructors with software tools that provide automated assessments[15]. Benefits derive from lessening the workload on instructors

by automating detection of mundane facets of performance, allowing instructors to devote time to more complex, higher-level considerations. Furthermore, automated measures provide a degree of standardization that is sometimes difficult to achieve otherwise. The current paper lays the groundwork and provides an initial quantitative evaluation of techniques for automated assessment of student performance within cyber security training exercises.

2.0 Methods

2.1 Subjects

Subjects consisted on a total of 26 individuals who consented to data collection during two separate Tracer FIRE cyber security training exercises. There were 11 subjects from the first event which occurred during the spring of 2014 and 15 subjects from the second event that occurred in the summer of 2014.

2.2 Procedure

The Tracer FIRE exercise consisted of a multi-day event that combined classroom instruction in the use of cyber security software tools, forensic analysis techniques, and adversary tactics and techniques with a team competition exercise. At the beginning of the competition, there was an announcement concerning the study and those willing to consent to data collection underwent the informed consent process. Data collection regarding human-machine transactions occurred non-intrusively through automated data logging as subjects participated in the exercise.

The exercise presented teams a multi-level challenge. At a low level, there was a series of puzzles that allowed participants to exercise their cyber forensic analysis skills, as well as the cyber security software tools. At a higher level, there was a complex scenario partially based on real-world events that involved multiple adversaries with differing objectives operating individually and in collaboration with one another. As participants solved the individual puzzles they received points that were tallied on a scoreboard and unlocked more puzzles. Additionally, by solving individual puzzles, participants obtained clues to the overall scenario that would be helpful in solving subsequent puzzles. At the conclusion, each team presented their interpretation of the overall scenario and the ultimate outcome hinged upon how closely the team interpretations corresponded with the ground truth of the actual events.

2.3 Tracer FIRE Instrumentation

Each student was provided with a laptop computer on which essential cyber security software tools had been installed which included EnCaseEnterprise, Wireshark, PDF Dissector and Volatility. Laptops also offered the basic tools available with the Microsoft Windows and Microsoft Office products. Students were free to download additional software tools and install them on computers used for the exercises. A web-based game server provided the basis for participants to access individual challenges, submit their answers and receive feedback indicating if their answers were correct. Additionally, a news server provided periodic announcements regarding events relevant to the overall scenario (e.g., press release from Hacktivist group).

A Sandia National Laboratories software tool known as Hyperion was used to capture human-machine transactions. This included the use of software applications (specifically, setting the keyboard/mouse input focus to the application), Internet accesses, windows events, keystrokes, and mouse clicks. The data collected from Hyperion was combined and synchronized with the game server logs and logs from of the news server to provide a combined record encompassing the activities of each individual participant. For each human-machine transaction, the data included:

- Participant UserID
- Timestamp
- Interval since previous transaction (i.e., duration)
- Challenge ID, for transactions involving the game server

- Event Type, for transactions involving game server
- Submission, answer submitted for transactions involving submitting answer to game server
- Points Awarded, for transactions involving submitting answers to the game server
- Software Tool, for transaction involving software tools
- Class of Event (Windows, Game Server or News Server)
- Article ID, for transactions involving the News Server

Note that this transaction-level data does not explicitly capture higher-level descriptions of activity, such as tasks and goals. Without addressing this need, it is not possible to associate transaction-level activity with subsequent success or failure. Nor is it plausible that transaction-level analysis contains the key to more effective analyst training (e.g., "you should use Internet Explorer more often.") For these reasons, our initial analysis is focused on associating specific challenge problems with the transactions undertaken to solve them.

2.4 Recognizing Task-Level Activity in Human/Computer Transactions

Initial data analysis focused on parsing data logs from the Tracer FIRE exercise into meaningful blocks of time in which participants were focused on a specific mid-level to high-level goal. Within the context of the Tracer FIRE exercise, these high-level goals would loosely map to the individual challenges. While blocks of time could be defined based on the times in which challenges were opened and when an answer was submitted, there would be drawbacks to this approach. For example, participants may work in collaboration with other members of their team without themselves accessing the game server. Furthermore, participants might take extended breaks during which there is no activity associated with a challenge. Ultimately, the mechanisms for parsing log entries into blocks of time during which participants are focused on specific high-level objectives would be applicable to contexts extending beyond post-event analysis of Tracer FIRE exercises, and be generalizable to operational settings.

In parsing logs into blocks of activity, the first condition involved periods of inactivity. It was assumed that a period of 15 minutes or more with no activity represented a boundary between two blocks. The one exception to this rule addressed situations in which no activities are logged because the participant is reading material accessed by searching the Internet. Accordingly, when periods of inactivity of up to 30 minutes were observed and the inactivity was immediately preceded by actions consistent with the participant accessing reading material (e.g., Firefox followed by Adobe reader consistent with downloading and reading a pdf document), the period of inactivity did not serve as a partition between blocks.

As described previously, challenges were accessed via a game server. When a participant opened a challenge, the action appeared in the log as a "Set" event. Likewise, when a participant submitted an answer, the action was recorded in the log as a "Submission" and when they abandoned a challenge, the log recorded an "Abandon." Activities occurring prior to a Set event were not included in the block of activities with the Set event, with it generally assumed that a Set event (i.e., opening a challenge) represented the beginning of a sequence of related activities.

However, there were three exceptions to this rule. First, if a participant had previously worked on a challenge or another member of their team had worked on a challenge, the participant could know and work toward the solution to a challenge without actually opening the challenge. Within the logs, this situation was reflected by instances in which there was a Set event immediately followed by a Submission. In these cases, the block of activity could extend to include activities prior to the Set event. Second, in solving a challenge, the answer could be recorded in an application such as Notepad or WordPad, or require the participant to access another software application (e.g., copy and paste a URL from Firefox). Consequently, a Set and Submit event would be separated by other activities. To addresses these situations, a rule was adopted that if Set and Submit events were separated by 3 or fewer actions, the block of activity could begin prior to the Set event. Third, participants would often make an incorrect submission for a specific challenge and soon thereafter, make another submission. Sometimes, this involved making minor modifications to their answer (e.g., changing the syntax) and other times, additional work was done. In the logs, these situations appeared as Set and Submit events involving the same challenge that were either successive or separated by other activities. For this case, when there were multiple Set and Submit events involving the same

challenge, it was assumed that each Set event corresponded to a continuation of preceding work on the challenge, resulting in blocks of activity that included multiple Set events.

Submission of a correct answer was considered the end of a block of activities. Likewise, abandonment of a challenge followed by opening a different challenge was considered the end of a block of activities.

While News items provided the context framing the individual challenges, they generally did not directly address the challenges. News events were pushed to participants, with participants free to access the News server to retrieve the news items at their discretion. It is unlikely that a participant would go to the News server to look for information to use in solving a specific challenge, but instead periodically check news items to see if there was anything of interest. Events associated with accessing the News server were not included within blocks of activity.

Sessions began with a series of activities associated with configuring the laptops computers used by participants and verifying their operation with these activities recorded in the logs. These activities generally involved command line activities (i.e., cmd.exe) and use Windows Explorer, as well as Internet browsers to download software or other files. Activities at the beginning of sessions were not included in the analysis if they involved use of the command line accompanied by Windows Explorer or an Internet browser.

Activities involving Hyperion, which is the software that supports the collection of data logs, were excluded. Likewise, instances in which participants engaged in activities that clearly did not relate to solving the challenges (e.g., game play with Minesweeper), and adjacent potentially related activities were excluded from the analysis of the data logs.

3.0 Results

Through the automated parsing of the data logs, a total of 379 blocks of activity were identified. As shown in Figure 1, the vast majority of blocks were less than 25 minutes duration, with some extending much longer. It should be noted that the number of blocks of activity varied significantly across subjects. On average, there were 14.5 blocks for each subject (sd=9.0). Table 1 provides descriptive statistics for several key variables concerning the automatically derived blocks of activity. On average, a block of activity extended for approximately 17-18 min and involved approximately 45 distinct actions. Within a block of activity, on average, participants used 4 to 5 different software tools, with there being approximately 22 transitions between software tools and 19 instances where a participant returned to a tool that had been previously used within the block of activity.

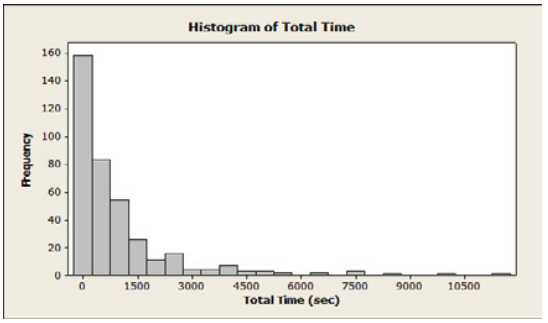

Figure 1.The distribution of blocks of activity relative to the duration of blocks.

A consideration of software applications, found that participants employed 62 distinct software tools. Figure 2 shows the nine software tools that were used by the most participants. The most frequently used software application was Explorer, however, it should be noted that the game server required the use of Explorer to access the exercise content. Yet, the utility of an Internet browser is evidenced by Firefox being the software tool used by the second most participants, with almost half the participants additionally using Chrome. This observation is further evidenced in Figure 3, which shows the total number of instances, summed across subjects that each software application was used.

Table 1. Descriptive statistics for automatically derived blocks of activity based on averaging the results across blocks for each subject and then, averaging these means across subjects

	Mean	Standard Deviation	Minimum	Median	Maximum
Duration	16 min 46 sec	9 min 54 sec	4 min 12 sec	15 min 48 sec	39 min 45 sec
Number Actions Per Block	45.0	27.5	9.3	45.0	137.0
Mean Time Per Action	17.8 sec	11.6 sec	2.5 sec	16.8 sec	56.0 sec
Number Different Software Tools	4.6	1.4	2.7	4.5	10.0
Number Transitions Between Software Tools	22.7	14.8	4.0	18.3	63.0
Number Returns to a Previous Software Tool	19.1	13.7	2.3	14.6	54.0

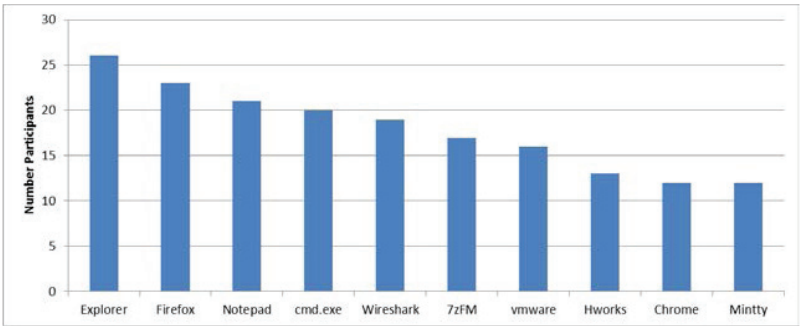

Figure 2. Software applications utilized by the most participants.

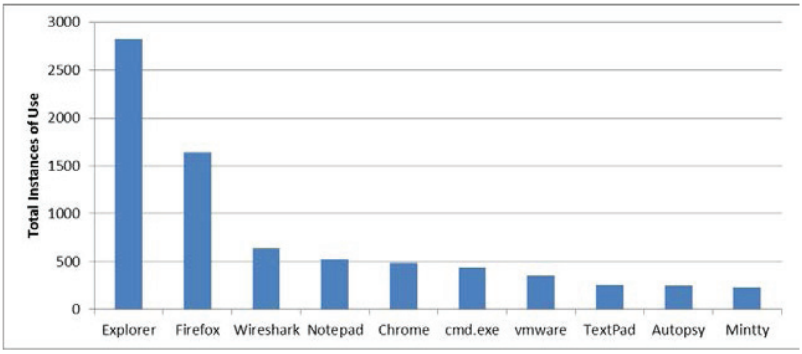

Figure 3. Total number of instances software applications were used summed across subjects.

4.0 Discussion

The emphasis of this paper has been to demonstrate the instrumentation of a cyber security exercise and the use of automated techniques to parse the resulting data logs into meaningful units that may provide the basis for further analysis and assessment of human performance. Previous studies have cast doubt upon the utility of performance measures derived based on the scores obtained within the context of competitive exercises[13]. Instead, more

meaningful insights may be gained from the work processes and the use of software tools to facilitate these work processes. For instance, it has been shown that the more effective performers tend to utilize general purpose tools in support of their use of specialized cyber security tools[12]. Automated parsing of logs is an essential step in development of techniques for automated performance assessment. However, at present, uncertainty exists concerning the appropriate metrics for assessing performance within cyber security exercises[1]. An accompanying paper (McClain et al.) addresses this topic through further analysis of the current data set to compare the behavioural characteristics of expert and novice participants.

Acknowledgements

Sandia National Laboratories is a multi-program laboratory managed and operated by Sandia Corporation, a wholly owned subsidiary of Lockheed Martin Corporation,for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-AC04-94AL85000. (SAND2014-2123 C)

References

[1]C. Forsythe, A. Silva, S. Stevens-Adams and J. Bradshaw, "Human Dimension in Cyber Operations Research and Development Priorities," in Proceedings of the Human-Computer Interaction International Conference, Las Vegas, NV, 2013.

[2]T. Sommestad and A. Hunstad, "Intrusion detection and the role of the system administrator," Information Management & Computer Security, vol. 21, no. 1, pp. 30-40, 2013.

[3] A. D'Amico, K. Whitley, D. Tesone, B. O'Brien and E. Roth, "Achieving cyber defense situational awareness: A cognitive task analysis of information assurance analysts," in Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 2005.

[4] R. F. Erbacher, D. A. Frincke, P. C. Wong, S. Moody and G. Fink, "A multi-phase network situational awareness cognitive task analysis," Information Visualization, vol. 9, no. 3, pp. 204-219, 2010.

[5] J. R. Goodall, W. G. Lutters and A. Komlodi, "I know my network: Collaboration and expertise in intrusion detection," in In Proceedings of the 2004 ACM Conference on Computer Supported Cooperative Work, Chicago, IL, 2004.

[6]J. R. Goodall, W. G. Lutters and A. Komlodi, "Developing expertise for network intrusion detection," Information Technology & People, vol. 22, no. 2, pp. 92-108, 2009.

[7] J. N. Haack, G. A. Fink, W. M. Maiden, D. McKinnon and E. W. Fulp, "Mixed-Initiative Cyber Security: Putting humans in the right loop," in The First International Workshop on Mixed-Initiative Multiagent Systems (MIMS) at AAMAS, 2009.

[8] S. Jariwala, M. Champion, P. Rajivan and N. J. Cooke, "Influence of Team Communication and Coordination on the Performance of Teams at the iCTF Competition," in Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 2012.

[9]J. T. McClain, A. Silva, G. Emmanuel, B. Anderson, K. Nauer and C. Forsythe, "Human Performance Factors in Cyber Security Forensic Analysis," in Proceedings of the Applied Human Factors and Ergonomics Conference, Las Vegas, NV, 2015.

[10]C. L. Paul and K. Whitley, "A taxonomy of cyber awareness questions for the user-centered design of cyber situation awareness," in Human Aspects of Information Security, Privacy, and Trust, 2013.

[11]T. Reed, K. Nauer and A. Silva, "Instrumenting competition-based exercises to evaluate cyber defender situation awareness," in Foundations of Augmented Cognition, 2013.

[12] A. Silva, J. McClain, T. Reed, B. Anderson, K. Nauer, R. G. Abbott and C. Forsythe, "Factors impacting performance in competitive cyber exercises," in Proceedings of the Interservice/Interagency Training, Simulation and Education Conference, Orlando, FL, 2014.

[13]T. Reed, R. G. Abbott, B. Anderson, K. Nauer and C. Forsythe, "Simulation of workflow and threat characteristics for cyber security incident response teams," in Proceedings of the 2014 International Annual Meeting of the Human Factors and Ergonomics Society, Chicago, IL, 2014.

[14]T. Sommestad and J. Hallberg, "Cyber security exercises and competitions as a platform for cyber security experiments," in Secure IT Systems, Berlin Heidelberg, Springer, 2012, pp. 47-60.

[15]S. Stevens, C. Forsythe, R. G. Abbott and C. Giesesler, "Experimental assessment of automated knowledge capture," in Proceedings of the Interservice/Interagency, Training, Simulation and Education Conference, Orlando, FL, 2009.

[16] S. Stevens-Adams, J. Basilico, R. G. Abbott, C. Gieseler and C. Forsythe, "Using after-action review based on automated performance assessment to enhance training effectiveness," in Proceedings of the Human Factors and Ergonomics Society, San Francisco, CA, 2010.

[17]S. Stevens-Adams, A. Carbajal, A. Silva, K. Nauer, B. Anderson, T. Reed and C. Forsythe, "Enhanced training for cyber situational awareness," in Foundations of Augmented Cognition, Berlin Heidelberg, Springer, 2013, pp. 90-99.