Factors Impacting Performance in Competitive Cyber Exercises

Austin Silva, Jonathan McClain, Theodore Reed, Benjamin Anderson, Kevin Nauer, Robert Abbott & Chris Forsythe

Sandia National Laboratories

Albuquerque, NM

aussilv@sandia.gov, jtmccl@sandia.gov, tmreed@sandia.gov, brander@sandia.gov, ksnauer@sandia.gov, rgabbot@sandia.gov, jcforsy@sandia.gov

ABSTRACT

Many opportunities are available for training that involves participation as either individuals or teams in competitive events. Cyber security has proven conducive to this form of training. In competitive cyber security exercises, participants are usually provided with standardized hardware and software, including various software tools for cyber forensic analysis. Generally, performance is assessed on the basis of points awarded for completing challenges presented to the participants. Ideally, through thorough instrumentation of the software environment, instructors and test coordinators would be provided with detailed data concerning the performance of individual students, as well as their unique training needs. The research described here provides an illustration of such instrumentation implemented within the context of a competition-based cyber security exercise (Tracer FIRE). The study considered factors that contributed to successful performance within the competition. Emphasis was placed on the use of software tools by participants, including tools provided by the exercise coordinators and tools acquired online by participants during the event. Resulting findings provide the basis for recommendations to competition coordinators regarding key facets of the software environment and cues that individual participants are struggling and there is need for training intervention.

ABOUT THE AUTHORS

Austin Silva is a cognitive scientist at Sandia National Laboratories in the Cognitive Modeling organization. With a background in electrical engineering and educational neuroscience, his current research focuses on the intersection of cybersecurity, novel technologies, and improving human performance.

Jonathan T. McClain is a member of the Cognitive Systems organization at Sandia National Laboratories. His research interests include instrumentation for automated knowledge capture and measuring human performance in cyber security incident responders.

Benjamin Anderson is a member of Sandia National Laboratories' Information Design Assurance Red Team (IDART) and a member of the RECOIL Lab where he conducts research and training in cyber forensics and incident response, including competition-based cyber security exercises.

Kevin Nauer has over ten years experience conducting forensic analysis and leading a team of analysts to conduct incident response operations. He leads a development effort to create a framework to support collaborative cyber security incident response operations, including training cyber analysts through competition-based exercises.

Dr. Robert G. Abbott is a Principal Member of the Technical Staff in the Cognitive Systems group of the Cyber Engineering Research Institute where he leads research in cognitive and behavior modeling.

Dr. Chris Forsythe is a Distinguished Member of the Technical Staff in the Human Factors organization at Sandia National Laboratories. He has over 25 years of experience conducting research regarding human performance and the use of technology to enhance performance in training and operational environments.

Factors Impacting Performance in Competitive Cyber Exercises

Austin Silva, Jonathan McClain, Theodore Reed, Benjamin Anderson, Kevin Nauer, Robert Abbott & Chris Forsythe

Sandia National Laboratories

Albuquerque, NM

aussilv@sandia.gov, jtmccl@sandia.gov, tmreed@sandia.gov, brander@sandia.gov, ksnauer@sandia.gov, rgabbot@sandia.gov, jcforsy@sandia.gov

INTRODUCTION

Within cyber security, a major challenge concerns the methods for providing professionals with training that is operationally-relevant and reflects current threats. Competitive exercises offer one method to address this challenge. These events have become popular with events varying in type (e.g., capture the flag, treasure hunt) and scope (Childers et al, 2010).

In developing a competition-based cyber security exercise, organizers face many considerations. The difficulty of the problems presented to students must be calibrated to their level of expertise so that experienced participants are sufficiently challenged, yet inexperienced participants are not overwhelmed (Werther et al, 2011). Exercises may allow participants to assume roles involving offense, defense or some combination (O'Connor, Sangster & Dean, 2010), while presenting extraneous events that simulate real-world operational demands (e.g., business service requests) (Corkin, 2005). There exist various approaches to score performance and provide feedback within the context of the event (Corkin, 2005). Additionally, competitive exercises may be combined with classroom training with the topics covered during lectures coordinated to varying degrees with the challenges presented during the hands-on competition.

While there has been extensive discussion of the opportunities, merits and methods for conducting competition-based cyber security exercises (e.g., Conklin, 2006; Fink, et al., 2013; Whitman & Mattord, 2008), there has been relatively little published empirical research concerning the behavioral performance of participants. Reed, Nauer & Silva (2013) presented data concerning the successful completion of challenges relative to unsuccessful completion and abandonment of challenges in discussing the application of principles from game development to assure participant engagement throughout a competition. Stevens-Adams, et al. (2013) compared the performance of teams given alternative classroom training and showed that participants who received training that emphasized adversary techniques and tactics performed better during the competition than teams who received more procedurally-based training focused on the use of cyber security software tools. Jariwala, et al. (2012) reported results regarding self-reported team communication and coordination, and leadership, and Malviya, et al. (2011) discussed the difficulties of assessing situation awareness in the context of competition-based cyber security exercises.

A distinction can be made between competition and training. Competition invovles skills measurement of individuals and perhaps, teams, whereas training concerns improvement in performance. The current research provides a consideration of factors that influence performance during competitive exercises, and specifically, addresses the use of cyber security software tools. A critical component of operational cyber security involves the use of software tools to conduct forensic analysis. A variety of products have emerged that provide a range of capabilities (e.g., analysis of packet captures, memory forensics, network analysis, etc.) Incorporation of appropriate software tools is essential to the operational relevance of competition-based cyber security exercises. Consequently, competition organizers should benefit from an understanding of how software tools are used during exercises and how patterns of tool use relate to performance in successfully completing the challenges presented to the participants.

Tracer FIRE

For the current study, data was collected as cyber security professionals participated in a competition-based cyber security exercise known as Tracer FIRE (Forensic and Incident Response Exercise). Tracer FIRE was originally developed to assure that U.S. Department of Energy cyber security professionals remained current with respect to threats, tools and techniques. Since its inception, the audience has been expanded to include other U.S. government agencies, law enforcement, industry and universities. A Tracer FIRE exercise generally begins with a series of lectures concerning pertinent topics followed by a multi-day competitive event. For the competition, participants form teams that work together to solve challenge problems. The challenges are presented using a *Jeopardy* board layout with categories appearing in the columns and the number of points awarded for successfully solving challenges in the rows. Participants are provided identical computers and a standard suit of cyber security software tools that includes ENCASE Enterprise, WireShark, IDA Pro, Volatility, Hex Workshop and PDF Dissector. Figure 1 offers a graphical depiction of the software architecture that provides the platform for conducting Tracer FIRE exercises.

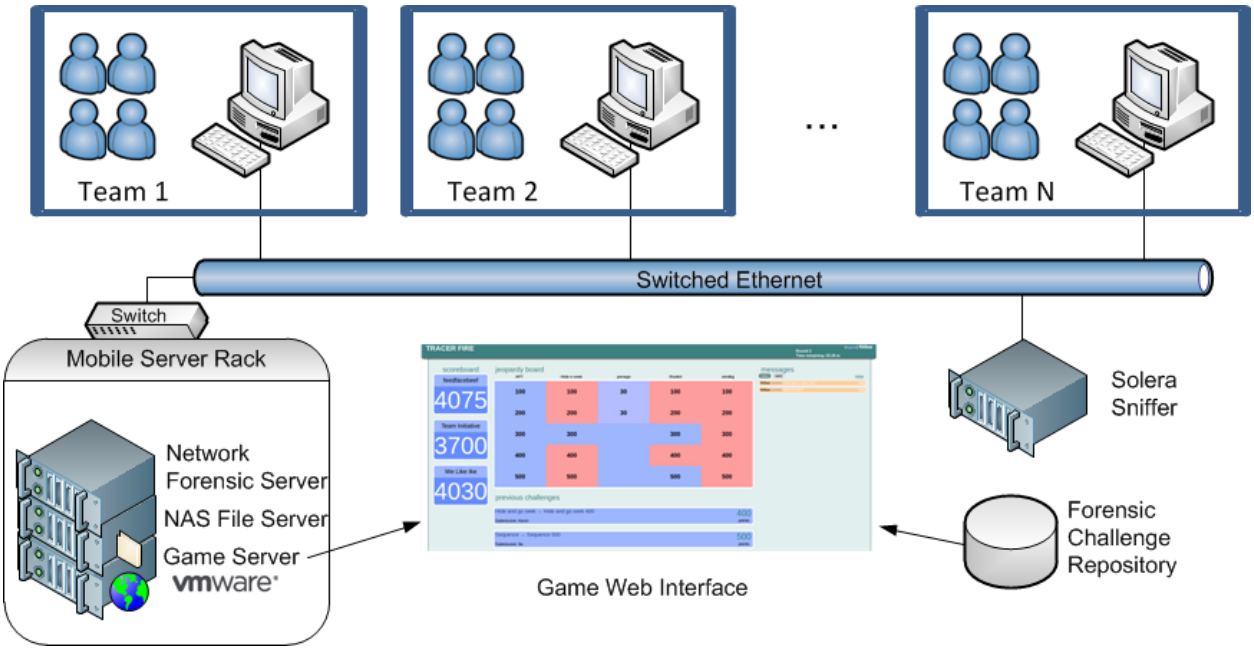

Figure 1. Graphical Depiction of Software Architecture Used to Conduct Tracer FIRE Exercises

The challenges presented during Tracer FIRE events are multi-level. At a high level, the overall challenge entails an attack against a fictitious organization with the attack involving multiple adversaries who have differing motives, and operate both independently and in coordination with one another. At a lower level, individual challenges (i.e., blocks on the Jeopardy board) present "puzzles" that must be solved with each puzzle providing a clue to the overall scenario. For example, in one challenge, participants must analyze the memory image from a flash drive to discover which archive file on the flash drive is unusual (i.e., dates have been modified). The objective for participants is to successfully complete the individual challenges to receive points while simultaneously trying to determine the overall scenario. The intent has been to emulate the operational experience of cyber forensic analysts who must analyze a collection of individual events in attempting to gain an understanding of a broader assault upon the network assets being protected. During the exercise, teams are allowed to choose how they allocate work across the members of the team, and the order in which they work and the time they commit to individual challenges. At any given time, each team member is allowed to have a single challenge open meaning that teams may simultaneously work on multiple challenges. Points are awarded for successful submission of answers to each challenge, with a small point deduction for submission of incorrect answers. Current point totals and team standings are continuously

displayed on a large scoreboard. The exercise ends with teams presenting their conclusions with regard to the overall scenario and an announcement of the winning team.

Tracer FIRE serves as both a training environment and a research platform. Research has focused on understanding factors affecting performance in cyber forensic analysis. To date, studies have considered the points received for successfully completing challenges (Reed, Nauer & Silva, 2013; Stevens-Adams, et al., 2013) and the correspondence of participant conclusions regarding the overall scenario to ground truth (Stevens-Adams, et al., 2013). For more detailed measures of moment-to-moment activities, the Tracer FIRE software environment has been instrumented to log the use of software tools, including opening and closing of windows, the content of windows and keystrokes and mouse clicks within each window. These logs provide a detailed record of participant behavior within the context of specific challenges that may be combined with data concerning correct/incorrect answer submissions, time committed to challenges and the abandonment of challenges.

METHODS

Subjects

Participants in the current study consisted of 11 cyber security professionals who consented to provide questionnaire data and allow their behavioral performance data to be collected during one of two multi-day Tracer FIRE exercises. Participation in the study was voluntary and during the two events for which data were collected, approximately half of the attendees agreed to take part in the study.

Procedure

At the beginning of the Tracer FIRE exercise, participants completed a survey in which they self-reported their professional experience in six topical areas of cyber security and their professional experience with eight software tools provided to students as components of the Tracer FIRE software environment. After completing the survey, subjects participated in the Tracer FIRE exercise. Data collection during the exercise was non-intrusive with the experience of study participants being no different from that of others who had declined to participate.

RESULTS

Participant Experience Surveys

The survey asked subjects to rate their professional experience with respect to topical areas within cyber security and their professional experience with specific cyber security software tools using a seven-point scale. The average ratings of participants are shown in Figure 2. It may be observed that while participants reported some experience in most areas, on average their professional cyber security experience was somewhat limited.

Challenge Performance

On average, participants submitted 28.8 answers (sd=20.2), with an average success rate (i.e., correct answer) of 25.0% (sd=13.2%). Participants had an average of 47.6 instances (sd=30.7) in which they closed a challenge without submitting a correct answer, which would have allowed them to open a different challenge (i.e., abandons). On average, participants committed 3.1 hours (sd=2.1 hours) working on challenges for which they submitted correct answers. Individual accuracy was correlated with the average time per submission (r= 0.639, t = 2.352, p<0.05), with individuals who devoted more time to submissions exhibiting a greater success rate (See Figure 3).

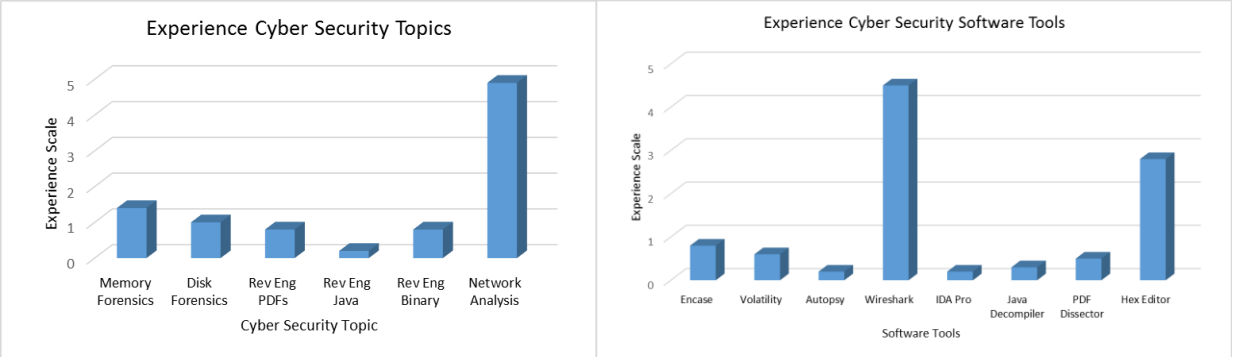

Figure 2. Self-Reported Experience with Cyber Security Topics and Software Tools Averaged Across Participants (0=No Experience; 1=1 month or less; 2=3 months or less; 4=6 months or less; 5=1 year or less; 6=3 years or less; and 7=3 years or more)

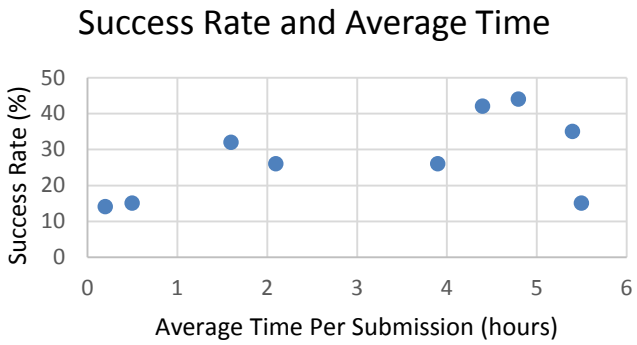

Figure 3. Participants with Greater Success Rates Devoted More Time to Submissions

While not statistically significant, it was noted that participants who had a low success rate also had a high incidence of abandons (r = 0.526, t = 1.748, NS). The individual with the highest abandon count (Total Abandons=106) had the second lowest submission accuracy of 15.1%. Conversely, the individual with the lowest abandon count (Total Abandons=20) had the second highest accuracy of 41.67% (the highest accuracy was 44.44% with Total Abandons=24). Thus, participants who frequently abandoned challenges tended to submit fewer correct answers.

Software Tool Use and Challenge Performance

The Tracer FIRE software environment provides participants a set of commonly-used cyber security software tools (e.g., Encase Enterprise, IDA Pro), which is combined with other more general purpose software tools (e.g., Notepad, Microsoft Excel, Cygwin). Also, participants are not restricted to the use of these tools and are allowed to download, install and use other software tools. Combined, participants in the current study were observed to have used 75 unique software tools. Figure 4 shows the ten most commonly used software tools and the average frequency each tool was used by the participants. It is noteworthy that Internet Explorer and Firefox accounted for the majority of tool use. This may reflect uncertainty regarding challenges and the use of a browser to search the Internet for relevant information. In fact, the frequency with which participants switched from the use of a non-browser software tool to an Internet browser was correlated with the number of incorrect submissions such that participants who frequently resorted to searching the Internet tended to make more incorrect submissions (r = 0.757,

t = 3.275, p<0.02). Likewise, while failing to reach statistical significance, there was a tendency for the use of Internet browsers to be related to the frequency of abandons (r = 0.583, t = 2.028, NS).

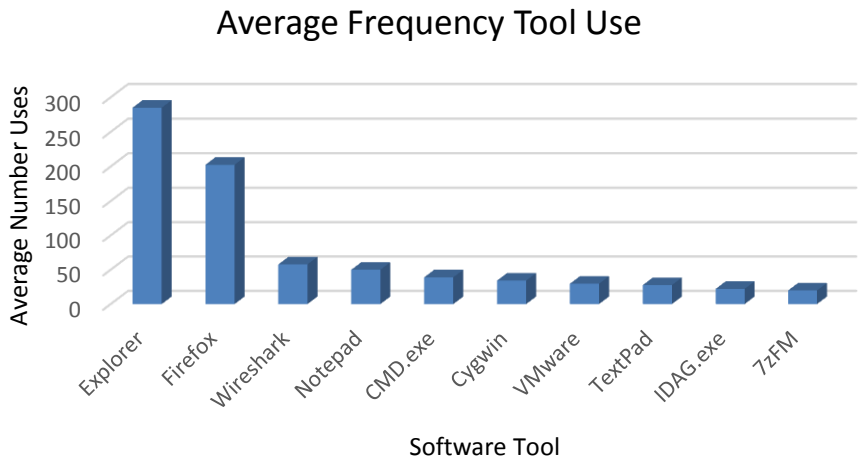

Figure 4. Ten Most Frequently Used Software Tools and the Frequency of Use Averaged Across Participants

As previously discussed, participants who devoted more time to challenges tended to make more correct submissions. This observation extended to the time devoted to specific tools and the extent to which participants switched between tools. Frequent switching between tools was associated with more incorrect submissions (r=0.877, t=5.1577, p<0.001). This may reflect a tendency for participants who are uncertain to switch between different tools in an attempt to gain insights into possible solutions.

For further analysis, each of the tools was assigned to one of nine categories based on their general function. Figure 5 shows the average frequency that each category of software tool was used by participants. Operating System (OS) tools (e.g., Windows Explorer, java, installers, help panels, etc.) and Internet browsers accounted for the overwhelming majority of tool use. Participants that more frequently used general-purpose tools (e.g., Cygwin to perform custom file manipulation, Excel to sort and filter data, etc.) submitted more successful answers (r=0.810, t=3.910, p<0.005) and earned more points (r=0.740, t=3.113, p<0.02). Interestingly, participants who spent the most time using a hex editor were more likely to submit incorrect answers (r=0.749, t=3.195, p<0.02). In this case, participants may have failed to recognize the appropriate level of analysis and mistakenly committed time to the lowest level of file analysis (i.e., hex code).

DISCUSSION

Based on findings from the current study, recommendations may be offered to organizers of competition-based cyber security exercises. The most successful participants were more likely to combine the use of specialized software tools with the use of general-purpose software tools. While much emphasis is placed on specialized tools designed to enable activities essential to cyber security forensic analysis, superior performance involves an integration of specialized and general-purpose software tools. This is consistent with previous findings that have indicated the importance of various artifacts (e.g., drawing, note taking and spreadsheet tools) in cyber forensic analysis (Singh, et al., 2011; Treude et al., 2011). Specialized tools enable detailed analysis that would be difficult, and perhaps, impossible otherwise. However, the general purpose tools support cognitive processes such as memory and pattern recognition that are integral to situation awareness and comprehending the broader picture. ***It is important that the software architecture used in competition-based exercises include a variety of general purpose tools and where appropriate, allow participants to download preferred tools***.

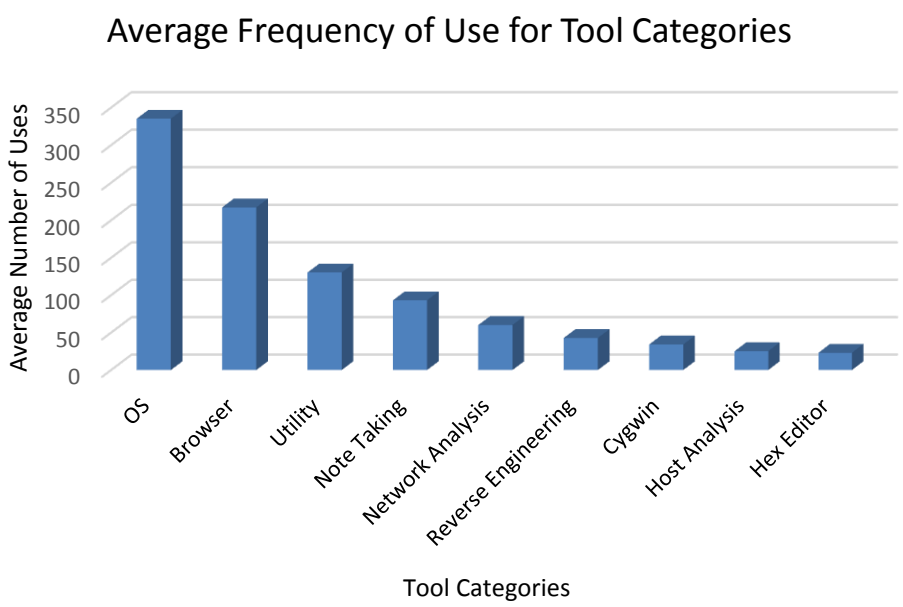

Figure 5. Frequency of Use for Each Category of Tool Averaged Across Participants

Individuals who exhibited superior performance devoted more time to individual challenges, and worked for longer blocks of time using specific software tools. In contrast, participants who committed shorter blocks of time to specific challenges and frequently shifted from one software tool to another performed less well. These data suggest clues that instructors may use to ascertain when individual participants are struggling and there may be a need to intervene and provide guidance. Specifically, ***when participants frequently switch from one challenge to another, or within a challenge, frequently shift from one software tool to another; this may be a sign that they are having difficulties***. In fact, it is not uncommon to observe students repeating activities that had previously proven unsuccessful or aimlessly trying different ideas due to a weak understanding of the problem or an inability to recognize a productive solution.

Similarly, it has been observed that while effective use of the Internet to identify information that is beneficial to solving challenges is a key to successful performance, students also frequently turn to the Internet when they are struggling. In the current study, Internet browser use was associated with less success in identifying and submitting correct answers. Consequently, by monitoring browser use, there is an opportunity for instructors to recognize when individual participants are having difficulties. In general, ***some browser use is essential, however extended browser use accompanied by frequent switching between the browser and other software tools is indicative of a student that is lost and searching for direction***.

The current study involved semi-naturalistic observation of performance as students applied knowledge and skills imparted during classroom instruction. The results obtained are observational and do not reflect the rigors typically found with laboratory experiments. For instance, the number of successful and unsuccessful submissions is a crude measure, especially given that submissions are either right or wrong and do not reflect gradations in a student's comprehension. Within a Tracer FIRE exercise, individuals work as teams and there are no restrictions on asking each other questions, sharing information, handing off challenges to more capable team members or collaboratively working on challenges. This arrangement is considered valuable in that it corresponds to real-world cyber operations where an individual's effectiveness is often a function of how well they work within a team that consists of individuals possessing various capabilities. Consequently, confounds arise when using data to draw conclusions regarding the knowledge and skills of individual participants. While the current study does not speak to the efficacy of competition-based exercises, it does provide an illustration of how properly instrumented competitive events may allow instructors to effectively gauge individual student performance within an operationally-relevant setting and intervene to improve training outcomes.

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

Childers, N., Boe, B., Cavallaro, L., Cavedon, L., Cova, M., Egele, M., & Vigna, G. (2010). Organizing large scale hacking competitions. In Detection of Intrusions and Malware, and Vulnerability Assessment (pp. 132-152). Springer Berlin Heidelberg.

Conklin, A. (2005). The use of a collegiate cyber defense competition in information security education. In Proceedings of the 2nd annual conference on Information security curriculum development (pp. 16-18). ACM.

Conklin, A. (2006). Cyber defense competitions and information security education: An active learning solution for a capstone course. In System Sciences, 2006. HICSS'06. Proceedings of the 39th Annual Hawaii International Conference on (Vol. 9, pp. 220b-220b). IEEE.

Fink, G., Best, D., Manz, D., Popovsky, V., & Endicott-Popovsky, B. (2013). Gamification for Measuring Cyber Security Situational Awareness. In Foundations of Augmented Cognition (pp. 656-665). Springer Berlin Heidelberg.

Jariwala, S., Champion, M., Rajivan, P., & Cooke, N. J. (2012, September). 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 (Vol. 56, No. 1, pp. 458-462). SAGE Publications.Klein, G., Calderwood, R., & Clinton-Cirocco, A. (2010). Rapid decision making on the fire ground: The original study plus a postscript. Journal of Cognitive Engineering and Decision Making, 4(3), 186-209.

Malviya, A., Fink, G. A., Sego, L., & Endicott-Popovsky, B. (2011, April). Situational awareness as a measure of performance in cyber security collaborative work. In Information Technology: New Generations (ITNG), 2011 Eighth International Conference on (pp. 937-942). IEEE.

O'Connor, T. J., Sangster, B., & Dean, E. (2010). Using hacking to teach computer science fundamentals. *American Society for Engineering Education, St. Lawrence Section*.

Reed, T., Abbott, R.G., Anderson, B., Nauer, K. & Forsythe, C. (in press). Simulation of workflow and threat characteristics for cyber security incident response teams. Proceedings of the Annual meeting of the Human Factors and Ergonomics Society, Chicago, IL.

Reed, T., Nauer, K., & Silva, A. (2013). Instrumenting Competition-Based Exercises to Evaluate Cyber Defender Situation Awareness. In Foundations of Augmented Cognition (pp. 80-89). Springer Berlin Heidelberg.

Singh, A., Bradel, L., Endert, A., Kincaid, R., Andrews, C., & North, C. (2011, July). Supporting the cyber analytic process using visual history on large displays. In Proceedings of the 8th International Symposium on Visualization for Cyber Security (p. 3). ACM.

Stevens-Adams, S., Carbajal, A., Silva, A., Nauer, K., Anderson, B., Reed, T., & Forsythe, C. (2013). Enhanced Training for Cyber Situational Awareness. In Foundations of Augmented Cognition (pp. 90-99). Springer Berlin Heidelberg.

Treude, C., Storey, M., & Salois, M. (2011, October). An exploratory study of software reverse engineering in a security context. In Reverse Engineering (WCRE), 2011 18th Working Conference on (pp. 184-188). IEEE.

Werther, J., Zhivich, M., Leek, T., & Zeldovich, N. (2011). Experiences in cyber security education: The MIT Lincoln Laboratory capture-the-flag exercise. Cyber Security Experimentation and Test, 8.

Whitman, M. E., & Mattord, H. J. (2008, September). The southeast collegiate cyber defense competition. In Proceedings of the 5th annual conference on Information security curriculum development (pp. 1-4). ACM.