Network Defense and Team Cognition: A Team-Based Cybersecurity Simulation

by

Aaron Bradbury

Thesis Presented in Partial Fulfillment
of the Requirements for the Degree
Master of Science

Approved November 2016 by the
Graduate Supervisory Committee:

Nancy Cooke, Chair
Rod Roscoe
Russell Branaghan

ARIZONA STATE UNIVESRITY

December 2016

ABSTRACT

This research evaluates a cyber test-bed, DEXTAR (Defense Exercises for Team Awareness Research), and examines the relationship between good and bad team performance in increasingly difficult scenarios. Twenty-one computer science graduate students (seven three-person teams), with experience in cybersecurity, participated in a team-based cyber defense exercise in the context of DEXTAR, a high fidelity cybersecurity testbed. Performance measures were analyzed in addition to team process, team behavior, and workload to examine the relationship between good and bad teams. Lessons learned are reported that will inform the next generation of DEXTAR.

DEDICATION

I dedicate this work to my mother who always encouraged my creativity, my father who taught me independence, and my sister who showed me perseverance. Thank you all for everything!

ACKNOWLEDGMENTS

I have sincere gratitude for my chair Dr. Nancy Cooke. Her role as my advisor has made my education creatively challenging and extremely rewarding. I also sincerely thank my committee member Dr. Rod Roscoe for his expert and thorough feedback, and Dr. Russell Branaghan for his expert and clear communication of human-centered design.

I sincerely thank Steven Shope, Paul Jorgenson, and Jessica Twyford for their work in the development of DEXTAR.

I sincerely thank Rachel Howes, Mike Becker, Laís Gonçalves de Lima, and Nikhil Patel for their help with data collection.

This research was sponsored by the Army Research Office under SBIR grant W911NF-14-C-0084. I thank Cliff Wang from the Army Research Office for his support.

TABLE OF CONTENTS

Page

Page

APPENDIX

LIST OF TABLES

Table Page

LIST OF FIGURES

Figure Page

INTRODUCTION

Cybersecurity relies on human analysts to recognize and resolve threats based on their knowledge of network vulnerabilities and overall healthy network performance (Champion, Rajivan, Cooke, & Jariwala, 2012). The field of team cognition strives to understand how humans work together to make better decisions in team-based environments, particularly through communication (Cooke, Gorman, Myers, & Duran, 2013). Recently, team cognition research has been conducted to understand effective processes and behaviors in cyber defense teams.

The success of a cyber-attack is determined by an ability to exploit some component of a cyber system, and the amount and complexity of these attacks grows every year. Such growth has triggered cybersecurity researchers to examine cyber vulnerabilities including sophisticated hacking techniques (Im & Baskerville, 2005) and human-user actions (Crawford, 2008). Much of this research focuses on how good the attacker is at deploying a hack or taking advantage of the human-users of the system being attacked. From a human systems engineering perspective, human analysts are a major component of the cyber system tasked with identifying and mitigating thousands of potential attacks every hour. This task contains processes of threat detection and escalation, through interactions between humans and technology, complicated by noisy network traffic data. This has led to questions about how effective humans are at performing such a task while experiencing cognitive fatigue as a result of task-work complexities (Champion, Jariwala, Ward, & Cooke, 2014; Champion et al., 2012; Knott et al., 2013). Research has shown that, compared to individual cyber analysts, cybersecurity teams are better equipped to complete complex tasks that may result in cyber-analyst cognitive fatigue (Rajivan, 2014). Organizations have demanded additional personnel in an attempt to prevent loss from cyber-threats, however more personnel does not directly translate to teamwork.

In cybersecurity, humans work with network tools and hardware components on different parts of the task. This socio-technical system requires that humans and technology work together to detect clues about an attack that are spread across the network, occur at different points in time, and displayed to different analysts. This is unquestionably a human systems issue (Gutzwiller, Fugate, Sawyer, & Hancock, 2015). DEXTAR (Defense Exercises for Team Awareness Research), a human-in-the-loop cybersecurity testbed, was launched to further such research. DEXTAR's capabilities were extended through the computing facility DETER, a virtual network cybersecurity testbed. With both testbeds working together, human-in-the-loop studies on cybersecurity can be conducted in realistic large scale environments to conduct research on team behavior, hardware performance, and software effectiveness, to name a few.

BACKGROUND

Defining Teams

Before discussing teamwork, it is important to define teams. As previously discussed, cybersecurity analysts are tasked with identifying threats to security that are spread across a network. This requires interactions with people and technology to collaborate and substantiate analysis. Teams do exhibit this behavior, but this alone does not define “team,” as teams are more than a group of individuals working together (Paris, Salas, & Cannon-Bowers, 2000). Specifically, teams are made up of members with explicit roles and responsibilities who work interdependently, adaptively, and dynamically towards a shared goal (Salas, Dickinson, Converse, & Tannenbaum, 1992). In cybersecurity, this ideally means different analysts with unique responsibilities all coordinating with each other to construct expertise on an identified problem, corroborate security issues, and prevent loss. However, research has found that cybersecurity personnel tend not to work in teams.

Teamwork in Cybersecurity

In the professional cyber-defense domain the human's role tends to be that of Information Security Analysts (Champion et al, 2014). Analysts must recognize that an attack is actually taking place in a low signal to noise environment, though the methods attackers use rapidly change, requiring expertise beyond that of a novice (Liu, Erbacher, Glodek, Etoty, & Yen, 2013). A common structure includes levels of personnel responsible for triage (identifying real for false alerts) who report to responders (ongoing threat identification and report correlation) who further escalate to forensics (large-scale emerging threat identification) (D'Amico, Whitley, Tesone, OBrien, & Roth, 2005). Such a structure encourages individual work and results in cybersecurity analysts feeling daunted by poor team organization and experiencing communication breakdowns, while analyzing a wide range of possible security alerts including attrition, web, email, removable media, lost or stolen equipment, and others – over 10,000 per hour in some professional settings. Traditionally, analysts in cybersecurity work independently and these analysts report feelings of cognitive fatigue (Champion et al., 2012). Cognitive fatigue has been shown to be related to reduced human perception and comprehension (Cain & Schuster, 2014) and reduced communication (Champion et al., 2012). When perception, comprehension, and communication are hindered, performance can be negatively affected. Teamwork, however, has been shown to relieve these effects (Rajivan, 2014), thus additional teamwork research in cybersecurity should be a priority.

Team Cognition

A better understanding of cyber-analyst team's cognitive processes may lead to better cyber-analyst team performance. Cyber-analyst work is mainly represented by cognitive tasks. If research seeks to better understand cyber-analyst teams then it must examine cognitive processes at the team level (Salas, Cooke, & Rosen, 2008), and to understand such processes requires observations about how teams learn, reason, problem solve, decide, and make judgments (Cooke et al., 2013). In cybersecurity, team cognition sheds light on such processes and allows researchers to measure those that result in teams making better assessments of cybersecurity events. In turn, this informs the design of training, tools, and team structures that improve team cognition in the cybersecurity context. As research has shown, team cognition and performance are often positively related (Cannon-Bowers & Salas, 2001; Cooke, Gorman, & Winner, 2007). There are various theories of what team cognition is, and the two most prominent theories discussed in the literature are covered in the next sections.

Shared Mental Models

Traditionally, team cognition has been measured using a Shared Mental Model (SMM) approach (Klimoski & Mohammed, 1994). By combining the individual knowledge structures from all team members, analysis can reveal how much information about a team task is shared. The theory supposes that when team members are on the same page – have the same mental model – their coordination is stronger, especially revealed in expert teams (Cannon-Bowers & Salas, 2001; Converse, 1993). However, evaluations of the shared mental model theory state that it is an oversimplification of team cognition as it is unlikely that individual team members, with specific roles and unique knowledge, have the same mental models (Cooke et al., 2007). This criticism showed how shared mental models fit in group work contexts (where all members are homogeneous, sharing the same roles and tasks) and proposed a new theory of team cognition called Interactive Team Cognition.

Interactive Team Cognition

Interactive Team Cognition (ITC) is a theory that views team cognition as a team-based activity, rather than a product of individual knowledge. (Cooke et al., 2013). Because teams are heterogeneous (its members have different roles and responsibilities) mental models of individual knowledge should differ. Thus, team cognition must be measured at the team level by observing team behaviors within a simulated or real-world context. Context is important from an ecological perspective, of which ITC theory adopts, as the environment should closely resemble real-world settings, including limited experimental interruptions (Cooke, Gorman, & Kiekel, 2008). ITC theory allows for observations of team behavior that are unobtrusive and retain context purity. Many experiments have been conducted based on the ITC perspective such as uninhabited aerial vehicles (Demir, McNeese, & Cooke, 2016), urban search and rescue (Bartlett & Cooke, 2015), non-combat military (Fouse et al., 2011), and cybersecurity (Rajivan, Janssen, & Cooke, 2013) simulations to examine team skill acquisition, human-agent teaming, and performance via team communication. These experiments show how equipped this perspective is for conceptualizing team experimentation in contextually different domains.

Developing DEXTAR

The cybersecurity literature is becoming more accessible as researchers publish their findings. Many are pointing to a need for cyber-analyst studies with skilled participants in more complex and realistic settings (Champion et al., 2012; Rajivan et al., 2013). The development of DEXTAR was an iterative process, informed by the literature and two studies of particular interest are discussed in detail.

To better understand the conditions experienced by cybersecurity analysts, researchers conducted a cognitive task analysis (CTA) on cyber-analysts and then simulated their findings in a synthetic task environment (Champion et al., 2012). The CTA revealed that real-world cybersecurity teams can often be described as groups of independently working individuals, and identified team structure, team communication, and information overload as possible contributing factors for low team performance. In the lab, researchers expanded on the synthetic task environment CyberCog to replicate these conditions. Teams of 3 participants classified 30 security alerts as either Reconnaissance, False Positive, Failed Attack, or Attack in a 30 minute session. In a second high-workload session, the same team would classify 144 security alerts in 60 minutes. Their research found that in the second session there was a 16% drop in accurate security event detection and reduced situation awareness, markedly due to further false alarms, a lack of communication, and incorrect security event classification. Their research showed a strong effect of workload on team performance and the problems associated with overloading a team's cognitive capacity. However, the design of the task was made simple enough so that participants were not required to have any previous knowledge of cybersecurity.

Further research, conducted using the synthetic task environment CyberCog, simulated cyber-analyst tasks with high workload to replicate conditions experienced by cybersecurity analysts (Rajivan et al., 2013). Network intrusion alerts that varied in complexity were displayed to participants who then classified each alert as either suspicious or benign. Participants were assigned to a “group” condition where all members had the same role, or a “team” condition where all members had specific roles. Both conditions had access to the same tools to aid participants in the alert-classification task. The groups and teams then classified 225 alerts in two 30 minute sessions. When classifying easy alerts, groups and teams both performed equally. However, when classifying difficult alerts, teams significantly outperformed groups. This research highlighted the benefits of teamwork when tasks are complex and workload is high, and called for higher-fidelity environments that led to the development of the DEXTAR testbed. Similar to Champion and colleagues' experiment, Rajivan et al. made their experiment simple enough so that participants were not required to have any prior knowledge about cybersecurity.

OVERVIEW

The cybersecurity domain is a complex system of human-computer interactions, including defensive and offensive operations, set in a wide array of contexts (Knott et al., 2013). This complex system must be realized in the lab in order validate the research and inform real-world cybersecurity practices. To simulate and conduct experiments in such an environment, human systems research needs theoretical approaches that are unobtrusive, and testbeds that are capable of mirroring real-world conditions. A controlled high-fidelity environment that matches the cognitive demands of information security analysis may be the best way to study cyber-analyst teams. This ultimately requires recruiting skilled participants with backgrounds in computer science and related professional fields.

Through an iterative development process informed by research, the DEXTAR testbed was designed to conduct human-in-the-loop cybersecurity experimentation; physically host and analyze six-member teams; collect user interaction and team performance data; capture audio, video, and computer screen recordings; and test hardware configurations and software. Through integration with DETER Lab, the DEXTAR testbed is also capable of large-scale virtual networks that are fully customizable; virtual machines and servers with Windows and Linux operating systems; physical testbed integration through virtual networks; human, scripted, and agent based cyberattacks; and temporal traffic and network performance data (Benzel, 2011).

Using the high-fidelity DEXTAR testbed, this research examines how high scoring cybersecurity teams differ from low scoring teams in terms of performance and process. Additionally, this research will evaluate the viability of the DEXTAR testbed in its first experiment. The experiment manipulates difficulty to increase the team's cognitive demand over four scenarios to elicit responses from participants regarding their perceived workload. In turn, this research assesses the DEXTAR testbed as a high-fidelity cybersecurity experimentation platform, and will make suggestions for future experimental designs. The following hypotheses were made:

H1: High-performance teams will exhibit significantly better team processes than low-performance teams as observed by an experimenter.

H2: Perceived workload will increase significantly from mission 1 through mission 3 as measured by the NASA TLX.

METHOD

Participants

Twenty-one students from Arizona State University's Computer Science graduate program and the surrounding community were recruited to participate in this study. Participants were required to have advanced knowledge about cybersecurity, and recruiting materials (Appendix A) specifically asked for experience or education in information technology/network administration, cybersecurity, or hacking. All participants were over 18-years old and had to be able to work at a computer for three hours. Due to the training materials being presented in English, and measures requiring experimenters to document verbal communication within the team, participants needed to be able to understand the spoken and written English language.

Materials and Design

Environment

Seven teams of three participants were exposed to a virtual network hosted by DETER Labs at the University of Southern California. The virtual network was a mock business network for a fake company named GEO. The network included 3 data servers, 1 webserver, 30 personal computers, and 3 IT computers with administrative rights to the network. Participants acted as the company's network administrators, each with full and equal access to make changes as needed to protect the network and its data. To access this network from the CERTT Lab at Arizona State University, participants were seated at one of three computers in the DEXTAR testbed (Figure 1) with remote desktop connections.

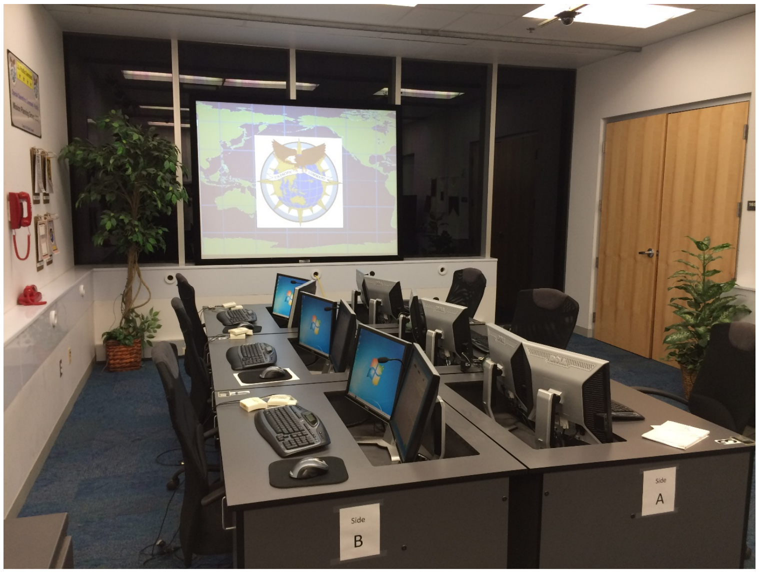
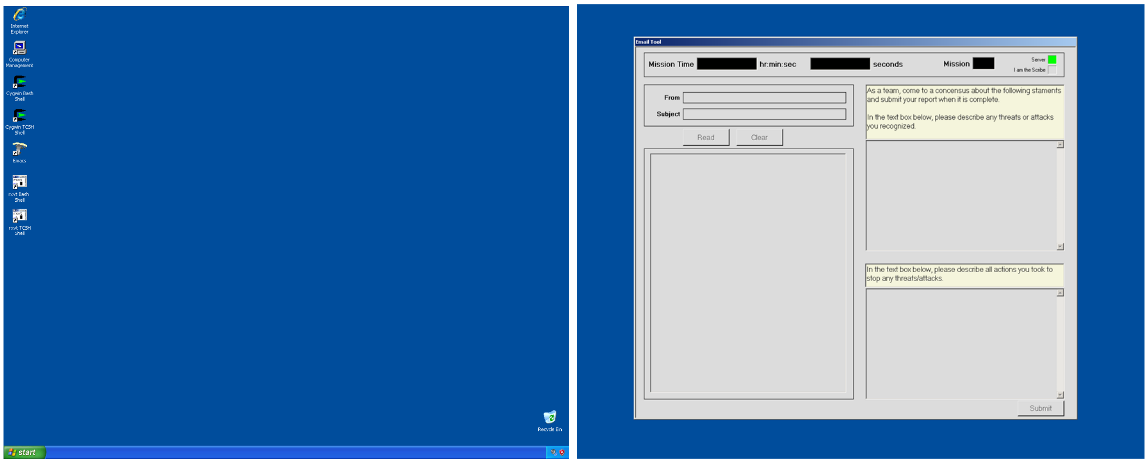


Figure 1. The DEXTAR Testbed. Participants sat at one of three stations (4, 5, or 6 from front to back) on the left.

Each participant station (labeled 4, 5, and 6) was responsible for specific regions of the virtual network. Station 4 was responsible for the webserver and compiling the team report. Station 5 was responsible for administration workstations and the admin data server. Station 6 was responsible for sales workstations and the sales data server. Each computer had two monitors. The left monitor contained the virtual IT machine's desktop, and the right contained email software and electronic measures built in the DEXTAR testbed (Figure 2).



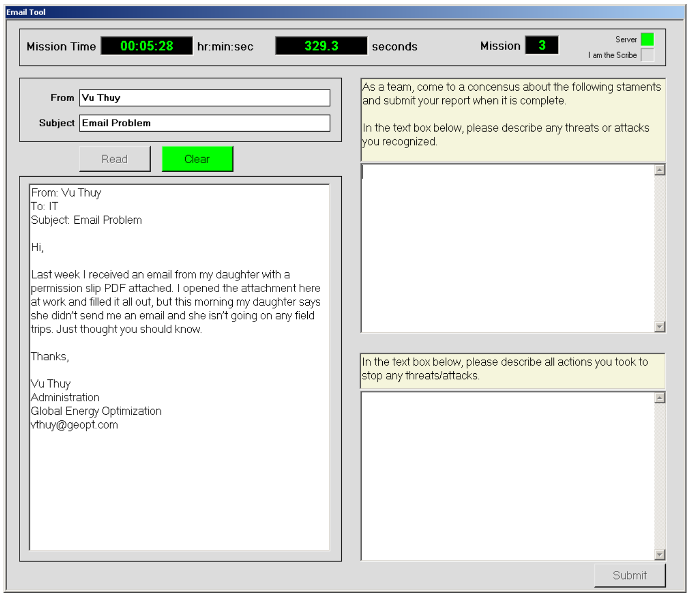
As a team, come to a concensus about the following staments and submit your report when it is complete.

In the text box below, please describe any threats or attacks you recognized.

In the text box below, please describe all actions you took to stop any threats/attacks.

Figure 2. The left and right monitor screens. The left screen displays the virtual machine's desktop, and the right displays electronic measures built in to the DEXTAR testbed.

The virtual network simulated a series of attacks separated into three increasingly difficult scenarios and a fourth scenario that was a repeat of the first. Participants would recognize the threat and then offer a resolution in their report. All attacks were scripted to maintain experimental control and executed by the experimenter at the start of each scenario. To increase difficulty, the scenarios began with a basic denial of service attack on the mock company's webserver. The second scenario consisted of an SQL injection in which malicious code was injected into a database on the webserver to gather customer financial information and return it to the agent-attacker. The third scenario emulated social engineering to gain access to a computer on the virtual network in the administration department and, from inside the virtual business network, send proprietary information to a virtual-attacker machine. Additionally, the third scenario included a situational awareness email sent to participants from an employee in administration informing them of a suspected security breach on the affected computer. The fourth and final scenario repeated the first denial of service attack.



Mission Time 00:05:28 hr:min:sec 329.3 seconds Mission 3

Server 🟩
I am the Scribe □

From: Vu Thuy
Subject: Email Problem

From: Vu Thuy
To: IT
Subject: Email Problem

Hi,

Last week I received an email from my daughter with a permission slip PDF attached. I opened the attachment here at work and filled it all out, but this morning my daughter says she didn't send me an email and she isn't going on any field trips. Just thought you should know.

Thanks,

Vu Thuy
Administration
Global Energy Optimization
vthuy@geopt.com

As a team, come to a concensus about the following staments and submit your report when it is complete.

In the text box below, please describe any threats or attacks you recognized.

In the text box below, please describe all actions you took to stop any threats/attacks.

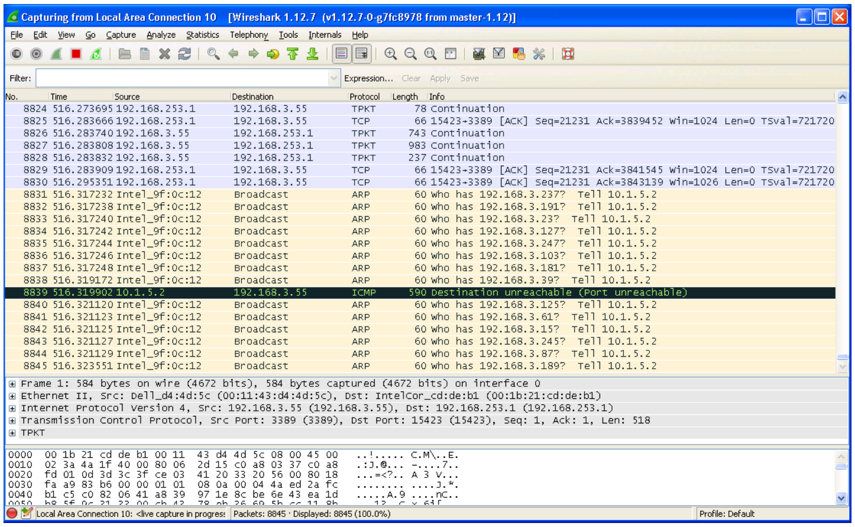
Figure 3. Reporting software. The banner displays mission time and mission number to the participant. The left side displays a canned email describing a suspicious activity on a computer. The right side is where the team fills out and submits their report.

All scenarios included other benign emails to the participants, and emails were accessible from reporting software by all participants (Figure 3). Each scenario ended when a detailed report was submitted from station 4 using the reporting software, or after 30 minutes had elapsed. The participant-submitted report included what type of threat was discovered and how it was stopped, and was considered as the team's objective scenario-completion time. To help participants keep track of time, a thirty-minute countdown timer was displayed on a projector screen.

Prior to any experimentation, a brief training session led by an experimenter introduced the participants to the environment (Appendix B), walked them through all provided tools, exposed them to their reporting software, and provided them with a topology of the virtual network (Appendix C).

Tools

Common tools were provided to the participants to aid in their work. Wireshark, a network protocol analyzer, showed participants what was happening on the network and helped participants monitor communication to network computers from outside attackers (Figure 4). The network security auditing tool nMap helped participants determine what services were operating on networked machines and what firewalls were active on them. Participants also had access to the command prompt on all virtual Windows XP machines, and the terminal on all virtual Linux machines. The command prompt and terminal are the operating system's command-line interpreter for Windows and Linux. Both allowed participants to execute features and tools built in to the operating system (Figure 5).


Capturing from Local Area Connection 10 [Wireshark 1.12.7 (v1.12.7-0-g7fc8978 from master-1.12)]
File Edit View Go Capture Analyze Statistics Telephony Tools Internals Help
Filter:
Expression... Clear Apply Save

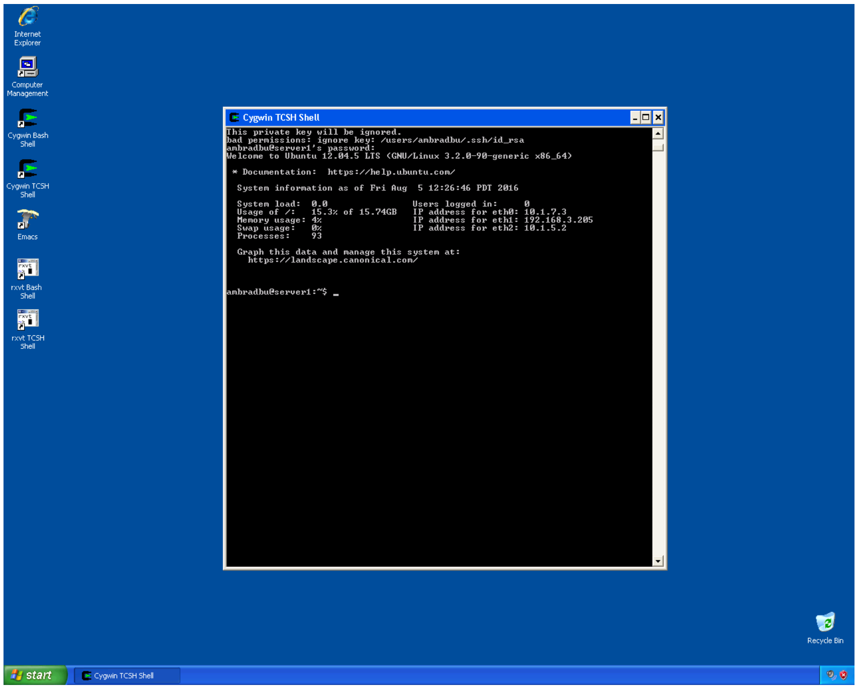
```
No.   Time       Source            Destination       Protocol Length Info
 8824 516.273695 192.168.253.1     192.168.3.55      TPKT       78 Continuation
 8825 516.283666 192.168.253.1     192.168.3.55      TCP        66 15423→3389 [ACK] Seq=21231 Ack=3839452 Win=1024 Len=0 TSval=721720
 8826 516.283740 192.168.3.55      192.168.253.1     TPKT      743 Continuation
 8827 516.283808 192.168.3.55      192.168.253.1     TPKT      983 Continuation
 8828 516.283832 192.168.3.55      192.168.253.1     TPKT      237 Continuation
 8829 516.283909 192.168.253.1     192.168.3.55      TCP        66 15423→3389 [ACK] Seq=21231 Ack=3841545 Win=1024 Len=0 TSval=721720
 8830 516.295351 192.168.253.1     192.168.3.55      TCP        66 15423→3389 [ACK] Seq=21231 Ack=3843139 Win=1026 Len=0 TSval=721720
 8831 516.317232 Intel_9f:0c:12    Broadcast         ARP        60 Who has 192.168.3.237?  Tell 10.1.5.2
 8832 516.317238 Intel_9f:0c:12    Broadcast         ARP        60 Who has 192.168.3.191?  Tell 10.1.5.2
 8833 516.317240 Intel_9f:0c:12    Broadcast         ARP        60 Who has 192.168.3.23?  Tell 10.1.5.2
 8834 516.317242 Intel_9f:0c:12    Broadcast         ARP        60 Who has 192.168.3.127?  Tell 10.1.5.2
 8835 516.317244 Intel_9f:0c:12    Broadcast         ARP        60 Who has 192.168.3.247?  Tell 10.1.5.2
 8836 516.317246 Intel_9f:0c:12    Broadcast         ARP        60 Who has 192.168.3.103?  Tell 10.1.5.2
 8837 516.317248 Intel_9f:0c:12    Broadcast         ARP        60 Who has 192.168.3.181?  Tell 10.1.5.2
 8838 516.319172 Intel_9f:0c:12    Broadcast         ARP        60 Who has 192.168.3.39?  Tell 10.1.5.2
 8839 516.319902 10.1.5.2          192.168.3.55      ICMP      590 Destination unreachable (Port unreachable)
 8840 516.321120 Intel_9f:0c:12    Broadcast         ARP        60 Who has 192.168.3.125?  Tell 10.1.5.2
 8841 516.321123 Intel_9f:0c:12    Broadcast         ARP        60 Who has 192.168.3.61?  Tell 10.1.5.2
 8842 516.321125 Intel_9f:0c:12    Broadcast         ARP        60 Who has 192.168.3.15?  Tell 10.1.5.2
 8843 516.321127 Intel_9f:0c:12    Broadcast         ARP        60 Who has 192.168.3.245?  Tell 10.1.5.2
 8844 516.321129 Intel_9f:0c:12    Broadcast         ARP        60 Who has 192.168.3.87?  Tell 10.1.5.2
 8845 516.323551 Intel_9f:0c:12    Broadcast         ARP        60 Who has 192.168.3.189?  Tell 10.1.5.2

⊞ Frame 1: 584 bytes on wire (4672 bits), 584 bytes captured (4672 bits) on interface 0
⊞ Ethernet II, Src: Dell_d4:4d:5c (00:11:43:d4:4d:5c), Dst: IntelCor_cd:de:b1 (00:1b:21:cd:de:b1)
⊞ Internet Protocol Version 4, Src: 192.168.3.55 (192.168.3.55), Dst: 192.168.253.1 (192.168.253.1)
⊞ Transmission Control Protocol, Src Port: 3389 (3389), Dst Port: 15423 (15423), Seq: 1, Ack: 1, Len: 518
⊞ TPKT

0000  00 1b 21 cd de b1 00 11  43 d4 4d 5c 08 00 45 00   ..!..... C.M\..E.
0010  02 3a 4a 1f 40 00 80 06  2d 15 c0 a8 03 37 c0 a8   .:J.@... -....7..
0020  fd 01 0d 3d 3c 3f ce 03  41 20 33 20 56 00 80 18   ...=<?.. A 3 V...
0030  fa a9 83 b6 00 00 01 01  08 0a 00 04 4a ed 2a fc   ........ ....J.*.
0040  b1 c5 c0 82 06 41 a8 39  97 1e 8c be 6e 43 ea 1d   .....A.9 ....nC..

Local Area Connection 10: <live capture in progres  Packets: 8845 · Displayed: 8845 (100.0%)  Profile: Default
```

Figure 4. Wireshark. This image shows a denial of service attack in progress that is being displayed by Wireshark.



```
This private key will be ignored.
bad permissions: ignore key: /users/ambradbu/.ssh/id_rsa
ambradbu@server1's password:
Welcome to Ubuntu 12.04.5 LTS (GNU/Linux 3.2.0-90-generic x86_64)

 * Documentation:  https://help.ubuntu.com/

  System information as of Fri Aug  5 12:26:46 PDT 2016

  System load:  0.0               Users logged in:     0
  Usage of /:   15.3% of 15.74GB   IP address for eth0: 10.1.7.3
  Memory usage: 4%                IP address for eth1: 192.168.3.205
  Swap usage:   0%                IP address for eth2: 10.1.5.2
  Processes:    93

  Graph this data and manage this system at:
    https://landscape.canonical.com/


ambradbu@server1:~$ _
```

Figure 5. The terminal. Command-line interpreters execute tools built in to the operating system. Cygwin is used to gain access to a Linux based operating system from Windows.

Measures

After each scenario, participants completed the NASA-TLX (Hart & Staveland, 1988) to measure their perceived difficulty of the task across four dimensions: Activity, Time Pressure, Success, and Emotion (Appendix D). Activity measures the amount of mental and perceptual activity required while performing the task. Time Pressure measures the pace at which participants perceived the task. Success is a measure of how successful participants thought they were at accomplishing their goals. Finally, Emotion measures the participant's enjoyment of the task. At the end of the experiment, participants completed the mini-marker personality test (Saucier, 1994) to measure their intro/extraversion (Appendix E), and a demographics survey (Appendix F). All measures were completed on the computer monitor to their right.

During each scenario, an experimenter took notes on the team's process using a process rating tool (Appendix G) that rated seven observable processes (coordination, communication, situation awareness, problem solving, decision making, plan revision, and overall process) from poor to excellent; and three observable behaviors (taking turns speaking, off topic conversations, and plan discussion) from never to always; and two additional questions regarding whether or not all team members agree on their report and whether or not team members used provided tools. Audio and video were captured as participants worked through the scenarios for verification of the process ratings. The participant's computer screens were captured for the duration of the experiment using the screen capturing software SnagIt, and a microphone mounted at each participant workstation captured verbal communication.

Procedure

Three participants located the experiment in the CERTT Lab on ASU's Polytechnic campus, were greeted, and were provided informed consent (Appendix H) that needed to be signed by all participants before experimentation could begin. Participation was reimbursed at $10 per hour, and was still paid should someone decide to drop from the experiment. Snacks and water were provided. An experimental session took approximately four hours.

Participants were first randomly seated at one of three workstations that were already remotely connected to the virtual network hosted at DETER Lab. The assigned station would determine the participant's network responsibilities (webserver and report scribe, admin machines, or sales machines). Due to a simultaneous experiment being conducted, participants were fitted with physiological monitoring devices which were then calibrated, and participants completed the N-Back task to capture inter-individual working memory (Jaeggi, Buschkuehl, Perrig, & Meier, 2010) as it related to physiology (data from these physiological measures are not addressed in this experiment). Experimenter-led training introduced participants to the environment, walked them through all tools, exposed them to the reporting software, and provided them with a network topology handout. After ensuring all participants understood their resources and environment the experiment began.

The experimenter started the attack script for the first scenario. Participants had 30 minutes to recognize the attack, resolve it, and submit a report. The scenario ended when the report was submitted by the scribe at station 4 (ideally after obtaining the team's agreement) or after 30 minutes had elapsed, whichever came first. A thirty-minute countdown was displayed on a projector screen to help participants keep track of time. At the end of the scenario, the experimenter ensured all participant tools were closed, that their left screen was on the virtual machine desktop, and that their right screen showed the Workload TLX survey which subjects then completed. This process was repeated for all remaining scenarios.

Once all scenarios were finished, participants completed a personality test and demographics survey. Afterwards, their physiological monitoring devices were removed. Finally, participants were debriefed, paid (for which they signed an acknowledgement (Appendix I) and received a receipt), and thanked for their time.

RESULTS

Missing Data

Missing data were discovered in both the mini-markers survey and demographics electronic questionnaire due to software issues that resulted in data being captured from only one team. These data were not analyzed.

Team Performance

Team scores were calculated per scenario by giving one point for discovering a threat and one point for resolving a threat. The total threat discovered and resolved points were weighted by dividing by the amount of time (in seconds) it took the team to complete the scenario, and then increased to an integer by multiplying by 10,000. Time was acquired from the scenario start time until the team submitted their final report. The weighted scores from all four scenarios were added together to obtain a total score for each team. Five bonus points were given for recognizing the situation awareness event in mission three. Five points was chosen because it matched the weight of recognizing one threat in 30 minutes, after accounting for time and multiplying to an integer.

Team performance data proved to be nonparametric. To separate the high- and low-performing teams, all scores above the median (21.91) were considered high. Four low teams (Team 2, 4, 5, and 6) scored 20.52, 21.91, 21.60, and 21.56. Three high teams (Team 1, 3, and 7) scored 29.85, 22.44, and 74.99. As expected due to the median split, a Kruskal-Wallis test confirmed that high-scoring teams significantly outperformed low-scoring teams, $X^2(1, N = 7)=4.5$, $p = 0.03$, with a mean rank of 2.5 for bad-team and 6 for good-team. It should be noted that no teams successfully recognized the social engineering attack in mission three and only Team 1 recognized the SQL-injection in mission two.

Table 1

Team Performance including Situation Awareness (SA) Points

Team	Threats Recognized	Threats Resolved	Bonus SA Points	Time (in seconds)
1	3	0	5	5,821.53
2	2	0	5	5,978.12
3	2	2	0	7,181.39
4	2	1	5	7,089.77
5	2	1	5	7,325.75
6	2	2	0	7,481.99
7	2	2	0	4,961.00

Note. The values presented aggregate all four missions. Scores (not presented in this table) were calculated per mission relative to mission time.

The results from Team 7 were particularly interesting and pointed to the ability for a team to inflate their score through time alone. Table 1 highlights the relatively similar performance, in terms of threats recognized and resolved, exhibited by all teams, as well as the score boost attributed by time. Additional analyses were not performed on these data. Table 2 displays the data from mission four, the mission that resulted in Team 7's disparate score. During debriefing, the team attributed their desire to be done with the experiment as a reason for the short mission time.

Table 2

Team Performance Results from Mission 4

Team	Threats Recognized	Threats Resolved	Time (in seconds)	Score
1	1	0	741.92	13.48
2	1	0	1,003.18	9.97
3	1	1	1,807.21	11.07
4	1	1	1,774.38	11.27
5	1	1	1,783.73	11.21
6	1	1	1,900.00	10.53
7	1	1	312.00	64.10

Note. Scores are calculated by dividing Threats Recognized and Threats Resolved by Time and multiplying both values by 10,000 to increase them to an integer. The combined integers represent the mission score.

Team Process and Team Behavior

This study was underpowered and significant findings related to performance were unlikely. These data were also nonparametric which limited the exploration of descriptive statistics. One behavior that stood out, however, was that all teams opted against using nMap. Additional analyses were not performed on these data.

Workload

Descriptive statistics were examined on all dimensions of the survey and across all four missions. On some workload dimensions, such as Mental Activity and Emotion, patterns between high and low performance opposed each other, while Time Pressure and Goal Achievement patterns were more in agreement. While examining perceived workload, a measure that has close ties to performance (Champion et al., 2012; Rajivan et al., 2013), the variance over the first three missions for good teams was less for the Mental Activity and Time Pressure dimensions than it was for bad teams who exhibited a large decline. It is important to explain that missions 2 and 3 were affected by network packet loss (normal network behavior) and threats failed to display for participants within the thirty minute mission time. From an experimenter's point-of-view this is a missed opportunity to collect meaningful data. However, from the participant's point-of-view this may have generated a very different perspective: either there is something there that is difficult to find, or there is nothing threatening happening. Future experiments may want to incorporate no-signal missions to attempt to reproduce and explain this behavior. Another interesting description came from the Emotion dimension. Overall, good teams reported more negative emotions as missions progressed, whereas bad teams reported feeling better as missions progressed. It is apparent that additional experimentation is required to attempt to replicate and understand these findings with significant differences between teams. It may then be beneficial to mix low- and high-signal with expert and novice conditions.

DISCUSSION

The DEXTAR testbed is capable of high-fidelity cybersecurity experimentation. The synthetic environment itself performed just as a real network does, demonstrated by network traffic data, packet loss, and analyst's tools that functioned as they would on real networks.

The expected finding between good and bad team performance scores comes with caveats due to the small sample size, the performance score median split, and missions 2 and 3 being affected by packet loss (discussed in detail in the following section). The lack of clear differences may suggest a need for more sensitive scoring or to increase the difficulty of the task. However, the exploration of descriptive statistics may have provided some insights for future designs.

Questions also remain about what truly makes an expert cyber-analyst. Cognitive task analyses (Champion et al., 2012) helped to start this conversation, however more work needs to be done in order to develop measures that test participant skill level. Keyword pairwise comparisons, developed by understanding expert analyst's mental models of the cybersecurity domain, may be a plausible solution.

Evaluation of DEXTAR: Lessons Learned

Lessons learned from this experiment regarding the efficacy of DEXTAR as a high-fidelity cybersecurity testbed are discussed in this section. The information is also presented in Table 3 at the end of this section for quick reference.

Mission two included an SQL injection and only one team successfully recognized this threat. No teams recognized the social engineering attack in mission three. Further investigation revealed that network traffic does not always display in real-time due to packet loss. When lost packets are picked up by cyber-analysis tools, the information is displayed at a later time. This typically took more than 30 minutes to display in our environment and ultimately meant that information regarding these threats never reached the participants within the allotted 30 minute mission time. With that in mind, packet loss is a normal characteristic of network performance. Future designs should consider this reality.

Across teams, the ability to recognize and resolve the threats in this experiment were very similar. The most decisive factor in the performance differences between high and low teams was the time it took the team to report on their missions. The results showed that high and low teams differed in their performance, but did not provide indications of the mechanisms behind the differences. Future designs may want to compare novice with expert teams, and reevaluate performance scoring to develop a more sensitive method.

Although the difficulty was designed to increase through missions two and three, normal packet loss eliminated the team’s ability to identify and report on these attacks. This raises questions about the difficulty of the missions, however only missions one and four (the denial of service attack) can be considered here. Thus, there is not a strong indication that the missions themselves actually were more difficult. This, along with the small sample size, may also be why the workload data from the Workload TLX survey did not provide expected results. Future designs may want to consider this when manipulating difficulty across missions.

Both good and bad teams came to a consensus on their mission reports and used all the tools provided to them with the exception of nMap. Interestingly, no teams used this tool. When asked about the decision to use Wireshark, participants stated there are many similarities between nMap and Wireshark and that they felt more comfortable with Wireshark. Future research should seek to understand why cyber-analysts prefer to use some tools and not others, along with heuristic evaluations of tools' user interfaces and comparisons of tools in experimental conditions.

On occasion, complications arose when accessing the virtual environment hosted by DETER Lab due to its resource sharing nature. Students, designers, and experimenters from around the world all share the resources offered by DETER Lab to host virtual networks for various purposes. Scheduled experiments discussed in this paper had to be cancelled when DETER was faced with a "tragedy of the commons" economic problem and resources were unavailable. Future designs may want to consider hosting virtual networks with cloud services that can guarantee environments are available when experimentation is ready to begin.

Evidenced by this experiment's small sample size, volunteers were difficult to find. This was likely due to the experiment seeking skilled individuals at a graduate student and professional level. Reimbursement offered to participants should better match their valuable time. Additionally, distributed-team experimental designs may make it possible to recruit from a much larger population. A virtual network that is hosted by cloud services can support globally distributed teams. Such a design ultimately allows for a host of new research questions, as well as requires effective methods that facilitate team communication (e.g. phone, messengers, email, etc.) when team members are not all in direct proximity with one another. Distributed designs will need to incorporate such communication methods, and research may want to examine the differences between communication methods as well as distributed and non-distributed cyber-teams. Finally, incorporating DEXTAR into cybersecurity education may help prepare students for cybersecurity careers after graduation. In turn, this may help continuing research, including the evaluation of DEXTAR as a training platform.

Table 3

Summary of Lessons Learned

Lesson Learned	Suggestions for Future Research
Packet loss	Mission time should be extended to allow network monitoring tools to pick up and display lost packets.
Time was the only component of the performance measure that resulted in significance	More sensitive team performance measures should be developed in the cybersecurity context. Scores may increase due to correct threat identification, correct resolution, and quality reporting. Scores may decrease due to incorrect threat identification, incorrect resolution, and reporting false alarms.
No indication of participant skill level	Keyword pairwise comparisons of expert mental models may help identify participant skill level.
No indication that missions increased in difficulty	Increase mission time to account for packet loss.
No teams used nMap	Evaluate tool interfaces and compare tools in experimental conditions.
Unable to access virtual environment due to unavailable resources	Virtual networks should be hosted with cloud services that can guarantee resources are available.
Small sample size	Reimbursement should match participants who qualify to volunteer for high-fidelity research. Distributed experimental designs also reach a larger population.

CONCLUSION

This thesis covered the examination of cybersecurity analysts within DEXTAR, a high-fidelity cybersecurity testbed. Although mission performance scores accounted for threats recognized, threats resolved, situation awareness, and total time spent on a mission, total time spent was ultimately the deciding factor in this experiment revealing that more sensitive performance measures may need to be developed. The DEXTAR testbed displayed evidence that it is capable of high-fidelity cybersecurity experimentation. Limitations were revealed that may aid in better high-fidelity cybersecurity testbed design, and further discussion made additional suggestions for future cybersecurity research.

REFERENCES

Ashkanasy, N. M. (2004). Emotion and performance. *Human Performance, 17*(2), 137-144.

Bartlett, C. E., & Cooke, N. J. (2015, September). Human-Robot Teaming in Urban Search and Rescue. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (Vol. 59, No. 1, pp. 250-254). SAGE Publications.

Benzel, T. (2011, December). The science of cyber security experimentation: the DETER project. In *Proceedings of the 27th Annual Computer Security Applications Conference* (pp. 137-148). ACM.

Cain, A. A., & Schuster, D. (2014, March). Measurement of situation awareness among diverse agents in cyber security. In *2014 IEEE International Inter-Disciplinary Conference on Cognitive Methods in Situation Awareness and Decision Support (CogSIMA)* (pp. 124-129). IEEE.

Cannon-Bowers, J. A., & Salas, E. (2001). Reflections on shared cognition. *Journal of Organizational Behavior, 22*(2), 195-202.

Champion, M., Jariwala, S., Ward, P., & Cooke, N. J. (2014, September). Using Cognitive Task Analysis to Investigate the Contribution of Informal Education to Developing Cybersecurity Expertise. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (Vol. 58, No. 1, pp. 310-314). SAGE Publications.

Champion, M., Rajivan, P., Cooke, N. J., & Jariwala, S. (2012, March). Team-based cyber defense analysis. In *Cognitive Methods in Situation Awareness and Decision Support (CogSIMA), 2012 IEEE International Multi-Disciplinary Conference on* (pp. 218-221). IEEE.

Converse, S. (1993). Shared mental models in expert team decision making. *Individual and group decision making: Current,* (1993), 221.

Cooke, N., Gorman, J. C., & Kiekel, P. A. (2008). Communication as team-level cognitive processing. In *Ashgate Publishing Ltd*.

Cooke, N. J., Gorman, J. C., Myers, C. W., & Duran, J. L. (2013). Interactive team cognition. *Cognitive science, 37*(2), 255-285.

Cooke, N. J., Gorman, J. C., Winner, J. L., & Durso, F. T. (2007). Team cognition. *Handbook of applied cognition, 2*, 239-268.

Crawford, M. (April 21, 2006). *Whoops, human error does it again. CSO Online.* Retrieved from http://www.cso.com.au/article/155941/whoops_human_error_does_it_again/

D’Amico, A., Whitley, K., Tesone, D., OBrien, B., & Roth, E. (2005). Achieving cyber defense situational awareness: A cognitive task analysis of information assurance analysts. *Human Factors and Ergonomics Society Annual Meeting Proceedings, 49*(3) 229-233.

Demir, M., McNeese, N. J., & Cooke, N. J. (2016, March). Team communication behaviors of the human-automation teaming. In *2016 IEEE International Multi-Disciplinary Conference on Cognitive Methods in Situation Awareness and Decision Support (CogSIMA)* (pp. 28-34). IEEE.

Folkman, S., & Moskowitz, J. T. (2000). Stress, positive emotion, and coping. *Current directions in psychological science*, *9*(4), 115-118.

Fouse, S., Cooke, N. J., Gorman, J. C., Murray, I., Uribe, M., & Bradbury, A. (2011, September). Effects of role and location switching on team performance in a collaborative planning environment. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (Vol. 55, No. 1, pp. 1442-1446). SAGE Publications.

Gutzwiller, R. S., Fugate, S., Sawyer, B. D., & Hancock, P. A. (2015, September). The human factors of cyber network defense. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (Vol. 59, No. 1, pp. 322-326). SAGE Publications.

Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. *Advances in psychology*, *52*, 139-183.

Im, G. P., & Baskerville, R. L. (2005). A longitudinal study of information system threat categories: the enduring problem of human error. *ACM SIGMIS Database*, *36*(4), 68-79.

Jaeggi, S. M., Buschkuehl, M., Perrig, W. J., & Meier, B. (2010). The concurrent validity of the N-back task as a working memory measure.*Memory*, *18*(4), 394-412.

Jorna, P. G. (1992). Spectral analysis of heart rate and psychological state: A review of its validity as a workload index. *Biological psychology*, *34*(2), 237-257.

Klimoski, R., & Mohammed, S. (1994). Team mental model: Construct or metaphor?. *Journal of management*, *20*(2), 403-437.

Knott, B. A., Mancuso, V. F., Bennett, K., Finomore, V., McNeese, M., McKneely, J. A., & Beecher, M. (2013, September). Human Factors in Cyber Warfare Alternative Perspectives. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (Vol. 57, No. 1, pp. 399-403). SAGE Publications.

Liu, P., Erbacher, R., Glodek, W., Etoty, R. E., & Yen, J. (2013). *Human Subject Research Protocol: Computer-Aided Human Centric Cyber Situation Awareness: Understanding Cognitive Processes of Cyber Analysts* (No. ARL-TR-6731). ARMY RESEARCH LAB ADELPHI MD COMPUTATIONAL AND INFORMATION SCIENCES DIRECTORATE.

Paris, C. R., Salas, E., & Cannon-Bowers, J. A. (2000). Teamwork in multi-person systems: a review and analysis. Ergonomics, 43(8), 1052-1075.

Rajivan, P. (2014). *Information Pooling Bias in Collaborative Cyber Forensics* (Doctoral dissertation, ARIZONA STATE UNIVERSITY).

Rajivan, P., Champion, M., Cooke, N. J., Jariwala, S., Dube, G., & Buchanan, V. (2013, July). Effects of teamwork versus group work on signal detection in cyber defense teams. In *International Conference on Augmented Cognition* (pp. 172-180). Springer Berlin Heidelberg.

Rajivan, P., Janssen, M. A., & Cooke, N. J. (2013, September). Agent-based model of a cyber security defense analyst team. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (Vol. 57, No. 1, pp. 314-318). SAGE Publications.

Richardson, R. (2008). CSI computer crime and security survey. *Computer Security Institute*, *1*, 1-30.

Salas, E., Dickinson, T. L., Converse, S. A., & Tannenbaum, S. I. (1992). Toward an understanding of team performance and training. Teams their Training and Performance, 3-29.

Saucier, G. (1994). Mini-markers: A brief version of Goldberg's unipolar Big-Five markers. *Journal of personality assessment*, *63*(3), 506-516.

APPENDIX A

RECRUITING MATERIAL

Paid Research Opportunity

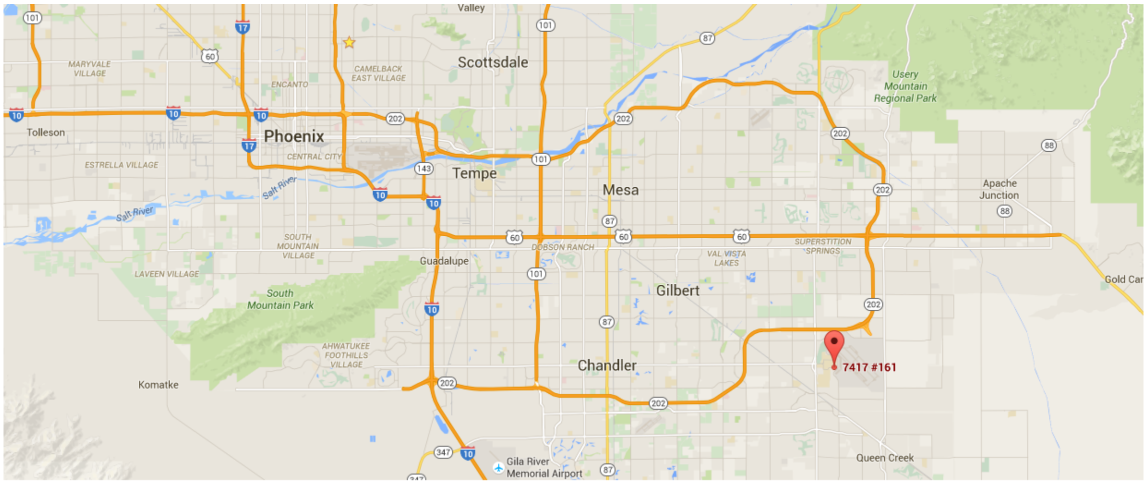
We are looking for individuals experienced in hacking, network forensics, and I/T administration to participate in a cyber security exercise.

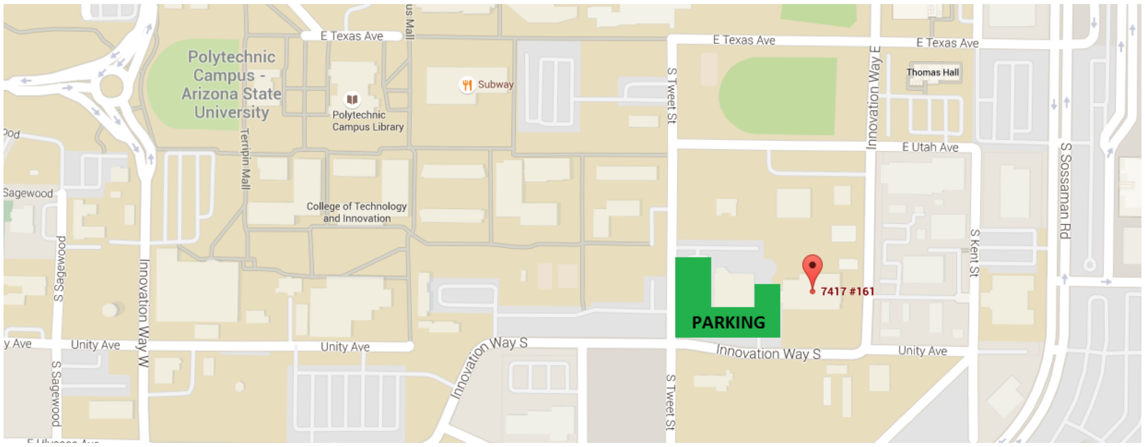
Participants will be compensated $10 per hour, and the experiment is expected to last 4 hours. Participation is limited to one session. Sessions are available Monday through Saturday starting 11/16/2015 and will be added as needed through December, 2015. Parking and temporary parking passes are provided upon arrival. Participants must have normal to corrected hearing and vision, and must be fluent in the English language.

Signup online now at cyberstudy.hfesasu.org! Try not to sign up with people you know well.

This study is being conducted at Arizona State University's Polytechnic campus by researchers in the Human Systems Engineering program's CERTT Lab. Contact us with any questions at asucyber2015@gmail.com.

Location: Interdisciplinary Science and Technology Bldg III, 7417 Innovation Way S #161, Mesa, AZ 85212





RECRUITING SCRIPT

A team of researchers at ASU Poly would like to invite your participation in a cyber security exercise at the Poly campus. Morning and afternoon sessions are available. You will be compensated $10 an hour, for approximately **4 hours** to be trained in the cyber simulator and to engage in cyber forensics exercises as part of a three-person team

You can participate in this study if you are:

- At least 18 years old
- Fluent in the English language
- Comfortable participating in team activities
- Have average or corrected to average hearing and vision

Please sign the sheet being passed around the room and provide your contact information so that we can get in touch with you and schedule a session.

APPENDIX B

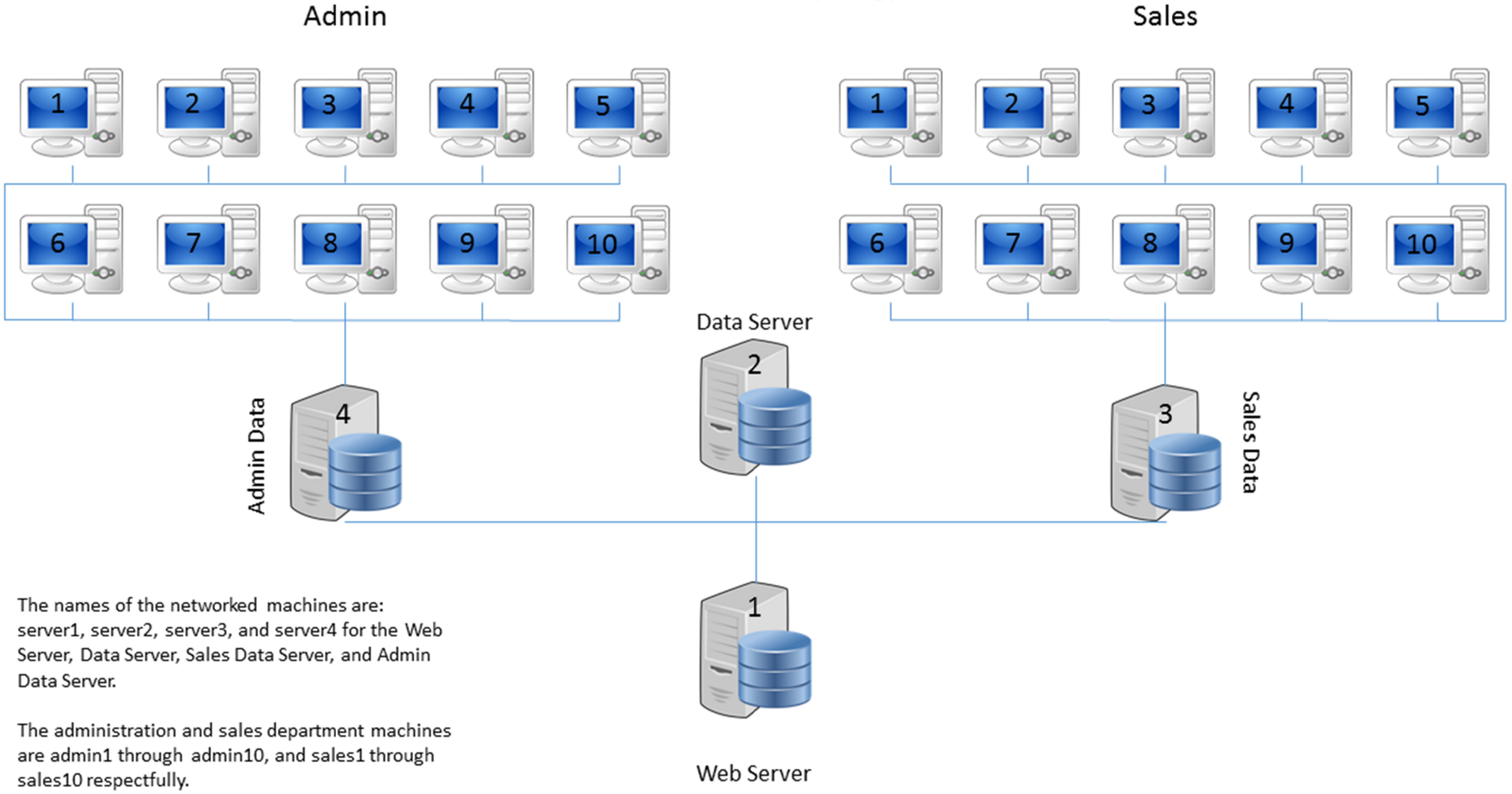
INTRODUCTION TO ENVIRONMENT

“Hello analysts. GEO, the company you work for, has decided to implement a new cyber-defense department and you three were the top candidates selected to be part of this team. The company has provided you with Wireshark and nMap as tools to monitor network traffic and packet data. All three of you have full administrator access to the network and will be able to make changes in order to defend against attack. GEO has seen an increase in cyberattacks recently and believes you three, working as a team, will provide much needed support. Because this is a new team, GEO would like four reports compiled, one every thirty minutes, regarding any suspicious activity you detect and the actions you took to resolve them. While you all have equal access to the whole network, GEO has also assigned roles to each of you. Here, [station 4] your primary responsibilities include the Web Server and compiling the team report. Here, [station 5] your primary responsibilities include workstations in Administration and the Admin Data Server. Finally, [station 6] your primary responsibilities include workstations in the company’s Sales Department and the Sales Data Server. Remember, you are working as a team and can share any information however needed. A network topology has been provided to aid as a map of GEO’s network. The names of machines on the network are server1 – spelled all lowercase with the numerical number 1 and no spaces – through server4, admin1 through 10, and sales1 through 10. This may be informative to you for navigation purposes and is located on the topology handout. The webserver is also accessible via your web-browser by typing server1 in the address bar, and has already been added as your browser’s default page.”

APPENDIX C

VIRTUAL NETWORK TOPOLOGY

GEO Network Topology



The names of the networked machines are: server1, server2, server3, and server4 for the Web Server, Data Server, Sales Data Server, and Admin Data Server.

The administration and sales department machines are admin1 through admin10, and sales1 through sales10 respectfully.

APPENDIX D

NASA TLX SURVEY

TLX Report

Instructions:

Below you will be asked some questions about the task you just completed. Please read each question and think about the information being requested. Then, respond on each scale about how you felt or what you experienced within the task. Please consider each scale independent of the previous or following scales. If you have any questions, please ask the experimenter.

1. How much mental and perceptual activity was required (e.g., thinking, deciding, calculating, remembering, looking, searching, etc.)?

The task was easy	**1 2 3 4 5 6 7 8 9 10**	The task was demanding
The task was simple	**1 2 3 4 5 6 7 8 9 10**	The task was complex
The task was forgiving	**1 2 3 4 5 6 7 8 9 10**	The task was exacting
The task was mentally effortless	**1 2 3 4 5 6 7 8 9 10**	The task was mentally difficult

2. How much time pressure did you feel due to the rate or pace at which the tasks or task elements occurred?

The task was slow	**1 2 3 4 5 6 7 8 9 10**	The task was rapid
The task was leisurely	**1 2 3 4 5 6 7 8 9 10**	The task was frantic

3. How successful do you think you were in accomplishing the goals of the task set by the experimenter (or yourself)?

Unsuccessful	**1 2 3 4 5 6 7 8 9 10**	Successful

4. Please rate the following emotional dimensions felt during the task

Insecure	**1 2 3 4 5 6 7 8 9 10**	Secure
Discouraged	**1 2 3 4 5 6 7 8 9 10**	Gratified
Irritated	**1 2 3 4 5 6 7 8 9 10**	Content
Stressed	**1 2 3 4 5 6 7 8 9 10**	Relaxed
Annoyed	**1 2 3 4 5 6 7 8 9 10**	Complacent

APPENDIX E

MINI-MARKER SET

How Accurately Can You Describe Yourself?

Please use this list of common human traits to describe yourself as accurately as possible. Describe yourself as you see yourself at the present time, not as you wish to be in the future.
Describe yourself as you are generally or typically, as compared with other persons you know of the same sex and of roughly your same age. Before each trait, please write a number indicating how accurately that trait describes you, using the following rating scale:

1	2	3	4	5	6	7	8	9
Extremely Inaccurate	Very Inaccurate	Moderately Inaccurate	Slightly Inaccurate	Neither Accurate nor Inaccurate	Slightly Accurate	Moderately Accurate	Very Accurate	Extremely Accurate

____ Bashful

____ Bold

____ Careless

____ Cold

____ Complex

____ Cooperative

____ Creative

____ Deep

____ Disorganized

____ Efficient

____ Energetic

____ Envious

____ Extraverted

____ Fretful

____ Harsh

____ Imaginative

____ Inefficient

____ Intellectual

____ Jealous

____ Kind

____ Moody

____ Organized

____ Philosophical

____ Practical

____ Quiet

____ Relaxed

____ Rude

____ Shy

____ Sloppy

____ Sympathetic

____ Systematic

____ Talkative

____ Temperamental

____ Touchy

____ Uncreative

____ Unenvious

____ Unintellectual

____ Unsympathetic

____ Warm

____ Withdrawn

APPENDIX F

DEMOGRAPHICS QUESTIONNAIRE

Date: ________ Team #: ________ Role/Station: ________

Please answer the following to the best of your ability. All answers will be kept confidential and will only be reported statistically (grouped with others' responses). Please feel free to leave a question blank if you feel uncomfortable answering it.

1. What is your age? _________

2. What is your gender? (circle):
    a. Male
    b. Female

3. Please specify your ethnicity.
    a. White
    b. Hispanic or Latino
    c. Black or African American
    d. Native American or American Indian
    e. Asian/Pacific Islander
    f. Other
    g. Prefer not to answer

4. What is your current level of education?
    a. Less than High School
    b. High School/GED
    c. Some College
    d. 2 year degree
    e. 4 year degree
    f. Master's
    g. Doctoral
    h. Professional (MD, JD, etc.)

5. If you have been or are enrolled in a post high school institution, what is your major?
____

6. Are you currently employed?
    a. Yes
    b. No

7. If yes to #5, what is your job title?
____

8. Do you have experience planning or coordinating events?
    a. Yes
    b. No

9. If yes to #10, please elaborate:
____
____
____
____

10. How often do you use a computer?
    a. Daily
    b. Every couple days
    c. Once a week
    d. Every couple weeks
    e. Less than once a month
    f. I do not use computers

11. Please rate the degree to which you agree with the following statement: I am proficient with computers.
    a. Strongly Agree
    b. Slightly Agree
    c. Neutral
    d. Slightly Disagree
    e. Strongly Disagree

12. In what way do you use computers? (Circle all that apply)
    a. I do not use computers
    b. Internet
    c. Email
    d. Word processing
    e. Spreadsheets
    f. Computer Games
    g. Other

13. Do you have any experience in cyber security?
    a. Yes
    b. No

14. If yes to #15, please describe:
____

____
____

15. Do you work with a team on a regular basis?
    a. Yes
    b. No

16. If yes to #17, in what context do you work with a team and how many individuals make up this team? (Circle all that apply)
    a. Work-related *If circled, provide number of individuals* _______
    b. Sports *If circled, provide number of individuals* _______
    c. Other Recreation *If circled, provide number of individuals* _______
    d. Other *If circled, provide number of individuals* _______

    Please specify other: ______________________

Please rate the degree to which you agree with the following statements. Consider your *team* to be made up of the other people on your side of the divider.

17. I feel like my individual contribution to the team was important.
    a. Strongly Agree
    b. Slightly Agree
    c. Neutral
    d. Slightly Disagree
    e. Strongly Disagree

18. Regardless of outcome, I feel like we performed well overall.
    a. Strongly Agree
    b. Slightly Agree
    c. Neutral
    d. Slightly Disagree
    e. Strongly Disagree

19. The other people on my team were good members.
    a. Strongly Agree
    b. Slightly Agree
    c. Neutral
    d. Slightly Disagree
    e. Strongly Disagree

20. If I were asked to participate in another project like this one, I would like to be with the same team member.
    a. Strongly Agree
    b. Slightly Agree
    c. Neutral
    d. Slightly Disagree
    e. Strongly Disagree

21. This task was complicated.
    a. Strongly Agree
    b. Slightly Agree
    c. Neutral
    d. Slightly Disagree
    e. Strongly Disagree

22. The strategy we employed were the most effective way to complete the task.
    a. Strongly Agree
    b. Slightly Agree
    c. Neutral
    d. Slightly Disagree
    e. Strongly Disagree

23. This task was boring.
    a. Strongly Agree
    b. Slightly Agree
    c. Neutral
    d. Slightly Disagree
    e. Strongly Disagree

24. The way we made decisions was the best way to make decisions.
    a. Strongly Agree
    b. Slightly Agree
    c. Neutral
    d. Slightly Disagree
    e. Strongly Disagree

25. Our group could have done better if we had worked more as a team.

a. Strongly Agree
b. Slightly Agree
c. Neutral
d. Slightly Disagree
e. Strongly Disagree

26. I did **not** like the way our team made decisions.
    a. Strongly Agree
    b. Slightly Agree
    c. Neutral
    d. Slightly Disagree
    e. Strongly Disagree

27. I was motivated to help the group complete missions.
    a. Strongly Agree
    b. Slightly Agree
    c. Neutral
    d. Slightly Disagree
    e. Strongly Disagree

28. This task was easy.
    a. Strongly Agree
    b. Slightly Agree
    c. Neutral
    d. Slightly Disagree
    e. Strongly Disagree

29. I liked to talk with other members of the group.
    a. Strongly Agree
    b. Slightly Agree
    c. Neutral
    d. Slightly Disagree
    e. Strongly Disagree

30. The user-computer interface was easy to use.
    a. Strongly Agree
    b. Slightly Agree
    c. Neutral
    d. Slightly Disagree
    e. Strongly Disagree

31. I enjoyed participating in this study.
    a. Strongly Agree
    b. Slightly Agree
    c. Neutral
    d. Slightly Disagree
    e. Strongly Disagree

APPENDIX G

TEAM PROCESS CHECKLIST

Date: Team: # Mission: # Experimenter

Rate all dimensions from 0 (poor/never) to 10 (excellent/always).

1. Coordination: The timely and adaptive push and pull of info among team members (e.g. getting the right information to the team at the right time).

   0 1 2 3 4 5 6 7 8 9 10
   Comments:

2. Communication: The verbal or non-verbal passing of relevant and necessary information and the recipient's acknowledgement of understanding that information.

   0 1 2 3 4 5 6 7 8 9 10
   Comments:

3. Team Situation Awareness: Team members perceive current environmental conditions, communicating perceptions of those conditions appropriately, and coordinating responses as necessary.

   0 1 2 3 4 5 6 7 8 9 10
   Comments:

4. Team Problem Solving: Team identifies problems and generates possible solutions.

   0 1 2 3 4 5 6 7 8 9 10
   Comments:

5. Team Decision Making: Occurs when all team members agree on one or a set of solutions over alternatives.

   0 1 2 3 4 5 6 7 8 9 10
   Comments:

6. Outcome and Revision: Teams analyze, test, and validate the agreed upon team solution against goal requirements.

   0 1 2 3 4 5 6 7 8 9 10
   Comments:

7. Overall Team Process: How good was the team process overall (taking all 6 process measures into account)?

   0 1 2 3 4 5 6 7 8 9 10

   Comments:

8. Team members take turns speaking: Not talking over each other, taking turns speaking, and/or waiting for a response and then answering.

   0 1 2 3 4 5 6 7 8 9 10

   Comments:

9. Off topic conversations: Anything that is unrelated to the mission.

   0 1 2 3 4 5 6 7 8 9 10

   Comments:

10. Team plan discussion: Any planning, correcting/editing, proposing an idea/plan of action?

    0 1 2 3 4 5 6 7 8 9 10

    Comments:

11. Team verified and agreed on report: Acknowledgement of ALL team members and acceptance by ALL team members.

    Yes No

    Comments:

12. Team members use provided tools: Any tools (e.g. software) given to the participants to aid in the mission.

    Participant 1 Participant 2 Participant 3 None

    Comments:

13. General Comments:

APPENDIX H

CONSENT FORM

CONSENT FORM
CYBER WARFARE SIMULATION ENVIRONMENT AND ADAPTER PHASE II

INTRODUCTION

The purposes of this form is to provide you (as a prospective research study participant) information that may affect your decision as to whether or not to participate in this research and to record the consent of those who agree to be involved in the study.

RESEARCHERS

Nancy J. Cooke, Professor, Aaron Bradbury, Graduate Researcher, and Mike Becker, Graduate Researcher have invited your participation in a research study.

STUDY PURPOSE

The purpose of the research is to better understand the team process and performance of computer security defense analysts and how various measures of individual and team state relate to team effectiveness. The testbed you will work in is used to test hypothesized needs, cognitive processes, and displays for enhanced situation awareness for cyber security analysis.

DESCRIPTION OF RESEARCH STUDY

If you decide to participate, you will join a study funded by the Army Research Organization through Sandia Research Corporation and the Air Force Research Laboratory through Charles River Analytics. You will participate in a simulated computer security task as part of a 3-person team. After signing the informed consent, you will be presented with materials that instruct you on the simulated task. After reaching a certain performance level through training, you will then interact (communicate) with other trained participants as a team to make decisions related to the computer security task. Measures of your team's task performance will be collected at the end of the task. Experimenters will also observe and evaluate the team's process behaviors such as communication, coordination, and leadership. Communications and team process behaviors will be measured through observations and audio and video recording. A personality scale will also be administered at the end of the study. In addition you will be asked to wear a sensor on your head and on either your arm or your calf. The head-worn sensor measures motion and oxygenated blood flow to the prefrontal cortex (the part of the brain behind the forehead). It is worn as a headband. The arm/calf-worn sensor measures galvanic skin response (the production of sweat on the skin), temperature, motion, cardiac information, and oxygenated blood flow. It is worn similarly to an mp3 player during exercise. We will use an alcohol prep pad to clean the surface of your skin to ensure that no lotion or makeup will interfere with data acquisition; however neither sensor requires adhesion to the skin surface. Neither sensor should interfere with experimental activities. In addition mouse activity will be recorded.

We anticipate that this study will require roughly 4 hours. Your participation is completely voluntary and you may cease participation at any time. There will be approximately 30-60 other teams in this study. Participants must be between 18 and 50 years of age.

RISKS

Pulse oximetry sensors such as those included in this study have been deemed safe by the FDA. People have sometimes reported the following side effects, however these side effects are relatively uncommon, and usually resolve quickly afterwards when they occur:

- Itchiness
- Tingling
- Skin irritation and redness under the electrode
- Transient (brief) headaches

We will periodically ask you about these side effects during the study. If you notice any side effects during your session or after you leave the laboratory, please contact the study doctor or the research staff to let them know. Your privacy will be protected in this study (see section on confidentiality).

BENEFITS

You will personally benefit by increasing your experience with research methodology in human systems engineering and others will benefit more generally from the findings pertinent to and the tools developed for enhancing situation awareness in cyber security.

CONFIDENTIALITY

All information obtained in this study is strictly confidential. The results of this research study may be used in reports, presentations, and publications, but the researchers will not identify you. The Department of Defense is supporting this research. They and our research partners at Charles River Analytics and will have access to research data. Federal government offices that oversee human research protection will have access to research records and identifiable subject records to ensure compliance. In order to maintain confidentiality of your records, Dr. Nancy J. Cooke will follow these procedures: (1) each participant will be assigned a number; (2) the researchers will record any data collected during the study by number, not by name; (3) any original data files will be stored in a secured location accessed only by authorized researchers; (4) consent forms will not link names to ID numbers; (5) video recordings will be kept apart from other study data; (6) only processed data from video recordings will be used for further analysis (no images). Consent forms will also be secured in a separate file.

WITHDRAWAL PRIVILEGE

Participation in this study is completely voluntary. It is ok for you to say no. Even if you say yes now, you are free to say no later, and withdraw from the study at any time. Your participation is voluntary and that nonparticipation or withdrawal is acceptable.

COSTS AND PAYMENTS

The researchers want your decision about participating in the study to be absolutely voluntary, yet they recognize that your participation may pose inconvenience. You will be compensated $10 per hour for your participation. Additionally, please ask us any research questions at the end as we hope this is a positive learning experience for you.

DISCOSURE

Experimenter Nancy J. Cooke is related through marriage to Steve Shope of Sandia Research Corporation, President of the prime company subcontracting to ASU.

VOLUNTARY CONSENT

Any questions you have concerning the research study or your participation in the study, before or after your consent, will be answered by Nancy J. Cooke at ASU Polytechnic, 480-988-2173.

If you have questions about your rights as a subject/participant in this research, or if you feel you have been placed at risk; you can contact the Chair of the Human Subjects Institutional

Review Board, through the ASU Office of Research Integrity and Assurance, at 480-965 6788.
This form explains the nature, demands, benefits and any risk of the project. By signing this form you agree knowingly to assume any risks involved. Remember, your participation is voluntary. You may choose not to participate or to withdraw your consent and discontinue participation at any time without penalty or loss of benefit. In signing this consent form, you are not waiving any legal claims, rights, or remedies. A copy of this consent form will be given (offered) to you.

Your signature below indicates that you consent to participate in the above study.

______________________ ______________________ ___________
Subject's Signature Printed Name Date

INVESTIGATOR'S STATEMENT
"I certify that I have explained to the above individual the nature and purpose, the potential benefits and possible risks associated with participation in this research study, have answered any questions that have been raised, and have witnessed the above signature. These elements of Informed Consent conform to the Assurance given by Arizona State University to the Office for Human Research Protections to protect the rights of human subjects. I have provided (offered) the subject/participant a copy of this signed consent document."

Signature of Investigator_________________________________ Date___________

APPENDIX I

PAYMENT ACKNOWLEDGEMENT

Cyber Awareness II Project Participant Acknowledgement of Payment

I, ____, participated in a collaborative team study that was conducted in Dr. Nancy Cooke's Cognitive Engineering Research Institute (CERI) Laboratory located at Arizona State University's Polytechnic Campus.

I was paid [$10/hr] a total of $________

(Circle one of the following that applies)

For completing that study session for which I was scheduled.

OR

For completing part of the study session for which I was scheduled.

OR

Because other participants did not show up and I was sent home.

OR

I was the fourth person, therefore was rescheduled for a guaranteed spot in a different study session and was sent home.

Print Name: ____________________________________

Sign Name: ____________________________________

Phone Number: _________________________________

Date: __________ Participant Number: ___________ Experimenter Initials: __________