Information Pooling Bias in Collaborative Cyber Forensics

by

Prashanth Rajivan

A Dissertation Presented in Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy

Approved November 2014 by the
Graduate Supervisory Committee:

Nancy Cooke, Chair
Gail-Joon Ahn
Marcus Janssen

ARIZONA STATE UNIVERSITY

December 2014

ABSTRACT

Cyber threats are growing in number and sophistication making it important to continually study and improve all dimensions of cyber defense. Human teamwork in cyber defense analysis has been overlooked even though it has been identified as an important predictor of cyber defense performance. Also, to detect advanced forms of threats effective information sharing and collaboration between the cyber defense analysts becomes imperative. Therefore, through this dissertation work, I took a cognitive engineering approach to investigate and improve cyber defense teamwork. The approach involved investigating a plausible team-level bias called the information pooling bias in cyber defense analyst teams conducting the detection task that is part of forensics analysis through human-in-the-loop experimentation. The approach also involved developing agent-based models based on the experimental results to explore the cognitive underpinnings of this bias in human analysts. A prototype collaborative visualization tool was developed by considering the plausible cognitive limitations contributing to the bias to investigate whether a cognitive engineering-driven visualization tool can help mitigate the bias in comparison to off-the-shelf tools. It was found that participant teams conducting the collaborative detection tasks as part of forensics analysis, experience the information pooling bias affecting their performance. Results indicate that cognitive friendly visualizations can help mitigate the effect of this bias in cyber defense analysts. Agent-based modeling produced insights on internal cognitive processes that might be contributing to this bias which could be leveraged in building future visualizations. This work has multiple implications including the development of new knowledge about the science of cyber defense teamwork, a demonstration of the advantage of developing tools

using a cognitive engineering approach, a demonstration of the advantage of using a hybrid cognitive engineering methodology to study teams in general and finally, a demonstration of the effect of effective teamwork on cyber defense performance.

DEDICATION

I dedicate my dissertation work to my family. If it had not been my sister Preethi's, my wife Aswathy's, my mother Shylaja's and my brother-in-law Chandramohan's encouragement and constant support, I would have not have pursued the doctoral degree and completed it. Whenever I lost my motivation and strength, my sister and my mother was there to pull me out of it, gave me strength to pursue and put me back on track. The last mile of my pursuit was the hardest and I could definitely not have completed it so well if I hadn't had the support, love and encouragement of my dear wife Aswathy. Thank you all for being there for me!

ACKNOWLEDGMENTS

This dissertation work was partly supported by the Army Research Office under MURI Grant W911NF-09-1-0525. I thank Cliff Wang from Army Research Office for his support.

I sincerely thank Dr Nancy Cooke for being a great mentor, advisor and a friend. She has been there for me always and have helped me throughout my time at Arizona State University. I thank my committee members Dr Marco Janssen & Dr Gail Ahn in guiding me and helping me with my dissertation.

I sincerely thank Verica Buchanan & Jessica Twyford for helping me with the experiment. I also thank David, Anirudh, Adriana & Ethan, high school students who helped me with the experiment

Finally, I thank my friends Vijaykumar Ravishankar and Balaji Kadambi for all our insightful discussions.

TABLE OF CONTENTS

Page

Page

Cyber threats are growing in number and sophistication. Cyber warfare is becoming a reality. Therefore, it is important to continually study and improve all dimensions of cyber defense. Although recent research has turned attention towards the human element of cyber defense, there is still considerably less emphasis on understanding teamwork in cyber defense. Through this dissertation, I investigated team cognition in cyber defense analysts, performing a threat detection task, using experiments involving human subjects and agent-based modeling. The information pooling bias is the specific team-level issue that will be addressed through this dissertation work.

Most organizations, small or large, now rely on computers and computer networks for their daily operations. These networks could include devices ranging from less critical personal computers to highly critical data servers and sometimes to even more critical nuclear and power control systems. A study conducted by the Ponemon institute and sponsored by HP (Ponemon Institute, 2013) reveals that there was a 78 percent increase in cybercrime cost from 2009 and that the average number of successful attacks per organization per week has risen to 122 from 102 attacks per week in 2012. This report also points out that the sophistication of attacks has grown because the adversaries now rely on intelligence to obtain sensitive data and to disrupt the services. Hence effective cyber defense capability becomes critical for any organization to protect against the growing number of cyber attacks. Towards this end, there is a sudden surge in demand from organizations across the globe for advanced tools, new services and additional personnel to solidify their cyber defense capabilities. However, simply adding more personnel and tools to the cyber security system does not automatically translate to better security, but instead could be detrimental to the existing system.

Cyber defense can be conceptualized as a complex socio-technical system comprised of many human and technological components working together on different parts of the task. Humans and their technology counterparts have to work together effectively to maintain the security stature of an organization. Therefore, a critical investigation into the social and cognitive aspects of the human analysts, along with investigation into the human systems integration aspects of the task is essential.

Cyber attacks have evolved from traditional isolated Denial of Service (DOS) and malware type attacks often launched by a single independent entity or a small group of hackers to coordinated large scale attacks by state sponsored organizations using stealth modes and advanced persistent threat (APT) types of attacks. Advanced persistent threat is a target-oriented and long-term attack (Liu, Chen, & Lin, 2013) in which attackers use customized malware and bot machines to gain control of network boundary systems in an organization. They use such systems as entry points to navigate by using multi-step attacks, reaching the specific information or system in a large enterprise network. The attackers using APT are very target centric, persistent and spend all of their time and effort to obtain the intended target information or system and hence the name advanced persistent threat. Such kinds of attacks involve social engineering, coordination among multiple individuals attacking multiple network entry points to gain access, attacking different parts of the network, and also happening over a long period of time at a snail's pace to avoid detection. It is common that such targeted attacks go unnoticed for several months.

A recent example of APT is the attack on the Target Corporation in which credit card and debit card information from millions of customers was stolen. It was a planned

attack in which the hackers prepared extensively and used techniques such as social engineering to accomplish spear phishing and insertion of malware in the Point of Sale (POS) terminal to steal the information. The corporation's network was in the compromised status for a long time, but still it went unnoticed. The individual pieces of such a large scale sophisticated attack would seem like isolated events or even seem like benign activity happening across the different parts of a network and happening at different points in time and therefore tend not to generate suspicion. However, when all the individual observations or pieces of evidence are put together they would indicate an attack and hence could be detected early on before severe damage is done to the infrastructure and information.

To detect such kinds of sophisticated attacks, effective and timely knowledge sharing through collaboration among cyber defense analysts becomes essential because the clues needed to detect the attack are spread across many networks, across different points in time, across shifts and across different analysts and system owners. Cyber experts often talk about improving communication about new threats and collaborative response between organizations and between organizations and the government and even on improving communication between different kinds of software products used in cyber defense. But what seems to be overlooked is the fact that the key amidst all these components are the human analysts who are conducting cyber defense. Investigating how the team of cyber defense analysts interact and collaborate can provide insights on their limitations and cognitive biases which in turn can lead to finding ways to mitigate them, thereby improving their overall efficiency.

The personnel conducting cyber defense tasks are named differently in each organization. Therefore, in this paper, I will be referring to them as cyber defense analysts or simply analysts. The analyst could be an employee of the same company or could be an employee of a company that is providing security services to other companies. The cyber defense analysis task involves high uncertainty, high information load, cyber attacks evolving at very high speeds (P. Liu, 2009), and thus little time for an analyst to detect and respond to an attack (Champion, Rajivan, Cooke, & Jariwala, 2012). Analysts are often placed under extreme time pressure. In some settings, they have to process the alerts given to them at a pace of one every two minutes. Thus, a combination of factors that include overwhelming amounts of data, numerous false alarms, and time stress leads to cognitive overload in cyber defense analysts (Champion et al., 2012). Because cyber defense analysis is a complex task, it is often performed by analysts as a group, with each analyst working on a different level of the task with specific domain knowledge and experience. However, simply bringing a group of people together to work on a task would not suffice. To work on such complex tasks we need effective teams of cyber defense analysts. Cyber defense analyst teams are in many cases, loose associations of individuals, rather than functioning and effective teams (Champion et al., 2012). For our definition, a team is a type of a group in which members of the team have diverse backgrounds, but work together in an interdependent manner towards a common goal (Salas, Dickinson, Converse, & Tannenbaum, 1992).

Cyber defense analyst groups display minimal teamwork (Champion et al., 2012) due to cognitive load, lack of motivation, time crunch and also due to institutional policies on employee rewards for cyber defense analysts. Analysts are often rewarded

though bonuses and employment advancement based on the number of critical attacks or intrusions they detect as an individual. Thus, a notion of "Knowledge is Power" is prevalent in the cyber defense community which inhibits analysts from sharing information with their peers in anticipation that they might use that information in the future for detecting attacks. This hampers information flow and communication among the analysts. Therefore, improving teamwork would likely reduce the cognitive overload and stress in analysts, improve information flow and communication, and in turn, improve the overall performance of the analysts.

The most common cyber defense analyst roles are (1) triage analysts or detectors and (2) senior analysts or responders and (3) forensics analysts. As the role name indicates, triage analysts scan the network for intrusion alerts generated from IDS (Intrusion Detection System) sensors to identify the suspicious alerts and reject the false alerts. They then filter associated data pertinent to those suspicious sets of alerts to analyze the data and to decide whether the alerts could actually correspond to an attack. The analysts eventually report their findings to their senior analysts (D'Amico, Whitley, Tesone, OBrien, & Roth, 2005). The reports are peer reviewed before being passed on to the senior analyst. The senior analyst collects these reports and correlates them to determine if there is an attack incident ongoing at a larger level to take the appropriate response (DAmico et al., 2005). The forensics analyst analyzes attack evidence from a longer time period to detect whether the attack evidence correlates to a larger story and whether those evidences also indicates an emerging threat. As described, the cyber defense analysis task is structured loosely in a layered manner in which analysts in one layer feed the analysts in the layer above them with attack pertinent reports for further

processing. Therefore, the quality of decisions made by the analysts working at higher levels of the task depends on the quality of the reports from the analysts working at the lower levels of the task.

Cyber defense analysts currently use e-mail and traditional chat systems as software tools to communicate with each other. They use wikis or generic document sharing tools to collaborate and share their findings. In some organizations, they even use software bug ticketing systems to report the intrusion and attack incidents by raising tickets which then get assigned to various personnel to take appropriate response. There is a lack of well-integrated, custom made, collaboration and reporting tools to assist cyber defense analysts even though effective collaboration is an important component for such critical tasks. The developmental focus in the cyber domain has been predominantly on developing detection tools and visualization tools that will assist in fusing information from multiple sources. Through this work I examine the impact of such a collaboration tool in improving information sharing among the analysts and their detection performance.

To summarize, cyber attacks are growing in number and sophistication and the cyber defense analysts who are designated to protect our organizations from these attacks are cognitively overloaded and do not work as a team. With growing attack sophistication (such as advanced persistent threats) there is a need for timely information and knowledge sharing, but there is a lack of institutional policies, training and tools that promote team work. Cyber defense is loosely structured as a hierarchical process and reports from low level analysts conducting detection tasks determine the overall security stature of an organization. Therefore in this thesis team processes in cyber defense

analyst teams conducting the forensics task are investigated using human-in-the-loop experiments and agent-based modeling. A prototype of collaboration and reporting software to assist cyber defense analysts in collaborating and sharing knowledge effectively with other analysts is developed and tested.

In the next section the literature on team cognition and information sharing is reviewed followed by a discussion of team cognition of cyber defense analysts. Research questions derived from the literature reviewed and past work are presented and a two-part methodology to address them is described. First a human-in-the-loop experiment is conducted. Second an agent based model is developed and used to simulate information sharing among analysts under different models of in-the-head search process. This is followed by a comparison of the experimental results and results from the model simulation. Finally I present a discussion of the findings from this work and conclusions.

BACKGROUND

In this section background on theories and perspectives of team cognition and specific team processes such as communication and team situation awareness are discussed as they are relevant to investigate teamwork among cyber defense analysts from a cognitive stand point. Because human factors based research in the cyber domain is nascent, the literature is reviewed from a related field: intelligence analysis, to identify the types of cognitive limitations and biases that operators face in such domains especially at the team level. Then the literature is reviewed on one such bias assumed to be relevant to cyber defense analyst teams conducting the detection task as part of forensics analysis: information pooling. The hidden profiles paradigm and the methods used in the past to investigate the information pooling bias are also reviewed.

Because agent-based models are applied in this thesis to extend the human-in-the-loop experiments and to study research questions that are difficult to study in the lab, the field of cognitive modeling is introduced along with a discussion of the limitations of existing cognitive modeling methods. Literature on agent-based models which have been predominantly used to study social systems is also reviewed, as well as other related organization and group modeling methodologies such as social network models. Finally, as a background for development of a collaborative tool, the literature on computer supported collaborative work (CSCW) and CSCW in cyber security is reviewed.

Team Cognition

Team cognition is defined as cognitive processes such as decision making and learning occurring at the team level (Salas, Cooke, & Rosen, 2008). Team cognition has a

significant effect on team performance (Cannon-Bowers & Salas, 2001; Cooke, Gorman, & Winner, 2007). The Iranian Airbus tragedy of 1988 in which a commercial flight full of passengers was mistakenly shot down by USS Vincennes (Collyer & Malecki, 1998) is a classic example of the effects of poor team cognition. The three major theoretical perspectives used for explaining team cognition are: shared cognition or shared mental models, transactive memory, and interactive team cognition.

Shared Cognition

The shared cognition or shared mental models view has been around for more than two decades and is the most widely adopted approach used to explain team cognition (Cannon-Bowers & Salas, 1993; Klimoski & Mohammed, 1994). It adopts the concept of mental models (individual) and extends it to explain cognition in teams. Mental models can be defined as "mechanisms whereby humans are able to generate descriptions of system purpose and form explanations of system functioning and observed system states, and predictions of future system states" (Rouse & Morris, 1986), p. 7). Cannon-Bowers, Salas, and Converse (1990) first developed the concept of team mental models based on their study of expert teams: "When we observe expert, high performance teams in action, it is clear they can often coordinate their behavior without the need to communicate" (Cannon-Bowers, Salas, & Converse, 2001) p. 196). Shared cognition (Cannon-Bowers et al., 2001; Cannon-Bowers & Salas, 2001) theory suggests that team performance is dependent on the degree to which the knowledge and understanding of the task and the situation is similar across the members of the team. In simple terms it requires the members of the team to be on the same page. This shared cognition model is often

critiqued for its simplistic view of team cognition given that it is unlikely that all individuals have identical knowledge structures (Cooke et al., 2007)

Transactive Memory

In everyday life, we often use memory systems outside of our own minds (i.e., calendars, notes and directories) to remember things such as meeting times and phone numbers. This is because, as humans, we have constraints on how much we can remember. Miller (1956) showed that there is a limit to how much information we can record in our working memory and that we can only hold up to seven plus or minus two chunks of information. Individual chunks could be a single letter, a group of letters, a word, a number and so forth depending on how an individual group's the information received. However, later studies have disputed the magical number seven proposed by Miller, but still endorse the fact that there is a limit to working memory (Cowan, 1988; Cowan, 2001).

To formalize this type of memory which is distributed across individuals and systems, Wegner (1987) introduced "transactive memory". Transactive memory is related to shared cognition theory (Hollingshead, 1998), where each individual in a group is considered a memory system holding distinct information and knowledge along with the awareness of what others in the group know. Transactive memory is similar to external memory, but instead of remembering to look at book for a certain information, we just remember that our teammate is an expert on a topic and that asking him or her will give us the same information, yet perhaps more quickly and accurately. Interaction and communication are critical group level processes involved in building a good transactive memory system. A transactive memory system is critical for teams. Individuals on the

team must leverage the expertise of others on the team to conduct their tasks (Lewis, 2003).

Interactive Team Cognition

Cooke et al (2013) proposed a theory of Interactive Team Cognition which is a more recent perspective on team cognition which states that team cognition is displayed in team interactions and that it is an activity, not a property or a product, and it needs to be measured at the team level and in the context of the task. This is in contrast to the theory of shared cognition which states that team cognition is the sum of the knowledge of individual team members. However it does not discount the importance of individual knowledge for effective performance, but argues instead that team cognition is not tied to the knowledge of the individual members of the team and that traditional methods to measure team cognition using introspection and subjective queries will not essentially capture the depth of team cognition. The authors also argue that team cognition has to be studied at the team level and not at the individual level and in addition, it has to be studied in the context in which the task is performed which could be a simulated context of the real world task like cyber defense. Training large teams for shared cognition through cross training is not practical and is also not sustainable for large teams and teams performing complex tasks for which each member has a specific set of skills and expertise. Perturbation training (Gorman, Cooke, & Amazeen, 2010) is the training approach associated with interactive team cognition. It involves presenting disruptions or roadblocks while the team performs its task which will consequently require them as a team to modify and coordinate their tasks in new ways.

Interactive team cognition adopts an ecological perspective (human-environment) or between-the-heads (BTH) approach (Cooke, Gorman, & Kiekel, 2008) and suggests that observing team communication is an unobtrusive and an easier method to measuring and understanding team cognition in the context of the task.

By taking the interactive team cognition perspective, it can be deduced that in order to improve cyber defense performance, developing tools and interventions focusing on just improving the individual's knowledge and decision making abilities will not be sufficient. Developing tools and training interventions that improve team level processes such as communication and information sharing in cyber defense analysts is equally important for improving system level performance.

Communication is the key medium through which a team of humans form relationships, collaborate and share information. Communication could be conducted through various forms such as face-to-face communication, non-verbal communication and even through virtual mediums such as telephone networks and internet networks. Whatever the form be, communication is a key element in the team process.

Team Communication

Early research on teams reported that team communication can be inhibitory to team performance and that it has to be restricted (Briggs & Naylor, 1965; Johnston & Briggs, 1968; Williges, Johnston, & Briggs, 1966), which led researchers to focus predominantly on improving individual efficiency. Later research (Brannick, Roach, & Salas, 1993; Salas et al., 1992; Stout, Salas, & Carson, 1994) however reported that team processes such as communication and interaction are also essential for performance.

Communication can be verbal communication or non-verbal communication (such as gestures), synchronous communication or asynchronous communication.

To investigate team cognition, from the interactive team cognition perspective, it is imperative to record all verbal communications taking place between the members of the team during the experiment session. The mode of communication could be through face-to-face when it will be recorded through a microphone or it could be through a computer chat system when the communication will be saved as text files. This communication data will then have to be transcribed and analyzed to identify patterns and gaps and consequently to gain important insights about team cognition.

Communication analysis traditionally involves transcribing and coding communication data manually. Such a manual process is strenuous, time consuming, rated subjectively and often analyzed outside the context (Cooke, et al., 2008). Automated methods to analyze communication data are becoming popular. Latent Semantic Analysis (Landauer, Foltz, & Laham, 1998) and keyword indexing are two automated methods for analyzing the content of the communication data. ProNet (Cooke, Neville, & Rowe, 1996), which is based on Pathfinder network scaling (Schvaneveldt, Durso, & Dearholt, 1989), is a method for analyzing flow patterns in communication data.

Team communication and collaboration have been identified as important for explaining performance differences between teams performing cyber defense analysis (Jariwala, Champion, Rajivan, & Cooke, 2012). Simply increasing communication among cyber defense analysts may not improve performance unless the communication is useful communication that can contribute to advancing team cognition such as team

situation awareness. Situation awareness in particular has gained wide interest in the cyber security domain because cyber security is a hyper-dimensional environment and it is important for analysts to be aware of key events happening in the network and prioritize them by filtering out irrelevant information in order to take appropriate response.

Situation Awareness

There are several definitions of situation awareness (SA), however the definition which is widely used is "the perception of the elements in the environment within a volume of time and space, the comprehension of their meaning, and the projection of their status in the near future" (Endsley, 1995, p. 97). Endsley's definition and model of SA has been widely adopted by researchers in the cyber domain because it is similar to the JDL (Joint Directors of Laboratories) data fusion model which is widely used in the cyber domain. JDL data fusion model is a five level model that describes how data from multiple sources can be integrated to get a unified view (Hall & Llinas, 1997; Blasch, Bosse, and Lambert, 2012). Level 0 in the JDL model is called Sub-Object Assessment and involves detecting signals in the incoming data. This is similar to perception phase of Endsley's model. Level 1 (called Object Assessment) and level 2 (called Situation Assessment) involves association and aggregation of the information from level 0 to form hypotheses and to understand the situation at hand. This is similar to the comprehension phase of Endsley's model. Finally Level 3 (called Impact Assessment) involves drawing inferences and impact estimation which is similar to the projection phase of Endsley's model. Level 4 and 5 are the more recent additions to the JDL model to take into account the process and user cognition aspects of data fusion.

Information quantity, which is excessive in the cyber domain, has been identified as one of the key factors impacting SA (Endsley, 2000; Taylor, 1990). Endsley (1995) suggests that situation awareness is a product of the situation assessment process, performed by analysts while working with large quantities of information. Technologies such as data filters (example: Wireshark and Snort), fusion algorithms (Stotz & Sudit, 2007) and visualizations (Shiravi, Shiravi, & Ghorbani, 2011) are being developed to assist in cyber analysis and to provide analysts a better picture of the complex cyber world. However, it is important to recognize that the "awareness" in situation awareness resides neither with the analyst alone, nor with the technology alone, but with the joint human-technology system (McNeese, Cooke, & Champion, 2011).

Situation awareness is a dynamic cognitive process whereby an individual continuously modifies and updates his or her SA with new information from the environment. This dynamic property makes SA difficult to assess and measure (Prince, Ellis, Brannick, & Salas, 2007). Factors found to affect SA in analysts are: Attentional tunneling, Requisite Memory Trap, Workload, Data Overload, Misplaced Salience, Complexity Creep, Errant Mental Model and Out-of-loop syndrome (M. R. Endsley, 2006).

Situation awareness conceptualized at the team level is called team situation awareness (Team SA). Team SA is viewed as an important factor to be considered in designing human-machine systems and interfaces (Shu & Furuta, 2005). Endsley defines team SA as "the degree to which every team member possesses the SA required for his or her responsibilities" (Endsley, 1989). According to this perspective, the team's performance depends on the level of situation awareness in each of the team members

and one member's poor SA can affect the team's performance. However, this model of team SA does not go far enough (Gorman, Cooke, & Winner, 2006). It may be relevant to homogenous groups, but not to heterogeneous teams and this perspective may not suffice as team increases in size (Cooke, Gorman, & Rowe, 2004). If a team is truly an interdependent group, then each team member will have different, though perhaps overlapping, perspectives on the situation. In a complex and dynamic world, it is likely that two or more perspectives on the team will need to be fused in order to have SA that extends beyond an analyst's screen of alerts. The fusion takes place through some form of team interaction – often communication. For example, one analyst may be aware of a denial of service attack on a network server and once this information is joined with another analyst's awareness of two other similar attacks on a different network a bigger picture emerges. Without the interaction, the team as a whole cannot perceive, comprehend, and project.

In short, team SA is much more than the sum of individual SA (Salas, Prince, Baker, & Shrestha, 1995). This follows from the perspective of Interactive Team Cognition (Cooke, Gorman, Myers, & Duran, in press) that espouses that cognitive processing at the team level occurs through team interactions situated in a rich context. This view of team cognition can be contrasted with others that focus on the aggregate of individual knowledge (e.g., Langan-Fox, Code, & Langfield-Smith, 2000). Thus by placing the focus on team interaction, team situation awareness can be described as the coordinated perception of change in the environment by team members that serve as the basis for effective action (Gorman et al., 2006). According to this view, team SA means, members of a team becoming aware of different aspects of the situation and knitting the

pieces of the puzzle together through communication or other interactions to achieve team situation awareness and to take appropriate actions. (Salas et al., 1995). This view (Cooke, Salas, Kiekel, & Bell, 2004) suggest that team members through team interactions transform individual knowledge to collective knowledge and in the process achieve team situation awareness.

Team cognition and its processes have been profusely investigated in other domains such as medical diagnosis, air traffic control and intelligence analysis. Hence there is a large collection of literature on cognitive biases that affect team cognition in those domains. However, as suggested earlier, researchers in the cyber security domain have predominantly focused on the technical side of the problem, even though it has been widely characterized as a socio-technical problem (Dutta & McCrohan, 2002; Kraemer, Carayon, & Clem, 2009). Studies to explore the human side of the cyber problem are minimal and most have focused on the individual analyst because the task on first sight seems to be an individual cognitive task. Champion et al., (2012) found that team processes such as communication and collaboration play important roles in the outcome performance, which is detecting potential cyber attacks. There is very little work done so far to explore the various aspects of team cognition of cyber defense. Therefore I will review literature from a related field of work: intelligence analysis, to identify the types of cognitive limitations and biases the analysts face in such domains especially at the team level. Intelligence analysis is a field of work similar to cyber defense analysis, but instead of looking at computer logs and intrusion alerts, analysts look at email intercepts, phone taps, and so forth to identify potential attacks on the nation (Puvathingal & Hantula, 2011).

Intelligence Analysis

Lowenthal (2002) defines intelligence analysis as "the process by which specific types of information important to the national security are requested, collected, analyzed and provided to policymakers" (p. 8). Therefore, to get a preliminary understanding on the probable cognitive limitations and biases in cyber defense, related work in the intelligence analysis domain was examined. The intelligence analysis task is also found to be cognitively demanding work, with considerable time pressure and is also considered risky because the decisions that the analysts take can either help the nation or could create unwanted chaos (Johnston, 2005). Similar to analysts in the cyber domain, intelligence analysts must deal with intentionally misleading information, missing information and incorrect information (Johnston, 2005). And similar to cyber defense, the notion that "knowledge is power" prevails and the culture is competitive whereby individuals are trying to get the first hand information before others for job bonuses and promotions which inhibits information sharing (Vogt et al., 2011).

Kahneman and Klein (2009) suggest that it may not be possible to achieve optimal decisions in complex, information-overloaded domains such as intelligence analysis. Loss in group level process such as communication and collaboration can also lead to suboptimal decision making in complex environments (Hill, 1982). Improving group level process in information overloaded environments could lead to more optimal decision making because the group will then be able to effectively use the diverse knowledge, experience and skills of the group members (Laughlin & Bonner, 1999; Mesmer-Magnus & DeChurch, 2009a). Factors such as the common knowledge effect

(CKE), confirmation bias, overconfidence, and group polarization are found to cause process loss in intelligence analysis teams (Straus, Parker, & Bruce, 2011).

The common knowledge effect or CKE (Gigone & Hastie, 1993) occurs when a part of a group knows the relevant information, but fails to communicate it to the rest of the members assuming it is common knowledge. Confirmation bias occurs when a group looks only for information that serves as supporting evidence to a preconceived hypothesis they developed about the situation instead of looking for information that would dispute the hypothesis (Heuer, 1999; Johnston, 2005). Heuer (1999) suggested tools used in intelligence analysis should challenge the analyst to reduce the confirmation bias effect.

Overconfidence is when individuals or groups overestimate their knowledge which leads them to make overconfident decisions and is found to exist in complex environments with uncertainty such as the intelligence analysis domain (Heuer, 1999; Yates, 1990). Teams with many overconfident individuals tend to have less interaction within the team because they do not find the need to seek additional information from other team members. Such teams are found to exhibit low performance in comparison to teams with not so confident members because they have to verify their decisions or findings (Puncochar & Fox, 2004; Sieck & Arkes, 2005).

Group polarization occurs when individuals change their decisions and attitudes they have towards the problem to match the rest of the team members' decisions (Brauer, Judd, & Gliner, 1995). Kelly, Jackson, & Hutson-Comeaux (1997) and Schulz-Hardt, Jochims, & Frey (2002) suggest that having heterogeneous teams (i.e. teams with individuals with diverse experience, skill and knowledge) in such domains could reduce

the effect of such biases because having individuals with diverse experience and skills in the same team will enable the team to view and think about the situation at hand from different perspectives and will enable the team to come up with different strategies.

Simply knowing about the existence of various biases is insufficient. It is necessary to get an understanding about why such biases are present and to find the source of the bias in order to effectively mitigate them. The next section reviews the literature on human reasoning to find answers to the questions about the origin and source of biases.

Reasoning and Biases

Reasoning is the underlying ability in humans that enables them to think, make sense of things, make arguments, form new beliefs and opinions and reinforce or reject existing beliefs and opinions (Kompridis, 2000). It is considered to be the distinguishing characteristic of human beings that enables them to innovate and conduct knowledge-based tasks. However a long literature indicates that humans are poor at reasoning and that decisions arising from our reasoning are often flawed due to biases (Kahneman, Slovic, & Tversky, 1982).

The traditional belief was that humans developed the reasoning ability to find truth, to reinforce personal beliefs and to improve individual cognition (Kahneman,2003, p. 699; Sloman,1996, p. 18). More recently, an evolutionary psychology perspective was taken to explain the origin of reasoning by Hugo Mercier and Dan Sperber through their argumentative theory (Mercier & Sperber, 2011). They state that humans developed reasoning to support social functions. Evolutionarily, humans started collaborating to

hunt, find food and to defend them from threats. To work in groups, humans had to develop agreements by resolving difference of opinion which required them to develop reasons for their opinions and to communicate this reasoning to others in the group.

According to the argumentative theory, humans in a group play two roles: convincer and convincee. Convincer, based on their intuitive beliefs, develop and communicate arguments to others in a group to persuade others to also be convinced about what they believe is the correct course of action. To effectively persuade, only confirming arguments that support their individual beliefs are collected and presented. In return, others in the social setting (convincee) may be ready to be persuaded because they have the same set of beliefs or might defend because they do not share the same set of beliefs. But from an evolutionary stand point, because working as a group was essential, people would take moderate stand points in which they resist new arguments initially, but later on accept arguments that are valid.

As it can be deduced, such a reasoning process that involves finding reasons to defend one's opinions to others, aimed at supporting social functions, is biased. If the beliefs and reasons are valid then they will lead to good decisions; otherwise they will lead to flawed decisions. Working alone can more often lead to biased decisions because there is now way to evaluate one's beliefs, opinions or hypotheses. So to mitigate such biases humans have to work in group with a heterogeneous set of people and at the group level, members should be able to produce competing arguments, evaluate arguments, accept good arguments and reject the bad ones.

Complementing the argumentative theory with interactive team cognition theory, it can be deduced that such biases have to be studied at the group/team level to identify

patterns of arguments communicated back and forth between the members of a team and find interventions to help individuals to produce new competing arguments and also to evaluate all hypotheses presented by others by using all the information available with them.

To identify cognitive biases that would be present in cyber defense analyst teams conducting the detection task, a closer look at the task of such analysts is necessary. As described earlier, the primary task of cyber defense analyst conducting the detection task as part of forensics analysis is to analyze attack evidence from a long time period to detect whether the evidence correlates and whether an emerging threat is indicated. Forensics analysis are sometimes performed in collaboration but mostly it is done in isolation. Therefore the existing process is ineffective and it would be beneficial if analysts pooled observations or evidence from their peer analysts' reports to find associations. Presumably, if the analysts were able to effectively pool and fuse their individual observations, attack detection performance would improve.

However the literature reviewed shows us that biases such as the common knowledge effect and confirmation bias can lead to a biased discussion and that having analysts to discuss and fuse information might not lead to better performance after all. Thus, careful investigation is necessary to identify whether such biases affect cyber defense analysts. Then techniques and tools to mitigate these biases have to be developed.

The common knowledge effect, information pooling bias, and confirmation bias are parts of a larger paradigm popular in social psychology called the hidden profile paradigm (Straus et al., 2011).

Hidden Profile Paradigm

Teams are employed to make complex decisions because they can expand the pool of available information and when the team members pool all of their diverse experience and information we intuitively assume they would achieve optimal decisions that would almost be impossible for an individual expert to achieve (Mesmer-Magnus & DeChurch, 2009b). Similarly when cyber defense analysts collaborate they would have to pool all the information available to them to make sense of the entire threat landscape pertinent to the network they are defending. The information pooling process involves sharing and receiving of information between members to update one's own mental model about the situation at hand, to make new connections, and for general sense making. If they do not share all of the information, especially their expert knowledge, with each other they cannot make the connections that might exist between their individual observations, identify the possible trends in their observations, and discover overlapping incidents happening at different parts of the network.

Intuitively one might think that when a group of people discuss they would share the novel or unique information available to them instead of the information already known to all because the novel information and arguments are more influential than that which is known to all (Burnstein & Vinokur, 1977). However, past research shows that groups are not so effective in pooling all of the available information (Lu, Yuan, & McLeod, 2012; Wittenbaum, Hollingshead, & Botero, 2004). The information pooling process in a group is known to be rife with cognitive biases (Puvathingal & Hantula, 2011; Stasser & Titus, 1985; Straus et al., 2011). One such cognitive bias is the shared information bias or information pooling bias in which the pre-distributed information and

analyst knowledge biases them to share information that is already known to others and prevents them from sharing new information (Stasser & Titus, 1985). The possible causes for this bias include preferences made by the members before the discussion causing them to confirm to initial beliefs, the memory recency effect, the frequency of mentioning the shared information, and the need for individuals to seek social validation for their initial beliefs (Lu et al., 2012; Stasser & Titus, 1985). In the past this paradigm has been mostly investigated from a social psychology point of view, but there could be underlying cognitive factors that are causing such a bias and that the social causes found could be mere manifestations of the individual cognitive limitations.

Strasser and Titus (1985) introduced the hidden profile paradigm. The research showed that group discussion might not be an effective means for exchanging new or novel information. In such information sampling studies the decision making groups are asked to make decisions by pooling each individual's information and discussing the different alternatives. In such studies the information is distributed across the team members such that some information is shared by all the team members, but there is some unique information given to each team member. The information that is shared is called "Shared" and the other is called "Unshared" (Wittenbaum et al., 2004). The goal set for the team is to pool all the unique information to achieve the optimal decision. However time after time the studies show that the groups do not pool the unique information available to individual members; rather they keep discussing the shared information (Lu et al., 2012). There is a difference between the groups studied under this paradigm and cyber defense analysts conducting the detection task. The original experiments focused on getting one optimum solution such as finding the murderer by a mock jury team or

finding a leader by a mock political caucus team. In these cases there is a clear presence of choices and only one choice is optimal. Also the shared information between the members of the team about the items to make decisions about would be exactly the same. But in cyber defense there is no one optimal choice given that there is a need to respond to all attacks. However, there will be different priorities between different attacks for which a higher priority could be given to attacks that are large in scale and which are stealthy because they usually lead to maximum damage. Also the individual analysts have similar information about similar kinds of attacks which would be the shared information, but the shared information is not exactly the same across all team members.

Groups with unequal information distribution were found to be eight times less likely to find the solution than were groups having full information (Lu et al., 2012). It was also found that percentage of unique information mentioned out of the total available information (the information coverage measure) and the percentage of unique information out of the total discussion (the discussion focus measure), were positively related to decision quality, but the effect of information coverage was stronger than that of discussion focus (Lu, et al., 2012). Stasser and Titus (1987) noted that when each member of a group discloses the same amount of shared and unshared information, in other words having no bias towards certain information, there is still a sampling advantage towards the shared information at the group level because more people know the shared information.

Groups tend to communicate and discuss information that is known to majority of the members of the team, but fail to communicate information that is uniquely available to each person. Therefore, simply getting a team of analysts to collaborate and discuss is

unlikely to provide a boost in the performance. Critical investigation is necessary to identify the cognitive issues involved when cyber teams collaborate, to determine whether they communicate and collaborate effectively, and to discover interventions to help them communicate and collaborate more effectively.

To study such cognitive biases, an environment with sufficient experimental control and which is representative of the real world cyber defense task is required. Field studies offer very little experimental control, but the findings from observations and interviews would be ecologically valid. Conducting field studies in the cyber defense context is difficult because the cyber defense task is highly technical and confidential in nature which inhibits participants from participating in the research. Also collecting cognitive measures is often difficult with field based studies.

Experiments on the other hand are a better option to study the human element of cyber defense. By nature they provide good experimental control and the task is relatively easy to simulate in the lab with good fidelity because it is mainly a computer based task with low human mobility as opposed to other tasks such as military warfare which involves external and environmental effects that would affect the human while the task is carried out.

Although human-in-the-loop experiments offer good experimental control, they offer less flexibility to understand the various cognitive processes involved in doing a certain task. Computational cognitive modeling methodology can supplement this limitation and can enable the experimenter to study the intricacies of cognitive processes which are difficult to infer from experiments (Newell 1990 and Sun 2009). But such cognitive models, however accurate, are not recommended to be used in isolation to

develop cognitive theories of a certain task. They should always be coupled with experiments and used to extend existing experiments, generate new hypotheses, develop future theories, and to explore a combination of parameters which are difficult to explore with a lab based experiment. (Sun 2008)

Cognitive Modeling

Cognitive models represent one or many human cognitive processes such as perception, decision making, and language comprehension. Cognitive models are mostly built for the purpose of understanding and predicting human cognition. Cognitive models come in a variety of forms from simple box and line based diagrammatic models to models that use mathematical equations and even to dynamic computational models that use software programs.

Computational cognitive modeling helps to describe specific cognitive processes, associated with a task or in general, using computer algorithms (Turing, 1950). Some have taken a strictly artificial intelligence perspective (Schank & Abelson, 1977; Minsky,1975) in which there is less emphasis on comparing the model output with human data.

Computational cognitive modeling using cognitive architectures has been receiving more traction recently because it provides more capabilities, allows testing and validation and even provides visual capabilities to observe the phenomenon modelled as it unfolds over time. Cognitive architectures such as ACT-R (Anderson, 1996) and SOAR (Laird, Newell, & Rosenbloom, 1987) are popular examples of computational cognitive modeling.

These architectures provide a framework and libraries to build models which make modeling simpler compared to building a model from the bottom up. ACT-R and SOAR are types of computational cognitive modeling methodologies that use mathematical formulations of cognitive process combined with the power of programming language and computational power of computers to run complex computer simulations of various cognitive processes. ACT-R and SOAR have grown over the years into a large collection of libraries of cognitive processes.

In addition to using such cognitive architectures and models for strictly comprehension and theory development purposes, they have also been used to develop intelligent applications such as intelligent assistants (Guerra, 2011), learning assistants (Koedinger, Anderson, Hadley, Mark 1997) and synthetic teammates (Ball et al., 2010).

Such cognitive architectures have predominantly focused on the individual's cognition and have not been extended to agent-based, groups and particularly teams. A need to combine agent-based simulations with intelligent agents has been expressed in the past (Sun, 2006b) and there has been some work in the past on developing agent-based cognitive architectures (Sun, 2006a). However, work in that direction has been slow and there is a dearth of research on integrating the individual intelligent agents with agent-based simulations.

Although new agent-based simulation environments are being developed by extending existing cognitive architectures, another approach would be to leverage existing proven agent-based simulations used in social science and build cognitive agents within them. The cognitive processes modelled in these agents need not be as comprehensive as in these architectures, but could be limited to processes that are

relevant to the intended research. This could be done instead of putting the effort in building a complete intelligent agent and extending it to agent-based. For example, for studying team cognition more effort could be put into modelling the interactions between the agents and less effort in developing a perfect intelligent agent. Though the approach of building a unified architecture is the ultimate goal, researchers could use this approach in the mean time to explore questions pertinent to team cognition and group cognition.

Agent-Based Modeling

Agent-based modelling is a computational modeling technique used for research in the social sciences research domain. It is often used by social scientists to study several social constructs such as hunter-gatherer problems (Barceló et al., 2013; Janssen & Hill, 2013), prisoner's dilemma (Wilensky, 2002) and so forth. It has also been used often to study epidemic diffusion in the population (Carley et al., 2006). The prime focus in developing agent-based models is in studying the interactions between the agents and to study the patterns and emergent properties produced by those interactions.

In agent-based models, the agents act autonomously to achieve set goals which require them to interact with other autonomous agents locally and also to develop adaptive behavior based on the current environmental state (Grimm & Railsback, 2005). Agents are assigned rules and algorithms to carry out the individual process and for interacting with other agents. Because the focus has been on the social interactions, the assumptions made about individual cognition have been very rudimentary (Sun, 2006a). Agent-based modeling can be extended by leveraging findings from cognitive sciences to model more intelligent agents. Hence the outputs from the cognitive modeling efforts can

be used as inputs for developing agent-based models for which the individual agents in the agent-based simulation could be developed based on the cognitive models of cyber defense operators.

Interestingly, agent-based models are described as a methodology to study macro-level patterns emerging from micro-level social interactions between agents. This definition has a stark similarity to the definition of interactive team cognition (Cooke, Gorman, Myers, & Duran, 2013) which is also characterized as a macro-level phenomenon emerging from micro-level interactions between the members of the team. Hence, agent-based modeling seems well-suited to modeling team cognition in cyber defense operators. Similar to cognitive modeling the output of agent-based models should be compared to that of an associated human-in-the-loop cyber defense experiment.

Agent-based models can extend experiments to explore new phenomenon that are difficult to investigate with human participants such as experimenting with very large teams and longitudinal experiments that extend over a long period. Agent-based models must be developed in close alignment with the human-in-the-loop experiment it is extending where the rules of individual process and rules of interactions must be developed based on the tasks performed by human participants in the lab.

Agent-based models can be useful in generating new hypotheses, developing future theories, and exploring a combination of parameters which are difficult to explore with a lab based experiment (Sun 2008). Also developing such computational models will allow easy sharing and reuse and extension by a larger community.

One of the objectives of this dissertation is to develop a prototype collaboration tool that will improve information sharing. Towards that objective I will be reviewing the

different collaboration tools used in the intelligence analysis domain and will also be reviewing existing collaboration tools used in cyber security.

Tools Aiding Collaboration

There are a range of tools (web based and standalone) in the market that help teams to communicate and collaborate, conduct discussions, build knowledge and develop hypotheses collaboratively. Tools in the form of chat interfaces, online forums and email clients are commonly used for communication and collaboration. Chat-based systems enable synchronous communication. Forum and email based systems enable asynchronous communication. Such generic tools would provide some collaboration assistance, but developing collaboration tools specifically for each domain considering all of the unique requirements is necessary in order to improve team performance in each domain. This is particularly important for domains that primarily involve knowledge work such as medical diagnosis, research and development, intelligence analysis and cyber defense. In such domains, the individuals or the groups have to construct new knowledge out of massive amounts of information, but humans have mental limitations that strain this process and hence require carefully designed tools that would enhance the ability of the groups to construct, organize and share knowledge (Stahl 2006).

Collaboration tools for knowledge sharing are popular in the educational domain. Tools such as Teacher's Curriculum Assistant (TCA), Hermes and webguide are used for collaborative knowledge building (Stahl, Koschmann, & Suthers, 2006). As transdisciplinary research is gaining more traction, collaboration tools are being considered essential for managing transdisciplinary research, to share knowledge across

transdisciplinary teams and teams that are geographically distributed (Bietz et al., 2012; Schnapp, Rotschy, Crowley & O'Rourke, 2012; Vogel et al., 2013).

Several collaboration tools are being developed for collaborative intelligence analysis which is a domain that is comparable to cyber defense in terms of both cognitive load and tasks performed. There have been efforts to develop stand-alone collaboration tools that are used strictly for collaboration and then there have been efforts to develop collaboration tools that are deeply integrated with the existing analysis task. Collaboration features integrated deep into the existing work process allow the analyst to use an integrated system and thus does not divert attention from the primary task of analysis (Bier, Card, & Bodnar, 2008).

POLESTAR (Pioch, & Everett, 2006) is a knowledge management tool for intelligence analysis with extended collaboration features. The tool suggests what other analysts have reported who were working on similar analyses by leveraging their notes and reports. This way of suggesting would lead to *ad hoc* group creation. It also allows analysts to share their reports with each other and assists in getting peer reviews from their team members. Cemberia (Isenberg & Fisher 2009) is a tabletop (Microsoft surface) visual analytic software that allows small groups of analysts to collaboratively forage for the information available and construct a story and hypothesis about the situation at hand. The software uses the visualization technique of *brushing* and *linking* (Buja, Mcdonald, Michalak & Stuetzlew, 1991) in which the changes made by one analyst are propagated on the other analyst's visualization, thereby improving awareness.

Commentspace (Willett, Heer, Hellerstein, & Agrawala, 2011) is a collaborative visual analytic tool that uses tags and links between individual comments on a forum

based visualization system to help analysts in information gathering and sensemaking. Individual analysts can post comments to a topic, tag comments as hypotheses, or a question and in return, other analysts can find existing evidence and link each to hypotheses or questions posted on the forum, consequently helping all analysts to make sense of the situation at hand. Then there are collaboration tools that help analysts organize information around entities such as people, places and things instead of having them collaborate on free textual information (Bier, Card, & Bodnar, 2008).

The other approach used to facilitate collaboration in the intelligence analysis domain is through large high resolution displays which can be used as collaboration tools for enabling co-located individuals to share information and to make sense of the information collaboratively (Vogt, et al, 2011). The software used by individual analysts is configured to receive information from multiple input devices and analysts, thereby facilitating information sharing with the team.

However, in the cyber domain, there has not been much focus on developing collaboration tools to improve collaboration and information sharing between cyber defense analysts. A few research projects that have come close to looking at the collaboration and information sharing aspects of cyber security include VULCAN (Hui et al., 2010) and TAXII (Connolly, Davidson, Richard, & Skorupka, 2012). The VULCAN project focusses on improving information and situation awareness between cyber analysts across organizational boundaries. They proposed to achieve this by tracking each analyst's work process and extracting data on the internet sources they search and questions they ask on the data they are analyzing. They use this information to assist other analysts working in other organizations during their analysis. TAXII or Trusted

Automated eXchange of Indicator Information is a community driven development effort which allows one organization to safely and in an automated manner share the threats they are observing in their organizations which might help other organizations to prevent such a threat from affecting them. Overall, there is a lack of effective tools and solutions to assist cyber defense personnel to collaborate and share information within an organization.

Summary of Background

Cyber defense analysis is a complex task in which analysts are overloaded with missing, incorrect, and intentionally altered information leading to cognitive overload, low situation awareness and stress that affects their performance and in turn affects organization's security posture. Though cyber defense analysts are set up to work as a team, there are a variety of factors that thwart teamwork. From expert interviews, surveys and from past literature, it was found that factors such as institutional rewards structures, lack of team training, lack of collaboration tools and in addition the human biases such as the common knowledge effect, the confirmation bias, overconfidence, the information pooling bias, and group polarization could be affecting their team work and performance.

RELATED WORK

This dissertation work was inspired not only by the relevant literature, but also by research conducted by the author and others. Surveys and interviews conducted with cyber defense analysts and subject matter experts from both the military and private organizations indicate that cyber defense analysts in general lack teamwork and that the rewarding structures employed to motivate them are also conducive to individual effort instead of team effort (Champion et al., 2012). Cyber defense analysts are offered individual level bonuses for good performance. This kind of rewarding structure could be leading to a notion of knowledge is power in cyber defense analysts which in turn would inhibit information flow and communication among them. Loss of information flow and communication would affect availability of essential knowledge for attack detection which will in turn deteriorate the overall security performance.

Recognizing team level efforts, in addition to providing individual level rewards, would encourage analysts to proactively collaborate and share information. Analysts in a team could leverage each other's expertise and knowledge during attack detection and share the rewards for their performance. This would lead to higher performance than conducting attack detection individually and keeping all the rewards for oneself.

A three person team, human-in-the-loop experiment (Rajivan et al., 2013), was conducted to investigate the effect of team level rewards in contrast to individual level rewards on attack detection performance in cyber defense analysts. Participants used the simulation environment CyberCog to conduct the cyber defense task. Participants were primed and rewarded to work either individually or as a team while triaging and detecting cyber attacks from the intrusion alerts. In the experiment, the participants were

overloaded with data and a time crunch to recreate the cognitive overload experienced by real world cyber defense analysts. Also participants in both experimental conditions could either choose to transfer unfamiliar alerts to other members of the team for analysis or learn to analyze those alerts themselves using the lookup system which provided a textual description of the analysis procedure.

The primary measure of team performance was based on the Signal Detection Theory (Stanislaw & Todorov, 1999). For the alerts analyzed the number of hits (number of suspicious alerts the team classified as suspicious), misses (number of suspicious alerts the team classified as benign), false positives (number of benign alerts the team classified as suspicious), and correct rejections (number of benign alerts the team classified as benign) were recorded. Subjective impressions of workload were measured using the NASA TLX (Hart & Staveland, 1988) at the end of each Mission.

It was found that team performance was significantly better than group performance on novel and difficult to analyze alerts. It is imperative that cyber defense analysts analyze such novel, non-intuitive "hard" type of alerts accurately because they are more often the real attacks which lead to destructive and expensive consequences. From the results it can be inferred that the cyber defense analysts can achieve higher performance by simply collaborating with other analysts to leverage each other's unique expertise and knowledge to analyze alerts that are novel and non-intuitive to them. Putting the extra effort to collaborate on everything might be detrimental to their performance.

In the experiment described previously, the participant teams were organized to be heterogeneous in terms of the knowledge they possessed from training to conduct the

tasks. However, cyber defense groups in the real world are usually composed of people with similar experience and knowledge with similar responsibilities. Therefore it is important to contrast this performance to that of a homogenous group of cyber defense analysts as in the real world. It is also important to investigate the effect of team size on cyber defense performance.

An agent-based model (Rajivan, Janssen & Cooke, 2013) that simulates the task of computer network intrusion detection and the interactions among analysts while conducting intrusion detection was developed. The model was an extension to the human-in-the-loop experiment described. The agent-based model extended that experiment to investigate the research questions: Does team heterogeneity affect attack detection performance in cyber defense analysts? Do large teams or small teams lead to better attack detection performance in cyber defense analysts?

Agents in the model were characterized by their technological capabilities they possess for cyber defense. Based on working memory literature (Miller, 1956) each agent was assigned a memory capacity because there is a boundary on the possible number of capabilities an agent can possess, given that the agents represent humans. All agents were also assigned an equal set of points which they can expand by receiving rewards for solving alerts.

Each agent can analyze the alert assigned and get rewards if the agent already had the required capability with them or if not the agent has two options: (1) Learn how to process the alert with a certain probability of accuracy and acquire the capability or (2)

Collaborate, if allowed, with the partners they found and acquire the capability from the partners instead of learning. Both of the options had a cost and payoff.

The three experimental conditions used in the model were: No collaboration, conservative collaboration, and progressive collaboration defined in terms of the way the agents find their partners. In the conservative condition, the agents searched for other similar agents in terms of capabilities (homogenous teams) and in the progressive condition, the agents searched for distinctly different agents (heterogeneous teams). The maximum number of partners an agent searches for depends on the maximum partnership team size (three, five or six).

Results indicated that collaboration had a significant effect on performance. Collaboration in comparison to no collaboration leads to better performance in terms of alerts solved. Furthermore, when agents collaborated with agents who were less similar to themselves they solved more alerts when compared to agents who collaborated with other agents who were very similar to themselves. This demonstrated that team performance would be better in a heterogeneous team. The size of the team was also found to have a significant effect on the performance in terms of rewards. Smaller teams fared better when compared to larger teams. The final take away from the model is that small teams of heterogeneous analysts would improve the overall cyber defense performance in terms of alerts solved and at the same time would prevent analysts from being under rewarded.

OVERALL TECHNICAL APPROACH

I adopted a multifaceted and multi-disciplinary methodological approach to investigate my research questions. I used human-in-the-loop experiments to observe and measure the effect of information pooling bias in cyber defense analyst teams. I developed a prototype visualization collaboration software interface from a cognitive engineering perspective to test whether such a visualization tool would help in mitigating the bias in cyber defense analyst teams. The tool was then tested in the same human-in-the-loop simulation environment that was used to measure the bias in the first place. Then I extended this empirical work computationally using agent-based simulations to explore the underlying cognitive process theories that might be contributing to the bias. The methods described draw from disciplines such as cognitive science, social science and computer science.

The overall technical approach and the outputs of this research are described in Figure 1. The cyber defense analysis process is unique in ways such as the highly technical nature of the domain, the type of data that need to be analyzed, the type of threats, the large variance in speed with which the attacks occur and the hyper dimensionality of the space in which attacks occur. But it is also similar in the analysis process, cognitive load and other cognitive characteristics to domains like intelligence analysis and physical threat sensing and detection.

Based on the parallels identified, I developed hypotheses about cyber defense teams to test using task centric simulations and human-in-the-loop experiments. The participants in my experiment used a simplified version of the synthetic task environment (Cooke, Rivera, Shope, & Caukwell, 1999) called CyberCog (Rajivan, 2012) to perform

the detection as part of forensics analysis task of a real world cyber defense analyst. Synthetic task environments are simulation environments built to recreate the real world tasks and cognitive aspects of the task with highest fidelity possible, giving less focus on the realistic appearance of the environment (Cooke & Shope, 2004).

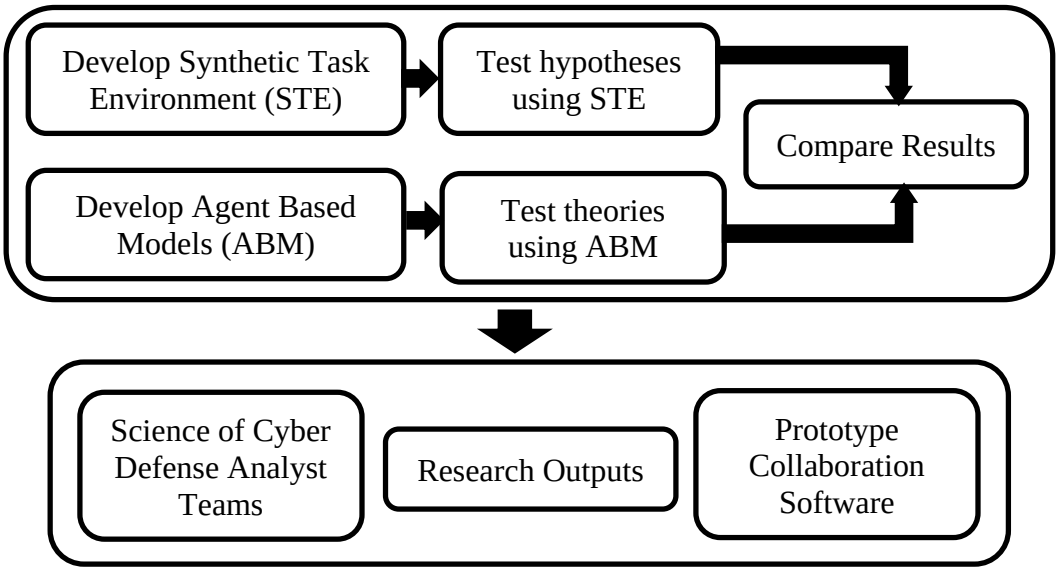

Figure 1. Pictorial representation of overall technical approach

CyberCog: Simulation Environment

CyberCog is a three-person synthetic task environment that simulates the cyber defense analysis process. The CyberCog system presents a simulated set of network and system security alerts to experimental participants who have to categorize these alerts as either benign or suspicious based on the analysis they conduct using other simulated information sources such as network and system activity logs, a user database, a security news website, and a vulnerability database. Participants must use one or more of these additional data sources to accurately analyze each alert presented in the cybercog system. For example the participant must use the network/system logs to get more information on

the activity that raised the alert, must use the user data base to identify if the user reported to have been responsible for the activity is authorized and needs to use the vulnerability database to know if the activity that caused the alarm was due to a vulnerability exploit.

Figure 2 is a screen capture of the CyberCog system in which the alerts are presented to the participants. Simulated intrusion alerts used in the system are of 15 different types based on real world intrusion alert types such as alert for malware attack, suspicious email messages, buffer overflows and so forth. However, the alerts used in this system were simplified versions of their real world counterparts to make them understandable to the experimental participants who were not familiar with the domain or the task. Simplified does not imply that the alerts are easy to analyze, but simply means that they are presented in a form that is free from technical jargon. For this dissertation work, the scenarios and scenario data such as the attack descriptions were based on the scenario data suite in CyberCog.

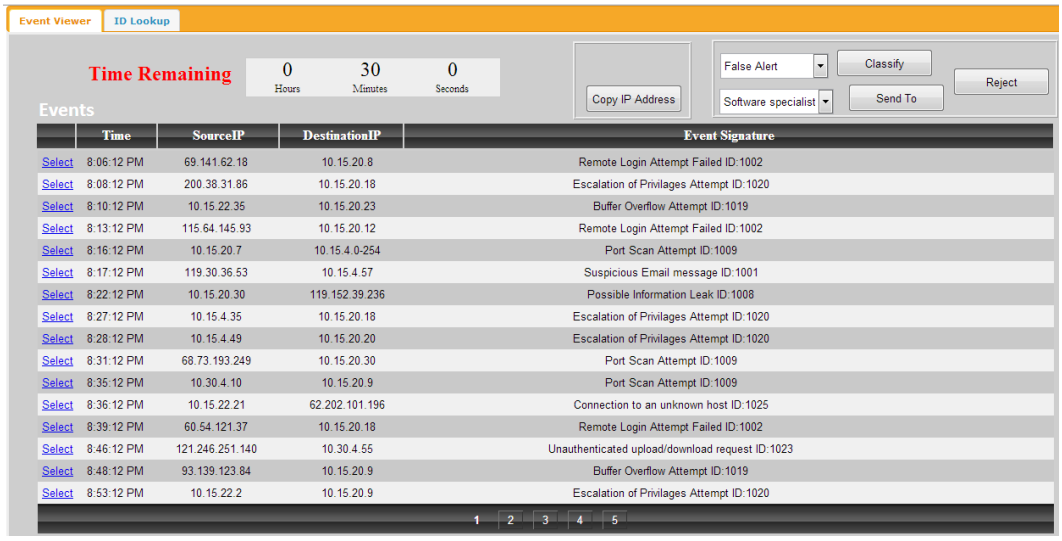

Figure 2. Screen capture of the web page presenting intrusion alerts in CyberCog.

Conducting human-in-the-loop studies on team cognition in the cyber defense context is a challenge because of the highly technical nature of the task. Finding participants with cyber defense skills and knowledge is a challenge when the task is recreated with all of its fidelities in the lab because access to cyber defense analysts is restricted. Recreating a simplified version of the cyber defense task such that participants with little to no cyber security knowledge can perform the task is a challenging process. Hence using agent-based models as a complimentary approach to human-in-the-loop experiments would make the experimental process more efficient because it will help extend the lab-based experiments to large sized teams and systems, to study the effect observed on teams on a longer duration and also allows the investigation of more hypotheses in a quicker and inexpensive manner.

Agent-based models representing analyst collaboration will be developed. The models will enable the study of macro level emergence from micro level interaction between the agents. The rules of interactions and behavioral characteristics will be modeled based on the findings from literature review, surveys, interviews and field observations. Agent-based models can be executed with different research questions based on hypotheses and gaps identified from conducting human-in-the-loop experiments and will be used to investigate research questions that are beyond the capabilities of laboratory experimentation.

Models by themselves are not very insightful and therefore the data from the model has to be finally compared with data from the experiment to make inferences. All the methods described in this section were employed in this dissertation work which will be described in the following sections.

HUMAN-IN-THE-LOOP EXPERIMENT

I assume that motivating and rewarding analysts to work as a team ***alone*** does not ensure effective team work and information flow. Pooling from each individual of a team the unique knowledge they possess and their expertise in making decisions is the key necessity for teamwork. However, past literature shows us that teams by default are ineffective in pooling novel information. Teams are known to repeatedly discuss and pool information that is commonly known to a majority of the team members. They are known to be ineffective in using the unique knowledge available to each team member to make decisions. This sort of an effect is popularly known as information pooling bias or hidden profile paradigm (Stasser & Titus, 1985). This effect has been observed in a wide array of teams such as medical teams (Christensen & Abbott, 2003), military teams (Natter et al., 2009), intelligence analysis team (Straus et al., 2011) and jury teams (Hastie, Penrod, & Pennington, 2013).

Premise 1

Cyber defense analysts would likely benefit from pooling novel information and knowledge available with their team members in detecting attacks. Other analyst members would have information that would confirm or reject one's initial inferences and hypotheses. Other analysts would have knowledge and expertise relevant to analyzing a certain kind of attack one is monitoring or they could have information that helps to make association between disparate observations. Other analysts might even have information that reveals an incident previously deemed to be an isolated event as an important event which needs immediate attention and response. However if they do not

share such novel information, the effort of the analysts to work as a team may not pay off in terms of improved performance.

Research Question 1

Does information pooling bias affect cyber defense analyst team discussions and decisions?

Hypothesis 1

I hypothesize that, cyber defense analysts, conducting the detection task during forensics, pool information in a biased manner during team discussions causing them to make sub-optimal decisions.

Rationale. Each analyst in the team conducting the detection task during forensics would be working on non-overlapping parts of the system. There would be conspicuous incidents such as denial of service and regular malware attacks occurring across the different parts of the networks and the analysts would want to talk more about these incidents during the discussion than the incidents that seem isolated because they have made some initial inferences about the conspicuous incidents and are looking for validation from other members. They would not discuss the unique events because they would have been unable to fit those unique events with their other observations and also would be unable to fit them into their mental model of the current network situation. Therefore cyber defense analyst teams would be affected by the information pooling bias.

Premise 2

Currently, cyber defense analysts are either using off-the-shelf collaboration tools in their work or no collaboration tools at all. Off-the-shelf collaboration tools such as wiki applications and chat interfaces may facilitate collaboration, but the development of

collaboration tools specifically designed to address the unique human knowledge and system requirements of cyber defense is necessary to improve analyst performance. The individual analyst or the team of analysts have to construct new knowledge about the emerging attacks out of massive amounts of information, but humans have mental limitations that strain this process and hence require carefully designed tools that would enhance the ability of the groups to construct, organize and share knowledge (Stahl, 2006).

Research Question 2:

Does a tailor made collaboration tool lead to superior analyst performance compared to using off-the-shelf collaboration tool such as wiki software?

Hypothesis 2:

I hypothesize that, tailored collaboration tools developed by considering the cyber defense analysts' cognitive requirements will lead to higher detection performance in analysts.

Rationale: When collaboration tools are developed by taking into consideration human strengths and limitations, then intuitively the human performance will be elevated. However, the extent of performance improvement depends on the level of thought and detail put into understanding the nuances of human behavior and cognition in a particular context such as the cyber defense. In this case, the higher level limitation is that cyber defense analysts do not collaborate which can be simply be solved by deploying generic collaboration tools such as the wiki or chat interfaces. However, the degree to which those tools help cyber defense analysts to effectively collaborate is often overlooked or is considered an afterthought. Rather than deploying generic tools and investing futile

efforts to customize those for collaboration in cyber defense, tools can be developed from the ground up. New collaboration tools for cyber defense can be developed through detailed thought beforehand about the cyber defense operator's cognitive requirements and limitations in addition to the system requirements to effectively improve team performance.

Wiki as a collaboration tool for forensics

Wikis are a type of online collaboration tool that enables sharing of data and allows collaboratively editing. They are already being used as a collaboration tool by cyber defense teams in some organizations. The analysts use wikis to write their individual reports and to archive them. Wikis are also used to share the individual reports with the rest of the team members. The members on the team can search and retrieve other members' reports. Similarly, in this study, wiki was used in one of the experimental conditions to present reports to the participants. The participants can look up others' reports and share one's own reports with the others. Wiki represents off-the-shelf tools that would be used for collaboration. A standard wiki application called *DokuWiki* meant for small scale companies was employed in the experiment.

Collaboration Visualization Tool for Forensics

The custom collaborative visualization tool was aimed at addressing the information pooling bias found in similar teams such as the medical teams, intelligence analysis teams, and presumably cyber defense teams. In the past, most tools proposed to address this specific team issue have not been based on the cognitive underpinnings of this bias, but rather have focused on trying to motivate the team members to spell out all

of the information. Other types of tools such as group decision making software that evaluate the decisions after they are made are available, but they do not address the problem upfront when the decisions are made. Also the solutions for each type of team have to take into consideration the specific needs of the particular domain and context. One solution will not fit all team types. Therefore, a tailor-made collaborative visualization tool is needed to truly improve collaboration in cyber defense teams. The screenshot of the prototype visualization is shown in Figure 3

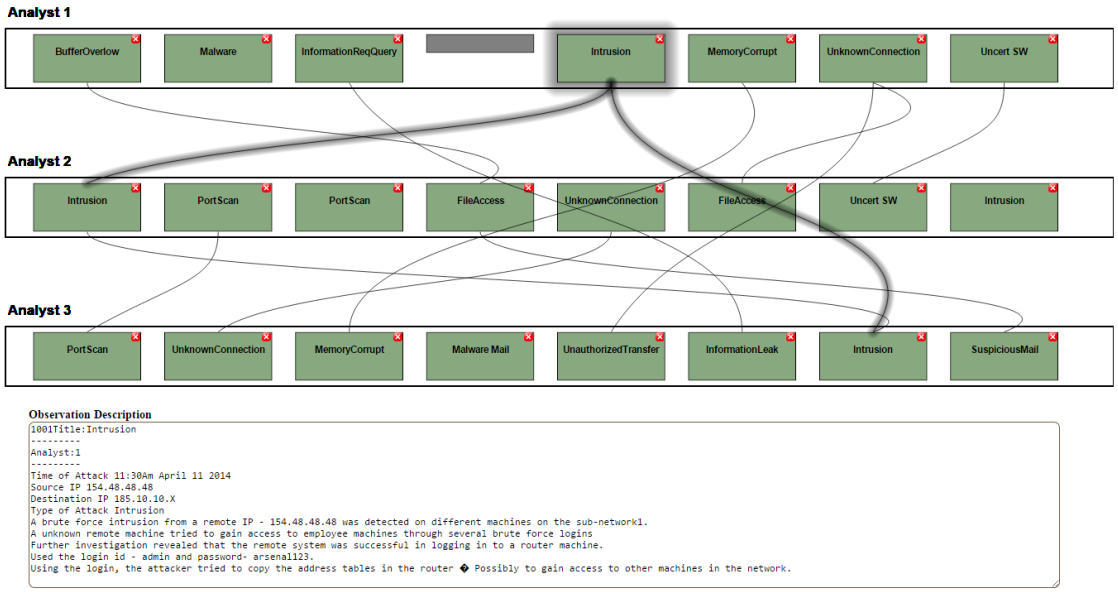

Figure 3. Screenshot of collaboration tool presenting all the three aspects of the tool.

The collaborative visualization leverages the individual text reports of attack descriptions available to each team member and finds possible connections between them based on certain attributes. The association could be based on the type of the attack, source IP address, possible attack paths and vulnerabilities. For example one team member could be reporting on the same malware as seen by another team member and this would create an association between the observations in their reports. Also, for

example, one team member could be seeing an intrusion attempt on a system and another team member could be seeing a similar intrusion attempt on another system present in another part of the network. If the two systems are connected with each other in some way and if there is software vulnerability in one system that could cause an intrusion into another system then that would create an association between those observations in their reports. Finding such associations manually will be very hard because separately these would seem to be isolated intrusion attempts. For the sake of this experiment, the associations were manually assigned based on the scenario developed.

The individual observations and possible associations were represented using a card based system (Keel, 2004) in which the individual observations were represented in card like formats and the connections between the observations were represented using lines that connect them. This way of finding connections between individual observations available with all team members would overcome the cognitive limitation of humans in finding association and fusing information manually. If one does not see these connections they might resort to discussing the ones they know are relevant and would downplay the observations which do not seem to indicate a high priority threat because they were unable to associate it with other such observations available with other team members. A screenshot of this tool is presented in Figure 3.

When a team member is talking about an observation, then that member chooses that observation on his or her screen and that will be emphasized along with its associations. Now based on the amount of time spent on the observation it will be automatically deemphasized by greying out those boxes and other observations and

associations will be emphasized automatically to prime and promote the analyst to discuss all observations and its associations. This is shown in Figure 3.

Then the analyst can choose to hide away or *mute* some of the observations as a process of data reduction. They can retrieve it back to the system (*un-mute*) if they need to look over it again. However this form of data reduction can help analysts to again discuss all pieces of information equally and will allow analysts to use all pieces of information in making their decisions. The grey box in row 1 under analyst 1 in Figure 3 shows the muted/greyed out representation.

In summary, the prototype collaboration tool employs three features to tackle the information pooling bias in cyber defense analysts. The three features include: visual representation of different associations between individual reports represented as cards, emphasizes/deemphasizes observations based on the discussion focus and finally the ability to *mute* or put away certain cards or observations. Next rationale for employing the three features and the cognitive limitations they address will be discussed.

The majority of work on information pooling bias or the hidden profile effect has focused on ways to get the team members to discuss the novel information available to each of them. The focus has been more on the social and communication aspect of the teams and very little on the cognitive aspects. Examination of the cognitive underpinnings of this bias, may suggest additional ways to mitigate it. Humans tend to easily communicate knowledge or information for which they have developed a good mental model and are restrictive in communicating when they are unsure or do not have a vivid mental connection for a piece of information or knowledge. When information and knowledge are spread across all team members, elaborate searching to find connections

would be necessary and it would take a lot of team effort and individual mental effort to incorporate all pieces of information to make a good mental model and to make effective decisions.

Software to generate possible connections between team member's observations leveraging individual reports is technically feasible and showing such associations between the team member observations would help the individual analysts to make the mental connections effortlessly in contrast to making such connections by oneself. This can lead to a more constructive discussion incorporating the novel pieces of information such as unique events in addition to discussing information that is known to all members of the team.

Presenting the possible associations between team member's individual observations may not ensure that the analysts would discuss all pieces of information equally. Humans tend to process information present in their field of vision that only affects or are related to their current train of thought. So they might suffer from a tunneled focus and spend a lot of time discussing some observations and not give priority to others and therefore causing the analysts to still conduct a biased discussion even though they have the ability to see other observations and their associations.

After having discussed certain observations and their connections, analysts as a team would be inching towards a satisficing position in which they would be mentally overloaded and would be reluctant to work on other observations and associations thoroughly. The presence of the discussed observations in their visual field would bias them to further discuss those instead of focusing on and discussing a new set of observations and associations. When there are two choices with one being the harder

option and the other being the easier option we would choose the easier option even if the harder option may be the better choice. Here the easier and lower information yielding choice would be the observations and associations that have been discussed enough and the harder and greater information yielding choice would the observations and associations that have not received much focus. The bias to choose the easier low information yielding choice instead of the harder high information yielding choice would persist unless the easier choice is removed even as a choice. Cutting off the easier choice will allow one to start working on options that yield more information.

When observations can be hidden away from the analyst's visual field it is a form of arriving at a state of *closure/completion* with those observations and that they would stop including them in their subsequent discussions. Hence this would allow the analyst teams to work on new observations without being biased about the ones already discussed. Allowing analysts to hide or box away the discussed observations would help them to focus on and discuss the other observations and associations which would lead to more effective decision making.

Experiment Description

A human-in-the-loop experiment was conducted to investigate research questions 1 and 2. In this study, the hypothesis that information pooling bias is present in cyber defense analyst teams conducting the forensics task was tested and along with the ability of a prototype collaboration tool to mitigate cognitive limitations. The key component of the experiment was the discussion that took place between the participants in each Mission or trial. There were two discussion session trials. At the start of each session the

participants were assigned reports of attack descriptions. They were asked to study the report individually for a short duration of 10 minutes. Then during the discussion the participants were asked to share and discuss the attack descriptions available to them to get the big picture of the network situation at hand. They conducted the discussion either by using the report files provided to them in the form of Microsoft PowerPoint or by using an off-the-shelf collaboration tool (wiki) or by using the prototype collaborative visualization software depending on the experimental condition they were randomly assigned. Figure 4 is a pictorial representation of the experiment process.

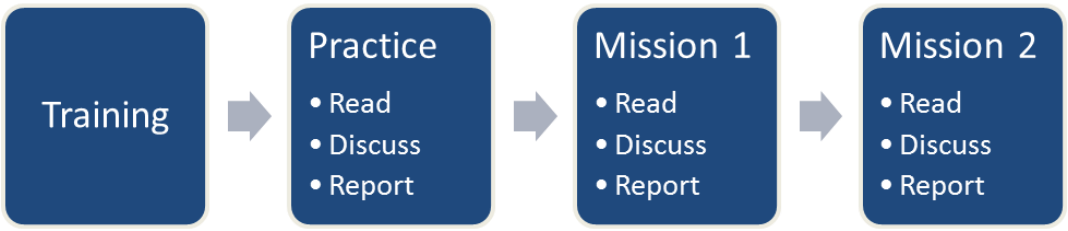

Figure 4. Summary of the experiment Process

Participants. Thirty teams comprised of three participants each were recruited to work as cyber defense analyst teams in the study for three hours. The three hours included the training period, breaks, practice session and the actual experiment trials. The participants were recruited through advertisements posted around the university campus and through an email list service of the university. The participants were asked to sign-up for a date based on their availability. The experimental condition to be run on each date was assigned randomly. The participants were given $10 per hour for their participation in the experiment. An informed consent form was presented to the participant and they were assigned to the experiment only if they provided their consent to participate in the

experiment. The participants were then provided the necessary training for performing the tasks in the experiment.

Materials. The training document used by the participants in presented in Appendix A and the attack reports used in both Mission 1 and Mission 2 is presented in Appendix B.

Training. Training is crucial because the cyber defense task and the terminologies are mostly unfamiliar to a majority of the population. This kind of training could reduce to an extent the individual differences between the participants in terms of knowledge and skills required to conduct the task. The four main learning objectives intended from the training include (1) Become familiar with computer networks and the associated terminologies (2) Develop an understanding of how an attacker/hacker can attack computer networks (3) Develop an understanding about the different cyber attacks used in the experiment (4) Learn how to discuss attacks with others on the team to get a big picture view of the network being analyzed.

First, the participants were given an overview of the cyber domain including an overview on computer networks, the Internet and its basic components such as IP address, software ports, and computer devices and how the communication flows between devices on the network. All descriptions were presented in a simple and jargon-less manner using examples such that the participants with little to no-training can quickly grasp and comprehend the material presented. They were also frequently quizzed to help them reflect the material they have studied. Similarly, the participants were also given a description of the network that they were going to analyze during the experiment trials and were also given description of how a cyber-attack is carried out, using graphical

examples. Then the participants were given descriptions of the cyber attacks used in the experiment task and also how to analyze those descriptions. Finally, the training involved information on how to conduct a discussion regarding individual observations with other team members and on how to make connections between different attack observations to detect threats such as multi-step attacks and APT. Such multi-step attacks were defined as large scale attacks in the training and the participant teams were instructed that their goal was discuss and detect attacks happening at a larger scale. The participants were given a 15-minute break after the training session.

Experimental Missions. After training and the subsequent break, the participants performed one short practice Mission for hands on experience at conducting a forensic discussion by reading the attack observations assigned to them and conducting a discussion on them later. After the practice Mission, the participants were shown an animated video with motivational background music describing their task and goals in the context of a military Mission. Such a back story was provided to get the teams to perform their tasks with some level of motivation. Later, the participants performed two trials of discussion based on the attack descriptions assigned to them.

During the discussion trials, each participant was assigned separate reports that contained a list of attack evidence descriptions. Each attack description contained the name of the attack, type of attack, time of attack, attack methodology, information/file involved in the attack and source and destination machine IP address. This simulates reports generated by cyber defense analysts in the real world. The reports assigned to each participant were different from each other, but each contained eight attack descriptions to analyze and discuss for experimental control.

The reports assigned to each participant were carefully constructed such that a majority of the attacks and the corresponding descriptions with each participant was associated with two or three fictional large scale attacks happening at a certain part of a fictional network. As per the scenario, there will be similar such fictional attacks happening at other parts of the network and the evidence for those attacks were assigned to other team members. Such attacks were termed *shared attacks* in the experiment. Additionally for each participant there were also clues of attacks that were disconnected from the rest of the attacks and seemed like isolated events happening in that part of the network. However such attacks were also constructed to be part of a large scale attack spanning different sub-networks and the clues about such attacks were spread across all three members of the team. Such attacks were termed u*nique attacks.* There were also alerts that were indeed isolated and had no connection whatsoever with other observations and alerts available with other team members and were simply termed as *isolated attacks*.

Each participant received four shared attacks, two unique attacks, and two isolated attacks totaling eight attack descriptions per team member. Figure 5 shows the distribution of the attack description per member analyst. As shown in Figure 5**,** there are five shared attacks of which all three participants each receive one copy of the two shared attacks and the remaining three shared attacks are shared by two of three team members. To clarify, as show in Figure 5 descriptions of shared attack 1 and 2 are shared among all three team members whereas description of shared attack 3 is shared between analyst 1 and 2, description of shared attack 4 is shared between analyst 2 and 3 and finally description of shared attack 5 is shared between analyst 3 and 1. Then there are two large

scale multi-step attacks which are represented in Figure 5 as unique A and unique B. There are three unique attacks that are part of each of these two large scale attacks represented as unique A1 to unique A3 and unique B1 to unique B3. These six unique attacks are equally distributed among all three team members such that each person has one part of the large scale attack needed to detect it. Finally, there are two isolated attacks per team member which are in no way connected to attacks in other participants. Each attack type is represented by a different shape in the Figure 5. In total there are five shared attacks, six unique attacks and six isolated attacks that are shared among the three team members.

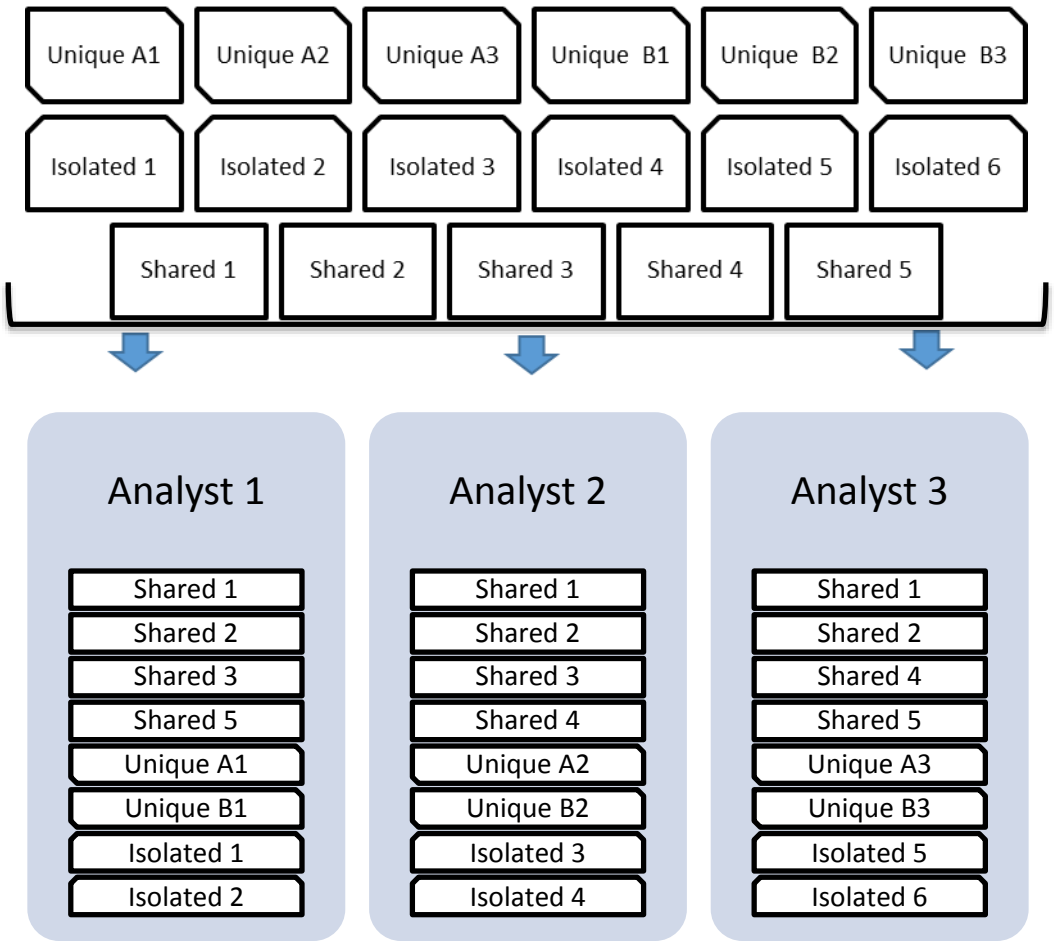

Figure 5. Pictorial representation of the attack distribution in the experiment

The participants had to discuss thoroughly, pooling all the information available to them, to find these connections and identifying large scale attacks whereas the evidence to detect other attacks were available to each of the team members and was conspicuous.

The participants were instructed that they would be reading descriptions of fictional network attacks observed. The participants were then asked to discuss, identify and report at the end all of the large scale attacks they detected through their discussion. They were also asked to ignore reporting the isolated attacks. They were then alerted to the fact that the reports were not identical and that there could be similarities and connections between their individual reports.

The training, attack reports, the tools, the measurements and the whole procedure was refined and practiced through several rounds of pilot testing. Emphasis was given to making the training material comprehensive, but at the same time concise and simple enough for all participants to understand and use the knowledge gained in conducting the task. Emphasis was also given to refining the attack descriptions. Attack descriptions had to be constructed to be at a good level of difficulty such that it was neither too easy nor too difficult to discuss and detect the large scale attacks in the Mission.

The aim of the experiment was to observe and measure whether the participants incorporate all of the information into their discussion and in making decisions and also whether they identify the large scale attack by pooling and fusing evidence that is spread across all the members of the team. They were also advised to take notes during the discussion to help them recall their findings to report at the end. They were given 25 minutes to discuss and at the end of the 25 minute duration the participants as a team

reported their findings. They were given 10 minute break and refreshments between the two trials.

Experiment Design. As shown in Table 1, the experiment was a 3X2 mixed factorial design. Type of tool was one of the independent variables with three levels. For each type of experimental condition, the participants performed two trials of discussion (a within subjects variable). All participant teams irrespective of the experimental condition conducted the discussion using Microsoft PowerPoint during the first trial. The data from the first trial served as the baseline measure of performance and baseline communication data. During Trial 2, the participant teams in the first experimental condition or control condition again used Microsoft PowerPoint for conducting the discussion whereas participant teams in the second experimental condition used *DocuWiki*, a wiki application, during the discussion and finally participant teams in the third experiment condition used the prototype collaborative visualization tool.

Table 1. Experiment Design of the experiment

	Mission1 (Baseline)	Mission2
Tool Type (Condition)	No Tool - Slide Based	No Tool - Slide Based
	No Tool - Slide Based	Wiki
	No Tool - Slide Based	Visualization

Measures

Performance. Team Detection Performance was based on the total number of attacks correctly identified and this was broken down by total number of shared attacks and total number of unique attacks detected. These numbers were based on the team report provided at the end. The report had a low chance of any confounds with memory

recall errors because during their report they used their notes from the discussion, had access to their attack descriptions and were not constrained by time.

Collaboration. To measure the team's focus of the discussion (Stasser & Stewart 1992), the team's communication during the discussion was coded in real-time by experimenters. Three experimenters were given a simple interface as shown in Figure 6 with buttons representing the eight attack descriptions per analyst. The coders were instructed to listen to the discussion and in real-time, based on the attack description being discussed, click on the respective buttons. Each click was recorded as one statement of the attack description in the discussion. The coders had around two to three weeks of practice doing this task while the experiment was pilot-tested. The practice included listening to the conversation and also clicking on the appropriate buttons.

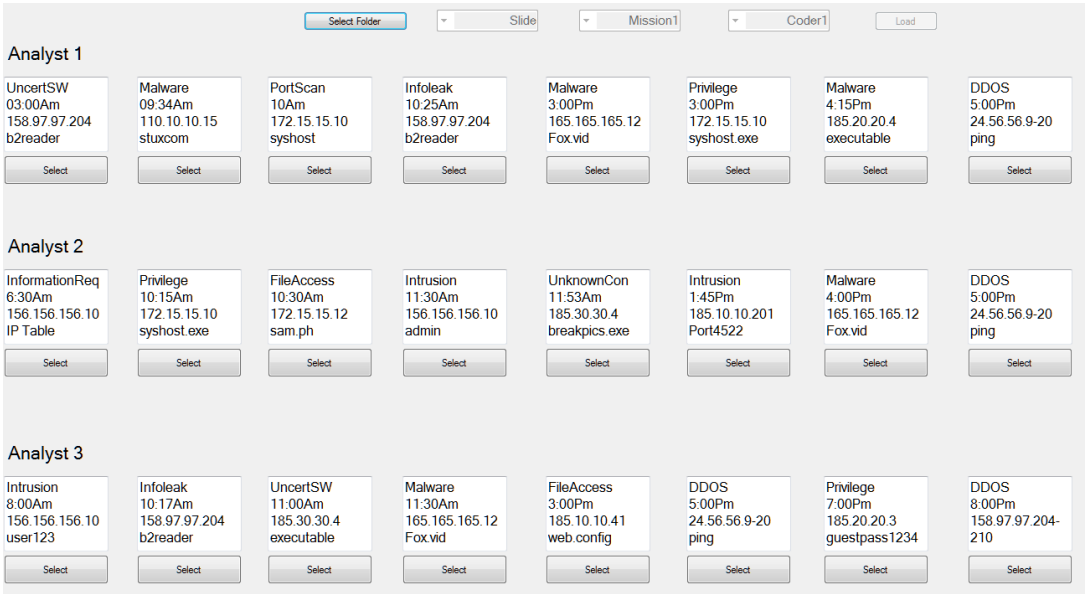

Figure 6. Screenshot of the communication coder interface

Such online coding methodology avoided the need to conduct any post-experiment communication coding, though the audio records were available if there were

any questions. Then communication coded identified the number of times the participants mentioned each attack description (including attacks unique to them). Then the number of times each attack type (shared, unique and isolated) was mentioned $X_{AttackType}$ was also computed. Also the total number of times all the attack descriptions were mentioned X_{Total} was calculated. All the measures were at the team level.

$$X_{Total} = \sum Mention\ of\ attack\ description$$

$$X_{AttackType} = \sum_{AttackType} Mention\ of\ attack\ description$$

Where AttackType = {Shared, Unique, Isolated}

Then a percentage *P* of each attack type (shared, unique and isolated) mentioned was calculated by taking the ratio against the total number of mentioning of all attack descriptions. $P_{AttackType} = \frac{X_{attacktype}}{X_{total}} \times 100$

This serves as a measure of the discussion focus for each kind of attack description. These measures were termed *shared percent, unique percent* and *isolated percent* for focus on shared type of attack, unique type of attacks (multi-step) and isolated type of attacks respectively.

Workload. In addition the NASA TLX workload questionnaire was administered after both trials and at the end of the experiment session to assess their perceived workload.

HUMAN-IN-THE-LOOP EXPERIMENT RESULTS

As described in the previous section, two main types of measures were collected from the experiment and analyzed. They include measures of performance and discussion focus. In addition, workload measures were collected once for each team at the end of the session and analyzed. The performance measure components include overall attack detection performance, performance in detecting attacks observed by two or more members of the team (termed as *shared attacks*), and performance in detecting attacks observed by only one of the team members, but which is associated with others attacks observed by other members of team because they are part of large scale attack (termed as *unique attacks*). The discussion focus measures included the percentage of discussion that involved discussing information shared by two or more members of the team (*shared percent*) and the percentage of discussion that involved discussing information that is only uniquely available to individual members of the team (unique percent). Workload measures included participant judgments of mental load, physical load and temporal load.

The analysis of the experimental results have been organized into four main sections. First, the descriptive statistics for the two main measures (performance and discussion focus) for each Mission are presented and across Mission and conditions are presented. Then a MANOVA was conducted for each mission to detect how the measured varied in each mission. Analysis was done in this manner because in Mission 1 no interventions were employed whereas in Mission 2 interventions were employed. Then relevant comparisons as required were conducted. Finally, analysis on the workload measures is presented.

Table 2 presents the descriptives of the discussion focus measures shared percent, unique percent and overall detection performance in Mission 1 and Mission 2 by combining the data from all three conditions (in other words data from all teams irrespective of the experimental condition) in each Mission. Skewness and kurtosis measures, as shown in Table 2, are calculated in terms of z value. For both Missions and for all three measures it is between -1.96 and 1.96 and therefore it can be inferred that the distribution of the data is normal and does not violate assumptions of normality.

Table 2. Descriptives of discussion focus and overall performance measure with Z values of skewness and Kurtosis

	Mission (All Teams)	N	Mean (%)	SD	Skewness (Z)	Kurtosis (Z)
Shared Percent	Mission 1	30	63.1	7.09	-0.34	0.99
	Mission 2	30	60.2	10.8	-1.46	1.67
Unique Percent	Mission 1	30	16.3	5.8	1.6	0.76
	Mission 2	30	22.08	8.9	1.7	-0.01
Overall Detection Performance	Mission 1	30	11.3	2.05	-0.05	-0.84
	Mission 2	30	12.44	2.7	0.59	-0.44

In Mission 1 all teams in all three conditions used only Microsoft PowerPoint slides during their discussions. However, in Mission 2, based on the experimental condition, teams in different conditions used different tools during their discussion where teams in the slide condition used PowerPoint slides, teams in Wiki condition used a wiki application and teams in the visualization condition used the visualization.

From the Table 2, it can be seen that mean percentage of shared information in Mission1 (combining data from all teams) is 63.1% whereas in Mission 2 it is 60.2%. Next, as it can be seen from the Table 2, mean percentage of unique information in the Mission1 is 16.3% whereas in Mission 2 it is 22.08%. Similarly it can also be seen that

detection performance in Mission 1 is 11.3 attacks out of 18 possible attacks whereas in Mission 2 it is 12.44 attacks.

Table 3. Descriptives of the discussion focus and overall performance measures in Mission 1

		N	Mean	Median	Standard Deviation
Shared percent	Slide condition	10	60.5	61.2	5.03
	Wiki condition	10	64.5	65	4.2
	Visual condition	10	64.3	65.2	10.6
Unique percent	Slide condition	10	17.1	16.6	5.8
	Wiki condition	10	15.2	16.7	5.4
	Visual condition	10	16.6	15.3	6.56
Detection performance	Slide condition	10	10.7	10.5	1.56
	Wiki condition	10	11.9	12	2.28
	Visual condition	10	11.5	12	2.27

Table 3 presents the descriptive statistics for the discussion focus measures: shared percent and unique percent, and the overall detection performance in Mission 1 by the three conditions. Mission1 was designed to be the baseline condition for detecting the presence of information pooling bias in cyber defense analyst teams. The descriptives presented in Table 3 show that the mean of all measures in all teams across all three conditions is very similar. These results show that participant teams while performing the cyber-attack detection and forensics analysis focused majorly on discussing shared information (around 60%) compared to the unique information (around 15%). The remainder of their discussion was focused on the noise data.

Table 4. Correlation between discussion focus measures and overall performance in Mission 1

		Shared_percent	Unique_percent	Detection_Perf
Shared_percent	Pearson Correlation	1	-.719**	-.417*
	Sig. (2-tailed)		0	0.022
Unique_percent	Pearson Correlation	-.719**	1	.380*
	Sig. (2-tailed)	0		0.038
Detection_Perf	Pearson Correlation	-.417*	.380*	1
	Sig. (2-tailed)	0.022	0.038	

Table 4 shows the correlation between the discussion focus measures and the detection performance measure. As it can be seen from table 4, the shared percent measure was significantly negatively correlated with performance (r(88)=-0.417) indicating that higher the focus on discussing shared information lesser was the performance. The unique percent measure was positively correlated with the performance (r(88)=0.380) which indicates that higher the focus on discussing unique information higher was the performance.

Table 5. Descriptives of the discussion focus and overall performance measures in Mission 2

		N	Mean	Median	Standard Deviation
Shared percent	Slide condition	10	63.14	61.39	5.4
	Wiki condition	10	67.2	66.09	8.02
	Visual condition	10	50.29	50.17	10.46
Unique percent	Slide condition	10	18.51	18.41	5.2
	Wiki condition	10	17.4	18.03	6.09
	Visual condition	10	30.14	31.92	9.2
Detection performance	Slide condition	10	11.4	12	2.5
	Wiki condition	10	11.8	12	1.3
	Visual condition	10	14.2	15	3.1

Participants in Mission 2 used different tools in each condition during their discussion. In the slide condition, participants used Microsoft power point slide, in wiki condition, participants used the wiki software and in the visualization condition, participants used a custom developed visualization during their discussion. Table 5 presents the descriptive statistics for the discussion focus measures: shared percent and unique percent, and the overall detection performance in Mission 2 by the three conditions.

Table 6. Correlation discussion focus measures and overall performance in Mission 2

		Shared_percent	Unique_percent	Detection_Perf
Shared_percent	Pearson Correlation	1	-.854**	-.450*
	Sig. (2-tailed)		0	0.013
Unique_percent	Pearson Correlation	-.854**	1	.585**
	Sig. (2-tailed)	0		0.001
Detection_Perf	Pearson Correlation	-.450*	.585**	1
	Sig. (2-tailed)	0.013	0.001	

Table 6 shows the correlation between the discussion focus measures and the detection performance measure in Mission 2 and as can be seen in Table 6, the results obtained are on par with correlation results found in Mission 1 data wherein the shared percent measure is again significantly negatively correlated with performance (r(88)=-0.450) indicating that the higher focus on discussing shared information lesser the performance. Also the unique percent measure is significantly positively correlated with the performance (r(88)=0.585) indicating that the higher the unique information discussed higher is the performance. As it can be seen, the correlation between the unique percent measure and overall performance is more strongly correlated in Mission 2 (r=0.585) in comparison to Mission 1 (r=0.38).

Thus far the descriptives were presented by each Mission but it is also important to look at how each measure fared across the two Missions and across the three conditions.

Towards that, first a mixed ANOVA was conducted on discussion focus measures: shared percent and unique percent and the performance measures to see the effect of the different interventions introduced in Mission 2 in comparison to Mission1 where all the teams used PowerPoint slides during their discussion. Therefore the within-subjects factor was the Mission and the between-subjects factor was the condition.

The mixed ANOVA on shared percent revealed that there was a significant interaction effect ($F=10.285$, $p<0.01$). This means that percentage of shared information in the discussion significantly varied between the Missions as a function of the condition. Figure 7 shows the comparison of mean shared percent measure across both Missions and three experimental conditions. As it can be seen in Figure 7, there is drop in shared percentage in Mission 2 in the visualization condition whereas there is an increase in shared information percentage in Mission 2 in the slide and wiki condition.

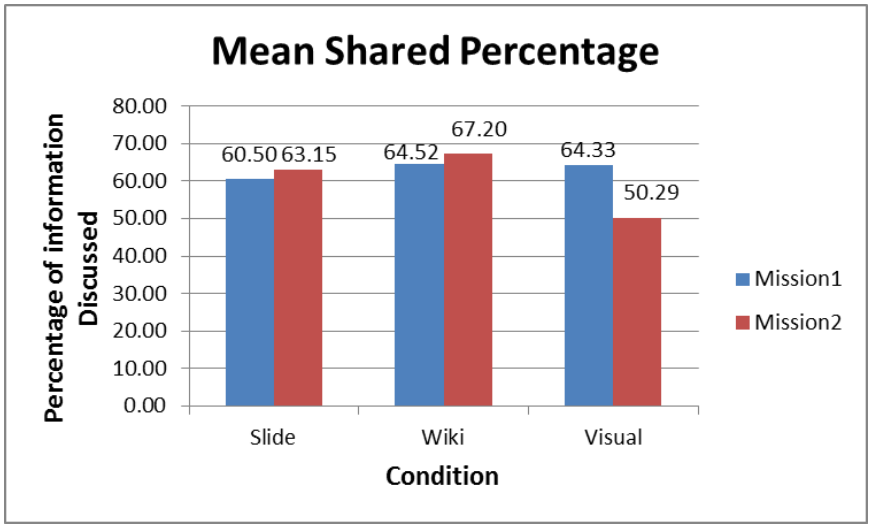

Figure 7. Bar graphs of shared percentage measure across both Missions and three conditions

Similarly, the mixed ANOVA on unique percent revealed that there was a significant interaction effect ($F=5.589$, $p<0.009$). This means that percentage of unique information in the discussion significantly varied between the Missions as a function of the condition. Figure 8 shows the comparison of unique percent measure across both Missions and three experimental condition. As it can be seen in Figure 8, there is an increase in focus on unique information in Mission 2 in all three conditions but the increase in visual condition seems greater.

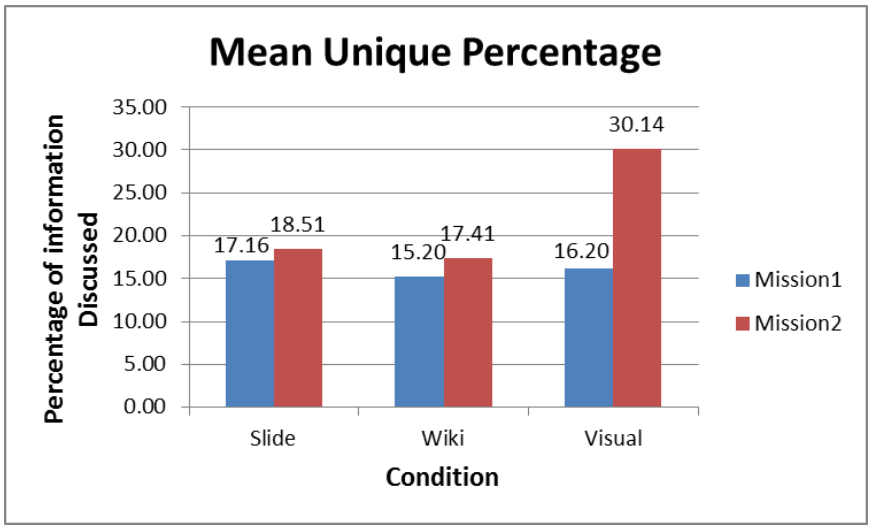

Figure 8. Bar graphs of unique percentage measure across both Missions and three conditions

The mixed ANOVA on overall detection performance revealed a non-significant interaction effect (F=3.136, p=0.060). Figure 9 shows the comparison of overall detection performance measure across both Missions and three experimental condition.

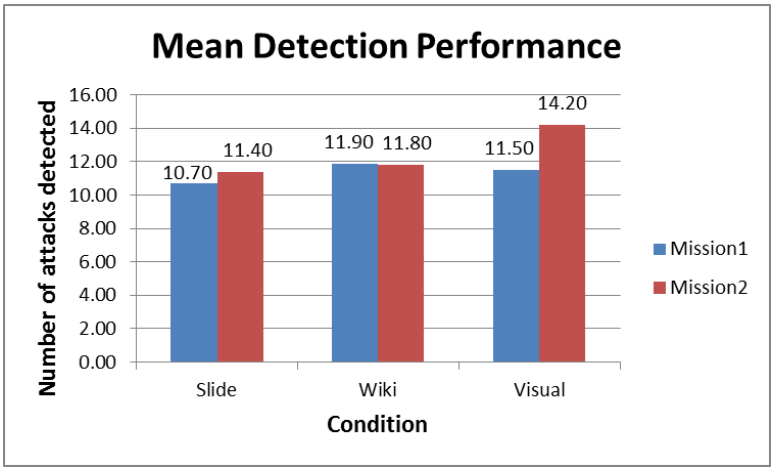

Figure 9. Graphs of detection performance across both Missions and three conditions

Since the interaction effect on the overall detection performance was non-significant, the overall detection performance was broken down to its constituents: performance from detecting shared attacks and performance from detecting unique attacks. The mixed ANOVA on detection performance of shared attacks revealed that

there was a non-significant effect interaction effect (F=0.480, p=0.960) whereas mixed ANOVA on detection performance of unique attacks revealed that there was a significant interaction effect (F=10.082, p=0.001). As shown in Figure 10, there is an increase in number of unique attacks detected in Mission 2 in the visualization condition and the slide condition. However there number of unique attacks detected in Mission 2 in the wiki condition decreases. Therefore, hereon, analysis will be done on both overall detection performance and its constituents.

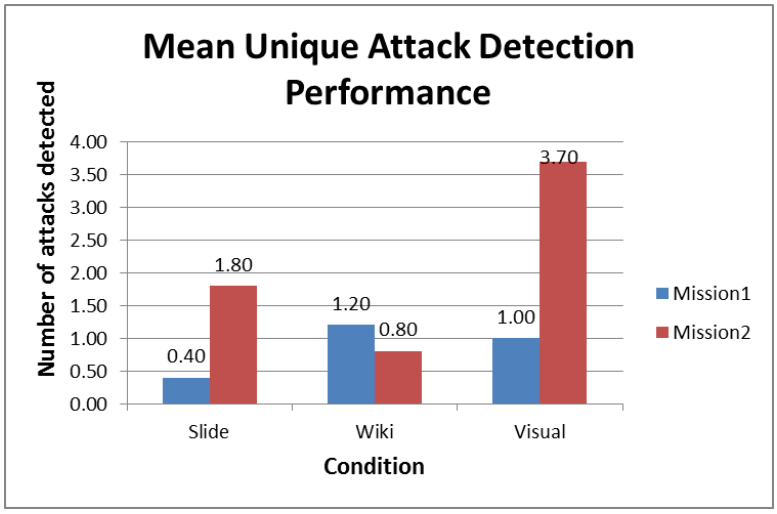

Figure 10. Graphs of performance on unique attacks across both Missions and conditions

Next a MANOVA was conducted for each mission to investigate the effect of condition on discussion focus measure and performance in each mission. A Multivariate ANOVA was conducted across the three conditions in Mission 1. The multivariate test (Hotelling's Trace) yielded a non-significant result: $F(8,46)=1.074$, $p=0.398$, partial $\eta^2 = 0.157$. This result indicates that the variables did not vary significantly across the three groups in Mission1 which is the desired outcome: no significant differences in team's discussion focus or in their performance and that all the teams in Mission 1 focused mainly on discussing shared information as opposed to the unique information.

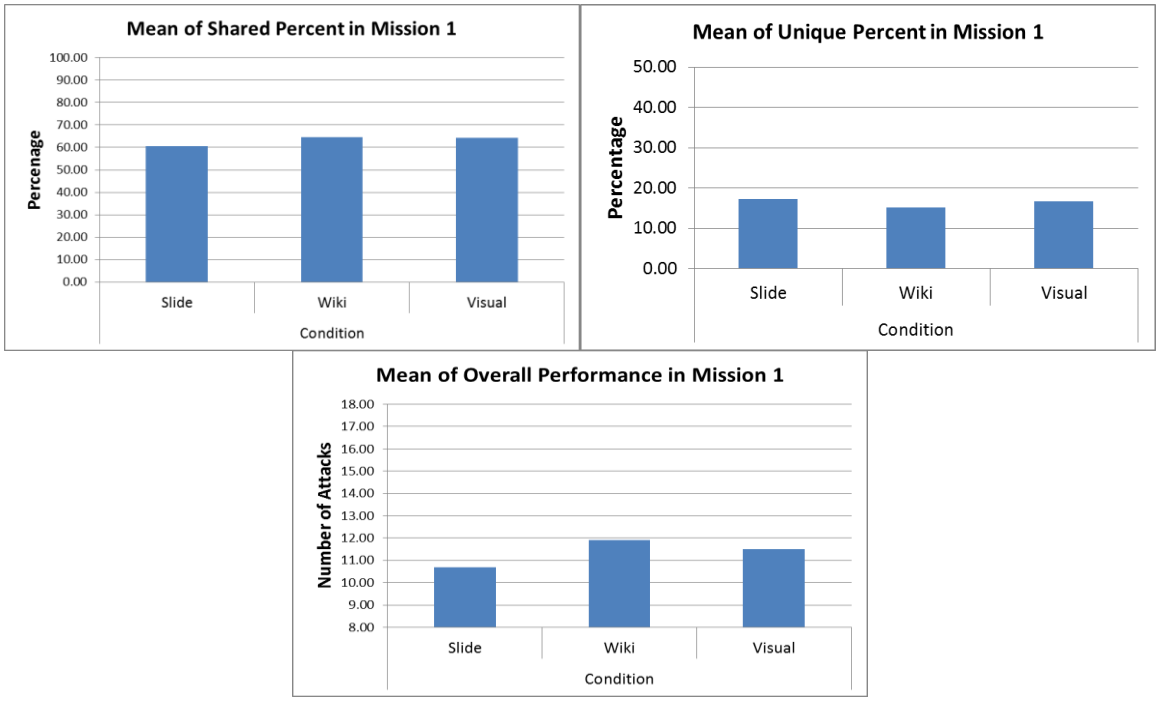

Figure 12. Bar graphs of discussion focus and performance measures in Mission1

As shown in figure 12, in all three conditions the team spent the majority (around 60 %) of their focus discussing shared information whereas only spent around 15% to 17% of their focus discussing the unique information. This bias is reflected in the performance. As shown in Figure 10 and Figure 11, most of the performance outcome is in detecting the shared attacks, whereas they detected only few unique attacks.

Then a Multivariate ANOVA was conducted across the three groups in Mission 2. The multivariate test (Hotelling's Trace) yielded a significant result: $F(8,46)=3.341$ $p=0.004$, partial $\eta^2 = 0.368$. This result indicates that there is a significant difference in the team's discussion focus and in their performance in Mission 2. As it can be seen in Table 7 and Figure 13, in the slide and wiki conditions, around 65% of the team's discussion focus was on shared information whereas they only 18% of the discussion focus was on discussing the unique information.

Table 7. Descriptives of discussion focus and performance measures in Mission2 by the three conditions

		Mean	Std. Deviation	N
Shared Percent	Slide	63.15	5.41	10
	Wiki	67.20	8.03	10
	Visual	50.29	10.47	10
Unique Percent	Slide	18.52	5.23	10
	Wiki	17.41	6.09	10
	Visual	30.15	9.24	10
Detection Performance Total	Slide	11.40	2.59	10
	Wiki	11.80	1.40	10
	Visual	14.20	3.12	10
Detection Performance Shared Attacks	Slide	9.60	2.12	10
	Wiki	11.00	1.33	10
	Visual	10.50	1.72	10
Detection Performance Unique Attacks	Slide	1.80	1.40	10
	Wiki	0.80	1.03	10
	Visual	3.70	2.36	10

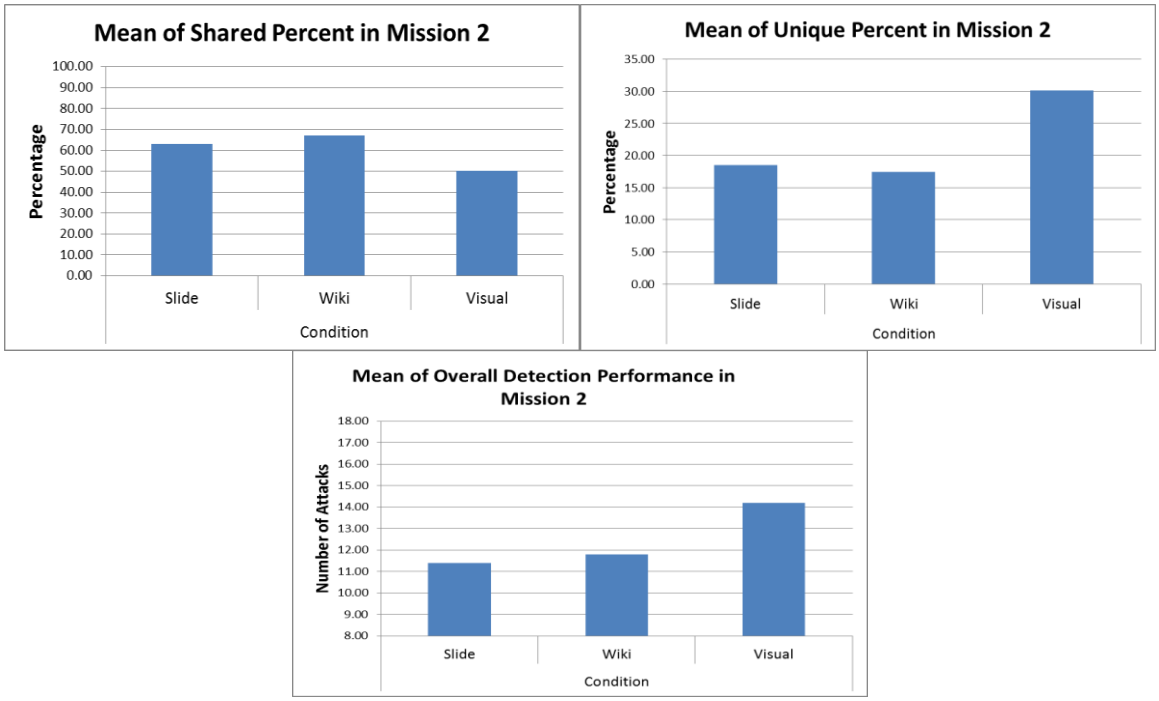

Figure 13. Bar graphs of discussion focus and performance measures in Mission 2

However it can be seen that teams in the visualization condition spent 50% of their focus discussing the shared information and spent around 30% of their focus discussing the unique information. As it can be seen from Table 7 and Figure 13 the bias is reflected in the performance in wiki and slide conditions where most of the performance outcome is in detecting the shared attacks (around 11.6 attacks); whereas they have detected only few unique attacks (around 1.3 attacks).

To further investigate the extent of difference detected in the variables in Mission 2, a test of between subjects effect on individual variables across the three conditions in Mission 2 was conducted.

Table 8. Results of the between-subjects analysis on discussion focus and performance measures in Mission 2 by the three conditions

	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared	Noncent. Parameter	Observed Power[f]
Shared Percent	1559.288	2	779.644	11.507	.000	.460	23.014	.987
Unique Percent	995.350	2	497.675	9.963	.001	.425	19.927	.973
Detection Performance Total	45.867	2	22.933	3.739	.037	.217	7.478	.634
Detection Performance Shared Attacks	10.067	2	5.033	1.639	.213	.108	3.279	.315
Detection Performance Unique Attacks	43.400	2	21.700	7.580	.002	.360	15.159	.919

As shown in Table 8, discussion focus measures: shared percent ($F=11.5$, $p<0.01$) and unique percent ($F=9.9$, $p=0.01$) varied significantly, overall detection performance varied significantly ($F=3.7$, $p=0.037$), detection performance of the shared attacks did not vary significantly ($F=1.639$, $p=0.213$) and detection performance of the unique attacks

varied significantly (F=7.5, p=0.002). This shows that the variance in the overall performance comes from the variance in detection of the unique attacks.

Now to investigate how the teams in the three conditions in Mission 2 fared between each other, a pairwise comparison of each variable (that was found to significantly vary) between all pairs of conditions: slides, wiki and visualization was conducted. As highlighted in Table 9 the variables: shared percent, unique percent and the detection performance on unique attacks in the visualization condition varied significantly from the slide and the wiki condition. However as it can be seen in Table 9, there is no significant difference between the conditions slide and wiki in Mission 2 for any of the variables.

Table 9. Multiple comparison on the measures in Mission 2

Measure	Conditions Compared		Mean Difference	Significance (p value)
Shared Percent	Slide	Wiki	-4.05	0.52
	Wiki	Visual	-16.91	<0.01
	Slide	Visual	-12.86	<0.01
Unique Percent	Slide	Wiki	1.1	1
	Wiki	Visual	-12.73	<0.01
	Slide	Visual	-11.62	<0.01
Overall Detection Performance	Slide	Wiki	-0.4	1
	Wiki	Visual	-2.4	0.09
	Slide	Visual	-2.8	0.04
Detection Performance Unique attacks	Slide	Wiki	1	0.39
	Wiki	Visual	-2.9	<0.01
	Slide	Visual	-1.9	0.04

The NASA TLX workload questionnaire was administered at the end of the experiment (after Mission 2). A multivariate analysis was performed on responses to the NASA TLX workload questionnaire. No significant difference ($F(26,148) = 1.009$, $p=0.461$) in workload perception was detected in participants across the three conditions.

This is a desired outcome as the introduction of the visualization did not increase their cognitive work load, nor did it seem to decrease it. All the measures except for physical stress and across all three conditions averaged around 7 with the maximum being 10 and minimum being 1. Hence it can be deduced that the participants perceived the task in general to be of high workload. The physical stress averaged around 3 which means that the participants perceived the task to less stressful physically.

Summary of Results

AGENT BASED MODEL

Premise 3

To represent such an information sharing paradigm computationally, both the team level social process that primes the analyst to search their memory for a piece of information to contribute to the current discussion and the cognitive process that the analysts use to look for that information piece must be considered. Because by default the members of a team would not discuss novel information enough, an assumption could be made that they could be using a heuristic search process internally with the goal of the search process being to communicate only information that validates or conforms to the current team discussion. Taking a deductive approach, such a search process can be represented in the model using meta-heuristics such as local search and hill climbing where the result of the search process in combination with the social process that encourage such a search process leads to sub-optimal solutions. Then such assumptions could be tested by running simulations and comparing them against the empirical data.

Research Questions 3

Can the cognitive process used by analysts to search and retrieve information that leads to a biased team discussion be represented in the agent-based model using heuristic search algorithms such as Hill climbing and local search?

Hypothesis 3

I hypothesize that, human analysts are using simple heuristics to search for information to contribute during a biased team discussion

Rationale. Heuristics are used instead of traditional algorithms when the optimal solution is computationally very expensive and that it is satisfactory to achieve sub-

optimal solutions. This could be the case with human teams when they discuss the shared information more often because communicating and finding connections between novel pieces of information is computationally expensive for humans and they are happy with achieving a sub-optimal solution using the shared information. Hence meta-heuristics such as hill climbing and local search algorithms can be used to represent the internal cognitive search process in combination with the external social stimuli motivating agents to search locally would represent such an information sharing paradigm.

Other memory based meta-heuristics such as the "*tabu search*" use memories of past searches to decide whether to search a certain search space again or move to different location in the search space to search for new optimums. This is similar to the expectation from teams conducting information pooling in which they search and communicate all pieces of information equally to achieve optimum solutions.

Research Question 3A

What type of model predicts the discussion pattern and performance of cyber defense analysts conducting a less biased team discussion? Is it memory-aided meta-heuristics?

Hypothesis 3A

I hypothesize that, memory-aided based meta-heuristics, will better represent the discussion pattern and performance of cyber defense analysts conducting a less biased team discussion

Rationale. When the agents can move around the search space so that they are not confined to local maxima it would mean they search other pieces of information to communicate as well, which is representative of the less biased team discussion. Hence

heuristics with memory aids to conduct such a search would represent the less biased team discussion.

Premise 3B

Agent-based modelling has traditionally been used to study several social processes with very little to no importance given to modeling the cognitive underpinnings of the agents. Through this work I am displaying the advantage of modelling the cognitive aspects of the agents in addition to the social processes to study the agent phenomenon in a more comprehensive manner so that it represents team cognition experiment conducted with human subjects in the lab.

Research Question 3B

Can agent-based models replicate the empirical results from experiments conducted with humans?

Hypothesis 3B

I hypothesize that, the results acquired by running an agent-based model that represents the key cognitive aspects of the human subjects will closely align with and predict the results obtained from the human-in-the-loop experiments.

To investigate the research 3 along with its sub-questions 3A, and 3B an agent-based model was developed.

Model Design

The model has four kinds of entities: Cyber defense analysts, their individual memory space, attack information and a social space. The individual memory space represents the human memory that has the pieces of information for agents to

communicate during a team discussion. The social space represents the information pooling happening at the team level during team discussions. Attack information represents a piece of attack description used by analysts in discussion and subsequently used to detect the attack itself. As shown in Figure 14, the cyber defense analysts and the attack information in the model are represented using agents. Both the memory space and the social space are represented using patches. Memory space and the social space are dimensionless and is simply representative of spaces in the head (ITH) and between the heads (BTH) respectively.

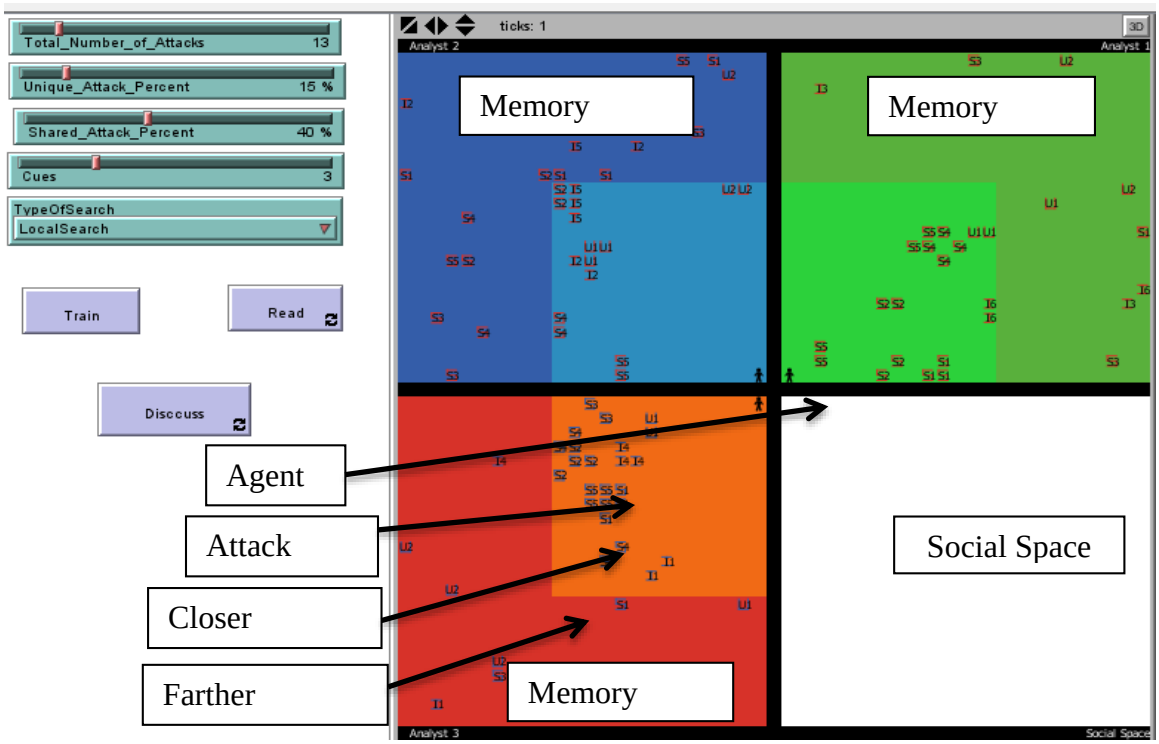

Figure 14. Snapshot of the model interface showing 3 colored quadrants representing individual memory space, the white colored quadrant representing the social space, agents and attack information

The whole rectangular patch space is divided into 4 quadrants. One quadrant of patches is assigned to each of the three cyber defense analysts as memory space and the remaining fourth quadrant is made the social space which is to capture the social interaction happening between the agents. As shown Figure 14, all three agents are put near the center of the entire rectangular patch space close to each other and at the border of their individual memory space. The four quadrants described are shown in the Figure 14 with each patch color representing the memory space of the three different analyst agents.

Input Parameter 1: Total number of attacks to be detected

Input Parameter 2: Number of Information pieces per attack

Input Parameter 3: Percentage of each attack type

Assumption 1: Each cyber-attack will have some number of information pieces and the human analyst needs to integrate these information pieces to comprehend the attack and to make good judgments about its connections to other attacks observed by others in the team.

Based on the parameters, total number of attacks and number of information pieces per attack, information pieces describing the attacks, are generated. Attack information is an agent breed that has state variables such as the attack type, attack number, attack information number, the percentage of information the attack information offers in detecting the overall attack, and a variable to track number of times it is mentioned. The number of attack information pieces per attack kind is based on attack type percentage parameter where the shared attack percent gives the percentage of shared

attacks with regard to the total number of attacks, and similarly the unique percent gives the percentage of unique attacks.

Assumption 2: Human memory is a space in the brain and this space has parts that are easily accessible to the individual and then there are parts that require more effort to access.

In the model, the analyst agent has two main state variables: patch boundaries of its memory space and capacity of its short-term memory. The locations of memory space are in discrete units of x and y coordinates of the Netlogo patches. There are in turn two parts to the memory space: (1) memory locations that are closer which are easily accessible and have information that have been consolidated by the agent. This is represented on the interface with light shaded patches (2) memory locations that are far, difficult to access, and that contain disparate information require several rounds of searching to reach. This is represented with darker shade patches. Both the closer and farther memory is shown in the Figure 14.

Input Parameter 4: Search Model – Random, Local or Memory-aided local search

Different search models (Random, Local or Memory-aided) are compared. Based on the chosen search model, the agent's search behavior to find information in its memory space varies. The total number of attacks detected and amount of mentioning of each kind of attack information during the discussion is compared between these different search models.

The random search model is used as null model for comparison. The remaining two models (Local and Memory-aided) are hypothesized to be representative of the two conditions (Biased and Less-Biased) tested in the experiment. As shown in Figure 15, the five cells (represented in yellow) in the experimental design for which the participants used slides or wiki software involved biased discussion whereas the one cell (represented in green) for which the participants used visualization involved less-biased discussion. The local search model is hypothesized (Hypothesis 3) to be representative of the five cells (in yellow) of biased condition in the experiment. The memory-aided local search model is hypothesized (Hypothesis 2) to be representative of the one cell of less-biased condition in the experiment (in green).

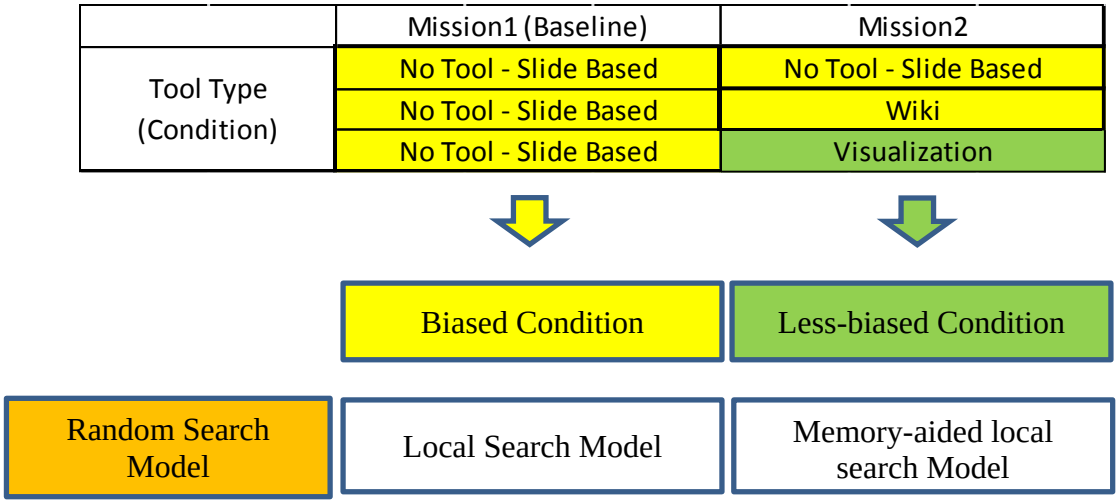

Figure 15. The three search models and the two mapped experiment conditions

The **goal** of the agents is to detect all the attacks. To detect an attack, the agents as a team have to integrate the attack information pieces belonging to an attack from each agent's memory space. To accomplish this, the agents must sense the ongoing team discussion, search for relevant information pieces in its own memory space and contribute the found information back to the social space (team discussion) where the

information pieces contributed are integrated and evaluated to know if the attack is detected.

The model is run for a maximum of 15000 ticks or until all the attacks are found. Every tick in the model represents one tenth of a second. So 10 ticks represents one second, 600 ticks represent 1 minute and 15000 ticks represents 25 minutes which is the duration of discussion in the complimentary human-in-the-loop experiment. Instead, considering one tick as one second and therefore 1500 ticks in total is a small duration for model execution. Therefore 15000 ticks was chosen

Process overview and scheduling

The overall simulation process involves three phases: Setup, Read and Discuss similar to the human-in-the-loop experiment.

The simulation begins with **setup** phase which involves the creation of agents, patches and also the distribution of information pieces for each kind of attack on the memory space of each cyber defense agent. The information at this stage is distributed to random locations on far side of the memory space (represented by dark shade patches on the interface). There are attack information pieces that are shared among all three agents and then there are information pieces that are uniquely available to each agent and finally there are information pieces which represent isolated attacks. Isolated attacks are simply noise and therefore no analysis was performed on that attack kind. The number of attack clues in the model is similar to that of the experiment. The shared clues represent the conspicuous attacks often seen in the networks, whereas the unique clues represent the large scale stealth attacks such as the APTs. Below is pseudo code of the setup process, highlighting the key steps involved in the process.

Setup

Create 3 Agents (Cyber Defense Analyst)
Create 3 Memory Spaces
$Number\ of\ X_{AttackType} = \frac{X_{Percent} * TotalAttacks}{100}$
where Attack Type = {Shared, Unique, Isolated}
Create attack information pieces for each attack type created
Distribute the attack information pieces to Each Memory Space

End

Assumption 3: When we read information, we consolidate the different parts of information and we store it in our memory in a way that is easy to accessible during future retrieval

After the setup phase, agents **read** the attack information pieces in their memory space as in the experiment. Each agent is assigned a short-term memory capacity based on Miller's work on short-term memory (Miller, 1956). Based on one's capacity, agents temporarily stores in their short-term memory the attack information pieces being read and when related information appears in the memory, the agent moves it to one of the closer memory space locations and also moves the associated evidence to one of its neighboring locations, thereby consolidating the attack information pieces in its memory space.

During discussions, team members have to use their recognition memory (Atkinson & Juola, 1974) to recognize the information being discussed and use that as a reference to find internally the relevant information to contribute to the discussion. However, when visual aids are used, they can augment this recognition memory and can help in locating the relevant information more quickly and easily.

To represent such an augmented human recognition memory in the model, each agent's memory space is mapped into four quadrants. This mapping will be used by the memory-aided local search model to quickly locate the region where certain attack information can be found instead of searching for it in a strictly uphill manner. Such an augmented recognition memory structure will be used by agents in the memory-aided local search model as it is hypothesized to represent the less-biased experimental condition in which visualization was used. Towards this, each agent creates a list of attack information pieces present in its memory space along with pointers to the mapped region in its memory space where it is present. As shown in Figure 16, each agent has a list of attack information pieces and each entry in the list has a pointer to a list which has the list of region identifier for the agent to lookup while searching for that attack information.

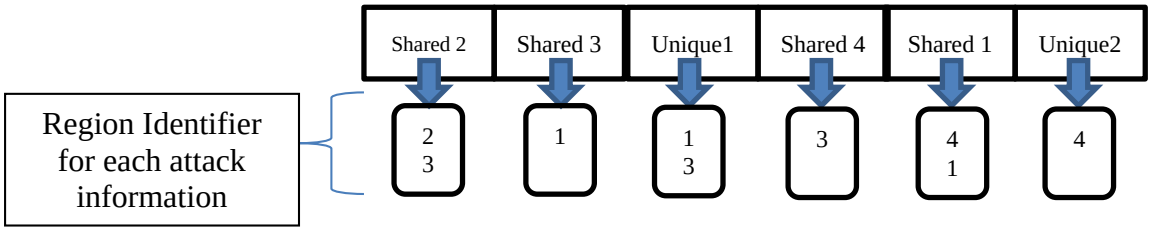

Figure 16. Representation of the recognition memory structure

Read
Each agent loops through each attack information in its Memory space
If the attack information matches another piece already in memory
Move both information pieces closer to each other and also closer to the center
End IF
-----Build the Recognition Memory----
Divide the Each Memory space further into 4 quadrants
Create a table of attack information pieces and its quadrant in Memory space
End

After reading, the discussion session begins. At the start of **discussion** phase, one analyst agent will be randomly picked to initiate discussion by triggering the agent to start the search process and to copy attack information it finds from its memory space to the social space. The search process employed by the analyst agent is based on the chosen search model: Random, Local Search or Memory-aided local search. The next agent in the order (agent 1 -> 2 -> 3 ->1) will be chosen to contribute towards the discussion during the next tick. The chosen agent in the next tick will again use the chosen search process to find attack information to add to the discussion. It would do two rounds of search to find relevant attack information to contribute to the social space. There are two outcomes to a search process: agent finds relevant information or agent does not find relevant information. If the agent fails to find relevant information it does not contribute anything to the space during the tick. On the other hand, if the agent finds relevant information then the agent copies it to the social space. It is then determined if the information pieces now present in the social space accounts to 90% of the information needed to detect the attack. If so, it is noted that the attack is detected and the social space is cleared, to make way for new discussions. Then the discussion will be handed over to the next randomly chosen agent. The agent in the next tick will look at the social space to see if there is an existing discussion and if there is a discussion ongoing in the social space the agent will perform the assigned search again to determine the attack information to contribute to the discussion. This continues until all the attacks are discovered or if the simulation completes 15000 ticks. The flowchart in Figure 17 captures this whole process.

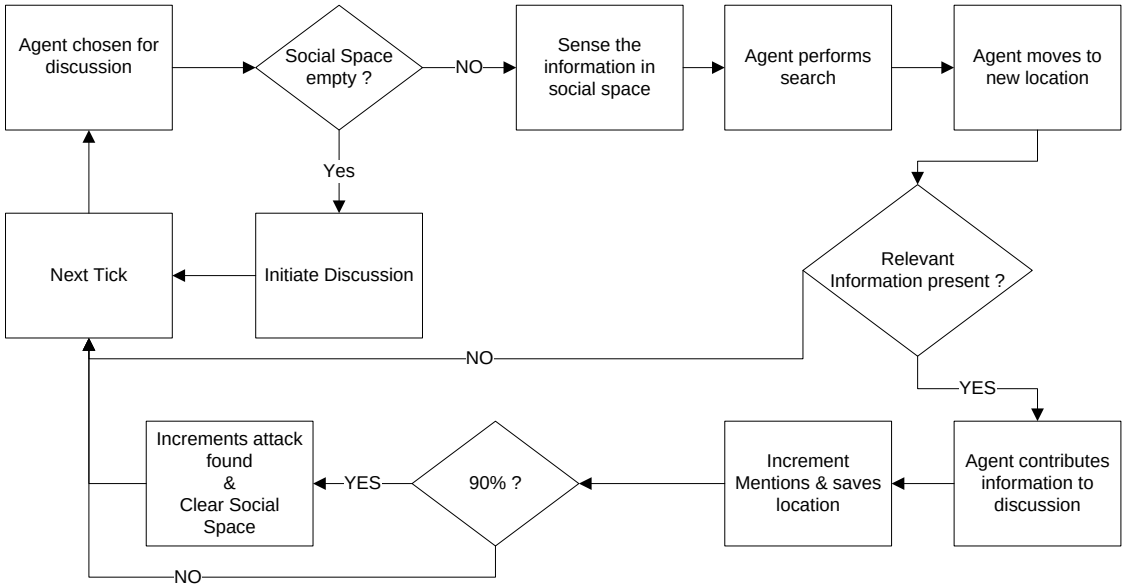

Figure 17. Flow chart of the discussion phase in the model

Go

Run until 15000 ticks (represents 25 minutes) or until all attacks are detected
If social Space empty
Randomly choose an agent
Ask the chosen agent to initiate discussion
If model = Memory-aided local search [choose pieces of attacks not detected]
Else [Choose one of the piece of attack in the closer memory]
Else
Choose the next Analyst
If model = Random Search [Perform Random Search]
If model = Local Search [Perform Local Search]
If model = Memory-aided local search [Perform Memory-aided local search]
Evaluate the social space
End IF

End

The three search models compared are random search, local search and memory search. In **random search model**, at each tick the agent moves to a random location in its memory space to find relevant information pieces to contribute to the current discussion. The agent attempts twice to find relevant information for the discussion.

```
Random Search
        Repeat twice
                Move to a random location in the Memory space
                If there is relevant information in the location
                        Contribute the information to the discussion
                        End Repeat
                End IF
        End Repeat
End
```

In **local search model,** at each tick the agent looks at its neighboring locations to find relevant information pieces to contribute to the current discussion. If there are no relevant information pieces in its immediate vicinity, it moves one step at a time in an uphill manner in the direction of the relevant information pieces. This uphill movement ensures that the agent will find some relevant information in the near future.

```
Local Search
        Repeat twice
            Identify neighboring information that was not mentioned in the last
instance
            If discussion relevant information was found
                Contribute the information to the discussion
                End Repeat
             Else
                Face towards the locations where relevant information is present in the
space
                Move 1 step in that direction (to avoid moving up hill in random
direction)
            End IF
        End Repeat
End
```

In the **memory-aided local search** model, the agents are given access to the recognition memory structure. To contribute to the discussion, the agent, using the recognition memory structure, moves to one of the regions containing the information.

The recognition memory structure is represented in Figure 16. After moving to the region, it does local search again to find the relevant information pieces for contributing to the discussion. Such a usage of recognition memory structure represents the stimulation of recognition memory by the visual aid. Also in this model, the agents have access to information on attacks that have been already discussed and therefore avoid initiating a discussion on attacks that has already been discussed. This represents a closure aid provided by the visualization tool in which the users are allowed to hide information pieces of the attacks already detected.

Memory-aided local search
Look up the recognition memory table
Find the sub-quadrant of the Memory space where relevant information is present
Move to one of the location in the sub-quadrant
Perform Local Search in the sub-quadrant
End

Each attack agent breed carries a variable for the percentage of information the attack information offers in detecting the attack. This variable's value ranges between 1 and 100. The value of this variable (Y_k) assigned to each information of an attack (A_i) is such that the sum of this variable across all the information pieces of an attack results in 100 (representing 100 percent of information). $\sum_{A_i} Y_k = 100$ where 'i' is the attack number and 'k' is the number of information under each attack. At the end of every successful search and after a new information is contributed to the discussion, the elements in the social space are **evaluated** to determine if the current set of attack information pieces account for 100% of the attack information. If 100% of attack information is accounted, it is determined that the team of agents has detected the attack and therefore clears the

social space to make way for new discussions. The current discussion continues if 100% of the attack information has not been contributed. In the case of shared attack kind, there are identical copies of the information pieces in all three memory spaces and hence all three agents have to contribute to the attack information to detect the shared attack. However all three agents do not have to contribute the 100% of information because once an agent contributes most of the information, the others agents simply have to confirm the presence of such information to detect the shared attack. Therefore instead of evaluating the summation to 300 (3 * 100), it is evaluated whether the contribution sums up to 270 and if so it is determined that the shared attack has been detected.

Evaluation

If 100 percent of attack is discussed
Increment Found Attack
End Discussion
End IF

End

Design concepts

Basic Principle. The model is based on the theory that teams by default tend to communicate information that is commonly known to all members of the team and fail to incorporate the novel pieces of information or the expert information in making team level discussion (Stasser & Titus 1985). Such an information pooling bias has been observed in many kinds of teams, but hasn't been investigated it in cyber defense teams where the type of information and the nature of information foraging and fusing are distinct from other types of teams. Past work on this bias has looked at the social aspects of the team in understanding the factors that cause the bias, but has not focused on the

cognitive underpinnings such as cognitive load, lack of perception about the information distribution and also the cognitive tunneling that leads to discussing and exploring information that is available with majority of the team members. Through this model we represent the cognitive search process and social priming that contribute to the bias.

Emergence. Agents are coded rules to look at the social space to search and contribute to the discussion. The local search process causes agents to look at common information initially and often thereafter because it is allocated closer to the agent and causes the agent to not perceive the novel information that is allocated far away in the memory space. This represents the internal process of team members demonstrating such a bias. The agents might reach the novel information, but this would require a lot of effort and is unlikely to happen often. Also, the agents might initiate a discussion with the novel information, but the other agents search location might be different and would be unable to reach to associated information using local search and even with few search rounds they would not have anything to contribute causing the agent to move on to other discussions. The restricted number of rounds of the search process represents the cognitive workload limitation. The whole simulation will be run for 15000 ticks and not for infinite time because the discussion in the real world has temporal constraints. So such simple rules will lead to an emergent phenomenon where the agents will more often discuss the common information and though they would reach the novel information it would not be discussed enough because at that point other agents would be looking at other information. Unless it leads to discussion, decisions will not be made using the novel information.

Adaptation. If the agent senses that there is an ongoing discussion in the social space then the agent will base its search process on the information present in the social space or the agent will initiate a discussion by using the information that is available close to its last search location. If there was no contribution to the discussion, the agent initiated then the agent will move to another space to start a new discussion using the information in its neighboring locations. To jump to a new space the agent would use random walk type search procedures.

Objectives. At each tick the objective of the agent is to find information to contribute to the ongoing discussion in the social space and also to detect the attack being discussed.

Learning. The agents at the beginning will read and consolidate the attack evidence.

Sensing. The agent senses the information present in the social space to conduct the internal local search process to find relevant information.

Interaction. The analyst agent hands off the discussion to the next agent in the order. The order is after analyst agent 1 it is handed off to Agent 2 who in turn hands it off to Agent 3 and comes back to Agent 1.

Stochasticity. There is randomness in the agent choice at the beginning of the simulation to initiate discussion. Then the choice of information by each agent to initiate a discussion and also to contribute to the discussion is also stochastic. In the random search model, the agent moves a random location to contribute to the discussion. In the local search model, the agent moves in an uphill style in the direction of one of the relevant information.

Observation

Measures similar to that collected from human-in-the-loop study were also collected from the model. The measures are based on the measures used in the hidden profile paradigm experiments.

To measure the agent's focus of the discussion (Lu, Yuan & McLeod 2012), the number of times the agent contributes each piece of attack information was recorded (X_i). Then a summation of X_i by each type of attack (shared, unique and isolated) was calculated giving the number of times a certain type of attack information was contributed to the discussion.

$$X_{attacktype} = \sum_{AttackType} X_i \qquad (1)$$

Where AttackType = {Shared, Unique, Isolated}

Then a summation of all X_i was calculated which gave the total number of times all the attack information pieces was mentioned in the discussion.

$$X_{Total} = \sum X_i \qquad (2)$$

Then using $X_{attacktype}$ and X_{Total}, percentage of shared information mentioned (aka *Shared Percent*) and percentage of unique information mentioned (aka *Unique Percent*) was calculated using the equation 3 and 4.

$$Shared\ Percent = \frac{X_{attacktype=Shared}*100}{X_{Total}} \qquad (3)$$

$$Unique\ Percent = \frac{X_{attacktype=Unique}*100}{X_{Total}} \qquad (4)$$

Finally, the total number of attacks detected by all the agents (aka *Detection Performance*) was also calculated. Because the measurements are conceptually in-line with the experiment they can be directly compared to validate the model.

SIMULATION RESULTS

Pretest

To determine the number of model repetitions i.e. the number of samples per model, a stability analysis on the model was conducted by running the model in increments of 100 repetitions for each model and plotting the cumulative averages of the dependent measures: Shared Percentage, Unique Percentage. As shown in Figure 18, the model generates stable results around 250 repetitions for shared percent, 700 repetitions for unique percent. Therefore the agent based model was run for 1000 repetitions.

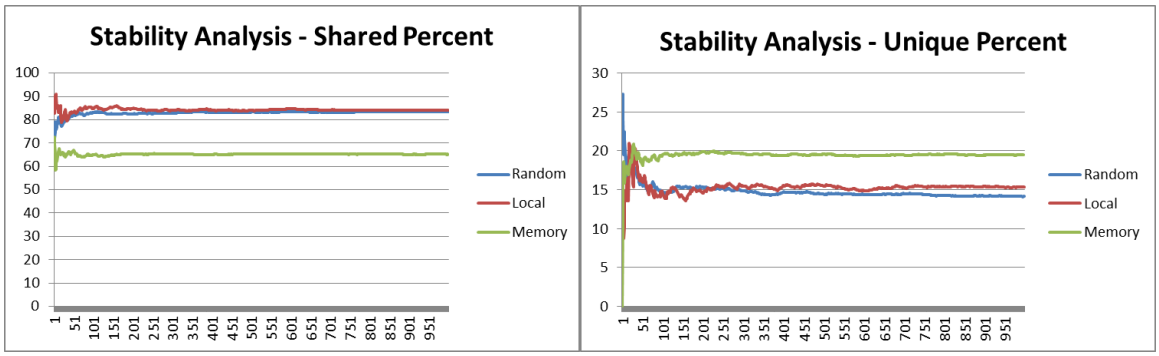

Figure 18. Graph of cumulative average of both shared percent measure (left) and unique percent measure (right) in all three models.

Experiment

Based on the findings from the pretest, the model was run for 1000 repetitions for the three search models: Random Search, Local Search and Memory-aided local search. Each repetition was run for 15000 ticks or rounds. Macro level performance measures such as Total number of attacks detected by all the agents (aka *Detection Performance*), percentage of shared information mentioned (aka *Shared Percent*) and percentage of

unique information mentioned (aka *Unique Percent*) were collected from each repetition. Table 10 gives the list of parameters and default values.

Table 10. Table of model parameter and values

Parameter	Default Value
Total Number of attacks	13
Percent of shared attacks	40%
Percent of unique attacks	15%
Number of information pieces	3
Type of Search	Random

Table 11. Descriptive statistics for each measure in all the three models

	Search Model	Mean	Median	SD
Shared Percent	Random	83.5	86.1	14.03
	Local Search	73.4	75.2	13.07
	Memory-aided Local Search	69.2	68.7	11.2
Unique Percent	Random	15.1	12.6	14.02
	Local Search	23	22.2	13.13
	Memory-aided Local Search	24.4	25.3	10.42
Detection Performance	Random	3.4	3	1.06
	Local Search	6.5	7	0.6
	Memory-aided Local Search	6.5	7	0.6

Table 11 shows the descriptive statistics for each measure based on the three models: Random Search, Local Search and Memory-aided local search. The distribution of all three measures from all three models is not-normal and is skewed. Therefore the median is a more appropriate measure than mean. The graph in Figure 19 is based on the median values for each measure for the three models. As shown in Figure 19, the medians are in the predicted direction wherein discussion of shared information in the "Memory-Aided Search" model is lesser than "Local Search" model and similarly the

discussion of unique information in the “Memory-Aided Search” model is greater than “Local Search” model.

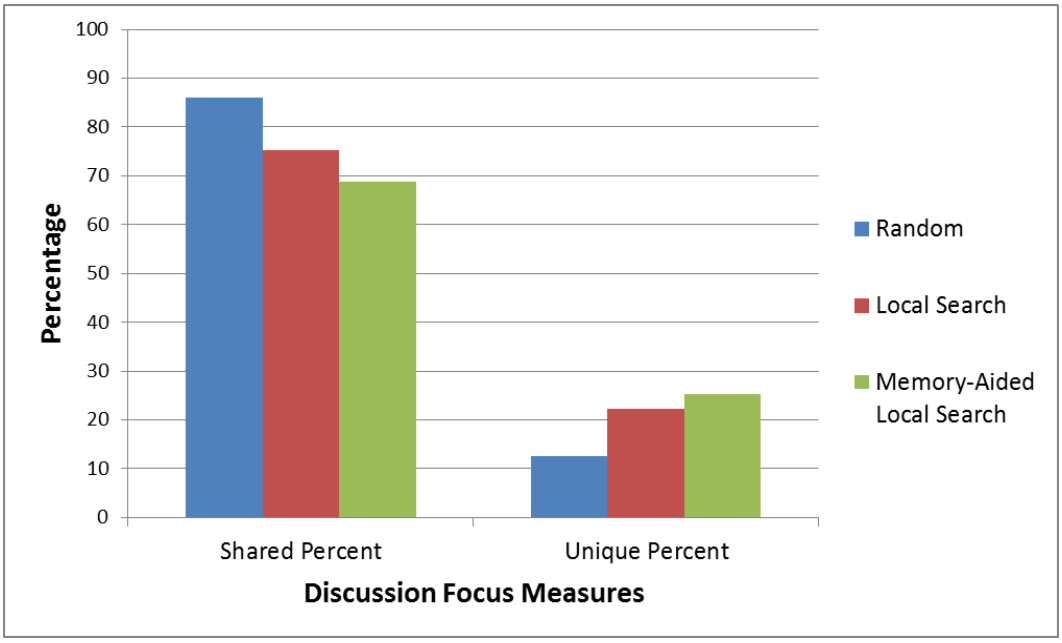

Figure 19: Bar chart of measures: Shared percent, Unique Percent and Detection performance in terms of median values across all three models

Because the distribution is not normal, a non-parametric analysis was performed on the measures from the model. Kruskal-Wallis, the non-parametric alternative to one-way ANOVA was employed on the measures. As shown in the Figure 20, the mean rank of *shared percent* in the memory-aided local search model is the lowest whereas the mean rank of *unique percent* in the memory-aided local search model is the highest and similarly the mean rank of the performance in both local search and memory-aided local search model is the highest.

Because the sample size is very large, effect sizes are a more accurate measure of comparison to investigate the significance of the difference between a pair of models. Grissom and Kim (2012) have provided an effect size estimator (like Cohen's d) for use

in association with nonparametric statistics which involves Mann-Whitney U statistic, Z-score and the sample sizes of the two models using formula $\rho_{a,b} = \frac{Z}{\sqrt{(n_a + n_b)}}$

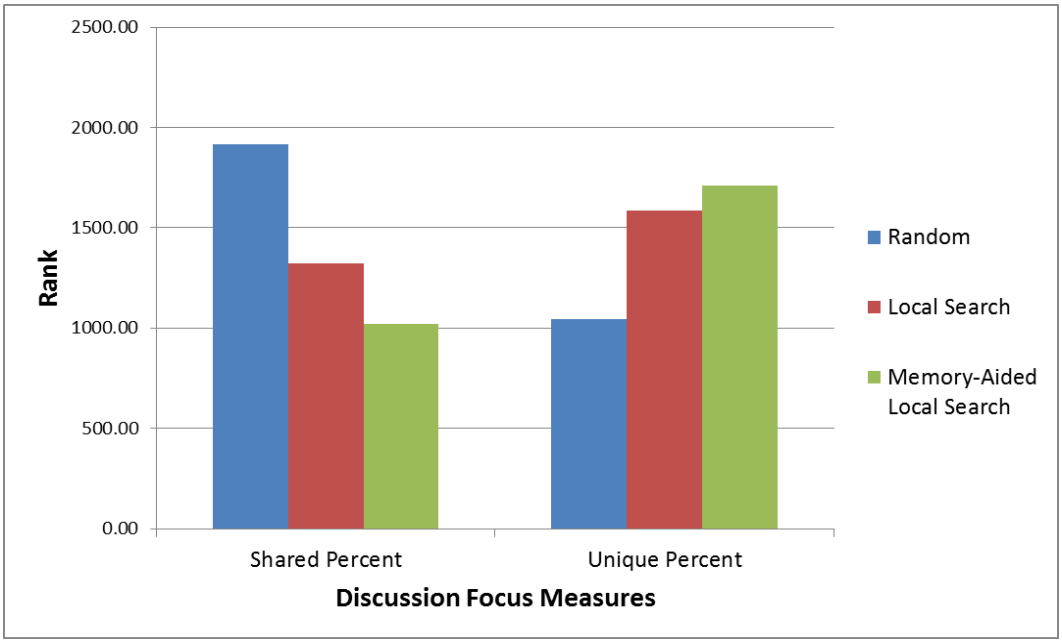

Figure 20. Bar chart of measures: Shared percent, Unique Percent and Detection performance in terms of mean rank across all three models

As it can be seen in Table 12, the effect size of shared percent between random search model and local search model is -0.38 which is a medium effect. The effect size of unique percent between random search model and local search model is -0.33 which is again a medium effect. The effect size of performance between random search model and local search model is -0.87 which is a high effect.

As it can be seen in Table 12, the effect size of shared percent between random search model and memory-aided local search model is -0.51 which is a medium effect. The effect size of unique percent between random search model and memory-aided local search model is -0.39 which is again a medium effect. The effect size of performance

between random search model and memory-aided local search model is -0.86 which is a high effect.

Table 12. Non-parametric effect size for each measure between pairs of models

Effect Size Comparison	Shared Percent	Unique Percent	Detection Performance
Random vs Local	-0.38	-0.33	-0.87
Random vs Memory-Aided	-0.51	-0.39	-0.86
Local vs Memory-Aided	-0.21	-0.08	-0.03

As it can be seen in Table 12, the effect size of shared percent between local search and memory-aided local search model is -0.20 which is a low effect. The effect size of unique percent between local search and memory-aided local search model is 0.084 which is a very low. The effect size of performance between local search and memory-aided local search model is 0.02 which is a very low effect.

Summary of Findings

From the effect size comparison, it can be determined that the random model is significantly different from the local search model and memory-aided local search model. The percent shared discussion from the local search model is significantly greater than percent shared discussion from the memory-aided local search model and the effect size statistic also reveals a medium effect. Though Mann-Whitney test revealed a significant difference in terms of unique percent discussed between local search model and memory-aided local search model, the effect size statistic reveals a very low effect size between them.

COMPARATIVE ANALYSIS: Experiment versus Model

Measures: shared percent and unique percent from the model have to be compared with similar measures obtained from the experiment to make an inference that the local search process model is indicative of biased team discussions observed in the experiment and that the recognition-based local search process model is indicative of the less-biased team discussions observed in the experiment.

To make such a comparison, the conditions tested in the experiment have to be mapped to two groups: Biased condition and less-biased condition. Towards that, as shown in Figure 21, all five cells involving slide and wiki software in the experiment (represented in yellow) had biased discussion and the one cell for which the participants used visualization in their discussion (represented in green) had less-biased model. Correspondingly, the local search model is hypothesized to be representative of the biased model and the memory-aided local search model is hypothesized to be representative of the less-biased model. The random search model is simply the null model.

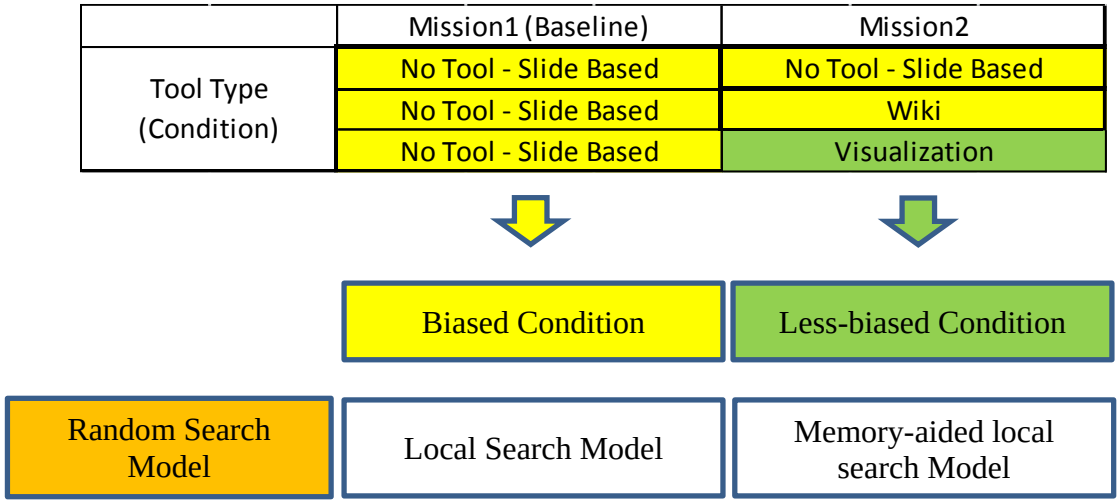

Figure 21. The three search models and the two mapped experiment conditions

First the distribution of the measures from two experimental conditions were compared against the measures from the three models, visually using graphs. Figure 22 shows the graphs of the data distribution of the **shared percent** measure from the random model (*represented in blue*) superimposed on shared percent measure from biased experimental condition (*in grey on left figure*) and less-biased experimental condition (*in grey on right figure*). As it can be seen clearly from Figure 22, data from random model do not represent the data from the biased or less-biased experimental condition.

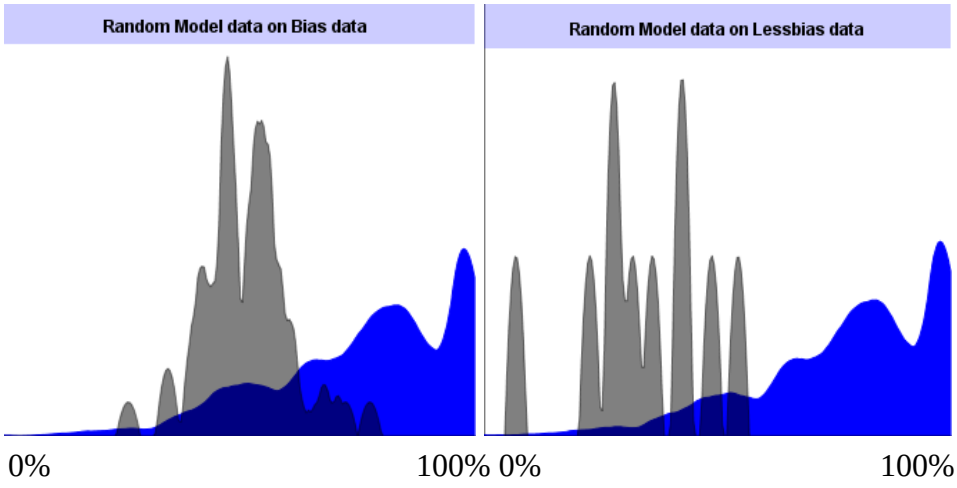

Figure 22. Shared percent data distribution from random model (*in blue*) superimposed on biased condition (*in grey on left figure*) and less-biased condition (*in grey on right figure*)

Figure 23 shows the graphs of the data distribution of **shared percent** measure from the local search model (*represented in blue*) superimposed on shared percent measure from biased experimental condition (*in grey on left figure*) and less-biased experimental condition (*in grey on right figure*). As it can be seen from Figure 23, some significant area of the data from local search model overlaps with the data from the biased experimental condition and does not overlap so much with the data from the less-biased experimental condition.

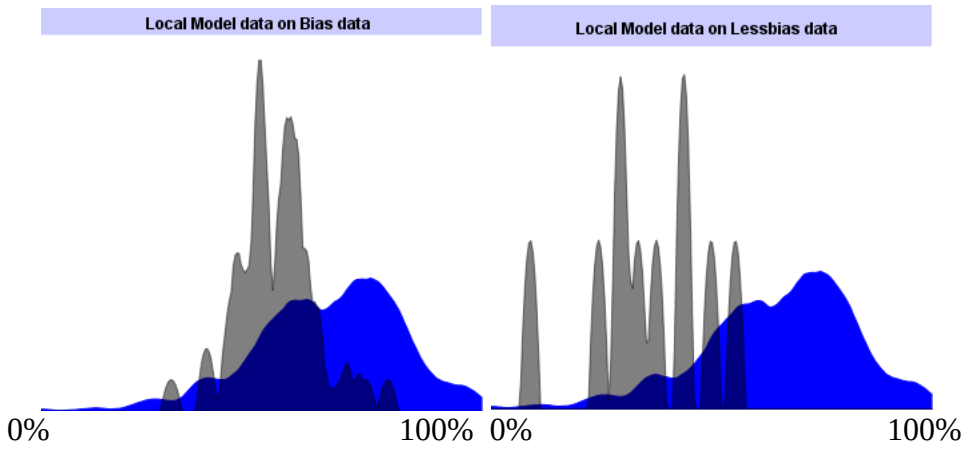

Figure 23. Shared percent data distribution from local search model (*in blue*) superimposed on biased condition (*in grey on left figure*) and less-biased condition (*in grey on right figure*)

Figure 24 shows the graphs of the data distribution of **shared percent** measure from the memory-aided local search model (*represented in blue*) superimposed on shared percent measure from biased experimental condition (*in grey on left figure*) and less-biased experimental condition (*in grey on right figure*).

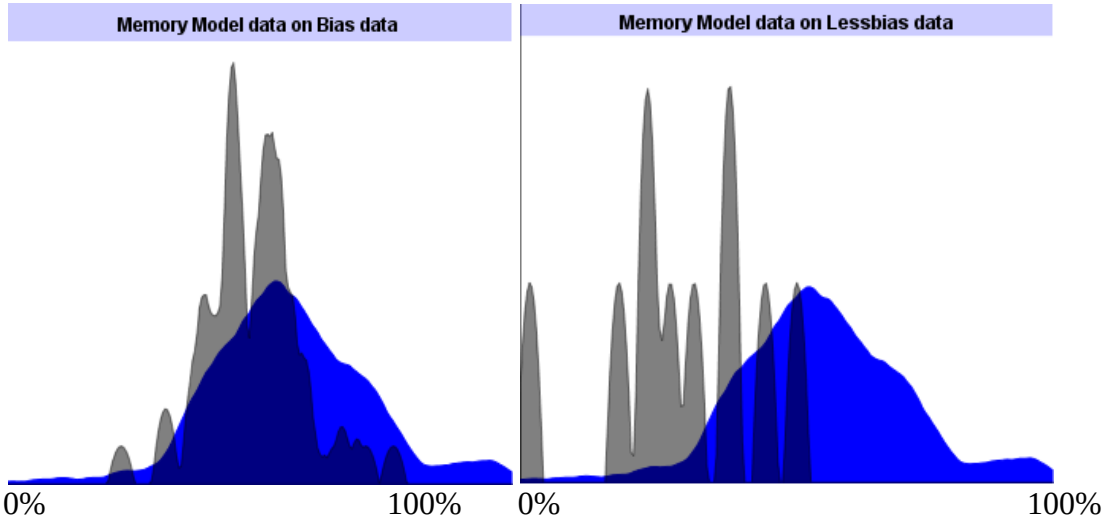

Figure 24. Shared percent distribution from memory-aided search model (*in blue*) superimposed on biased condition (*in grey on left figure*) and less-biased condition (*in grey on right figure*)

As it can be seen from Figure 24, some significant area of the data from local search model overlaps with the data from both the biased experimental condition and less-biased experimental condition.

Though such a visual comparison gives some information about the fit between the experimental data and the model data, a more standardized analysis is required to compare the two distributions to make any concrete claims of similarity between the model data and the experimental data.

Even though the trend of decreasing shared percent between biased and less-biased condition and increasing unique percent is observed between local search and the memory-aided local search model, the distribution of the percentage data from the models seems to be shifted from distribution of percentage data from the experiment either in the right or the left direction. Hence to make things more clear and standardized, a single measure combining the shared and unique percent was calculated and was used to compare the experimental data and data from model simulations.

Magnitude of difference between the shared percent and unique percent is the standardized measure used which shows the magnitude of difference between the percentage of shared information discussed to unique information discussed and it is calculated using the ratio between shared percent and unique percent as shown in the equation below.

$$M_{SU} = \frac{Shared\ Percent}{Unique\ Percent}$$

Magnitude of difference was calculated on data from the experimental conditions: Biased and Less-Biased and also on data from the three search models: Random, Local and Memory-Aided.

		Magnitude of Difference (M)
Model	Random Search	6.83
	Local Search	3.38
	Memory-Aided Local Search	2.72
Experiment	Biased	3.78
	Less-Biased	1.67

Figure 25. Magnitude of difference value for all three models and two experiment conditions

As shown in Figure 25, the magnitude of difference between shared percent and unique percent in random is 6.8 which is very high and does not match up to data from either conditions. The magnitude of difference in local model is 3.38 and is very similar to magnitude of difference in biased model which is 3.78. The magnitude of difference in cognitive model is 2.72 and magnitude of difference in less-biased is condition is 1.67. Though the decreasing trend observed in less-biased condition is also observed in the memory-aided local search condition, it is about 1 unit away from the magnitude of difference value in both biased and less-biased condition. Hence a more formal analysis needs to be conducted to correctly compare models and the experiment conditions.

Traditional statistics such as the t-test cannot be employed to compare the results from the experiment with the model simulation results because the results from the experiment follows a normal distribution with a small sample size, whereas the results from the model are not-normal and are based on a large sample size. Therefore a

Bayesian estimation method was employed to compare the data from the experiment with the model to identify the difference/similarity between them.

Bayesian estimation is an alternative to t-test to compare two groups when the distributions are skewed and have different sample sizes (Kruschke 2013). "Bayesian estimation for 2 groups provides complete distributions of credible values for the effect size, group means and their difference, standard deviations and their difference, and the normality of the data" (Kruschke, 2013, p1). In simple terms, the method generates a large set of new values based on the provided input data, computed based on Bayesian method, which is credible and not simply resampled as done with traditional statistics. The new values are all the possible values generated based on several combinations of mean and standard deviations of the input data of both groups being compared. Furthermore, Bayesian estimation can also be used to accept the null hypothesis that is there is no difference between the groups.

Therefore Bayesian estimation procedure was first used to compare the distributions of magnitude of difference measure between the local search model and biased condition. This analysis provided the result as shown in Figure 26. Figure 26 is the distribution of mean value difference between local search model and biased condition for the magnitude of difference measure. The Bayesian estimation procedure produces a distribution of all possible mean values for each group and then takes difference between them and then provides a distribution of those difference values. If the distribution contains a zero in the 95% high density interval (HDI) of the distribution then it can be deduced that there is no difference in their means. In this case the zero slightly misses the 95% HDI. This mean that distribution of the magnitude difference measure from the

biased experiment condition is different from the distribution of the magnitude measure from local search model. The mean of this difference distribution is -0.601 and that the difference distribution does contain a zero. Hence it can be inferred that local search model is somewhat representative of the biased experiment condition

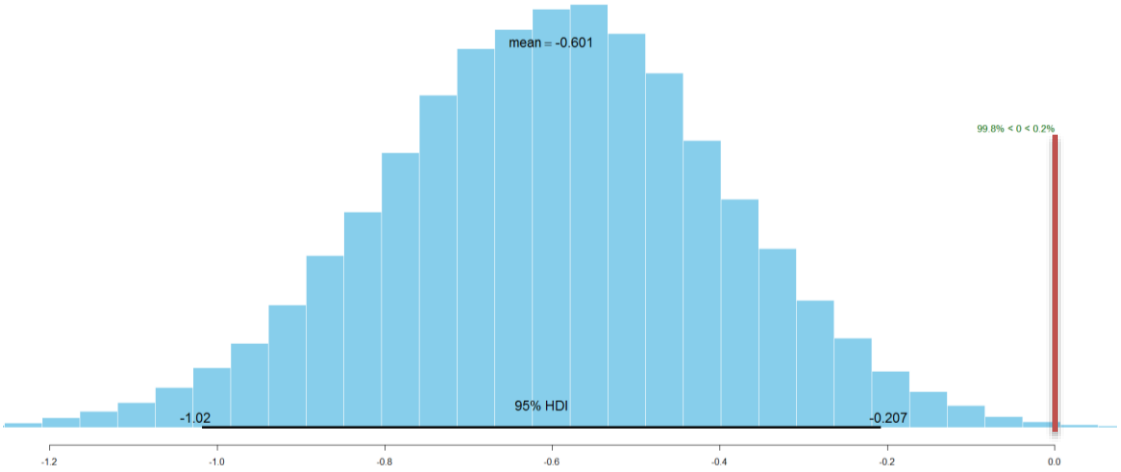

Figure 26. Distribution of mean value difference between local search model and biased condition for the magnitude of difference measure. The zero slightly missed 95% HDI interval

Next the Bayesian estimation procedure was used to compare the distributions of magnitude of difference measure between the local search model and the less-biased condition. This analysis provided the result as shown in Figure 27. Figure 27 is the distribution of mean value difference between local search model and less-biased condition for the magnitude of difference measure. In this case the zero misses the 95% HDI. This means that distribution of the magnitude difference measure from the less-biased experiment condition is different from the from local search model. However this time the mean of this difference distribution is 1.32. Hence it can be inferred that local search model is better representative of the biased experiment condition in comparison to the less-biased experiment condition.

Next the magnitude of difference measure between the memory-aided local search model and biased experimental condition was compared. As can be seen in Figure 28, the zero is away from the 95% HDI of the mean difference distribution. This means that distribution of the magnitude of difference measure from the memory-aided local search model is significantly different from the biased experiment condition. The mean of this difference distribution is -1.16

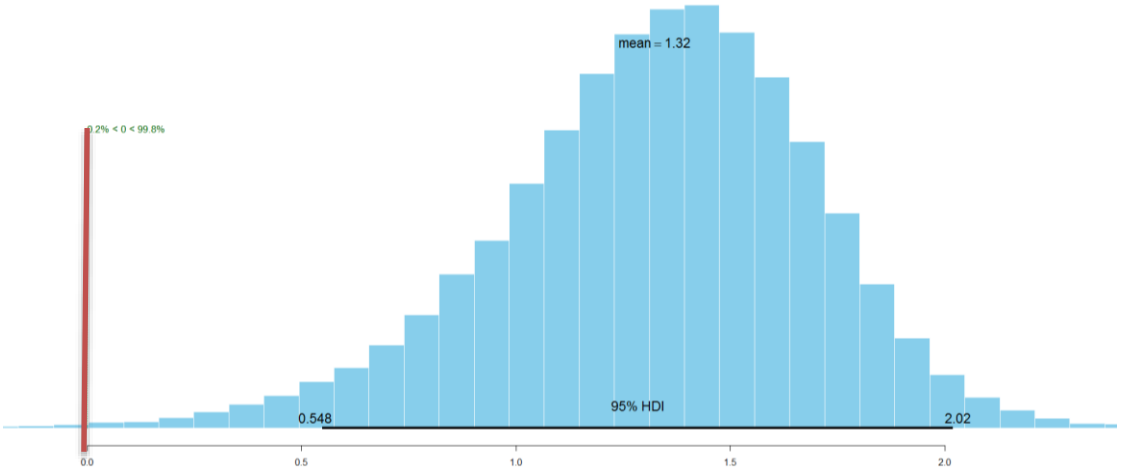

Figure 27. Distribution of mean value difference between local search model and less-biased condition for the magnitude of difference measure. The zero misses 95% HDI interval

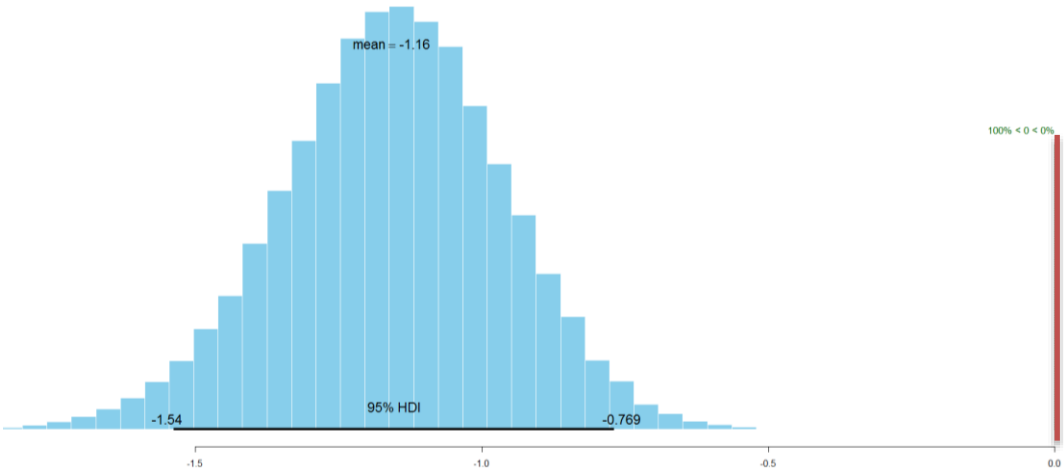

Figure 28. Distribution of mean value difference between memory-aided local search model and biased condition for the magnitude of difference measure. The zero misses 95% HDI interval

Finally, the magnitude of difference measure between the memory-aided local search model and less-biased experiment condition was compared. As can be seen in figure 29, the zero is almost in the 95% HDI of the mean difference distribution. This means that distribution of the magnitude of difference measure from the memory-aided local search model is representative of the less-biased experiment condition in comparison to the biased experiment condition with mean of this difference distribution being 0.767.

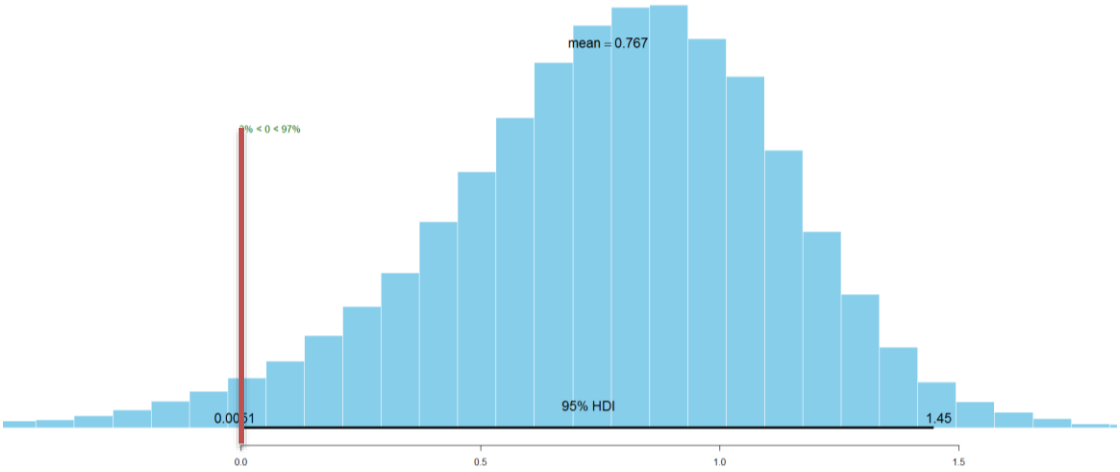

Figure 29. Distribution of mean value difference between memory-aided local search model and less-biased condition for the magnitude of difference measure. The zero is almost in 95% HDI interval

Table 13 shows the mean of the distribution of mean difference of M for the four comparisons conducted using the Bayesian estimation method. The two cells highlighted in the table contain the value of M for the comparison that was found the most similar where the local search model is most representative of the bias experiment condition and that memory-aided local search model is the most representative of the less-biased condition.

Table 13. Table of M values for all the four comparisons made. Cells highlighted has the values of the comparisons that revealed most similarity

	Biased Condition	**Less-Biased Condition**
Local Search Model	-0.601	1.32
Memory-Aided Local Search Model	-1.16	0.76

Summary of findings

It was found, through visual comparison of the data distributions and through the magnitude of difference measure, that data from the random model was not representative of either of the experimental conditions: Biased or less-biased. However it was found, through a Bayesian estimation procedure on the magnitude of difference measure, that data from the local search model were better representative of the data from the biased condition in comparison to less-biased experiment condition. Similarly it was found that data from the memory-aided local search model were better representative of the data from the less-biased experiment condition in comparison to the biased experiment condition. It was also observed that data from the agent-based model simulation is fairly good representative of the empirical results obtained from the human-in-the-loop team cognition experiments.

DISCUSSION

This dissertation work investigated the presence of the information pooling bias in cyber defense analyst teams conducting detection tasks as part of forensics analysis and also demonstrated that collaborative visualizations, designed considering human cognitive processes, can be effective in minimizing this bias and improving cyber defense analyst team performance. Furthermore, agent-based modeling was used to theorize about internal cognitive search processes in human analysts that result in such biases during their team discussions.

Results strongly indicate that all the teams who participated in the experiment exhibited the bias while performing the detection task by spending a majority of time discussing attacks that were also observed by other members of the team, whereas they spent a low percentage of time discussing attack that were uniquely available to each team member and which were part of a large scale attack.

Specifically, it was observed that when teams did not receive the visualization during their discussion, they discussed shared attack information 63.9% of time which is 3.8 times higher than the 16.9% of time spent discussing unique attack information. However, during Mission 2, when the participant teams in the visualization condition used the prototype collaborative visualization, they discussed shared attack information only 50.3% of time which is only 1.7 times higher than 30.2% of time spent on discussing the unique attack information. This demonstrated a stark increase in the amount of time spent discussing the unique pieces of information when the cognitively friendly visualization was introduced. However bias was observed to still exist in Mission 2.

These findings strongly show that participant teams demonstrated the information pooling bias and this indicate that if forensics analysts collaborate to analyze evidence they may also be affected by the information pooling bias as hypothesized (Hypothesis 1) in this dissertation.

Detection performance of the teams was also observed to improve in teams who used the tailor-made collaborative visualization tool during their discussions. Teams without visualization on an average detected 11 attacks, whereas teams with the visualization on an average detected 14 attacks. This improvement in detection performance comes from the detection of increased number of unique attacks as opposed to the detection of the shared attacks where the average number of shared attacks detected with or without visualization remained the same, but the average number of unique attacks that was part of a large scale attack detected with visualization was significantly higher than unique attacks detected without visualization.

These findings indicate that the information pooling bias can be minimized (not completely mitigated) in cyber defense analyst teams conducting the detection task as part of forensics analysis by using tailor-made collaboration tools developed taking into consideration the cyber defense analyst's cognitive requirements as hypothesized (Hypothesis 2) in this dissertation.

A strong positive correlation between the percentage of time spent on discussing unique attacks and detection performance was detected whereas a strong negative correlation between the percentage of time spent discussing shared attacks and detection performance was found. This result indicates that spending more time discussing shared information could be detrimental to team performance. It was also found that the

percentage of time spent discussing the unique information was a significant predictor of the detection performance of cyber defense analyst team. This result confirms the natural intuition that bringing into the discussion unique and expert information available to each team member could lead to superior team performance.

The mixed ANOVA analysis revealed a statistically significant interaction effect on all three measures: shared percentage, unique percentage and detection performance from detecting unique attacks. This means all three measures varied between the two Missions as a function of the condition. There was a significant drop in shared percentage in the visualization condition between Mission 1 to Mission 2 in comparison to a non-significant change in shared percentage in slides and wiki condition. Similarly there was a significant increase in unique information percentage and detection performance in the visualization condition between Mission 1 to Mission 2 in comparison to the corresponding non-significant change in slides and wiki condition. Interestingly in conditions in which the participant teams used slides or wiki, a slight non-significant increase in shared percent was detected in their discussion in Mission 2 which indicated of the possibility that even experience in the task may not mitigate this bias in cyber defense analyst teams.

No statistically significant difference was found in amount of shared attack information discussed, unique attack information discussed or the detection performance among the teams across the three conditions in Mission 1 because all the teams in all conditions in Mission 1 used only Microsoft PowerPoint slides during their discussion.

On the contrary, a statistically significant difference was found in amount of shared attack information discussed, unique attack information discussed and the detection

performance among the teams in Mission 2 across the three conditions when different interventions were employed. Specifically the significant difference in performance stemmed from the difference in the number of unique attacks detected.

No difference between slides and wiki condition across all three measures was detected, whereas much of the variance that contributed to the significance came from the difference between the slide and visualization condition, and wiki and visualization condition across all three measures. These results indicate that using off-the-shelf collaboration tools such as the wiki application which is commonly used in the cyber defense arena for collaboration is no better than no collaboration tool because it was found to be ineffective in reducing information pooling bias and also ineffective in improving performance of the team.

Often visualizations or any new collaborative tools are perceived to cause more cognitive load or perceived to be an impedance to the existing work process. The NASA TLX workload measures indicate that the participants perceived no significant mental load difference between using Microsoft PowerPoint slides or wiki application or the prototype collaborative visualization tool. All of the participants perceived the task to be moderately hard with average mental work load rated as about seven on the scale of one to ten with one meaning no mental load and ten indicating a very high mental load. Interestingly a significant variance was detected in participants' perception of the time pressure they felt doing the task. The participants in the wiki condition (7.2) rated the task to be around one unit more time pressure than participants in the slide (6.2) and visualization condition (6.5). The participants in the wiki condition were observed to be switching between different pages of attack descriptions during the discussion search for

information to contribute which might have contributed to this perception of higher time pressure.

Hence from the results, it can be inferred that when cyber defense analyst conduct collaborative detection task as part of forensics, they may be plagued with an information pooling bias which prevents the individual members of the team from communicating the isolated and unique events which only they observe and resort to repeated communication and discussion of the information that is known to everyone on the team. The results obtained complement the results found in teams in other domains such as medical teams, intelligence analysis and jury teams (Lu, Yuan, & McLeod, 2012; Wittenbaum, Hollingshead, & Botero, 2004; Stasser & Titus, 1985).

Such a biased team discussion could lead to ineffective forensics analyses because sometimes integrating the seemingly disparate unique and isolated events could be crucial to detecting large scale multi-step attacks such as advanced persistent threats (APT). Currently there is a scarcity of methods to proactively detect APTs even though the breadcrumbs of the attack emerging in a network are available, observed, and most often reported by the analysts. It will be hard to program an expert system to integrate such seemingly disparate information and detect an emerging large scale attack because it is difficult for the systems to leverage and integrate contextual information. However it is possible for the human analyst to incorporate the contextual information to integrate the seemingly disparate events that are part of a large scale emerging attack. On the other hand, humans have biases and cognitive limitations that prevent them from doing such complex integration. Therefore instead of trying to achieve perfectly intelligent experts systems to do such tasks and trying to keep the human analysts out of the loop, it would

be more effective to leverage human strengths such as contextual-based decision making and pattern recognition and alleviating their cognitive limitations through tools that will lead to sustained superior performance and at a lower software development cost.

Past work, investigating this bias in other contexts has mostly explored the social and motivational causes of this bias and there is very limited work done on exploring the cognitive processes that could be causing this bias (Wittenbaum et al., 2004). It is therefore imperative to understand the cognitive processes underlying the bias in order to design cognitive-friendly collaboration tools in the cyber defense context. Towards that, agent based modelling (ABM) can be used to theorize about the underlying cognitive processes. An ABM can be developed to help theorize about the effect of individual cognitive processes (coded as rules) on social/team level processes (observed as macro-level behaviors) (Rajivan, Janssen & Cooke, 2013). Such agent based models were developed as part of this dissertation work to theorize about the cognitive search processes used in the head of an analyst who is trying to search for information to contribute to an ongoing discussion. Cognitive search processes were particularly chosen for exploration because they were suspected to be the key component behind the bias because if the team members conducted a depth first kind of search, it would lead to a tunneled and narrow focused discussion spending most of the time discussing the same topic and being myopic about other potential large scale attacks. Hence it was hypothesized in that humans, by default, use heuristics based on local search/uphill search process to search for information in their memories in order to contribute to the ongoing discussion, leading to the information pooling bias. Furthermore, when the ongoing topic of discussion does not appear in the current search horizon, it can cause

humans to not recognize the presence of the related information in their memory spaces. Therefore assistance is needed in the form of visual interventions to stimulate recognition memory to help find that relevant information to bring to the discussion. It was also hypothesized that visualization used in the experiment aids memory recognition of the appropriate information to contribute to the discussion and also information to be avoided that is already well discussed, thereby leading to a less biased team discussion and improved performance.

Three search models were developed and explored: Random, Local and Memory-Aided Local. The random search model was the null model for which agents' do random walks in search of information to contribute to the discussion and was developed for comparison purposes to evaluate whether the models of interest (local and cognitive) were not producing a stochastic behavior. Results indicated that both local search models and the memory-aided local search model deviated significantly from the null model (random search) and therefore it can be inferred that local and memory-aided local search models were not behaving in a random fashion.

In the local search model, agents conducted local neighborhood search and moved in an uphill manner in search of information to contribute to the discussion. In the memory-aided local search model, agents were aided in finding regions in its memory space where it would be possible to find relevant discussion information and once they knew the region to examine, they did local/uphill search in that region in search of information to contribute to the discussion. The local search model and memory-aided local search model were found to deviate significantly from each other in terms of the discussion focus measured through shared percent and unique percent measures. It was observed

that agents in the local search model spent more time discussing shared information more than agents in the memory-aided local search model. Similarly it was observed that agents in the local search model spent less time discussing unique information compared to agents in the memory-aided local search model.

The models themselves do not convey much information and hence have to be compared with the complementary human-in-the-loop experiment to know if the behavior of the agents observed in the local search model was representative of the biased team discussion observed in the human-in-the-loop experiment and also to know if the behavior of the agents observed in the memory-aided local search model was representative of the less-biased team discussion observed in the human-in-the-loop experiment. This trend between local search and memory-aided local search in terms of discussion focus is in parallel to the trend observed between biased discussion and less-biased discussion (for which the teams used visualization) in the human-in-the-loop experiment. Instead of conducting a comparison based on raw percentage values which might lead to making incorrect inferences, a comparison was conducted based on a standardized value which was a ratio between the percentage of time spent discussing shared information to the percentage of time spent discussing unique information and this ratio was called magnitude of difference.

Bayesian statistics (Kruschke 2013) were used to compare the magnitude of difference values from the experiment against the magnitude of difference values from the model. Based on the results the local search model was found to be representative of the biased discussion observed in the experiment as hypothesized (Hypothesis 3) and that memory-aided local search model was found to be representative of the less biased

discussion observed in the experiment as hypothesized (Hypothesis 3A). These results are particularly insightful because we can now suspect that humans could be using simple heuristics and local/uphill kind of search process when they are undergoing such a bias and that they could be lacking a global view due to low recognition memory which is essential to see the connections between seemingly disparate but associated information. Therefore in such contexts, we need tools and visualizations that will enhance human search processes and will stimulate their recognition memory which could lead to a more global view of the situation at hand. Moreover results from the experiment conducted with human subjects could be replicated with an agent-based model by implementing the key cognitive aspects of the human subjects. It was observed that the results closely align with the empirical results obtained from the human-in-the-loop experiments as hypothesized (Hypothesis 3B).

Limitations

The visualization used in the experiment was built strictly for this experiment and assumes that the relevant meta-data needed to develop the visualization would be available. The experiment was conducted with students with little to no cyber defense experience who were trained to perform the synthetic cyber defense and forensics task. Therefore, for more ecologically valid results, the experiment needs to be conducted with actual cyber forensics teams.

This work supports the use of visualization tools that consider human cognitive limitations and biases in design. These can be very effective in improving team performance. However the tool built here was built exclusively for the experiment and the task designed.. Such a visualization tool could be built by leveraging and mining

analyst reports and archived attack reports, so that forensics analysts can effectively detect emerging large scale attacks. Usability studies should also be employed to make the tool being developed user- and cognitive- friendly.

Future Directions

It was observed that some of the participants were strongly holding on to some of their pre-conceived theories about the attacks even though other team member deemed them to be unlikely and in most cases their pre-conceived theories were indeed incorrect in that context. This effect still remained even when other team members provided reasons that their theories were incorrect. This has a stark similarity to the confirmation bias seen in other domains and may also plague the cyber defense task. Hence it would be worthwhile to explore and investigate the confirmation bias in the cyber defense context.

It would be interesting to observe and measure how experience working with the same team confounds with information pooling bias and also to measure whether the bias increases or decreases in the cyber defense context with time and experience. Such a study can be done by simply extending this existing study with more scenarios and missions and running it as a longitudinal study, measuring the bias at each discussion trial.

The agent based model was simplistic in terms of the cognitive features coded into the agent. Only few key cognitive features were used to construct agents. Therefore to make the model more ecologically valid it would be worthwhile to explore how ACT-R cognitive models can be integrated into agent based modelling methodologies so that social/group/team processes emerge from theoretically strong cognitive agent models. The existing model also explored only two models of a heuristic search processes. It

would be worthwhile to compare experimental data with models on algorithmic search processes such as breadth-first search and depth-first search.

Conclusion

This dissertation work has multiple implications. Foremost, this dissertation work contributes to the knowledge about of the science of team-based cyber defense which is severely limited. Specifically, this work contributes insights on plausible cyber defense analysts' biases when they share information with each other. The dissertation was carried out using a combination of a human-in-the-loop experiment, agent-based modeling, and software prototyping to investigate team cognition in cyber defense and therefore demonstrate how such a multi-faceted, multidisciplinary approach is effective and insightful for team cognition research. The collaboration software prototyping was done from a cognitive standpoint considering human strengths and limitations. This demonstrates the advantage of developing tools using a cognitive engineering approach to mitigate the human operator's cognitive limitations. Finally, this work is a demonstration of the advantages of effective team work on cyber defense performance.

REFERENCES

Anderson, J. R. (1996). ACT: A simple theory of complex cognition. *American Psychologist, 51*(4), 355.

Anderson, J. R., & Lebiere, C. (1998). *The atomic components of thought* Psychology Press.

Atkinson, R. C., & Juola, J. F. (1974). Search and decision processes in recognition memory. WH Freeman.

Ball, J., Myers, C., Heiberg, A., Cooke, N. J., Matessa, M., Freiman, M., & Rodgers, S. (2010). The synthetic teammate project. *Computational and Mathematical Organization Theory, 16*(3), 271-299.

Barceló, J. A., Bernal, F. D. C., del Olmo, R., Mameli, L., Quesada, F. M., Poza, D., & Vilà, X. (2013). Social interaction in hunter-gatherer societies: Simulating the consequences of cooperation and social aggregation. *Social Science Computer Review*, , 0894439313511943.

Bietz, M. J., Abrams, S., Cooper, D. M., Stevens, K. R., Puga, F., Patel, D. I., Olson,G.M., & Olson, J. S. (2012). Improving the odds through the Collaboration Success Wizard. Translational behavioral medicine, 2(4), 480-486.

Bier, E. A., Card, S. K., & Bodnar, J. W. (2008). Entity-based collaboration tools for intelligence analysis. *Visual Analytics Science and Technology, 2008. VAST'08. IEEE Symposium on*, 99-106.

Blasch, E., Bosse, E., & Lambert, D. A. (Eds.). (2012). High-Level Information Fusion Management and Systems Design. Artech House.

Brannick, M. T., Roach, R. M., & Salas, E. (1993). Understanding team performance: A multimethod study. *Human Performance, 6*(4), 287-308.

Brauer, M., Judd, C. M., & Gliner, M. D. (1995). The effects of repeated expressions on attitude polarization during group discussions. *Journal of Personality and Social Psychology, 68*(6), 1014.

Briggs, G. E., & Naylor, J. C. (1965). Team versus individual training, training task fidelity, and task organization effects on transfer performance by three-man teams. *Journal of Applied Psychology, 49*(6), 387.

Buja, A., McDonald, J. A., Michalak, J., & Stuetzle, W. (1991, October). Interactive data visualization using focusing and linking. In Visualization, 1991. Visualization'91, Proceedings., IEEE Conference on (pp. 156-163). IEEE.

Burnstein, E., & Vinokur, A. (1977). Persuasive argumentation and social comparison as determinants of attitude polarization. *Journal of Experimental Social Psychology, 13*(4), 315-332.

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

Cannon-Bowers, J. A., Salas, E., & Converse, S. (2001). Shared mental models in expert team decision making. *Individual and Group Decision Making,* , 221-246.

Cannon-Bowers, J., & Salas, E. (1993). Shared mental models in expert team decision making. *Individual and Group Decision Making,* , 221-246.

Cannon-Bowers, J., Salas, E., & Converse, S. (1990). Cognitive psychology and team training: Training shared mental models and complex systems. *Human Factors Society Bulletin, 33*(12), 1-4.

Carley, K. M., Fridsma, D. B., Casman, E., Yahja, A., Altman, N., Chen, L., Nave, D. (2006). BioWar: Scalable agent-based model of bioattacks. *Systems, Man and Cybernetics, Part A: Systems and Humans, IEEE Transactions on, 36*(2), 252-265.

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

Christensen, C., & Abbott, A. S. (2003). 10 team medical decision making. *Decision Making in Health Care: Theory, Psychology, and Applications,*, 267.

Chung, H., Yang, S., Massjouni, N., Andrews, C., Kanna, R., & North, C. (2010). Vizcept: Supporting synchronous collaboration for constructing visualizations in

intelligence analysis. *Visual Analytics Science and Technology (VAST), 2010 IEEE Symposium on*, 107-114.

Collyer, S. C., & Malecki, G. S. (1998). Tactical decision making under stress: History and overview.

Connolly, J., Davidson, M., Richard, M., & Skorupka, C. (2012). The trusted automated eXchange of indicator information (TAXII™).

Cooke, N. J., Gorman, J. C., & Rowe, L. J. (2004). An ecological perspective on team cognition. *An Ecological Perspective on Team Cognition*,

Cooke, N. J., Gorman, J. C., & Winner, J. L. (2007). Team cognition. *Handbook of Applied Cognition*, , 239-268.

Cooke, N., Gorman, J., & Kiekel, P. (2008). Communication as team-level cognitive processing. *Macrocognition in Teams*, , 51-64.

Cooke, N. J., Gorman, J. C., Myers, C. W., & Duran, J.L. (2013). Interactive Team Cognition, 37, Cognitive Science, 255-285, DOI: 10.1111/cogs.12009.

Cooke, N. J., Neville, K. J., & Rowe, A. L. (1996). Procedural network representations of sequential data. Human-Computer Interaction, 11(1), 29-68.

Cooke, N. J., Rivera, K., Shope, S. M., & Caukwell, S. (1999). A synthetic task environment for team cognition research. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, , *43*(3) 303-308.

Cooke, N. J., Salas, E., Kiekel, P. A., & Bell, B. (2004). Advances in measuring team cognition. *Team Cognition: Understanding the Factors that Drive Process and Performance*, , 83–106.

Cooke, N. J., & Shope, S. M. (2004). Designing a synthetic task environment. *Scaled Worlds: Development, Validation, and Application*, , 263-278.

Cowan, N. (1988). Evolving conceptions of memory storage, selective attention, and their mutual constraints within the human information-processing system. Psychological bulletin, 104(2), 163.

Cowan, N. (2001). Metatheory of storage capacity limits. Behavioral and brain sciences, 24(01), 154-176.

D'Amico, A., & Whitley, K. (2008). The real work of computer network defense analysts. *Vizsec 2007,* , 19-37.

DAmico, 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.

Dutt, V., Ahn, Y. S., Ben-Asher, N., & Gonzalez, C. (2012). Modeling the effects of base-rate on cyber threat detection performance. In N. Rußwinkel, U. Drewitz & H. v. Rijn (Eds.), Proceedings of the 11th International Conference on Cognitive Modeling (ICCM 2012) (pp. 88-93). Berlin, Germany: Universitaetsverlag der TU Berlin.

Dutta, A., & McCrohan, K. (2002). Management's role in information security in a cyber economy. *California Management Review, 45*(1), 67-87.

Endsley, M. R. (1995). Toward a theory of situation awareness in dynamic systems. *Human Factors: The Journal of the Human Factors and Ergonomics Society, 37*(1), 32-64.

Endsley, M. R. (2000). Theoretical underpinnings of situation awareness: A critical review. *Situation Awareness Analysis and Measurement,*, 3-32.

Endsley, M. R. (2006). Situation awareness. *Handbook of Human Factors and Ergonomics,* , 528-542.

Endsley, M. (1989). Final report: Situation awareness in an advanced strategic Mission (NOR DOC 89-32). *Hawthorne, CA: Northrop Corporation*

Gigone, D., & Hastie, R. (1993). The common knowledge effect: Information sharing and group judgment. *Journal of Personality and Social Psychology*, *65*(5), 959.

Gilbert, N. (1997). A simulation of the structure of academic science.

Gonzalez, C., Lerch, J. F., & Lebiere, C. (2003). Instance-based learning in dynamic decision making. Cognitive Science, 27(4), 591-635. doi:10.1016/S0364-0213(03)00031-4

Gorman, J. C., Cooke, N. J., & Winner, J. L. (2006). Measuring team situation awareness in decentralized command and control environments. *Ergonomics*, *49*(12), 1312-1325.

Guerra, A. (2011). A framework for building intelligent software assistants for virtual worlds. *ETD Collection for Pace University.Paper AAI3462987*,

Hall, D. L., & Llinas, J. (1997). An introduction to multisensor data fusion. *Proceedings of the IEEE*, *85*(1), 6-23.

Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (task load index): Results of empirical and theoretical research. *Human Mental Workload*, *1*, 139–183.

Hastie, R., Penrod, S., & Pennington, N. (2013). *Inside the jury* The Lawbook Exchange, Ltd.

Heuer, R. J. (1999). *Psychology of intelligence analysis* United States Govt Printing Office.

Hill, G. W. (1982). Group versus individual performance: Are N 1 heads better than one? *Psychological Bulletin*, *91*(3), 517-539.

Hui, P., Bruce, J., Fink, G., Gregory, M., Best, D., McGrath, L., & Endert, A. (2010). Towards efficient collaboration in cyber security. *Collaborative Technologies and Systems (CTS), 2010 International Symposium on*, 489-498.

Isenberg, P., & Fisher, D. (2009, June). Collaborative Brushing and Linking for Co-located Visual Analytics of Document Collections. In Computer Graphics Forum (Vol. 28, No. 3, pp. 1031-1038). Blackwell Publishing Ltd.

Janssen, M. A., & Hill, K. (2013). Benefits of grouping and cooperative hunting among ache hunter-gatherers: Insights from an agent-based foraging model.

Jariwala, S., Champion, M., Rajivan, P., & Cooke, N. J. (2012). Influence of team communication and coordination on the performance of teams at the iCTF competition. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting, , 56*(1) 458-462.

Johnston, R. (2005). Analytic culture in the US intelligence community: An ethnographic study. *Washington, DC: Center for the Study of Intelligence.*,

Johnston, W. A., & Briggs, G. E. (1968). Team performance as a function of team arrangement and work load. *Journal of Applied Psychology, 52*(2), 89.

Kahneman, D. (2003). A perspective on judgment and choice: mapping bounded rationality. American psychologist, 58(9), 697.

Kahneman, D., & Klein, G. (2009). Conditions for intuitive expertise: A failure to disagree. *American Psychologist; American Psychologist, 64*(6), 515.

Kahneman, D., Slovic, P., & Tversky, A. (1982). Judgment under uncertainty: Heuristics and biases. New York: Cambridge.

Keel, P. (2004). Electronic card wall (EWall). Collaboration and knowledge management workshop proceedings, San Diego, CA.

Kelly, J. R., Jackson, J. W., & Hutson-Comeaux, S. L. (1997). The effects of time pressure and task differences on influence modes and accuracy in decision-making groups. *Personality and Social Psychology Bulletin, 23*(1), 10-22.

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

Koedinger, K. R., Anderson, J. R., Hadley, W. H., & Mark, M. A. (1997). Intelligent tutoring goes to school in the big city. International Journal of Artificial Intelligence in Education (IJAIED), 8, 30-43.

Kompridis, N. (2000). So we need something else for reason to mean. International journal of philosophical studies, 8(3), 271-295.

Kraemer, S., Carayon, P., & Clem, J. (2009). Human and organizational factors in computer and information security: Pathways to vulnerabilities. *Computers & Security, 28*(7), 509-520.

Kruschke, J. K. (2013). Bayesian estimation supersedes the t-test. Journal of Experimental Psychology: General, 142(2), 573.

Laird, J. E., Newell, A., & Rosenbloom, P. S. (1987). Soar: An architecture for general intelligence. *Artificial Intelligence, 33*(1), 1-64.

Landauer, T. K., Foltz, P. W., & Laham, D. (1998). An introduction to latent semantic analysis. *Discourse Processes, 25*(2-3), 259-284.

Langan-Fox, J., Code, S., & Langfield-Smith, K. (2000). Team mental models: Methods, techniques and applications. Human Factors, 42(2), 1-30.

Laughlin, P. R., & Bonner, B. L. (1999). Collective induction: Effects of multiple hypotheses and multiple evidence in two problem domains. *Journal of Personality and Social Psychology, 77*(6), 1163.

Lewis, K. (2003). Measuring transactive memory systems in the field: Scale development and validation. *Journal of Applied Psychology, 88*(4), 587.

Liu, S., Chen, Y., & Lin, S. (2013). A novel search engine to uncover potential victims for APT investigations. *Network and parallel computing* (pp. 405-416) Springer.

Liu, P. (2009.). Computer-aided human centric cyber situation awareness.

Lowenthal, M. M. (2002). Intelligence from secrets to policy. *Congressional Quarterly Press (Washington, DC),* (second edition), 8.

Lu, L., Yuan, Y. C., & McLeod, P. L. (2012). Twenty-five years of hidden profiles in group decision making: A meta-analysis. *Personality and Social Psychology Review : An Official Journal of the Society for Personality and Social Psychology, Inc, 16*(1), 54-75.

McNeese, M., Cooke, N.J., Champion, M.A. (2011) Situating Cyber Situation Awareness. Proceedings of the 10th International Conference on Naturalistic Decision Making.

Mercier, H., & Sperber, D. (2011). Why do humans reason? Arguments for an argumentative theory. Behavioral and brain sciences, 34(02), 57-74.

Mesmer-Magnus, J. R., & DeChurch, L. A. (2009a). Information sharing and team performance: A meta-analysis. *Journal of Applied Psychology, 94*(2), 535.

Mesmer-Magnus, J. R., & DeChurch, L. A. (2009b). Information sharing and team performance: A meta-analysis. *Journal of Applied Psychology, 94*(2), 535.

Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review, 63*(2), 81.

Natter, M., Bos, N., Ockerman, J., Happel, J., Abitante, G., & Tzeng, N.A C2 hidden profile experiment.

Newell, A. (1990). Unified theories of cognition. *Cambridge, MA: Harvard University,*

Pioch, N. J., & Everett, J. O. (2006, November). POLESTAR: collaborative knowledge management and sensemaking tools for intelligence analysts. In Proceedings of the 15th ACM international conference on Information and knowledge management (pp. 513-521). ACM.

Ponemon Institute. (2013). *2013 cost of cyber-crime study: United states.*

Prince, C., Ellis, E., Brannick, M. T., & Salas, E. (2007). Measurement of team situation awareness in low experience level aviators. *The International Journal of Aviation Psychology, 17*(1), 41-57.

Puncochar, J. M., & Fox, P. W. (2004). Confidence in individual and group decision making: When" two heads" are worse than one. *Journal of Educational Psychology, 96*(3), 582.

Puvathingal, B. J., & Hantula, D. A. (2011). Revisiting the psychology of intelligence analysis: From rational actors to adaptive thinkers.

Rajivan, P. (2012). *CyberCog A Synthetic Task Environment for Measuring Cyber Situation Awareness,*

Rajivan, P., Champion, M., Cooke, N. J., Jariwala, S., Dube, G., & Buchanan, V. (2013). Effects of Teamwork versus Group Work on Signal Detection in Cyber Defense Teams. In Foundations of 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.

Rouse, W. B., & Morris, N. M. (1986). On looking into the black box: Prospects and limits in the search for mental models. *Psychological Bulletin, 100*(3), 349.

Salas, E., Cooke, N. J., & Rosen, M. A. (2008). On teams, teamwork, and team performance: Discoveries and developments. *Human Factors: The Journal of the Human Factors and Ergonomics Society, 50*(3), 540-547.

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.

Salas, E., Prince, C., Baker, D. P., & Shrestha, L. (1995). Situation awareness in team performance: Implications for measurement and training. *Human Factors: The Journal of the Human Factors and Ergonomics Society, 37*(1), 123-136.

Schnapp, L. M., Rotschy, L., Hall, T. E., Crowley, S., & O'Rourke, M. (2012). How to talk to strangers: facilitating knowledge sharing within translational health teams with the Toolbox dialogue method. Translational behavioral medicine, 2(4), 469-479.

Schulz-Hardt, S., Jochims, M., & Frey, D. (2002). Productive conflict in group decision making: Genuine and contrived dissent as strategies to counteract biased information seeking. *Organizational Behavior and Human Decision Processes, 88*(2), 563-586.

Schvaneveldt, R. W., Durso, F. T., & Dearholt, D. W. (1989). Network structures in proximity data. The psychology of learning and motivation, 24, 249-284.

Shiravi, H., Shiravi, A., & Ghorbani, A. (2011). A survey of visualization systems for network security. *Visualization and Computer Graphics, IEEE Transactions on*, (99), 1-1.

Shu, Y., & Furuta, K. (2005). An inference method of team situation awareness based on mutual awareness. *Cognition, Technology & Work, 7*(4), 272-287.

Sieck, W. R., & Arkes, H. R. (2005). The recalcitrance of overconfidence and its contribution to decision aid neglect. *Journal of Behavioral Decision Making, 18*(1), 29-53.

Sloman, S. A. (1996). The empirical case for two systems of reasoning. Psychological bulletin, 119(1), 3.

Stahl, G., Koschmann, T., & Suthers, D. (2006). Computer-supported collaborative learning.

Stahl, G. (2006). Group cognition (Vol. 106). Cambridge, MA: MIT Press.

Stanislaw, H., & Todorov, N. (1999). Calculation of signal detection theory measures. *Behavior Research Methods, 31*(1), 137-149.

Stasser, G., & Titus, W. (1985). Pooling of unshared information in group decision making: Biased information sampling during discussion. *Journal of Personality and Social Psychology, 48*(6), 1467.

Stasser, G., & Titus, W. (1987). Effects of information load and percentage of shared information on the dissemination of unshared information during group discussion. *Journal of Personality and Social Psychology, 53*(1), 81.

Stotz, A., & Sudit, M. (2007). INformation fusion engine for real-time decision-making (INFERD): A perceptual system for cyber attack tracking. *Information Fusion, 2007 10th International Conference on,* 1-8.

Stout, R. J., Salas, E., & Carson, R. (1994). Individual task proficiency and team process behavior: What's important for team functioning? *Military Psychology*, *6*(3), 177-192.

Straus, S. G., Parker, A. M., & Bruce, J. B. (2011). The group matters: A review of processes and outcomes in intelligence analysis. *Group Dynamics: Theory, Research, and Practice, 15*(2), 128.

Sun, R. (2006a). The CLARION cognitive architecture: Extending cognitive modeling to social simulation. *Cognition and Agent-based Interaction*, , 79-99.

Sun, R. (2006b). Prolegomena to integrating cognitive modeling and social simulation. *Cognition and Agent-based Interaction:From Cognitive Modeling to Social Simulation*,3-26.

Sun, R. (2008). Introduction to computational cognitive modeling. Cambridge handbook of computational psychology, 3-19.

Sun, R. (2009). Theoretical status of computational cognitive modeling. Cognitive Systems Research, 10(2), 124-140.

Taylor, R. (1990). Situational awareness rating technique (SART): The development of a tool for aircrew systems design. *The Situational Awareness in Aerospace Operations AGARDCP478*, (Situational Awareness in Aerospace Operations)

Vogel, A. L., Hall, K. L., Fiore, S. M., Klein, J. T., Michelle, B. L., Gadlin, H., ... & Falk-Krzesinski, H. J. (2013). The team science toolkit: enhancing research collaboration through online knowledge sharing. American journal of preventive medicine, 45(6), 787.

Vogt, K., Bradel, L., Andrews, C., North, C., Endert, A., & Hutchings, D. (2011). Co-located collaborative sensemaking on a large high-resolution display with multiple input devices. In Human-Computer Interaction–INTERACT 2011 (pp. 589-604). Springer Berlin Heidelberg.

Wegner, D. M. (1987). Transactive memory: A contemporary analysis of the group mind. *Theories of Group Behavior, 185*, 208.

Wilensky, U. (2002). Netlogo pd n-person iterated model. *Center for Connected Learning and Computer-Based Modeling, Northwestern University, Evanston, IL,*

Willett, W., Heer, J., Hellerstein, J., & Agrawala, M. (2011, May). CommentSpace: structured support for collaborative visual analysis. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (pp. 3131-3140). ACM

Williges, R. C., Johnston, W. A., & Briggs, G. E. (1966). Role of verbal communication in teamwork. *Journal of Applied Psychology, 50*(6), 473.

Wittenbaum, G. M., Hollingshead, A. B., & Botero, I. C. (2004). From cooperative to motivated information sharing in groups: Moving beyond the hidden profile paradigm. *Communication Monographs, 71*(3), 286-310.

Yates, J. F. (1990). Judgment and decision making. Prentice-Hall, Inc. Grissom, R. J., & Kim, J. J. (2012). Effect sizes for research: Univariate and multivariate applications. (2nd ed.). New York, NY: Taylor & Francis

APPENDIX A

CYBERCOG TRAINING MATERIAL

In this appendix, only the content from the training presentation is presented.

Background Information
Cyber security is a serious national threat in the modern era.
This research is in collaboration with the federal government working to improve the overall efficiency of the national cyber security response system.

Hello !
In this training module, you will learn:
- The basic concepts of Computer Networks
- About Cyber Attacks
- Types of Cyber Attacks
- And how to discuss and analyze cyber attacks as a team

Learning Objectives
Become familiar with computer networks and its terminologies
Get an understanding on how an attacker/hacker can attack computer networks
Learn how to discuss attacks with others on your team to get a big picture view of the network you will be analyzing.

Computer Networks
The internet is a vast network of computers. It connects billions of computer systems distributed world wide. This is a public network (accessible to anyone).
Every organization has its own private network of computers for its own purposes
And such individual private networks are connected via the Internet (public)
There is a lot of information exchanged in these networks and they contribute to network traffic

The Computer Network
A computer network is simply a collection of computers connected together
Computer networks allows information exchange
A network can connect different kinds of computers
- Personal computers
- Computer Servers – file server, website server
- Gateways & Routers

A network is often large and hence divided into several pieces for easy management, these are called sub-networks

Quiz Time
Question 1
What is done to make it easy to manage a large network?
- Shut down low priority computers
- Divide a network in to sub-networks
- Connect every computer with each other
- Nothing can be done

IP Address

Each computer device in a network is identified using a standard address called an IP address

It is simply a string of numbers separated by periods

EX: 74.125.224.179

This is the IP address of a Google's server

Anything that begins with 74.X.X.X belongs to Google

…

Another example relevant to the current study is an IP address that begins with 185.X.X.X This information will be important to you later on as this is the IP address of the organization you are going to defend.

Port

Port as the name suggests is the outlet for transmitting information.

A unique port number will be used by software programs to communicate with other devices.

Ports are used by software programs to send information, request information, etc.

They are usually represented by two or four digit numbers EX: 80, 64, 8081, 8084

Backdoor: ports used by computers/programs outside the private network to illegally access a computer/software.

Quiz Time

Question 2

What is the IP address of the organization you are going to defend?

165.X.X.X

74.X.X.X

185.X.X.X

10.X.X.X

Types of computer devices

Personal computers

The most common kind used for personal computing purposes (Low performance) EX: Desktops, laptops.

Networking Devices

Routers: Helps in correctly routing the computer data to the intended destination. Like Google map application but for networks

Gateway: As the name indicates, it is the machine that controls computer traffic to and from a private network. It's the end point of a private network.

Computer Server

As the name indicates it serves requests from other computers (High performance). Not for personal computing.

WebServer: Runs programs that renders websites on request.

Database Server: Responds to requests from webserver for retrieving stored data

File server/Data Server: For storing all private files. Responds to request to retrieve a certain file.

Load Balancer: It is a server that balances request load to any server to avoid overloading one server
Quiz Time
Question 3
How does information sent from software reach the network for transmission?

- Sent via port
- Sent via gateway
- Sent directly in to the network
- Sent via router

Network Map
A private network is often represented using a network diagram / map for easy understanding of how the different machines are connected with each other
It will show the different computer devices in the network, its connections along with its unique IP address
Remember your organization will have an IP address beginning with 185.X.X.X
It will also represent the subnetworks (pieces of networks) using definite boundaries encapsulating each piece of the network.

Network Map
Now lets take a look at the network map of the organization you are going to defend today.

You are going to observe attacks from sub-net 2
(Others in team wont be seeing attacks from this sub-net)

Your Computer Network Map
The whole Network
Your Computer Network Map
Your Computer Network Map
Your Computer Network Map
Your Computer Network Map
To Do
Go back to the previous slide, and identify the following:

- Internet
- Servers accessible via internet
- Servers accessible internally
- Employee desktop machines

Cyber Attack
Cyber Security
There are rules and restriction on how the computers in public and private networks can access and exchange information.

However hackers/cyber attackers (The bad guys !) circumvent these rules and regulations to access private networks and private information without authorization or by faking authorization.

Cyber Attacks
Cyber attacks can originate from an external location or can sometimes originate from inside the organization (aka insider attack)
The attackers exploit security gaps, holes in software programs or use rogue software (aka malware) to launch attacks on computer devices.

Cyber Analyst
Cyber analysts are personnel who are constantly defending our computer networks from such hackers/attackers
They are like police/warriors of the Internet
In this experiment you are going to be a cyber analyst
They monitor pieces of networks (sub-networks) using special software programs to detect and respond to cyber attacks

Real world Cyber Analyst
Computer Virus (Malware)
Malware, short for malicious software, is any software used to disrupt computer operation, gather sensitive information, or gain access to private computer systems.
Malware is a general term used to refer to hostile or intrusive software.
Software & its vulnerabilities
When large complex software programs are developed it is very common that some parts of it are not well tested for security flaws.
Those are identified by hackers to exploit the program and instruct it to do unexpected things such as

- Retrieve sensitive information
- Create fake authorized users in the computers
- Create rogue connections from that device to an unknown location and exfiltrate sensitive information

Such flaws are called vulnerabilities of software

Quiz Time
Question 4
Each analyst will be monitoring the whole network

- Yes
- No

Quiz Time
Question 5
What do hackers do when they discover a vulnerability in the software?

- Reports the vulnerability

Exploits the vulnerability
Does nothing
Fixes the vulnerability

Type of Attacks
Type of Attacks
Now lets look at different types of attacks used by hackers/attackers
For each attack type we will see:
What the attack means
Method of attack
And what is achieved through the attack

Suspicious Email Message
Email message will be sent to the employees of the organization with:
Malicious web links to steal information
Malicious attachments (such as malware) to gain access to private network or disrupt the computer
Email asks to visit a fake or modified website to also steal information
Malicious emails can be just an isolated attack for stealing personal or financial information
Or it could be the first step to a large scale attack

Remote Login Attempt
An attacker tries to gain access to an internal system from a remote location.
Method:
Attacker uses brute force method to gain access to a private machine
Attacker will enter several usernames and passwords until they succeed.
Gaining access to a machine in a private network can lead to large scale attacks

Connection to an Unknown Host
A machine from within the company is trying to establish an unauthorized connection to a remote suspicious location
Method:
After gaining access to a machine, attackers create backdoor ports
Through these backdoor ports they establish connections to a remote location
And would transfer confidential information via the unauthorized connection to the remote location
File Access Attempt
Someone was trying to access or modify an important file.

Method:
Hacker gains access to a machines and attempts to modify one or more files in that machine
Modifying the contents of document type files can lead to loss of important information

Modifying the contents of software files can lead to malicious behavior
Note: Some of these attacks fail because the files are locked for modification and the contents of the file can only be read (aka read-only)

Buffer Overflow Attempt
Buffer is a type of computer memory
Method:

A vulnerable software program can sometimes be modified to overflow the memory allocated
Through memory overflow the program can instruct the machine to do undesirable things

Simple buffer overflow can lead to computer crashing
But sometimes the overflow can cause new programs to be executed which can lead to creating fake users, modifying files, etc.

Possible Malware
Malware is software used to disrupt computer operation, gather sensitive information, or gain access to private computer systems.
Method:

Installed via email attachment
Transferred via USB drive
Or embedded within another program or file

Malware behavior varies vastly. And most them are very disruptive.
Information Request Query
Hacker is trying to get information about a network, system, or software to initiate an attack.
Hackers use this kind of simple attacks to get information about the private computers and network
It is a kind of reconnaissance
They use the information to develop attacks later
Method:

Sends request to network devices to return information about its network

Possible Information Leak
A file, message or data which could be sensitive to the organization is exfiltrated from the company's private network.
Method:

After gaining access to a machine the attacker can transfer information through a variety of techniques (such as back door)

Loosing confidential information means losing intellectual property, national secrets, etc.
Loosing financial information can cause financial loss

Port Scan Attempt
The process of checking every port on a computer

It is used by hackers

- To identify open ports to gain access
- Learn about the software using the different ports to identify vulnerable software running in the machine

Method:

- Hackers scan the computer networks to identify soft spots in the network
- If a port is open they could try gaining access to the machine
- Similarly they can identify vulnerable software to exploit by scanning the ports

DDOS

DDOS or distributed denial of service attack. One hacker or a group of hackers can send tons of traffic from a lot of machines to bring down a particular service or a machine.

Why is it an attack:

- This is often used to disrupting a service or a computer
- Disrupting essential services can be disastrous
- Disrupting websites can be a loss to the company
- Disrupting Army networks can handicap the security of a nation

“Ping” is a data packet that is often exploited in this attack. “Ping” is used to determine if a remote computer is powered ‘on’ or ‘off’

Quiz Time

Question 6

An attacker is trying out several usernames and passwords on a system. Why ?

- To get information about the machine
- To gain remote access to the machine
- To transfer a file to the machine
- To identify if there any open ports in the machine

Quiz Time

Question 7

Attacker created an unknown user name with high privileges on a private machine. How ?

- By installing a malware
- By sending a suspicious email
- By a buffer overflow attack
- By remotely connecting to the machine

Quiz Time

Question 8

How can an attacker send a sensitive file from a private machine to a remote location?

- Using a software port
- Using a backdoor port
- Using an unauthorized connection
- All of the above

Cyber Report and Attack Observations

Cyber Reports

A cyber report is a collection of attacks observed by one cyber analyst from one sub-network.

In each team there are 3 cyber defense analysts and hence there will be 3 different reports of cyber attacks which have to be discussed and analyzed.

Attack Observation

An attack observation is the description of the attack observed in a particular sub-network.

Each attack observation in a report will have descriptions of

- Time of attack
- Details of source (IP address, Port) of attack
- Details of the machine (IP address) being attacked
- Type of attack
- Attack process description
- Details of any immediate response taken

Example Attack Observation

Time of attack: 9Am 10 April 2014
Source IP: 152.160.160.12
Destination IP: 185.10.10.34 (Employee machine) (Subnet 1)
Type of attack: Intrusion attempt/port scan
Unauthorized access to an employee machine (185.10.10.34) through a open port that allows remote access
Open port that allows remote access was detected
Open port allowed remote access to the aerielview service
Accessed from a remote host 152.160.160.12
Aerielview service shut down in response

Attack Diagnosis

As described before, reports contain attack observations from each sub-network.
But attacks identified from the individual sub-networks need to be further analyzed to identify if they are part of any large scale attacks and are not just isolated events.
Such an analysis will give a big picture view of the network

Isolated Attacks

Isolated attacks are attacks that are targeted at only one machine in one sub network
They are often less damaging

Large scale attacks involve attacks happening on different parts of the network launched by the same attacker/ group of attackers

Large scale attack are usually of three kinds:

- Same type of attacks from same source occurring at different parts of the network
- Same attacks at different sub-net
- Same attacks at different locations

Large scale attacks

Large scale attacks involve attacks happening on different parts of the network launched by the same attacker/ group of attackers

Large scale attack are usually of three kinds:

- Same type of attacks from same source occurring at different parts of the network
- Attack migrating from one part of the network to the other by using different exploits and attacks
- Attacks migrating at a very slow pace (Stealth attacks)

Large scale attacks

Large scale attacks often lead to dire consequences. Few Examples:

- Heavy loss (Ex: Cyber attack on Target which lead to large financial loss and defacement)
- Large scale disaster (Iranian Nuclear Plant attack – Was deemed as the first indication of cyber warfare)

Therefore it is very important to detect the presence of large scale attacks in one's network

Detecting Large scale attacks

Detecting large scale attacks is difficult

They often look as isolated events to the analyst looking at only one part of the network

However there will be subtle evidence in these seemingly isolated events that will indicate they are connected and that are part of a large scale attack.

Hence effective information sharing between analysts is necessary to see the big picture view of the attack

Detecting Large scale attacks

Detecting large scale attacks require team effort

Requires extensive discussion about each attack observed by each analyst

Each analyst describes the attack to the rest of team to help in finding the clues that indicates connections between different attack observations

Requires "Connecting the dots" effort

It's like solving a puzzle !

Cues to discover Large scale attack

There will be pieces of information in each observation that indicates similarity to observations in other team members reports

Sample cues

Similarities could be based on the following

Same source IP address
Same type of attack (Ex: Intrusion)
Similar time of attack (Ex 9:00Am)
Similar attack method
Information type / Files involved in the attack
Port number (Ex: 8081)
Type of destination machine (Eg.: Server, Desktop)

Example – same attack at different locations
Unknown remote access to an employee machine (185.10.10.34) through a port that allows remote access
Open port that allows remote access was detected
Open port allowed remote access to the aerielview service
Accessed from a remote host 152.160.160.12
Aerielview service shut down in response

Type of attack Intrusion attempt/port scan
Example – Migrating Attack

Quiz Time
Question 9
An actual isolated attack can be observed on only machine in one sub-network
Yes
No

Quiz Time
Question 10
Which of the following information can be used to identify similarities between reports ?
File involved in the attack
Time of attack
Port number used in the attack
All of the above

Quiz Time
Question 11
A large scale attack is
Attack migrating from one machine to other
Same kind of attack on several machines
Attack migrating at a slow pace
Same attack on the same machine over several days
Only 1,2,and 3
All of the above

Discussion Process

Before Discussion – Reading Report

You are expected to read and understand the information present in the report handed to you during the reading time.

The training provided will help you to understand the material

During Discussion –

Pick and Describe

Take turns and describe each attack from the report you read to your team members

Pick an attack observation of your choice

- Describe the attack in overall
- Specify the source of the attack (IP address, machine name)
- Specify the machine being attacked
- Say the time of attack
- And describe how the attacker carried out the attack

Example – same attack at different locations

Unknown remote access to an employee machine (185.10.10.34) through a port that allows remote access

Open port that allows remote access was detected

Open port allowed remote access to the aerielview service

Accessed from a remote host 152.160.160.12

Aerielview service shut down in response

During Discussion - Listen

When your team member is describing the attack

- Listen Keenly
- Look for cues which might indicate some connection
- Discuss further and make the connections between the individual attacks.

Remember

You wont be able to discover the large scale attacks by simply reading out the information

During the discussion you need to describe each attack observations in your individual report

Talk effectively

Listen keenly

Lead an effective discussion

Find the Attacks

Report the Attacks

Save the day !

Your Goals

Discuss all your individual observations thoroughly with your team members

Identify the large scale attacks

Report your team’s findings.

APPENDIX B

ATTACK REPORTS PRESENTED TO PARTICIPANTS

Mission 1Attack Evidences
Uncertified Software
Time of Attack: 3Am April 10 2014
Source IP: 158.97.97.204
Destination IP: 185.10.10.25
Type of attack Uncertified Software Installed
An uncertified software was installed on an employee machine with IP: 185.10.10.25
Details:

A software called "Corpusrecorder" was installed
Source of software: corpusrecorder.com (unverified site)
Antivirus reported it as suspicious and quarantined it

Malware/Mail
Time of Attack: 0934Am on April 10 2014
Source IP: 110.10.10.15
Destination IP: 185.10.10.201
Type of attack: Malware/Mail
A malware was detected in an employee machine
Details:

An employee (in managerial level) reported her desktop to have slowed down after downloading an email attachment called “stuxcom”.
A malware was detected in that attachment .
The malware was later quarantined

Port Scan
Time of Attack: 10Am April 10 2014
Source IP: 175.15.10.10
Destination IP: 185.10.10.X (all machines)
Type of Attack: Port Scan
A remote host (175.15.10.10) performed a port scan on all the desktop machines
Details:

The port scan was trying to identify machines running the “remote desktop” service.

InfoLeak
Time of Attack: 1025Am April 10 2014
Source IP: 185.10.10.45
Destination IP: 165.165.165.15
Type of attack: Information Leak
Information leak to a known rogue location 165.165.165.15 has been detected.
Details:

A port was opened by a software called "b2reader" (seems like a rogue software).
Then an unauthorized connection to 165.165.165.15 was created
Then a file named "password.txt" was transferred over that established connection

Malware
Time of Attack: 1115Am on April 10 2014
Source IP: 185.20.20.4
Destination IP: 185.10.10.14
Type of attack: Malware
Trojan detected on an employee machine (185.10.10.14).
Connection to the shared file server (185.20.20.4) from a local system (185.10.10.14) at 1115Am on April 10 2014.
A file called "breakpics.pdf" was downloaded from the file server (185.20.20.4)
File breakpics.pdf was detected to contain a trojan
Trojan quarantined

Malware
Time of Attack: 3Pm April 10 2014
Source IP: 165.165.165.12
Destination IP: 185.10.10.X
Type of attack: Malware
Employees of the organization received a malware from an email message.
Details:

The email was received from a blacklisted IP address 165.165.165.12 at 3pm on April 10 2014.
There was an attachment called "Fox.vid".
The software had an embedded malware which tried to install itself in the background.
Anti-virus quarantined it.

Privilege Escalation
Time of Attack: 3pm on April 10 2014.
Source IP: 172.15.15.10
Destination IP: 185.10.10.3
Type of attack: Privilege Escalation / Buffer Overflow
A remote machine (172.15.15.10) gained admin privileges on a local machine.
Details:

The attacker used port scan and found a vulnerable service called "syshost.exe".
Then launched a buffer overflow on "syshost.exe"
This lead to the creation of new user with admin privileges in the machine.
The user was deleted in response.

DDOS
Time of Attack: 5Pm April 10 2014
Source IP: 24.56.56.9-20
Destination IP: 185.10.10.X
Type of attack: Denial of Service
Several of our employer systems faced a denial of service attack at 5pm (April 10 2014).
Details:

The attack originated from rogue machines in the IP range 24.56.56.9-20
A flood of ping requests on an open port caused the attack.
Lead the systems to be unresponsive for several hours.

End of Report

Intrusion/local
Time of Attack: 0945Am April 10 2014
Source IP: 185.10.10.201
Destination IP: 185.20.20.3
Type of attack: Intrusion / File transfer
An unauthorized Intrusion from a local user machine was detected on the Internal server
Details:

User has no authorization to access the Internal Server
A file called "stuxcom“ was transferred from local machine 185.10.10.201 To internal server 185.20.20.3
The "stuxcom" program opened a port number 4522 on the web server
Port was closed after detection

Privilege Escalation
Time of Attack: 1015Am on April 10 2014
Source IP: 172.15.15.20
Destination IP: 185.20.20.4
Type of attack Privilege Escalation
A remote machine (172.15.15.10) gained admin privileges on the file server
Details

The attacker used port scan and found a vulnerable service called "syshost.exe".
Then launched a buffer overflow on "syshost.exe"
This lead to the creation of new user with admin privileges in the machine.
The user was deleted in response.

Unknown Connection
Time of Attack: 1103Am on April 10 2014.14
Source IP: 185.30.30.4
Destination IP: 185.20.20.4
Type of attack: Unknown Connection
An unauthorized connection from a database server was detected on the file server (185.20.20.4)
Details:

Database server (185.30.30.4) established an unauthorized remote connection to the shared file server (185.20.20.4).
Connection was established via port 8081
Then a file called "breakpics.pdf" was copied to the file server through this connection

Intrusion/File
Time of Attack: 1130Am on April 10 2014
Source IP: 156.156.156.10
Destination IP: 185.20.20.4
A brute force intrusion was attempted on the file server.
Details:

Remote machine tried to gain access on the file server using several login attempts using different usernames and passwords.
Successfully gained access using login id "admin" and "password" combination
Using the login a scan on the files in the system was initiated . The scan was stopped in response .

Malware
Time of Attack: 3Pm April 10 2014
Source IP: 165.165.165.12
Destination IP: 185.10.10.X
An email message with a malware attachment was detected. The message was sent from a blacklisted IP address 165.165.165.12 at 3pm on April 10 2014.
Details:

The email message was targeted to just our employees.
It came with an attachment called "Fox.vid".
The software had an embedded malware which tried to install itself in the background.
Anti-virus quarantined it.

File Access
Time of Attack: 3:30Pm April 10 2014
Source IP: 76.15.245.12
Destination IP: 185.20.20.2
Type of attack: File Access
File on the database server was attempted to be edited
Details:

File name root/security/sam.ph
The file has software program to get username and password from the server
File was locked for any modification. No known damage was done

DDOS
Time of Attack: 5pm April 10 2014
Source IP: 24.56.56.9-20
Destination IP: 185.20.20.4
Type of attack: Denial of Service
Rogue machines in the range 24.56.56.9-20 launched a denial of service attack on the Internal Database server.
Details:

A flood of ping requests on our internal Database Server was detected

The requests were immediately rejected by the server reducing the impact of the attack.

Information Request
Time of Attack: 0630Pm April 10 2014
Source IP: 225.153.160.62
Destination IP: 185.30.30.1
Type of attack: Information Request Query
A remote machine requested IP information about all internal server machines
Details:

A request from 225.153.160.62 to the gateway router to return IP table information about the internal servers
Request from 225.153.160.62 to conduct port scan
IP table was returned but port scan request denied

End of Report

Intrusion/Net
Time of Attack: 8Am April 10 2014
Source IP: 156.156.156.10
Destination IP: 185.30.30.1
Type of attack: Intrusion / Network scan
A brute force intrusion on the gateway machine was detected.
Details:

Remote machine tried to gain access on the gateway machine through several login attempts using different usernames and passwords.
Successful gained access using the login id "admin" and password "password" combination.
Using the login, the hacker tried to scan the network.

Privilege Escalation
Time of Attack: 10Am on April 10 2014
Source IP: 185.20.20.3
Destination IP: 185.30.30.4
Type of attack: Privilege Escalation / Buffer Overflow
The internal server (185.20.20.3) gained admin privileges on the database master server (185.30.30.4)
Details:

Several failed connections from 185.20.20.3 (via Port number:4522)
Led to buffer Overflow on the login service on Database server
A new User added on the database server due to buffer overflow

Information Leak
Time of Attack: 1017Am April 10 2014
Source IP: 185.30.30.3
Destination IP: 165.165.165.15

Type of attack Information Leak
Information leak to a remote location with IP: 165.165.165.15 has been detected.
Details:

Port 5621 was opened by a software called "b2reader“
Then a connection was established to 165.165.165.15 using the port 5621.
Then a file named "pass.txt" was transferred over this established connection

Suspicious connection
Time of Attack: 11Am April 10 2014
Source IP: 185.30.30.4
Destination IP: 185.20.20.4
A suspicious connection from the database server 185.30.30.4 was observed on the File server 185.20.20.4
Details:

A file from an unknown source (possibly from a USB stick) was installed on the database server 185.30.30.4.
This software installation caused a buffer overflow
Later a new user was created
The new user opened a back door port 8081
A new Connection attempted via back door 8081 to the shared file server (185.20.20.4).
The user and the connection deleted on detection

File Access
Time of Attack: 3Pm April 10 2014
Source IP: 40.40.40.12
Destination IP: 185.30.30.3
Type of attack: File Access
A remote user attempted to modify a read-only file on the webserver 185.30.30.3.
Details:

A remote machine (40.40.40.12) repeatedly attempted to modify a read only file "web.config".
The attempt was unsuccessful

Malware
Time of Attack: 3Pm April 10 2014
Source IP: 165.165.165.12
Destination IP: 185.10.10.X
Our employees received a harmful email message from a blacklisted IP address 165.165.165.12.
Details:

A Suspicious email message with attachment called "Fox.vid" was received by our employees
The software had an embedded malware which tried to install itself in the background.

It was detected by antivirus and was defended.

DDOS
Time of Attack: 5Pm April 10 2014
Source IP: 24.56.56.9-20
Destination IP: 185.30.30.1
Type of attack: Denial of Service
Rogue machines (in IP range 24.56.56.9-20) launched a denial of service attack on our gateway system (185.30.30.1).
Details:

A flood of ping requests on the border gateway system (185.30.30.1) was detected.

Response: Configured the router to drop the requests from the IP 24.56.56.9-20

DDOS
Time of Attack: 8Pm April 10 2014
Source IP: 125.125.10.10-25
Destination IP: 185.30.30.2
Type of attack: Denial of Service (DDOS)
A series of remote machines (in IP range from 125.125.10.10 to 125.10.10.25) launched a denial of service attack on the load balancer 185.30.30.2
Details:

A flood of ping connections was sent to load balancer system

But the pings were rejected by the router

End of Report

Mission 2 Attack Evidences
Intrusion
Time of Attack 9Am April 11 2014
Source IP 154.48.48.48
Destination IP 185.X.X.X
Type of Attack Intrusion
A brute force intrusion from a remote IP - 154.48.48.48 was detected on different machines on the network.
Several failed login failures was observed. The remote machine tried to gain access through several logins.
On further investigation, it was found that the remote system was successful in logging in to a router machine. Used the login id - "admin" and password- "starwars123".
Using the login, the attacker tried to copy the address tables in the router – Possibly to hop to other machines in the network.
Login ID and password was changed in response

Information Request Query
Time of attack 9:30Am April 11 2014
Source IP 128.128.128.15

Destination IP 185.30.30.1
Type of attack Suspicious Information Request Query
An unauthorized request to send IP addresses and port numbers of all systems in the network was detected
Unauthorized Source IP (128.128.128.15)
Gateways and routers received the request to send all IP address information
Request dropped

Port Scan
Time of Attack 9:45Am April 11 2014
Source 135.128.128.10
Destination 185.20.20.4 & 185.20.20.2
Type of attack Port Scan
File server and the internal database server in the sub-network was port scanned by a remote machine 135.128.128.10
Port scan was to identify all the systems running bingbar service

Uncertified Software
Time of Attack 11Am April 11 2014
Source IP Unknown
Destination IP 185.30.30.1
Type of attack uncertified software installed
An unknown software was installed on the gateway machine
Software called "confikergetter.exe" was installed.

Memory Corrupt
Time of Attack 12:15Pm April 11 2014
Source IP 175.45.65.65
Destination IP 185.30.20.3
Type of attack Memory Corrupted
Memory on the system was corrupted due to a suspicious script.
Found Script data on the system fdskdskhsdkhfdhfdshfkddskkdskd <script type textjavascript> function ex() for(i = 0; i<0; i++) (buffer2 += buffer;) document.title = buffer2;) <script>sfdsdhdhsdkhfsdfhk

Malware
Time of Attack 2pm April 11 2014
Source IP Unknown
Destination IP 185.30.30.1
Type of attack Malware Software
Possible malware was detected by anti-virus on the gateway machine (185.30.30.1).
Name of the of malware software – "adzap"
Quarantined by anti-virus

Connection Redirect
Time of Attack 4:15pm April 11 2014
Source IP 185.30.30.3
Type of attack Connection Redirect
All web requests to the main website file (aka the landing page) is being redirected to another page called "newlanding.html"
Webserver (185.30.30.3) machine was overloaded due to the several redirects
The redirects was later observed to have ended

Buffer Overflow
Time of Attack 1130pm April 11 2014
Source IP Unknown
Destination IP 185.30.30.1
Type of attack Buffer Overflow
A Buffer Overflow on the gateway machine (185.30.30.1) was detected.
It lead to the creation of a new user.
New user's login id "trackme" password "startrekfan"
The buffer overflow vulnerability was later patched and the new user created was removed in response

Intrusion
Time of Attack 9Am April 11 2014
Source IP 154.48.48.48
Destination IP 185.20.20.4
Type of Attack Intrusion
A brute force intrusion from a remote IP - 154.48.48.48 was detected on the file server.
Several failed login failures was observed. The remote machine tried to gain access through several logins.
On further investigation, it was found that the remote system was successful in logging in to file server. Used the login id - "admin" and password- "starwars123".
Using the login, the attacker tried to modify the files.
Login ID and password was changed in response

Port Scan
Time of Attack 9:45Am April 11 2014
Source 135.128.128.10
Destination 185.X.X.X
Type of attack Port Scan
Several machines in the sub-network was port scanned by a remote machine 135.128.128.10
Port scan was to identify all the systems running bingbar service

Intrusion
Time of Attack 9:55Am April 11 2014
Source IP 152.160.160.12

Destination IP 185.20.20.3
Type of attack Intrusion attempt/port scan
An unknown remote access to the internal server (185.20.20.3) using an open port was detected.
The open port allowed remote access to the aerielview service
The Aerielview service was shut down in response

Port Scan
Time of Attack 9:57Am April 11 2014
Source IP 145.138.138.10
Destination IP 185.20.20.X
Type of attack Port scan
Several machines in the subnet was scanned by 145.138.138.10.
The attacker was looking for the service “Aerielview”

Uncertified Software
Time of Attack 11Am April 11 2014
Source IP USB Drive
Destination IP 185.20.20.2
Type of attack uncertified software installed
An unknown software was installed on the database machine
Software called "confikergetter.exe" was installed.

Unknown Connection
Time of Attack 11Am April 11 2014
Source IP 185.20.20.3
Destination Machine 172.132.132.12-25
Type of attack Unknown Remote Connection
Detected several unauthorized connection requests to remote hosts in the ip range 172.132.132.12-25.
Connections were not established

File Access
Time of Attack 4:00pm April 11 2014
Source IP 185.10.10.21
Destination IP 185.30.30.3
Type of attack File Integrity
A file on the web server (185.30.30.3) was modified
Configuration file "web.config" was modified by an internal user (from ip 185.10.10.21)
Change made to file is untraceable
In addition to the change, a new file was also added to the webserver (185.30.30.3) –
New file name "newlanding.html"

File Access
Time of Attack 1140Pm April 11 2014

Source of Attack 185.30.30.1
Destination IP 185.30.30.3
Type of attack File Integrity
The main file (aka the landing page) of the organization's website was modified on the webserver.
The user modified the page logging in from the gateway machine (185.30.30.1)
login \"trackme\" password \"startrekfan\"
The "Default.html" file on the website was found to be modified
Intrusion
Time of Attack 9Am April 11 2014
Source IP 154.48.48.48
Destination IP 185.20.20.4
Type of Attack Intrusion
A brute force intrusion from a remote IP - 154.48.48.48 was detected on the file server.
Several failed login failures was observed. The remote machine tried to gain access through several logins.
Especially the login id "admin" and password- "starwars123" was attempted
No machines were compromised

Port Scan
Time of Attack 9:45Am April 11 2014
Source 135.128.128.10
Destination 185.10.10.X
Type of attack Port Scan
Several machines in the sub-network was port scanned by a remote machine 135.128.128.10
Port scan was to identify all the systems running bingbar service

Unknown Connection
Time of Attack 11Am April 11 2014
Source IP 185.10.10.X
Destination Machine 172.132.132.12-25
Type of attack Unknown Remote Connection
Detected several unauthorized connection requests from several machines to remote hosts in the IP range 172.132.132.12-25.
Connections were not established

Memory Corrupt
Time of Attack 12:15Pm April 11 2014
Source IP 175.45.65.65
Destination IP 185.10.10.4
Type of attack Memory Corrupted
Memory on the system was corrupted due to a suspicious script.
Found Script data on the system fdskdskhsdkhfdhfdshfkddskkdskd
<script type textjavascript> function ex() for(i = 0; i<0; i++) (buffer2 += buffer;)

document.title = buffer2;) <script>sfdsdhdhsdkhfsdfhk

Suspicious Email
Time of Attack 3pm on April 11 2014
Source IP 165.165.165.12
Destination IP 185.10.10.X
Type of attack Suspicious Email/Phishing
A possible phishing email message was detected.
The message was sent from a blacklisted IP address:165.165.165.12
The email message was targeted to just our employees.
The message contains a link to bank of America requesting login information

Unauthorized Transfer
Time of Attack 4:20pm April 11 2014
Source IP 185.30.30.4
Destination IP 185.10.10.21
Type of attack Unauthorized information transfer
Unauthorized user Information was transferred from the database server (185.30.30.4) to a local machine (185.10.10.21)
This occurs when users access the webpage "newlanding.html"
It was user information that was transferred to local machine (185.10.10.21)
Transfer to the local machine is unauthorized

Information Request Query
Time of Attack 9:30pm April 11 2014
Source IP - 165.165.165.12
Destination IP 185.10.10.X
Type of attack Suspicious Information Request Query
A Request to send information about all IP addresses of machines in the network was received by several machines in the sub-network
The request was masqueraded as a legitimate request.
Request flagged as suspicious because it originated from a blacklisted IP
Request was rejected in response

Malware/mail
Time of Attack 9Am on April 12 2014
Source IP 165.165.165.15
Destination IP 185.10.10.X
Type of attack Malware through mail
Employees received suspicious email messages
A Suspicious email message with link to their own company’s website ("Default.html") and requesting them to accept the new terms and conditions of the site
A rogue script on the website suspected of exfilterating user information to a remote location (165.165.165.15)