Situational Awareness in Computer Network Defense:

Principles, Methods and Applications

Cyril Onwubiko
Research Series Limited, UK

Thomas John Owens
Brunel University, UK

Information Science
REFERENCE

Managing Director: Lindsay Johnston
Senior Editorial Director: Heather Probst
Book Production Manager: Sean Woznicki
Development Manager: Joel Gamon
Development Editor: Myla Harty
Acquisitions Editor: Erika Gallagher
Typesetter: Jennifer Romanchak
Cover Design: Nick Newcomer, Greg Snader

Published in the United States of America by
Information Science Reference (an imprint of IGI Global)
701 E. Chocolate Avenue
Hershey PA 17033
Tel: 717-533-8845
Fax: 717-533-8661
E-mail: cust@igi-global.com
Web site: http://www.igi-global.com

Library of Congress Cataloging-in-Publication Data

Situational awareness in computer network defense: principles, methods and applications / Cyril Onwubiko and Thomas Owens, editors.
p. cm.
Includes bibliographical references and index.
ISBN 978-1-4666-0104-8 (hbk.) -- ISBN 978-1-4666-0105-5 (ebook) -- ISBN 978-1-4666-0106-2 (print & perpetual access) 1. Cyberinfrastructure--Security measures. 2. Computer networks--Security measures. 3. Computer security. 4. Situational awareness. I. Onwubiko, Cyril, 1972- II. Owens, Thomas.
TK5105.59S566 2012
005.8--dc23

2011043980

British Cataloguing in Publication Data
A Cataloguing in Publication record for this book is available from the British Library.

Chapter 8
Cyber Situation Awareness through Instance-Based Learning:
Modeling the Security Analyst in a Cyber-Attack Scenario

Varun Dutt
Carnegie Mellon University, USA

Cleotilde Gonzalez
Carnegie Mellon University, USA

ABSTRACT

In a corporate network, the situation awareness (SA) of a security analyst is of particular interest. The current work describes a cognitive Instance-Based Learning (IBL) model of an analyst's recognition and comprehension processes in a cyber-attack scenario. The IBL model first recognizes network events based upon events' situation attributes and their similarity to past experiences (instances) stored in the model's memory. Then, the model comprehends a sequence of observed events as being a cyber-attack or not, based upon instances retrieved from its memory, similarity mechanism used, and the model's risk-tolerance. The execution of the model generates predictions about the recognition and comprehension processes of an analyst in a cyber-attack. A security analyst's decisions in the model are evaluated based upon two cyber-SA metrics of accuracy and timeliness. The chapter highlights the potential of this research for design of training and decision support tools for security analysts.

INTRODUCTION

With the prevalence of WikiLeaks hacks and other threats to corporate and national cybersecurity, guarding against cyber-attacks today is becoming a significant part of IT governance, especially because most government agencies have moved to online systems (Sideman, 2011). In order to protect national cybersecurity, leaders from the Defense Department, NATO, and the European Union assembled in Brussels recently to discuss a plan to prevent, detect, defend, and recover

DOI: 10.4018/978-1-4666-0104-8.ch008

from cyber-attacks (Sideman, 2011). The leaders there agreed that existing cybersecurity measures were incomplete and decided to fast-track a new plan for cyber-incident response. Similarly, the Department of Homeland Security (DHS) has recently launched a national campaign called, "Stop|Think|Connect," aiming to cultivate a collective sense of cyber–civic duty among personnel in organizations and enterprises that help preserve cybersecurity (Lute & McConnell, 2011). The DHS' message begins with the following wisdom:

Senior management in each and every office, company and department, whether private or public, must take responsibility for the protection of its own systems and information, by fielding up-to-date security technology, training employees to avoid common vulnerabilities, and reporting cybercrime when it occurs. (Lute & McConnell, 2011, p. 1)

As 80%-90% of what individuals and the government do using the Internet today depend upon private corporate networks provided by organizations and enterprises (Sideman, 2011), according to DHS, corporate networks that ensure our cybersecurity have much bigger responsibilities than previously thought (Lute & McConnell, 2011). Thus, meeting the DHS' objectives in a corporate network requires cyber situation-awareness (SA), a three stage process which includes recognition (or the awareness of the current situation in the network); comprehension (or the awareness of malicious behavior in the current situation in the network); and projection (assessment of possible future courses of action resulting from the current situation in the network) (Endsley, 1995; Tadda, Salerno, Boulware, Hinman, & Gorton, 2006).

The ability of a corporate network to protect itself from a cyber-attack using cyber-tools and algorithms without any interventions from human decision-makers is still a distant goal (Jajodia, Liu, Swarup, & Wang, 2010). Thus, the role of human decision-makers in security systems is one that is crucial and indispensible (Gardner, 1987; Johnson-Laird, 2006).

In the absence of perfect cyber-SA tools to recognize, comprehend, and project about cyber-attacks (PSU, 2011), a key role in the cybersecurity process is that of a security analyst. The security analyst is a human decision-maker who is in charge of protecting the online operations of a corporate network (e.g., an online retail company with an external webserver and an internal fileserver) from threats of random or organized cyber-attacks. However, very little is currently known about the role of the cognitive processes of the security analyst (like memory, risk-tolerance, similarity etc.) that might influence the cyber-SA of the analyst and his ability to detect cyber-attacks in corporate networks under different scenarios (Jajodia et al., 2010; PSU, 2011). Also, currently there seems to be a big gap between how security analysts function in the real world according to their cognitive processes and how cyber-SA tools and algorithms function that intend to replace human analysts, sometime in the future (Jajodia et al., 2010; PSU, 2011). Due to these reasons, it becomes important to investigate the influence of cognitive processes of a security analyst on his cyber-SA in popular cyber-attack scenarios.

Past literature shows there has only been one known cognitive attempt, through an expert system called R-CAST, to understand the cognitive decision-making aspects about a security analyst's cyber-SA (Fan & Yen, 2007; Jajodia et al., 2010). The R-CAST is a team-oriented cognitive-agent architecture that is a computational implementation of Klien's Recognition-Primed Decision (RPD) model (Klien, 1989). R-CAST, being a computational implementation of RPD, is a rule-based system which requires *a priori* knowledge base about cyber-attacks in a scenario in which it makes decisions (Fan & Yen, 2007). The *a priori* knowledge base is used during a mental simulation in the RPD. In the mental simulation,

the R-CAST applies rules that are constrained by the cyber-attack scenario in which the R-CAST operates to determine the future courses of action (Fan & Yen, 2007). The cognitive approach taken in this chapter (more details below) does not incorporate dependencies about an existing knowledge-base and future courses of action as assumed in the R-CAST.

The main purpose of this chapter is to describe a cognitive model of the recognition and comprehension processes in a security analyst's cyber-SA. The model is based on Instance-Based Learning Theory (IBLT; Gonzalez, Lerch, & Lebiere, 2003). Furthermore, we evaluate the performance of the IBL model of the security analyst using two cyber-SA measures: accuracy and timeliness (Jajodia et al., 2010) on a popular simple cyber-attack scenario about an island-hopping attack (Ou, Boyer, & McQueen, 2006; Xie, Li, Ou, Liu, & Levy, 2010). IBLT is well suited to modeling the security analyst's decisions as the theory provides a generic decision-making process that starts by recognizing and generating experiences through interaction with a changing decision environment, and closes with the reinforcement of experiences that led to good decision outcomes through feedback from the decision environment. Unlike the R-CAST, the IBLT neither assumes a rule-based cognitive process nor needs an existing knowledge-base to choose future courses of action and make decisions; rather, experiences in IBLT are generated overtime as a result of interaction of an IBL model with its decision environment (e.g., a cyber-attack scenario).

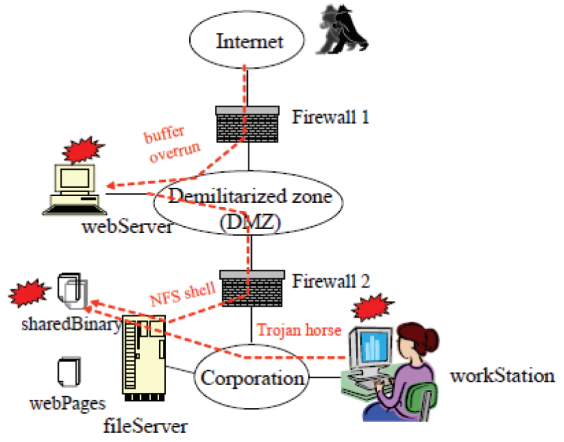
In the next section, we describe a popular cyber-attack scenario of an island-hopping attack in a corporate network. Then, we describe a model based upon IBLT that is used to make predictions about the cyber-SA of a security analyst in the scenario. Finally, we discuss the predictions from the IBL model and explain the implication of the model's predictions when designing training and decision support tools for security analysts.

A SIMPLE SCENARIO OF A CYBER ATTACK

The cyber-infrastructure in a corporate network typically consists of a webserver and a fileserver (Ou et al., 2006; Xie et al., 2010) that are protected by two firewalls in the Demilitarized zone (or DMZ) (where the DMZ separates the external network ("Internet") from the company's internal LAN network). The webserver handles customer interactions on a company's webpage. The fileserver is a repository for many workstations that are internal to the company and that allow company employees to do their daily operations. These operations are made possible by enabling workstations to mount executable binaries from the fileserver. An external firewall ('firewall 1' in Figure 1) controls the traffic between the Internet and the DMZ. The firewall 1's rules are configured to allow a bidirectional flow of the incoming "request" traffic and the outgoing "response" traffic between the Internet and company's webserver. Generally, an attacker is identified as a computer on the Internet and thus firewall 1 protects the path between the attacker's computer on the Internet and the company's website hosted by the webserver. Another firewall ('firewall 2' in Figure 1) controls the flow of traffic between the webserver and the fileserver (i.e., company's internal LAN network). Firewall 2 allows a Network File System (NFS) protocol access between the fileserver and webserver. For this cyber-infrastructure, most attackers follow a sequence of an "island-hopping" attack (Jajodia et al., 2010; pp. 30), where the webserver is compromised first, and then the webserver is used to originate attacks on the fileserver (through venerability in the NFS protocol) and other company workstations (by mounting executable binaries from the fileserver).

The security analyst is in charge of overseeing the cyber-infrastructure of the company (consisting on the two firewalls, DMZ, webserver, fileserver, and workstations) from cyber-attacks

Figure 1. A simple scenario of a cyber-attack. The attacker using a computer on the Internet tries to gain access of a company's fileserver indirectly through the company's webserver. Source: Xie et al. (2010).



originating from computers on the Internet. Ou et al. (2006) and Xie et al. (2010) defined a simple scenario of an island-hopping cyber-attack within this cyber-infrastructure. In the simple scenario, a security analyst is exposed to a sequence of 25 network events (consisting of both threat and non-threat events), whose nature (threat or non-threat) is not precisely known to a security analyst. Out of the total of 25 events, there are 8 predefined threat events in the sequence that are initiated by an attacker. The attacker, through some of these 8 events, first compromises the webserver by remotely exploiting vulnerability on the webserver and getting local access to the webserver. If the cyber-attack remains undetected by the 8th event, then the attacker gains full access to the webserver. Since typically in a corporate network and in the simple scenario, a webserver is allowed to access the fileserver through only a NFS event, the attacker then modifies data on the fileserver through the vulnerability in the NFS event. If the cyber-attack remains undetected by the security analyst by the 11th event, then the attacker gains full access of the fileserver. Once the attacker gets access to modify files on the fileserver, he then installs a Trojan-horse program (i.e., a malicious code) in the executable binaries on fileserver that is then downloaded and used by different workstations (event 19th out of 25). The attacker can now wait for an innocent user on workstation to execute the Trojan-horse program and obtain control on the machine (event 21st out of 25).

During the course of this simple scenario, a security analyst is able to observe all 25 events corresponding to file executions and the packets of information transmitted on and between the webserver, fileserver, and different workstations. He is also able to observe alerts that correspond to some network events using an intrusion-detection system (IDS) (Jajodia et al., 2010). The IDS raises an alert for suspicious file executions or suspicious packet transmission events that is generated on the corporate network. Among the alerts generated by the IDS here, there is both a false-positive and a false-negative alert, and one alert that correspond to the 8th event but is received by the analyst after the 13th event in the sequence (i.e., a time-delayed alert). Most importantly, due to the absence of a precise alert corresponding to a potential threat event, the analyst does not have precise information on whether a network event and its corresponding alert (from the IDS) are initiated by an attacker or by an innocent company employee. Even through the analyst lacks this precise information, he needs to decide, as early as possible and most accurately, whether the sequence of events in the simple scenario constitutes a cyber-attack. The earliest possible or proportion of timeliness is determined by subtracting the percentage of events seen by the analyst before he makes a decision to the total number of events (25) in the scenario from 100%. The accuracy of the analyst is determined by whether the analyst's decision was to ignore the sequence of events, or declare a cyber-attack based upon the sequence of observed network events.

BACKGROUND

We believe that a security analyst's accurate and timely classification of a sequence of network events as a cyber-attack or not (or analyst's cyber-SA) is based upon the following three factors:

1. The knowledge level of the analyst in terms of the mix of threat and non-threat experiences stored in analyst's memory.
2. The analyst's risk-tolerance level, i.e., the willingness of an analyst to classify a sequence of events as a cyber-attack.
3. The analyst's similarity model, i.e., the process that the analyst uses to compare network events with prior experiences that are stored in his memory.

Prior literature has shown that the cyber-SA of a security analyst is a function of *a priori* experiences in an analyst's memory about a cyber-attack scenario (Jajodia et al., 2010) and the analyst's risk-tolerance (McCumber, 2004; Salter, Saydjari, Schneier, & Wallner, 1998). Similarly, Dutt, Ahn, & Gonzalez, (2011) and Dutt & Gonzalez, (2011) have provided *a priori* predictions about the cyber-SA of a simulated analyst in an IBL model and demonstrated that these predictions are influenced by the experiences in memory of a simulated analyst and the risk-tolerance of the simulated analyst.

Recent research in judgment and decision making (JDM) has also discussed how our experiences of events in the environment shape our decision choices (Hertwig, Barron, Weber, & Erev, 2004; Lejarraga, Dutt, & Gonzalez, 2011). Typically, having a greater number of bad experiences in memory about an activity (e.g., a cyber-attack) makes a decision-maker (e.g., analyst) avoid the activity; whereas, good experiences with an activity boost the likelihood a decision-maker will underestimate the same activity (Hertwig et al., 2004; Lejarraga et al., in press).

Similarly, past research has found the role of similarity to be critical in problem solving, judgment, decision making, categorization, and cognition (Goldstone, Day, & Son, 2010; Vosniadou & Ortony, 1989). Essentially, two potential and competing models of human similarity judgments have been proposed. These models include the geometric model (Shepard, 1962a, 1962b) and the feature-based model (Tversky, 1977). In the geometric model, similarity between a pair of objects (e.g., a situation event in decision environment and an experience in memory) is taken to be inversely related to the distance between two objects' points in the space. The distance could be either a linear difference (linear-geometric) or a squared difference (squared-geometric) between two objects' points in the space (Shepard, 1962a, 1962b). In contrast, the feature-based similarity model characterizes similarity in terms of a feature-matching process based on weighting common and distinctive features between a pair of objects (Tversky, 1977).

Although there is literature that discusses the role of prior experiences of threats in general and the relevance of risk-tolerance in network security (Jajodia et al., 2010; McCumber, 2004; Salter, et al., 1998), it is difficult to find research that empirically investigates the role of both these factors together on a security analyst's cyber-SA. Similarly, although there is research that applies both models of similarity to human judgments in general (Goldstone, Day, & Son, 2010), research is needed that evaluates the effects of similarity models on the cyber-SA of a security analyst in cyber-attack scenarios.

The above three factors, as well as many other cognitive factors that may limit on enhance the cyber-SA of an analyst, can be studied through computational cognitive modeling. In this chapter,

we use IBLT to develop a model of the security analyst, and we assess the effects of the three factors (analyst's knowledge level, risk-tolerance, and similarity model) on the accuracy and timeliness of the analyst to detect a cyber-attack in the simple scenario.

INSTANCE-BASED LEARNING THEORY AND IBL MODEL OF THE SECURITY ANALYST

IBLT is a theory of how people make decisions from experience in dynamic environments (Gonzalez et al., 2003). In the past, computational models based on IBLT have proven to be able to generate *a priori* predictions of human behavior in many dynamic decision making situations like and including those faced by the security analyst (Dutt, Ahn, & Gonzalez, 2011; Dutt, Cassenti, & Gonzalez, 2010; Dutt & Gonzalez, 2011; Gonzalez & Dutt, 2010).

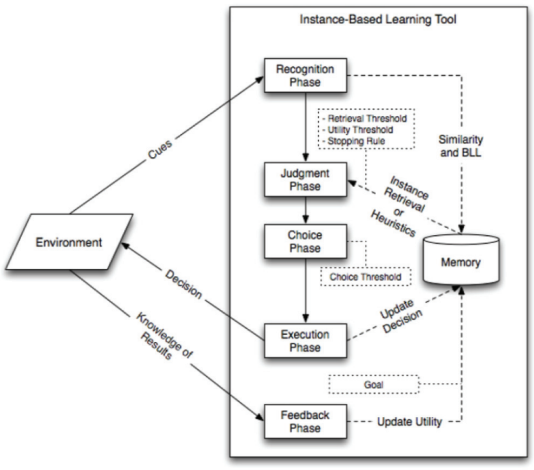
IBLT proposes that people represent every decision making situation as *instances* that are stored in memory. For each decision-making situation, an instance is retrieved from memory and reused depending on the similarity of the current situation's attributes to the attributes of instances stored in memory. An instance in IBLT is composed of three parts: situation (S) (the knowledge of situation attributes in a situation event), decision (D) (the course of action to take for a situation event), and utility (U) (i.e., a measure of the goodness of a decision made for a situation event).

In the case of the decision situations faced by the security analyst, these attributes are those that characterize potential threat events in a corporate network and that needs to be investigated continuously by the analyst. The situation attributes that characterize potential threat events in the simple scenario are the *IP* address of the location (webserver, fileserver, or workstation) where the event took place, the *directory* location in which the event took place, whether the IDS raised an *alert* corresponding to the event, and whether the *operation* carried out as part of the event (e.g., a file execution) by a user of the network succeeded or failed. However, as there are inherent uncertainties present in any scenario, one could think of other attributes that might characterize the simple scenario. Thus, we admit that the list of these four attributes might not be exhaustive and open to inclusion of other attributes or a different set of attributes. However, for the purpose of analysis in this chapter, we assume the above described four attributes to characterize the simple scenario.

In the IBL model of the security analyst, an instance's S slots refers to the situation attributes defined above; the D slot refers to the decision, i.e., whether to classify a sequence of events as constituting a cyber-attack or not; and, the U slot refers to the accuracy of the classification of an situation as a threat. IBLT proposes five mental phases in a closed-loop decision making process: recognition, judgment, choice, execution, and feedback (Figure 2). The five decision phases represent a complete learning cycle where the theory explains how knowledge is acquired, reused, and learnt by human decision-makers. Because the focus of this study is on the recognition and comprehension process in the SA of a security analyst, we will only focus on and discuss the recognition, judgment, choice, and execution phases in the IBLT (for details on the feedback phase refer to Gonzalez and Dutt (2010); and Gonzalez, Lerch, and Lebiere (2003)). In addition to the IBLT's decision-making process, IBLT borrowed some of the proposed statistical-learning mechanisms from a popular cognitive architecture called ACT-R (Anderson & Lebiere, 1998, 2003). Thus, most of the previous cognitive models that have used IBLT were developed for the ACT-R architecture.

The IBLT's process starts in the recognition phase in search for alternatives and classifies the current situation as *typical* or *atypical*. The current situation is typical if there are memories of similar situations (i.e., instances of previous trials that are similar enough to the current situation). If the

Figure 2. The five phases of IBL theory (right) and an environment, i.e., a decision task with which a model developed according to the IBLT interacts (left).



situation is typical, then the most similar instance is retrieved from memory in the judgment phase and is used to determine the expected utility of the situation being evaluated. In the IBL model, the decision alternatives refer to whether a sequence of events constitutes a cyber-attack or not. The actual determination of the utility is based upon the value in the utility slot of an instance retrieved from memory. The decision to retrieve an instance from memory for a situation event is based upon a comparison of the instance's memory strength, called *activation*. Thus, an instance is retrieved from memory if the instance has the highest activation among all instances in memory.

If the situation event in the network is atypical, then a judgment heuristic rule is applied to determine the utility of a new instance corresponding to a decision alternative. In the IBL model, we prepopulate the memory of a simulated analyst with certain instances to start with. These are assumed to be pre-stored experiences of past situations in the analyst's memory, and thus all situation events are treated by the model as typical.

Next, in the choice phase, a decision alternative is selected based upon the utility determined in the judgment phase (above). Thus, the choice phase in the IBL model consists of whether to classify a set of network events seen up to the scenario's current event as constituting a cyber-attack, or whether to accumulate more evidence by further observing incoming situation events before such a classification could be made. According to IBLT, this decision is determined in the "necessity level," which represents a satisficing mechanism to stop search of the environment and be "satisfied" with the current evidence (e.g., the *satisficing strategy*, Simon & March, 1958). We will call this parameter in the model, the "risk-tolerance level" (a free parameter) to represent the number of events the model has to classify as threats before the model classifies the scenario as a cyber-attack. For the risk-tolerance level, each time the model classifies a situation event in the network as a threat (based upon retrieval of an instance from memory), a counter increments and signifies an accumulation of evidence in favor of a cyber-attack. If the value of the accumulated evidence (represented by the counter) becomes equal to the analyst's risk-tolerance level, the analyst will classify the scenario as a cyber-attack based upon the sequence of already observed network events; otherwise, the model will decide to continue obtaining more information from the environment and observe the next situation event in the network. We manipulate the risk-tolerance parameter in this study at different number of events: 2, 4, or 6 (more details ahead). Regardless, the main outcome of the choice phase in the model is whether to classify a set of network events as a cyber-attack or not.

The model's choice phase is also based upon a property of the analyst to exhibit "inertia," i.e., simply not to decide to classify a sequence of observed network events as a cyber-attack due to lack of attention and continue to wait for the next situation event. The inertia in the model is governed by a free parameter called *probability of inertia* (Pinertia) (Gonzalez & Dutt, 2010;

Gonzalez, Dutt, & Lejarraja, 2011). If the value of a random number derived from a uniform distribution between [0, 1] is less than Pinertia, the model will choose to observe another network event in the scenario and will not classify the sequence of already observed events as a cyber-attack; otherwise, the model will make a decision to classify the observed events based upon the set risk-tolerance level. We assumed a default value of Pinertia at 0.3 (or 30%).

The choice phase is followed by the execution of the best decision alternative. The execution phase for the IBL model means either to classify a sequence of observed events as a cyber-attack and stop online operations in the company, or *not* to classify the sequence of events as a cyber-attack and to let the online operations of the company continue undisrupted.

In IBLT, the activation of an instance i in memory is defined using the ACT-R architecture's activation equation:

$$A_i = B_i + \sum_{l=1}^{k} P_l \times M_{li} + {}_i \qquad (1)$$

where, i refers to the i^{th} instance that is pre-populated in memory where i = 1,2, …, Total number of pre-populated instances; and, B_i is the base-level learning parameter and reflects the recency and frequency of the use of the i^{th} instance since the time it was created, which is given by:

$$B_i = \ln\left(\sum_{t_i \in \{1,\ldots,t-1\}} \left(t - t_i\right)^{-d}\right) \qquad (2)$$

The frequency effect is provided by $t-1$, the number of retrievals of the i^{th} instance from memory in the past. The recency effect is provided by $t-t_i$, the event since the t^{th} past retrieval of the i^{th} instance (in Equation 2, t denotes the current event number in the scenario). The d is the decay parameter and has a default value of 0.5 in the ACT-R architecture, and it is the value we assume for the IBL model of the security analyst.

The $\sum_{l=1}^{k} P_l \times M_{li}$ term is the similarity component and represents the mismatch between a situation event's attributes and the situation (S) slots of an instance i in memory. And k is the total number of a situation's attributes that are used to retrieve the instance i from memory. In the IBL model, the value of k = 4, as in the simple scenario, there are 4 attributes that characterize a situation event in the network and that are also used to retrieve instances from memory. As mentioned above, these attributes are *IP*, *directory*, *alert*, and *operation* in an event. The match scale (P_l) reflects the amount of weighting given to the similarity between an instance i's situation slot l and the corresponding situation event's attribute. P_l is generally a negative integer with a common value of -1.0 for all situation slots k of an instance i. The M_{li} or match similarities represents the similarity between the value l of a situation event's attribute that is used to retrieve instances from memory and the value in the corresponding situation slots of an instance i in memory. In this chapter, the $\sum_{l=1}^{k} P_l \times M_{li}$ term has been defined by using both a squared-geometric similarity model and a feature-based similarity model (Shepard, 1962a, 1962b; Tversky, 1977). In the squared-geometric model the $\sum_{l=1}^{k} P_l \times M_{li}$ is defined as:

$$\sum_{l=1}^{k} P_l \times M_{li} = \sum_{l=1}^{4} -1 \times (l_i - l_{event})^2 \qquad (3)$$

However, in the feature-based similarity model, the $\sum_{l=1}^{k} P_l \times M_{li}$ is defined as:

$$\sum_{l=1}^{k} P_l \times M_{li} = \theta \times f(i \cap event) - \alpha \times f(i - event) - \beta \times f(event - i) \quad (4)$$

The similarity of instance *i* to situation event is expressed as a linear combination of the measure of the common and distinctive features. The term *f*(*i* ∩ *event*) represents the number of features that the four slots of instance *i* and the four attributes in a situation event have in common. The term *f*(*i*−*event*) represents the features in the instance *i*'s four slots that are missing from the four attributes in the situation event. The term *f*(*event*−*i*) represents the features of the four attributes in the situation event that are missing from the instance *i*'s four slots. Furthermore, θ, α, and β are weights for the common and distinctive components. We assumed default values of the weights and thus, θ=2, α =1, and β =1. The default value assumption is because it balances out the effects of the common features (1st term in Equation 4) and the uncommon features (2nd and 3rd terms in Equation 4). Thus, the default assumption is a safe assumption to make both from literature (Tversky, 1977) and because we make predictions about the working of an analyst where we don't know about the real behavior of an analyst.

In order to find the value of the $\sum_{l=1}^{k} P_l \times M_{li}$ term, the situation events' attributes and the values in the corresponding slots of instances in memory were coded using numeric codes. Table 1 shows the codes assigned to the SDU slots of instances in memory and the situation events' attributes in the simple scenario. The assumption of on these codes is made to yield a nontrivial contribution of the similarity term in the activation equation (Equation 1).

Due to the $\sum_{l=1}^{k} P_l \times M_{li}$ specification, instances that encode a similar situation to the current situation event's attributes, receive a less negative activation (in Equation 1). In contrast, instances that encode a dissimilar situation to the current situation event's attributes receive a more negative activation.

Furthermore $_i$, is the noise value that is computed and added to an instance *i*'s activation at the time of its retrieval attempt from memory. The noise value is characterized by a parameter *s*. The noise is defined as,

$$_{\mathrm{i}} = \mathrm{s} \times \ln\left(\frac{1-\eta_{\mathrm{i}}}{\eta_{\mathrm{i}}}\right) \quad (5)$$

where, η_i is a random draw from a uniform distribution bounded in [0, 1] for an instance *i* in memory. We set the parameter *s* in an IBL model to make it a part of the activation equation (Equation 1). The *s* parameter has a default value of 0.25 in the ACT-R architecture and we assume the default value of s in the IBL model of the security analyst.

IMPLEMENTATION AND EXECUTION OF THE IBL MODEL

The IBL model of the security analyst was created using Matlab software. The IBL model goes over a sequence of 25 network events in the simple scenario (Figure 1). The memory of a simulated analyst in the model was pre-populated with instances encoding all possible sequences of network events based upon values of events' attributes. Some of these instances contained a threat value as the utility and some did not (more information below). Unbeknownst to the model (but known to the modeler), out of the 25 events in the scenario (mentioned above), there are 8 pre-defined threat events that are executed by an attacker outside the company (Ou et al., 2006; Xie et al. 2010). For each event in the scenario, the IBL model uses Equations 1, 2, {3 or 4}, and 5 to retrieve an instance that is most similar to the encountered event. Based upon the value of

Table 1. The coded values in the slots of an instance in memory and attributes of a situation event

Attributes	Values	Codes
IP (S)	Webserver	1
	Fileserver	2
	Workstation	3
Directory (S)	Missing value	-100
	File X	1
Alert (S)	Present	1
	Absent	0
Operation (S)	Successful	1
	Unsuccessful	0
Decision (D)	Cyber-attack	1
	No Cyber-attack	0
Threat (U)	Yes	1
	No	0

the utility slot of a retrieved instance, the situation event is classified as a threat or not a threat. Depending upon the inertia mechanism and the risk-tolerance level of a simulated analyst in the model, a decision is made to classify a sequence of observed events as a cyber-attack and stop company's online operations, or to let the company continue its online operations (no cyber-attack).

The IBL model was executed for a set of 500 simulated analysts on the same simple scenario where each simulated analyst encountered 25 or less situation events in the network. For each of the 500 simulated analysts, we manipulated the mix of threat and non-threat instances in memory, i.e., experience of the analyst, the risk-tolerance level of the analyst, and the similarity model used by the analyst.

The mix of threat and non-threat instances in the model's memory could be one of the following three kinds: ambivalent analyst (Ambi): 50% of threat instances and 50% non-threat instances for each situation event in the scenario; an extra-careful analyst (Extra): 75% of threat instances and 25% of non-threat instances for each situation event in the scenario; and a less-careful analyst (Less): 25% of threat instances and 75% of non-threat instances for each situation event in the scenario. The risk-tolerance level of analyst was manipulated on the following three levels: low (2 events out of a possible 25 event need to be classified as threats before the analyst classifies a sequence of observed events as cyber-attack); medium (4 events out of a possible 25 event to be classified as threats before the analyst classifies a sequence of observed events as cyber-attack); and high (6 events out of a possible 25 event to be classified as threats before the analyst classifies a sequence of observed events as cyber-attack). Please note that the values of 2, 4, and 6 events for the risk-tolerance is a reasonable and balanced manipulation given that there are only 8 total threat events (whose threat identity is unknown to the model) in the scenario. Finally, the similarity model was manipulated at two levels and could be either squared-geometric (Equation 3) or feature-based (Equation 4).

We wanted to derive predictions of the effect of the above manipulations in the model upon the cyber-SA of the analyst. The cyber-SA of a simulated analyst was measured using the accuracy and timeliness of the analyst. The accuracy was evaluated using two different cyber-SA metrics, recall and precision, and the timeliness was evaluated in the model using a single timeliness cyber-SA metric (Jajodia et al., 2010). *Recall* is the percent of events correctly detected as threats out of the total number of known threat events observed by the model before the model stopped in the scenario (Recall is the same as hit rate in Signal Detection Theory; Jajodia et al., 2010). *Precision* is the percentage of events correctly detected as threats out of the total number of threat events detected by the model before it stopped in the scenario. *Timeliness* is 100% minus percentage of events, out of a total 25, after which the model stops in the scenario and classifies the scenario to be a cyber-attack (the timeliness could be defined as the number of events out of 25, but defining it as a percentage allows us to compare it to other

two cyber-SA measures). A point to note is that if the model is unable to stop before the 25 events elapse in the scenario, then the denominators of the above cyber-SA metrics equal 25.

For both similarity models, we expected best performance for the IBL model representing an extra-careful analyst with a low risk-tolerance, and the worst performance for the IBL model representing a less-careful analyst with a high risk-tolerance. This fact is because an extra-careful analyst with a low risk-tolerance will be classifying network events more cautiously compared to a less-careful analyst with a high risk-tolerance. Also, as both similarity models, squared-geometric and feature-based, aim to search for the most similar instance in memory to a situation event in the simple scenario, we expect a similar performance in the IBL model for both similarity models.

RESULTS

Figure 3 shows the predictions of the cyber-SA measures (recall, precision, and timeliness) of an average security analyst from the IBL model due to the effects of manipulating the memory, risk-tolerance, and the similarity model used. First, for both similarity models, the effect of memory manipulation on cyber-SA measures (panel A and D) was stronger compared to the risk-tolerance measure (panel B and E). Thus, although there was a pronounced change in the three measures, recall, precision, and timeliness, as a result of the memory manipulation (Less, Ambi, and Extra), the change in the three measures was little due to the risk-tolerance manipulation (High, Medium, and Low). Furthermore, as per our expectation for both similarity models, an extra-careful analyst with a low risk-tolerance did better on all three performance measures compared to a less-careful analyst with a high risk-tolerance (panel C and F). Also, the precision was higher in the feature-based model compared to that in the squared-geometric model, but in general, the precision was less than the recall and timeliness in different manipulations. This latter observation is due to the fact that a model that has a greater recall and timeliness need not have a greater precision simultaneously. That is because it is not necessary that a model that is able to retrieve more threat instances from memory and rapidly, is able to retrieve them accurately for each situation event in the scenario (thus, there are chances of false-alarms).

Figure 3. The effect of experience (memory) on cyber-SA of an analyst in the squared-geometric similarity model (A) and in the feature-based similarity model (D). The effect of risk-tolerance on cyber-SA of an analyst in the squared-geometric similarity model (B) and in the feature-based similarity model (E). The interaction effect of memory and risk-tolerance on cyber-SA of an analyst in the squared-geometric similarity model (C) and in the feature-based similarity model (F). A greater percentage on all three cyber-SA measures, recall, precision, and timeliness is more desirable as it makes the simulated analyst more efficient.

DISCUSSION

In this chapter, we have shown that computational models based on the IBLT can be used to make predictions of a security analyst's cyber-SA in a cyber-attack scenario. Particularly, the model can make concrete predictions of the level of recall, precision, and timeliness of a security analyst given some level of analyst's experiences about network events (in memory), analyst's risk-tolerance, and the model that an analyst uses to compute similarity of network events with experiences in his memory.

We created an IBL model of the security analyst for a simple scenario of a typical island-hopping cyber-attack. The island-hopping attack portrayed in the simple scenario is one of the most common methods of cyber-attack in the real world (Ou et al., 2006; Xie et al., 2010). Then, using the simple scenario, we evaluated the performance of a simulated analyst on three commonly used

measures of cyber-SA. These measures are based upon accuracy of analyst (precision and recall) and the timeliness of the analyst to react to cyber-attacks (timeliness). Our results revealed that both the risk-tolerance level of an analyst and the mix of experiences of threat and non-threat instances in analyst's memory affect the analyst's cyber-SA; with the effect of the analyst's experiences (in memory) more impacting compared to risk-tolerance. One reason for the lesser impact of the risk-tolerance manipulation could be due to the nature and working of IBL models that are strongly dependent upon retrieval of instances from memory to make choice decisions. Another reason could be the presence of inertia in the model, which drives the model to observe more network events before the model could make a stop decision and where the risk-tolerance will only come to play a role in the model if the probability of inertia (set at 30%) is exceeded.

Also, when the simulated analyst is less careful, then for any situation event the model has only a 25% chance of retrieving threat instances from memory and a 75% chance of it retrieving non-threat instances from memory. As a consequence, the model has a lesser chance to classify actual threat events in the simple scenario as threats. Furthermore, it takes more time for the model to accumulate evidence that equals the risk-tolerance level that causes the model to make a decision in favor of a cyber-attack and stop work (decreasing the timeliness). However, when the simulated analyst is more careful, then for any situation event there is a 75% chance of the model retrieving threat instances and 25% chance of it retrieving non-threat instances. As a consequence, the model has a greater chance to classify actual threats in the simple scenario as threats and also takes less time to accumulate evidence that is equal to the risk-tolerance level (increasing the Timeliness).

The most important aspect of the model is the fact that although the recall and timeliness increase as a direct function of the model's ability to retrieve threat instances from the memory and its risk-tolerance, there is not a substantial increase in its precision when either of the two manipulations (memory and risk-tolerance) is favorable (Figure 3). The slow increase in precision is expected because a model that is able to retrieve more threat instances from memory and is less risk-tolerant might not necessarily be more precise in its actions. However, there is still an increase in precision with a manipulation of both memory and risk-tolerance and this suggests that making a security analyst less risk-tolerant as well as extra-careful might help increase his job efficiency. Because the IBL model is a process model that observes events, and makes decisions by retrieving experiences from memory, these are only some of the many predictions that the IBL model can make regarding the cyber-SA of human analysts.

Furthermore, although the current model is able to make *a priori* predictions, these need to be actually validated with human data. We plan to run laboratory studies in the near future to assess human behavior in this simple scenario. An experimental approach will allow us to validate our model's predictions and improve the relevance of the model and the assumptions made in it on its free parameters. In these experimental studies, we believe that some of the interesting factors to manipulate would include the experiences of the human analyst (stored in memory). One method we are currently considering is to make participants read or watch examples of more and less threatening scenarios before they participate in the act of detecting cyber-attacks in the simple scenario (i.e., priming the memory of the model with more or less threat instances as we did in the IBL model). Also, we plan to record the risk-seeking and risk-averse behavior of participants using popular measures involving gambles to control for the risk-tolerance factor (typically a

risk-seeking person is more risk-tolerant compared to a risk-averse person). Also, once we calibrate the current predictions of the model with an empirical study data (that we plan to collect in the future), we can evaluate the efficacy of different similarity models. Thus, our next goal will be to validate the predictions from the IBL model.

IMPLICATIONS FOR TRAINING AND DECISION SUPPORT OF SECURITY ANALYSTS

If our model is able to represent the cyber-SA of human analysts accurately, this model would have significant potential to contribute towards the design of training and decision support tools for security analysts. Based upon our current predictions, it might be better to devise analyst training and decision support that primes them to have experienced more threat rather than non-threat network events. The analyst's cyber-SA is also impacted by how tolerant he/she is to cyber-attacks. Thus, companies recruiting security analysts for network monitoring operations could measure the risk-seeking/risk-aversion character of a potential analyst (by using different risk-orientation measures that use gambles). Doing so would help evaluate a humans fit for the security analyst's position. Furthermore, although risk-orientation is a characteristic of a person (like his personality) that comes about as a result of his day-to-day experience and education, but there might be training interventions that could make analysts conscious of their risk-orientation or alter their risk-orientation. Based upon our results, making analysts less risk-tolerant (or more risk-averse) would help in increasing their efficiency in their job. Finally, based upon our results, training security analysts about the similar and dissimilar features between threats and non-threats in different cyber-attacks will benefit making analyst more precise on their job.

CONCLUSION

Due to the growing threat to our cyber infrastructure and the heightened need to implement cybersecurity, it becomes important to evaluate the cyber situation awareness (cyber-SA) of security analysts in different cyber-attack scenarios. In this chapter, we suggest a memory-based account, based upon instance-based learning theory, of the decisions of a security analyst who is put in a popular cyber-attack scenario of an island-hopping attack. Our results indicate that the cyber-SA of an analyst is a function of his memory of threat and non-threat events, his risk-tolerance, and the similarity methods he uses to compare network events to prior experiences of events in his memory. Based upon our predictions, it might be helpful to devise analyst job training that makes analysts cautious about the possibility of cyber threats, less risk-tolerant, and that enable them to look for features in attributes of network events that communicate the indication of potential threats.

ACKNOWLEDGMENT

This research was a part of a Multidisciplinary University Research Initiative Award (MURI; # W911NF-09-1-0525) from Army Research Office for a research project on Cyber Situation Awareness. We would like to thank Dynamic Decision Making Laboratory for providing the computing support for this chapter. Furthermore, we would like to thank Young-Suk Ahn, Dynamic Decision Making Laboratory for helping with compiling some of the results reported in this chapter and providing helpful comments. Finally, we are grateful to Hau-yu Wong, Dynamic Decision Making Laboratory for providing her editorial comments on this chapter.

REFERENCES

Anderson, J. R., & Lebiere, C. (1998). *The atomic components of thought*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Anderson, J. R., & Lebiere, C. (2003). The Newell test for a theory of mind. *The Behavioral and Brain Sciences*, *26*(5), 587–639. doi:10.1017/S0140525X0300013X

Dutt, V., Ahn, Y., & Gonzalez, C. (2011). *Cyber situation awareness: Modeling the security analyst in a cyber-attack scenario through instance-based learning*. Manuscript submitted for publication.

Dutt, V., Cassenti, D. N., & Gonzalez, C. (2010). Modeling a robotics operator manager in a tactical battlefield. In *Proceedings of the IEEE Conference on Cognitive Methods in Situation Awareness and Decision Support* (p. xx). Miami Beach, FL.

Dutt, V., & Gonzalez, C. (2011). Cyber situation awareness: Modeling the security analyst in a cyber attack scenario through instance-based learning. In *Proceedings of the 20th Behavior Representation in Modeling & Simulation (BRIMS) Conference*. Sundance, Utah, USA.

Endsley, M. R. (1995). Toward a theory of situation awareness in dynamic systems. *Human Factors Journal*, *37*(1), 32–64. doi:10.1518/001872095779049543

Fan, X., & Yen, J. (2007). R-CAST: Integrating team intelligence for human-centered teamwork. In

Gardner, H. (1987). *The mind's new science: A history of the cognitive revolution*. New York, NY: Basic Books.

Goldstone, R. L., Day, S., & Son, J. Y. (2010). Comparison. In B. Glatzeder, V. Goel, & A. von Müller (Eds.), *On thinking: Volume II, towards a theory of thinking* (pp. 103-122). Heidelberg, Germany: Springer Verlag.

Gonzalez, C., & Dutt, V. (2010). Instance-based learning: Integrating decisions from experience in sampling and repeated choice paradigms. *Psychological Review*, *118*(4).

Gonzalez, C., Dutt, V., & Lejarraja, T. (2011). *How did an IBL model become the runners-up in the market entry competition*? Manuscript in preparation.

Gonzalez, C., Lerch, J. F., & Lebiere, C. (2003). Instance-based learning in dynamic decision making. *Cognitive Science*, *27*(4), 591–635. doi:10.1207/s15516709cog2704_2

Hertwig, R., Barron, G., Weber, E. U., & Erev, I. (2004). Decisions from experience and the effect of rare events in risky choice. *Psychological Science*, *15*(8), 534–539. doi:10.1111/j.0956-7976.2004.00715.x

In *Proceedings of the2010IEEE/IFIP International Conference on Dependable Systems and Networks (DSN)* (pp. 211 - 220). Hong Kong, China: IEEE Press.

Jajodia, S., Liu, P., Swarup, V., & Wang, C. (2010). *Cyber situational awareness*. New York, NY: Springer.

Johnson-Laird, P. (2006). *How we reason*. London, UK: Oxford University Press.

Klein, G. A. (1989). Recognition-primed decisions. In Rouse, W. B. (Ed.), *Advances in man-machine system research* (*Vol. 5*, pp. 47–92). Greenwich, CT: JAI Press.

Lejarraga, T., Dutt, V., & Gonzalez, C. (in press). Instance-based learning: A general model of decisions from experience in repeated binary choice. *Journal of Behavioral Decision Making*.

Lute, J. H., & McConnell, B. (2011). *A civil perspective on cybersecurity*. Retrieved February 28, 2011, from http://www.wired.com/threatlevel/2011/02/dhs-op-ed/

McCumber, J. (2004). *Assessing and managing security risk in IT systems: A structured methodology*. Boca Raton, FL: Auerbach Publications. doi:10.1201/9780203490426

Ou, X., Boyer, W. F., & McQueen, M. A. (2006). A scalable approach to attack graph generation. In

Proceedings of the 13th ACM Conference on Computer and Communications Security (pp. 336–345). Alexandria, VA: ACM.

Proceedings of Twenty-Second AAAI Conference on Artificial Intelligence (pp. 1535 – 1541). Vancouver, British Columbia, Canada.

PSU. (2011). *Center for cyber-security, Information Privacy, and Trust*. Retrieved March 1, 2011, from http://cybersecurity.ist.psu.edu/research.php.

Salter, C., Saydjari, O., Schneier, B., & Wallner, J. (1998). Toward a secure system engineering methodology. In *Proceedings of New Security Paradigms Workshop* (pp. 2-10). Charlottesville, VA: ACM.

Shepard, R. N. (1962a). The analysis of proximities: multidimensional scaling with an unknown distance function: Part I. *Psychometrika*, *27*, 125–140. doi:10.1007/BF02289630

Shepard, R. N. (1962b). The analysis of proximities: Multidimensional scaling with an unknown distance function: Part II. *Psychometrika*, *27*, 219–246. doi:10.1007/BF02289621

Sideman, A. (2011). *Agencies must determine computer security teams in face of potential federal shutdown.* Retrieved March 1, 2011, from http://fcw.com/articles/2011/02/23/agencies-must-determine-computer-security-teams-in-face-of-shutdown.aspx

Simon, H. A., & March, J. G. (1958). *Organizations*. New York, NY: Wiley.

Tadda, G., Salerno, J. J., Boulware, D., Hinman, M., & Gorton, S. (2006). Realizing situation awareness within a cyber environment. In *Proceedings of SPIE Vol. 6242* (pp. 624204). Orlando, FL: SPIE.

Tversky, A. (1977). Features of similarity. *Psychological Review*, *84*, 327–352. doi:10.1037/0033-295X.84.4.327

Vosniadou, S., & Ortony, A. (1989). *Similarity and analogical reasoning*. New York, NY: Cambridge University Press. doi:10.1017/CBO9780511529863

Xie, P., Li, J. H., Ou X., Liu, P., & Levy, R. (2010). Using Bayesian networks for cyber security analysis.

ADDITIONAL READING

Busemeyer, J. R., & Diederich, A. (2009). *Cognitive modeling*. New York, NY: Sage.

Endsley, M. R. (2004). Situation awareness: Progress and directions. In Banbury, S., & Tremblay, S. (Eds.), *A cognitive approach to situation awareness: Theory, measurement and application* (pp. 317–341). Aldershot, UK: Ashgate Publishing.

Endsley, M. R., & Garland, D. J. (Eds.). (2000). *Situation awareness analysis and measurement*. Mahwah, NJ: Lawrence Erlbaum Associates.

Gonzalez, C., & Dutt, V. (2010). Instance-based learning models of training. In *Proceedings of the 54th AnnualMeeting of the Human Factors and Ergonomics Society (pp. 2319-2323)*. San Francisco, CA: Human Factors and Ergonomics Society.

Li, J., Ou, X., & Rajagopalan, R. (2009). Uncertainty and risk management in cyber situational awareness. Retrieved February 28, 2011, from http://www.hpl.hp.com/techreports/2009/HPL-2009-174.html

KEY TERMS AND DEFINITIONS

Cyber-Attack: Also known as cyber-warfare and is the use of computers and the Internet in conducting warfare in cyberspace.

Cyber-Situation Awareness: When a security accident occurs, the top three questions security administrators would ask are in essence: What has happened? Why did it happen? What should I do? Answers to these questions form the "core" of Cyber Situational Awareness.

Dynamic Decision-Making: The interdependent decision making that takes place in an environment that changes over time either due to the previous actions of the decision maker, or due to events that are outside of the control of the decision maker.

Instance-Based Learning Theory: A theory of how humans make decisions in dynamic tasks. According to the theory, individuals rely on their accumulated experience to make decisions by retrieving past solutions to similar situations stored in memory. Thus, decision accuracy can only improve gradually and through interaction with similar situations.

Intrusion-Detection System: A device or software application that monitors network and/or system activities for malicious activities or policy violations and produces reports to a security analyst.

Network Events: Events that take place over a network like opening of a file by a user on a workstation that resides on a remote server. These events could be further classified as threats (executed by a cyber-attacker) or non-threats (executed by a normal user of the network without any malicious intensions).

Security Analyst: A decision-maker who is in charge of observing the online operations of a corporate network (e.g., an online retail company with an external webserver and an internal fileserver) from threats of random or organized cyber-attacks.