SAND2014-1495C

Simulation of Workflow and Threat Characteristics for Cyber Security Incident Response Teams

Type name and affiliation information here

Within large organizations, the defense of cyber assets generally involves the use of various mechanisms, such as intrusion detection systems, to alert cyber security personnel to suspicious network activity. Resulting alerts are reviewed by the organization's cyber security personnel to investigate and assess the threat and initiate appropriate actions to defend the organization's network assets. While automated software routines are essential to cope with the massive volumes of data transmitted across data networks, the ultimate success of an organization's efforts to resist adversarial attacks upon their cyber assets relies on the effectiveness of individuals and teams. This paper reports research to understand the factors that impact the effectiveness of Cyber Security Incidence Response Teams (CSIRTs). Specifically, a simulation is described that captures the workflow within a CSIRT. The simulation is then demonstrated in a study comparing the differential response time to threats that vary with respect to key characteristics (attack trajectory, targeted asset and perpetrator). It is shown that the results of the simulation correlate with data from the actual incident response times of a professional CSIRT.

As illustrated by recent, high-profile attacks on familiar organizations (e.g. Google, RSA and Target), there exists an ongoing effort by groups with varying motivations to take advantage of vulnerabilities in cyber systems to advance their objectives (Coviello, 2011; Finkle & Hosenball, 2014; Zetter, 2010). In response, a substantial industry has arisen focused on research and development of software products to monitor and scan data networks, and detect events indicative of potential attacks. While such products are essential to an organization's ability to defend their cyber assets, the eventual resolution of incidents relies on cyber professionals to investigate and assess individual alerts, initiate appropriate responses and compile data from individual incidents to recognize larger patterns of events.

Within large organizations, the investigation and resolution of cyber incidents is often the responsibility of a Cyber Security Incidence Response Team (CSIRT). The composition of CSIRTs differ with respect to the number of cyber professionals, the levels and areas of expertise and the work processes and practices. Generally, the primary responsibility of a CSIRT is to review information from a variety of sources (e.g., intrusion detection systems, automated queries, user reports, notifications from other cyber professionals) to identify evidence of potential cyber threats. Once a potential threat has been identified, the CSIRT undertakes various activities to understand and assess the threat and initiate measures to resolve the incident. The corresponding tasks rely on general knowledge of computer and network systems and domain-specific knowledge of the local infrastructure, and adversary tactics and techniques, as well as various cognitive processes (e.g. inferential reasoning, pattern recognition, procedural memory, communication, etc.). Consequently, the ability of organizations to effectively defend their cyber assets depends upon the performance of individual cyber professionals and their ability to coordinate activities to function as effective teams.

Various capabilities have been reported to simulate information networks and their vulnerability to cyber attacks, at varying scales and levels of fidelity (Chi et al., 2001; Futoransky et al., 2009; Kuhl, et al., 2001; Van Leeuwen et al., 2010; Yun et al., 2005). LeBlanc et al. (2011) provide a review of several of the more prominent efforts in this regard. This includes models focusing on adversary tactics (Eom et al., 2008; Lee et al., 2005; Zakrzewska & Ferragut, 2011), as well as attacks by botnets (Kotenko, Kanovalov & Shorov, 2010). Approaches based on game theory have been demonstrated in modeling the interplay between attackers and defenders, including the inferences made by each side regarding the other (Hamilton & Hamilton, 2008; Shen et al., 2007; Shiva et al., 2010). Other efforts have involved relatively sophisticated models of the cognitive processes of adversaries and defenders. For example, Dutt, Ahn & Gonzalez (2013) reported a model implemented within the ACT-R framework that incorporated parameters for risk aversion and experience with threats. Kotenko (2005) described a simulation of the two-way interaction between attackers and cyber defenders that utilized an agent framework to capture the activities of teams. The individual agents monitored the network for suspicious activity and developed evolving plans to thwart the activities of the attackers based on models of each of the other agents on their team, and their respective beliefs.

Previous efforts to simulate the interplay between cyber adversaries and defenders have generally focused on scenarios in which an organization is undergoing an active attack, or otherwise, must cope with high demand situations. In contrast, the current paper describes a model-based simulation that accounts for both mundane day-to-day operations, as well as the response of a CSIRT to incidents with potential to result in substantial consequences. In this regard, it should be noted that the overwhelming majority of alerts investigated by cyber defenders are concluded to be the product of either legitimate user activity, false alarms generated by automated intrusion

detection algorithms or network technical problems that cannot be attributed to nefarious activities.

CSIRT Workflow Simulation

Analysts from the Sandia National Laboratories CSIRT were interviewed and based on these discussions, the workflow diagramed in Figure 1 was prepared. This workflow was then implemented within a discrete event simulation using MicroSaint Sharp. The following describes the model developed to simulate the workflow shown in Figure 1.

The model begins by generating a specified number of alerts. Each alert corresponds to an event that has triggered some form of automated network monitoring or other alerting function. This step is meant to simulate the experience where on any given morning, analysts arrive to find a queue containing some number of alerts that were either generated overnight or left unresolved from the previous day.

At the onset of a simulation, based on likelihood estimates derived from records of events encountered by the Sandia National Laboratories CSIRT, alerts are assigned an alert type (See Table 1), and a ground truth level of threat, which is an additive function of three threat characteristics: (1) trajectory of the attack; (2) the asset targeted; and (3) the perpetrator. Next, for each analyst, the perceived level of threat is calculated. The perceived level of threat varies for each analyst with the values randomly drawn from analyst-specific distributions in which the actual level of threat serves as the mean and the standard deviation is inversely related to the level of expertise assigned to the analyst. The perceived level of threat is further modified on the basis of two additional factors: (1) novelty, or how unique is the activity, and (2) recency, or the extent to which activity resembles other recent substantive threats. On average, experienced analysts are assigned a perceived level of threat that is closer to the actual level of threat than less experienced analysts. Thus, a less experienced analyst is more likely to either overestimate or underestimate the threat represented by a given alert.

Table 1. Types of alerts incorporated into the simulation.

Type of Alert
Legitimate Network Traffic
Network Technical Problem
Anomalous Firewall Activity
Anomalous Web Activity
Sandbox Activity
Beacon Activity
Email Attachment
Signature match

Once the threat associated with a given alert has been defined, the next step in the model simulates the process whereby individual analysts scan the queue of alerts and select an alert to open for investigation. The current model simulates a five-member CSIRT team, with each individual assigned an integer from 1-10 that reflects their level of expertise with each type of activity that might generate an alert. An individual analyst will appraise alerts on the basis of the perceived level of threat and their expertise, favoring alerts that pose a high level of threat and correspond to their area of expertise. Each analyst scans the list of unopened alerts until they reach an alert that exceeds an analyst-specific threshold of interest. If no alert exceeds their threshold, their threshold is lowered and they again scan the list.

Once an alert has been opened for investigation, it is determined which of thirteen tasks must be performed (See Table 2). Given the type of activity, the likelihood of each task being performed is based on data recorded from incidents encountered by the Sandia National Laboratories CSIRT. A specific software tool that would ordinarily be used to accomplish the task is assigned to each task. Individual analysts are assigned integer values from 1-10 to reflect both their level of expertise with a particular task and their expertise with the associated software tool. Expertise with tasks and tools serve as factors in determining the time required for an analyst to perform a given task and the effectiveness with which they will perform the task. Consequently, superior performance results when an analyst has both high expertise with a task and associated software tool. Likewise, intermediate levels of performance result when an analyst has expertise with a task, but not the associated software tool, or vice versa.

Once the task has been determined, there is next consideration of which analyst should do the work. This simulates the practice in which an analyst looks at an alert and realizes that a task needs to be performed for which another analyst is more skilled. However, there is an incumbency bias whereby given two analysts with similar levels of expertise, the analyst who opened the investigation will conduct the task. If it is determined that another analyst should perform a given task and that analyst is busy, the task goes into a queue and is worked once the analyst has completed their current task. Thus, at any given time, more experienced analysts may have numerous tasks waiting in their queue.

Completion of a task results in generation of evidence toward resolution of the alert. The time required to complete a given task and the evidence generated are both a function of the analyst's level of expertise with the task and the associated software tool. The threshold of evidence that must be reached to resolve an alert varies for each analyst based on the perceived level of threat attributed to an alert by each analyst. Following completion of a task, the total level of evidence is compared to the threshold. One of four outcomes may result: (1) the alert is correctly resolved; (2) the alert is erroneously resolved (i.e., false positive); (3) the alert is correctly unresolved; or (4) the alert is incorrectly unresolved (i.e., false negative). Where an alert is unresolved, there is a determination of the next task to be performed, with this process continuing until the alert is eventually resolved.

Table 2. Tasks incorporated into the simulation.

Tasks
Submit to sandbox
Submit to analysis
Retrieve machine proxy
Reverse engineer executable
Reverse engineer protocol
Retrieve forensics data
Analyze memory image
Retrieve network data
Retrieve email
Add network signature
Retrieve SSL keys
Implement network block
Implement additional alerts

Evaluation of the Model

To evaluate the model, a collection of 136 actual alerts were obtained from records generated by the Sandia National Laboratories CSIRT. These records included data concerning the nature of each alert and tasks performed by each analyst that worked on the alert. This allowed a determination of the time to resolve each alert, the number of analysts that worked on each alert and the number of entries, with entries corresponding to separate tasks performed in resolving alerts. Additionally, an experienced analyst from the Sandia National Laboratories CSIRT reviewed each alert and assigned values from 1-3 regarding characteristics of the associated threat (i.e., threat trajectory, targeted asset and perpetrator).

Threat characteristics are summaries of the alert data based on the MITRE Cyber Prep Methodology and threat attributes described by Mateski et al. (2012). The summaries are defined simply as a 3-pair describing the threat actions. Resource, family, and profiles are condensed by actions and context into each of the 3-pair attributes. The first attribute is an observed trajectory: (1) targeting no specific entity, (2) targeting a specific single entity, or (3) targeting multiple entities or high value entities. The second attribute is the affected or potentially affected asset set: (1) no asset, (2) a client or set of client assets, or (3) an infrastructure, service, or critical asset. The final attribute describes the threat potential: (1) a careless or unknown entity, (2) an action associated with criminal activity, or (3) an action associated with advanced theat.

Table 3 provides correlation results for the variables derived through the analysis of actual alerts. It can be observed that the three measures of the level of effort devoted to individual alerts (i.e., total time, number of entries and number of analysts) are each strongly related to one another. As might be expected, alerts requiring longer to resolve also have more entries indicating that more tasks are performed, and their resolution involves work by more analysts. Each of the threat characteristics were strongly associated with each measure of the level of effort required to resolve alerts. This suggests that for each threat characteristic, as the level of threat increases, a greater level of effort is expended to resolve the alert. There were positive relationships between the three threat characteristics, although these correlations were weaker than those found between the measures of the level of effort, with the association between the characteristics asset and perpetrator failing to reach statistical significance. A stepwise regression was performed to assess the contribution of threat characteristics to each measure of the level of effort required to resolve alerts. For total time, the only threat characteristic that attained statistical significance was asset (t=3.98, p<0.001, *R-Square*=10.65). With number of entries, all three threat characteristics attained significance (perpetrator – t=6.92, p<0.001; asset – t=3.38, p<0.001; trajectory – t=2.84, p<0.005; *R-Square*=41.26). Finally, for the number of analysts, perpetrator and asset were both statistically significant (perpetrator – t=6.36, p<0.001; asset – t=3.39, p<0.001, *R-Square*=30.85).

Evaluation of the model involved simulating a series of 136 alerts with threat characteristics equivalent to those of the actual events drawn from the records of the Sandia National Laboratories CSIRT. The simulation utilized an intermediate value of expertise (i.e., 5 on a scale of 1-10) for each of five analysts, with respect to the type of activity, tasks and software tools. For each alert, the simulation generated a value for the total time required to resolve the alert. There was a statistically significant correlation between the total times for the simulated alerts and the corresponding total times for the actual alerts (r=0.185, p<0.03). This result suggests that the mechanisms incorporated into the model to simulate the investigation of alerts produce a differential response to varying threat characteristics that is comparable to that occurring with actual events. However, it was also noted that the current model does not have mechanisms to account for the situation in which analysts suspend work on an alert for some period of time and then, later resume work on the alert based on having gained new insights or a lull in ongoing demands on their time. In the current model, once an investigation of an alert begins, it continues until the alert is resolved. It is believed that the correlation would have been even stronger had these mechanisms existed in the model.

Conclusion

The research described in the preceding sections provides an account of the workflow within a CSIRT and the manner in which varying threats differentially affect this workflow. In particular, it is proposed that analysts employ thresholds whereby a certain level of evidence must be attained before they are satisfied that they have a sufficient understanding of suspicious network activity to close an investigation. It may be further asserted that these thresholds are a function of characteristics of the corresponding threat.

The current analysis does not take into consideration the expertise of the analyst. It is conjectured that expertise influences the workflow in three ways. First, as analysts gain expertise, they more accurately assess the nature of threats and are better able to calibrate the level of effort devoted to an individual alert to the threat posed by the related activity. Consequently, inexperienced analysts are expected to either underestimate or overestimate the level of threat, and as a result, tend to commit insufficient resources to their investigation or continue investigations beyond the point of diminishing returns. Secondly, a richer understanding of tasks should allow analysts possessing greater expertise to perform those tasks more efficiently and productively. Third, greater knowledge of the procedures entailed in using software tools combined with a better conceptual knowledge of the application of the software tools should similarly result in superior efficiency and productivity. As a result, on average, the two latter factors should result in experienced analysts requiring less time to perform tasks and generating more evidence toward resolution of an investigation through their task performance.

Of the factors described above, experience with software tools is expected to become increasingly important for domains such as cyber security. With cyber security, and similar domains, a situation is arising in which the available data is so immense that it is unrealistic for an individual to learn to recognize meaningful patterns through the implicit learning processes that have traditionally been associated with attaining domain expertise (Klein, Calderwood & Clinto-Cirocco, 2010). Consequently, it is asserted that in these domains, a trend will emerge in which expertise is rooted in conceptual and practical understanding of software tools and the ability to effectively apply software tools to unique, and often, unexpected, circumstances.

Acknowledgement

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

References

Chi, S.D., Park, J.S., Jung, K.C. & Lee, J.S. (2001). Network security modeling and cyber attack simulation methodology. Information Security and Privacy *Lecture Notes in Computer Science*, *2119*, 320-333.

Coviello, A. W. (2011). Open letter to RSA customers. *RSA* [database online].

Dutt, V., Ahn, Y. S., & Gonzalez, C. (2013). Cyber situation awareness modeling detection of cyber attacks with instance-based learning theory. *Human Factors*, *55*(3), 605-618.

Eom, J.H., Han, Y.J., Park, S.H. & Chung, T.M. (2008). Active cyber attack model for network system's vulnerability assessment. *Proceedings of the International Conference on Information Science and Security*, Seoul Korea, 153-158.

Finkle, J. & Hosenball, M. (2014). Exclusive: More U.S. retailers victims of cyber attacks – sources. *Reuters*, *Sun January 12th*.

Futoransky, A., Miranda, F., Orlicki, J. & Sarraute, C. (2009). Simulating cyber attacks for fun and profit. *Proceedings of the 2nd International Conference on Simulation Tools and Techniques.*

Hamilton, S.N. & Hamilton, W.L. (2008). Adversary modeling and simulation in cyber warfare. *Proceedings of the 23rd International Information Security Conference, 278*, 461-475.

Klein, G., Calderwood, R., & Clinton-Cirocco, A. (2010). Rapid decision making on the fire ground: The original study plus a postscript. *Journal of Cognitive Engineering and Decision Making*, *4*(3), 186-209.

Kotenko, I. (2005). Agent-based modeling and simulation of cyber warfare between malefactors and security agents in internet. *Proceedings of the 19th European Conference on Modeling and Simulation.*

Kotenko, I., Konovalov, A. & Shorov, A. (2010). Agent-based modeling and simulation of botnets and botnet defense. *Proceedings of the Conference on Cyber Conflict,* Tallinn Estonia, 21-44.

Kuhl, M.E., Kistner, J., Costantini, K. & Sudit, M. (2001). Cyber attack modeling and simulation for network security analysis. *WSC '07 Proceedings of the 39th Conference on Winter Simulation, IEEE Press:* Piscataway NJ, 1180-1188.

LeBlanc, S.P., Partington, A., Chapman, I. & Bernier, M. (2011). An overview of cyber attack and computer network operations simulation. *Proceedings of the 2011 Military Modleing and Simulation Symposium*, 92-100.

Lee, J.K., Lee, M.W., Lee, J.S., Chi, S.D. & Ohn, S.Y. (2005). Automated cyber attack scenario generation using the symbolic simulation. *Artificial Intelligence and Simulation Lecture Notes in Computer Science*, 3397, 380-389.

Mateski, M., Trevino, C.M., Veitch, C.K., Michalski, J., Harris, J.M., Maruoka, S., Frye, J. (2012). *Cyber threat metrics*. Sandia National Laboratories.

Shen, D., Chen, G., Haynes, L. & Blasch, E. (2007). Strategies comparison for game theoretic cyber situational awareness and impact assessment. *Proceedings of the 10th International Conference on Information Fusion*, Quebec, Quebec, 1-8.

Shiva, S., Roy, S., Bedi, H., Dasgupta, D., & Wu, Q. (2010). A stochastic game model with imperfect information in cyber security. *Proceedings of the 5th International Conference on Information-Warfare and Security*.

Van Leeuwen, B., Urias, V., Eldridge, J., Villamarin, C. & Olsberg, R. (2010). Cyber security analysis testbed: Combining real, emulation and simulation. *Proceedings of the IEEE Carnahan Conference on Security Technology*, San Jose CA, 121-126.

Yun, J.B., Park, E.K., Im, E.G. & In, H.P. (2005). A scalable, ordered scenario-based network security simulator. *Systems Modeling and Simulation: Theories and Applications Lecture Notes in Computer Science*, *3398*, 487-494.

Zakrzewska, A.N. & Ferragut, E.M. (2011). Modeling of cyber conflicts using an extended Petri Net formalism. *Proceedings of the IEEE Symposium on Computational Intelligence in Cyber Security*, Paris, 60-67.

Zetter, K. (2010). Google hack attack was ultra sophisticated, new details show. *Wired Magazine*, *14*.

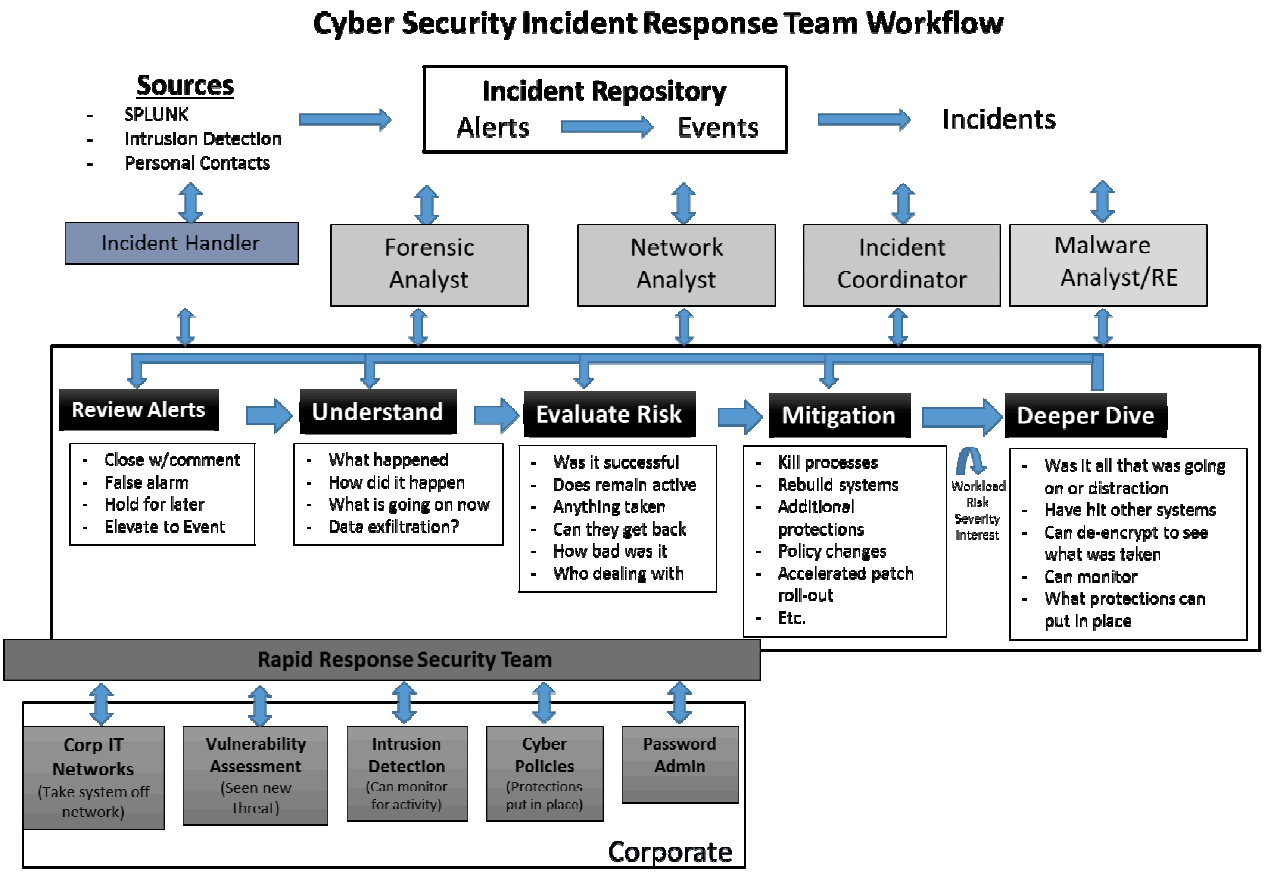

Figure 1. Workflow for Cyber Security Incident Response Team implemented in simulation.

Table 3. Correlation analysis of variables assessed from actual cyber security incidents

	Total Time	# Entries	# Analysts	Trajectory	Asset
# Entries	r=0.513 p<0.001				
# Analysts	r=0.524 p<0.001	r=0.860 p<0.001			
Trajectory	r=0.171 p<0.048	r=0.348 p<0.001	r=0.229 p<0.008		
Asset	r=0.326 p<0.001	r=0.352 p<0.001	r=0.311 p<0.001	r=0.241 p<0.005	
Perpetrator	r=0.171 p<0.048	r=0546 p<0.001	r=0.498 p<0.001	r=0.192 p<0.026	r=0.136 NS