Toward Automated Reduction of Human Errors based on Cognitive Analysis

Daisuke Miyamoto*, Takeshi Takahashi†
*The University of Tokyo, Tokyo, Japan, daisu-mi@nc.u-tokyo.ac.jp
†National Institute of Information and Communications Technology, Tokyo, Japan, takeshi_takahashi@ieee.org

***Abstract*—Following the immense development of cyber society where various activities including e-commerce take place, the demands for security is rapidly growing. Among major causes of security flaws is human error, which is unintentionally caused by humans. To cope with that, we intend to build a human error database that automatically develops further. We conducted a survey on human factors and concluded that the root causes of human errors are related to the internal mental processes, and the cognitive-psychological methodology is a feasible for the estimation of them. Based on that this paper proposes a framework that consists of data collection methods and data structure. It also explores the usability of the data by presenting use cases of human error prevention and incident handling.**

***Keywords*-Cyber Security, Human Error, Mental Workload, Cognitive Psychology, Ontology**

I. Introduction

The human factor has long been recognized as the weakest link in information security. Misjudgment often causes information leakage and malware infection, and misoperation damages valuable data, including personal information as well as business secret. Therefore, studies on human error becomes a hot topic in the field of cyber security.

The root causes of human error were analyzed with various aspects to protect assets from the errors. Particularly, principles of psychology are related to the analysis [1]; people underestimate risk, people have limited time and mental resources, security consequences are hard for people to assess since they are abstract and hypothetical and that losses are perceived as higher magnitude than gains.

The past studies primary challenged to understand, categorize and analyze the users behavior and their foundation. Their contribution illustrated that the organization must implement concrete security policies, e.g., Annex A. in ISO27001 [2], educate their employee to have knowledge of information security, and persuade them to follow the organization's best practice.

Unfortunately, lots of issue still need to be handled to cope with human errors. In some cases, researchers have tried to capture the cognitive models that users develop to understand the threats, with the goal of improving educational materials that users are more likely to understand. However, the effectiveness of education was limited to the small number of users; there can be much number of users who were uneducated. Moreover, in real life, security is rarely user's primary goal [3]. The user is primarily concerned with other tasks and hence, security cannot be foremost in the users' mind.

Our study aims at developing an automated system for thwarting the damage caused by human errors. The key idea is to store both users' operational history and their internal mental processes. To avoid human errors, we will develop the interface to support users for their make decision in regard to their mental model. As a first step, this paper discusses the suitable methodology for recording users' mental data. It then considers the construction of the resilient defense system.

The rest of this paper is organized as follows. Section II explains our related work, and describes our proposal in section III. Section IV present several use cases. Section V discusses the consideration for the system, and finally summarizes our contribution in section VI.

II. Related Work

This section provides the overview of failure analysis studies, with the special focus on human errors in the context of computer security, reported until now.

Failure analysis is the process of investigating the reason of failure. Its process also collects and analyzes data, and develops methods and/or algorithms to eliminate the root causes of the failure. Zahran et al.[4] summarized the categorization techniques for such analysis and introduced the component-based categorization; the failure can be caused of the components of information systems, namely, hardware, software, communications networks, people, data resources and organization. In case of the hardware errors, past researches have analyzed the failure mechanism causing specific devices and developed test models for hardware improvements. Preventing software errors, many models of the software development have been contributed, and the typical research vector is source code analysis.

The analysis of human error is also important part in failure study, as well as in cyber security. For instance, social engineering, one of the most severe threats in the cyber spaces, is used for both identity theft and malware infection in which the targets are human rather than computer systems. The research vectors against the attack are development of educational materials [5], [6], user interface for end users [7], [8], and detection methods [9]. Especially, the interface studies investigated the reasons of users' misoperations [10] and misjudgments [11]. Based on their subjects experiments, they clarified the mental model of users and indicated the way for improving the user interfaces.

In the context of the people in enterprise, human factors were analyzed to mitigate risks in the organization. According to Hawkey et al. [12], [13], challenges of IT security managements were classified into technical factors, organizational factors, and human factors. They also introduced that the lack of security training and culture, the difference of perception of risks, and "communication", that is, security responsibilities interact and communicate with other stakeholders within the organization. To understand human behavioral model, Parkin et al. [14] showed five behavioral foundation, namely cultural, ethical, temporal, mindset, and capability difference. Based on the foundation, they developed ontology which aims at maintaining compliance with ISO27002 standard [15] while considering the security behaviors of individuals within the organization.

Alfawaz et al. [16] classified the characteristics of organizational subjects involved in these information security practices. They analyzed the participants' activities and categorized individual security behaviors into four modes, (i) Knowing-Doing mode, (ii) Knowing-Not doing mode, (iii) Not knowing-Doing mode and (iv) Not knowing-Not doing mode. Term "Knowing" means that the participants know the organization's requirements for information security of behavior and have security knowledge. "Doing" also means that they are doing the right behavior. The cases of (i) and (iv) said that the participants (do not) know the requirements and (do not) have the knowledge, therefore, they are (not) doing the right behavior. The example of the mode (ii) is that the participant is unaware of the requirements, but asks someone before taking certain actions. The mode (iii) is serious, that the participants do not perform the right behavior even they know the requirements. In the context of the study of misuse [17], the mode (i) was labeled as "accidental" and the mode (iii) was labeled as "intentional", that is, deliberate ignorance of rules. Aspect from the risk management [18], the non-compliant people may possess a limited understanding of the security threats, but are more motivated toward immediate performance gains and hence, circumvent security policies or refuse to follow the organization rule. In addition, IT saboteurs, who are malicious and their goal is to disrupt the system, are analyzed aspect from the insider IT risk management [19].

The earlier researches can be summarized that understanding both the personal knowledge and his/her internal mental processes is necessary for thwarting the impact of human error; the organizations should educate their employee to have enough knowledge of security risks, and also persuade the non-compliant employee. The most predominant methodology is the development of the educational materials in the former cases, the analysis of the human behavior in the latter cases.

III. Automated mental data collection

Our motivation is to store personal operation records, as well as their mental model, in the automated manner. The difficulty lies in the process of collecting data on human's internal mentality. Past researches tended to employ questionnaire to people, but there still remains concern for false return. To the view of this, the non-repudiable methods for data collection are desired. We summarize the requirements in section III-A and illustrate the collection method in section III-B.

A. Requirement

To address the problem in collecting internal mental processes, this paper borrows three concept of the information security, namely confidentiality, integrity, and availability, for the description of our requirement.

REQ 1: Confidentiality

Mental data typically includes personal information, which is essentially privacy sensitive. Thus the use and sharing of such data needs to be careful. Organizations thus need agreement on the use of such information with their employees. The range of such information sharing must be controlled; for instance, a person under unhealthy mental condition can be identified by the employee's section, division, department, or the organization following the employment contract. Organizations need to customize the rules for individual employee. Note that this issue is beyond the scope of this paper.

REQ 2: Integrity

Regardless of the impact of the fear of negative evaluation, it can be naturally assumed that some of people will conceal their mistakes. In fact, disclosing mistakes often damage their own self-image and professional standing. Instead of the questionnaire-based methods, the proposed mechanism needs to employ the behavioral observation-based methods that estimate the mental model as the root causes of the human error.

REQ 3: Availability

The observation should employ the method which is easily applicable to people. It should not take much effort to start collecting data or disturb the handling of people during the tasks performance. Further more, people will not carry implants or needles or other devices which may hurt them in any way. The collected data also should be useful to analyze and hence, it should be formatted corresponding to the international standards.

B. Data collection

Based on the requirements explained in Section III-A, we explored the method for collecting mental data. In the

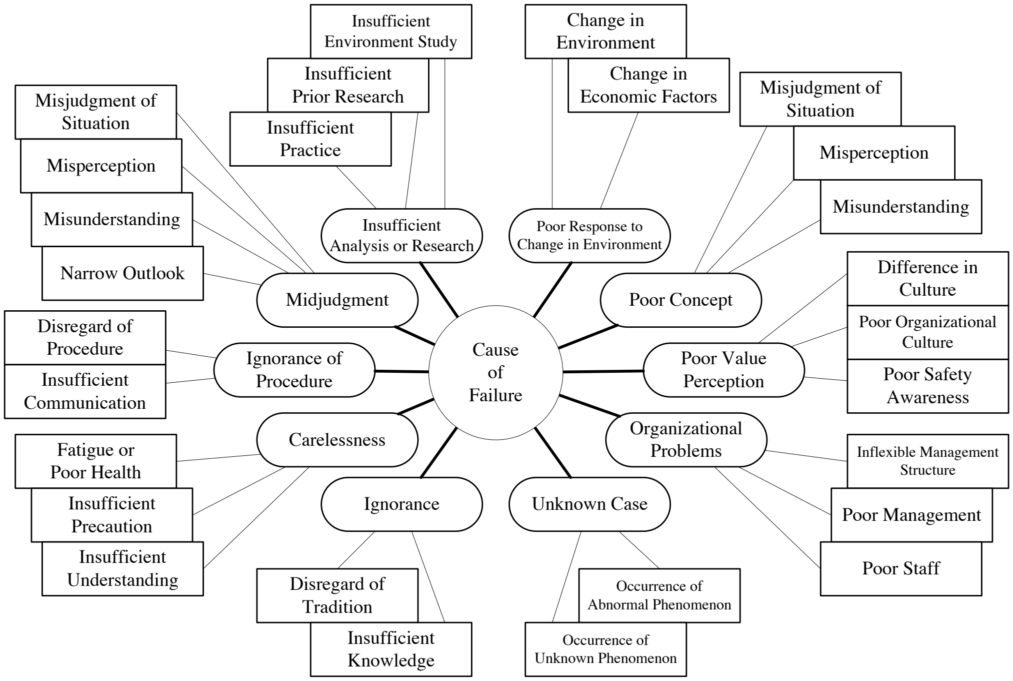

Figure 1. Cause-of-Failure map proposed by Hatamura (cited from [20])

context of cognitive psychology, it can be explored the internal mental processes by observation of human behavior. Herein, this paper introduces the following research domains that might be helpful for the observation.

Eye Movements

Research on experimental psychology has evidenced a strong link between eye movements and mental disorders [21], [22]. Leigh et al. [23] classified the eye movements into four categories – Saccades, Fixations, Smooth pursuit movements, and Vestibulo-ocular reflex. Generally, the saccadic eye movement changes with what they are seeing. In the context of mental model, it is reported that the mental rotation is suppressed during the movements [24], and that mental workload, which is the indicator of a person's mental/cognitive busyness, can be estimated with saccadic intrusions [25]. Note that the eye movements are linked with visual brain areas and hence, it does not reflect spatial relations in meta models [26].

Facial Skin Temperature

Variation of facial skin temperature has received some attention as a physiological measure of mental status [27]–[29]. According to Genno et al. [30], their experiments showed that temperature change in nose area when subjects experienced sensations like stress and fatigue. Further more, the thermography, when combined with other modes of measurement provides a highly automated and flexible means to objectively evaluate workload [27].

Aside from the above methods, brain activity [31], skin conductivity [32], heart measure, and blood pressure [33] are feasible due to the sensitivity to workload changes, but they tend to require much obtrusiveness for people [34]. In regard to the availability, the non-intrusive methodology is necessary.

IV. Utilization of collected data

This section discusses the utilization of the collected data that consists of operational and metal data. In the field of psychology, some research employed the statistical techniques, e.g., machine learning and cluster analysis, to develop an automatic analysis of eye tracking data [35]. Aside from such data oriented analysis, our study may be able to utilize the correlation of the two distinct data. Within the scope of the cyber security, the following sections show several use cases.

A. *Extension for Failure Knowledge Base*

The Failure Knowledge Base [20] is constructed by Hatamura et al, and mainly details failures made in the machinery, materials, chemical substance/chemical plant, and construction areas. There have been great responses

```
<?xml version="1.0" encoding="UTF-8"?>
<xsd:schema targetNamespace="http://.../mentalinfo"
 xmlns:xsd="http://www.w3.org/2001/XMLSchema"
 xmlns:incident-mentalinfo="http://.../mentalinfo"
 xmlns:ai="http://scap.nist.gov/schema/asset-identific
ation/1.1">

 <xsd:complexType name="EmployeeType">
  <xsd:sequence maxOccurs="unbounded">
   <xsd:element ref="Name"/>
   <xsd:element ref="Organization"/>
  </xsd:sequence>
 </xsd:complexType>

 <xsd:complexType name="MentalType">
  <xsd:sequence maxOccurs="unbounded">
   <xsd:element ref="Saccadic"/>
   <xsd:element ref="Fixations"/>
   <xsd:element ref="Temperature"/>
  </xsd:sequence>
 </xsd:complexType>

 <xsd:complexType name="TargetAsset">
  <xsd:complexContent>
   <xsd:extension base="ai:it-asset-type"/>
  </xsd:complexContent>
 </xsd:complexType>

 <xsd:complexType name="Operation">
  <xsd:sequence maxOccurs="unbounded">
   <xsd:element name="Saccadic" type="xsd:boolean"/>
   <xsd:element name="Fixations" type="xsd:boolean"/>
   <xsd:element name="Temperature" type="xsd:boolean"/>
  </xsd:sequence>
 </xsd:complexType>
</xsd:schema>
```

Figure 2. Preliminary schema for mental mode

from people not only in these areas, but also in the finance and insurance industries. Figure 1 summarizes the causes of failure, and shows that an individual has responsibility in the cases of Ignorance, Carelessness, Ignorance of Procedure, Misjudgment, and Insufficient Analysis or Research.

Our approach will extend this knowledge base by adding mental models in each record of the knowledge base. For instance, in the case of the ”Careless”, people cannot pay sufficient attention due to extreme busyness or poor physical condition. Provided the organization had such knowledge base and stored the mental model for each employee, it can realize the risk of human error when it has an employee whose mental mode is impaired. It then can notify the operation team before any incident is caused by the employee.

B. Incident Handling

In general, cyber security operations consist of three domains: Incident Handling Domain (IHD), IT Asset Management Domain (ITAMD) and Knowledge Accumulation Domain (KAD) [36]. IHD detects and responds incidents occurred in cyber society. Their set of the investigation are called forensics, including monitoring incidents, computer events that composes the incidents, and attack behaviors caused by the incidents. ITAMD runs cyber security operations inside each user organization such as installing, configuring, and managing IT assets in the organization. It covers both incident preventive operations and damage controlling operations in each organization. KAD researches cyber security-related information. Then it generates reusable knowledge for the other organizations and accumulates them.

Since the basic of forensics is log analysis, ITAMD usually requires recording the operation history to the assets; such history reveals the unwanted changes in the assets. Recall that our approach stores the operational log, as well as the mental model. When IHD investigate the root causes, the mental data will give hints to identify human-error.

For such incident handling, several international standards have been proposed [37]–[40]. Toward the development of cooperating these standards, preliminary schema for the mental model can be defined as shown in Figure 2. The set of the operator’s identifier, the operational target (assets), and the mental data all of which can be described. Within the schema, the operators mental can be identified as anomaly or not. Comparing this scheme with using numeric variable instead of Boolean is our future work.

V. Discussion

This section discusses the methodology for implementing our approach. This paper aims at developing automated systems for thwarting the damage caused by human errors, thus it showed how we will estimate the internal mental processes by collecting data based on the cognitive psychology. Hereafter, we describe our consideration for the implementation.

We explored the research domains in which the damage caused by human error is rapidly growing, and found that the network operations, especially backbone network operations, could not overcome the effect of human factor. Several case studies [41], [42] pointed that a routing problem designation reflects errors with the configuration or interaction of routing protocols (BGP), and the predominant problems stem from human error and misconfiguration of equipment. The recent report also said that BGP incidents are generally caused by human error and configuration errors [43].

Generally, network operators launch terminal application, login to some servers, and enter commands in which they often use command line interfaces. Due to that the context of the operation can be easily extracted by monitoring keyboard input events to the application, our system will collect both operation logs and mental data.

Aside from the context of network operations, cyber security operations are also related to human errors. In particular, the modern attack vector called advanced persistent threat often attempts to deceive people to infect malware. The observation of the personal mentality is feasible to find when people deceived.

In any contexts, toward the practical use, the indication of the risk might be considered. For convincing users to understand the human errors, the risk should be indicated so that users can easily understand. There exist the area of visualization study, where assorted studies such as [44] are reported, and these techniques need to be incorporated.

Investigation of the individual difference about the mental process is also important work item. The difference of the skills, motivations, knowledge among people will affect our systems to reduce human errors. Thresholds for anomaly judgment should differ for each person. In regard to these points as well as the work items explained in section IV, we will design and implement our systems to thwart the damages caused of human error.

VI. Conclusion

Everyone knows the phrase, "Don't drink and drive", which tells you that if you are drunk, never get behind the wheel. In a metaphorical sense, our approach is an installation of an alcohol analyzer to avoid drink-driving.

This paper introduced a framework that consists of data collection methods and data structure. Although it was difficult to completely eliminating the damage caused by human errors, our methodology aims at thwarting them in aspects from prevention and forensics. Based on our survey, this paper focused on estimating the internal mental processes as the root causes of the human error, and argued our future challenge based on the cognitive psychology.

This is merely a first step toward the establishment of a human error database that autonomously develops further, and various other schemes need to be incorporated. As a future work, we will implement its prototype and analyze its feasibility and clarify further issues.

References

[1] R. West, "The Psychology of Security," *Communications of the ACM*, vol. 51, pp. 34–41, 2008.

[2] British Standards Institution, "ISO/IEC 27001:2005 - Information Technology - Security Techniques - Information Security Management Systems – Requirements," 2005.

[3] A. Whitten and J. D. Tygar, "Why Johnny Can't Encrypt," in *Proceedings of the 8th USENIX Security Symposium*, August 1999.

[4] S. M. Zahran and G. H. Galal-Edeen, "A Categorization Technique for Resolving Information System Failures Reasons," *International Journal of Electrical and Computer Science*, vol. 12, no. 5, pp. 67–77, 2012.

[5] S. Sheng, B. Magnien, P. Kumaraguru, A. Acquisti, L. F. Cranor, J. I. Hong, and E. Nunge, "Anti-Phishing Phil: the design and evaluation of a game that teaches people not to fall for phish," in *Proceedings of the 1st Symposium On Usable Privacy and Security*, Jul. 2007.

[6] P. Kumaraguru, J. Cranshaw, A. Acquisti, L. F. Cranor, J. I. Hong, M. Ann, and T. Pham, "School of Phish: A real-World Evaluation of Anti-Phishing Training," in *Proceedings of the 5th Symposium On Usable Privacy and Security*, Jul. 2009.

[7] R. Dhamija and J. D. Tygar, "The Battle Against Phishing: Dynamic Security Skins," in *Proceedings of the 1st Symposium On Usable Privacy and Security*, Jul. 2005.

[8] Y. Oiwa, H. Takagi, H. Watanabe, and H. Suzuki, "PAKE-based Mutual HTTP Authentication for Preventing Phishing Attacks," in *Proceedings of the 17th World Wide Web Conference*, Apr. 2009.

[9] G. Xiang, J. Hong, C. Rose, and L. Cranor, "CANTINA+: A Feature-Rich Machine Learning Framework for Detecting Phishing Web Sites," *ACM Transactions on Information and System Security*, vol. 14, pp. 21–28, 2011.

[10] C. Birge, "Enhancing research into usable privacy and security," in *Proceedings of the 27th ACM international conference on Design of communication*, October 2009, pp. 221–226.

[11] F. Raja, K. Hawkey, and K. Beznosov, "Revealing hidden context: improving mental models of personal firewall users," in *Proceedings of the 5th Symposium on Usable Privacy and Security*, July 2009, pp. 1–12.

[12] K. Hawkey, D. Botta, R. Werlinger, K. Muldner, A. Gagne, and K. Beznosov, "Human, Organizational, and Technological factors of IT security," in *Extended Abstracts on Human Factors in Computing Systems*, April 2008, pp. 3639–3644.

[13] R. Werlinger, K. Hawkey, D. Botta, and K. Beznosov, "Security Practitioners in Context: Their Activities and Interactions with Other Stakeholders within Organizations," *International Journal of Human-Computer Studies*, vol. 67, pp. 584–606, 2009.

[14] S. E. Parkin, A. van Moorsel, and R. Coles, "An information security ontology incorporating human-behavioural implications," in *Proceedings of the 2nd international conference on Security of information and networks*, October 2009, pp. 46–55.

[15] British Standards Institution, "ISO/IEC 27002:2005 - Information Technology - Security Techniques - Code of Practice and Information Security Management," 2005.

[16] S. Alfawaz, K. Nelson, and K. Mohannak, "Information security culture: a behaviour compliance conceptual framework," in *Proceedings of the 8th Australasian Conference on Information Security*, July 2010, pp. 47–55.

[17] G. B. Magklaras and S. M. Furnell, "Insider Threat Prediction Tool: Evaluating the probability of IT misuse," *Computers and Security*, vol. 21, no. 1, pp. 62–73, 2002.

[18] S. Vidyaraman, M. Chandrasekaran, and S. Upadhyaya, "Position: The User is the Enemy," in *Proceedings of the Workshop on New Security Paradigms*, September 2007, pp. 75–80.

[19] S. R. Band, D. Cappelli, L. F. Fischer, A. P. Moore, E. D. Shaw, and R. F. Trzeciak, "Comparing insider it sabotage and espionage: A model-based analysis," Software Engineering Institute, CarnegieMellon University, Tech. Rep. CMU/SEI-2006-TR-026, December 2006.

[20] Y. Hatamura, “Structure and Expression of Failure Knowledge Database,” Available at: http://www.sozogaku.com/fkd/en/infen/mandara.html, March 2005.

[21] T. Crawford, S. Higham, T. Renvoize, J. Patel, M. Dale, A. Suriya, and S. Tetley, “Inhibitory control of saccadic eye movements and cognitive impairment in alzheimer’s disease,” *Biol Psychiatry*, vol. 9, no. 57, pp. 1052–1060, 2005.

[22] B. Noris, K. Benmachiche, J. Meynet, J.-P. Thiran, and A. Billard, “Analysis of Head-Mounted Wireless Camera Videos for Early Diagnosis of Autism,” *Advances in Soft Computing*, vol. 45, pp. 663–670, 2007.

[23] R. J. Leigh and D. S. Zee, *The Neurology of Eye Movements*, 4th ed. Oxford University Press, 1991.

[24] D. E. Irwin and J. R. Brockmole, “Mental rotation is suppressed during saccadic eye movements,” *Psychonomic Bulletin and Review*, vol. 7, no. 4, pp. 654–661, 2000.

[25] S. Tokuda, G. Obinata, E. Palmer, and A. Chaparro, “Estimation of mental workload using saccadic eye movements in a free-viewing task,” *Proceedings of the 33rd Annua International Conference of the IEEE Engineering in Medicine and Biology Society*, pp. 4523–4529, August 2011.

[26] J. F. Sima, M. Lindner, H. Schultheis, and T. Barkowsky, “Eye Movements Reflect Reasoning with Mental Images but do not with Mental Models in Orientation Knowledge Tasks,” in *Spatial Cognition VII, Lecture Notes in Computer Science*, vol. 6222, 2010, pp. 248–261.

[27] C. K. Ora and V. G. Duffyb, “Development of a facial skin temperature-based methodology for non-intrusive mental workload measurement,” *Occupational Ergonomics*, vol. 7, pp. 83–94, 2007.

[28] L.-M. Wang, V. G. Duffy, and Y. Du, “A composite measure for the evaluation of mental workload,” in *Proceedings of the 1st International Conference on Digital Human Modeling*, 2007, pp. 460–466.

[29] J. Voskamp and B. Urban, “Measuring Cognitive Workload in Non-military Scenarios Criteria for Sensor Technologies,” in *Proceedings of the 5th International Conference on Foundations of Augmented Cognition*, June 2009, pp. 304–310.

[30] H. Genno, K. Ishikawa, O. Kanbara, M. Kikumoto, Y. Fujiwara, R. Suzuki, and M. Osumi, “Using facial skin temperature to objectively evaluate sensations,” *International Journal of Industrial Ergonomics*, vol. 19, pp. 161–171, 1997.

[31] G. F. Wilson, “An analysis of Mental Workload in Pilots during flight using multiple psychophysiological measures,” *International Journal of Aviation Psychology*, vol. 12, pp. 3–8, 2002.

[32] A. Haag, S. Goronzy, P. Schaich, and J. Williams, “Emotion Recognition Using Bio-Sensors: First Steps Towards an Automatic System,” in *Proceedings of Affective Dialogue Systems, Tutorial and Research Workshop*, June 2004, pp. 36–48.

[33] S. Miyake, “Multivariate workload evaluation combining physiological and subjective measures,” *International Journal of Psychophysiology*, vol. 40, pp. 233–238, 2001.

[34] M. Grootjen, M. A. Neerincx, and J. C. van Weert, “Task Based Interpretation of Operator State Information for Adaptive Support,” ACI/HFES-2006, Tech. Rep., 2006.

[35] F. Galgani, Y. Sun, P. L. Lanzi, and J. Leigh, “Automatic Analysis of Eye Tracking Data for Medical Diagnosis,” in *Proceedings of the IEEE Symposium on Computational Intelligence and Data Mining*, March 2009, pp. 195–202.

[36] T. Takahashi, Y. Kadobayashi, and H. Fujiwara, “Ontological Approach toward Cybersecurity in Cloud Computing,” in *Proceedings of the 3rd international conference on Security of information and networks*, September 2010, pp. 100–109.

[37] The MITRE Corporation, “Common Weakness Scoring System,” Available at: http://cwe.mitre.org/cwss, June 2011.

[38] Forum of Incident Response and Security Teams, “Common Vulnerability Scoring System,” Available at: http://www.first.org/cvss, June 2007.

[39] National Institute of Standards and Technology, “The Asset Reporting Format,” Available at: http://scap.nist.gov/specifications/arf, June 2011.

[40] A. Montville, “Asset Identification,” Available at: http://www.ietf.org/id/draft-montville-sacm-asset-identification-00.txt, September 2012.

[41] C. Labovitz, A. Ahuja, and F. Jahanian, “Experimental Study of Internet Stability and Backbone Failures,” in *Proceedings of the 29th Annual International Symposium on Fault-Tolerant Computing*, June 1999, pp. 278–285.

[42] R. Mahajan, D. Wetherall, and T. Anderson, “Understanding BGP misconfiguration,” *SIGCOMM Computer Communocation Review*, vol. 32, no. 4, pp. 3–16, 2002.

[43] Internet Society, “Some Perspectives on Cybersecurity: 2012,” Available at: http://www.internetsociety.org/doc/some-perspectives-cybersecurity-2012, November 2012.

[44] T. Takahashi, K. Emura, S. Matsuo, and T. Minowa, “Risk Visualization and Alerting System: Architecture and Proof-of-Concept Implementation,” in *International Workshop on Security in Embedded Systems and Smartphones*, May 2013.