Proceedings of the INMM & ESARDA Joint Virtual Annual Meeting
August 23-26 & August 30-September 1, 2021

Perspectives of Worker's Human Data Collection to Prevent Insider Threat and Implications to Safety/Security Culture in Nuclear Facilities

Chul Min Kim
Department of Nuclear and Quantum Engineering, Korea Advanced Institute of Science and Technology

Man-Sung Yim*
Department of Nuclear and Quantum Engineering, Korea Advanced Institute of Science and Technology

*Corresponding author: msyim@kaist.ac.kr

ABSTRACT

As IoT technology expands and big data analytics become more active, more data will likely be collected at future nuclear facilities. Recently, image and video analysis and bio-signal data processing technologies are rapidly developing. In nuclear facilities, worker's behavior or bio-signal data may be used as supplementary indicators for security purposes. There exist various activities, including, but not limited to, fitness-for-duty (FFD) exams, trustworthiness evaluation, access control, and insider threat mitigation. However, the legal and ethical problems of personal information collection are likely to arise, and how these technologies can affect worker's safety and security culture has not yet been identified. This study summarized previous research using bio-signal data in nuclear facilities and proposed an integrated human data analysis system. It includes the considerations of how the data should be measured and analyzed to be applied in practice. In addition, the relationship between personal information collection and the safety/security culture is discussed.

INTRODUCTION

Insider threat is the most vulnerable part of nuclear security. System improvement has been made considering the insider attack, but workers' trustworthiness is still evaluated based on subjective factors. Current trustworthiness analysis focuses on finding motivation or personal characteristics that are evaluated to be related to the threat. Behavior observation is also performed by trained personnel or, like security culture, everyone is obliged to monitor everyone, it is difficult to provide objective information.

We suggested a framework for an integrated human data analysis system using human behavior and biosignal data that can be observed in nuclear facilities. First, we re-categorized the measures of insider threat mitigation and re-categorized the framework of trustworthiness analysis. Then we analyzed the opportunities of using biosignal to provide objective information to mitigate insider threat. We categorized what information can be extracted from behavioral data and biosignal in the existing trustworthiness analysis framework. Finally, for policy suggestions, we analyzed the legal and ethical problems related to biosignal collection. Finally, we proposed how the analyzed signals (or cues) should be utilized to positively affect safety/security culture and increase the acceptance of workers for human data collection and analysis.

INSIDER THREAT MITIGATION IN NUCLEAR FACILITIES

In nuclear facilities, insider is defined as a person with authorized access to items that an organization wishes to protect – information, people, and dangerous or valuable materials, facilities, and equipment [1]. Insider threat refers to the act of using such authority and knowledge to leak information, steal nuclear material, or sabotage facilities. It is perhaps the most serious challenge that nuclear security systems face, due to complacency of an organization ("Our employees are trustworthy!"), secrecy of security measures, and the small number of cases [2].

As summarized in Figure 1, insider threat mitigation system in nuclear facilities could be categorized in two ways: system-based approaches and human-based approaches. System-based approaches focus on improving the physical/cyber security system to detect, delay and respond against the insider threat effectively. It includes updating design basis threats (DBT) to consider insider attacks and managing the confidentiality of facilities by effective compartmentalization. On the other hand, human-based approaches focus on evaluating and maintaining the trustworthiness of workers to deny and remove malicious insiders. It includes evaluating the personality, mental status, and background of workers ("individual characteristics-based evaluation") and behavioral observation ("behavior-based evaluation").

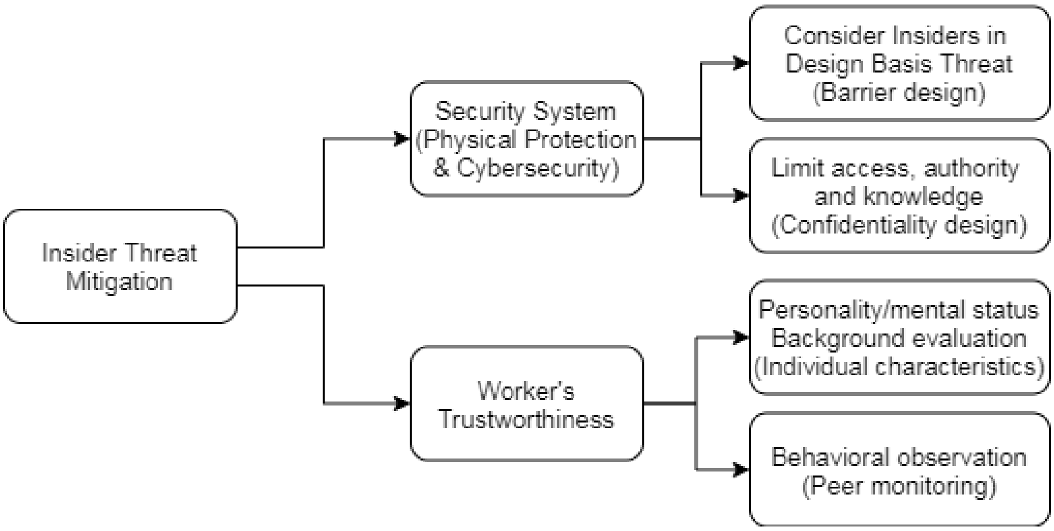
Insider Threat Mitigation
Security System (Physical Protection & Cybersecurity)
Worker's Trustworthiness
Consider Insiders in Design Basis Threat (Barrier design)
Limit access, authority and knowledge (Confidentiality design)
Personality/mental status Background evaluation (Individual characteristics)
Behavioral observation (Peer monitoring)

Figure 1. Categorizing the Means of Mitigating Insider Threat

Individual characteristics-based evaluation includes methods of not hiring people with specific personalities or psychological states or moving them to less sensitive positions to exclude insiders from the organization. A typical example is a dark triad, including narcissism, Machiavellianism, and psychopaths. After the recent emergence of the Islamic State (IS), the concept of violent extremists (VE) has also been proposed [3].

The evaluation also includes background checks by exploring the motivation of insider threat, as summarized in Figure 2. For example, if a person's deposit balance suddenly decreases, whether they are addicted to gambling will be investigated. Likewise, if someone suddenly takes a vacation and repeatedly travels to a particular country, further investigation is conducted to determine the motive.

Although this traditional approach is best, for now, there are fundamental limitations. Terrorists, for example, do not necessarily have personality problems. Moreover, because there are human rights issues, it is impossible to monitor and verify all motivations.

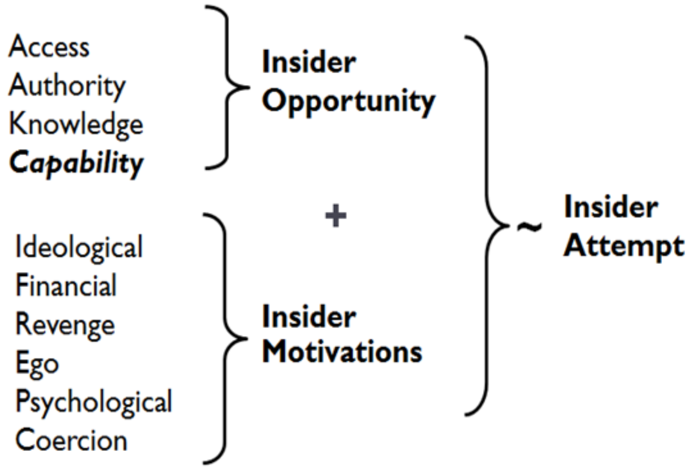

Figure 2. Opportunity and Motivations of Insider Threat [1]

The importance of behavioral observation has been emphasized to compensate for the limitations of individual characteristics-based evaluation. Behavior, rather than personality, could be a more direct indicator of trustworthiness. Behavioral observation defines and captures behaviors that potential insiders unconsciously take in daily life based on psychosocial modeling. Previous studies, especially in the field of cybersecurity, have defined representative psychosocial indicators of insider threat: Disgruntlement, not accepting feedback, anger management issues, disengagement, disregard for authority, low performance, stress, confrontational behavior, personal issues, self-centeredness, lack of dependability, and absenteeism [14].

Although many studies argue the importance of maintaining the workers' trustworthiness [4-9], most current policies focus on security system improvements [10-13]. The system-based approaches could provide practical improvements and are easy to be approved by the regulators. Also, the current behavioral observation method is difficult to overcome the limitations of the famous phrase, "Quis custodiet ipsos custodes?" (who guards the guardians?). A specific group should be designated, trained, and obligated to monitor the behavior of their colleagues. More recently, nuclear operators promote a security culture, emphasizing the need for everyone to monitor everyone actively. Typical examples of the policy include regular peer review and two-person rule. Other critical challenges include:

1) The assessments are subjective, potentially biased, and infrequently administered [15].
2) The assessments could be misused to kick out people – personal bias.
3) The assessments could be confused with performance evaluation.
4) Behaviors could be easily concealed.
5) It is difficult to doubt their colleagues (especially their boss). Even when insider activities are observed, they were mostly unreported [6].

OPPORTUNITIES OF BIOSIGNAL-BASED APPROACHES

Biosignals could be a tool that connects intentions to behaviors since it is difficult to conceal or manipulate one's biosignals. For this purpose, various researches have suggested the opportunities of biosignals to mitigate the insider threat. While the majority of studies focus on cybersecurity [16,17], several studies considered security-sensitive industries, especially nuclear facilities [15,18-20]. Their approaches could be divided into three categories – fitness-for-duty evaluation, deception detection, and malicious intention detection.

First, fitness-for-duty (FFD) is "to provide reasonable assurance that nuclear facility personnel are trustworthy, will perform their tasks in a reliable manner, are not under the influence of any substance, legal or illegal, that may impair their ability to perform their duties, and are not mentally or physically impaired from any cause that can adversely affect their ability to safely and competently perform their duties" [21]. Several studies showed that biosignals could help understanding the mental/physical status related to FFD.

For example, Suh and Yim classified five representative unhealthy states during eye close/eye open resting state – alcohol use, sleep deprivation, moderate/severe depression, moderate/severe anxiety, and heavy stress/workload – by using electroencephalography (EEG), heart rate variability (HRV), and other biosignals [17]. They achieved 97.7% accuracy for multi-class (6 states) problem, and 99.5% for binary class (healthy/unhealthy) problem with the Support Vector Machine (SVM) algorithm. In addition, Kim et al. classified fatigue, alcohol intake, and normal states during resting states and three tasks (memory matrix test, a chalkboard challenge, and the train of thought test) [20]. They achieved 81.8-100% accuracy using LSTM algorithm, indicating the feasibility of FFD classification during work.

The biosignal-based FFD classification could provide an effective screening system based on individual characteristics. However, an additional monitoring system should be needed to overcome the limitations of behavioral observation. The evaluation based on the biosignal has the potential to directly detect the concealed malicious intentions.

In this aspect, some studies have explored the possibility of biosignal-based deception detection. For example, Noonan et al. (2018) tried to correlate personality by word uses [8], Greitzer et al. (2012) and Fuller et al. (2013) tried to focus on the linguistic-based cues of deception [13,21], and Buller et al. (1996) and Moore et al. (2010) focused on the cues of nonverbal communication [22,23]. Also, some studies have predicted the likelihood of a crime by collecting these cues based on big data - what words are used, what their gestures are, whether their voice is shaking, and how they treat their co-workers. However, even these approaches are ultimately classified as assessing a person's trustworthiness rather than detecting maliciousness of specific behavior.

The other studies focused on the actual intention of the potential insiders. Suh and Yim (2018) used EEG signals to identify insider threats based on observing signature indicators while a potential insider threat is contemplating a malicious act [14]. Kim et al. (2020) also tried to detect malicious intention when reading the insider threat-related scenarios, using subject-wise classification [17]. The approach to detect intention directly using biosignal can be considered the closest to the original purpose of behavioral observation. However, there is a limit in that the noise increases because the biosignal is continuously measured for a long time while performing a mock crime or reading a scenario. Also, since the prediction results are highly likely to depend on the individual emotional states of the subjects, it is not easy to establish policies such as using the measurement results to exclude them from the work of the day. That is, as the number of false positives increases, the probability of an innocent employee being excluded from work instead of a malicious insider increases.

TOWARDS THE EFFECTIVE BEHAVIORAL OBSERVATION

In order to secure the objectivity of biosignal-based evaluation, it is necessary to measure the instantaneous physiological reaction. The simplest and most applicable method would be testing the participants' knowledge using repetitive short questions. If the questions are designed to induce initial recognition based on their knowledge and require false responses (lies) from guilty participants, the event-related potential (ERP) could be measured and analyzed to distinguish guilty from the innocents. This paradigm could be used only limitedly for the purpose of lie detection, such as criminal investigation. It is difficult to assume that someone has malicious intention by knowing any prohibited information or being aware of a crime-related object.

However, since workers in nuclear facilities have an obligation to report, the following behaviors could be assumed as cues of malicious intention: Concealing critical information related to facility safety or security despite knowing it; and falsely reporting personal background information required for trustworthiness checks. If the physiological responses indicate that the participant recognized a specific information but declared it as unknown, it can be suspected of having malicious intentions. Although this method can show the highest accuracy in predicting instantaneous intentions in real time through EEG, there has been no studies combined EEG with other biosignals, such as electrocardiogram (ECG) or electrodermal activity (EDA) - skin conductivity.

In order to design an effective question set, it is necessary to analyze the pathways of insider attack to sort out what information the malicious insider needs to hide. In this study, according to the authority of the insider, the attack pathway was categorized into three types: operators and regular employees, cooperation between insiders and outsiders, and secret organizations of outsiders during construction/maintenance periods. Most actions would be immediately detected on the system, but it could be assumed that insiders would know how to cause delayed sabotage without being detected. Such an action may cause the reactor to shut down after a long period of time or make it difficult to effectively respond in the event of an accident.

First of all, regular employees, may commit malicious actions. In particular, nuclear power plant operators and a small number of subcontractor employees have unescorted access to vital areas, making it easy to hide their actions. Representative actions include routine logging, misreporting values when performing numerical observations, deliberately pressing a different button, turning a switch, or turning a valve less or more. The second is a scenario where insiders and outsiders collaborate. Typically, there could be actions such as a security guard cooperating with an outsider, a janitor entering with a master key and turning the RCP pump switch, or an insider hiding an error even when an outsider enters to work and recognizes an error. The third is a situation in which the two-person rule is difficult to follow because everyone is busy during construction or maintenance. There may be terrorist infiltration, or some daily workers may steal or replace important parts, and someone may do something wrong and hide it.

Therefore, the integrated biosignal-based insider threat mitigation system would include three components. The first is to check if the worker is friendly with any person or thing. For example, it tests whether a worker knows a specific person's face and checks for secret organizations between insiders and outsiders or between outsiders. The second is to check if any confidential information has been leaked by presenting objects and situations that the worker should not be aware of. The third is an experiment to see if an error-related potential appears through a biosignal when an abnormal situation is observed and not reported during work. In other words, it examines violations of reporting obligations. In addition, it is possible to block the intrusion of unauthorized persons through continuous authorization using the collected biosignals. This can also be used as a concept of delay in physical protection.

POLICY SUGGESTION FOR BIOSIGNAL-BASED SYSTEM

How can we make the biosignal-based system to help enhance safety/security culture?

Although a small number of employees create insider threats, monitoring activities make the majority of innocent employees uncomfortable. Monitoring can increase employee stress, reduce commitment, and lower productivity [24]. Excessive monitoring may be perceived as a lack of trust, which may lower employee job satisfaction, and may lead to dissatisfaction in dismissing suspected employees [25]. In addition, an unintentional false report may harm the whistleblower. As a result, the insider threat may increase. This property is also well described in "Trust Trap" [26].

However, for employers, monitoring is worth the investment since the potential risk of insider threats outweighs the cost of monitoring. Employees could also welcome monitoring is welcome if it increases productivity and reduces the burden of behavioral observation and whistleblowing. The most important thing is trust. The monitoring process must be fully open to employees, explained, and fairly managed to maintain the trust. Therefore, appropriate privacy safeguards are required to utilize employee personal information, and monitoring should be done publicly as part of the performance measurement process.

A biosignal-based system can enhance security culture by using it to weaken insiders' motivation. In particular, the placebo effect of biosignal analysis cannot be ignored, which induces active reporting by employees. It was observed that even a group of graduate students with more than a certain level of knowledge took the fact that a lie detector was attached to their body very seriously and tried not to lie.

How should a biosignal-based system deal with the problems of personal information collection?

Ethics and information security issues regarding monitoring using biosignals have been steadily raised. However, even for sensitive technologies such as Drug & Alcohol (D&A) tests and stem cell experiments, R&D has been active as social consensus has been reached on the scope of use of the technology. This suggests that immediate social awareness may delay technology development, but it will eventually spread due to the utility of technology.

However, when restricting the employee's rights, it is essential to evaluate whether the reason for the restriction conforms to the legal provisions. On the other hand, it may be possible to limit basic rights to those engaged in hazardous work environments directly related to public life and safety. This is because they require a high level of concentration that is different from that of ordinary people. At this time, an objective balance should be made between their human rights and the benefits of monitoring.

For example, in some industries, it is possible to enter into an employment contract with the provision that a polygraph test is taken once every six months. In addition, the 2009 US DOE report "Predictive Modeling for Insider Threat Mitigation" mentions the Privacy and Ethical Concerns of monitoring activities as follows: "State agencies may only use this information under certain circumstances, but employers are being given the right to monitor employee cyber activity and share personal employee data internally." [27]

How should the data be measured and analyzed to be applied in practice?

In order to increase the acceptability of employees, a biosignal-based system for fitness-for-duty evaluation and human error mitigation, which many people are already familiar with, maybe introduced first. It is possible to collect biosignals to measure fatigue before entry and during work for all personnel, and at the same time use a method that selects some personnel in a similar way to drug testing and asks them to answer specific questions. The equipment should also be simplified as much as possible to increase convenience. Representative examples include a headband EEG device, an earplug EEG device, glasses-based eye trackers, and wrist band-based ECG/EDA sensors. Rather than increasing the number of channels, reducing the signal-to-noise ratio for practical comfort is necessary.

Even if the biosignal-based system prediction accuracy does not reach a level close to 100%, it is expected that the demand from the state and regulatory agencies will be sufficient. States, regulators, and operators require all three FFD evaluations, human error mitigation and insider threat mitigation. FFD requires high prediction accuracy, but some errors can be tolerated in the other two tasks. For example, in safety-related tasks, even if a false alarm sounds, it could be interpreted well to mean that additional confirmation is required before performing work. In the case of an insider threat, it could be used as an indicator to assist intelligence agencies in profiling. In particular, despite the risk of false positives in technology, the objectivity and cultural advantage of biosignal could be noted compared with human-based subjective evaluation. Also, legal and ethical issues could be overcome relatively easily in some emerging nuclear power countries. The higher the concern about nuclear security and the stronger the state has over individuals, the greater the room for the successful introduction of biosignal-based systems.

Furthermore, the biosignal-based system can be used in aircraft operations or military facilities. This can reduce the financial risk of the industry and improve public awareness by strengthening safety and security.

ACKNOWLEDGMENTS

This work was supported by the Nuclear Safety Research Program through the Korea Foundation Of Nuclear Safety (KoFONS) using the financial resource granted by the Nuclear Safety and Security Commission (NSSC) of the Republic of Korea (No. 2003023). This research was also supported by the KUSTAR-KAIST Institute, KAIST, Korea.

REFERENCES

1. International Atomic Energy Agency, IAEA Nuclear Security Series No.8 (Rev.1), Preventive and Protective Measures against Insider Threats: Implementing Guide. Vienna: International Atomic Energy Agency, 2008/2020.
2. Bunn, M., & Sagan, S. D. (2014, April). A worst practices guide to insider threats: lessons from past mistakes. Cambridge, MA: American Academy of Arts and Sciences.
3. World Institute for Nuclear Security, Countering Violent Extremism and Insider Threats in the Nuclear Sector, WINS International Best Practice Guide 3.8, 2020.
4. World Institute for Nuclear Security, Human Reliability as a Factor in Nuclear Security, WINS International Best Practice Guide 3.2, 2019.
5. Dominguez, D., & Duran, F. A. (2013). Security Modeling and Simulation to Address the Insider Threat (No. SAND2013-5211C). Sandia National Lab.(SNL-NM), Albuquerque, NM (United States).Chapter 5, "Preventing Insider Theft: Lessons from the Casino and Pharmaceutical Industries" by Matthew Bunn and Kathryn M. Glynn, was previously published in Journal of Nuclear Materials Management 41, no. 3 (2013): 4–16.
6. Greitzer, F. L., Strozer, J., Cohen, S., Bergey, J., Cowley, J., Moore, A., & Mundie, D. (2014, January). Unintentional insider threat: contributing factors, observables, and mitigation strategies. In 2014 47th Hawaii International Conference on System Sciences (pp. 2025-2034). IEEE.
7. Ho, S. M., Kaarst□Brown, M., & Benbasat, I. (2018). Trustworthiness attribution: Inquiry into insider threat detection. Journal of the Association for Information Science and Technology, 69(2), 271-280.
8. Noonan, C. F. (2018). Spy the Lie: Detecting Malicious Insiders (No. PNNL-SA-122655). Pacific Northwest National Lab.(PNNL), Richland, WA (United States).
9. Dominguez, D., & Duran, F. A. (2013). Security Modeling and Simulation to Address the Insider Threat (No. SAND2013-5211C). Sandia National Lab.(SNL-NM), Albuquerque, NM (United States).
10. Healey, A. N. (2016). The insider threat to nuclear safety and security. Security Journal, 29(1), 23-38.

11. Kim, K. N., Yim, M. S., & Schneider, E. (2017). A study of insider threat in nuclear security analysis using game theoretic modeling. Annals of Nuclear Energy, 108, 301-309.
12. Zou, B., Yang, M., Guo, J., Wang, J., Benjamin, E. R., Liu, H., & Li, W. (2018). Insider threats of Physical Protection Systems in nuclear power plants: Prevention and evaluation. Progress in Nuclear Energy, 104, 8-15.
13. Greitzer, F. L., Kangas, L. J., Noonan, C. F., Dalton, A. C., & Hohimer, R. E. (2012, January). Identifying at-risk employees: Modeling psychosocial precursors of potential insider threats. In 2012 45th Hawaii International Conference on System Sciences (pp. 2392-2401). IEEE.
14. Suh, Y.A.; Yim, M.-S. "High risk non-initiating insider" identification based on EEG analysis for enhancing nuclear security. Ann. Nucl. Energy 2018, 113, 308–318.
15. Almehmadi, A. Micromovement behavior as an intention detection measurement for preventing insider threats. IEEE Access 2018, 60, 40626–40637.
16. Hashem, Y., Takabi, H., Ghasemigol, M., Dantu, R., Towards insider threat detection using psychophysiological signals. In Proceedings of the 7th ACM CCS International Workshop on Managing Insider Security Threats, Denver, CO, USA, 12–16 October 2015; pp. 71–74.
17. Kim, J. H., Kim, C. M., & Yim, M. S. (2020). An investigation of insider threat mitigation based on EEG signal classification. Sensors, 20(21), 6365.
18. Suh, Y. A., & Yim, M. S. (2020). A Worker's Fitness-for-Duty Status Identification Based on Biosignals to Reduce Human Error in Nuclear Power Plants. Nuclear Technology, 206(12), 1840-1860.
19. Kim, J. H., Cho, Y., Suh, Y. A., & Yim, M. S. (2021). Development of an Information Security-Enforced EEG-Based Nuclear Operators' Fitness for Duty Classification System. IEEE Access, 9, 72535-72546.
20. U. S. NRC, "Fitness-for-Duty Programs," <https://www.nrc.gov/reactors/operating/ops-experience/fitness-for-duty.html>
21. Fuller, C. M., Biros, D. P., Burgoon, J., & Nunamaker, J. (2013). An examination and validation of linguistic constructs for studying high-stakes deception. Group Decision and Negotiation, 22(1), 117-134.
22. Buller, D. B., & Burgoon, J. K. (1996). Interpersonal deception theory. Communication theory, 6(3), 203-242.
23. Moore, N. J., Hickson, M., & Stacks, D. W. (2010). Nonverbal communication. New York: Oxford University Press.
24. Brown, K. A. (1996). Workplace safety: a call for research. Journal of operations management, 14(2), 157-171.
25. Shaw, E. D., & Fischer, L. F. (2005). Ten tales of betrayal: The threat to corporate infrastructure by information technology insiders analysis and observations. DEFENSE PERSONNEL SECURITY RESEARCH CENTER MONTEREY CA.
26. Band, S. R., Cappelli, D. M., Fischer, L. F., Moore, A. P., Shaw, E. D., & Trzeciak, R. F. (2006). Comparing insider IT sabotage and espionage: A model-based analysis. CARNEGIE-MELLON UNIV PITTSBURGH PA SOFTWARE ENGINEERING INST.
27. Greitzer, F. L., Paulson, P., Kangas, L., Franklin, L. R., Edgar, T. W., & Frincke, D. A. (2009). Predictive modelling for insider threat mitigation. Pacific Northwest National Laboratory, Richland, WA, Tech. Rep. PNNL Technical Report PNNL-65204.