SANDIA REPORT
SAND2020-8253
Printed Click to enter a date

Sandia National Laboratories

Phase I Closeout Report: *Invoking Artificial Neural Networks to Measure Insider Threat Mitigation*

Adam D. Williams, Shannon N. Abbott
Sandia National Laboratories/Center for Global Security and Cooperation

William S. Charlton
University of Texas/Nuclear Engineering Teaching Laboratory

Prepared by
Sandia National Laboratories
Albuquerque, New Mexico
87185 and Livermore,
California 94550

Issued by Sandia National Laboratories, operated for the United States Department of Energy by National Technology & Engineering Solutions of Sandia, LLC.

Available to DOE and DOE contractors from
U.S. Department of Energy
Office of Scientific and Technical Information
P.O. Box 62
Oak Ridge, TN 37831

Telephone: (865) 576-8401
Facsimile: (865) 576-5728
E-Mail: reports@osti.gov
Online ordering: http://www.osti.gov/scitech

Available to the public from
U.S. Department of Commerce
National Technical Information Service
5301 Shawnee Rd
Alexandria, VA 22312

Telephone: (800) 553-6847
Facsimile: (703) 605-6900
E-Mail: orders@ntis.gov
Online order: https://classic.ntis.gov/help/order-methods/

ABSTRACT

Researchers from Sandia National Laboratories (Sandia) and the University of Texas at Austin (UT) conducted this study to explore the effectiveness of commercial artificial neural network (ANN) software to improve insider threat detection and mitigation (ITDM). This study hypothesized that ANNs could be "trained" to learn patterns of organizational behaviors, detect off-normal (or anomalous) deviations from these patterns, and alert when certain types, frequencies, or quantities of deviations emerge. The ReconaSense® ANN system was installed at UT's Nuclear Engineering Teaching Laboratory (NETL) and collected 13,653 access control data points and 694 intrusion sensor data points over a three-month period. Preliminary analysis of this baseline data demonstrated regularized patterns of life in the facility, and that off-normal behaviors are detectable under certain situations—even for a facility with anticipated highly non-routine, operational behaviors. Completion of this pilot study demonstrated how the ReconaSense® ANN could be used to identify expected operational patterns and detect unexpected anomalous behaviors in support of a data-analytic approach to ITDM. While additional studies are needed to fully understand and characterize this system, the results of this initial study are overall very promising for demonstrating a new framework for ITDM utilizing ANNs and data analysis techniques.

ACKNOWLEDGEMENTS

The authors would like to thank the National Nuclear Security Administration's Office of International Nuclear Security (NNSA/INS) for funding this study. We also acknowledge the feedback and insights gained from discussions with various colleagues across the NNSA laboratory complex and members of the Institute for Nuclear Materials Management. Lastly, we thank Ms. Jacqueline Hoswell for her efforts throughout this phase of the project.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

Acronyms and Definitions

Abbreviation	Definition
AI	Artificial intelligence
ANN	artificial neural network
CAS	central alarm station
ITM	insider threat mitigation
ITDM	insider threat detection and mitigation
HRP	human reliability programs
NETL	Nuclear Engineering Teaching Laboratory
NNSA/INS	National Nuclear Security Administration/Office of International Nuclear Security
SMOT	Structurational Model of Technology
UT	University of Texas at Austin

1. INTRODUCTION

Insider threat mitigation (ITM) programs have traditionally focused on identifying characteristics of individuals and using preventative and protective measures to mitigate insider threat. These behavioral characteristics have remained the focus of ITM programs instead of shifting the focus to collective behaviors observed in the workplace that could be used in a more comprehensive "health-monitoring" system to improve ITM programs. These comprehensive approaches would use more empirical data and integrate workplace behavior-related insights into traditional ITM programs.

The difficulty in using an empirical, data-driven, collective-behavior-focused ITM program is that the program must differentiate between malicious intent and natural organizational evolution to explain anomalies in workplace behavior. Having such a program requires the presence of defined observables on which to build measures of behaviors that represent insider potential manifesting into malicious action. The goal of such an approach is not to attempt to identify and define all possible organization-level insider threat indicators, but rather to argue that undesired deviations from expected (or normal) patterns of organizational behavior may indicate insider threat potential.

In response, researchers from Sandia National Laboratories (Sandia) and the University of Texas at Austin (UT) conducted this study to explore the effectiveness of commercial artificial neural network (ANN) software to improve insider threat detection mitigation (ITDM). This study hypothesized that ANNs could be "trained" to learn the patterns of organizational behaviors and alert when certain types, frequencies, or quantities of deviations emerge. If successful, the application of ANNs to ITDM would result in a new approach for understanding, detecting, evaluating, and mitigating insider threats, including a more advanced evaluation framework and set of measures.

This report is the final deliverable to close out Phase I of this project. After summarizing the research project proposal and study objectives, this report situates this ANN for ITDM approach within relevant literature from several disciplines and illustrates support for the hypothesis that ANNs are capable of detecting insider threat deviations from expected organizational behaviors. Section 3 reviews the approach and methods used in the study, before Section 4 reviews the data, summarizes the analysis, and presents the results of the study. Finally, Section 5 concludes the report with insights and implications derived from the study and recommendations for future research.

1.1. Proposal Summary

To complete this study, the commercially available ANN software was installed at UT's Nuclear Engineering Teaching Laboratory (NETL)—a TRIGA MARK II research reactor facility. More specifically, the ANN algorithm—part of the ReconAccess software suite provided by ReconaSense®[1] —was installed as part of the NETL access control system. The goal was to explore the use of this particular ANN to "learn" the operational patterns of NETL to examine the software's ability to detect off-normal personnel activities and identify elevated risk levels for suspected regions of the facility. This project hypothesizes that leveraging a better understanding of workplace dynamics—sometimes called operational patterns of a facility—will improve the ability to identify, detect, and forecast insider threat potential.

[1] ReconaSense, "ReconAccess: Risk-Adaptive and Intelligent Access Control," 2019. Available at: https://reconasense.com/wp-content/uploads/ReconAccess_datasheet.pdf

In addition to collecting data for training the ANN, this new ITDM approach was tested against three scenarios: attempted access to the intrusion detection system panel, attempted off-hour access to the reactor bay, and scouting potential access to the fuel storage facility. Signals were collected from door access readers, video surveillance, area radiation monitors, and personnel radiation detection portals. This proposal focused on demonstrating a proof of concept that ANNs could learn operational patterns at a nuclear facility—and identify deviations from expected behaviors—by recording and learning from data signals already being collected at NETL. Funded for completion in FY19, this project was granted a no-cost extension to overcome installation-related delays.

1.2. Study Objectives

The overall goal of this project is to demonstrate the efficacy of using ANN-based approaches for understanding, detecting, evaluating, and mitigating insider threats, including a more advanced evaluation framework and set of measures. To understand such patterns, project objectives included:

- Collecting empirical data on operational patterns at NETL (based on data signals already recorded for other purposes)
- Training the ANNs to define expected operational patterns at NETL
- Evaluating the ability of the ANNs to identify operational patterns at NETL
- Analyzing the ability of these ANN-learned operational patterns at NETL against a set of hypothesized insider threat scenarios
- Recommending next steps for expanding this research, improving this ANN-based approach to ITDM, and developing a new framework for addressing insider threats

2. LITERATURE REVIEW

2.1. Current Approaches to Insider Threat

Though much analytical work on insider threat has focused on cyber-related applications,[2] this project focused on insider threats in physical security applications.[3] While various iterations exist, the National Insider Threat Task Force (2016) defined an insider threat as "the risk [that] an insider will use their authorized access, wittingly or unwittingly, to do harm to their organization. This can include theft of proprietary information and technology; damage to company facilities, systems or equipment; actual or threatened harm to employees; or other actions that would prevent the company from carrying out its normal business practices."[4] This is largely mimicked in definitions of insider threat used in the nuclear energy domain, most commonly attributed to the International Atomic Energy Agency[5] and the World Institute for Nuclear Security.[6]

The underlying logic for these definitions suggests that insider threat *opportunities* manifest when a combination of access, authority, and knowledge of a nuclear facility exists. These definitions also argue that potential insider threats materialize when such opportunities meet motivations to act maliciously against a facility. Traditional responses have focused on preventative and protective measures to mitigate opportunities and motivations within insider threat mitigation programs.

Preventative measures aim to reduce the likelihood that bad actors will gain opportunities to act maliciously against a facility and include human reliability programs (HRP) as well as other employment screening programs. HRPs (while important) cannot eliminate the potential for an insider threat to exist in a facility; there have been numerous examples of this.[7] [8] Protective measures aim to reduce opportunities for malicious insider acts through access controls, contraband detection, and other physical security measures. While protective measures can provide some deterrence to an insider, their application as point detectors can render them more susceptible to a knowledgeable insider.

Despite recent advances in HRPs, there are still significant challenges to comprehensively identifying all possible motivations or triggers that could initiate an insider act. Similarly, protective measures made modest advances but the rigidity of the underlying logic struggles to adjust to natural evolutions in expected operational patterns at nuclear facilities. There clearly is a need for developing a new framework for insider threat mitigation utilizing modern technological advances in data analysis. If insider *opportunity* is often considered a function of personnel access, authority, and knowledge,[9] then perhaps there is benefit in understanding operational patterns (or, expected

---

[2] Cappelli, Dawn et al., *The CERT Guide to Insider Threats* (Pearson Education, Inc., 2012).
[3] The authors acknowledge that lessons learned or insights gained from this physical security-focused project may have some application for insider threats in cyber applications.
[4] National Intelligence Task Force, *Protecting Your Organization from the Insider Out: Government Best Practices* (National Counterintelligence and Security Center, 2016). Available at: https://www.dni.gov/files/NCSC/documents/products/Govt_Best_Practices_Guide_Insider_Threat.pdf.
[5] International Atomic Energy Agency, *Preventive and Protective Measures Against Insider Threats*, IAEA Nuclear Security Series No. 8: Implementing Guide (Vienna: IAEA, 2008).
[6] World Institute for Nuclear Security, "Countering Violent Extremism and Insider Threats in the Nuclear Sector," 2018.
[7] J.E. Landers, "Psychological Profiles of the Malicious Insider," PNNL SA 102669, Pacific Northwest National Laboratory, 2014.
[8] A. Kolaczkowski et al., *Good Practices for Implementing Human Reliability Analysis (HRA) (NUREG-1792)* (U.S. Nuclear Regulatory Commission, 2005).
[9] A.D. Williams., S.N. Abbott, and A.C. Littlefield, "Insider Threat," in Encyclopedia of Security and Emergency

behaviors) related to access, authority, and knowledge. This approach reframes individually focused insider *opportunity* to facility-focused insider potential—where unacceptable deviations from expected patterns of access, authority, and knowledge relate to the likelihood of an insider act. This perspective may provide the framework to reconcile several issues with current approaches, including how human errors can be conflated with malicious or intentional acts and challenges related to adequately (and accurately) attributing potential motivations or triggers for insider actions.

2.2. Operational Patterns & Insider Potential

By exploring how individuals construct institutions, processes, and practices to achieve a common goal—which often manifest in observable patterns of expected behaviors—organization science provides insights for understanding various influences on ITM. For example, one popular concept of organization science describes how differences between designed and as-built organizations can lead to unexpected outcomes. For insider threat, this suggests that understanding the relationship between *designed* mitigations and daily work practices (as built) can help better explain observed operational patterns.[10,11] This concept further supports the argument that a better understanding of operational patterns can improve insider threat analysis and ITM performance.

To the extent that organizational behaviors can influence operational patterns, it is useful to invoke the organization science concept of investigating behaviors from three distinct perspectives or *lenses*: the strategic design lens, political lens, and cultural lens.[12] Each of these lenses represents shared ideas about human nature, the meaning of organizing, and the information required to make sense of an organization—each of which impacts observed operational patterns. The strategic design lens argues that with the right plan, information flow, and resource distribution, the organization can be rationally optimized to achieve its goal. The political lens interprets organizations as diverse coalitions of stakeholders with different (and sometimes conflicting) interests whose performances are influenced by ever-changing power dynamics that impact decisions. Lastly, the cultural lens describes organizational behavior in terms of the tacit knowledge of "this is how we do things around here" and the processes used to share this knowledge with newcomers. Table 1 summarizes the conceptual focus areas for each of these lenses and provides examples describing their ITM applicability. The *three-lens* approach suggests that performance assumptions underlying both preventive and protective ITM designs are influenced by both the independent focal areas of, and the interactions between, each lens.

Management, eds. Shapiro L., and Maras M.H. . Springer, Cham, 2019.

[10] C. Argyis, "Single-Loop and Double-Loop Models in Research on Decision Making," *Administrative Science Quarterly* 21(September 1976): 363-375.

[11] R.M. Cyert and J.G. March, *A Behavioral Theory of the Firm* (Englewood Cliffs, NJ: Prentice-Hall, 1963).

[12] John S. Carroll, "Introduction to Organizational Analysis: The Three Lenses" [Unpublished manuscript] (Cambridge, MA: MIT Sloan School, 2006).

Table 1. Carroll's (2006) Three Lens Organization Science Approach and Representative Insider Threat Mitigation Examples

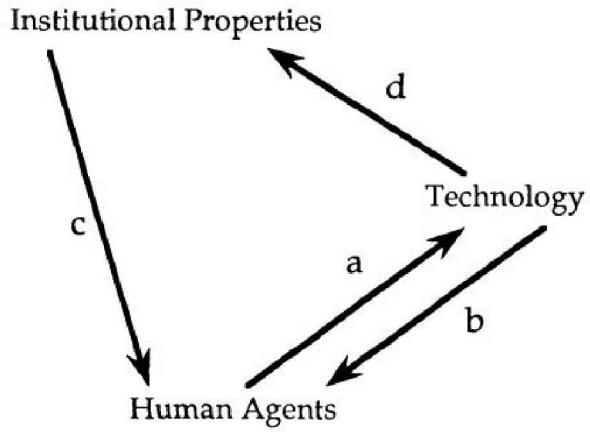
Lens	Conceptual Focus Areas	Insider Threat Mitigation-Related Examples
Strategic Design	• Organizational charts • Roles and responsibilities • Establishing formal communications channels • Resource allocation	• Clear background process* • Established investigation procedures** • Distinct access controls within facility** • Clearly communicating reinvestigation requirements to personnel**
Political	• Formal roles or accumulation of power (e.g., bringing in more money) • Informal roles or accumulation of power (e.g., level of experience)	• ITM manager gaining additional authority/decision-making responsibilities (e.g., via threat escalation)** • Lobbying management to increase background investigation requirements*
Cultural	• Artefacts (visible expressions) • Espoused values (stated intentions) • Attitudes/assumptions (*taken for granted* understanding)	• "If you see something, say something" signage/posted procedures** • Self-assessment results on security-related beliefs, assumptions, work practices, and norms** • An organizational reputation for taking ITM seriously (e.g. clearly communicated successes)*
		* denotes a traditional "preventive" ITM measure ** denotes a traditional "protective" ITM measure

Using these three lenses—and the dynamics between them—seems useful for more comprehensively describing operational patterns that impact ITM. Here, structuration theory asserts that organizational behavior results from recurrent human action that is both (and simultaneously) shaped by artefacts and constructed by their interpretation. The origins of structuration theory (or the enactment of structure)[13] expanded into a spectrum of descriptions related to recursive relationships within organizations to describe observed behaviors, including operational patterns. For ITM, this suggests organizational influences can impact the likelihood of an insider act just as much as individual access, authority, and knowledge—supporting a shift from a focus on insider *opportunity* to a focus on insider *potential*. Such a perspective also aligns well with the complex systems theory phenomenon in which interdependence among components is influenced by—and influences—the surrounding environment.[14] In other words, both structuration and complex systems theories logically support the importance of understanding how interactions between humans, artefacts, policies, procedures, and organizations result in operational patterns. From this perspective, evaluating such patterns in terms of interactions between humans, artefacts, policies, procedures, and organizations can help support ITM.

If desired levels of ITM are produced (*emerge* in systems theory terminology) through recurrent human interaction (influenced by institutional properties) with preventive and proactive measures, then a new analytical approach is necessary. The Structurational Model of Technology (SMOT)[15] serves as an instructive example. SMOT (Figure 1) offers a recursive description of how *human agent* actions situate the uses of *technology*, which then shapes the enacted organizational structure that

[13] Anthony Giddens, *The Constitution of Society: Outline of the Theory of Structuration* (University of California Press, 1984).
[14] Joseph V. Tranquillo, *An Introduction to Complex Systems* (Springer International Publishing, 2019).
[15] Wanda J. Orlikowski, "The Duality of Technology: Rethinking the Concept of Technology in Organizations," *Organization Science* 3(August 1992): 398-427.

produces *institutional properties* that enable or constrain those *human agent* actions. By replacing *technology* with *preventive and protective measures*, the common understanding of insider threat mitigation gives way to perceiving performance as human agents interacting with these ITM measures under the influences of institutional properties.



Arrow	Influence Description	Representative ITM Examples
a	Technology as a Product of Human Action	Use of two-factor authentication for accessing a NM vault
b	Technology as a Medium of Human Action	Type of physical lock supporting two-person access rules
c	Institutional Conditions of Interaction with Technology	Frequency & severity of reporting of two-person rule violations
d	Institutional Consequences of Interaction with Technology	Type/scope of recording & reporting tools supporting two-person access rules

Figure 1. Orlikowski's SMOT (above), and SMOT Descriptive Table with representative examples (at right).

From the SMOT perspective, ITM is not a function of a singular measure, but rather a descriptor of the capability of a human agent to use a given mitigation measure (under the influence of institutional properties) to complete either preventive or protective functions. For example, the two-person rule in and of itself does not counter potential insider acts. Rather, completion of two-person rules as part of normal, observed operational patterns counters potential insider acts. Yet, these patterns—as articulated in SMOT—are subject to how well existent institutional properties and resources (e.g., available technology) support completing the two-person rule. The measures that compose ITM programs also shape the enacted structure of the organization and the resulting institutional properties related to security performance. For example, facilities that rely on advanced and aggressive background reinvestigations have different enacted structures than those that rely on voluntary reporting for prevention—which result in *different* operational patterns.

These organizational science insights offer new insights for better describing ITM. For example, there is a need to better incorporate how daily work behaviors and operational patterns impact—and are impacted by—ITM. For example, to the extent that security culture is an ITM measure, its effectiveness is related to how one expert noted that security culture can be described in terms of "whether an operation is done with efficiency or lackadaisically."[16] Similarly, where ITM tasks require the use of technology, their execution emerges from the interactions of those technologies, human agents, and institutional properties. For example, consider simply *having* WPA2 encryption on facility servers versus the extent to which *personnel* regularly employ increasingly complex passwords to *enable* WPA2 encryption as *normal job behaviors*. Because "everyday actions are

[16] National Academy of Sciences, *Brazil-U.S. Workshop on Strengthening the Culture of Nuclear Safety and Security: Summary of a Workshop* (Washington, DC: The National Academies Press, 2015).

consequential in producing the structural contours of everyday life,"[17] ITM can be framed as the operational patterns resulting from recursively completing work tasks as human agents interact with preventive and proactive measures. To the extent that institutional properties influence these operational patterns, they also provide a new perspective by which to identify (and, potentially, prescriptively analyze) ITM. Taken together, these insights from organization science suggest ITM can be described in terms of operational patterns that emerge from dynamic and recursive relationships between preventive/protective measures, institutional properties, and daily work practices.

Traditionally, ITM programs have relied on reducing insider *opportunity* by emphasizing observation, evaluation, and implementation at the *level of the individual*, looking for a person with the intent and motivation to commit an insider attack. Despite the relative success of such individual-focused programs, examining ITM from the *level of the organization* incorporates the insights from organization theory—discussed above—to describe ITM in terms of operational patterns. If operational patterns can be captured by data signals *already* recorded (often for quality assurance, physical protection, or personnel safety reasons), then this provides a new perspective for describing ITM. Here, insider *potential* can be defined in terms of measured deviations from operational patterns—or, in other words, the likelihood for a successful insider act can be described in terms of unexpected behaviors registered in facility-related data signals. Thus, the ability to identify and measure operational patterns can help derive thresholds by which deviations from expected behaviors can be monitored and help determine appropriate and effective ITM.

By way of metaphor, this approach to ITM is similar to the concept of how medical doctors evaluate individual health. In humans, no one biological or medical measure comprehensively describes the overall health of a patient. High blood pressure in an otherwise healthy patient may not represent a serious medical concern if all other vital signs appear to be normal. High blood pressure in combination with other worrisome indicators, however, may lead a doctor to diagnose patients with a more serious medical condition. No doctor would diagnose a patient with a serious medical condition based on one vital sign by itself. Instead, they would rely on the cumulative results of a number of vital signs that lead to such diagnoses. The same principle can be applied to using organizational-level data for ITM and the idea of insider *potential*—there is no *silver bullet* indicator. Thus, similar to how a doctor checks multiple vital signs for a better understanding of the health of a patient, there may be combinations of organization-level signals that can serve as accurate indicators of insider potential.

Unlike individual health monitoring, there are no easy threshold lines that automatically determine likelihood of insider act success. Rather, and consistent with insights from organization theory, monitoring relevant, facility-level data signals over time will identify natural operational patterns. These baseline operational patterns establish measures for expected organizational behaviors—as determined from the set of continuously collected facility-level data signals—and can help determine thresholds of undesired deviations in two different ways. First, acceptable deviation ranges can be established on an absolute scale (e.g., only 5 individuals a day should be accessing a sensitive area). This approach has the advantage of increased clarity of identifying an anomalous event, but has the disadvantage of more false positive results. Second, acceptable deviation ranges can be placed *around* the baseline patterns (e.g., +/- 5% change in the number of people accessing a sensitive area per

[17] M.S. Feldman and W.J. Orlikowski, "Theorizing Practice and Practicing Theory," *Organization Science*, 22(September-October 2011), 1240-1253.

day). Again, any single undesired deviation from operational patterns does not suggest that the facility has an increasing insider threat potential, but rather that a significant enough change has occurred and should be investigated. The collection, processing, and evaluation of large, diverse data streams from multiple facility-level signals provide the opportunity for describing insider potential—and, by extension, the ITM program—in terms of operational patterns that more comprehensively describe expected organizational behaviors.

3. APPROACH & METHODOLOGY

This work focused on a new approach to insider threat mitigation based on evaluating operational patterns that describe behaviors at the level of the organization. By shifting focus away from the level of the individual, it is possible to detect deviations from expected operational patterns of behavior across the facility. Such operational patterns form as personnel at a facility settle into routine sets of daily or weekly practices. This is a natural human trait that is expected to exist in both commercial facilities (which have very routine operations) and research facilities (which can have very irregular operations). In addition, these operational patterns can be captured and described by using signals already collected at many operational nuclear facilities (including access control data, intrusions sensor data, camera video, area radiation monitoring data, personal radiation monitoring data, material control data, etc.). If there are empirical bounds to these operational patterns that represent the range of expected—or "normal"—behavior, then it is hypothesized that an insider threat attempt would manifest as a deviation from these bounds. This project further hypothesizes that these anomalies could be detected using an ANN analyzing existing data signals.

3.1. ReconaSense® & Artificial Neural Networks for Security

The ReconaSense®[18] AI Platform for Physical Security[19] relies on its proprietary ANN to improve facility-level security by providing real-time analytical processing of aggregated data from sensors, access controls, video systems, data repositories, and expected operating procedures. The ANN treats everything as a data point from an individual entity—where a person, a sensor, a camera, etc. is an entity—and works on a few basic principles. The ANN receives an array of input data, evaluates the data based on internal information, sends results to several output variables, and "learns" through back-propagation to gain intelligence. Ultimately, this ANN "system can be configured to automatically adjust risk levels as events occur and in so doing, eliminate threats before they occur."

More specifically, the ReconaSense® ReconAccess software uses an ANN to identify abnormal events and alert an operator to a possible threat. By taking local policies into account and learning the flow of people and processes at a facility, ReconAccess is able to identify when abnormal activities occur at a facility. The software integrates into the facility's existing security posture and implements role-based access controls with risk-adaptive access controls to ensure that employees, visitors, and others with access to a facility only visit the appropriate areas. Upon sensing abnormal activities, ReconAccess alerts the appropriate responders and can even initiate an emergency lockdown if necessary. The company claims that, "By leveraging artificial intelligence (AI), ReconaSense® identifies and mitigates potential threats and attacks before they happen giving security teams the ability to go beyond managing data and individual alerts to achieving true situational awareness and rapid response capabilities."[20]

The ReconaSense® software was installed—with some necessary adaptations to existing controllers, servers, and communication hardware—to control and monitor the NETL access control system. In the first phase of installation, a duplicate NETL access control server was installed with the ReconaSense® software implemented on this server. Controllers on the access control panel were

[18] ReconaSense®—which is located in Austin, TX—had previously interacted with UT regarding some of its products capabilities, which was a driver in selecting them as the commercial vendor for this project.
[19] John Carter, "The ReconaSense AI Platform for Physical Security." (ReconaSense, 2018).
[20] ReconaSense, "ReconAccess: Risk-Adaptive and Intelligent Access Control."

also replaced to allow for the duplicate server to provide access control functions and insider threat monitoring for only a small subsection of the NETL facility. This allowed for testing of the system to ensure that access control functions including alarm communication, display, and assessment to the central alarm station (CAS) were maintained before the ReconaSense® software was implemented on the complete NETL access control system. The Phase I system was tested for simple functionality, with minor errors in communication and access discovered and corrected. The Phase I system had only limited testing for insider threat mitigation.

In the second phase of installation, the complete NETL access control system was converted over to the ReconaSense® implementation. This included the access controls as well as monitoring of the intrusion alarm system. The implementation was carefully conducted to ensure that the UT CAS maintained alarm communication, display, and assessment at all times including door alarms and access control alarms, and that the system was installed with firewall protection. This installation phase did not include complete integration of the camera signals (which are fed to the CAS through a separate software implementation) or monitoring of the area radiation monitoring instrumentation throughout NETL. (NOTE: Each of these data signals could easily be included in a future phase, particularly as the area radiation monitor data is networked in the facility and feeding those signals to the ANN should be possible with no hardware modifications). Testing of the system included performance testing to ensure acceptable performance for the access control system and alarm system followed by an initial collection of baseline data for workplace patterns at the UT NETL. Installation of the hardware and software was successfully completed in early November 2019, and initial performance testing completed later in the month. Installation of the software and modification to the access controllers and communication system was more time consuming than originally anticipated. This impacted the amount of time available for initial collection of baseline data. As a partial solution, the period of performance for this research was extended to continue collecting training data during all of the NETL normal operations.

3.2. UT's Nuclear Engineering Teaching Laboratory

The NETL is an innovative facility with unique capabilities. NETL's multifaceted mission includes educating the next generation of leaders in nuclear science and engineering, conducting leading research at the forefront of the national and international nuclear community, applying nuclear technology for solving multidisciplinary problems, and providing service to the citizens of Texas, the United States, and the international community. In addition to its educational role, NETL has several primary research thrusts. These include (but are not limited to) nuclear forensics but also include robotics applications along with other areas; trace element analysis using neutron activation analysis and prompt gamma activation analysis; measuring distribution of elements in material using neutron depth profiling; and imaging materials with neutron radiography. NETL can also produce a variety of radioisotopes for use in research, nuclear medicine, and industrial processes, as well as support the design and development of experiments, processes, and products.

The NETL reactor was designed by General Atomics and is a TRIGA Mark II nuclear research reactor. The NETL is the newest of the current fleet of U.S. university reactors, reaching initial criticality in March 1992. Upgrading from its initial TRIGA Mark I reactor resulted in a substantial increase in NETL's research capability.[21] The NETL reactor has in-core irradiation facilities and five beam ports. The reactor is capable of steady-state operation at power levels up to 1 MW or pulsing

[21] Sean O'Kelly, "Ten Years of TRIGA Reactor Research at the University of Texas" (IAEA, 2002).

mode operation where powers as high as 1500 MW are achieved for about 10 msec. Under normal operating conditions, NETL hosts a range of personnel, who can include permanent operational staff, administrative staff, faculty, post-doctoral and staff researchers, graduate students, undergraduate students, contractors, and visitors.

Because NETL is under the authority of the Nuclear Regulatory Commission, it is important to note that the Code of Federal regulations 10 CFR 73 dictates security responsibilities and capabilities for the facility.[22]

[22] 38 FR 35430, Dec. 28, 1973

This page left blank

4. DATA COLLECTION AND ANALYSIS

Baseline (training) data was collected for a period of 90 days at NETL during normal operations. This time period included the December holiday break, which is a period of very limited activity at NETL. For Phase I activities, this project focused on collecting signals from access control and intrusion detection processes.

For the purpose of observing trends in the access control data, the period from December 23, 2019 to January 9, 2020 was excluded from analysis. A total of 13,653 access control data points and 694 intrusion sensor data points were collected. These data points were loosely organized for analysis to observe trends in the bounds of the NETL operational patterns. The data was organized into the following groupings:

- **Single access point operational pattern bound in time**: All access control data was organized by access point, date, and time of allowed access, and then by identity used for access. This allowed for observation of access patterns in time, including bounds for when general access is expected to occur for an average individual, as well as for specific individuals.

- **Time-sequenced, multiple access points operational patterns bound in time**: All access control data was organized by identity used for access, by date and time of allowed access, and then by access point. This allowed for observation of access patterns of individuals including bounds for when particular individuals would be expected to complete a particular access sequence.

This categorization of the data was performed simply to provide a rough estimate of the application of this ANN system to insider threat mitigation—more as a proof of concept than a complete or formal analysis. For example, there are other categorizations that could provide insight for organizing and analyzing the data. Grouping the data by personnel type is one categorization likely to generate a better understanding of operational patterns and produce useful conclusions—since student access likely has a very different profile from administrative staff access, which likely has a very different profile from operational staff access.

Because of the NETL research and education mission and the diversity of personnel (including operational/administrative staff, faculty, graduate/undergraduate students, and visitors), it was expected that there would be difficulty in the ANN being able to establish patterns of behavior. However, even from the limited baseline data, there were clear operational patterns for most personnel. The ANN was also able to establish bounds for the facility operation rhythms that supported the potential detection of insider attempts via deviations from these bounds.

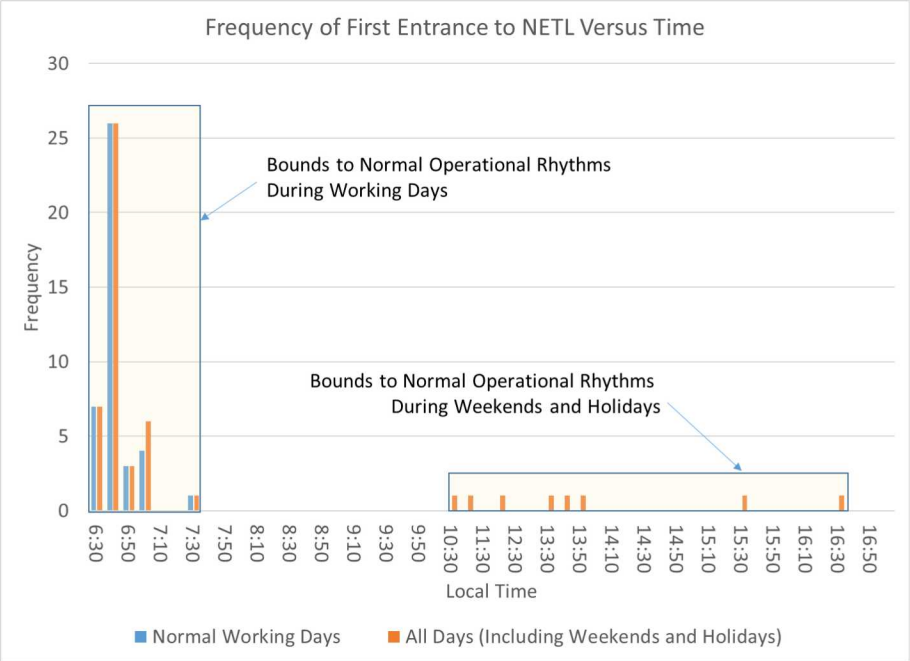


Figure 2. Frequency distribution showing time of first entrance to NETL facility, comparing "working days" and "all operational days."

The frequency distribution of the first allowed access to the NETL facility versus the time of access is shown in Figure 2. This distribution is shown for working days only and for all days (working days plus weekends and holidays). As can be seen, there are clear bounds on the normal time of first entry to the facility on all working days. Similarly, we can see that there are even bounds to the normal time of first entry on weekends and holidays (though admittedly the amount of data to establish those bounds is still low and the bounds themselves are wide). The ANN is capable of establishing these bounds as expected access profiles and analyzing future attempts to access for deviations outside of this expected profile. In addition, parsing the data by identification credential used for access indicates that the first access to the NETL during normal working days is performed by the same two individuals[23] in all but one instance. Thus, the ANN can also build individual access profiles based on expected bounds in NETL access and assess deviations in behavior outside of these bounds.

Another example of this analysis in time- and personnel-spaces is that of analyzing the data associated with first entry to the reactor control room. The frequency distribution of the first allowed access to the NETL reactor control room versus the time of access is shown in Figure 3. This distribution is shown for working days only and for all operational days (working days plus weekends and holidays). Similar to the data for first access to NETL writ large, there are clear bounds on the normal time of first entry to the facility on all working days, including less variation of first entrance on weekends and holidays. This is due to the fact that—even though the NETL reactor operates only during normal business days—laboratory facilities continue to operate on weekends and holidays. Similar to the data for the first NETL access, additional analysis of the data using identification used for access indicates that the first access to the NETL reactor control room

[23] To be more precise, this data indicates that that same two *identification credentials* completed the first NETL access in all but two cases. For this proof of concept, we make the simplifying assumption that such credentials are used by the individuals to whom they have been dispensed.

during normal working days is performed by the same three individuals. The ANN is capable of establishing these bounds and analyzing for deviations in behavior around first access to the reactor control room outside of these bounds.

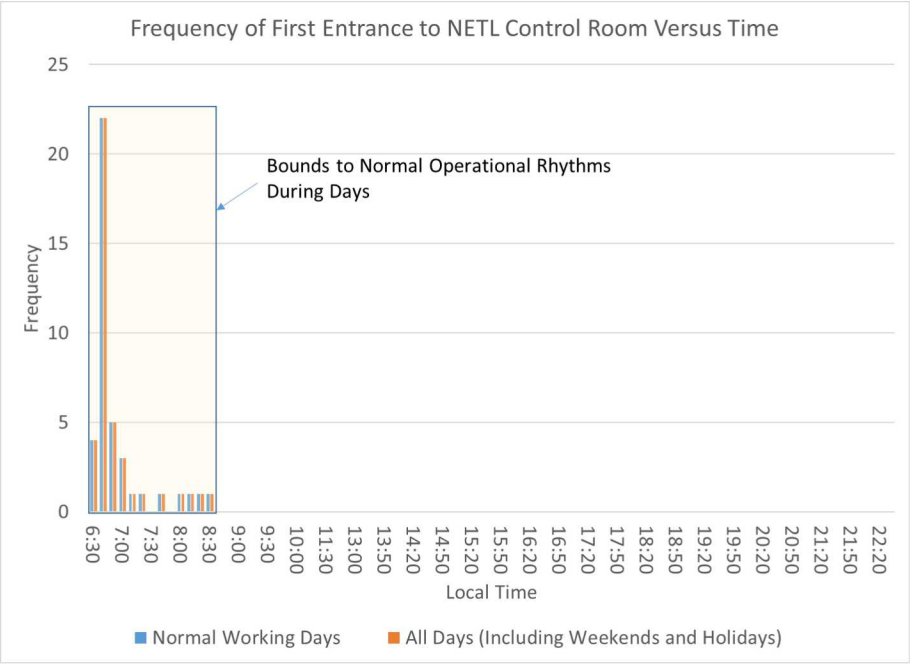


Figure 3. Frequency distribution showing time of first entrance to NETL control room during data collection time and delineated by working days only and all days.

Additional fidelity was achieved by further parsing the data. The data was further analyzed to determine the time of first entry to the NETL by each individual and organizing them by personnel group. The results are shown in Figure 4. As can be seen, each personnel group has specific patterns in terms of their time of first entry to the facility. In some cases, these patterns are very tightly bounded in time (for example for the administrative and operational personnel), and in other cases these patterns have wide distributions (for example the faculty, undergraduate students, and graduate students). These distributions—which are also found for other access points within NETL—represent *expected* arrival profiles per personnel group.

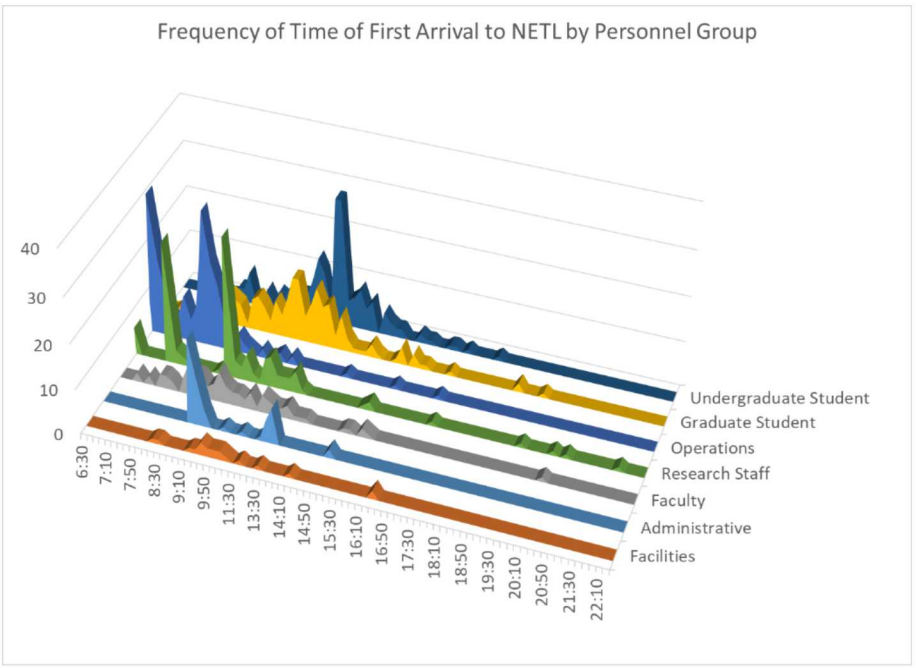


Figure 4. Frequency distribution showing time of first entrance to NETL during data collection time and separated by personnel group.

Further, even for these broad distributions, the ANN can identify more tightly bound access times for each individual within a given personnel group. Figure 5, for example, shows the frequency distribution of time of first arrival to NETL by four undergraduate students. As illustrated in the plot, even though the sum of all students shows a wide bound for this time of first access (the dark blue distribution near the top of Figure 4), the bounds are much tighter for any individual student. Thus, we expect that the ANN would be able to identify deviations in behaviors both from the norm for their personnel group and from the norm for their individual activities.

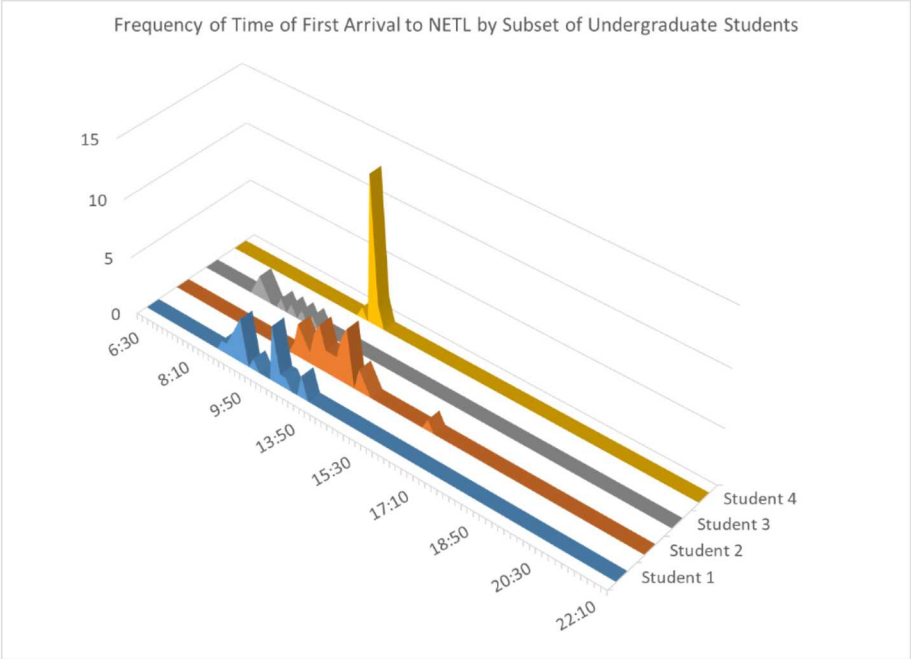


Figure 5. Frequency distribution showing time of first entrance to NETL during data collection time for four specific undergraduate student individuals.

These examples—time of first access to NETL and the reactor control room—represent just one data stream and demonstrate how baseline data can identify operational patterns. Using the ANN, collected data can then be evaluated for deviations from these operational patterns—or expected behaviors—in both the time and personnel-identification domains. Yet, the ANN can construct a more nuanced understanding and identify more complex patterns of facility operations than these relatively simple examples. Consider how the baseline data collected shows that initial entry to the facility is typically followed by entry to the reactor control room and then entry to the reactor bay. The baseline data collected indicates that these operations are all completed by the same individual in a very specific order of access within an expected bounding in time. Looking at the distributions in Figure 5 as a profile of expected behaviors, the ANN would be able to clearly identify deviations from this time-sequenced bounding. For example, the ANN similarly expects an additional outer door entry during the first entry to the reactor control room. Deviations from such expected behaviors related to accessing sensitive areas of the facility may represent a possible insider threat. More specifically, consider either an "off-time" first access by the first individual or the absence of the second individual while the first individual is in the reactor control room. In each of these scenarios—which the ANN would classify as deviations or anomalies—an individual with authorized access is unexpectedly alone in the reactor control room. Again, this situation in and of itself does not indicate an insider act is taking place, but such scenarios do greatly increase the potential for them.

4.1. Testing Scenarios Developed

For the purpose of testing the capability of the ReconaSense® ANN, NETL researchers in, collaboration with SNL staff and ReconaSense® engineers, developed several insider threat detection scenarios[24] for evaluation. In Scenario 1—**Intrusion System and Access Control**

[24] The researchers also developed three additional scenarios for testing safety/security synergies that are possible using

Closet—the insider is attempting to gain access to the closet that contains NETL's intrusion detection system panel. This panel is where all intrusion sensor data is fed and processed, and where signals are communicated to the CAS. If an insider gains access to this panel, they presumably could sabotage the panel to eliminate/falsify alarm signals at the CAS. Such an act would then allow an outsider or insider to attack the facility at some later date/time without the possibility of assessed detection. This panel is protected by NETL physical security and access control systems. In this scenario, the ANN acts as an insider threat detection system by attempting to detect off-normal activity for access to this panel from three prescribed pathways:

- an unauthorized individual (testing both single attempts and repeated attempts during both normal working hours and off-hours)
- an unauthorized individual during off-hours using stolen or falsified credentials of an authorized individual but using their own credentials to access the NETL building
- an unauthorized individual during off-hours using stolen or falsified credentials of an authorized individual to access both the NETL building and the panel

The ANN-based insider threat detection system is looking for abnormal patterns of behavior regarding both the insider actor and the stolen/falsified credentials.

In Scenario 2—**Intrusion Detection**—the insider is trying to gain access to the reactor bay during off-hours. During off-hours, the reactor bay is locked and alarmed. This scenario assumes that the insider has authorized access to the NETL building but not to the reactor bay inside the NETL building. The insider uses their credentials to access the NETL building and then attempts unauthorized entry to the reactor bay. In this scenario, the ANN acts as an insider threat detection system looking for off-normal activity that would include not only attempts at unauthorized access (which is similar to the results from scenario 1), but also early detection of the insider moving toward the reactor bay before they have reached any related access control station(s). NETL deploys sensors that can detect motion as individuals move in a direction toward the access bay. When combined with ANN-generated insights like the time (e.g., off-normal hours) or expected profile of access (e.g., credential that entered NETL does not have reactor bay access), this motion could be flagged as a possible elevated risk to the facility and relayed to the CAS for assessment.

In Scenario 3—**Fuel Storage Surveillance Activity**—the insider is trying to acquire knowledge pertaining to the security system for NETL's fuel storage facility. NETL fuel is stored in pits built into the floor of the reactor bay. This is essentially a locked and alarmed floor vault. To complete this task, the insider needs to surveil the area around the fuel and then test the alarm systems to determine what level of activity will set off the alarms while accessing the storage location. Testing of the alarm system includes the intrusion detection sensors, area radiation sensors, cables/conduits for those sensors, and the alarm panel. The ANN acts as an insider threat detection system that will analyze motion sensor info, reactor bay access control data, surveillance data, and intrusion detection data to assess personnel off-normal activity, in particular for signs of multiple attempts to access the fuel storage area. This scenario includes evaluating the ANN's ability to identify potential insider surveillance and potential insider testing alarms/sensors to determine sensitivity levels (NOTE: The second task is likely much easier to detect than the first).

---

the ReconaSense® ANN. These scenarios were beyond the scope of Phase I activities, so no testing for these scenarios was completed and thus they will not be detailed in this report.

Following collection of baseline data, an initial testing of the system for these three scenarios was completed. Since only limited testing data was collected, this evaluation primarily consisted of demonstrating the validity and acceptability of the scenarios to provide useful results in evaluating the ANN's capability as an insider threat detection system. However, this testing on the baseline data does show promise for the ANN as an insider threat detection system. Table 1 shows a summary of the preliminary results from this evaluation. Preliminary assessment of these results is that the scenarios do appear to be valid scenarios that will fully test the ANN capabilities. For example, consider the success in detecting and denying access for all attempts for unauthorized access to the panel closet and the reactor bay, as well as the success rate for detection and denial of panel closet access for use of an authorized credential by an unauthorized individual with access to the NETL building.

Table 2. Preliminary Results from Testing Scenarios

No.	Name	Test Description	Preliminary Results
1	Intrusion System and Access Control Closet	Unauthorized Access Attempt	Detected and Access Denied in All Cases
		Authorized Access Credentials Used by Unauthorized Individual Who Entered Building Using Their Own Credentials	Detected and Access Denied in Most Cases
		Authorized Access Credentials Used by Unauthorized Individual Who Entered Building Using Authorized Individual's Credentials	Not Detected and Access Granted in All Cases
2	Intrusion Detection	Unauthorized Access to Reactor Bay	Detected and Access Denied in All Cases
		Early Detection by Motion Sensor	Not Tested
3	Fuel Storage Surveillance	Insider Surveillance	Difficult to Detect Without Additional Sensing Input
		Insider Alarm Testing	Not Tested

Yet, this preliminary analysis also highlighted several areas for additional exploration in future phases of this research. Consider, for example, the challenges for adequately distinguishing "insider surveillance" from normal operational behaviors described with the collected data signals. In Scenario 3, it became obvious that because of the number of people in the reactor bay during the testing, the system had difficult determining what motion was attributed to what person. This system would be greatly enhanced if the individual staff member's location was being tracked (for instance, there are personal dosimeters that maintain position information about the dosimeter and this could be used to uniquely identify the position of an individual in the facility).

This page left blank

5. CONCLUSIONS AND RECOMMENDATIONS

Overall, Sandia and UT were successful in meeting the Phase I "proof-of-concept" research goals of installing the ReconaSense® system, collecting baseline data, and performing some initial testing of the system for operational patterns at NETL. The results indicated that ANN can identify and define obvious patterns of life for personnel, in this case based on time-series access control data. These Phase I results also showed the capability for ANNs to establish boundaries on expected operational patterns across personnel types. Even from the limited baseline data collected, this ANN-based approach illustrated that the potential detection of insider attempts through deviations from expected operational patterns is feasible. More succinctly, completion of this pilot study demonstrated how the ReconaSense® ANN could be used to identify expected operational patterns and detect unexpected anomalous behaviors in support of a data-analytic approach to ITDM.

The installed ANN does show promise in improving ITDM, but additional data collection is needed to fully test the capability. Training data needs to continue to be collected on normal facility behaviors for the ANN to learn and to identify more robust and nuanced operational patterns. More robust operational patterns will also support more detailed and rigorous scenario testing to expand the insights gained in the Phase I work. For example, a set of controlled experiments based on the scenarios in Table 2 could be designed and conducted to more formally assess how well this ANN's anomaly detection capabilities support insider threat detection and mitigation. For example, asking a graduate student to enter the reactor control room at 11:30 pm on the second Tuesday of the month and assessing (1) how easily this known anomaly is to observe in the data and (2) did the ANN register the anomaly. If successful, follow-on efforts should include sensitivity analysis of these anomaly detection capabilities through a series of more varied, less controlled experiments based on these same scenarios. For example, asking a graduate student to enter the reactor control room sometime after 9 pm during the first week of the month to assess ANN anomaly detection sensitivity in the same manner described above. Additional data collection and ANN learning could also yield a deeper understanding of insider potential introduced in Phase I, including the development (and validation) of new ITDM performance metrics.

Phase I completion identified an additional set of recommendations for expanding the insights gained in applying an ANN to a data-analytic approach to ITDM. One recommendation is to incorporate NETL's networked area radiation monitor data streams into the ANN analysis. This would provide real-time visibility of radiological sources movement and lead to more detailed operational patterns and—by default—more comprehensive anomaly detection for improved ITDM.

Similarly, it is recommended to incorporate the NETL camera system into the data analysis. Yet, the current system at NETL is relatively old technology and is only being used by the CAS for alarm assessment. There clearly are additional patterns of behavior in the camera feed data that could be useful if processed through image recognition software and integrated into the ANN. For example, when coupled with the access control data, camera feed data could not only identify who is where in a facility but also their relationships to other people and sensitive objects/areas. This would be especially useful for identifying individuals who are under duress by an insider or outsider adversary. Implementing and testing these additional features may show added ITDM capability for commercial facilities.

Another recommendation stems from the Phase I conclusions in Scenario 3: Fuel Storage Surveillance. The difficulty in uniquely identifying the position of individuals using simply motion detector information correlated with access control information suggests a need to incorporate data related to individual location. For example, NETL personnel could carry electronic personal dosimeters with facility position sensors. Though not currently used at NETL, they are commercially available, and Phase I insights suggest acquiring, integrating, and evaluating their impact on ANN operational pattern identification and anomaly detection capabilities.

While additional studies are needed to fully understand and characterize this system, the results of this initial study are overall very promising for demonstrating a new framework for ITDM utilizing ANNs and data analysis techniques.

6. REFERENCES

Argyis, C., "Single-Loop and Double-Loop Models in Research on Decision Making," *Administrative Science Quarterly,* 21(September 1976), 363-375.

Carroll, J. S. "Introduction to Organizational Analysis: The Three Lenses" [unpublished manuscript]. Cambridge, MA: MIT Sloan School, 2006.

Cappelli, Dawn et al., *The CERT Guide to Insider Threats.* Pearson Education, Inc., 2012.

Cyert, R., & March, J. *A Behavioral Theory of the Firm.* Englewood Cliffs, NJ: Prentice-Hall, 1963.

Feldman, M.S. and Orlikowski, W.J., "Theorizing Practice and Practicing Theory," *Organization Science* 22(September-October 2011), 1240-1253.

Giddens, A. *The Constitution of Society: Outline of the Theory of Structuration.* University of California Press, 1984.

International Atomic Energy Agency, *Preventive and Protective Measures Against Insider Threats,* IAEA Nuclear Security Series No. 8: Implementing Guide. Vienna: IAEA, 2008.

Kolaczkowski, A. et al. *Good Practices for Implementing Human Reliability Analysis (HRA) (NUREG-1792).* Washington, D.C.: US Nuclear Regulatory Commission, 2005.

Landers, J. *Psychological Profiles of the Malicious Insider.* PNNL SA 102669, Pacific Northwest National Laboratory, 2014.

National Academy of Sciences. Brazil-U.S. Workshop on Strengthening the Culture of Nuclear Safety and Security: Summary of a Workshop. Washington, DC: The National Academies Press, 2015.

National Intelligence Taskforce. *Protecting Your Organization from the Insider Out: Government Best Practices.* National Counterintelligence and Security Center, 2016.

O'Kelly, Sean. "Ten Years of TRIGA Reactor Research at the University of Texas." IAEA, 2002.

Orlikowski, W.J. "The Duality of Technology: Rethinking the Concept of Technology in Organizations," *Organization Science,* 3(August 1992), 398-427.

ReconaSense. "ReconAccess: Risk-Adaptive and Intelligent Access Control" 2019. https://reconasense.com/wp-content/uploads/ReconAccess_datasheet.pdf

Tranquillo, J.V. A*n Introduction to Complex Systems.* Springer International Publishing, 2019.

Williams A.D., Abbott S.N., Littlefield A.C. "Insider Threat." *Encyclopedia of Security and Emergency Management.* Shapiro L., Maras M.H, eds. Springer, Cham, 2019.

This page left blank

DISTRIBUTION

Email—Internal

Name	Org.	Sandia Email Address
Dominic Martinez	06812	dmartin@sandia.gov
Sondra Spence	06812	sspence@sandia.gov
Tina Hernandez	06832	therna@sandia.gov
Gerald Hendrickson	00023	gahendr@sandia.gov
Technical Library	01977	sanddocs@sandia.gov

Email—External[redacted]

Name	Company Email Address	Company Name
Kathrine Holt	katherine.holt@nnsa.doe.gov	National Nuclear Security Administration
Pratap Sadasivan	pratap.sadasivan@nnsa.doe.gov	National Nuclear Security Administration
Cary Crawford	crawfordce@ornl.gov	Oak Ridge National Laboratory
Melinda Lane	lane14@llnl.gov	Lawrence Livermore National Laboratory
William Charlton	wcharlton@austin.utexas.edu	The University of Texas at Austin

This page left blank

This page left blank

Sandia National Laboratories

Sandia National Laboratories is a multimission laboratory managed and operated by National Technology & Engineering Solutions of Sandia LLC, a wholly owned subsidiary of Honeywell International Inc. for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA0003525.