# NUCLEAR FACILITY ACADEMIC JOURNAL

## TO MEASURE INSIDER THREAT DETECTION & MITIGATION

RECONASENSE

# TABLE OF CONTENTS

# INTRODUCTION

## Identifying Characteristics

Traditionally, insider threat mitigation (ITM) programs at nuclear facilities focus on identifying characteristics of individuals to identify and address possible malicious actions. This paradigm often manifests in the application, over the lifecycle of potential employees ,of different countermeasures to address these potential risks. Consider, for example, the National Insider Threat Task Force (NITTF) definition for insider threat:

"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." [1, p. 3]

In the United States, the 2011 creation of the NITTF with Executive Order 13587 launched a coordinated effort to establish detection and prevention programs across governmental agencies.[2] In its 2017 publication on best practices, the NITTF also highlighted the need for vigilance and diligence as successfully mitigating insider threats is "a dynamic effort requiring constant evaluation, fresh perspectives, and updated approaches." [3]

ITM approaches at nuclear facilities tend to focus on identifying and deterring problematic or malevolent behaviors of individuals. These approaches have traditionally focused on preventative (measures implemented before access is granted) and protective (measures taken after access is granted and throughout employment) strategies to mitigate unwanted individual behaviors [4]. Such approaches may result in an overreliance on generic job task analysis and detection of aberrant individual behavior that may not fully account for workplace behavior patterns or may inadvertently ignore facility recovery operations. Yet, there may be an advantage in shifting this focus toward collective behaviors observed in the workplace for use in a more comprehensive "health-monitoring" paradigm. In such a paradigm, observed—and empirically measured—patterns of expected operational activities can serve as a baseline from which to detect potential insider threat activities. For example, data collected to monitor quality assurance or safety activities could be used to establish expected workplace trends and dynamics likely to be disrupted in the event of a malicious insider activity. Simply put, anomalies in expected operational patterns may provide additional useful insight to mitigate and minimize insider threats.

Building on best practices exhibited in the nuclear industry (e.g., by the U.S. Nuclear Regulatory Commission) and lessons learned from other industries (e.g., the casino industry), there seems to be a benefit in invoking an empirical, data-driven, collective-behavior-focused program to counter insider threats. From this perspective of ITM, undesired deviations from expected (or normal) patterns of organizational behavior may indicate an increased likelihood or opportunity for a malicious insider act. One key challenge for such an approach is the ability to distinguish between malicious intent and natural organizational evolution to explain anomalies in expected operational workplace patterns. Thus, a collective-behavior-based approach to ITM requires the presence of defined, observable measures and organization-level insider threat indicators on which to build metrics of behaviors that represent insider potential manifesting into malicious action.

To this end, researchers from Organization 1 and Organization 2 conducted a multi-phase empirical study to explore the effectiveness of using commercial artificial neural network (ANN) software to improve insider threat detection mitigation (ITDM) with this collective-behavior-based approach. The 2019–2020 study hypothesized that ANNs could be "trained" to identify and learn operational workplace patterns 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 comprehensive evaluation framework and set of measures.

This article begins by situating this collective-behavior-based approach for ITDM within relevant literature from several disciplines and describing the efficacy of ANNs for this application. It then presents the hypothesis that ANNs are capable of detecting insider threat deviations from expected operational workplace patterns and reviews the approach and methods used in the empirical study. Following a summary of the analysis and evaluation of the study, this article concludes with insights and implications for ITDM and recommendations for future research.

# 2 SITUATING COLLECTIVE BEHAVIOR-BASED INSIDER THREAT DETECTION & MITIGATION

## 2.1 Traditional Approaches to Insider Threat Mitigation

By definition, traditional ITM programs at nuclear facilities have focused on countering "an individual with authorized access to [nuclear material] associated facilities or associated activities or to sensitive information or sensitive information assets, who could commit, or facilitate the commission of criminal or intentional unauthorized acts involving or directed at nuclear material, other radioactive material, associated facilities or associated activities or other acts determined by the State to have an adverse impact on nuclear security" [5]. This globally accepted definition of insider threat from the International Atomic Energy Agency (IAEA) is similar to what is touted by the World Institute for Nuclear Security [6] (WINS) and is generally consistent with the broader NITTF definition presented in the previous section.

Many ITM programs focus on the idea that an insider threat opportunity will present itself when any person has access, authority, and knowledge of a specific nuclear facility [7]. The assumption is that threat opportunities will materialize into attacks when an insider has the motivation to act maliciously. In other words, an insider will attempt a malicious action when opportunity aligns with motivation. Up to this point, ITM programs have focused on implementing preventive and protective measures to counter such attacks. Preventive measures—typically implemented before access to a nuclear facility is granted—aim to reduce the likelihood that bad actors will gain opportunities to act maliciously. Preventive ITM measures include pre-employment screening, human reliability programs (HRPs), and other monitoring approaches based on behavioral reporting mechanisms. Though a critical element of ITM in general, HRPs cannot completely eliminate the potential for insider threats to a facility [8,9]. Protective measures—typically implemented after access is granted to a nuclear facility—aim to reduce opportunities for malicious insider acts through access controls, contraband detection, and other physical or cyber security measures. While protective measures provide some deterrence, their intended use as "point-in-time behavior" detectors can render them more susceptible to exploitation by a knowledgeable insider. Many traditional protective measures were developed to deter or capture adversaries without access to the facility. Because insiders have legitimate facility access, they are not likely to set off a "point-in-time" detector. For example, an insider is unlikely to set off a perimeter motion detector if they have authorized access to use the front door. Thus, protective measures, as traditionally applied, might need to be reconsidered to successfully identify and stop insider attempts.

Traditional ITM programs at nuclear facilities (e.g., [7]) have improved in recent years with additional study. Improvements include the use of mixed-integer programming to simulate insider actions in terms of advanced game-theoretic models [10], estimating insider-related behaviors in terms of the common cause failures of protective devices [11], and identifying sociotechnical indicators of insider threat risk [12]. These programs, however, are still driven by peer-to-peer reporting and individual behavioral observation mechanisms. Facilities are better off with these human-oriented ITM programs than with no ITM programs—but there is a need for a new insider threat framework that utilizes advances in data analysis to better characterize deviations from expected operational workplace patterns. If insider opportunity is often considered a function of personnel access, authority, and knowledge [4], then there should be a benefit in understanding operational patterns (or expected behaviors) related to access, authority, and knowledge. As shown in Figure 1, this approach shift focus from individually focused insider opportunity to facility-focused insider potential and asserts that unacceptable deviations from


Figure 1: Venn diagram comparing the collective-behavior-based approach versus the individual-behavior-based approach to ITM from the international best practices for nuclear facilities in [5].

expected patterns of access, authority, and knowledge relate to the likelihood of a malicious insider act. Such a collective-behavior-based ITDM program could help to address issues in current approaches, such as inaccurately conflating human error with malicious acts, more adequately attributing triggers for insider actions, and more quickly communicating abnormal behavior within a facility to initiate a proper response.

## 2.2 Operational Patterns for Collective Behavior-Based Insider Threat Detection & Mitigation

A collective-behavior-based approach begins with exploring how individuals construct institutions, processes, and practices to achieve a common goal, which can invoke tenets of organization science to describe the resultant observable patterns of expected behaviors. Previous research demonstrates the utility of applying different organization science theories and insights to ITDM [13]. For example, one popular concept would describe the performance of ITDM in terms of the relationship between planned insider threat mitigations ("as designed") and daily work practices with those mitigations ("as built") [14], whereas another organization science theory would argue that ITDM performance results from recurrent human action (e.g., individual access, authority, and knowledge) that is both (and simultaneously) shaped by artifacts and constructed by their interpretation (e.g., organizational behaviors and patterns) [15]. Similarly, Rasmussen's [16] collective-behavior-based framework is popularly used in complex systems safety to describe how an organization can promote "campaigns for safety culture" to act as a "counter gradient" to various dynamics driving individuals toward the "boundary of functionally acceptable behavior," as illustrated in Figure 2. In other words, organizational patterns can be used to account for individual behaviors that tend toward states of higher risk. The results of this research imply that a better understanding of organizational patterns could improve ITM programs by defining—and measuring—insider threat potential in terms of deviations from expected operational workplace patterns.

While organization science has identified that observed patterns of organizational behaviors play a key role in mitigating the natural tendency for organizational pressures to drive individuals toward increasingly risky behaviors, no such approach has been developed to frame the likelihood of insider act success. Consistent with insights from organization theory, the monitoring of relevant, facility-level data signals over time can identify natural operational patterns that compose both "perceived boundary of acceptable performance" and the "Brownian movements" of individuals. Establishing baseline operational patterns for expected organizational behaviors from the set of continuously collected facility-level data signals can help determine the "error margin" between the perceived and actual boundary of functionally acceptable behavior—including establishing thresholds of undesired deviations in two different ways.

First, an absolute scale can be used to establish acceptable deviation ranges (e.g., only five individuals a day should be accessing a sensitive area). This approach clearly identifies an anomalous event but is likely to produce 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 day). This allows bit more flexibility so that a single undesired deviation from operational patterns does not indicate 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 ITDM program—in terms of operational patterns that more comprehensively describe expected organizational behaviors. It is theoretically possible that any type of recorded data can serve as either a signal of organizational patterns or as a way to collect relevant information about an individual's patterns and how they shift over time. In reality, the degree to which ITDM programs do not add any burden to current operational and/or security operations is likely to significantly increase acceptance and usability of such programs. Therefore, this initial exploration of a collective-behavior-based approach to ITDM focused on the impact of leveraging data signals already being collected within a facility (see Section 3.1).

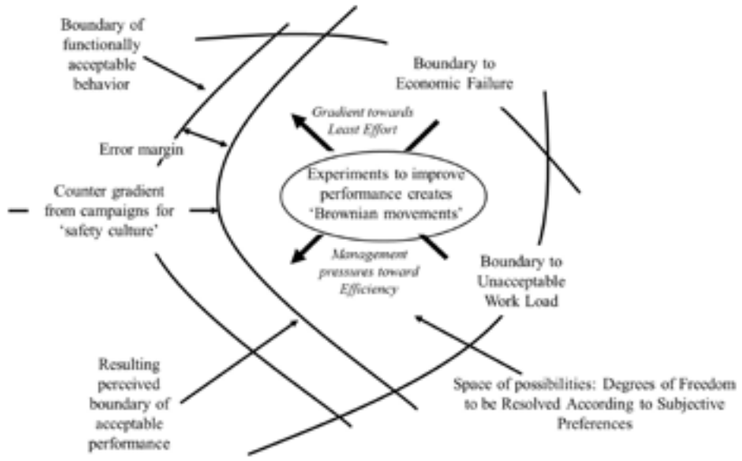


Figure 2: Rasmussen's model of how organizational influences (the "counter gradient" that creates the "resulting perceived boundary of acceptable behavior") can help mitigate natural tendency for organizational pressures to drive individuals toward increasingly risky behaviors (toward the "boundary to functionally acceptable behavior"), recreated from [16].

## 2.3 Data-Driven Insights for Insider Threat Detection & Mitigation

Similarly, leveraging advances in machine learning (ML) can support a data-driven approach to describing—and identifying anomalies within—observed operational workplace patterns. The ML domain introduces various mechanisms to support the aggregation of multiple, disparate data signals to detect abnormalities through deviations from an expected baseline. While direct quantification of insider threat is likely to be difficult, introducing ML insights and tools provides additional capabilities to enhance ITDM.

ML methods are often considered a black box that is difficult to interpret, but generally ML is an algorithm that performs a task without explicit programming. One of the most well-known examples of ML, the artificial neural network (ANN), was inspired by modeling neurons in the biological brain, an example of which is illustrated in Figure 3.

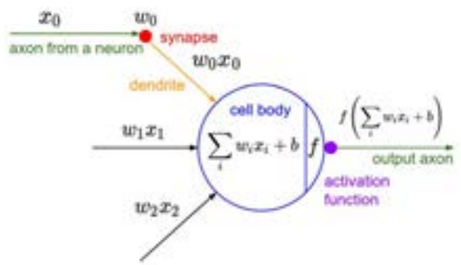


Figure 3: Representative model of a single neuron processing input signals into an output activity without explicit programming for the activation function, from [17].

Neurons are typically organized into layers that form a network, and at a high-level, neurons receive multiple input signals and produce some output. In this representation, signals from neurons in the previous layer (x_i) are multiplied by a given weighting factor (w_i). These signals are summed together with a given bias term (b) before being applied within an activation function (f), which ensures that the neural network can learn a nonlinear function. Output signals are propagated through the neurons in the network until the final output is calculated. The derivative of the loss between the target value and predicted value is back propagated through the neural network using the chain rule, where the gradients are used to adjust the weighting factors to give predictions closer to the true value. This process is referred to as training. Using this generic process, neural networks can, theoretically, learn any nonlinear function and offer several key concepts that might be leveraged to improve ITDM.

ANNs have been applied to a wide range of domains including pattern detection, routine task performance reduction, and sensor attack mitigation—each of which implies benefits for ITDM. Relevant ITDM patterns could exhibit complex temporal dependencies in high dimensional data making it difficult to detect with traditional statistical approaches. ANN-based approaches enable the capture of these subtle changes within larger, multisensory datasets.

Specific ML examples of this capability include a method to simultaneously train two neural networks to allow for recurrent attempts at "fooling" each other to identify anomalies within discrete event sequences [18] and an algorithm for effectively capturing multiscale sequential patterns in order to identify anomalies [19]. Similarly, ML approaches can reduce the number of routine tasks performed by humans in support of ITDM, including rapid response image recognition algorithms [20] and ML tools with high degrees of accuracy (~94%) for categorizing digitally captured visual images [21]. Lastly, ML approaches ITDM by mitigating insider attempts to attack sensor systems, as demonstrated by the ability of convolutional autoencoders to protect against attacks on fingerprint readers for access control [22].

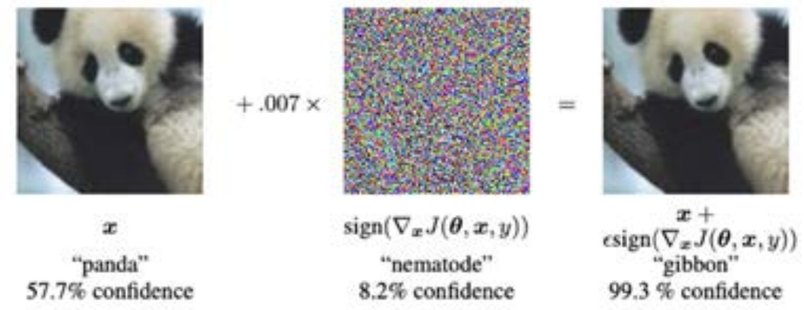


Figure 4: Representative example of "attacking" machine learning and artificial neural network algorithms to distort analytical outputs, taken from [23].

While the application of ML and ANN approaches to ITDM is promising, it is not a panacea solution. Shortcomings that would need to be addressed for successful application to ITDM include (but are not limited to) the infrastructure need for large amounts of data for training, the potential difficulties in transferring algorithms between facilities, calibrating for background signal noise, sensor drift (or misalignment), and protecting the ML algorithms themselves from attack. For example, consider the three images in Figure 4, below. While the algorithm correctly classifies the left-most image as a panda with approximately 58% confidence, adding a noise pattern that is imperceptible to the human eye (the middle image) results in the algorithm incorrectly classifying the modified image as a gibbon with over 99% confidence [23].

# 3 METHODS & DATA COLLECTION

## 3.1 Research Approach

To focus on this new approach to ITDM, the empirical focus shifted away from the level of the individual and toward the possibility of detecting deviations from expected operational patterns of behavior across a given nuclear facility. Such operational patterns emerge as facility personnel settle into routines of regular (e.g., daily or weekly) practices. This is a natural human trait expected to exist in both commercial facilities (that have very routine operations) and research facilities (that can have very irregular operations). Yet, these operational patterns can often be captured and described by using signals already collected at many nuclear facilities, including (for example) access control data, intrusions sensor data, camera video, area radiation monitoring data, personal radiation monitoring data, and material control data. If empirical bounds to these operational patterns exist, they could be considered representative of the range of expected—or "normal"—operational behaviors at a facility. By extension, then, deviation from these bounds could be an indicator of unwanted behaviors—including possible malicious insider acts.

Organization 2 (ORG 2) provided a facility within which to identify and test the empirical bounds of operational workplace patterns. ORG 2 is an innovative facility with unique capabilities and a multifaceted mission that includes educating next generation leaders in nuclear science and engineering. In addition to its educational role, ORG 2's research focuses on several primary thrusts, including (but not limited to) nuclear forensics; trace element analysis using neutron and prompt gamma activation; neutron depth profiling-based measurements of elemental distribution in materials; and imaging materials with neutron radiography. ORG 2 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. ORG 2 also houses a TRIGA Mark II nuclear research reactor designed by General Atomics. The reactor is the newest of the current fleet of U.S. university reactors that reached initial criticality in March 1992 [26]. The reactor is capable of steady-state operation at power levels up to 1 megawatt (MW) or pulsing mode operation in which much higher power levels (as high as 1500 MW) can be achieved for short durations (on the order of 10 microseconds).

Under normal operating conditions, ORG 2 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 ORG 2 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. ORG 2 provided a unique venue to gather empirical data to evaluate the efficacy of this proposed new perspective for ITDM. For example, the range of educational and research-based activities at ORG 2 provided opportunities to empirically describe operational workplace patterns in both more regular activities (e.g., those associated with course meeting times) and more irregular activities (e.g., those associated with individual experiments). More specifically, ORG 2 provided an opportunity to explore the use of an ANN to "learn" related operational workplace patterns to examine its ability to detect off-normal personnel activities—and (potentially) identify elevated insider threat risk levels across the facility.

To capture the set of data signals related to operational workforce patterns, the ReconaSense® Platform for Physical Security [27]—a commercially available ANN platform—was selected (and installed) at ORG 2. (NOTE: ReconaSense®— located in Austin, TX—had previously interacted with ORG 2 regarding some of its technical capabilities, which was a driver in selecting them as the commercial vendor for this project.) This initial investigation was predicated on exploring the efficacy of unsupervised neural network learning. More specifically, the ReconaSense® ANN system collects information from facility sensors for a period of time in order to generate a training dataset. The dataset size increases until an internal performance metric, likely based on validation error, is met. Validation error refers to the ANN prediction error on a dataset not seen in training. Evaluating performance using validation error helps avoid common ML problems such as overfitting. Once a sufficiently large dataset is collected, the ANN weights are tuned through an optimization algorithm such as stochastic gradient descent. More specifically, the ReconaSense® ReconAccess [28] software uses its ANN algorithm to collect data from sensors already deployed to learn the flow of people and processes at a facility, identify abnormal events, and alert an operator to a possible threat. This ANN integrates into a facility's existing security posture and implements role-based access controls with risk-adaptive access controls to ensure that all personnel with access to a facility only visit the appropriate areas. The software manages security by managing risk and flagging and reporting anamolous events.

The ReconaSense Platform incorporates an ANN that provides real-time analytical processing of aggregated data from a wide array of intelligent sensors, access control activity, video systems, and big data repositories (Figure 5). The ANN works on a few basic principles. It receives a wide array of input data that affect internal variables. The ANN evaluates the data based on internal information and sends results to several output variables. The output response can back-propagate and start a new process with that additional intelligence. The system treats everything as a data point (access control data, intrusion sensor data, etc.). It continually records all events from the sensors as individual entities in a database. The ANN then evaluates any new events with respect to the training dataset and determines the degree to which that event correlates with the expected behaviors shown in the training dataset. The degree to which an event is inconsistent with the training data determines a degree of automated risk modification for the facility (or area in a facility) and can flag the event as an anomaly and notify facility personnel. Since all new events get incorporated into the existing data, if an event that is flagged as an anomaly gets cleared as a correct access, then this new event changes the bounds of the training data and the system then learns that this new event is not a threat event. From this standpoint, the ANN evolves to incorporate new data and learns how to identify anomalies and ignore false alarms.

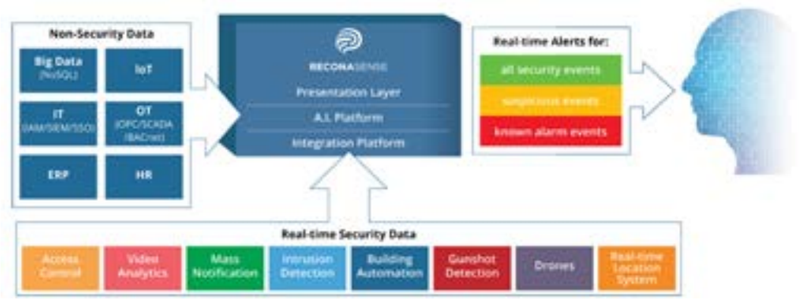


Figure 5: Data Flow Schematic for ReconaSense Platform.

To Consistent with an exploratory approach, the ANN  was used to capture ORG 2 facility data already being collected and focused signals related to access control and intrusion detection. Access control data collection focused on understanding when and how frequently authorized personnel enter a given area, as well as attempts at unauthorized access. At ORG 2, this data is collected via identification badge readers assigned to authorized personnel and corresponding databases that contain each individual's different levels of access. Collected as independent data streams, simple statistical tests could be applied to measure deviations from baseline patterns of individual access points. Yet, taking these access control points as interdependent data streams represents a more complex scenario that could leverage ANNs to help ascertain a deeper understanding of expected profiles of movement throughout ORG 2. Similarly, intrusion detection collection focuses on understanding the implications of registered movement in protected areas under different ORG 2 operational states. Measuring the deviation of individual intrusion detection sensors—which consisted of area motion sensors and balanced magnetic switches—does not require sophisticated ANN-based approaches. Considering the intrusion detection data in conjunction with other data signals (the access control data, for example) posits a more interesting and complex problem that supports an ANN-based solution. Table 1 summarizes the sensor and data type collected by the ANN, as well as representative activities expected to be captured by each sensor at a facility such as ORG 2. (NOTE: While this phase of the research did not include integration of camera signals or area radiation monitoring instrumentation, these data signals could easily be included in a future phase. See Section 5.)

Table 1: Description and categorization of data related to a representative set of expected organizational activities at ORG 2.

| ITDM Category | Sensor Type | Data Type | Representative Organizational Activity |
|---|---|---|---|
| Access Control | Badge reader<br>• ORG 2 entry<br>• Security control panel<br>• Limited area<br>• Reactor control room | Badge readers:<br>• # authorized attempts<br>• # unauthorized attempts (false negative + false positives)<br>• Time of access attempts | • Personnel arrival to facility<br>• Researchers approaching the reactor<br>• Reactor operator arriving for shift |
| Intrusion Detection | Balanced magnetic switch<br>• Limited area<br>• Security control panel<br>• Reactor control room<br><br>Area motion sensor<br>• Reactor bay<br>• Fuel storage surveillance | Balanced magnetic switches:<br># times switch opened<br>Time at which switch opened<br><br>Area motion sensors:<br># times change in physical phenomena registered<br>Time at which change in physical phenomena registered | • Researchers approaching the reactor<br>• Maintenance of security control panel<br>• Reactor operator arriving for shift<br><br>• Custodial services around the reactor<br>• Transfer of fresh/used fuel into/out of ORG 2 |

## 3.2 Data Collection Strategy

Baseline data to train this ANN on normal ORG 2 operations was collected from late 2019 to early fall 2020. For the purpose of observing trends in the access control patterns, data was collected in multiple phases. Phase I data was collected from October 12, 2019 to January 10, 2020 (but excluded December 23, 2019 to January 9, 2020—a university holiday when access to ORG 2 was significantly more sparse). Phase II data was collected from January 11, 2020 to September 25, 2020. To describe operational workplace patterns at ORG 2, access control data points were collected primarily from access credential readers, while intrusion detection data points were collected from area motion sensors and balanced magnetic switches throughout ORG 2. In total, 32,636 access control data points and 1,617 intrusion sensor data points were collected during Phase I and Phase II (Table 2). The data collected from XX to XX was used for very basic "training" of the ANN.

Table 2: Summary of artificial neural network data collected from ORG 2.

| Data Characteristic | Phase I Data Set | Phase II Data Set |
|---|---|---|
| Date range | Oct. 12, 2019 to Jan. 10, 2020[a] | Jan. 11, 2020 to Sep. 25, 2020[b] |
| Access control data points | 13,653 | 18,986 |
| Intrusion detection data points | 694 | 923 |
| Categories for organizing data points[c] | SAP<br>TSMAP | SAP<br>TSMAP |

[a] Excluding 12/13/19 to 01/09/20
[b] Acknowledging the near complete lack of student accesses between Mar. 15, 2020 and Jul. 1, 2020 due to COVID-19 precautions
[c] SAP = single-access-point operational patterns; TSMAP = time-sequences, multiple access point operational patterns

These data points were loosely organized into two categories to observe trends in the bounds of the ORG 2 operational patterns. The first category consisted of single-access-point operational patterns, in which the access control data were organized by access point, date and time of allowed access, and identity used for access. This data category used sensor observations to produce ANN-reported patterns in time to identify bounds for when general access is expected to occur for an average individual as well as for specific individuals. Deviations manifest as attempted accesses outside of these empirically defined time bounds. The second category consisted of time-sequenced, multiple-access-points operational patterns, in which the access control data were organized by identity used for access, date and time of allowed access, and by access point. This data category used sensor observations to produce ANN-reported patterns in time to identify bounds for when particular individuals would be expected to complete a particular access sequence. In this data category,

[1] The ReconaSense® was used to control and monitor the ORG 2 access control system, where operations were conducted in such a manner that supported testing of the ANN while ensuring that required security functions were not disturbed.

deviations manifest as either attempted accesses outside of empirically defined time boundaries or attempted accesses in a different order than in empirically defined baseline patterns.

All collected data in this study was anonymized; any "individual" analysis was conducted on a genericized average of several individuals; and a majority of the analysis was conducted at higher levels of data aggregation. (Note: Anonymization of individuals was conducted according to best research practices [29]. If such an ANN were deployed, then individual identities would be another data stream incorporated into the learning process—and mitigation strategies.) This includes evaluating operational workplace patterns of all facility personnel or those of facility personnel categories—small subsets of facility personnel with similar functional roles within the facility (e.g., faculty, operational staff, or students). While some empirical detail was lost, incorporating these facility personnel categories provided the opportunity to explore different operational workplace pattern profiles across personnel roles while respecting the anonymity of individuals associated with ORG 2.

## 3.3 Experimental Scenario Development

In addition to this "proof of concept" data categorization, the ANN was evaluated against three different scenarios related to insider threat mitigation using data collected from sensor types described in Table 1. In each scenario, ORG 2 personnel were tasked with carrying out a specific action within a set window of time. This allowed the research team to observe the ANN's performance and evaluate its ability to detect anomalous behaviors. In Scenario (1) the insider is attempting to gain access to the closet that contains ORG 2's intrusion detection system panel, which is protected by the ORG 2 physical security and access control systems. This panel is comprised of all incoming intrusion sensor data and processing, as well as alarm signal communication to the central alarm station (CAS). In this scenario, if an insider gains access to this panel, it is assumed they could sabotage the panel to eliminate/falsify alarm signals at the CAS—reducing the possibility for assessed detection and greatly increasing the success of a subsequent outsider or insider attempting an attack on the facility at some later date/time.

In Scenario (2) the insider is trying to gain access to the reactor bay during off-hours. During such off-hours, the reactor bay is locked and alarmed. This scenario assumes that the insider has authorized access to the ORG 2 building, but not to the reactor bay inside the ORG 2 building. It is hypothesized that the insider uses their authorized credentials to access the ORG 2 building and then attempts entry to the reactor bay. When combined with other ANN-generated insights like the time (e.g., off-normal hours) or expected profile of access (e.g., credential that entered ORG 2 does not have reactor bay access), this behavior could be flagged as a possible elevation in insider potential to the facility and sent to the CAS for assessment.

In Scenario (3) the insider is trying to acquire knowledge pertaining to the security system for ORG 2's fuel storage facility. To complete this task, the insider needs to surveil the area around the stored fuel and then test the alarm systems to determine what level of activity will set off storage location access alarms. Such testing of the alarm system could include the intrusion detection sensors, area radiation sensors, cables/conduits for those sensors, and the alarm panel. This scenario was designed to include evaluating the ANN's ability to identify potential insider surveillance and potential insider testing alarms/sensors to determine sensitivity levels.

In these three scenarios, both single-access-point and time-sequenced, multiple-access-point operational patterns were incorporated to evaluate the ANN's ability to support ITDM. Within these analyses, deviations from expected operational patterns included single-access-point patterns outside of ANN-defined bounds (e.g., anomalous access control sensor data) and time-sequenced, multiple-access-point patterns (e.g., unusual combinations of sensor signals). Overall, the data collected across the two phases of this research project supported training the ANN to identify and define expected operational patterns at ORG 2, developing additional models and performance measures for ITDM, and analyzing these ANN-learned operational patterns against a set of hypothesized insider threat-related scenarios.

When combined with other ANN-generated insights like the time (e.g., off-normal hours) or expected profile of access (e.g., credential that entered ORG 2 does not have reactor bay access), this behavior could be flagged as a possible elevation in insider potential to the facility and sent to the CAS for assessment.

# 4 RESULTS & ANALYSIS

## 4.1 Operational Workplace Pattern Learning Analysis for the Artificial Neural Network

Despite the anticipated irregularity in the operations of ORG 2 as a teaching facility composed of a diversity of personnel, including operational/administrative staff, faculty, graduate/undergraduate students, and visitors, Phase I data illustrated the ability to establish bounds for the operational workplace patterns that supported the potential detection of insider actions by deviations from these bounds (e.g., anomaly detection). One anticipated organizational pattern relates to the time when personnel first enter the ORG 2 facility—with the expectation that the data would be fairly tightly clustered around 7:00 a.m., the traditional start of a professional workday. For example, consider Figure 5 showing the frequency of distribution of the first allowed access to the ORG 2 facility versus the time of day for Phase I data as a representative single-access-point metric. As illustrated, there are clear bounds on an expected first facility access, for both normal working days and non-working days (e.g., holidays and weekends). Even though the bounds illustrated in Figure 6 are wider than anticipated, it still demonstrates the ability of the ANN to define expected operational workplace patterns and establish a baseline to measure potential deviations.

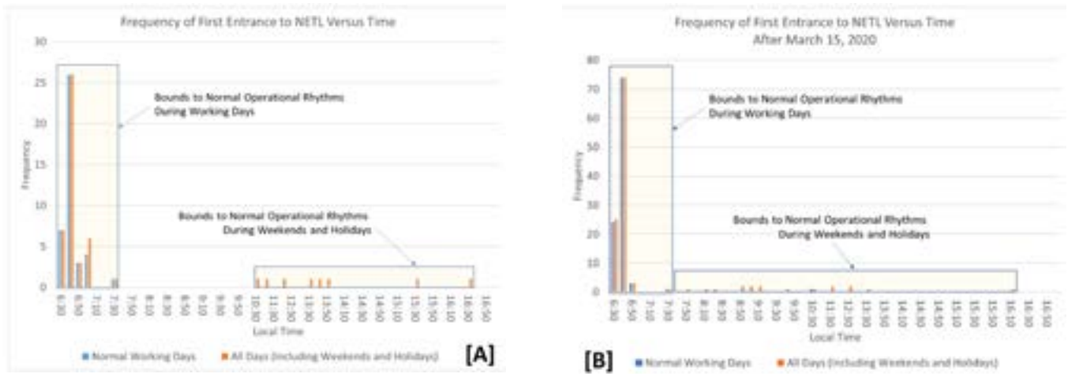


Figure 6: Frequency distribution showing time of first entrance to ORG 2 facility, comparing "working days" and "all operational days" with Phase I data in [A] and Phase II data in [B].

Finer, group-level details were observed by further parsing the single-access-point data. As shown in Figure 4, the Phase I data was disaggregated to determine the time of first entry to ORG 2 by each personnel group. As illustrated, visualization helps identify category-specific operational workplace 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 measured for other access points within ORG 2—represent expected arrival profiles per personnel group and provide another collective behavior-based profile to detect deviations that may reflect a potential insider act.

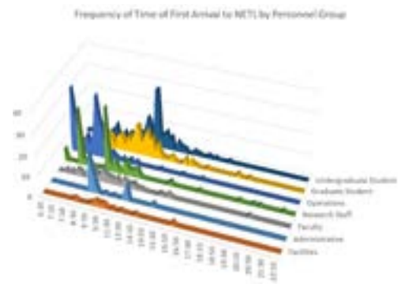

Figure 7: Frequency distribution showing time of first entrance to ORG 2 during data collection time and separated by personnel group.

With the Phase I data, time-sequenced, multiple-point operational patterns were analyzed with respect to the time between accessing entry points along a given path. Consider, for example, focusing on the observed behaviors of a single operator who is normally the first person to arrive at the facility. The ANN would register these observations and determine an expected routine of accessing entry points A, B, C, D, and E—resulting in an emergent, dynamic pattern of expected behavior. This individual usually arrives at entry points within a short window of 450 seconds—97% of the time for the first four entry points and 71% of the time for the fifth entry point. Figure 7 shows the observed frequency distribution of the time delay between successful authenticated access at each entry point. More specifically, this frequency distribution shows that, based on this representative training data set, the ANN would expect this individual to follow access point A with an authentication to access point B within 42-66 seconds. It similarly expects the individual to continue to access point C and D with only very short delay times (less than 12 and 6 seconds, respectively). Note, however, that there are outlier data points in which the individual did not immediately proceed to access point C after clearing access point B.

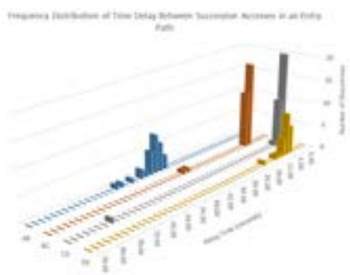

Figure 8: Frequency distribution showing time-series plot of time delays between authenticated access along a series of entry points (A, B, C, D, E) the compose the expected behavior of the first person to enter the facility.

## 4.2 Scenario Analysis Results for ANN-Based Insider Threat Detection & Mitigation

A series of scenario-based experiments were designed to further explore the efficacy of this ANN-based approach to ITDM. Across three different target locations within ORG 2—the closet housing the security system's control panel, the reactor bay, and nuclear fuel storage area—the scenarios introduce varying levels of experimental sophistication. In addition to the variability (and complexity) in the scenarios, the experiments incorporated additional insights by exploring the ANN's efficacy against a range of potential insider pathways to the target locations. The pathways are summarized in Table 3 below.

Table 3: Possible insider threat pathways to support experimental scenario analysis got ANN-based insider threat detection & mitigation.

| Possible Insider Pathway | Logical Description | Notional Example |
|---|---|---|
| [A] | Attempt of direct access by an unauthorized individual (testing both single attempts and repeated attempts both during normal working hours and on off-hours) | Bob (who is not allowed to access the panel) uses his own credential in an attempt to access the intrusion detection system panel |
| [B] | Attempt to access by an unauthorized individual during off-hours using credentials from an authorized individual but using their own credentials to access the ORG 2 building | Bob (who is not allowed to access the intrusion detection system panel) uses his credential to enter the building and uses Fred's credential (who is allowed access to the panel) in an attempt to access the intrusion detection system panel |
| [C] | Attempt to access by an unauthorized individual during off-hours using credentials from an authorized individual and using that individuals credentials to access the ORG 2 building | Bob (who is not allowed to access the intrusion detection system panel) uses Fred's credential (who is allowed access to the panel) to enter the building and in an attempt to access the intrusion detection system panel |

For Scenario (1), analysis was conducted on data collected from access control readers at the facility entrance and near security-related control systems. In Scenario (1), the ANN acted as an ITDM system by attempting to detect such off-normal activity as an insider attempting to gain access to the closet containing the ORG 2 intrusion detection system panel—and by electronically engaging locking mechanisms to deny access to any unauthorized attempt. Using a combination of single-access-point and time-sequenced, multiple-access-point operational patterns, these three tests of this hypothesized insider scenario were evaluated with the ANN. Because each case for test (1A) resulted in the insider attempt at unauthorized access being detected and denied, these results benchmark this ANN approach to well established capabilities (and performance) of traditional access control systems. The other two hypothesized manifestations of this scenario were similarly evaluated, where unauthorized access was detected and denied in most cases for test (1B) but in no cases for test (1C). Scenario (1) results are summarized in Table 4.

These Scenario (1) results match intuition, as tests (1A) to (1C) represent increasingly complex insider tactics. These results further support the use of traditional protective measures—in this scenario, layers of access controls—to mitigate the potential of unauthorized access to key locations within a nuclear facility. While not all attempts at unauthorized access were detected and (electronically) denied in tests (1B) and (1C), this illustrates two additional key findings. First, the increased success of increasingly complex insider tactics to gain unauthorized access suggests a need to augment applications of multi-layer, credential-based access controls. Second, an ANN-based solution that incorporates additional data streams could augment the ability of the credential-based access controls evaluated in Scenario (1) to detect and deny unauthorized access.

In Scenario (2) analysis was conducted on data collected from multiple access control readers leading to the reactor bay and motion sensors within the reactor bay itself. Here, the ANN acted as an ITDM system looking for off-normal activity that would include not only attempts at unauthorized access (similar to the results from the first scenario), but also early detection of the insider moving toward the reactor bay. More specifically, test (2A) used the ANN to identify unauthorized access to the reactor bay itself, while test (2B) was designed to identify suspicious movement toward the reactor before any attempt to bypass reactor bay access control station(s). Again, using a combination of single-access-point and time-sequenced, multiple-access-point operational patterns, these tests of Scenario (2) were analyzed with the ANN. In test (2A), for example, multiple-access-point patterns based on access control reader data were used to identify unacceptable deviations in expected operational workplace patterns related to unauthorized reactor bay access. Similar to test (1A), the insider attempting unauthorized access into the reactor bay was detected and denied in every case. Test (2B) was not evaluated with Phase I data due to time constraints, but the results from the Phase II data indicate that such motion detection sensors augment the data used by the ANN to improve its capability, resulting in the ANN detecting attempts at unauthorized reactor bay access before reaching a reactor bay entry point in most cases. All Scenario (2) results are summarized in Table 4.

Similar to Scenario (1), these results match intuition, where tests (1A) and (2A) each represent the simplest insider tactic and further support the use of traditional protective measures like credential-based access controls to mitigate the potential of unauthorized access to key locations within a nuclear facility. One additional key finding from Scenario (2)—particularly in test (2B)'s ability to leverage motion detector data outside of the reactor bay—is the demonstration of ANN-identified deviations to highlight potential suspicious insider activities (e.g., unauthorized reactor bay access) before they occur. Coupling this motion detector data and reactor bay access control data formed a time-sequenced, multiple-point operational pattern that could be used as a learned "profile" of deviation for the ANN. This demonstration introduces the ability to create more sophisticated profiles of operational workplace patterns and better train the ANN to detect deviations using this time-sequenced, multiple-access-point operational pattern approach—suggesting that incorporating more data streams into such patterns would provide higher fidelity learning to the ANN and more nuanced detection of deviations.

Lastly, in Scenario (3), analysis was conducted on data collected from motion detection sensors, reactor bay access controls, intrusion detection sensors, area radiation sensors, and alarm panel sensors. The ANN acted as an ITDM system by analyzing these data to surveil for off-normal personnel activity, in particular for signs of multiple attempts to access the

fuel storage area. Scenario (3) was designed as a more nuanced set of tests to evaluate the ability of an ANN-based ITDM approach to detect precursor activities to unauthorized access. More specifically, test (3A) sought to demonstrate the ability of the ANN to detect evidence of insider surveillance activities within the reactor bay, and test (3B) aimed to demonstrate the ANN's ability to detect insider testing of various alarms related to accessing nuclear fuel storage. For Phase I data, the ANN struggled to sufficiently detect insider surveillance activities in test (3A), and despite additional data points from Phase II, these results hold true and indicate the need for an additional data course for the ANN to better detect possible insider surveillance. Similar to test (2B), test (3B) was not evaluated with Phase I data due to time constraints, but the results from the Phase II data indicate that such multiple-point patterns derived from multiple sensor types can detect and deny insider alarm testing in many cases. However, the ANN is limited in its ability to detect this type of sensor testing due to hardware constraints—since the sensor signal has a very high nuisance rate and because identification of who is located where in a room is not possible with the existing system.

The Scenario (3) results introduce a new element to insider threat mitigation. Though both scenario tests had challenges—including adequately distinguishing "insider surveillance" from normal operational workplace patterns—these early results illustrate the potential benefit of ANN-based approaches for detecting a range of activities precursory to an attempted insider action. Yet, the use of an expanded set of data streams would establish more complex time-sequenced, multiple-access-point operational patterns of behavior that could better distinguish insider surveillance or alarm testing activities. For example, this ANN-based ITDM approach would be greatly enhanced with the use of personal dosimeters—which record position information—to uniquely identify the position of individuals in the facility.

Since Phase I consisted of limited testing data, its evaluation primarily consisted of demonstrating the validity and acceptability of the scenarios to provide useful results in evaluating the ANN's capability to support ITDM. For example, consider the success rate in detecting and denying all attempts at unauthorized access to the panel closet and the reactor bay, as well as the success rate in detecting and denying access to the panel closet by an unauthorized individual (with access to the ORG 2 building) using an authorized credential. The results of testing the scenarios against Phase II data showed both strong consistency with Phase I results and more nuance and higher fidelity—for example the ANN's ability to detect abnormal motion and flag that as a potential cause for concern.

Table 4 summarizes the ANN performance results of all hypothesized insider-threat-related testing scenarios.

Table 4: Summary results from insider-threat-related testing scenarios.

| Scenario Name [#] | Test Description | Phase I Results* | Phase II Results |
|---|---|---|---|
| Security Closet Access (1) | Unauthorized Access Attempt (1A)<br>Authorized Access Credentials Used by Unauthorized Individual Who Entered Building Using Their Own Credentials (1B)<br><br>Authorized Access Credentials Used by Unauthorized Individual Who Entered Building Using Authorized Individual's Credentials (1C) | Detected & Denied in ALL Cases [SAP]<br><br>Detected & Denied in MOST Cases [SAP; TSMAP]<br><br>Detected & Denies in NO Cases [TSMAP] | Detected & Denied in ALL Cases [SAP]<br><br>Detected & Denied in MOST Cases [SAP; TSMAP]<br><br>Detected & Denies in NO Cases [TSMAP] |
| Reactor Bay Access (2) | Unauthorized Access to Reactor Bay (2A)<br>Early Detection by Motion Sensor (2B) | Detected & Denied in ALL Cases [TSMAP]<br><br>Not Tested | Detected & Denied in ALL Cases [TSMAP]<br><br>Detected in MOST Cases |
| Fuel Storage Surveillance (3) | Insider Surveillance (3A)<br>Insider Alarm Testing (3B) | Difficult to Detect Without Additional Sensing Input [TSMAP]<br><br>Not Tested | Difficult to Detect Without Additional Sensing Input [TSMAP]<br><br>Difficult to Detect Without Additional Sensing Input [TSMAP] |

*SAP = single-access-point operational patterns; TSMAP = time-sequenced, multiple-access-point operational patterns

These combined results suggest that a "new" ITDM framework could be successful at detecting insider actions (or precursor activities) based on deviations in operational workplace patterns. More specifically, a new insider threat framework seems to emerge from these empirical insights—namely that insider threat can be identified based on a measured (and therefore quantifiable) change in operational workforce patterns. Such a data-driven framework can help build upon more traditional ITDM approaches that are deeply reliant on human judgment of the behaviors of others. For example, deviations from ANN-determined (e.g., "learned") baseline operational workplace patterns can lead to undesired personnel (or organizational) behaviors and facility performance. Small deviations may result from innocuous causes (e.g., personnel complacency) or indicate organizational evolution. Large deviations, however, may be indicative of more intentional or malicious activities. Using physical sensors to understand deviations can also potentially provide insight into the motivations of a potential insider threat. For example, an insider "testing" doors that they are not authorized to enter or coming into the facility long after their normal hours could suggest a stark behavioral change—possibly indicating a turn toward malicious intent—that should be investigated by management. By measuring these deviations, an ANN-based ITDM approach can quantify anomalous behavior that may be indicative of insider potential—thereby suggesting the ability to communicate insider risks in (near) real-time. Further, this quantified ITDM concept also allows organizations to target security measures and investigations towards specific deviations most likely to constitute a threat. While broader behavioral monitoring and HRP should remain in place, this quantifiable ITDM program would support more efficient and effective use

of security resources.

In addition, an ANN-based ITDM framework can be tailored to specific facilities based on their own operational workplace patterns and threshold of insider risk acceptance. Moreover, these thresholds for delineating between acceptable and problematic deviations can be calculated in two distinct manners. First, deviations from expected operational patterns can be defined as the absolute difference between expected and observed operational workplace pattern profiles (e.g., custodial staff with fairly rigid routine operations). This calculation would provide lower overall uncertainty in measuring the actual deviation—akin to common uses of moving averages and seasonality approaches—but may experience higher false positive rates given the propensity for human behaviors to drift over time [16]. Second, deviations can be defined as relative differences to learned operational workplace patterns. This calculation considers the likelihood that baseline personnel behavior profiles evolve over time and utilizes ANN machine learning to regularly update its definition of "expected" operational workplace patterns. This approach would likely provide slightly higher uncertainty in measuring deviations but far fewer false positives. These multiple levels of customization would allow facilities to adjust to their own staff and facility needs and utilize collected data signals already in place without a significant investment in new security hardware. While additional research needs to be completed to fine tune best practices and recommendations for this new ITDM, these results suggest that such an ANN-based approach can enhance current ITM, and the concept of insider potential should continue to be investigated (and refined).

Most of the scenarios analyzed were relatively simple and consisted of deviations using only a few features—where traditional statistical tests are applicable. Yet, these approaches would likely be ineffective in more complex scenarios that measure additional data-derived deviations from expected operational patterns. The ability of the ANN configuration deployed for this research to both detect and deny insider actions in the simpler scenarios is a proof-of-concept. In the scenarios where the ANN was unable to adequately detect the insider action, the results indicated that incorporating additional sensor data could better capture deviations in operational workplace patterns in order to detect potential insider actions. For example, the results of test (1C) show a successful insider act in which an unauthorized individual gained access to the reactor bay using the credentials of an authorized person was not detected by the ANN. While this deviation of expected operational workplace patterns was not detected by the current ANN configuration, the ReconaSense® software has the ability to incorporate additional data streams—including video feeds and facial recognition technology—that could build a more nuanced profile of time-sequenced, multiple-access-point operational patterns to better detect such a deviation. From this ITDM perspective, additional data would expand the ANN's understanding of expected operational workplace patterns and enhance the ANN's learning to detect—and deny (where it can) continued actions for—deviations from those patterns. Additional data would also expedite the ANN's ability to communicate the existence of an anomaly, and potentially an insider action, for appropriate response.

While elements of traditional ITM will still be important within the context of an updated framework, these results show that traditional protective ITM measures can be supplemented with those proposed in the new framework to enhance overall ITDM at nuclear facilities. For example, more traditional ITM programs would likely have similar access controls but would not be able to understand when abnormal access requests suggest a new pattern of behavior. The demonstrated ability of the ANN to detect deviations from expected operational workplace patterns supports a collective-based approach to identify new patterns of behavior that could represent malicious insider actions. This ability separates the proposed ITDM framework from traditional approaches, and this study's results suggest an improvement in identifying insider attempts.

ReconaSense® software has the ability to incorporate additional data streams—including video feeds and facial recognition technology—that could build a more nuanced profile of time-sequenced, multiple-access-point operational patterns to better detect deviations in operational workplace patterns in order to detect potential insider actions.

# 5 CONCLUSIONS, INSIGHTS, & IMPLICATIONS

The results of evaluating Phase I and Phase II data indicate that an ANN can identify and define obvious operational workplace patterns—based, in this case, on time-series access control data—in support of advancing ITDM. This ANN-based approach illustrated how collections of single-access-point and time-sequenced, multiple-access-point operational patterns provide data-driven profiles of expected personnel behavior, thus forming a baseline from which deviations or anomalous behaviors can be detected. Further, applying this ANN-based approach to a representative set of insider threat scenarios demonstrated the efficacy of this data-driven approach across a range of increasingly complex tests. Moving to a data-driven framework for identifying insider threat potential would help organizations improve upon their sole reliance on ITM programs based on human judgment, personnel behavioral observation programs, or individual psychological stressors/indicators. In addition to leveraging many elements of traditional ITM programs into a more data-driven, quantitative approach to ITDM, ANN-based approaches also provide additional benefits. First, a more data-driven approach could help remove bias in observing individual psychological stressors/indicators. For example, humans are prone to influence from both hindsight bias (e.g., the human tendency for humans for perceive past events as being more predictable [30]) and the prevalence effect (e.g., the phenomenon by which humans are more likely to dismiss a target with low frequency rather than one with high frequency [31]). Whereas neural networks can make decisions based on data analytics. Second, a more data-driven approach might mitigate how personal relationships and cultural attitudes surrounding reporting tend to negatively impact traditional ITM programs. Historically, after-action reports from many insider attacks suggest that obvious signs were ignored or unreported by people close to the individual perpetrating the insider action [32]. Even when behavioral reporting systems were in place, obvious signs were dismissed, rationalized, or disregarded on the grounds of existing personal or professional relationships. By quantifying deviations from expected operational workplace patterns, ITDM programs can more objectively investigate such changes, understand the motivations behind them, and present empirical evidence to further legitimize their investigations.

While these results suggest that the ANN provides an opportunity to move to a new ITDM framework based on quantitatively identified deviations from operational workplace patterns, future research needs to address a number of limitations. These limitations tend to fall into two categories: technical limitations and human impact (policy) limitations. Technical limitations include the need to test additional sensing data, understand how additional data will impact ANN learning, and evaluate ANN performance in defining deviations. While these results also, in many cases, relied on numerical measures for deviations from expected operational workplace patterns, analytical complications may arise from incorporating new data streams.

Follow-on research in a potential Phase III would look to incorporate ORG 2's area radiation monitor data, which is already networked in the facility. Feeding those signals to the ANN should be possible with no hardware modifications. In addition, Phase III could incorporate closed-circuit television camera footage, and facial recognition technology could help to identify important deviations—like an employee using a badge that does not belong to them (test [1A])—that do not necessarily fit neatly within a quantitative definition of a deviation.

Such new data streams would clearly improve the accuracy of defining insider potential, but also offer challenges in terms of transparently translating these impacts into quantitative measures. Yet, even a new data-driven ITDM approach cannot operate without human impacts and bias—necessitating a strong understanding of the impacts on security personnel, the workforce, and morale as deviations in operational workplace patterns are uncovered. For example, employees who know such a system is in place may act differently than those who do not know. This could lead to potential insiders purposefully changing their behaviors in order to "teach" the ANN new operational workplace patterns—an extension of the classic challenge of addressing the fact that insiders may know the protective measures in place. In this manner, this persistent challenge to countering insider actions will need additional empirical evaluation and updating of the logic supporting this data-driven, ANN-based approach to ITDM.

As indicated in the trends between Phase I, Phase II, and combined results, the more training data ingested by the ANN, the better it can learn and describe more robust and nuanced operational patterns. More robust operational workplace patterns also serve to extend this research by offering higher fidelity results on more rigorous scenario testing. Next steps could include, for example, a set of controlled experiments (including control groups) extending the three scenarios evaluated in this study to more formally assess how well anomaly detection capabilities support insider threat detection across different commercially available ANN options. For example, asking a graduate student to enter the reactor control room at 11:30 p.m. on the second Tuesday of the month and asking (1) how easily is this known anomaly observed in the data, (2) did the ANN register the anomaly as a deviation from expected operational workplace patterns, and (3) did the ANN correctly classify the anomaly against pre-determined operational workplace patterns defined as profiles of increased insider potential. If successful, follow-on efforts could include series of more varied, less controlled experiments based on these same scenarios—in which a graduate student would enter the reactor control room sometime after 9 p.m. during the first week of the month—in order, for example, to assess ANN anomaly detection sensitivity.

While additional studies are needed to fully understand and characterize the benefits of such an approach, the results of this initial study show promise for commercially available ANNs to improve ITDM from this collective-behavior-based approach. This data-driven approach is also beneficial in terms of leveraging data signals already collected at nuclear facilities to build profiles of expected operational workplace patterns. Further, introducing this new insider potential framework also provides an opportunity to use traditional concepts of risk management to quantitatively describe the susceptibility of facilities to malicious insider actions. Ultimately, this research indicates a very promising approach to expand and extend ITDM utilizing ANNs and data analysis techniques.

# ACKNOWLEDGMENTS

SAND2020-PEER REVIEW

## REFERENCES

[1] National Insider Threat Task Force. 2016. Protect Your Organization from the Inside Out: Government Best Practices. NITTF Report.

[2] Federal Register Vol. 76, No. 198. 2011. Presidential Documents. Retrieved from: https://www.dni.gov/files/NCSC/documents/nittf/EO_13587.pdf

[3] National Insider Threat Task Force. 2017. Insider Threat Guide: A Compendium of Best Practices to Accompany the National Insider Threat Minimum Standards. Washington, DC.

[4] Anonymous, 2019.

[5] International Atomic Energy Agency. 2008. Preventive and Protective Measures Against Insider Threats. IAEA Nuclear Security Series No. 8: Implementing Guide.

[6] World Institute for Nuclear Security. 2018. Countering Violent Extremism and Insider Threats in the Nuclear Sector.

[7] Carolynn P Scherer, Christy E. Ruggiero. 2019. Overview of Tools for Insider Threat: Analysis and Mitigation. Retrieved from: https://permalink.lanl.gov/object/tr?what=info:lanl-repo/lareport/LA-UR-19-22069

[8] John E. Landers. 2014. Psychological Profiles of the Malicious Insider. PNNL SA 102669, Pacific Northwest National Laboratory, Richland, WA. Retrieved from: https://www.osti.gov/servlets/purl/1130774

[9] A. Kolaczkowski, J. Forester, E. Lois, S. Cooper. 2005. Good Practices for Implementing Human Reliability Analysis (HRA). NUREG-1792. U.S. Nuclear Regulatory Commission. Retrieved from: https://www.nrc.gov/docs/ML0511/ML051160213.pdf

[10] Kyo-Nam Kim, Man-Sung Yim, & Erich Schneider. 2017. A Study of Insider Threat in Nuclear Security Analysis Using Game Theoretic Modeling. Annals of Nuclear Energy, 108 (October 2017), 301-309. DOI: https://doi.org/10.1016/j.anucene.2017.05.006

[11] Bowen Zou, Ming Yang, Jia Guo, Junbo Wang, Emi-Reynolds Benjamin, Hang Liu, & Wei Li. 2018. Insider Threats of Physical Protection Systems in Nuclear Power Plants: Prevention and Evaluation. Progress in Nuclear Energy, 104 (April 2018), 8-15. DOI: https://doi.org/10.1016/j.pnucene.2017.08.006

[12] Greitzer, Frank & Imran, Muhammad & Purl, Justin & Axelrad, Elise & Leong, Yung & Becker, Sunny & Laskey, Kathryn & Sticha, Paul. 2016. Developing an Ontology for Individual and Organizational Sociotechnical Indicators of Insider Threat Risk. In Proceedings from The Eleventh International Conference on Semantic Technology for Intelligence, Defense, and Security (STIDS 2016). November 15-16,2016, Fairfax, VA, 19-27.

[13] Anonymous, 2020.

[14] Richard M. Cyert and James G. March. 1963. A Behavioral Theory of the Firm (2nd. Ed.). Prentice-Hall, Malden, MA.

[15] Anthony Giddens. 1984. The Constitution of Society: Outline of the Theory of Structuration. University of California Press, Berkley, CA.

[16] Jens Rasmussen. 1997. Risk management in a dynamic society: A modelling problem." Safety Science 27, 2 (November-December 1997), 183–213. https://doi.org/10.1016/S0925-7535(97)00052-0

[17] Fei-Fei Li, Ranjay Krishna, Danfei Xu. 2020. CS231n: Convolutional Neural Networks for Visual Recognition. Retrieved from https://cs231n.github.io/neural-networks-1/#feedforward

[18] Dan Li, Dacheng Chen, Jonathan Goh, and See-kiong Ng. 2018. Anomaly Detection with Generative Adversarial Networks for Multivariate Time Series. arXiv: 1809.04758. Retrieved from: https://arxiv.org/abs/1809.04758

[19] Zhiwei Wang, Zhengzhang Chem, Jingchao Ni, Hiu Liu, Haifeng Chen, and Jiliang Tang. 2020. Multi-Scale One Class Recurrent Neural Networks for Discrete Event Sequence Anomaly Detection. arXiv: 2008.13361. Retrieved from: https://arxiv.org/abs/2008.13361

[20] Alexey Bochkovskiy, Chien-Yao Wang, Hong-Yuan Mark Liao. 2020. YOLOv4, Optimal Speed and Accuracy of Object Detection. Retrieved from: arXiv: 2004.10934. Retrieved from: https://arxiv.org/abs/2004.10934

[21] Alexander Kolesnikov, Lucas Beyer, Xiaohua Zhai, Joan Puigcerver, Jessica Yung, Sylvain Gelly, and Neil Houlsby. 2019. Big Transfer (BiT): General Visual Representation Learning. arXiv: 1912.11370. Retrieved from: https://arxiv.org/abs/1912.11370

[22] Jascha Kolberg, Marcel Grimmer, Marta Gomez-Barrero, and Christoph Bush. 2020. Anomaly Detection with Convolutional Autoencoders for Fingerprint Presentation Attack Detection. arXiv: 2008.07989. Retrieved from: https://arxiv.org/abs/2008.07989

[23] Ian J. Goodfellow, Jonathon Shlens and Christian Szegedy. 2015. Explaining and Harnessing Adverserial Examples. arXiv: 1412.6572. Retrieved from: https://arxiv.org/abs/1412.6572

[24] Matthias Hein, Maksym Andriushchenko, and Julian Bitterwolf. 2019. Why ReLU networks yield high-confidence predictions far away from training data and how to mitigate the problem. arXiv: 1812.05720. Retrieved from: https://arxiv.org/abs/1812.05720

[25] Alexander Meinke and Matthias Hein. 2019. "Toward neural networks that provably know when they don't". arXiv 1909.12180. Retrieved from: https://arxiv.org/abs/1909.121180

[26] Anonymous. 2002

[27] John Carter. 2018. The ReconaSense AI Platform for Physical Security. ReconaSense Report, Austin, TX.

[28] ReconaSense. 2020. ReconAccess: Risk-Adaptive and Intelligent Access Control. Retrieved from: https://reconasense.com/reconaccess/

[29] Robert K. Yin. 2016. Qualitative Research from Start to Finish, 2nd Edition. Guilford Press, New York.

[30] Roese, Neal J., and Kathleen D. Vohs. 2012. Hindsight bias. Perspectives on psychological science, 7(5), 411-426. DOI: https://doi.org/10.1177/1745691612454303.

[31] J. Wolfe, D. Rubinstein, and T. Horowitz. 2013. Prevalence effects on newly trained airport checkpoint screeners: Trained observers miss rare targets, too. Journal of Vision, 13(3), DOI: https://doi.org/10.1167/13.3.33.

[32] Anonymous. 2019.

About ReconaSense

ReconaSense® is the first COTS risk-adaptive framework with FICAM certification that can uniquely help identify Insider Threats via real-time data transformation, evaluation and automated policy adjustments.  We protect federal, military, state, and other critical infrastructures requiring top-grade security with insider threat detection and our unification platform.

ReconaSense is privately held company headquartered in Austin, Texas.

Interested in a demonstration?
Email us at insider@reconasense.com or call us at +1 512.220.2010.

Learn more at www.reconasense.com

## Location

CORPORATE HEADQUARTERS
9130 Jollyville Road
Suite 150
Austin, TX 78759

RECONASENSE



ReconaSense is an American company headquartered in Austin, Texas.

EB-1-1