SAND2013-7011C

Risk-Informed Management of Enterprise Security: *Methodology and Applications for Nuclear Facilities*

Felicia A. Durán[*], **G.D. Wyss, S.E. Jordan, and B.B. Cipiti**
Sandia National Laboratories,[†] Albuquerque, New Mexico, 87185-0757, USA

Abstract. Decision makers wish to use risk analysis to prioritize security investments. However, understanding security risk requires estimating the likelihood of attack, which is extremely uncertain and depends on unquantifiable psychological factors like dissuasion and deterrence. In addition, the most common performance metric for physical security systems, "probability of effectiveness at the design basis threat" [*P(E)*], performs poorly in cost-benefit analysis. This makes it difficult to prioritize investment options on the basis of *P(E),* especially across multiple targets or facilities. To overcome these obstacles, work at Sandia National Laboratories has developed a risk-informed security analysis method. This methodology, Risk-Informed Management of Enterprise Security (RIMES), characterizes targets by how difficult it would be for adversaries to exploit each target's vulnerabilities to induce consequences. Adversaries generally have success criteria (e.g., adequate or desired consequences and thresholds for likelihood of success), and choose among alternative strategies that meet these criteria while considering their degree of difficulty in achieving their "successful" outcome. RIMES has been applied to evaluate the theft and sabotage risks for two types of nuclear fuel cycle facilities – used nuclear fuel (UNF) storage and small modular reactors (SMRs).

1. Introduction

For many years, safety investment decisions have been made using risk-based cost-benefit analysis in which the benefit metric is heavily based on a quantitative estimate of risk reduction. Many seek to perform similar analyses to prioritize security investments, but this has met with limited success, in part because the "attack likelihood" component of risk is often extremely uncertain and not considered when conditional security risk is assessed. Therefore, Sandia National Laboratories has developed a risk-informed security analysis method. This methodology, Risk-Informed Management of Enterprise Security (RIMES), characterizes targets by how difficult it would be for adversaries to exploit each target's vulnerabilities to induce consequences. The goal of this work was to enable security analysts to describe the benefits of security risk reduction measures based on the degree to which they increase the difficulty for an adversary to successfully prepare and execute an attack that can produce a given level of consequences. The resulting method is highly scalable and enables robust risk-based cost-benefit security investment prioritization to be performed at levels of granularity ranging from a single target up to multiple targets or facilities across an enterprise. Recently, RIMES has been applied to evaluate the theft and sabotage risks for two types of nuclear fuel cycle facilities – used nuclear fuel (UNF) storage and small modular reactors (SMRs). This paper discusses the development of the RIMES method and summarizes it application for UNF storage and SMR security.

2.0 Probabilistic Risk Assessment – A Brief History, Current and Extended Use for Security

In 1974, Norm Rasmussen from the Massachusetts Institute of Technology led a team from the Atomic Energy Commission to conduct the reactor *safety* study (WASH-1400) [5] in which they developed the concept of societal risk. The WASH-1400 study was published in 1975, and although widely criticized, it nonetheless established the foundational principles of probabilistic risk assessment (PRA) still widely used today. Shortly thereafter, a modified version of the societal risk model was first proposed for nuclear *safeguards (security)* [6]. Known as the ERDA-7 proposal, this approach was evaluated by Rasmussen, who concluded that safeguards (security) risk could not be quantified using the WASH-1400 developed societal risk approach [7]. Rasmussen said that he did not believe that risks involving malevolent human action could be quantified by traditional risk assessment methods like fault tree and event tree analysis because attack probability estimates could not meet important statistical requirements [7]. Over the years, the ERDA-7 proposal has been subject to reintroduction and modification [1, 8, 9, 10]. Similar to Rasmussen's conclusions, subsequent critical reviews stated an approach like ERDA-7 proposal based on traditional risk assessment not be

[*] Sandia National Laboratories, P.O. Box 5800 MS-0757, Albuquerque NM 87185-0757, United States
[†] Sandia National Laboratories is a multi program laboratory managed and operated by Sandia Corporation, a wholly owned subsidiary of Lockheed Martin Corporation, for the U.S. Department of Energy's National Nuclear Security Administration under Contract DE-AC04-94AL85000. SAND2013-5095C, Unclassified/Unlimited Release.

used for security risk [11, 12]. The ERDA-7 approach is problematic for intentional malevolent acts because the terms in the equation are interdependent and data is lacking, which results in large uncertainties. Instead of using of the ERDA-7 approach, performance-based standards for the effectiveness of security systems as well as addressing consequences were recommended as useful tools [7, 13].

2.1. Current Definition of Risk

Kaplan and Garrick [14] stated the definition of risk that is most commonly used among modern risk analysts as, "Fundamentally... a risk analysis consists of an answer to the following three questions: *(1) What can happen? (2) How likely is it that [it] will happen? and (3) If it does happen, what are the consequences?* To answer these questions we would make a list of outcomes or 'scenarios' [where each line in the list] can be thought of as a triplet $<s_i, p_i, c_i>$ where s_i is a scenario identification or description; p_i is the probability of that scenario; and c_i is the consequence or evaluation measure of that scenario, i.e., the measure of damage. If this table contains all the scenarios we can think of, we can then say that it (the table) is the answer to the questions and therefore is the risk."‡ When this definition is placed in security terms, the scenario s_i represents a specific threat T with particular characteristics (e.g., number of attackers, weapons, tools, etc.) exploiting particular vulnerabilities to produce consequences c_i. The scenario likelihood p_i includes both the likelihood for a threat to attempt an attack (P_T), and the conditional likelihood that the attack by this threat will be successful ($P_{S|T}$).

Fundamentally, P_T can only be estimated in a Bayesian sense and is enormously uncertain because we cannot know the true intentions of all adversary groups. In addition, historical attacks indicate that adversary choices are not random. Instead, adversaries assemble resources that they believe are sufficient to ensure a high likelihood of a successful attack, or they select targets and plan attacks that they believe they can successfully achieve within their available resources and abilities. Hence, even a Bayesian estimate of P_T depends strongly on unquantifiable psychological factors like dissuasion, deterrence, and the adversary's level of commitment to their goals. Furthermore, P_T can change wildly over time as adversary groups are influenced by local and global political and social events of which we may not even be aware. Thus, the uncertainties in P_T are very large and can span several orders of magnitude for extreme but very rare attacks. Hence, investment decisions that are based on such risk estimates often cannot be supported with reasonable statistical confidence, even in the short term, to say nothing about decisions whose effects are expected to be felt for years or even decades to come. Ironically, these uncertainties are caused in large part by the use of the Kaplan-Garrick definition of risk.

Using conditional risk for security assessment can also lead to an important unintended side effect. By focusing on the adversary's successes and failures during the hypothesized attack, the analyst can be led to focus only on security risk mitigation options that make the observed adversary successes less likely. In so doing, the analyst may not recognize risk mitigation opportunities outside of the actual attack execution. For example, it may be possible to deny the adversary certainty of information that is critical to attack planning, or to minimize the consequences of the attack through resiliency and redundancy. A holistic perspective is required to ensure that the most cost-effective security mitigation options are discovered and pursued.

2.2. Extending the Definition for Security Risk

To overcome these obstacles, we propose a modified definition of risk where, instead of considering the highly uncertain likelihood or probability of an attack, one considers its difficulty for an adversary to successfully accomplish against the target(s) under consideration. Thus, a security risk analysis consists of answers to the following three revised questions: *(1) What can happen? (2) How difficult is it for an adversary to make this event happen? and (3) If it does happen, what are the consequences?* The triplet for security risk then becomes $<s_i, d_i, c_i>$ where d_i is the degree of difficulty

‡ Prior to Kaplan and Garrick, the most common definition of risk related to loss expectancy. Risk was defined as "probability *times* consequence." Kaplan and Garrick assert that risk is really "probability *and* consequence."

for an adversary to successfully accomplish attack scenario s_i at a specific target in order to cause consequence c_i.[§] Attacks are "higher risk" when they are attractive to an adversary because they are less difficult and/or lead to greater consequences than other candidate attacks. This definition explicitly acknowledges the observed adversary attack planning behaviors described above and addresses the problems associated with using probabilities to describe the intentional actions of both known and unknown intelligent actors. Risk evaluations using this definition do not require revision as adversary motivations change because this risk definition characterizes scenarios and targets rather than estimating the adversary's probability of attack.

This work uses the proposed definition by focusing on estimating the minimum threat capabilities and degree of difficulty required for an adversary to accomplish a specific attack scenario that exploits a target's vulnerabilities and induces specific consequences with a reasonably high likelihood of adversary success $P_{S|T}$. Adversary attack preparation activities are viewed as a project planning exercise, wherein a planner has success criteria (e.g., adequate or desired consequences and thresholds for likelihood of success), and chooses among alternative strategies that meet these criteria (e.g., achievable resources and plausible attack scenarios), while considering the degree of difficulty that will be encountered in order to achieve a successful outcome. Investments reduce security risk as they either (a) increase the difficulty for an adversary to successfully execute the most advantageous attack scenario, or (b) reduce the severity of the scenario's expected consequences. The latter can be measured through existing consequence metrics, but measuring the former requires development of a reasonable and robust metric to characterize the adversary's degree of difficulty in achieving a "successful" attack. Thus, the proposed definition and metric build upon the well-known P_E-based assessment and design methods, but do not exhibit the strong nonlinear behavior that has been observed for $P_{E|DBT}$. Such a metric and specific criteria for its scoring have been developed for the RIMES methodology in order to to compare and aggregate the relative degree of difficulty for disparate adversaries to successfully prepare for (e.g., acquire the requisite resources) and execute an attack (employ those resources in specific ways against specific targets). The metric is described in Section 3.

Using the metric as a measure of scenario difficulty, an analyst can compare security risks by comparing attack scenarios' levels of difficulty and consequences. The insights from such comparisons can provide important and useful security risk management insights for a broad range of applications. The objective of a security decision maker might be thought of as follows: to make the easiest attack path as difficult as possible within the constraints imposed by cost, operational and programmatic considerations. Consider a decision maker who is responsible for several sites where each attack leads to similar consequences. Figure 1a shows how results from this method can be applied to security decision making. Each light-colored bar represents the difficulty of the *easiest attack scenario* at a notional site in its original (2007) configuration. Note how it was much easier for an adversary to achieve a successful attack at Site D than at any other site. Note also how security at Site B was already significantly better than the original (2008) goal level. The decision maker focused on improving security at Site D, and in 2010, security is much more balanced across the enterprise as the difficulty of the easiest attack is now roughly comparable across all sites (the top of the dark bar in the graph). The decision maker can justify to the funding source *why* particular security investments were made and describe the specific benefits that the investments produced. Further, if policy changes cause the security goal to change, the decision maker can explain in simple terms to the funding source why additional security investments are necessary. Prioritizing investments is straightforward for this application, and the method is compatible with computerized optimization programs.

[§] This definition of risk, and specifically d_i, is a characteristic of scenario s_i for the specific *target*. The reader should not assume that d_i characterizes any specific adversary group or DBT. Rather, d_i incorporates the threshold threat characteristics needed for an adversary to have a high likelihood of success (i.e., a low value of $P_{E|TT}$) when attempting to execute scenario s_i at the specific target. It also incorporates the characteristics and complexities of the scenario that might make the scenario difficult for an adversary to accomplish successfully even if they had the requisite threshold threat characteristics.

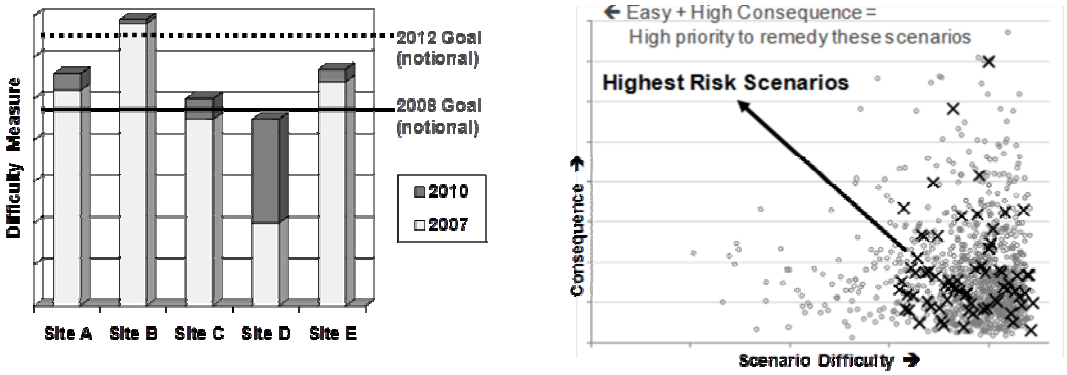

FIG. 1a (left) and 1b (right). (a) Comparing the relative difficulty of the easiest attack scenarios at five notional facilities where each attack leads to similar consequences. (b) Relative difficulty and consequences of attack scenarios at a notional facility (X symbols) compared with scenarios at other facilities within the enterprise (circles).

The situation where a variety of consequences are possible at a facility (or within an enterprise) is shown in Figure 1b. Here each identified attack path or scenario is represented as a circle on the scatter plot, with coordinates that represent the scenario's difficulty d_i and consequences c_i. Scenarios that produce higher consequences *and* are easier to accomplish are more attractive to an adversary because they represent a more efficient use of resources. Thus, they pose a greater risk and should be a higher priority for remediation. A scenario's risk can be reduced by reducing its consequence potential (moving the circle down), increasing its difficulty (moving it to the right), or a combination of these actions. Note that if one reduces the risk of a scenario s_j that is near the center of the pack of circles without also addressing scenarios that are more attractive (those that produce greater consequences *and* are easier to accomplish, *i.e.*, scenarios whose circles are above and to the left of s_j), the overall security risk may be unaffected by the investment because the most attractive scenarios remain available for adversary exploitation. Thus, the security investments should generally address those scenarios that are non-dominated (i.e., that represent the easiest way to produce consequences greater than or equal to c_j).

From the perspective of the security decision maker for an enterprise, the X symbols in Figure 1b represent the attack scenarios available at one facility, and the facility's manager wishes to mitigate the scenarios that are most attractive. The enterprise manager might use this graph, with circles representing attack scenarios available at other facilities in the enterprise, to inform the facility manager that only minimal security improvements will be supported because the enterprise has greater security risks that must be addressed first. However, if it is known from other sources that the facility is specifically targeted by credible threats, the enterprise manager may decide to support security upgrades at the facility anyway, believing that the easiest attack is not yet difficult enough.

3.0 Risk-Informed Management of Enterprise Security (RIMES)

3.1. General Characteristics of the RIMES Method

The RIMES approach described above starts by identifying a scenario that would offer an adversary a reasonable expectation of success[**] against the target(s) under consideration, i.e., a scenario for which the conditional likelihood that the attack by this threat will be successful $P_{S|T}$ exceeds an established threshold. Such scenarios can be developed by any number of means that are commonly used by the security analysis and vulnerability assessment community. Specific to each scenario, either explicitly or implicitly, are the resources (personnel, materiel, and knowledge) an adversary would need and the manner in which they would be employed, in order for the adversary to have a reasonable likelihood of success $P_{S|T}$ when executing the scenario.

[**] For most attack scenarios, "success" means inducing a specific consequence of the adversary's choosing from the target.

Considerations of the difficulty for an adversary to mount this scenario are partitioned into the two essential phases of adversary efforts for any attack scenario - Preparation and Execution. Since adversary success in the scenario requires successful completion of both phases, they are viewed with comparable significance. The primary factors that are generally key to adversary success in each phase of attack have been identified through discussions with subject matter experts, review of various ranking schemes for adversaries or threats or scenarios, and analysis of a diverse set of specific scenarios. Since we require a metric that characterizes the relative difficulty of successfully (inducing and) exploiting target vulnerabilities, we express scenario success factors in terms of their manifestation at the interface between target and threat. For example, while level of funding can be important to adversary success, this is manifested at the target in other factors, such as quality and size of the toolkit used in the scenario. We have developed these factors so that they can be considered as roughly independent dimensions of generally equivalent importance.

In addition to reflecting key factors for scenario success, the required metric must also reflect the relative level of difficulty for adversaries to be successful in the scenario against the target(s) under consideration. To do this, five discrete levels of difficulty have been defined for each success factor dimension. Guidelines have been developed for analysts to consistently assign the appropriate levels to each success factor dimension in order to reflect the relative difficulty that an adversary would encounter to successfully achieve or acquire the characteristics required in that dimension for the scenario to succeed. It is important to note that this process does not assign adversaries to a particular level, nor imply that all dimensions of a scenario are at the same level. Rather, the process dissects a successful scenario into the minimum levels of difficulty associated with each of the key factors that generally underlie adversary success. Since the scenario is specific to the target(s) under consideration, this process characterizes targets in terms of the levels of adversary difficulty to recognize, induce, and exploit vulnerabilities that enable scenario success.

The levels of difficulty for the dimensions have been calibrated so that a particular level for one dimension roughly correlates to an equivalent level of difficulty for any other dimension. In general, the levels of difficulty correlate with the size of the portion of the spectrum of generalized potential adversaries that could reasonably expect to achieve or acquire the associated level characteristics. Level 1 characteristics are easily accessible or achievable by the general population, while Level 5 characteristics would typically be accessible or achievable only by elite forces or state supported operations. Different levels of difficulty are distinguished by different levels of costs, quality of leadership, law enforcement or intelligence signatures, time to achieve, availability, ingenuity, and/or sophistication.

3.2 Dimensions of Success for Attack Preparation and Attack Execution

As a basis for the difficulty of attack metric, Table I presents the dimensions of success for preparation and execution of adversary attacks. The dominant challenges for adversaries in the Preparation phase of efforts are in developing, acquiring, and preparing the resources – personnel, materiel, and knowledge - required for the scenario without being detected or interdicted by authorities. The dominant resource attributes that are keys to scenario success, and the primary considerations that differentiate levels of difficulty for the adversary to succeed, are described in Column 1. In the Execution phase, the manner in which adversaries employ their resources can also be critically important to their ability to succeed. The dominant success factor dimensions for attack execution, and the primary considerations that differentiate levels of difficulty for the adversary to succeed, are described in Column 2.

3.3 Calculating the Metric

Generalized guidelines (not presented here) have been developed for assigning one of five levels of difficulty to each of the attack Preparation and Execution dimensions for any particular scenario and target(s). A scenario for which an adversary is considered to have a reasonable expectation of success against the target(s) under consideration is evaluated according to these guidelines. A numerical value is associated with each of the five levels of difficulty (currently, these are integer values 0 to 4). A dimension's values could also be weighted to reflect that dimension's relative general significance to

Table I. Dimensions of Difficulty for Attack Preparation and Execution

Attack Preparation	Attack Execution
Active Outsiders: # of Fully Engaged Participants: the difficulty an adversary faces to successfully muster and prepare team(s) without alerting authorities, which increases with the number of participants.	*Ingenuity / Inventiveness:* the degree to which an adversary must be creative or ingenious in order to discover and/or induce, and exploit the vulnerabilities required for a successful attack. Low levels are associated with simple, straightforward attacks that can easily conceived by most adversaries, while high levels are associated with attacks that reflect unique, imaginative approaches that are more likely to surprise and befuddle even very well prepared defenses.
Active Outsiders: Training & Expertise of Fully Engaged Participants: the depth and diversity of expertise required of participants, and by the rehearsal required for tasks.	*Situational Understanding & Exploitation:* the level of acuity required by the adversary to recognize the occurrence of exploitable conditions and the flexibility required to leverage those opportunities. Levels of difficulty are differentiated by the transience, unpredictability and observability of vulnerabilities upon which success of the scenario depends.
Support Structure: Size, Complexity, and Commitment: the contributions required of a support base during attack preparation, e.g., intelligence, safe haven, training or staging facilities, finances, scientific or technological R&D, and manufacturing. Difficulty varies with the extent, diversity, and quality of contributions required, and the degree of engagement and awareness of purpose for these contributions.	*Stealth & Covertness:* the degree to which scenario success depends on the concealment or masking of attack execution activities in order to delay the point of initial detection and recognition by authorities. Levels of difficulty are differentiated by the existence, duration and multiplicity of undetected adversary operations that must be conducted within the observational purview of authorities.
Tools: Availability reflects the difficulty associated with acquiring the tools required to successfully execute a scenario. Tools can include weapons, transportation, breaching equipment, electronics, fixtures, armor, disguise, etc. The levels of difficulty are distinguished by factors that influence their availability: rarity, law enforcement / intelligence signatures associated with their acquisition or staging, and level of controls in place to protect against illicit usage.	*Outsiders: Dedication / Persistence / Commitment:* the significance of consequences at risk for the attackers, their support base, and/or their cause, the persistence of their risk exposure, and the degree of adversary certainty of those consequences.
Insiders: # of Contributors: one of three dimensions (key factors for adversary success) associated with contributions from insiders. Difficulty varies with the necessity for insider contributions, the number of contributors required, and the necessity of collaboration among multiple insiders.	*Insiders: Degree of Engagement & Risk:* the equivalent significance, persistence, and certainty of risk exposure required of insiders contributing to the attack.
Insiders: Security Controls on Contributors: contributions required from insiders that have greater levels of access to security-sensitive features are generally more difficult for adversaries to confidently acquire due to the security controls in place to mitigate the potential for such occurrences.	*Operational Composition / Complexity:* the required number, modalities, and orchestration of separate avenues of adversary attack execution operations. Modalities refer to the nature of vulnerabilities and exploitation operations required for the scenario: e.g., physical, cyber, procedural, etc.

adversary success, although research to date has not indicated a rationale for other than uniform weighting. Since the dimensions are roughly independent and span the most significant challenges that are key to adversary success, the level of difficulty for each of the phases of the scenario is calculated as the length of the vector described by the values along each of the phase's dimensions (an L_2 norm), although other aggregation methods (e.g., power law-based methods) have also been used.

3.4 A Practical Method for Security Risk Management

The preceeding sections describe how RIMES is used to assess the difficulty of attack scenarios. It is important to place this description within the context of an overall risk management process. Since risk can be thought of as the potential for loss or consequence, a natural place for risk management to begin is to identify the consequences, or negative outcomes, that might come about because of the system or facility in question. Consequences can involve loss of money or loss of functionality, or they can be broader, including human health and safety, environmental effects, and even sociopolitical effects. One scenario can cause effects in more than one consequence category, as exemplified by a theft of nuclear material, which might cause effects in all of these categories if it were to be used in an improvised nuclear device. Where possible, metrics should be developed for each consequence, although a monotonic series of qualitative state descriptions can also be used.

Risk management then considers the ways by which each type of consequence can be achieved from the system or facility. Safety risk relates to the accidental and environmentally-induced ways to produce consequences, while security risk relates to the deliberate malevolent scenarios by which consequences can be induced. Several methods exist for developing attack scenarios, and these are described in the next section. It is important that the reasonably expected consequences be estimated for each malevolent scenario, in addition to its RIMES difficulty, so the risk manager can place the attack scenario in proper context. It is also important that the attack scenaros (a) cover the breadth of attack types and consequences that would be available to an adversary, and (b) represent the least difficult attack opportunities for each consequence type *and consequence level*, as it may be much easier to perform an attack that produces small consequences (say, injures 10 people) than large consequences (say, injures 1000 people) even at the same system or facility.

The attack scenarios, with their difficulty and consequences, can be represented on a scatter plot for security risk screening and prioritization. An overall risk management guideline is to make the easiest attacks for each consequence level difficult enough to deter the adversaries of concern. Thus, screening can occur as a decision maker accepts the risk of particular scenarios because they are either so difficult that they would not be attractive to the adversaries of concern, or because the expected consequences are low enough that they can be tolerated (e.g., covered by insurance). For attack scenarios that cannot be screened, a risk manager focuses on scenarios that represent the most attractive adversary opportunities: generally, those scenarios that are on or near the higher-risk frontier of the scatter plot. Since different consequence types may attractive to different types of adversaries, and these may be represented on different scatter plots, the risk manager looks for scenarios across the different scatter plots to ensure that scenarios that might be attractive to *any* adversary type are considered for mitigation.

As attack scenarios are identified for possible mitigation, the risk manager looks for opportunities to manipulate the physical, cyber and human aspects of the system or facility in order either to make the more-attractive higher-risk scenarios more difficult for an adversary to accomplish or to cause them to result in reduced consequences. Persons with physical security expertise often gravitate to the former, while there is often greater leverage in the latter because reduced consequences may reduce the risk of a larger population of scenarios – even those that yet remain undiscovered! As potential mitigation options are evaluated, the analyst must ensure that the "next easiest" attack is considered in the analysis. After all, it is of little use to make an easier attack scenario incredibly difficult if another attack that is almost as easy – and attractive – remains unaddressed. A key tool for the risk manager to use in choosing cost-effective risk mitigation is to observe the shift in the scenario difficulty-consequence population map on the scatter plot as "what if?" games are played with combinations of mitigation options.

Once the risk manager decides to implement a particular suite of mitigation options, it is important that those options be designed and built using good systems engineering principles. Many examples could be cited in which a poorly-designed "risk mitigation" activity actually increased security risks instead of reducing them. Note also that the proces of risk assessment, mitigation evaluation, and mitigation design/installation is an ongoing and iterative process throughtout the lifecycle of the system or facility because *everything* can change over time: the physical characteristics of facility or system itself, the environment in which it operates, the operational and security procedures, and the characteristics and identity of potential adversaries. Thus, security risk management, like safety risk management, is truly a job that is never complete.

3.5 Practical Approaches for Attack Scenario Development

The main basis of a RIMES assessment is to understand the difficulty an adversary would encounter to plan and execute any of the available attack scenarios. Thus, developing appropriate attack scenarios is of primary importance to the method. For persons who do not have attack scenario planning experince, this process can be unfamiliar and daunting. For this reason, we provide several suggestions to help analysts begin learning this discipline.

A safety analysis integral to the design and operation of most systems or facilities. Safety analyses are often developed in part to prevent the occurrence of nightmare consequences. If an accident scenario can cause these important consequences, there may be ways for an adversary to deliberately induce a similar scenario. These can be investigated as attack scenarios. In addition, if the safety analysis includes a probabilistic risk assessment (PRA), the PRA results can be used as starting material for attack scenarios, and can even be mathematically transformed so that they directly produce the locations an adversary would need to visit to cause consequences, resulting in target sets upon which scenarios can be developed.[i]

Another source for attack scenarios that should never be ignored is any existing security analysis for the system or facility, or for any other similar facility. Security analysts will likely already have candidate attack scenarios in mind if not already documented. Likewise, regulators may have already specified particular targets or operations for which security analyses are required, and attack scenarios related to those targets may also be available from security analysts. The same may be true as owner groups, user groups, or professional societies promulgate security "best practices" lists.

The above methods focus on a combination of compliance activities and deductive analysis methods, in which the analyst begins by examining consequences and seeks to deduce attack scenarios that might enable an adversary to produce those consequences. Another useful approach uses inductive reasoning wherein it is *assumed* that a particular "failure" has occurred or been induced, or that a particular "vulnerability" has been exploited. Using inductive logic, the analyst examines whether this situation (a) can be caused by an adversary, (b) could be used by an adversary as *part* of an attack scenario that would lead to consequences, and (c) might be in some way attractive to a particular type of adversary because of its ease or the opportunities it provides. Failure Modes and Effects Analysis (FMEA) provides a systematic method by which the initial list of "failures" can be identified. Security best practice lists can also be a useful starting point. Remember that particular entries are found on these lists precisely because in some previous instance either their presence has narrowly prevented a security event or their absence has enabled one.

In reality, many adversaries use a combination of inductive and deductive methods to plan their attacks. An adversary knows of a potentially useful condition, and analyzes deductively to see if it can be deliberately caused and inductively to see if it can be made to lead to consequences. A team that is specifically trained to mimic adversary behavior, often called a "Red Team," can recognize these opportunities and develop security scenarios from them. It is important to note that the designers of a system or facility often perform poorly as a Red Team – even if they have been properly trained. In practice, they have focused so hard on the difficult task of getting the sysem or facility to perform its *intended* functions that they can no longer step back enough to see how the

system might be *misused* for malicious purposes. For this reason, it is important for system designers to enlist independent analysts to review the system or facility and develop attack scenarios.

4.0 Application of RIMES for Nuclear Facilities

RIMES has been applied to evaluate the theft and sabotage risks for two types of nuclear fuel cycle facilities – used nuclear fuel (UNF) storage and small modular reactors (SMRs).

4.1 RIMES for Used Nuclear Fuel Storage Security

For UNF, increased emphasis is being placed on extended storage, especially dry storage, potentially for many decades. As part of this emphasis, technical analyses and guidance documents are needed to assure that the security risks associated with extended storage are understood and minimized. Any assessment of security over a very long timeframe is a challenge. The security assessment needs to consider protection provided by a storage container (cask) as well as the facility protection measures and to address identified security issues over the timeframe of extended storage. RIMES is being applied to provide a framework within which to evaluate security risks that may change and evolve over the timeframe of extended storage. Attack scenarios have been developed for sabotage and theft and the difficulty of these scenarios evaluated. In general, the relative difficulty of attack for sabotage was moderate to high and for theft was very high. Evaluation of consequences, in terms of potential radiological releases, will be incorporated in future analyses. Additional scenario development will also consider changes in future conditions and alternative storage facility design concepts.

4.2 RIMES for Small Modular Reactor Security

A generic integral pressurized water reactor (iPWR) design [19] was developed to provide a basis for the RIMES analysis. This design pulled from many of the common features of iPWR designs currently available today without representing any one specific design. The work for SMRs identified a preliminary list of safety and support systems necessary for safe shutdown and then applied RIMES for example theft and sabotage scenarios. A total of 14 scenarios were evaluated to cover a range of attack types and consequences. Consequences were loosely binned into economic damage only, economic damage with release, core melt with little/no release, and core melt with release. Both outsider and insider attack scenarios were considered. Subject matter experts in reactor design, reactor safety, physical security, and response forces participated in the assessment. A long-term goal is to use these results to better inform physical security system design for plant designers. In many cases, rather simple design changes can either significantly increase the difficulty or reduce the consequence of a particular scenario.

In general, core melt (high consequence) scenarios were found to result in high difficulty levels. Multiple systems would need to be disabled, some of which are redundant. One scenario, which was found to be at a more moderate difficulty rating, could easily be remedied with a simple design change to the reactor building. Lower consequence property damage scenarios can be achieved with relative low difficulty—these scenarios do not lead to core melt or any release, but could cost the operator a significant amount of money in lost operational time. Scenarios with lower levels of difficulty can be addressed through design changes or improvements to the physical protection system that increase difficulty or mitigate consequences. The RIMES methodology made it much easier to examine cost-effective design changes. However, it should be noted that all of these scenarios will change when applied to specific vendor designs.

5.0 Conclusions

The RIMES methodology has been developed to address some of the key issues associated with applying traditional risk analysis to security. RIMES is an objective risk-informed method that is based on characterizing targets in terms of an adversary's degree of difficulty to prepare for and execute successful attacks. A focus on the level of difficulty of a particular attack as opposed to the probability of attack will enable decision makers to balance competing security interests (e.g., multiple facilities) and provide objective and unbiased justification for investment decisions, resulting in more robust and cost-effective security systems. This shift allows for designers to manage risk

better by balancing increased security against those threats that require lower difficulty for an adversary to produce higher consequences.

This work has provided a preliminary examination of attack scenarios for two types of nuclear fuel cycle facilities – UNF storage and SMRs. For an individual facility, the RIMES methodology helps designers to focus on the attack scenarios of concern and the threats that can accomplish those scenarios, but RIMES can also examine how those scenarios and threats compare to those that could be executed in other parts of the nuclear fuel cycle. This work has also investigated and demonstrated how lower difficulty attacks for consequences of concern can be addressed by facility or security designs changes that can eliminate or mitigate the consequences or increase the difficulty of attack. The longer-term vision is to apply RIMES across the fuel cycle to examine most likely attack scenarios across various facility types to target investments to address security risks where they are needed most.

REFERENCES

[1] GARCIA, M.L., The Design and Evaluation of Physical Protection Systems, Second Edition, Butterworth-Heinemann (Elsevier), Burlington MA (2008).

[5] U.S. NUCLEAR REGULATORY COMMISSION, WASH-1400 Reactor Safety Study: An Assessment of Accidental Risks in U.S. Commercial Nuclear Power Plants, NUREG-75/014, U.S. Government Printing Office, Washington DC (1975).

[6] MURPHEY, W.M., SHERR, T. S., BENNETT, C.A., Societal Risk Approach to Safeguards Design and Evaluation, ERDA-7, Energy Research and Development Administration, Washington DC (1975).

[7] RASMUSSEN, N., Probabilistic Risk Analysis – Its Possible Use in Safeguards Problems, Proc. 17th Ann. Mtg. of Inst. of Nucl. Matls. Mgmt., Deerfield IL (1976).

[8] UDELL, C.J., CARLSON, R.L., Risk Evaluation System for Facility Safeguards and Security Planning, Proc. 30th Ann. Mtg. of Inst. of Nucl. Matls. Mgmt., Deerfield IL (1989).

[9] UDELL, C.J., et al., Short Form Risk Evaluation Method, Proc. 34th Ann. Mtg. of Inst. of Nucl. Matls. Mgmt., Deerfield IL (1993).

[10] BIRINGER, B.E., et al., Security Risk Assessment and Management – A Professional Practice Guide for Protecting Buildings and Infrastructures, John Wiley & Sons, Inc., Hoboken NJ (2007).

[11] RICHARDSON, J.M., Comprehensive Safeguards Evaluation Methods and Societal Risk Analysis, SAND82-0366, Sandia National Laboratories, Albuquerque NM (1982).

[12] COX, JR., L.A., Some Limitiations of Risk = Threat × Vulnerability × Consequence for Risk Analysis of Terrorist Attacks, Risk Analysis, **28** (2008) No.6.

[13] SNELL, M.K., GARDNER, B.H., Determining System Effectiveness Against Outsiders using ASSESS, Proc. 32nd Ann. Mtg. of Inst. of Nucl. Matls. Mgmt., Deerfield IL (1991).

[14] KAPLAN, S., GARRICK, B.J., On the Quantitative Definition of Risk, Risk Analysis, **1**:1, (1981).

[16] WYSS, G.D., et al., Risk-Based Cost-Benefit Analysis for Security Assessment Problems, Proc. 51st Ann. Mtg. of Inst. of Nucl. Matls. Mgmt., Deerfield IL (2010).

[17] WYSS, G.D., et al., Risk-Based Cost-Benefit Analysis for Security Assessment Problems, Proc. 44th Ann. Intl. Carnahan Conf. on Security Tech., San Jose CA (2010).

[18] WYSS, G.D., et al., Risk-Based Cost-Benefit Analysis for Security Assessment Problems, Proc. 50th Ann. Mtg. of Inst. of Nucl. Matls. Mgmt., Deerfield IL (2009).

[19] LEWIS, T., et al., Generic Small Modular Reactor Plant Design, SAND2013-10406, Sandia National Laboratories, Albuquerque, NM (2012).

[i] Reference to Vital Area Analysis methods documentation.