SAND2015-3632C

The Treatment of Blended Attacks in Nuclear Security Effectiveness Assessments

M. Snell[1], J. Rivers[2]

[1]ISandia National Laboratories, Albuquerque, United States of America

[2]United States Nuclear Regulatory Commission, Rockville, United States of America

E-mail contact of main author: mksnell@sandia.gov

Abstract. Cyber and information security have become more important in nuclear security in recent years due to the increasing reliance on networked computer systems as part of physical protection systems (PPS's) and nuclear material accounting systems. One of the concerns is protection against so-called "blended cyber-physical attacks" where cyber and information attacks are used to support physical attacks by outsider or insider threats. This paper describes a methodology for evaluating PPS effectiveness against blended attacks. The approach is based on a general methodology for evaluating PPS effectiveness against physical attacks either by an outsider or insider threat. Critical PPS components and subsystems would be identified as part of this methodology. Cyber security evaluation techniques would then be applied to the cyber components of the PPS to determine which, if any, of these critical PPS components and subsystems could be compromised user cyber-attacks. The complete blended attack would then be identified within a modeling framework used for evaluating violent insider attacks where the cyber threat is treated as just another type of insider who could collude with other (real) insiders or outsiders. It is recognized that combined cyber/physical systems are too complex to be evaluated thoroughly by such an effectiveness methodology. For this reason, we suggest that countries apply this approach within a framework of basic cyber security regulations, such as the U.S. Nuclear Regulatory Commission's Critical Digital Asset (CDA) regulatory approach, focused on protecting, in a graded fashion, those cyber systems that can cause critical PPS elements to fail.

Key Words: Blended attacks, effective evaluation techniques.

1. Introduction

A security effectiveness evaluation can be viewed as consisting of a planning phase, a conduct phase, and a closure phase where recommendations and reports are made. The conduct phase includes defining where the security requirements for the system under study, characterizing that system, and then applying an effectiveness evaluation methodology to determine how well the requirements are met by the system and what the strengths and weaknesses of that system are.

This paper focuses on describing an effectiveness evaluation methodology to address so-called "blended cyber-physical attacks" where cyber and information attacks are used to support physical attacks against physical protection systems (PPS's) by outsider or insider threats. The methodology was designed with two considerations in mind: first, that it be consistent with existing computer security and physical protection evaluation methods taught by the International Atomic Energy Agency (IAEA) and that it would require a relatively low

level of additional training; second, that it would be produce credible blended scenarios, be reasonably systematic and provide insight into both the cyber and physical protection systems. Note that there are a number of IAEA recommendations and guidance documents related to protection against blended attacks: see [1], [2], [3], and [4].

The training issue is important because there is already a significant amount of training required just to address cyber security and physical protection evaluations separately. This limitation would seem to rule out covering completely new approaches such as attack graphs.

There are a number of significant technical challenges to be addressed before a systematic approach to evaluating blended attacks can be introduced into the international nuclear security community. One issue is that there is a range of methodologies for evaluating PPS's across the international community but no standard methodology. Another issue is that while IAEA cyber security technical documents, guides, and training courses do consider defeat of physical protection equipment, the evaluation is typically limited to components and subsystems of the PPS and not the entire PPS. An important related issue is how to systematically determine what the objective(s) of the cyber-attack should be in supporting the physical portion of the attack. This is especially true given the complexity of the combined cyber security-PPS. Finally, there is a need for a more systematic methodology to integrate the cyber-attack scenarios into physical attack scenarios (heuristic methods, based on red-teaming do exist but depend on the ingenuity of the participants as the methods are not systematic).

Note that this paper does not address blended attacks on instrumentation and control systems although some of the same concepts could be adapted to address such attacks.

The rest of the paper is divided into three sections. Section 2 provides an overview of an effectiveness evaluation methodology for evaluating PPS's against insider and outsider threats that is consistent with current IAEA courses. Section 3 discusses approaches for identifying critical PPS components and subsystems based on the results of the effectiveness evaluation. Section 4 discusses how cyber security evaluation approaches can be applied to identify potential attacks on these critical PPS components and subsystems. Section 5 then covers how to combine the results of the cyber security evaluation back with the PPS evaluation. Finally, section 6 discusses the U.S. Nuclear Regulatory Commission's Critical Digital Asset (CDA) regulatory approach that can serve as a regulatory framework within which to apply the complete methodology discussed in Sections 2-5. Finally, section 7 provides some conclusions.

2. A Proposed Evaluation Methodology for Evaluating Physical Protection Systems

IAEA recommendations concerning PPS effectiveness evaluations are contained in [1]. Traditionally, outsider attacks against PPS's have been modeled and evaluated using a combination of what is called path analysis and scenario analysis. This section will briefly describe these evaluation methods and will discuss how those analysis approaches are applied to non-violent insider attacks.

The set of potential physical paths that an outsider threat can use can be represented in network form, for example, as an adversary sequence diagram (ASD). Evaluations of non-violent insiders do not use ASDs but base the evaluation on what are called adversary action sequences that have similar properties to outsider paths. A set of action sequences can be represented as an adversary action sequence diagram (AASD); see Figure 1.

Currently, for outsider threats, response timeliness for a path is based on PPS response times or PRT's. An adversary timelines can be compared to the response timeline, to determine Probability of Interruption, P_I, for the outsider threat. Evaluations for non-violent insiders determine a cumulative Probability of Detection, P_D along an adversary action sequence.

Scenario analysis determines whether the PPS effectiveness, P_E, is adequate across a range of detailed adversary attack scenarios that might be credibly planned and conducted by adversaries operating within the scope of the Design Basis Threat (DBT). Such scenarios may be created manually using teams of experts familiar with a PPS or can be based on an outsider path or insider action sequence. In the latter case, the outsider path suggests a sub-plan that serves as the main or direct part of the attack (direct in the sense of going to the target). Such plans might be based on the minimum delay, minimum probability of detection, or minimum P_I for the paths. Details can be added to these path descriptions to fill out the scenario. For example, instead of the step "Penetrate Fence" found in the path analysis, the scenario description might consist of: "Four adversaries bridge fence using ladder carried in from vehicle parked outside at night during a storm. Last adversary monitors radio traffic." For non-violent insiders a similar process elaborates on the insider action sequence to create a scenario description.

3. Identify Critical PPS Elements

This section discusses approaches for identifying critical PPS components and subsystems based on the results of the effectiveness evaluation that was described in section 2 and on other sources of information. We will describe PPS components and subsystems as critical elements if their compromise will cause the PPS effectiveness to drop below a level of performance that is acceptable to the competent authority.

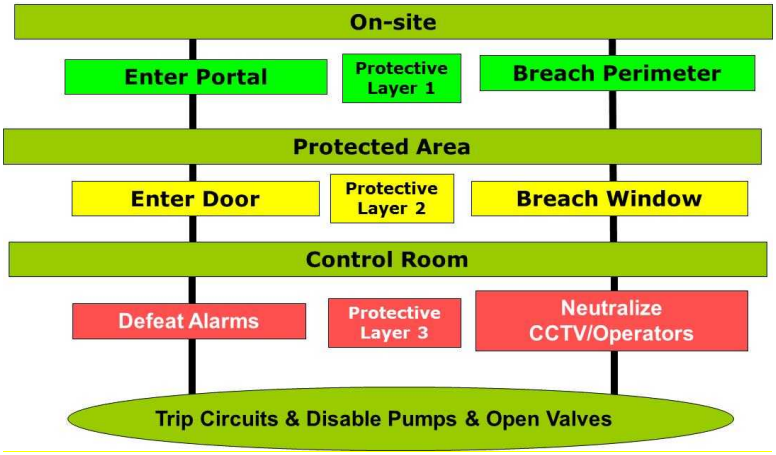

FIG. 1. Example of an Adversary Action Sequence Diagram.

There are several methods to identify critical elements. One method is by inspection: for example, if there is only one means of communication with offsite protective forces then that communication system would be a critical element. Another method is based on applying importance measures[1] to path/adversary action sequence analysis to identify those physical protection components and subsystems that are most critical.

Critical elements can also be identified using conventional collusion analysis by assuming an active insider of some kind is helping the outsider group (or is colluding with another insider). For example, an attack on a field distribution box by a maintenance technician may results in a similar reduction in security as a cyber-attack on box. As another example, an insider who can successfully defeat access authorization procedures to help an outsider gain access to a protected area would provide a similar advantage as a cyber-attack that accomplishes the same thing. The central concept followed here is that existing collusion evaluation approaches help identify the critical elements that should be considered.

Critical elements are also identified as part of the process for developing facility Performance Testing Program Plans (PTPP's); see Figure 2. Our evaluation methodology would take advantage of that source of information to reduce workload on the evaluation team. Other potential critical elements can be identified by reviewing contingency plans as well as non-security test and quality plans for components, software, and subsystems, whether these tests come from a vendor of a component or subsystem or by the facility itself.

4. Apply Cyber Security Evaluation Techniques to Critical Physical Protection System Components and Subsystems

Cyber security evaluation techniques, as found in existing IAEA technical documents and courses, would then be applied to the cyber components of the PPS by cyber security experts involved in the evaluation. A major focus of the evaluation process would be to determine which, if any, of the critical elements in the PPS can be defeated using cyber-attacks.

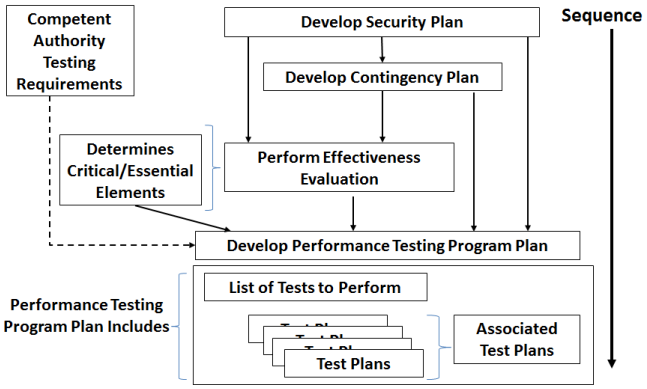

FIG. 2. Process for Developing Performance Testing Program Plans.

[1] For a discussion of importance measures, from the perspective of probabilistic safety analysis, see [5]. In the PPS context, we are most interested in finding those physical protection measures that require the least decrease in performance to cause P_I or P_D to drop below some threshold.

There is a need, then, to have methods for identifying and evaluating ways to compromise individual components and subsystems and then assessing how the PPS would be affected. There are a set of tools that can be used for this purpose: deductive methods such as logic diagrams (which include fault trees, physical protection logic trees, and attack trees) as well as inductive methods that start with some adversary action, such as defeat of a particular sensor, and then deduce the effects on the PPS that may occur. The cyber security experts would then identify elements that they could defeat that the PPS experts would evaluate to determine the effect on the PPS. Figure 3 provides a hypothetical attack tree for defeating authentication for an equally hypothetical prox card. Attack sequences through attack trees can have various measures applied, such as difficulty, financial cost or time involved in the accomplishing the defeat, in order to determine which attack sequences are most attractive from the adversary's perspective.

Note that fault trees and attack trees are already covered in courses on vital area identification and cyber security, respectively.

Creation and application of logic diagrams would have to be performed by some combination of experts from a PPS evaluation team and from a cyber-security evaluation team. Reference [6] discusses the composition and activities of PPS evaluation teams. This combined approach would, of course, require additional training.

5. Combine the Results of the Cyber security Evaluation with the PPS Evaluation

The complete blended attack would then be identified within a modeling framework used for evaluating violent insider attacks that is a relatively straightforward extension of the methodology discussed in section 2. The cyber-attack would be treated as just another type of insider who can be active and/or violent, has (or can gain) access, authority, and knowledge, can collude with other (real) insiders or outsiders, and can perform stealthy activities to degrade the PPS. If desired, scenarios could be mitigated through redesign of the PPS and/or cyber-systems, and/or developing appropriate compensatory measures as part of contingency plans.

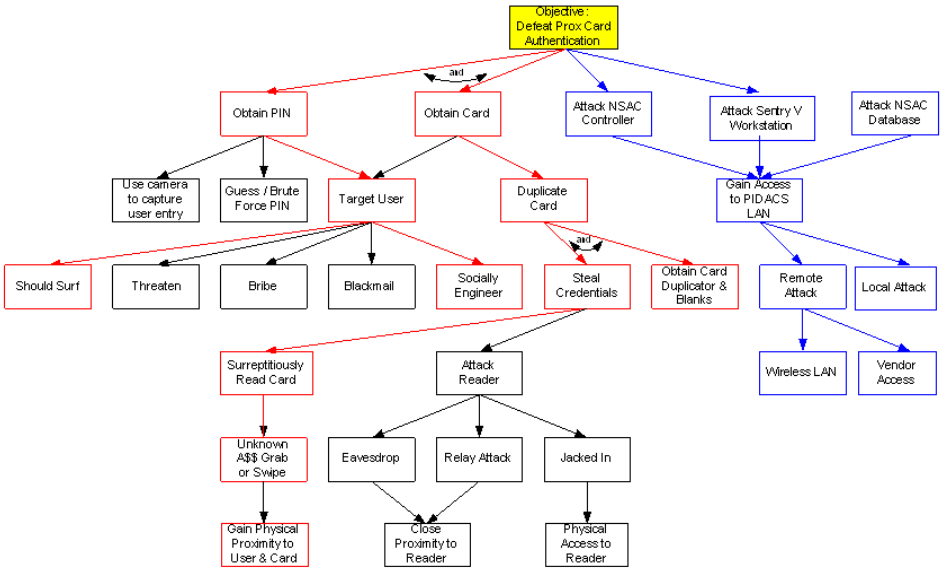

FIG. 3. Example of an Attack Tree

Overall system effectiveness would be characterized by two metrics:

1) A system effectiveness metric treated as a probability of system effectiveness, $P_{E(Total)}$, of the form

$$P_{E(Total)} = 1-(1-P_{DS})(1-P_{EO}) \quad (1)$$

where P_{DS} represents probability of detection of the cyber-physical attack during a prolonged stealthy preparation for an attack and P_{EO} measures effectiveness of the PPS during a more "overt" phase concerned with timely detection (in some sense) of the adversary attack over a more limited time frame. P_{EO} would be estimated assuming any physical and cyber preparatory attacks proposed for the PPS had successfully degraded the PPS. Note, though, that the adversary activities would incur the risk of being detected with probability P_{DS} resulting in a trade-off between performing preparatory attacks that would increase P_{DS} but potentially decrease P_{EO}.

2) Some standard measure, M_C, of the difficulty/cost to the adversary performing the cyber portion of the attack. This measure typically involves other factors besides systems effectiveness so it should be treated as a separate variable. Alternatively, M_C might be some qualitative risk measure associated with the adversary attack.

For the purposes of this step in the methodology, an adversary action sequence (AAS) would be defined more generally as a time-ordered sequence of n tasks that the adversary has to complete. This more general AAS can be thought of as a detailed plan of what an adversary team (outsiders) or individual insiders would need to accomplish to effect theft of nuclear or other radiological material, sabotage, or dispersal of radioactive material. Each task will have an associated probability of detection, P_D, and task time, T. Each task can have a location associated with it (as might be indicated on an ASD) or it might not: for example, the action might be "bribe the facility manager." Task times may be well defined, for example when they are based on the use of a specific adversary attack tool against a specific barrier but they may also be defined in an indefinite way, for example, "wait until the next material shipment." Probabilities of detection may be defined quantitatively or qualitatively[2]. Networks similar to ASDs and AASDs can be constructed to represent the range of AAS's that the adversary might use.

Timelines for both the adversary and one or more response organizations can be built based on the AAS; see Figure 4. Figure 4 shows response timelines for the security system, with an associated PPS response time (PRT), and for a system controlled by operations with an associated Operational System response time. Potentially, additional response organizations might participate and consequently have their own timelines represented.

The adversary timeline and (possibly multiple) response timeline(s) can be compared to determine something analogous to Probability of Interruption, P_I, for the outsider threats. In this more general approach, sensing opportunities on the adversary timeline would be considered timely if they were timely against at least one of the response timelines. Note that overall system effectiveness, P_{EO}, would include a term analogous with Probability of Neutralization, P_N, to go with this more general P_I. In practice, any approach to characterize P_{EO} would probably be based on a simulation of some kind rather than attempt to determine P_I and P_N individually.

[2] There are ways to combine both qualitative and quantitative metrics but we will not address that issue in this paper.

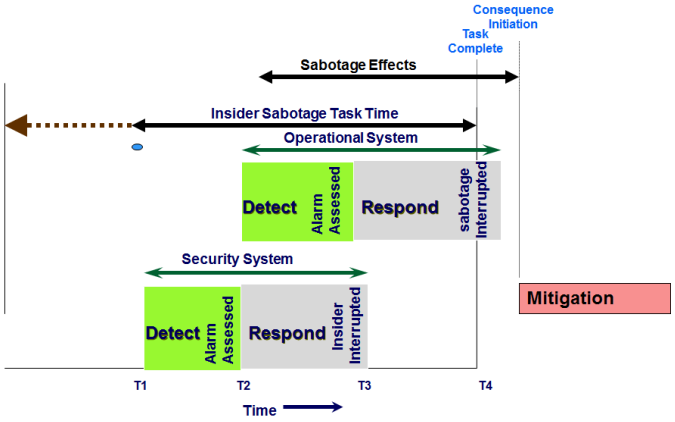

FIG. 4. Adversary, Security Response, and Operational Response Timelines.

It is recognized that these more general timelines may include ambiguous or even indefinite adversary or response task times. As one example, the PRT might be represented as an interval rather than a point value. The P_I and P_{EO} metrics can still be calculated but the results would consequently consist of intervals rather than point values.

Finally, scenarios would be ranked in terms of the pair of metrics ($P_{E(Total)}$, M_C).

6. The U.S. NRC's Regulatory Approach to Critical Digital Assets

The effectiveness evaluation methodology described in sections 2 through 5 has definite limitations and cannot possibly cover all cyber security issues related to blended attacks on PPS's. For this reason, we suggest that states apply this evaluation approach within a framework of basic cyber security regulations. The NRC's CDA regulatory approach will be discussed here; the approach has some merit as an example of such a framework because it has been applied to a large number of nuclear power plants and because much of the regulatory guidance documents are open source and can be readily adopted and/or modified.

Historically, shortly after 9/11, the NRC issued the first cyber security requirements within physical security and DBT regulations. Based on these requirements, the NRC has had programs in place to protect CDA's that interconnect plant systems performing safety, security, and emergency preparedness. In 2009, NRC issued a Power Reactor Cyber Security Rule, 10 CFR 73.54, and since that time a number of guidance documents have been developed. For example, the NRC published Regulatory Guide, RG 5.71[7] which provides guidance on an acceptable way to meet the requirements found in the Cyber Security Rule. The guidance includes "best practices" from such organizations as the International Society of Automation, the Institute of Electrical and Electronics Engineers, and the National Institute of Standards and Technology (NIST), and the Department of Homeland Security. The Nuclear Energy Institute also prepared guidance, endorsed by the NRC, on how to protect CDAs; see references [8] and [9]. The NRC is also considering the need for similar cyber security requirements for fuel cycle and spent fuel storage facilities, non-power reactors, decommissioned nuclear facilities, and materials licensees.

The following figure depicts the conceptual approach that the NRC has taken, based on NIST documents and concepts.

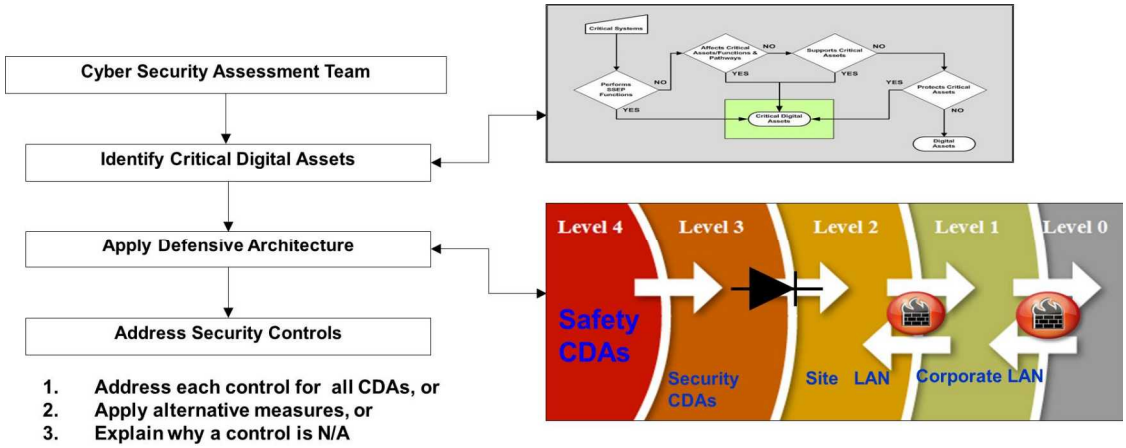

FIG. 5. Conceptual Approach Taken in NRC RG 5.71.

When evaluating the PPS, the Cyber Security Assessment Team would include experts on the PPS and familiar with the associated critical PPS components and subsystems identified earlier in the effectiveness methodology. It may turn out that certain cyber systems associated the PPS might be identified as CDA's; in this case, some defensive architecture[3] such as that shown in Figure 5 would be applied and certain security controls would be developed. Beyond some basic security measures, additional cyber security controls could be applied in a graded approach, combining the results of the evaluation approach covered in the earlier sections along with the graded assessment approach discussed in [9]. Using a graded approach is highly recommended: when NRC first started doing cyber inspections they discovered an unexpectedly high number of CDAs, with as many as 2500 per site. This brought to light the sheer magnitude of the cyber problem as assessment of each CDA under the original, non-graded, NRC approach involved assessment of each of 148 controls, for each CDA.

7. Conclusions

An effectiveness methodology has been presented to address blended cyber-physical attacks aimed at a PPS. While new training would be required, the approach would incorporate existing techniques used by cyber security evaluators as well as existing PPS techniques. The interface between the cyber-security evaluation and the PPS evaluation is explicitly addressed to make the approach more systematic and complete.

REFERENCES

[1] INTERNATIONAL ATOMIC ENERGY AGENCY, Nuclear Security Recommendations on Physical Protection of Nuclear Material and Nuclear Facilities (INFCIRC/225/Revision 5), IAEA Nuclear Security Series No. 13, IAEA, Vienna (2011).

[3] The NRC security level numbering scheme starts at 0 (least protection needed) and increases as more stringent security is needed up to level 4, which is reverse to the numbering scheme used in [3] where the most stringent protection is provided for computer systems assigned to protection level 1.

[2] INTERNATIONAL ATOMIC ENERGY AGENCY, Nuclear Security Recommendations on Radioactive Material and Associated Facilities, IAEA Nuclear Security Series No. 14, IAEA, Vienna (2011).

[3] INTERNATIONAL ATOMIC ENERGY AGENCY, Computer Security at Nuclear Facilities, IAEA Nuclear Security Series No. 17, IAEA, Vienna (2011).

[4] INTERNATIONAL ATOMIC ENERGY AGENCY, Security of Nuclear Information, IAEA Nuclear Security Series No. 23-G, IAEA, Vienna (2015).

[5] VAN DER BORST, M., SCHOONAKKER, H. "An overview of PSA importance measures", Reliability Engineering and System Safety 72 (2001) 241-245.

[6] GARCIA, M. L., Vulnerability Assessment of Physical Protection Systems, Elsevier Butterworth-Heinemann (2006).

[7] UNITED STATES NUCLEAR REGULATORY AGENCY, Cyber Security Programs for Nuclear Facilities, Regulatory Guide (RG) 5.71, (2010).

[8] NUCLEAR ENERGY INSTITUTE, Cyber Security Plan for Power Reactors, NEI-08-09 Revision 6, (2010).

[9] NUCLEAR ENERGY INSTITUTE, Cyber Security Assessments, NEI-13-10, (2014).

[10] PARK, J., SUH, Y., "A Development Framework for Software Security in Nuclear Safety Systems: Integrating Secure Development and Systems Security Activities", Nuclear Engineering and Technology **46** No. 1 (2014) 47-54.