Construction of a Cyber Attack Model for Nuclear Power Plants

Athi Varuttamaseni and Robert A. Bari
Brookhaven National Laboratory
Building 817, Upton, NY 11973
avarutta@bnl.gov; bari@bnl.gov

Robert Youngblood
Idaho National Laboratory
PO Box 1625, Idaho Falls, ID 83415
robert.youngblood@inl.gov

ABSTRACT

The consideration of how a compromised digital component can impact neighboring components is critical to understanding the progression of cyber attacks. The degree of influence that one component may have on another depends on a variety of factors, including the sharing of resources such as network bandwidth or processing power, the level of trust between components, and the inclusion of segmentation devices such as firewalls. The interactions among components via mechanisms that are unique to the digital world are not usually considered in traditional probabilistic risk assessment (PRA). This means potential sequences of events that may occur during an attack may be missed if one were to only look at conventional accident sequences.

This paper presents a method where, starting from the initial attack vector, the progression of a cyber attack can be modeled. The propagation of the attack is modeled by considering certain attributes of the digital components in the system. These attributes determine the potential vulnerability of a component to different classes of attack and the capability gained by the attackers once they are in control of the equipment. The use of attributes allows similar components (components with the same set of attributes) to be modeled in the same way, thereby reducing the computing resources required for analyzing large systems.

Key Words: cyber attack scenario, cybersecurity, critical digital asset, risk assessment

1 INTRODUCTION

In recent years, there has been an increase in the number of targeted attacks against computer systems in industrial facilities. As existing nuclear power plants (NPPs) incrementally upgrade their old analog systems to digital, and with new plants employing digital equipment in critical systems, the number of attempted cyber attacks on NPPs are expected to increase. Whereas traditionally, industrial control systems (ICS) generally use proprietary equipment with proprietary software, many of the new software and equipment are becoming standardized. This means that the knowledge and experience needed to attack ICS are becoming less specialized. This further reduces the barrier for an attacker to mount a successful attack against ICS, including those used in NPPs.

Recognizing the threat of cyber attacks on NPP, the Nuclear Regulatory Commission (NRC) has been active in updating the regulations and guidance to include a cybersecurity component. Regulatory Guide (RG) 5.71 [1], published in 2010, provides the industry with an acceptable method for meeting the NRC cybersecurity requirements. However, the focus of RG 5.71 is to ensure that critical digital assets (CDAs) which are important to the safe operation of NPPs are properly secured. It addresses to a lesser extent the protection against scenarios where the breach of safety barriers may not be the attacker's goals. For example, the attacker may want to force a plant trip or cause equipment damage to non-safety equipment and thus incur financial impact to the owner/operator without public health impacts. In a review of the draft RG 5.71, the Advisory Committee on Reactor Safeguards (ACRS) recommended that the NRC pursue a longer-term goal of investigating the use of probabilistic risk assessment (PRA) (especially in determining event sequence) in cybersecurity assessment as well as the interaction between safety and security considerations [2].

Probabilistic risk assessment (PRA), which is extensively used in the nuclear industry to consider the impact of equipment failure on plant safety, can be used to inform licensees about the relative importance of systems to plant safety. However, by itself, PRA is insufficient for analysis of cyber attacks. Cyber attacks are initiated by humans, and the progression of the attacks (in some cases) can be altered mid-course in response to defensive measures. Multiple pieces of equipment which may be independent (and thus provide a defense-in-depth against random failures) may be simultaneously targeted in an attack, rendering all of them ineffective. While PRA methodologies could, in principle, be adapted to describe such scenarios, traditional PRA for accidents assumes non-deliberate failures of components and systems.

Despite the weakness of traditional PRA in analyzing targeted cyber attackers, it is still a useful tool to be used as part of the cybersecurity analysis [3]. Consideration of common cause failure would now have to be expanded to account for the fact that redundant components which are seemingly independent may be individually targeted by an attacker if the goal of that attacker is the disabling of the relevant system. Furthermore, the sharing of resources such as network bandwidth, processing power, or programming workstation need to be considered during the analysis.

This paper presents an approach where the progression of different cyber attack scenarios can be predicted based on structural and functional dependencies of components in the system. Section 2 reviews related work in the area of scenario generation and attack progression modeling. Section 3 presents a new approach that is being developed for attack modeling that can be scaled to analyze NPP systems. The output from this modeling gives analysts insight into how compromised components can behave (which can be very different from the usual failed behavior considered in PRA). This knowledge, in turn, will allow considerations of the system response (e.g., timing of events and anticipated indicators of compromise). Finally, section 4 describes the application of this technique to a very simplified pressurizer pressure control system.

2 RELATED WORK

To gain insights about how cyber attacks can impact NPPs, it is useful to look at past events to see the potential impact of misbehaving digital equipment on system operations. Table 1 summarizes some of the reported cyber-related events (not all of them are deliberate attacks) at nuclear facilities worldwide [4].

Table I. Cyber-related events at nuclear facilities

Date	Facility	Description
January 2003	Davis-Besse NPP (US)	Slammer worm originating from an external network infected the licensee's corporate network (through a backdoor), then propagated to control system network. The Safety Parameter Display System (SPDS) and Plant Process Computer (PPC) were inaccessible for several hours. The plant was offline at the time of the incident.
August 2006	Browns Ferry NPP (US)	Malfunction of network equipment generated a large volume of network traffic which locked up the variable frequency drive controllers for the plant's water recirculation pumps. Per procedure, the operators initiated manual scram.
March 2008	Hatch NPP (US)	An engineer performed a software update on a computer on the enterprise network. However, this computer had two-way

		communication with another system on the control network. When the updated computer restarted, data on the control network system reset, causing the RPS to interpret the event as a drop in water level. This led to an automatic trip.
Identified in 2010	Natanz nuclear facility (Iran)	Targeted attack on the PLCs to disable centrifuges used for enrichment. The worm was believed to breach the air gap via removable media.
January 2014	Monju NPP (Japan)	A software update on a computer at the plant was believed to introduce a malware that sent data on the compromised machine to a command and control server in another country. Data on the compromised machine included emails and staff training reports. The malware also attempted access to a control room computer.
April 2016	Gundremmingen NPP (Germany)	Malware, including Conficker and W.32Ramnit, were discovered on several computers and removal media (e.g., USB flash drives) at the plant.

From the events shown in Table I, it is evident that for a cybersecurity assessment methodology to be successful, the method has to be able to account for the following dependencies:

1. Shared resources such as network bandwidth, memory, and CPU cycles,
2. Presence of vectors where code/data can be introduced to the system (e.g., removable media, local serial links),
3. Vulnerabilities introduced by the presence of communication pathways (e.g., network interface),
4. Trust relationship (e.g., firewall configuration, network segmentation).

There are many methods that have been introduced to address the unique characteristics of the nature of dependencies among digital systems [5-9]. A scenario graph, for instance, allows the sequence of attack to be depicted graphically. The attacker's initial foothold on the network (including the initial access level) can be represented as the leaf node of the graph. Multiple leaves correspond to the different ways in which the attacker can initiate the attack (i.e., the set of leaf nodes represents the set of feasible attack vectors). Each node in this graph represents the new capability (or new component that is compromised) gained by the attacker during the attack propagation. Probabilistic values may be assigned to each node to model the difficulty of compromising different components. However, in its simplest form (and to be conservative), non-zero success probability can often be assumed to be 1 (i.e., every vulnerability that is present is exploited). Scenario graphs have been used to represent attack scenarios of NPPs [7].

A related approach that is widely used in cybersecurity risk assessment is the attack graph [8,9]. Here, the transition between nodes can be expanded to include specific vulnerability. This means that for a transition to occur, a vulnerability has to be present (either postulated or discovered) and the condition for exploiting that vulnerability has to be satisfied by the attacker. To be more realistic, vulnerability databases can be used to identify the possible transitions between nodes. However, this approach has a drawback that undiscovered vulnerabilities would be modeled as invalid transitions (e.g., transition probability of 0). If a vulnerability were to be discovered in the future, the model would have to be modified and re-analyzed.

Another approach that is often used is simulation [10]. The attacker can be modeled as an agent with specified characteristics (e.g., resources, knowledge, time). The simulation attempts to trace a path from the initial interaction of the agent with the system to the final state. The way in which the agent interacts with different components in the system is specified by the properties of both the agent (e.g., level of knowledge) and the component (e.g., level of defense). The main drawback of this approach is that the extent of the interactions between the various agents has to be specified by the modeler and these can vary significantly in the real world.

To a degree, simulation attempts to understand how a given class of attacker can interact with the system and how the system responds. The conventional approach in cybersecurity assessment is to simulate an attack on a target system by using penetration testing. In this approach, experts with knowledge of the system and its vulnerability attempt to breach and manipulate the system. Vulnerabilities or design flaws that have been exploited are tracked and mitigated or removed after the tests. However, since penetration testing can be expensive, its use is often more limited.

The brief survey of the methods outlined in this section shows the need for an alternative approach to cybersecurity assessment that is not only easy to perform but also can be scaled to large systems. This new approach should fully consider the unique aspects of the types of dependencies that digital components can have on each other, and also allows for the inclusion of more traditional dependencies (e.g., power supply) which are already modeled in traditional risk assessments. One such approach is presented in the next section.

3 APPROACH

Although the number of digital components in modern nuclear power plants may be in the tens of thousands, they can often be grouped into a much fewer number of groups. The components in each group possess similar characteristics in terms of their vulnerabilities to attacks and how they behave if compromised. There have been some efforts at identifying important properties of common digital components that are used in nuclear facilities [11]. The approach presented in this paper uses these attributes to determine how an initially compromised component can be used to advance an attack. Figure 1 illustrates the key steps in the approach.

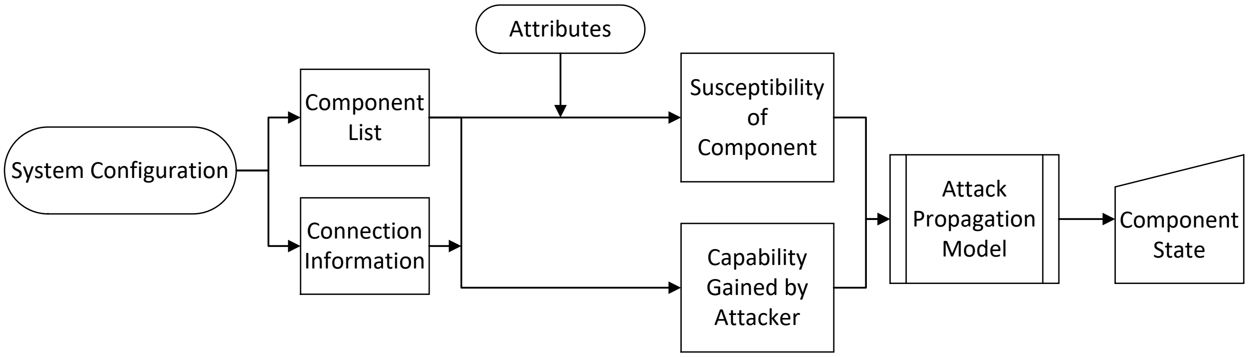

Figure 1. Flowchart for the generation of attack scenario.

The system configuration, which includes the network architecture and component information, is used to generate rules on how attacks can propagate. Each component in the system (as specified in the system configuration) has an associated set of attributes. These attributes determine the susceptibility of a component to an attack, the capability gained by the attacker if that attacker manages to gain complete control of the component, and how different components interact with one another during an attack. For instance, a component that has the capability to generate network traffic can, if compromised cause degradation of the performance of neighboring components that are vulnerable to malicious traffic. Similarly, a component that has user-updatable firmware will be vulnerable to malicious firmware introduced during the update. Tables II and III list several attributes that are relevant to control systems.

Table II. Relationship between attributes and capabilities gained by the attacker

Attribute	Capability Gained	Example Use by the Attacker
Local programming interface	Change code, Change data (given local access)	Insert malicious subroutine
Remote programming interface	Change code, Change data	Insert malicious subroutine
Network interface - output	Generate arbitrary network packets	Denial of service (DoS)
Network interface - input	Capture arbitrary network packets	Intelligence gathering
Firmware	Hardware control	Bypass security mechanism
Operating System	Hardware control	Bypass security mechanism
Writable data storage	Persistence	Stealth, persistent storage
Local user display	Fake output	Man-in-the-middle
Memory	Change data	Alteration of calibration data

Table III. Relationship between attributes and potential susceptibility of the component to an attack class

Attribute	Potential Susceptibility to Compromise
Network interface – input	Increased in CPU cycle and memory usage
Firmware	Alteration of hardware control or access
Operating System	Memory corruption, Execution of arbitrary code, Loss of access control.
Writable data storage	Persistence of malware, Data or code corruption.
Memory	Data or code corruption.

Once the set of attributes has been attached to each component, the potential interactions between components (in the context of the attack propagation) are completely specified. This allows the attack propagation to be modeled (i.e., modeling of how the attacker can compromise other connected components). The output of this stage of the analysis is an attack scenario that contains complete information on how the initial attack vector leads to the final set of compromised components. This information can then be used for further analysis on system response.

The approach outlined above has the advantage that it reduces the number of components that need to be independently analyzed into a smaller number of groups. This will allow the procedure to be applied to systems with a large number of components. Furthermore, unlike some of the approaches outlined in Section 3, the analysis does not depend on discovered vulnerabilities. In a sense, the approach gives the worst-case scenario of how the attack vector can be used to compromise the system.

4 APPLICATION TO A SIMPLIFIED SYSTEM

This section presents the analysis of a simplified pressurizer pressure control system of a typical pressurized water reactor as shown in Figure 2. The system consists of three pressure sensors, each of which is connected to two programmable logic controllers (PLCs). The PLCs, through the control network link, are connected to two pressurizer heater banks. Two of the pressure transmitters are analog while one is digital with an integrated network interface. The control network backbone is assumed to use Ethernet as the lower level communication protocol. This network is also used by other components (not part of the pressurizer control system) such as the data historian and engineering workstation. Unidirectional gateways are used to segment the control network from the enterprise network and from the reactor protection system network.

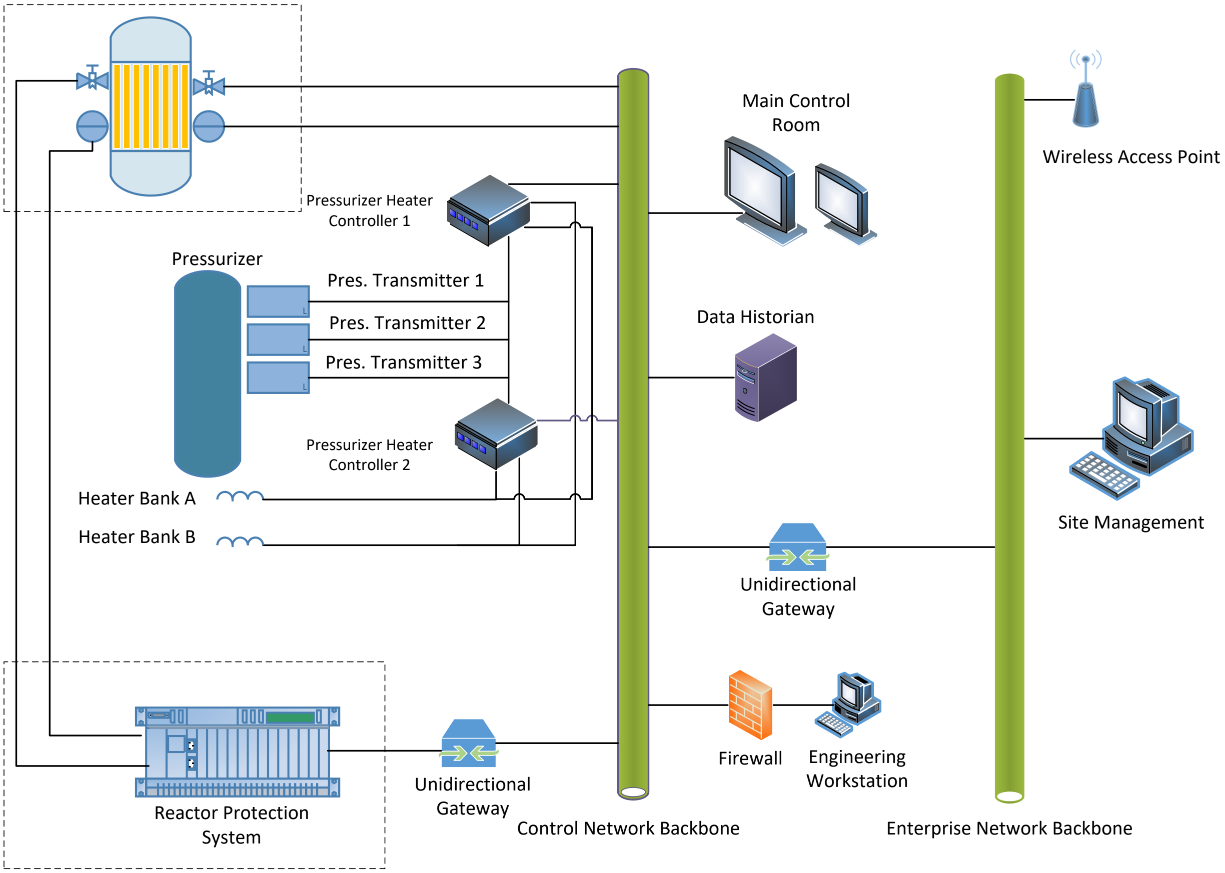

Figure 2. Simplified pressurizer pressure control system.

Each of the components in the system is assigned a set of attributes. For example, the digital pressure sensor has the attribute “Network interface – output”. This means that if this component were to be compromised (for instance, from a supply chain attack through changes in the network control stack), then it could cause a denial of service (DoS) to components connected to the control network by flooding the network with data. (Note that this was the primary attack mechanism on the Davis-Besse and Browns Ferry NPPs events described in Table I.) Figure 3 shows some of the scenarios obtained through the analysis.

In scenarios 1 and 3, two attacks (each using a different attack vector) lead to increased network latency. The first scenario is initiated by a software update of the PLC using a compromised firmware. This altered firmware causes the PLC’s networking hardware to send network packets at a high rate, using up the bandwidth of the communication link. This causes legitimate traffic to experience larger than normal delays. Scenario 3 leads to the same effect (i.e., increased latency), but instead of the corrupt PLC firmware, the attack is initiated through the supply chain of the digital pressure sensor (e.g., using hardware with maliciously added functionalities). In each of these scenarios, the impact of the increased latency on the performance of plant systems using this communication link will need to be studied using system modeling (e.g., thermal-hydraulics code or simulator).

In scenario 2, the memory content of the PLC is altered using a local debugging interface on the PLC. For the example presented here, the heater on/off setpoint is altered. As before, the impact on the system (e.g., burnout of the heater or under/over-pressurization will need to be studied using system modeling.

Note that these scenarios are generated automatically based on the attributes and connection information entered into the model. For the simple example here, the scenarios may appear obvious. However, the method scales well with the number of components and can be used for large systems. As more information on the properties of the components is obtained, the attributes can be refined so that the scenarios that are generated reflect available information.

Scenario 1 (PLC):
Firmware update -> PLC network stack compromise -> DoS -> Increased latency for control network communication

Scenario 2 (PLC):
JTAG interface -> PLC Memory access -> Change in heater setpoint

Scenario 3 (Digital P sensor)
Supply chain – network stack alteration -> Dos -> Increased latency for control network communication

Figure 3. Some attack scenarios leading to misbehavior of components in the pressurizer pressure control system.

One observation that we can make from this example is that the behavior of the two types of pressure sensor (one has a built-in analog-to-digital converter, the other does not) is different. Even though both types perform the same function (sending pressure information to the PLCs), their attributes are different (one containing a programmable network stack, the other does not). This demonstrates that for cybersecurity assessment, it is advantageous to define the components in terms of their attributes rather than their function in the system.

5 CONCLUSIONS

We have presented an approach where sets of attributes are used to determine

1. How a component will impact other connected components if it is compromised,
2. How a component will be impacted by other connected components,
3. The behavior of a compromised component.

These attributes are used to generate sequences of events (i.e., attack propagation model) delineating how an initial foothold on the system can lead to compromise of multiple components. The simple example presented in Section 4 shows how digital assets that perform a similar function may respond completely differently during an attack. This behavior is governed by properties (i.e., the attributes) that are not related to the function of that component in the system.

The formulation of the attributes can generally be done with consultations from experts and from analysis of past compromises. Work is ongoing to expand the list to make them applicable to a wider class of ICS components. The output from the analysis can be used to give insights into how the components will behave when compromised and ultimately how the system will respond. This allows systems behaviors that may indicate compromise to be promptly investigated. Follow-up work is in progress to use the information from the analysis to study responses at the system level. For instance, how the increased network latency impacts the performance of the pressure control system and other systems sharing the network resources will be investigated with the help of plant models. A longer-term goal is to extend the techniques presented in this paper so that they can be easily used as a baseline assessment of the cybersecurity of digital control and protection systems used in nuclear facilities.

6 ACKNOWLEDGMENTS

The work described here is sponsored by the United States Department of Energy, Office of Nuclear Energy (DOE-NE) through the Nuclear Energy Enabling Technologies (NEET) program. The work was performed at Brookhaven National Laboratory, operated by Brookhaven Science Associates for the U.S. Department of Energy under Contract No. DE-FOA-0001129.

7 REFERENCES

1. Regulatory Guide 5.71, "Cyber security Programs for Nuclear Facilities," U. S. Nuclear Regulatory Commission, (2010).
2. Letter from the Advisory Committee on Reactor Safeguards (ACRS) to the Chairman of the U.S. NRC, "*Draft Final Regulatory Guide 5.71, 'Cyber Security Programs for Nuclear Facilities'*," dated November 12, 2009, Agencywide Document Access and Management System (ADAMS) Accession No. ML093130111, (2009).
3. F. Mitch McCrory, "Cyber Informed Risk Analysis (CIRA) for Nuclear Power Cyber Security," *Proceeding of the Institute of Nuclear Materials Management Annual Meeting*, Washington, DC, March 17-18 (2015).
4. "Outpacing Cyber Threats: Priorities for Cybersecurity at Nuclear Facilities," http://www.nti.org/analysis/reports/outpacing-cyber-threats-priorities-cybersecurity-nuclear-facilities (2016).
5. Y. Cherdantseva, et. al., "A Review of Cyber Security Risk Assessment Methods for SCADA Systems," *Computers & Security,* **56**, pp 1-27 (2016).
6. Jung-Woon Lee and Kee-Choon Kwon, "A Cyber Security Risk Assessment for the Design of I&C Systems in Nuclear Power Plants," *Nuclear Engineering and Technology*, **44** (2012)
7. W. Ahn, et. Al., "Development of Cyber-Attack Scenarios for Nuclear Power Plants Using Scenario Graphs," *International Journal of Distributed Sensor Networks*, **32** (2015).
8. S. Jajodia and S. Noel, "Topological Vulnerability Analysis: A Powerful New Approach for Network Attack Prevention, Detection, and Response," *Algorithms, Architectures and Information Systems Security*, World Scientific, New Jersey (2009).
9. I. Kotenko and A. Chechulin, "A Cyber Attack Modeling and Impact Assessment Framework," *Proceeding of the 5th International Conference on Cyber Conflict*, Tallinn, Estonia, June 4 – 7, (2013).
10. F. Baiardi and D. Sgandurra, "Assessing ICT Risk Through a Monte Carlo Method," *Environment Systems and Decisions*, **33**, pp. 486-499 (2013).
11. "Cyber Security Technical Assessment Methodology: Vulnerability Identification and Mitigation: Vulnerability Identification and Mitigation," EPRI Technical Report 3002008023, EPRI, Palo Alto, CA (2016).