Development of a Quantitative Method for Evaluating the Efficacy of Cyber Security Controls in NPPs based on Intrusion Tolerant Concept

Chanyoung Lee [a], Poong Hyun Seong [a*]
[a] *Department of Nuclear and Quantum Engineering, Korea Advanced Institute of Science and Technology, 291 Daehak-ro, Yuseong-gu, Daejeon, 34141, Republic of Korea*
[*] *Corresponding author: phseong@kaist.ac.kr*

1. Introduction

Digital I&C systems have been developed and installed in nuclear power plants (NPPs). However, due to installation of digital I&C systems, cyber security concerns are increasing in the nuclear industry [1]. In order to provide useful information about cyber security issues, many regulatory documents, guides and standards were already published in the nuclear industry. The documents include cyber security plans, methods for cyber security assessments and comprehensive set of security controls. However, there are still difficulties when it comes to deciding which security controls are needed and to defining appropriate security control requirements [2]. It is because that practical examples for the application of security controls have not been available to system designers, and methods that can help assess how much security is improved if a specific control is applied are not included. In practice, evaluating security controls has been heavily based on human experts' experiences in the nuclear industry. For a more scientific approach, this study developed a measure of 'cyber security improvement' and suggests a way to quantitatively evaluate the effectiveness of security measures.

2. Development of a measure of 'cyber security improvement' based on intrusion tolerant concept

2.1 A measure of 'cyber security improvement'

The extent of cyber security improvement caused by security enhancement is defined as reduction ratio of the failure probability to secure the system from cyber-attack.

$$\textit{Cyber Security Improvement} = 1 - \frac{P_{Enhanced}}{P_{Current}}$$

Based on the aim, to protect availability of essential functions, the intrusion tolerant concept is applied to the measure of 'cyber security improvement' for evaluating security controls establishing defense-in-depth protective strategies [3]. Based on intrusion tolerant strategies: resistance strategy, detection strategy, mitigation strategy and practical assumptions, an event tree was constructed. Using the event tree shown as Fig1, the failure probability to secure the system from cyber-attack can be estimated as follows.

$$P_a(P_d + P_m - P_d P_m) \qquad (1)$$

Where, P_a is the probability that resistance strategy fails, P_d is the probability that detection strategy fails, and P_m is the probability that mitigation strategy fails. Based on definition of 'cyber security improvement': reduction ratio of the probability that a cyber attack damages a target system, it can be estimated as following Eq. 2.

$$1 - \frac{P_a'(P_d' + P_m' - P_d' P_m')}{P_a(P_d + P_m - P_d P_m)} \qquad (2)$$

Where, P_a', P_d', and P_m' are the probabilities with respect to enhanced system.

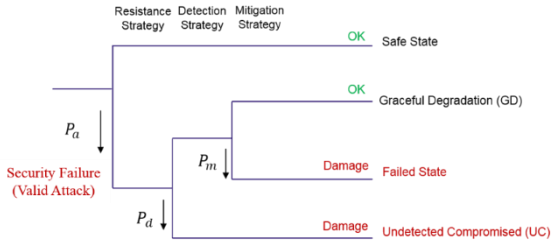

Fig. 1. Event tree based intrusion tolerant concept

2.2 Adoption of the concept of 'mean time to compromise

Estimating the failure probability of resistance strategy is more challenging than quantifying the failure probability of detection and mitigation strategies. Because exploitation of vulnerabilities has a strong dependence on not only vulnerable degree of a target system but also, attacker's factors such as skill-level and accessibility to information of a target system. In spite of these limitations, there have been several attempts to estimate the difficulty of actions taken by an attacker. Several cyber security researchers observed that cyber security level can be increased as the effort expended by an attacker increases [4]. The concept of 'mean time to compromise (MTTC)' was investigated in this work as a measure of effort expended by an attacker [5]. The value of MTTC is the estimation of the time required to a valid attack assuming that the effort is expended uniformly. The focus of MTTC is not to obtain the amount of actual time, but to quantify the difficulty of actions.

However the main limitation of the initial model lies in their lack of distinction of between different vulnerabilities and overly simplified attack model. For this problem, another model which links 'mean time to compromise' to the well-known 'common vulnerability scoring system (CVSS) [6]' was adopted in this work [7].

The equations estimating values of MTTC are used in this work.

2.3 Revision of the adopted model

The adopted model uses a vulnerability-specific attack graph. The attack graph represents knowledge about vulnerabilities' interdependence and potential sequences of attacks, so that attack paths with different MTTC values may encounter in series or parallel structure. For the parallel structure case, in the adopted model, there is an assumption that an attacker chooses exploitation based on relative difficult, and a probability that a path will be chosen is also based on relative difficulty of the attack path [7]. However, the assumption can lead to contradictory results when two versions (baseline and enhanced) of a same system are compared. Compensating for a relatively insecure attack path, having lower MTTC value, normally results in an increase in the MTTC value. On the other hand, complementing a relatively secure attack path, having higher MTTC value, leads to contradictory result that the MTTC value decreases. For example, in the case that three attack path A, B, C encounter in parallel structure like as Fig. 2, the MTTC value decreases.

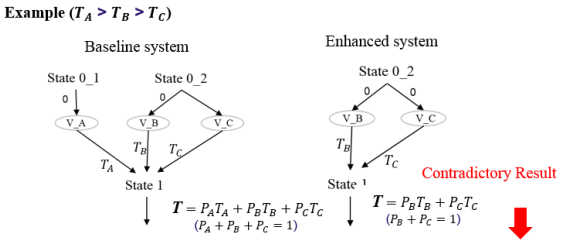

Fig. 2. Example of contradictory case

To reflect the argument that the direction to reduce the attack path is to improve security, a literature that describes a vulnerability as one attack surface are referred [8]. The smaller the attack surface is, the more secure the system is. In the same vein as covering the degree of system security, the two concepts (MTTC and attack surface) are to an extent related, and it is assumed that they are inversely proportional to each other. Attack surface in an attack process can be obtained by the summation of attack surfaces created by each vulnerability. With this regards, MTTC values can be obtained by referring to the way of attack surface. This revised case is illustrated in Fig. 3.

$$\frac{1}{\frac{1}{T_A}+\frac{1}{T_B}+\frac{1}{T_C}} \tag{3}$$

In this way, compensating for relatively insecure vulnerabilities and relatively secure vulnerabilities can result in increased MTTC values depending on the magnitude of each effect.

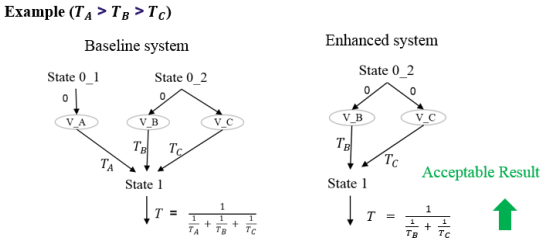

Fig. 3. Example of acceptable case

2.4 Replacement of the failure probability of resistance strategy with total MTTC

The completed attack graph can be used to determine the value of total MTTC. From 'User' to 'Malicious Activity' the value of total MTTC can be obtained following attack paths. When vulnerabilities are connected in series structure, the values of MTTC are summation of each MTTC of vulnerability, and when vulnerabilities are connected in parallel structure, the value of MTTC are calculated by referring to the way of attack surface concept. In this way, the value of total MTTC in the baseline system and the value of total MTTC in the enhanced system can be obtained. Because the probability of valid attack is inversely proportional to the value of MTTC, the ratio of probability of valid attack ($\frac{P_{a'}}{P_a}$) in the measure of 'cyber security improvement' can be replaced with the ratio of MTTC value ($\frac{T}{T'}$). T is the value of total MTTC in the baseline system and T' is the value of total MTTC in the enhanced system. Therefore the measure of 'cyber security improvement' can be replaced by Eq. 4.

$$1-\frac{T(P_{d'}+P_{m'}-P_{d'}P_{m'})}{T'(P_d+P_m-P_dP_m)} \tag{4}$$

3. A case study

In this study, the target system is the digital plant protection system (DPPS) which is a safety-critical I&C system of NPP. In this work, only the case of malware implementation was considered among possible threats under insecure maintenance devices. According to a study that conducted a cyber risk assessment on the nuclear safety system, possible attack types in the CDA malware implementation situation are 'DoS attack', 'Improper Command', 'Data Modification' [9]. Only these three kinds of possible attack types are considered in this work. To propose a more detailed description of the attack type, a simplified target system is created. In addition, the above-mentioned attack types are limited and exemplified in Fig. 4.

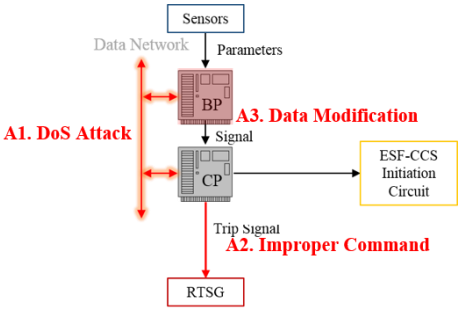

Fig. 4. Example of attack types

For the case study, security controls: vulnerability management and application of intrusion detection system (IDS) are assumed. In order to obtain the MTTC value of the baseline system, existing vulnerabilities of the sample operating system and data network were searched at the national vulnerability database (NVD) [10]. In order to avoid exposing the actual plant information, different situation from the actual nuclear power plant is assumed. Based on the information of the analyzed vulnerabilities and the situation of the system, we created an attack graph for each attack type. Fig. 5 is the attack graphs for 'Improper Command', in the baseline system.

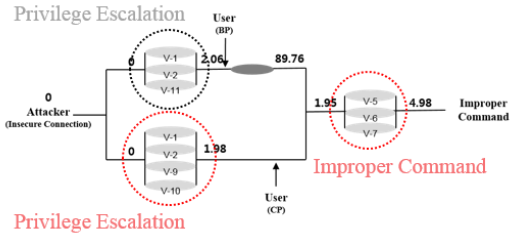

Fig. 5. Attack graph for 'Improper Command' in baseline system

The MTTC value was obtained based on the CVSS 'Base Score' of each vulnerability. The value of total MTTC of each attack graph of the baseline system was obtained by the method presented in the previous chapter. Total MTTC value for 'DoS Attack' is 3.88 days, for 'Improper Command' is 4.98 days, and for 'Data Modification' is 6.76 days.

The effects of two kinds of vulnerability managements were compared. Management 1: It complements the vulnerabilities related with 'buffer overflow' error among the investigated vulnerability. Management 2: It changes the system configuration by transferring vulnerabilities into new additional hosts. The attack graphs in the enhanced system are created after applying the two kinds of vulnerability management to the baseline system. Fig. 6, Fig. 7 are the attack graphs for 'Improper Command' when vulnerability management 1 and 2 are applied to the baseline system.

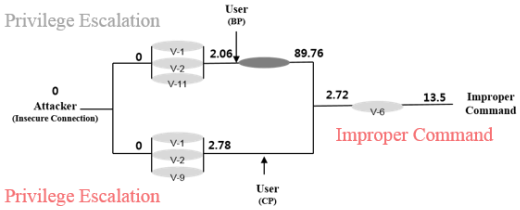

Fig. 6. Attack graph for 'Improper Command' when management 1 is applied

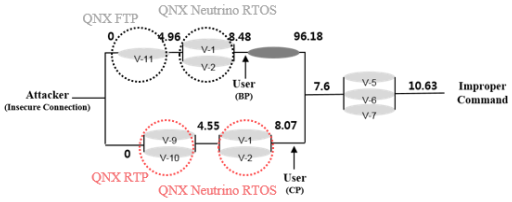

Fig. 7. Attack graph for 'Improper Command' when management 2 is applied

Total MTTC value for 'DoS Attack' is 4.24 days, for 'Improper Command' is 13.5 days, and for 'Data Modification' is 8.92 days when vulnerability management 1 is applied. For the case of application of vulnerability management 2, total MTTC value for 'DoS Attack' is 8.35 days, for 'Improper Command' is 10.63 days, and for 'Data Modification' is 12.69 days. They are summarized in Fig. 8. In Management 1, where only specific attack types were targeted, only one type showed a big improvement. In management 2, where changed the overall system structure, it showed proper effect in all types. In Management 1, where only specific attack types were targeted, only one type showed a big improvement. In management 2, where changed the overall system structure, it showed proper effect in all types.

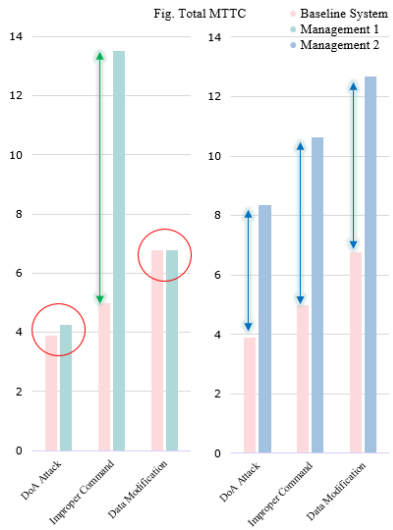

Fig. 8. Total MTTC values in baseline system and enhanced systems

The ratios of probability of valid attack ($\frac{P_a'}{P_a}$) in the measure of 'cyber security improvement' were obtained

through the ratio of MTTC value ($\frac{T}{T'}$) for each attack type and for each applied vulnerability management. They are summarized in Table 1.

Table 1: Ratios of $\boldsymbol{P_a}$ and $\boldsymbol{P_a}'$ ($\frac{P_{a'}}{P_a} = \frac{T}{T'}$)

	Dos Attack	Improper Command	Data Modification
Management 1	0.92	0.37	1
Management 2	0.46	0.47	0.53

Assumed failure probabilities of detection and mitigation strategies in baseline system are referred to the document modeling an intrusion tolerant system [3]. The probability that detection strategy fails (P_d) is 20%. The probability that mitigation strategy fails (P_{m}) is 40%.

In the enhanced system probability of detection was increased due to IDS. Referring to the experiments using the IDS based on 'Snort Rule' algorithm [11], the probability of detection were obtained in case by attack types. Although same detection algorithm is used, the results can be changed depending on used hardware, software and system state [12]. Hence, further researches are needed for acceptable values, but they suffice as a proof of the concept. In this work, mitigation process was not improved.

Table II: Detection probability and mitigation probability according to attack types in enhanced system

Attack Type	Probability of Detection P_d'	Probability of Mitigation P_m'
DoS Attack	100%	60%
Improper Command	89.8%	60%
Data Modification	93%	60%

The 'cyber security improvement' for all kinds of attack types are summarized in Fig. 9 when one of assumed security controls is applied to the system and one of vulnerability managements and IDS are applied to system at the same time.

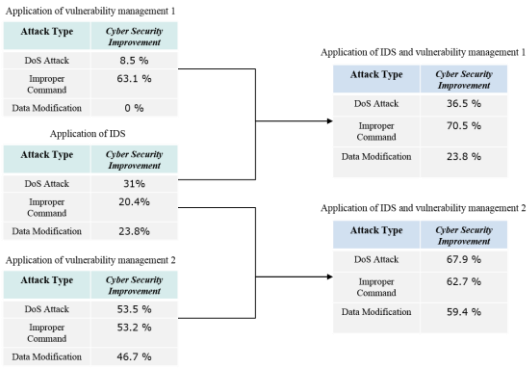

Fig. 9. Results of case study

With these results, different approaches are recommended depending on whether a particular attack type is prioritized or the overall rise is prioritized

4. Conclusions

In order for useful information about cyber security issues, many regulatory documents, guides and standards have been already published in the nuclear industry. However, there are still difficulties when it comes to deciding which security controls are needed and to defining appropriate security control requirements. It is because practical examples for the application of security controls have not been available to system designers and there is a lack of means for estimating the effectiveness of security controls. In this regard, this paper suggested a framework to quantitatively evaluate how much cyber security is improved when specific cyber security controls are applied in NPPs.

The result values of the 'cyber security improvement' can help assess how much system security can be improved if specific cyber security controls are applied, and which types of additional cyber security controls should be taken. In addition, it is expected that the suggested method can be applied to select appropriate security controls among various options in advance. Furthermore, by evaluating cyber security controls quantitatively, it can be also applied to establish a specific target of efficacy level that system designers can achieve. However, there are some limitations in this work to estimate the efficacy of cyber security controls. It is because the methods for obtaining probabilities of detection strategy and mitigation strategy need to be elaborated. Also the verification and validation of the suggested method need to be improved.

REFERENCES

[1] 한국 원자력 학회 "원자력 사이버보안 현안과 정책 제언" 2015. 8. 21.
[2] Song, Jae-Gu, et al. "An analysis of technical security control requirements for digital I&C systems in nuclear power plants." Nuclear Engineering and Technology 45.5 (2013): 637-652.
[3] Madan, Bharat B., et al. "A method for modeling and quantifying the security attributes of intrusion tolerant systems." Performance Evaluation 56.1 (2004): 167-186.
[4] McQueen, Miles A., et al. "Time-to-compromise model for cyber risk reduction estimation." Quality of Protection. Springer US, 2006. 49-64.
[5] Dacier, Marc, Yves Deswarte, and Mohamed Kaâniche. "Quantitative assessment of operational security: Models and tools." Information Systems Security, ed. by SK Katsikas and D. Gritzalis, London, Chapman & Hall (1996): 179-86.
[6] Mell, Peter, Karen Scarfone, and Sasha Romanosky. "Common vulnerability scoring system." IEEE Security & Privacy 4.6 (2006): 85-89.

[7] Nzoukou, William, et al. "A Unified Framework for Measuring a Network's Mean Time-to-Compromise." 2013 IEEE 32nd International Symposium on Reliable Distributed Systems. IEEE, 2013.
[8] Howard, Michael, Jon Pincus, and Jeannette M. Wing. "Measuring relative attack surfaces." Computer Security in the 21st Century. Springer US, 2005. 109-137.
[9] Song, Jae-Gu, et al. "A cyber security risk assessment for the design of I&C systems in nuclear power plants." Nuclear Engineering and Technology 44.8 (2012): 919-928.
[10] National Vulnerability Database (NVD), http://nvd.nist.gov
[11] Gao, Wei, and Thomas H. Morris. "On cyber attacks and signature based intrusion detection for modbus based industrial control systems." The Journal of Digital Forensics, Security and Law: JDFSL 9.1 (2014): 37.
[12] Kim, Man Cheol, and Seung Jun Lee. "Important factors affecting fault detection coverage in probabilistic safety assessment of digital instrumentation and control systems." Journal of Nuclear Science and Technology 51.6 (2014): 809-817.