Development of a Multi-layer State Estimation Method for Functional Impact and Recoverability Analysis Under Cyber Attacks in NPPs

Chanyoung Lee [a], Young Ho Chae [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

The application of digital and automation technologies to NPP I&C systems has raised the cyber security concerns in the nuclear industry. Several cyber-attack cases on nuclear facilities proved that cyber-attacks are possible in NPPs even if they are separated from the external network. Regulatory bodies require all nuclear facilities to have sufficient prevention and response capability [1]. Since cyber-attacks tend to be implemented with malicious intentions and through persistent attempts, it is impossible to have a complete prevention plan. The deployed security controls may become insufficient against the evolving cyber-attacks. The security controls can be breached or easily bypassed by information leakage or insiders. Immediate security patch is not allowable in NPPs, systems may be exposed to threats for a long time. For this reason, prevention and response plans should be established to complement each other.

However, there are several limitations. The existing fault diagnosis techniques are limited in detecting and analyzing symptoms of cyber-attack. Since pattern of cyber-attacks is unpredictable, developing a detailed response procedure is impossible. Response plans and techniques, developed in other industries, cannot be applied to NPPs. Without detailed procedures, operators will be faced with tasks of diagnosing the uncertain situations and taking response actions in a short time [2]. In order to solve the problems, an integrated cyber-attack response plan and an integrated response support system need to be developed.

2. Analysis of an Integrated Cyber-attack Response Plan in NPPs

Although several safety features have been implemented in NPPs, physical damage can be caused by cyber-attacks in various ways. Fortunately, it is possible to take safety response manually, as long as operators are able to recognize the current situation correctly. However, unexpected cyber-attack situations can make the human reliability of operators degraded. Confused operators may be hard to perceive the current situation and error-prone [3]. In addition, cyber-attacks can deliberately prevent operators from perceiving the current situation. Even, a cyber-attack with specific NPP knowledge could increase the possibility of wrong operator actions and cause serious consequences [4]. Without security analysis, the existing safety response plans have limitations. Operators may believe the manipulated information without suspicion of cyber-attacks. The detected cyber anomaly may be regarded as a simple device failure. Unobservable and potential cyber-attack damage cannot be identified and responded. The IAEA insisted that security analysis activities should be included in the scope of EOPs [5].

In the safety-critical industry, safety-related security analysis must be able to be conducted on site. With this regard, the primary purpose of cyber-attack response plan in NPPs is to ensure the integrity and rapid recovery of essential system functions associated with safety, security, emergency preparedness (SSEP). Based on the response purpose, the safety-related security analysis could be:

- Determining the impact of the incident on SSEP and identifying safety actions to place the organization or facility in a safe condition.
- Identifying the extent of the incident to establish an adequate security response.

Safety-related security analyses for safety and security response planning are not separated. It is because that identifying unobservable and potential SSEP damage is based on the analysis of security incident. The safety-related security analysis activities are divided and correlated as Fig. 1. However, the defined analytical activities could exceed the cognitive capability of human operators. An integrated security analysis support system needs to be developed. Not only helpful, but quantitative and intuitive information needs to be provided for individual analysis. The individual analysis activities need to be supported in an integrated manner aligning with the cognitive process model.

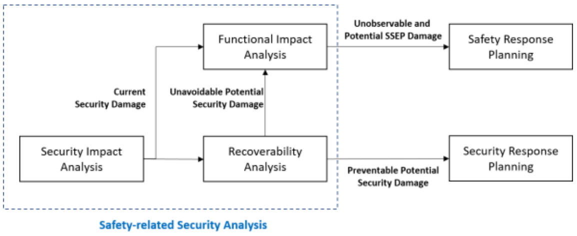
Functional Impact Analysis
Unobservable and Potential SSEP Damage
Safety Response Planning
Current Security Damage
Unavoidable Potential Security Damage
Security Impact Analysis
Recoverability Analysis
Preventable Potential Security Damage
Security Response Planning
Safety-related Security Analysis

Fig. Safety-related Security Analysis in NPPs

3. Development of a Multi-layer State Estimation Method for Functional Impact Analysis

Cyber-attacks tend to compromise security conditions sequentially required for their specific attack goals by security exploiting vulnerabilities. Defining a transient system security state as a set of compromised security conditions, current security state can be interpreted as the cyber-attack progress and security impact [6]. The current security state cannot be observed directly and should be inferred using partially observed cyber anomaly alarms. Assuming security state transitions and alarm generations follow probabilistic distributions, cyber-attack process can be modeled as a hidden Markov Model. However, since sufficient training data sets are required to construct HMMs, its application has been limited in areas where training data sets are unavailable. In this study, a method is developed to construct HMMs using the available system and security knowledge. In addition, an online model update method is developed.

A security condition dependency graph is used to determines the feasible states and state transitions [6]. An example of security condition graph is described in Fig. 2. A dependency graph defines logical relationships of security conditions and vulnerability exploitations. The security state transition probability depends on the exploitable security vulnerabilities and implemented attack techniques. The CVSS is used to quantify the impact and exploitability of vulnerabilities [7]. The concept of activity level of cyber-attack is suggested for shaping the state transition probability. Cyber anomaly alarms are generated when certain types of vulnerabilities are being exploited. Each security state determines exploitable vulnerabilities and alarms to be generated. Detection systems have their inherent false-negative and false-positive error rates. The concept of stealthy level of cyber-attack is suggested for shaping the false-negative error rate.

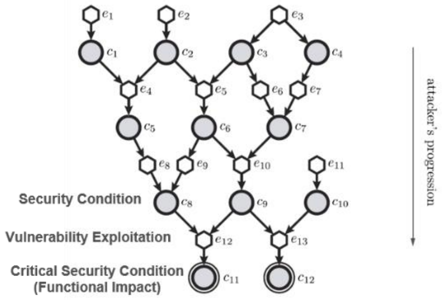

Fig. Security Condition Dependency Graph [7]

Several HMMs can be constructed within the pre-defined adversary range. The suitability of constructed HMM can be evaluated by the evaluation module with online observations. Since, whether a selected model is the optimal model cannot be judged by a few initial observations, Models constructed within the adversary range must be re-evaluated upon new observations. In addition, the adversary range needs to be shifted and narrowed gradually as observations are accumulated. The developed online update method can update the transition and observation patterns of unreached security states following the investigated system and security knowledge. The online model update algorithm enables to estimate security state using the most optimal model. In addition, it enables to maintain acceptable estimation performance under false positive error cases and false negative error cases. The estimated current security state and state transition path enables to functional impact and recoverability analysis. Elaborated methods will be developed in future work.

4. Summary and Conclusion

In this study, an integrated cyber-attack response plan and an integrated response support system are developed. The developed support system is based on the HMMs based state estimation method. A method is developed to construct HMMs using the available system and security knowledge. In addition, an online model update method is developed. The online model update algorithm enables to estimate security state using the most optimal model. In addition, it enables to maintain acceptable estimation performance under false positive error cases and false negative errors cases. Elaborated methods for functional impact analysis and recoverability analysis will be developed in future work.

ACKNOWLEDGEMENTS

This research was supported by the National R&D Program through the National Research Foundation of Korea (NRF) funded by the Korean Government. (MSIP: Ministry of Science, ICT and Future Planning)
(No. NRF-2016R1A5A1013919)

REFERENCES

[1] USNRC, “REGULATORY GUIDE 5.71 Cyber Security Programs for Nuclear Facilities,” no. January, pp. 1–105, 2010.
[2] Y. Zhao, L. Huang, C. Smidts, and Q. Zhu, “Finite-horizon Semi-Markov Game for Time-sensitive Attack Response and Probabilistic Risk Assessment in Nuclear Power Plants,” Reliab. Eng. Syst. Saf., p. 106878, 2020.
[3] P. Gontar, H. Homans, M. Rostalski, J. Behrend, F. Dehais, and K. Bengler, “Are pilots prepared for a cyber-attack? A human factors approach to the experimental evaluation of pilots’ behavior,” J. Air Transp. Manag., vol. 69, no. February 2017, pp. 26–37, 2018.
[4] H. E. Kim, H. S. Son, J. Kim, and H. G. Kang, “Systematic development of scenarios caused by cyber-attack-induced human errors in nuclear power plants,” Reliab. Eng. Syst. Saf., vol. 167, no. May, pp. 290–301, 2017.
[5] IAEA, Computer Security Incident Response Planning at Nuclear Facilities. 2016.
[6] E. Miehling, M. Rasouli, and D. Teneketzis, “A POMDP Approach to the Dynamic Defense of Large-Scale Cyber Networks,” IEEE Trans. Inf. Forensics Secur., vol. 13, no. 10, pp. 2490–2505, 2018.
[7] P. Mell, K. Scarfone, and S. Romanosky, “A Complete Guide to the Common Vulnerability Scoring System Version 2.0,” FIRSTForum Incid. Response Secur. Teams, 2007.