Improvements of PSA-Based Cybersecurity Assessment in Nuclear Power Plants

Dayoung Jung, Joonseok Lim, Gyunyoung Heo*
Department of Nuclear Engineering, College of Engineering, Kyung Hee University, Yongin, Republic of Korea
**Corresponding author : gheo@khu.ac.kr*

***Keywords : nuclear cybersecurity, probabilistic safety assessment, critical digital asset, attack vector**

1. Introduction

In recent years, the digitalization of instrumentation and control (I&C) systems in nuclear power plants (NPPs) has improved reliability and operational efficiency, but it has also increased the potential for cyberattacks. The 2010 Stuxnet attack demonstrated that digital intrusions can cause physical damage and lead to severe safety consequences. [1]

In response, regulatory authorities have established cybersecurity requirements. The U.S. Nuclear Regulatory Commission (NRC) issued 10 CFR 73.54 and Regulatory Guide (RG) 5.71, while the Korea Institute of Nuclear Nonproliferation and Control (KINAC) introduced RS-019 and RS-015, requiring the identification and protection of Critical Digital Assets (CDAs) supporting safety, security, and emergency preparedness (SSEP) functions. However, in modern NPPs, the number of CDAs can reach thousands, making it impractical to secure all assets equally. [2]

Probabilistic Safety Assessment (PSA), originally developed for quantifying risks from hardware failures, human errors, and external hazards, has been extended to incorporate cyber threats. By mapping cyberattack scenarios into PSA event trees and fault trees, and quantifying their impact on core damage frequency (CDF) or conditional core damage probability (CCDP), PSA-based cybersecurity assessment provides a risk-informed framework.

This study reviews existing PSA-based cybersecurity assessment methodologies, analyzes their limitations, and identifies gaps. The goal is to propose directions for PSA model revision that enhance regulatory applicability of cybersecurity risk assessment in NPPs.

2. Application and Limitations of Cybersecurity PSA

2.1. Methodologies for Cybersecurity Assessment

In cybersecurity assessment, three essential processes are required: (1) Identification of Critical Systems CS/CDA, (2) Selection of attack vectors and threat scenarios, and (3) Quantification of security measures.

CS/CDA identification is the process of determining digital assets that perform essential functions for SSEP within an NPP. This step defines the minimum scope of assets that must be protected against cyber threats. Over-identification of CS/CDAs may reduce management efficiency, whereas under-identification may compromise safety and security functions. Therefore, a systematic and quantitative methodology is required. [2]

Selection of attack vectors and threat scenarios specifies the actual pathways through which the identified CS/CDAs may be attacked, such as networks, wireless channels, portable media, supply chain, or physical access. This process goes beyond a simple list of vulnerabilities and derives realistic threat scenarios that reflect the characteristics and environment of the assets. This enables the incorporation of cyber-attacks into PSA models as external events.

Quantification of security measures evaluates the effectiveness of protective, detective, responsive, and recovery controls in preventing or mitigating cyber-attacks within the threat scenarios. This step moves beyond the binary verification of whether security controls exist, instead producing a quantitative estimation of attack success probability. The quantified results can be reflected in PSA risk indicators such as CDF and CCDP. [3]

Accordingly, these three processes constitute the fundamental pillars of nuclear cybersecurity risk assessment. This study reviews and compares various methodologies corresponding to each process and analyzes their gaps.

2.2. Cybersecurity Studies Based on PSA Models

First, in the CS/CDA identification stage, importance measures (e.g., Fussell–Vesely, RAW) are applied to identify functions that significantly contribute to plant risk when a cyberattack succeeds, along with the digital assets responsible for those functions. In actual studies, the control rod drive module, safety injection system logic processor, instrumentation signal processors, and reactor protection system communication interfaces have been identified as critical CDAs. Moreover, a reverse-tracing approach from accident-contributing functions in existing PSA models has also been suggested to specify the corresponding digital devices as CDAs. [1]

Second, for attack vector derivation and scenario selection, event trees (ETs) are employed to filter out only those cyberattack scenarios that are meaningfully connected to accident progression. By combining attack success probability, detection and blocking likelihoods, and safety system failure probabilities, accident sequences are derived and eventually linked to PSA metrics such as CDF. [4] In some studies, Bayesian Network inference results have been integrated into ET branch probabilities to enable more refined scenario analyses. [1]

Third, for quantification of security measures, the effects of security controls are reflected as changes in the basic event probabilities of fault trees (FTs) or in ET

branch probabilities. For example, multi-factor authentication reduces the probability of command spoofing events, while network segregation significantly decreases the probability of malicious packet injection. Incorporating these into FT/ET models allows estimation of the reduction in CDF.[3] Furthermore, approaches such as the "cyber-informed FT," which incorporates results from System-Theoretic Process Analysis (STPA), have been proposed to systematically model cyber threats and quantitatively compare risk reduction before and after security measures. [5]

In this way, ET/FT-based PSA approaches integrate cyber threats into conventional safety assessments and can be utilized for identifying vulnerabilities, prioritizing attack scenarios, and quantifying the effectiveness of security measures. However, in practice they do not fully encompass all CS/CDAs or exhaustively model all potential attack vectors, and the results are therefore interpreted in terms of representative, meaningful scenarios. Despite these limitations, ET/FT-based approaches significantly enhance the systematicity of cybersecurity PSA, which underscores their importance, while the following section discusses their inherent limitations.

2.3. WINS Guidance's Strengths

WINS guidance is designed to be practitioner-friendly and decision-oriented. It (i) provides a common language that enables non-specialists and specialists to communicate effectively about cyber risks; (ii) uses a self-assessment questionnaire mapped to a five-level maturity scale to structure program benchmarking and improvement; and (iii) stresses assurance of effectiveness through audits, peer reviews, and exercises, thereby focusing attention on outcomes rather than paper compliance. These features support sustainable capacity building through training and certification delivered by the WINS Academy and align with IAEA objectives by helping bridge global recommendations to practical implementation.[6]

3. Limitations of Existing Approaches

PSA traditionally represents risk through the risk triplet — what can go wrong (events and scenarios), how likely it is to occur (frequency/probability), and what the consequences are (outcomes). When cybersecurity factors are integrated into PSA, limitations emerge with respect to each of these three components. [3]

3.1. Event and Scenario Identification Limitations

The first element of the risk triplet concerns what can go wrong, that is, the ability to comprehensively identify potential events and scenarios. In cybersecurity PSA, this element is constrained by the following issues.

3.1.1. Incompleteness of CS/CDA Identification

PSA models are typically constructed around the safety functions defined in existing PSA and the major digital assets supporting those functions, but they do not comprehensively cover all CDAs. Although the control system itself is a digital asset, factors such as EOC(Error of Commission) that may occur during control operations are not sufficiently represented in current PSA models. This may lead to the omission of assets or operational conditions that could significantly influence accident progression. [7]

3.1.2. Simplification of Attack Vector Modeling

Real-world cyberattacks consist of multi-stage intrusions, concealment, evasion, and combinations of complex behaviors. Yet, PSA often reduces them to a single functional failure event or a single probability value. As a result, the sequential and strategic characteristics of realistic attack scenarios are not adequately captured, and the cyber risk derived from PSA tends to be oversimplified compared to actual threats.[2][4]

3.1.3. Lack of Cyber Impact Consideration on Initiating Events

Conventional PSA emphasizes initiating events caused by mechanical or electrical failures.[1][3] However, cyberattacks can themselves trigger initiating events or alter the nature of existing ones. For instance, falsified pressurizer pressure signals or blocked safety injection pump start signals may directly create new initiating events. Current models do not sufficiently account for this, and thus the unique characteristic that cyber threats can intervene at the starting point of accident sequences is not fully integrated into PSA.

3.2. Uncertainty in Probability Assessment

The second element of the risk triplet concerns how likely an event is to occur, which requires a credible quantification of probabilities or frequencies. In cybersecurity PSA, this quantification suffers from the following limitations.

3.2.1. Uncertainty in the Quantification of Security Measures

The effects of security controls are usually represented as reductions in PSA basic event probabilities or event tree branch probabilities. However, these probability values are often based on expert judgment or abstract indicators such as "difficulty," rather than empirical data. Thus, likelihood is often defined in terms of difficulty rather than frequency, leading to criticisms that the results may not adequately reflect actual occurrence probabilities.

When applied to operating plants, WINS BPGs are intentionally technology-agnostic and program-focused, which means they do not prescribe plant-specific

quantification of risk metrics such as CDF/CCDP nor provide parameter values for PSA models. As a result, maturity scores and qualitative assurance evidence may be variably interpreted across licensees; program-level assessments do not directly translate into initiating-event frequencies or basic-event probabilities; and self-assessment can introduce subjectivity without independent verification. These are not deficiencies of the BPGs per se but reflect their scope; license holders must supplement them with plant-specific modeling and data to support risk-informed decisions.[6]

3.2.2. Lack of Consideration for Dynamic and Human Factors

Actual cyber incidents are strongly influenced by time-dependent and human-related factors such as detection delays, operator response times, and interactions between security controls. However, most PSA studies remain limited to static probabilistic models, so these dynamic and behavioral factors are insufficiently reflected. [7][8]

3.3. Constraints in Consequence Analysis

The third element of the risk triplet concerns what the consequences are, focusing on the outcomes that follow from an event once it occurs. Cybersecurity PSA exhibits the following limitation in this respect.

3.3.1. Limitations in Consequence Analysis

Current PSA focuses primarily on severe accident outcomes such as core damage or radiological release, while the significant safety, economic, and operational consequences of plant trips themselves are insufficiently considered. In practice, even a single trip can result in substantial economic losses and a decline in operational reliability. The inability of cybersecurity PSA to incorporate such operational consequences represents an important limitation.

In conclusion, existing ET/FT-based cybersecurity PSA studies face multiple issues, including the incompleteness of CDA identification, oversimplification of attack scenario representation, uncertainty in security measure quantification, limitations in consequence analysis, insufficient consideration of initiating events, lack of data, inadequate treatment of dynamic and human factors, and constraints in practical application. Accordingly, future research needs to adopt expanded approaches that incorporate more sophisticated scenario modeling, data-driven probability estimation, dynamic PSA techniques, and the inclusion of human and operational consequences.

4. Proposed Improvements

4.1. Event and Scenario Identification Enhancements

4.1.1. Re-evaluation of Initiating Events

Initiating events should be reconsidered beyond mechanical and electrical failures to explicitly account for cyberattack perspectives. Attackers may falsify signals, block control commands, or alter system settings, thereby directly triggering or modifying initiating events. Such cyber-induced events should be newly defined as initiating events and incorporated into the PSA model.

4.1.2. Extension of Basic Events

The basic events of fault trees have traditionally been limited to random hardware failures. To address this limitation, cyber events such as command spoofing, malicious packet injection, and integrity verification failures should be introduced as new basic events. By doing so, PSA can evolve from a simple failure-based model into a comprehensive risk model that also encompasses cyber threats.

4.1.3. Explicit Modeling of Control Systems and EOC

The control system, as a critical digital asset, also plays a pivotal role and should be more explicitly represented in PSA models. In particular, EOCs that may occur during control operations need to be incorporated to reflect vulnerabilities at the control system level. This would allow PSA to capture not only safety system failures but also the contribution of control system vulnerabilities to accident progression. [7]

4.2. Enhancing Probability Assessment

4.2.1. Data-driven Quantification of Security Measures

In quantifying the effectiveness of security measures, current approaches heavily rely on expert judgment and difficulty-based indicators. These must be supplemented by data-driven assessments. Simulation testbeds, penetration testing, and cyber threat intelligence can be leveraged to obtain empirical data. In addition, probabilistic inference techniques such as Bayesian Networks can be applied to integrate empirical evidence with expert judgment, thereby reducing uncertainty.

4.3. Expanding Consequence Analysis

4.3.1. Expansion of Risk Metrics

Risk metrics should be extended beyond conventional safety indicators such as CDF and Large Early Release Frequency (LERF). Expanded measures should include plant trip probability, availability losses, and economic impacts. This reflects the fact that cyberattacks can have significant implications not only for safety but also for operational reliability and economics.

In essence, improvements such as the re-evaluation of initiating events, the extension of basic events to include cyber scenarios, the explicit modeling of control system vulnerabilities and EOC, the expansion of risk metrics, data-driven probability estimation, and the integration of dynamic and human factors will enable cybersecurity PSA to evolve into a more precise and reliable analytical framework. With such enhancements, cybersecurity PSA can move beyond being a mere accident modeling tool and serve as a practical risk management framework that supports decision-making in both plant operations and regulatory contexts. Furthermore, the limitations and corresponding improvements are summarized below according to the three elements of the risk triplet (event/scenario, probability, and consequence) (Fig 1).

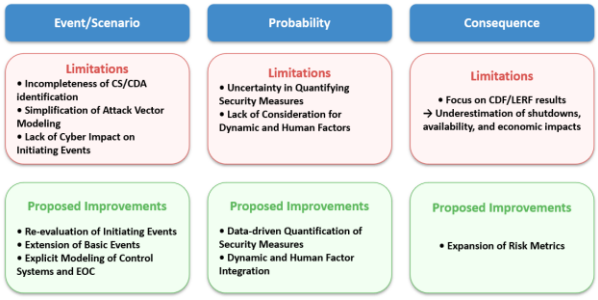

Fig. 1. Limitations and Improvements of Cybersecurity PSA Organized by the Risk Triplet.

5. Conclusions

This study reviewed PSA-based approaches for cybersecurity assessment in nuclear power plants, analyzed their limitations, and proposed directions for improvement. Although PSA has been extended to incorporate cyber threats, challenges remain in terms of CDA identification, attack vector modeling, quantification of security measures, consequence analysis, initiating event representation, data availability, and dynamic/human factors.

To address these gaps, this study suggests enhancing PSA models by including cyber-induced initiating and basic events, explicitly modeling control system vulnerabilities and EOC, broadening risk metrics beyond CDF/LERF, adopting data-driven quantification, and integrating dynamic and human reliability considerations.

In short, cybersecurity PSA should evolve into a practical risk management framework that supports both regulatory oversight and operational decision-making.

Acknowledgement

This work was supported by the Nuclear Safety Research Program through the Korea Foundation OfNuclear Safety(KoFONS), granted financial resource from the Nuclear Safety and Security Commission(NSSC), Republic of Korea. (No. 2106012)

REFERENCES

[1] Shin, J., Son, H., & Heo, G. (2017). Cyber security risk evaluation of a nuclear I&C using BN and ET. Nuclear Engineering and Technology, 49(3), 517-524.

[2] Shin-woo Lee, & Jung-hee Lee (2024). Improving the Efficiency of Cybersecurity Risk Analysis Methods for Nuclear Power Plant Control Systems. Journal of the Korea Institute of Information Security & Cryptology, 34(3), 537-552.

[3] Park, J. W., & Lee, S. J. (2019). Probabilistic safety assessment-based importance analysis of cyber-attacks on nuclear power plants. Nuclear Engineering and Technology, 51(1), 138-145.

[4] Son, K. S., Song, J. G., & Lee, J. W. (2023). Development of the framework for quantitative cyber risk assessment in nuclear facilities. Nuclear Engineering and Technology, 55(6), 2034-2046.

[5] Kwon, K., Kim, A., & Lee, S. A Comparative Study on Nuclear Power Plant Cyber Security Assessment Models Based on Risk Assessment Standard Guideline.

[6] World Institute for Nuclear Security. (2025). Assuring the effectiveness of cybersecurity at nuclear facilities [International Best Practice Guide].

[7] Kim, H. E., Son, H. S., Kim, J., & Kang, H. G. (2017). Systematic development of scenarios caused by cyber-attack-induced human errors in nuclear power plants. *Reliability Engineering and System Safety*, 167, 290–301.

[8] Vaddi, P. K., Zhao, Y., & Smidts, C. (2022). Dynamic Probabilistic Risk Assessment for Cyber Security Risk Analysis in Nuclear Reactors. In Proceedings of the Probabilistic Safety Assessment & Management Conference—PSAM (Vol. 16).