Gap Analysis of Probabilistic Safety Assessment for Risk-Informed Cyber Security of Nuclear Facilities

Minho Hwang, Joonseok Lim, Dayoung Jung, Gyunyoung Heo*
Department of Nuclear Engineering, Kyung Hee Univ., Deogyeong-daero, Giheung-gu, Yongin-si, Gyeonggi-do 17104, Republic of Korea
**Corresponding author: gheo@khu.ac.kr*

***Keywords : Cyber security, Probabilistic safety assessment, Graded approach, Facility lifecycle, Gap analysis**

1. Introduction

Cyber security measures for nuclear facilities should be differentiated according to risk rather than applied uniformly. Grading presupposes a measure of risk. Risk assessment is a basic principle of nuclear security regulation. IAEA guidance calls for computer security measures graded in proportion to the potential consequences of compromise[1]. U.S. NRC regulatory guidance endorses a consequence-based, graded approach as an acceptable method for risk-informed cyber security programs[2]. Korean regulation obligates licensees to classify critical digital assets (CDAs) into security grades and to apply measures by grade (NSSC Notice No. 2021-9, Annex 4, issued under the Act on Physical Protection and Radiological Emergency).

Because the grade is determined by the consequences of compromise within the facility rather than by the nature of the attack, this grading cannot be produced by examining cyber attacks alone. All three schemes above rank assets by severity, that is, by how much damage results when an asset is compromised and the safety, security, or emergency response function it supports is lost or malfunctions. The greatest damage comes from assets supporting safety functions. When such an asset is compromised, the effect of the lost or malfunctioning function propagates through facility systems. How far it propagates determines the magnitude of the damage. For conventional safety analysis, the principal end points of that propagation are fuel damage and radiological release. Tracing that propagation is the function of facility-level safety analysis models. Probabilistic safety assessment (PSA) is the established model of this kind. Its logic structure connects component states to core damage and release end states. Its minimal cut sets show which combinations of components can defeat safety functions. With an explicit mapping between CDAs and model events, this information can support the identification of safety-significant digital assets for grading. The mapping should account for basic events that may aggregate multiple components or failure modes.

Cyber security grading is already implemented at operating power reactors. In particular, for reactors with digital safety systems, previous studies have explored differentiating inspection and training according to the risk significance of individual digital assets. However, these studies have focused primarily on operating power reactors, while probabilistic methods have not yet been sufficiently incorporated into facility-level risk assessments for other nuclear facilities to support a comparable graded approach. Moreover, even at power reactors, existing PSA models are structured primarily for safety assessment of random failures and therefore require further development before they can support cyber security applications. This paper therefore examines whether risk models with adequate coverage exist for the facility types and lifecycle phases to be protected.

2. Risk Assessment Coverage across Facility Types and Lifecycle Phases

Regulatory documents, standards, and technical literature were reviewed from the IAEA, OECD/NEA, the U.S. NRC, U.S. DOE, EPRI, and the regulators of Finland, Canada, Japan, France, the U.K., and Korea. A methodology was counted regardless of its name, provided it assesses facility risk. Integrated safety analysis (ISA), safety case, and performance assessment (PA) were therefore included alongside PSA. A safety case is the assembled body of evidence, analysis, and argument supporting the claim that a facility is safe. Review, inspection & testing, and education & training were not treated as separate application areas; they are mentioned only as downstream examples of risk-informed grading. The shutdown / defueled column covers the whole transition between the end of operation and the start of decommissioning. Table 1 summarizes the result.

Four observations follow:

First, these methodologies are unevenly developed, whether read down the rows of Table 1 or across its columns. By facility type, quantification of accident sequences is furthest developed for power reactors, then spent fuel storage. Research reactors and fuel cycle facilities rely mainly on deterministic or semi-quantitative methods. Disposal facilities are quantified in a different mode altogether: long-term performance rather than accident sequences. By lifecycle phase, design is rated Established for all five facility types, whereas construction-specific methodologies are limited or absent. Table 1 therefore treats design/construction as a single phase group while

retaining construction-specific information within each cell. Operation is rated Established for all five facility types, whereas the shutdown transition is rated Partial in every facility type that has one. Decommissioning is rated Established for two types, both reactors. That rating covers deterministic assessment, with PSA-type quantification only at the pilot stage.

Second, full quantitative PSA producing risk metrics such as core damage frequency (CDF) is established regulatory practice only for power reactors, essentially in the combined design/construction phase and operation. Spent fuel storage PSA exists as a technical basis, but the governing requirements (10 CFR 72) remain deterministic. Elsewhere the established products differ in kind. ISA sorts each accident scenario into one of a few likelihood categories and designates items relied on for safety (IROFS). Deterministic assessment computes the radiation dose a person would receive in an assumed accident. PA yields, as its long-term performance measure, the annual dose to a representative person after a repository is closed.

Third, the risk profile changes fundamentally after operation. Once fuel is permanently removed from the core, core damage sequences no longer apply, and the dominant risk shifts to the spent fuel pool before diminishing with decay heat[3]. An unmodified operating-phase model therefore does not represent shutdown or decommissioning.

Fourth, disposal facilities reverse the usual order. At other facilities the operating phase is the most quantified. At disposal facilities the best-developed assessment is the post-closure PA, which runs to a million years. Post-closure is not an end-of-life phase. The sealed repository performs its containment and isolation function throughout it, so it is in effect the facility's operating period. It therefore falls outside the shutdown and decommissioning columns of Table 1, which are left blank for disposal. Accident-sequence quantification for the operational phase of Table 1 is not standard practice. Yet that phase is when digital systems run.

3. Gaps for Risk-Informed Graded Cyber Security

Table 1 shows whether an established risk assessment methodology exists for each of five facility types across four lifecycle-phase groups, with design and construction treated together. Table 2 shows, for the same five types, whether grading is required and whether cyber-side assessment means and safety-side accident-sequence models exist.

Functional and consequence assessments provide a basis for cyber security grading[1,2]. For the PSA-based approach considered here, the model must cover the relevant facilities and phases and support assessment of individual assets (Section 3.1). It must account for intentional compromise as well as random failure (Section 3.2). It must cover the end states and components relevant to the grading assessment (Section 3.3).

3.1 Scope required for grading versus scope where models exist

Grading is required across all five facility types (Table 2). Accident-sequence quantification, by contrast, is established regulatory practice in two of the eighteen cells that Table 1 rates: the power reactor design and construction group, and operation. The remaining cells yield likelihood categories, dose estimates, or long-term performance measures. Deterministic assessment can support consequence-based grading. A facility can be exempted, for example, when the consequences of a bounding accident stay below the applicable threshold. ISA identifies components as IROFS. Accident-sequence models provide information on combinations of modeled events leading to particular end states through their minimal cut sets.

Regulation abroad concentrates on power reactors in the same way, though not on the same element. In Korea the grading requirement covers all five facility types, and established quantitative PSA applications are concentrated in power reactors. In the United States the cyber security requirement is confined as well. The U.S. cyber security rule applies only to power reactor licensees, and independent spent fuel storage installations and U.S. research reactors have no cyber-specific requirement. For fuel cycle facilities, a dedicated cyber rule was proposed in 2017 (10 CFR 73.53). That proposal would have anchored consequence assessment to the performance-requirement thresholds of the existing ISA. It was withdrawn from Commission consideration in 2025 without taking effect. Outside the two established cells of Table 1, the scope of available accident-sequence models needs to be examined for PSA-based assessment of individual assets.

3.2 Differences from Safety and Security Point of View in PSA

Classical PSA presumes randomness. Its basic events traditionally represent stochastic, non-malevolent occurrences: equipment failures, human errors, and common cause failures. Their probabilities are estimated from operating experience. Intentional cyber attacks violate this statistical basis in three ways. The population of scenarios cannot be fully enumerated, the adversary adapts to defenses, and nuclear-sector cyber event data are too sparse for frequency estimation[4]. Regulatory documents draw the same line between random and intentional events. The Canadian regulatory document on PSA (REGDOC-2.4.2) confines external hazards within PSA to non-malevolent events. An NRC technical report on PSA scope (NUREG-2201) likewise places security-related scenarios outside PSA applications.

Practice has responded by avoiding probability quantification. The 2004 NRC self-assessment methodology used a susceptibility rating in place of probability, stating that the information needed to quantify attack probability was insufficient[5]. Korean practice similarly fixes the attack likelihood at 1 and grades by consequence alone. Attempts to couple attack models with PSA include attack trees joined to fault trees, Bayesian networks, and dynamic simulation with adversary models. All remain at the research stage, and no technique is established for industry use[4].

3.3 What current power reactor PSA does not represent

Power reactor operation is the best-covered area, yet even there current PSA models were not built to represent cyber compromise. The limitations appear at three levels.

The first is the risk metric. PSA quantifies core damage and large early release, but hostile objectives are not limited to these end states. Forced reactor trip, operational disruption, and the resulting economic loss can themselves be attack objectives. Metrics of the CDF or large early release frequency (LERF) type alone can underestimate the risk such attacks exploit. Risk metrics for security grading must therefore cover end states short of core damage as well.

The second is model scope. Initiating events induced by cyber compromise are not explicitly represented in conventional initiating event analysis, so attack-induced scenarios must be grafted onto models never designed for them. Components whose single failure or malfunction alone can shut the plant down are called single point vulnerabilities (SPVs). They are natural targets of shutdown-oriented attacks, yet PSA does not usually model them. An SPV causes a reactor trip. A trip is the entry point of a PSA model, not one of its end states. PSA represents the trip as an initiating event. Its frequency comes from operating experience, not from a model of the components that cause it. SPVs are mostly non-safety-class, and the PSA model covers mainly safety-class equipment. Identification of CDAs is likewise incomplete when checked against the PSA model. Digital equipment also tends to fail in unexpected ways, which leaves the enumerated set of failure modes incomplete[4].

The third is quantification. Section 3.2 outlined the difficulty of estimating attack probabilities for quantifying attack-induced accident frequencies. Consequence-based grading instead evaluates the effects of postulated compromise.

4. Conclusions and Future Work

Cyber security grading is required at every type of nuclear facility under Korean regulation. Quantitative accident-sequence modelling is established regulatory practice in only two cells of Table 1: the power reactor design and construction group, and operation. Elsewhere, PSA-based assessment of individual assets requires further examination of model scope and the links between digital assets and accident sequences. Even in the two established cells the model was built for random failures and for core damage and release end states. It may therefore require further development for cyber security applications.

Work proceeds in three steps.

First, each cell of Table 1 is matched to the PSA or risk assessment cases that already exist internationally. They include Level 1 PSA at HANARO and other research reactors, as well as the pilot dry cask PSA (NUREG-1864).

Second, domestic development follows those cases cell by cell. The models should provide the detail needed for the proposed PSA-based assessment, including explicit links between CDAs and the relevant model events.

Third, the power reactor model is extended for cyber security use. The extension adds cyber-induced initiating events, single point vulnerabilities, and end states short of core damage. It also maps the CDA list onto the PSA logic model. Because attack probability cannot yet be estimated, the ranking remains consequence-based.

ACKNOWLEDGMENT

This work was supported by the Nuclear Safety Research Program through the Korea Foundation of Nuclear Safety(KoFONS) using the financial resource granted by the Nuclear Safety and Security Commission(NSSC) of the Republic of Korea.(RS-2026-25537603)

REFERENCES

[1] IAEA, Computer Security Techniques for Nuclear Facilities, IAEA Nuclear Security Series No. 17-T Rev. 1, IAEA, Vienna, 2021.
[2] U.S. NRC, Cybersecurity Programs for Nuclear Power Reactors, Regulatory Guide 5.71 Rev. 1, 2023.
[3] U.S. NRC, Technical Study of Spent Fuel Pool Accident Risk at Decommissioning Nuclear Power Plants, NUREG-1738, 2001.
[4] S. Eggers, K. Le Blanc, Survey of cyber risk analysis techniques for use in the nuclear industry, Progress in Nuclear Energy, Vol. 140, 103908, 2021.
[5] U.S. NRC, Cyber Security Self-Assessment Method for U.S. Nuclear Power Plants, NUREG/CR-6847, 2004.

Table 1. Established risk assessment methodologies by facility type and lifecycle phase (a)

Facility type (b)	Design/construction	Operation	Shutdown / defueled	Decommissioning
Nuclear power plant	Established (design-stage PSA updated and used during construction)	Established (full-power, LPSD, and SFP PSA)	Partial (case studies; no consensus standard)	Established (deterministic; PSA pilots)
Research reactor	Established (deterministic design analysis with supplementary PSA; SAR updated during construction)	Established (deterministic SAR with supplementary PSA cases)	Partial	Established (dose-based deterministic)
Fuel cycle facility	Established (design-stage ISA; no construction-specific methodology identified)	Established (ISA maintenance; limited-scope PSA cases)	Partial	Partial (dose-based)
Spent fuel storage	Established (design-stage safety case and cask PSA; no construction-specific methodology identified)	Established (pool PSA; dry storage pilot PSA)	Partial (U.S. technical basis, NUREG-1738; Korean review deterministic)	Partial
Waste disposal (c)	Established (safety case / PA, updated during construction)	Established (safety case; no accident-sequence quantification)		

(a) Established: backing (a consensus standard, guidance, or regulatory requirement) and applications both found. Partial: only one of the two, or no methodology of its own, the preceding phase's model or SAR carried forward. Commissioning is included in the combined design-and-construction phase. LPSD: low power and shutdown; SFP: spent fuel pool; SAR: safety analysis report. Blank: the facility type has no phase corresponding to that column.

(b) Representative bases. Nuclear power plant: IAEA SSG-3/SSG-4 Rev.1, NSSC Notice No. 2017-34, IAEA GSR Part 6. Research reactor: IAEA SSR-3, NUREG-1537, IAEA SRS-77. Fuel cycle facility: 10 CFR 70.62/70.65, NUREG-1520 Rev.2, IAEA SSR-4. Spent fuel storage: IAEA SSG-15 Rev.1, NUREG-2161, NUREG-1864. Waste disposal: IAEA SSR-5, IAEA SSG-23, NUREG-1573. The ratings are based on all the regimes reviewed. Only the power reactor row lists a Korean requirement.

(c) The disposal lifecycle is pre-operational, operational, and post-closure. Post-closure is not an end-of-life phase. The sealed repository performs its containment and isolation function throughout it, and in that sense post-closure is the facility's true operating period. It has no counterpart in the shutdown or decommissioning columns, which are therefore left blank. Post-closure performance assessment is discussed in the text.

Table 2. Cyber security grading requirement and assessment means by facility type (d)

Facility type	Cyber-side assessment means	Safety-side sequence model (Table 1)
Nuclear power plant	Consequence-based qualitative grading established for operation; quantitative methods at research stage	Established, design/construction and operation
Research reactor	Research-stage cases only; no established methodology	Supplementary cases only
Fuel cycle facility	No facility-specific methodology identified	Limited-scope cases only
Spent fuel storage	No facility-specific methodology identified	Technical basis, not a requirement
Waste disposal	No methodology or application identified	None; PA is not sequence-based

(d) Grading is required for every facility type under the Korean definition of nuclear facilities (NSSC Notice No. 2021-9). Cyber-side methodology concentrates on operation; none was identified for construction, shutdown, or decommissioning. Where no facility-specific methodology exists, IAEA's general graded framework still applies. The safety-side column records accident-sequence quantification, not the verdicts of Table 1, whose established ratings include deterministic methods. Domestic implementation at storage and disposal facilities was not confirmed in open literature.