

Threat Level-Based Quantitative Evaluation of Cyber Incident Response Capabilities in Nuclear Facility Exercises

Dong Jun Choi^a, Jung Taek Seo^{b*}

^aDepartment of Information Security, Gachon University, Seongnam-si, Republic of Korea

^bDepartment of Smart Security, Gachon University, Seongnam-si, Republic of Korea

*Corresponding author: seojt@gachon.ac.kr

***Keywords :** nuclear facility, cyber incident response, quantitative evaluation, response exercise, threat level

1. Introduction

Nuclear facilities are national critical infrastructure in which a cyberattack can escalate beyond information leakage into physical consequences such as sabotage or a radioactive release, as the 2010 Stuxnet incident demonstrated for industrial control systems [1]. Because security controls alone cannot prevent every intrusion, the operator's capability to rapidly detect and respond to an incident serves as the last line of defense, and this capability is examined through cyber incident response exercises [2]. Existing evaluations, however, simply enumerate stage-wise checklist items and depend on the evaluator's qualitative judgment, so the results of the same exercise vary among evaluators and insufficient capabilities are difficult to identify objectively. Without quantitative criteria it is also hard to track improvement over repeated exercises or to decide which capabilities to reinforce first under limited resources. A prior study established a framework of response-capability requirements and evaluation criteria [3], but a methodology that uses the framework to quantify the capability level actually demonstrated in an exercise has not yet been presented.

To fill this gap, this paper proposes a methodology that quantitatively evaluates the response capability demonstrated in cyber incident response exercises at nuclear facilities. The methodology derives the scenario threat level from six standards-based risk factors, adopts it as the target level of response capability together with priority weights assigned by incident type, and compares normalized requirement scores against the target to identify the unmet requirements and the overall response-capability level. Applying it to two attack scenarios realized on a nuclear power plant simulator, we identify insufficient capabilities at the requirement level and confirm that the capabilities requiring reinforcement differ by incident type.

2. Background and Related Work

2.1 Cyber Incident Response at Nuclear Facilities

As the digitalization of instrumentation and control systems advances, the cyberattack surface of nuclear facilities has widened, and assets that perform or affect safety, security, and emergency-preparedness functions

are managed as critical digital assets. Following the defense-in-depth principle, these assets are graded into cybersecurity levels and isolated by unidirectional communication from higher to lower levels and by boundary protection, a scheme required by regulation [4].

Consequently, cyber incident response at nuclear facilities differs from that in general information systems in three ways. First, an incident may propagate beyond information leakage or service disruption to physical safety consequences such as a reactor trip or a radioactive release. Second, because assets are protected differentially by security level according to their safety importance, the severity of an incident varies with the targeted asset's level. Third, incident detection, notification, and response must occur within regulatory time limits [5]. Quantitative scoring methods from outside the nuclear domain that ignore security-level grading or regulatory deadlines therefore cannot be applied directly.

2.2 Response-Capability Evaluation Framework

Following NIST SP 800-61 [6] and the IAEA incident response guidance [2], cyber incident response is divided into six stages: preparation, identification and detection, containment, eradication, recovery, and post-incident activities. The prior framework [3] evaluates each stage against the performance objectives of timeliness, consistency, effectiveness, and expertise, and expresses response capability as 19 requirements, each consisting of one to six evaluation criteria judged by whether their performance indicators are satisfied. Because the number of criteria per requirement is uneven, simply summing satisfied criteria would not fairly compare requirements; the proposed methodology corrects for this with a normalized score.

2.3 Related Work

Ahn et al. [7] proposed scenario graphs that analyze attack paths of nuclear power plants, but the resulting risk is not connected to response-capability evaluation. Park and Lee [8] quantified the difficulty and impact of attack scenarios using Bayesian belief networks, and Son et al. [9] quantified cyber risk through security controls; both address the risk of the facility itself, not

the operator's response capability. Maturity models such as the ENISA CSIRT maturity framework [10] assess organizational response maturity, but they evaluate a standing capability independent of any scenario or threat level and are thus unsuitable for evaluating an individual exercise. Existing work therefore remains at quantifying threat risk or organization-level posture, and no methodology links the scenario threat level to the operator's response level for quantitative evaluation.

3. Proposed Methodology

The proposed methodology consists of five steps: it derives the scenario threat level, sets it as the target level, assigns stage-wise priority weights by incident type, computes normalized scores, and derives the unmet requirements and the overall response-capability level.

3.1 Scenario Threat-Level Derivation

The scenario threat level is derived by scoring each of the six risk factors in Table I from 0 to 10, assigning a higher score when the attack is easier and its impact is greater: a more remote attack surface (L_1), a higher CVSS severity (L_2), less required privilege (L_3) or user interaction (L_4), a higher-graded target asset (L_5 , e.g., the control and protection systems at level 4), and a shorter attack path (L_6). Because these criteria are defined on a standard basis, the scheme reduces score variance across evaluators for the same attack conditions.

Table I: Threat-level scoring factors

ID	Factor	Scoring criteria
L1	Attack surface	Internet 10, Wireless 8, Supply chain 6, Local 4, Physical 2
L2	Vulnerability severity	CVSS 9.0-10: 10, 7.0-8.9: 8, 5.0-6.9: 6, 1.0-4.9: 4, below: 2
L3	Required privilege	None 10, Low 6, High 2
L4	User interaction	None 10, Operations network 6, Control network 2
L5	Asset criticality	Security level 4: 10, 3: 8, 2: 6, 1: 4, 0: 2
L6	Attack path length	1 asset 10, 2 assets 8, 3 assets 6, 4 assets 4, 5 or more assets 2

The threat level is the arithmetic mean of the six factor scores on a 0-10 scale, as in Eq. (1), where L_j is the score of risk factor j and the six factors are reflected equally.

$$O_1 = (1/6) \sum_{j=1..6} L_j, \quad L_j \in [0, 10] \quad (1)$$

3.2 Target Level and Type-Specific Priority Weights

The operator's target level R_i must be at least the scenario threat level, since the threat level is the minimum response level the scenario requires. Next, the

response stages and performance objectives to be prioritized are assigned by incident type, as in Table II. Among the four performance objectives, expertise is assigned to the preparation stage and consistency to the post-incident stage, while the identification-and-detection, containment, eradication, and recovery stages, which directly respond to an ongoing incident, are evaluated against timeliness and effectiveness. The facility-sabotage type prioritizes the identification-and-detection and containment stages because it is important to block damage before it spreads physically, whereas the facility-security and system-failure type prioritizes the identification-and-detection and eradication-and-recovery stages because restoring the impaired function is important.

Table II: Prioritized stages and performance objectives by scenario type

Scenario type	Prioritized stages	Performance objectives
Facility sabotage	Identification and detection, Containment	Timeliness, Effectiveness
Facility security, System failure	Identification and detection, Eradication and recovery	Timeliness, Effectiveness

A weight W_i is then added to the requirements of the prioritized stages, adjusting the target level to R'_i as in Eq. (2). The weight raises the pass threshold by a fixed amount on the same point scale as the target level rather than as a multiplier, and is assigned in three tiers so that a higher target level receives a larger weight: 0.1-0.2 for a low target level (1-3), 0.2-0.3 for a medium level (4-6), and 0.3-0.4 for a high level (7-9). Since an excessively large weight would let the result be dominated by it, the evaluator adjusts it within the recommended range according to the scenario's intent.

$$R'_i = R_i + W_i \quad (2)$$

3.3 Normalized Scoring and Result Derivation

The response score of each requirement is computed from whether its evaluation criteria are satisfied. Because the number of criteria per requirement differs, the full mark of 10 is divided by the number of criteria n_i , assigning an equal score to each satisfied criterion; the normalized score S_i of requirement i is given by Eq. (3), where c_i is the number of satisfied criteria. Comparing S_i with the adjusted target level R'_i , a requirement is judged met if S_i is at least R'_i and unmet otherwise, and an unmet requirement indicates a response capability the operator must reinforce. The overall response-capability level for the scenario is the average of the 19 normalized scores. The stage-wise priority weights act only on the per-requirement verdict, not on the overall average, so the stage whose capability must be reinforced is identified separately from the aggregate level.

$$S_i = (c_i / n_i) \times 10 \quad (3)$$

4. Scenario-Based Application

The methodology was applied illustratively to two scenarios realized on the Compact Nuclear Simulator (CNS) of the Korea Atomic Energy Research Institute, which models a Westinghouse three-loop 900 MW pressurized water reactor [11]. The simulator confirms each attack's progression and physical effects, and the methodology then evaluates the operator's response to it. Both scenarios yield a threat level of 5.0, so the target level R_i is 5.

4.1 Scenario A: Pressurizer Spray Valve Opening

Scenario A is a facility-sabotage type in which opening the pressurizer spray valve causes a loss of reactor cooling capability. After infecting the engineering workstation via a USB device containing malware, the attacker abuses remote services to alter the control logic and opens the spray valve; as spraying lowers the pressurizer pressure below the low-pressure reactor-trip setpoint, the reactor protection system actuates and the reactor trips. Applying Table I, the six risk factors score 2, 8, 2, 2, 10, and 6 in order, reflecting physical access via a USB device, the high severity of the exploited remote-service vulnerability, a high-privilege requirement, control-network interaction, a security level-4 asset, and three attack paths, so the threat level is $(2+8+2+2+10+6)/6 = 5.0$. As a facility-sabotage type, the identification-and-detection and containment stages are prioritized per Table II, and, since a safety system is directly targeted, a weight of 0.3 (the upper end of the medium tier) is assigned, raising the pass threshold of the prioritized stages to 5.3.

4.2 Scenario B: Neutralization of Access Control and IP CCTV

Scenario B is a facility-security and system-failure type in which the access-control device and IP CCTV are neutralized. The attacker clones the signal of an RFID access card to pass access control, reaches the operations network through a network port exposed on the outer wall, and abuses a remote-code-execution vulnerability of the IP CCTV (CVSS 9.8) to stop the video-surveillance function. This scenario has the highest vulnerability severity but requires substantial prior access, and the final target is the video-surveillance system rather than a safety system, so its asset criticality is relatively low. The six factors score 2, 10, 2, 6, 4, and 6 in order, so the threat level is again 5.0. As a facility-security and system-failure type, the identification-and-detection and eradication-and-recovery stages are prioritized, with a weight of 0.2 raising their pass threshold to 5.2.

4.3 Result Analysis

Observing each exercise, the evaluator judges every criterion as met or unmet by whether it satisfies the performance indicator defined in the prior framework [3]; for example, detection timeliness (D5) has a single criterion that is met if the operator detects the incident within the prescribed time. The per-requirement target levels, normalized scores, and verdicts of both scenarios are summarized in Table III, and the overall results in Table IV.

Table III: Per-requirement target level, normalized score, and verdict for the two scenarios

Req.	A target	A score	A verdict	B target	B score	B verdict
P1	5.0	10.00	Met	5.0	3.33	Unmet
P2	5.0	10.00	Met	5.0	10.00	Met
D1	5.3	8.33	Met	5.2	5.00	Unmet
D2	5.3	5.00	Unmet	5.2	10.00	Met
D3	5.3	5.00	Unmet	5.2	5.00	Unmet
D4	5.3	10.00	Met	5.2	10.00	Met
D5	5.3	0.00	Unmet	5.2	10.00	Met
D6	5.3	10.00	Met	5.2	10.00	Met
C1	5.3	6.67	Met	5.0	6.67	Met
C2	5.3	10.00	Met	5.0	5.00	Met
E1	5.0	6.67	Met	5.2	6.67	Met
E2	5.0	10.00	Met	5.2	10.00	Met
E3	5.0	10.00	Met	5.2	5.00	Unmet
E4	5.0	10.00	Met	5.2	10.00	Met
R1	5.0	10.00	Met	5.2	10.00	Met
R2	5.0	10.00	Met	5.2	3.33	Unmet
I1	5.0	7.50	Met	5.0	7.50	Met
I2	5.0	5.00	Met	5.0	5.00	Met
I3	5.0	10.00	Met	5.0	10.00	Met

Table IV: Summary of evaluation results

Item	Scenario A	Scenario B
Incident type	Facility sabotage	Facility security, system failure
Threat / target level	5.0 / 5	5.0 / 5
Priority weight	0.3	0.2
Weighted target of prioritized stages	5.3	5.2
Overall response level	8.11	7.50
Unmet requirements	D2, D3, D5	P1, D1, D3, E3, R2

In Scenario A the overall response-capability level, the average of the 19 normalized scores, is 8.11. Although this far exceeds the target level, the unmet requirements are D2, D3, and D5. Detection timeliness (D5) scores zero because the incident was not detected within the prescribed time [5], and the co-unmet D2 (infection scope and propagation path) and D3 (impact on essential systems) indicate that the operator did not adequately grasp how far the attack had spread or which systems were affected; as all three belong to the identification-and-detection stage, early recognition and impact-scope assessment are the priority reinforcement targets. In Scenario B the overall level is 7.50 and the unmet requirements are P1, D1, D3, E3, and R2: P1

reflects a lack of the basic capability required before the exercise, D1 and D3 an insufficient grasp of the incident's cause and scope, and E3 (patch application) and R2 (recovery verification) shortfalls that make the eradication-and-recovery stage the priority reinforcement target. Note that D2 in Scenario A, scoring 5.00, would pass the unweighted target of 5.0 but is judged unmet under the weighted threshold of 5.3.

To confirm the validity of the type-specific prioritization, the priority weights of the two types were each applied to the same Scenario B scores: under the facility-sabotage type the unmet set includes containment (C2), whereas under the facility-security and system-failure type it includes eradication (E3). The requirements to be reinforced first thus differ by scenario type even for the same response performance, which a single aggregate score cannot reveal.

5. Discussion

Methods that quantitatively score individual incidents or exercises exist outside the nuclear domain: the National Cyber Incident Scoring System (NCISS) of CISA rates incident severity on a weighted mean of eight categories [12], the Locked Shields scoring system of NATO scores team performance in cyber defense exercises [13], and Bianchi et al. automatically score a blue team's response by comparing attack-defense graphs [14]. As all of these use weighting and normalization, they share an approach with the proposed methodology, and their quantitative scoring schemes are compared in Table V.

Table V: Comparison of existing quantitative scoring methods and the proposed methodology

Method	Scoring scheme	Measured object	Limitation for nuclear use
NCISS [12]	Weighted mean of 8 categories, normalized to 0-100	Incident severity	Not a response-performance evaluation; safety and regulation not reflected
Locked Shields [13]	5 weighted categories, S-curve normalization	Team defense performance	Fixed weights; threat and security level not reflected
Bianchi et al. [14]	Scores from attack-defense graph comparison	Blue-team response	No type-specific priority or safety characteristics
This work	Normalized scores with threat-level linkage and type-specific weights	Operator response performance	-

Compared with the proposed methodology, these methods fall short in three respects. First, their pass criteria and weights are fixed, so the target level does not change with the scenario threat level. Second, they

cannot evaluate differently by incident type via priority weights, so they always apply the same criterion value. Third, most produce a single aggregate score and cannot identify which stage and requirement are deficient. Indeed, as shown in Section 4, the proposed methodology flags the detection-timeliness requirement (D5) even when the overall average exceeds the pass line and, by scenario type, distinguishes the reinforcement target into containment (C2) and eradication (E3), a detailed result that single-score methods cannot readily provide. More fundamentally, existing methods do not reflect the characteristics of nuclear facilities described in Section 2.1: the proposed methodology reflects the differential protection by security level in risk factor L_5 , so that targeting a safety system yields a higher threat and target level, and assigns priority weights by incident type to stages such as containment that block the physical spread of damage.

The proposed methodology defines the threat and response levels through explicit formulas, improving reproducibility, and reflects the scenario type and security level in the evaluation. However, this study is limited to an illustrative application to two scenarios, the assignment of the priority weight W_i involves evaluator judgment, the threat level averages the six risk factors equally, and a requirement with a single criterion yields a score of only 0 or 10, limiting precise comparison.

6. Conclusion

This paper proposed a methodology that quantitatively evaluates the cyber incident response capability demonstrated in nuclear facility exercises by combining threat-level derivation from six risk factors, incident-type priority weights, and normalized scoring. Applying it to two simulator-based scenarios identified insufficient capabilities at the requirement level and confirmed that the reinforcement targets differ by incident type, and comparison with existing IT-domain scoring methods showed that the methodology reflects nuclear-specific characteristics such as type-specific prioritization and security-level grading. Future work will validate the methodology on real exercise data covering more incident types, systematize factor-specific weighting of the threat level, and subdivide single-criterion requirements for finer comparison.

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-2021-KN051410).

REFERENCES

- [1] R. Langner, Stuxnet: Dissecting a Cyberwarfare Weapon, *IEEE Security & Privacy*, Vol. 9, No. 3, pp. 18-22, 2011.
- [2] International Atomic Energy Agency, Computer Security Incident Response Planning at Nuclear Facilities, IAEA-TDL-005, Vienna, 2016.
- [3] H. A. Choi, C. Park, J. Lee, S. Jeon, and J. T. Seo, Framework for Evaluating Cyber Incident Response Capabilities of Nuclear Facility Operators through Operation-Based Exercises, *Nuclear Engineering and Technology*, Vol. 57, No. 11, 103772, 2025.
- [4] U.S. Nuclear Regulatory Commission, Cyber Security Programs for Nuclear Power Reactors, Regulatory Guide 5.71, 2023.
- [5] U.S. Nuclear Regulatory Commission, Cyber Security Event Notifications, Regulatory Guide 5.83, Rev. 1, 2023.
- [6] National Institute of Standards and Technology, Computer Security Incident Handling Guide, NIST SP 800-61 Rev. 2, 2012.
- [7] W. Ahn, M. Chung, B. Min, and J. Seo, Development of Cyber-Attack Scenarios for Nuclear Power Plants Using Scenario Graphs, *International Journal of Distributed Sensor Networks*, Vol. 11, 836258, 2015.
- [8] J. W. Park and S. J. Lee, A Quantitative Assessment Framework for Cyber-Attack Scenarios on Nuclear Power Plants Using Relative Difficulty and Consequence, *Annals of Nuclear Energy*, Vol. 142, 107432, 2020.
- [9] K. S. Son, J. G. Song, and J. W. Lee, Development of the Framework for Quantitative Cyber Risk Assessment in Nuclear Facilities, *Nuclear Engineering and Technology*, Vol. 55, No. 6, pp. 2034-2046, 2023.
- [10] European Union Agency for Cybersecurity, ENISA CSIRT Maturity Framework, 2022.
- [11] Y. H. Chae, C. Lee, S. M. Han, and P. H. Seong, Graph Neural Network Based Multiple Accident Diagnosis in Nuclear Power Plants, *Nuclear Engineering and Technology*, Vol. 54, No. 8, pp. 2859-2870, 2022.
- [12] Cybersecurity and Infrastructure Security Agency, National Cyber Incident Scoring System (NCISS), CISA, 2023.
- [13] S. Māses, K. Maennel, and A. Brilingaitė, Trends and Challenges for Balanced Scoring in Cybersecurity Exercises: A Case Study on the Example of Locked Shields, *Frontiers in Education*, Vol. 7, 958405, 2022.
- [14] F. Bianchi, E. Bassetti, and A. Spognardi, Scalable and Automated Evaluation of Blue Team Cyber Posture in Cyber Ranges, *Proceedings of the 39th ACM/SIGAPP Symposium on Applied Computing (SAC '24)*, pp. 1539-1541, 2024.