Development of a Quantitative Method for Evaluating Security Controls Based on Intrusion Tolerant Concept: Consideration of Adverse Effects

Chanyoung LEE[1], Poong Hyun SEONG[2]

1. Department of Nuclear and Quantum Engineering, KAIST, Daejeon, Korea (Tel: +82-42-350-5860, E-mail: lcy5228@kaist.ac.kr)

2. Department of Nuclear and Quantum Engineering, KAIST, Daejeon, Korea (Tel: +82-42-350-3820, E-mail: phseong@kaist.ac.kr)

Abstract: The introduction of digital techniques has brought up a new issue of cyber security; concerns are continuously growing in the nuclear industry. Considering that security techniques to be applied were not considered together when I&C systems were developed, it is necessary to analyze not only security enhancement, but also the influence of applied security techniques on the target system. In this study, a quantitative method for performing such evaluating is developed. The method can provide information including the degree to which security level can be improved and how the reliability of existing systems will be affected. In addition, a two-dimensional index called the *Hybrid Security Index (HSI)*, which includes an *Incompatibility Index* and a *Performance Index* is developed. Validity of the suggested method was proven by conducting a case study.

Keyword: Security Technique, Hybrid Security Index, Incompatibility Index, Performance Index

1 Introduction

Digital Instrumentation and Control (DI&C) systems have been developed and installed in Nuclear Power Plants (NPPs). This introduction of DI&C has brought up a new issue of cyber security; concerns about this system are continuously growing in the nuclear industry. Actually, NPP DI&C systems are physically isolated from external networks, and thus NPPs are regarded as stand-alone systems that are safe from external cyber-attack. Consequently, cyber security has received less attention than other safety problems have. However, continuous cyber-attack attempts against NPPs signify that NPPs are as susceptible to cyber-attack as is other safety critical infrastructure, and so public perception of cyber security for NPPs has been changing [1].

The *Chatham House Report* investigated the range of cyber security challenges at nuclear facilities [2]. According to this report, one of the major problems in the nuclear industry is that the industry is in a very early stage of dealing with cyber security issues. The main reason why this is the case is that cyber security has received less attention than other safety problems have. In addition, late adoption of DI&C systems has resulted in a level of cyber security advancements in the nuclear industry lower than those in other industries. Also, limited incident disclosure and unclear collaboration among related corporations have made it difficult to assess the true extent of cyber security problems.

In order to provide useful information on cyber security issues in the nuclear industry, several regulatory documents such as RG 5.71 and RS-015 were published. These documents include cyber security plans and comprehensive sets of security techniques. However, there are still difficulties when it comes to applying regulatory guidelines [3] because practical examples for the application of such guidelines have not been available to system designers, and methods that can help assess the degree to which security is improved if a specific technique is applied are not included.

For these reasons, quantitative methods of evaluating security techniques should be developed to help system designers understand security techniques and arrive at optimal design

solutions. Several methods have been proposed to evaluate security improvements when security techniques are introduced into the I&C system [4]. However, considering that the security techniques to be applied were not considered when the existing I & C systems were developed, it is necessary to analyze not only security enhancement, but also the influence of the applied security technique on the target system.

In this study, an evaluation method that can provide information including the degree to which the security level can be improved and how the reliability of existing systems is affected, is developed. The method focuses on providing meaningful quantitative indicators that can be used to arrive at an optimal solution considering both safety and security aspects.

2 Development of HSI

It should be possible to perform security techniques in harmony with the operation of the existing system. In order to select appropriate security techniques, it is important to analyze not only the degree to which the security level is improved by application of security techniques, but also the extent to which the security technique affects the reliability of the existing system. In addition, it is necessary to consider the above two factors and provide them to the system in a quantitative manner so that system designers can make an optimal decision. In this study, we developed a two - dimensional index called the *Hybrid Security Index* (*HSI*). *HSI* includes an *Incompatibility Index* and a *Performance Index*.

HSI = (*Incompatibility Index*, *Performance Index*)

The *Incompatibility Index* is a measure of the extent to which the reliability of the existing system is affected by security techniques. It can be defined as the change of the failure rate of the target system or the rate of change of the core damage frequency (CDF) in terms of PSA. The *Performance Index* is defined as the rate of increase in percentage of errors in the safe-state among errors caused by an intended fault. In safe-state, availability of safety functions are protected by applied security techniques in an intrusion situation. Details of the quantifying processes for each index are explained in chapter 2.1 and 2.2.

2.1 Estimation of *Incompatibility Index*

The definition of *Incompatibility Index* is based on the concept that security techniques should not affect the reliability of existing systems. There are roughly two ways in which security techniques can cause system failures. First, failures of security techniques may directly lead to system failure. However, the portion of system failures due to security technique failures is negligibly small because security technique failure itself cannot invoke physical problems. The second way is an indirect way in which the operation of the security techniques adversely affects system operation. The introduction of security techniques can complicate the system in terms of system structure and data communication, and the complexity of the system can lead to errors such as network congestion, time delay, and data loss. In addition, these errors can lead to increases in the software failure probability. In this study, among the two mentioned methods, the ways in which security techniques affect system reliability are restricted only to the indirect path.

In the proposed quantification method, using the metrics of system complexity and verification and validation (V&V) level, indirect evidence is applied to estimate the software failure probabilities [5]. For this, the safety integrity level (SIL) is used as an estimator of the V&V process. The relationships between SIL and software failure probability are summarized in Table 1 and Table 2.

Table 1 Standard for SIL

Consequence	Frequency of error occurrence in software			
	Reasonable	Probable	Occasional	Infrequent
Catastrophic	4	4	4 or 3	3
Critical	4	4 or 3	3	2 or 1
Marginal	3	3 or 2	2 or 1	1
Negligible	2	2 or 1	1	1

Table 2 Baseline failure probability estimates

for application software modules

SIL	Complexity of software		
	High	Medium	Low
0	1.0E-01	1.0E-02	1.0E-03
1	1.0E-02	1.0E-03	1.0E-04
2	1.0E-03	1.0E-04	1.0E-05
3	1.0E-04	1.0E-05	1.0E-06
4	1.0E-05	1.0E-06	1.0E-07

With regard to SIL class of software implemented in the NPP safety systems, the frequency of error occurrence in the safety systems of NPPs can be considered to be infrequent, and the consequence of software failure can be considered to be critical. The software implemented safety system exhibits low complexity because it is focused on the activation of safety-critical functions [6], [7]. However, the complexity level can be increased to medium or high level when the safety system includes security techniques. Factors affecting the increase in software complexity can be the number of applied security technologies, the range of detectable coverage, and the inspection period. An increase in the software failure probability also increases the failure rate of the associated software module implemented in system.

To calculate the *Incompatibility Index* for the target system, software modules that affect the failure of the target system are identified through fault tree analysis. The relationship between the failure of a software module and software failure probability should be analyzed. In addition, the software failure probability is allocated according to the assigned complexity level. Increased failure rates of software modules can change the failure rate of the target system. Furthermore, failures of the safety system may increase the core damage frequency (CDF), which is used in probabilistic safety assessment (PSA).

The degree of influence on system reliability of the application of security techniques can be quantified as the rate of change of the CDF value. Based on the newly calculated CDF and the previous value of the CDF, the *Incompatibility Index* can be defined as the rate of change due to the application of security techniques.

$$Incompatibility\ Index = \frac{CDF' - CDF}{CDF} \quad (1)$$

where CDF' is the newly calculated CDF.

2.2 Estimation of *Performance Index*

The nuclear industry is a safety critical industry that considers the availability of safety functions as the most important priority. In this regard, the *Performance Index* is defined as the rate of increase in percentage of errors in the safe-state among errors caused by an intended fault. In safe-state, availability of safety functions are protected by applied security techniques in an intrusion situation. Rather than the scenario in which an attacker infiltrates malicious codes directly into multi-protected operating NPPs, the scenario used in this study is limited to the following situation. In the scenario, an attacker inserts a latent digital fault into the target system in the software development life cycle (SDLC) environment, and the faults cause intended digital errors during operation. Many security experts also say that access to the SDLC environment is more possible, than that to NPPs, which are physically isolated and strictly controlled [1].

As an abstraction, a set of predicate block diagrams is used. Fig. 1 shows an example of a fault-tolerance predicate block diagram. The predicate diagram can be summarized as follows [8].

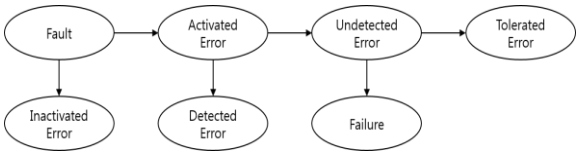

Fig.1 Fault-tolerance predicate block diagram

(1) *Inactivated Error*: A fault cannot be activated as an error if the faulty location is not read by the specific input.
(2) *Detected Error*: An error can be detected by a certain error detection method.

(3) *Tolerated Error*: If a fault is activated as an error, but is not detected, and the procedure output is correct, then it is a tolerated error.
(4) *Failure*: The parser processes its input and assigns a wrong value to the output; no error is detected.

The predicate block diagram has been used to classify experimental errors caused by digital faults. Using the diagram, the percentage of detected errors caused by fault tolerant techniques (FTTs) and the percentage of errors causing system failure can be estimated. FTT is a technique that can detect faults and make the system generate a fail-safe signal. However, in this study, faults that can be detected in FTTs, such as stuck-1 or 0 errors, are excluded. Only intelligently intended faults are assumed. Because of this assumption, the diagram can only verify the performance of the security technique by excluding the duplicated coverage with FTTs.

In addition, the process of activating the fail-safe signal generation (FSSG) after detecting errors is omitted in the fault-tolerate predicate block diagram, as shown in Fig. 1. However, there have been studies in which the activating FSSG process has been analyzed together because the process can fail due to human error in manual operation [9]. In addition, there was one study in which the probability of human error in the activation process was found to increase in an intrusion situation [10]. Based on these studies, the block diagram has been modified by adding a mitigated error node, as shown in Fig. 2.

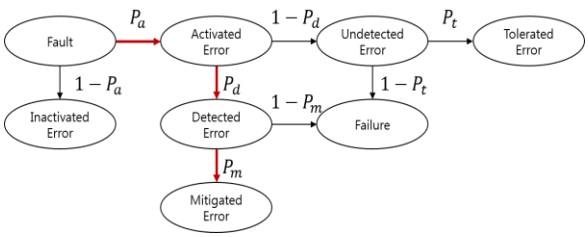

Fig.2 Intrusion-tolerance predicate block diagram

(5) *Mitigated Error*: If a detected error is mitigated safely by activating FSSG with success of manual operation, then it is a mitigated error.

where P_a is the percentage of activated errors among errors caused by an intended fault, P_d is the percentage of detected errors among the activated errors, P_t is the percentage of tolerated errors among the undetected errors, and P_m is the percentage of mitigated errors among the detected errors. In this study, processes that can be improved by application of security techniques are limited to three kinds of processes, shown as red arrows in Fig. 2.

Safe-state and unsafe-state are determined whether or not the target system can perform the minima functions or can be safely mitigated. In the diagram, the unsafe-state is only when errors caused by an intended fault are classified into the *Failure* state. Therefore, the percentage of safe-state, P_s, is the percentage of errors classified into categories of *Inactivated Error*, *Mitigated Error*, and *Tolerated Error*. P_s can be estimated according to the following equation.

$$P_s = (1 - P_a) + P_a P_d P_m + P_a (1 - P_d) P_t \quad (2)$$

The percentage of each branch and percentage of the safe-state depend not only on the performance of the security techniques, but also on the conditions of the target system. In more detail, the percentages may depend on which fault is in which processor. The safety systems in NPP I&C systems consist of several processors. These processors have specific functions, and a can be unit hosts in terms of cyber security [11].

In the percentages of each branch, the processor and the system effects are considered. This is based on the fact that those systems are able to perform evaluation by weighting the results of each processor. In some studies dealing with digital system reliability, a comprehensive result is provided by weighting the result of each component with its relative failure rate [8]. However, the approach of using failure rate is not appropriate from the viewpoint of cyber security dealing with various types of intended incidents.

To solve this problem, several documents related to SDLC are reviewed. According to these documents, the complexity and weakness of SDLC verification and validation (V&V) processes lead to residual faults in the target system [12]. In this regard, it can be assumed that intended faults tend to be implied because the SDLC V&V process is complex and vulnerable. With this assumption, the V&V based residual fault estimation model is used to calculate the number of residual faults in the processor. The number is calculated by analyzing various factors in the SDLC V&V process using a 'Bayesian Network'. Therefore, after determining the number of residual faults, weighting the processors is possible.

$$P_{x,sys} = \sum W_{proc} \cdot P_{x,proc} \quad (3)$$

$$W_{proc} = \frac{R_{proc}}{R_{sys}} \quad (4)$$

where $P_{x,proc}$ is the percentage for the processor (x = a, d, m, or t), $P_{x,sys}$ is the percentage for the system, $W_{proc}$is the weighting value for the processor, R_{proc} is the number of residual faults in the processor, and R_{sys} is the number of residual faults in the system. Based on Eq. 2, the percentage of the safe-state in terms of a baseline system, without security techniques, can be estimated using the following equation.

$$P_{s,sys} = (1 - P_{a,sys}) + P_{a,sys}P_{d,sys}P_{m,sys} + P_{a,sys}(1 - P_{d,sys})P_{t,sys} \quad (5)$$

Using the percentage for the safe-state, the *Performance Index* can be estimated as follows.

$$Performance\ Index = \frac{P'_{s,sys} - P_{s,sys}}{P_{s,sys}} \quad (6)$$

where $P'_{s,sys}$ is the percentage for the safe-state in terms of a system using security techniques.

2.3 Interpretation of HSI

The two-dimensional *HSI* includes the *Incompatibility Index* and the *Performance Index*. The *Incompatibility Index* has a positive value on the assumption that a security technique does not have a positive effect on the reliability of the existing system. The *Performance Index* also has a positive value on the assumption that security techniques are introduced only in the direction of improving security. If the *HSI* is displayed on a two-dimensional coordinate plane, it can be located in the upper right area. If the maximum limited value of the incompatible side and the minimum expected value of the performance side are set, allowable area can be created as shown in Fig. 3, and the system designer can obtain some insight from this meaningful information.

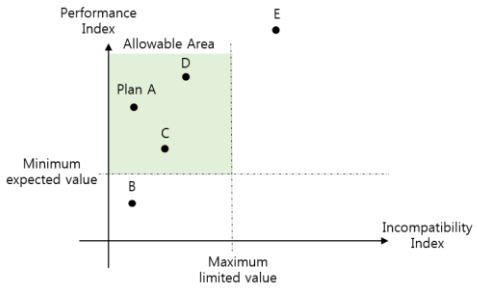

Fig. 3 Example of allowable area

3 Case Study

3.1 Target system and examples of security techniques

In this study, the target system is the digital plant protection system (DPPS), which is a safety-critical I&C system of NPPs [13]. The purpose of the DPPS is automatic generation of a trip signal in an emergency. In order to detect an emergency, system monitors various process parameters using independent instrumentation and processing channels. The signal-processing layout is shown in Fig. 4.

Conceptually, in a multi-channel digital protection system, the dominant cut-sets, which consist of the CCF probabilities of the digital modules and the error probability of the human operator, can be expressed mathematically as follows:

q_1 = Pr(OP) × Pr(AI CCF)
q_2 = Pr(OP) × Pr(DO CCF)
q_3 = Pr(OP) × Pr(PM CCF) × Pr(WDT CCF)
…

where

q_i is the probability of minimal cut-set i

Pr(OP) is the probability that a human operator fails in a manual operation

Pr(AI CCF) is the probability of the CCF of the analog input)

Pr(DO CCF) is the probability of the CCF of the analog input modules

Pr(PM CCF) is the probability of the CCF of the processor modules and

Pr(WDT CCF) is the probability of the CCF of the watchdog timers

Among these failure probabilities, only Pr(PM CCF) depends on the software failure probability [14].

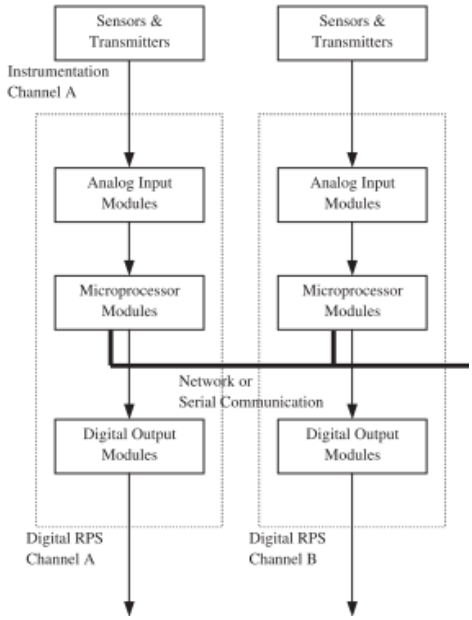

Fig. 4 Signal-processing layout

For the case study, examples of security techniques A,B (STA, STB) are assumed. STA and STB can detect intrusive traces, but the inspection periods are different. The specifications for STA and STB are summarized in Table 3.

Table 3. Examples of security techniques

Security Techniques	Inspection Period	System Complexity	S/W failure probability
STA	Real-Time	High	1.0E-03
STB	8 hour (period of ATIP)	Medium	1.0E-04

* ATIP: Automatic Test and Interface Processor

Factors affecting the increase in software complexity can be the number of applied security technologies, the detectable coverage, and the inspection period of security techniques. However, in this case study, it is assumed that only the inspection period determines the system complexity and the detecting performance. The software failure probabilities are assigned as shown in Table 3 based on the values in Table 2.

3.2 Calculation of Compatible Index

The calculated CDF results are summarized in Table 4. To calculate the CDF, fault trees for the DPPS were developed, and CCF events contributing 95% to the DPPS unavailability are investigated. Regarding the software failure probability, the values of 1.0E-03, 1.0E-04, and 1.0E-05 are used. Regarding the human failure probability, 0.5 is used. In addition, regarding the watchdog-timer coverage, 0.7 is used. With these calculated CDF values, *Incompatibility Indexes* are calculated.

Table 4. CDF calculation results [13]

CDF	WDT coverage 0.7		
OP failure	SW failure		
	1.0E-03	1.0E-04	1.0E-05
0.5	3.9	3.0	2.9

$$Incompatibility\ Index_{STA} = \frac{3.9 - 2.9}{2.9} = 34\%$$

$$Incompatibility\ Index_{STB} = \frac{3.0 - 2.9}{2.9} = 3.4\%$$

3.3 Calculation of *Incompatible Index*

Percentages in the intrusion-tolerance predicate block diagram are assumed according to the values in Table 5.

Table 5. Percentages in intrusion-tolerance predicate block diagram

Activated	Mitigated	Detected P_d			Tolerated P_t	
P_a	P_m	Base	STA	STB	Direct	Indirect

0.7	0.9	0	0.5	0.8	0.1	0.3

Among the percentages in the diagram, the percentage for activation and the percentage for mitigation are assumed as fixed values of 0.7 and 0.9. However, those values can be improved when security techniques that can make the system resistant or that can help operator mitigate errors in an intrusion situation are applied. Regarding the percentage of detection, it is determined according to which security technique is applied. The value of 0.8 is assumed for STA and 0.5 is assumed for STB. These assumptions are based on the concept that a security technique with shorter inspection period detects intrusive traces better. In addition, because intended faults are not detected without security techniques, the percentage of detection in the baseline system is 0. Regarding the percentage of tolerance, the value of 0.1 is assumed for processors; this value is directly related to the safety functions such as the bistable processor (BP) or the coincidence processor (CP). In addition, the value of 0.3 is assumed for processors; this value is indirectly related to the safety functions such as the ATIP or the cabinet operator module (COM).

For the number of residual faults in a processor, the results of research proposing a SDLC V&V system are used [12]. The number of estimated faults in the final software is summarized in Table 6.

Table 6. Number of estimated faults

Number of faults by BN				
DPPS total	BP	CP	ATIP	COM
3.69	1.51	1.40	0.38	0.40
1 (weighting)	0.41	0.38	0.1	0.11

With the percentage values for each branch and weighting values of the estimated residual faults, the percentages of the safe-state are estimated according to the system states. The system states are should one of these be the system without STA, and the baseline system without security techniques. $P_{s,sys}$ is 39.94% $P'_{s,sys-A}$ is 82.39%, and $P'_{s,sys-B}$ is 66.47%.

$$Performance\ Index_{STA} = \frac{82.39 - 39.94}{39.94} = 106\%$$

$$Performance\ Index_{STB} = \frac{66.47 - 39.94}{39.94} = 66\%$$

3.3 Results

HSI includes the *Incompatibility Index* and the *Performance Index*. Both indexes are calculated for systems with security techniques A and B.

$$\boldsymbol{HSI_A = (0.34, 1.06)}$$

$$\boldsymbol{HSI_B = (0.034, 0.66)}$$

If the maximum limited value of the incompatible side and the minimum expected value of the performance side are assumed to be 0.3 and 0.6, the allowable area can be described as shown in Fig. 5. Both values of *HSI* are displayed on a two-dimensional coordinate plane according to their *Incompatibility Index* and *Performance Index*. Values.

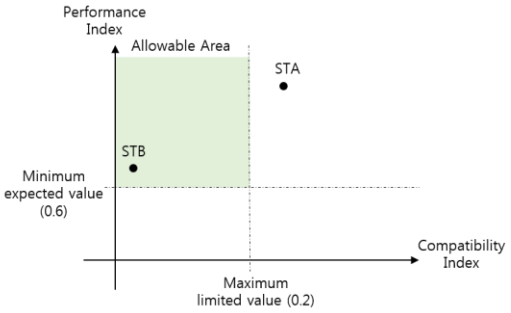

Fig.5 Results of HSI and allowable area

Although STA has a higher Performance Index than STB has, it can make the target system too complex; the effect is quantified as a high *Incompatibility Index* value exceeding the maximum limited value. However, although STB has a lower *Performance Index*, it can be considered an appropriate solution due to its lower *Incompatibility Index*. This result can help system designers obtain some insight from this meaningful information.

4 Conclusion

Security techniques to be applied were not considered when the operating I&C systems were developed. Therefore, it is necessary to analyze not only security enhancement, but also the influence of the security techniques on the target system in terms of system reliability. In this study, we developed a two-dimensional index called the *Hybrid Security Index (HSI)*. *HSI* includes the *Incompatibility Index* and the *Performance Index*.

The *Incompatibility Index* is estimated based on the concept that security techniques should not affect the reliability of existing systems because the introduction of security techniques can complicate the system in terms of software structure and data communication. In the proposed quantification method, indirect evidence is applied to estimate the failure probability of software modules.

The *Performance Index* is defined as the rate of increase in percentage of errors in the safe-state among errors caused by an intended fault. In safe-state, availability of safety functions are protected by applied security techniques. As an abstraction, a predicate block diagram is used to classify experimental errors caused by digital faults. However, in this study, excepting the faults that can be detected in FTTs, only intelligently intended faults are assumed. In addition, the block diagram is modified by adding a mitigated error node. With the modified diagram, the percentage of safe-state can be estimated. Furthermore, in order to provide comprehensive information, the results for each processor are weighted with the number of residual faults, which can be estimated by using the V&V based residual fault estimation model.

The result values of *HSI* can help assess not only the degree to which system security can be improved if specific cyber security techniques are applied, but also the influence of the security techniques on the target system in terms of system reliability. 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 techniques quantitatively, the method can also be applied to establish a specific target of efficacy level that system can achieve.

However, there are some limitations in this work in terms of the estimation of the percentages in the intrusion-tolerance predicated block diagram. This is because methods of obtaining percentages for each branch need to be elaborated. Also, verification and validation of the suggested method need to be conducted

References

[1] J. Song, J. Lee, C. Lee, K. Kwon, and D. Lee, “A Cyber Security Risk Assessment for the Design of I&C Systems in Nuclear Power Plants,” *Nucl. Eng. …*, vol. 44, no. 8, pp. 919–928, 2012.

[2] C. Baylon, R. Brunt, and D. Livingstone, “Cyber Security at Civil Nuclear Facilities Understanding the Risks,” *Chatham House*, p. 53, 2015.

[3] J. G. Song, J. W. Lee, G. Y. Park, K. C. Kwon, D. Y. Lee, and C. K. Lee, “An analysis of technical security control requirements for digital I&C systems in nuclear power plants,” *Nucl. Eng. Technol.*, vol. 45, no. 5, pp. 637–652, 2013.

[4] J. Shin, H. Son, R. Khalil Ur, and G. Heo, “Development of a cyber security risk model using Bayesian networks,” *Reliab. Eng. Syst. Saf.*, vol. 134, pp. 208–217, Feb. 2015.

[5] O. Bäckström, J.-E. Holmberg, M. Jockenhoevel-Barttfeld, M. Porthin, and A. Taurines, “Quantification of reactor protection system software reliability based on indirect and direct evidence,” *PSAM 2014 - Probabilistic Saf. Assess. Manag.*, no. June, 2014.

[6] “Development of the Digital Reactor Safety System,” 2007.

[7] S. H. Lee, H. E. Kim, K. S. Son, S. M. Shin, S. J. Lee, and H. G. Kang, “Reliability modeling of safety-critical network communication in a digitalized nuclear power plant,” *Reliab. Eng. Syst. Saf.*, vol. 144, pp. 285–295, 2015.

[8] J. S. Lee, M. C. Kim, P. H. Seong, H. G. Kang, and S. C. Jang, “Evaluation of error detection coverage and fault-tolerance of digital plant protection system in nuclear

power plants," *Ann. Nucl. Energy*, vol. 33, no. 6, pp. 544–554, 2006.

[9] B. G. Kim, H. G. Kang, H. E. Kim, S. J. Lee, and P. H. Seong, "Reliability modeling of digital component in plant protection system with various fault-tolerant techniques," *Nucl. Eng. Des.*, vol. 265, pp. 1005–1015, 2013.

[10] 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.

[11] J. Cho, S. J. Lee, and W. Jung, "Fault-weighted quantification method of fault detection coverage through fault mode and effect analysis in digital I&C systems," *Nucl. Eng. Des.*, vol. 316, pp. 198–208, 2017.

[12] H. S. Eom, G. Y. Park, S. C. Jang, H. S. Son, and H. G. Kang, "V&V-based remaining fault estimation model for safety-critical software of a nuclear power plant," *Ann. Nucl. Energy*, vol. 51, no. December 2016, pp. 38–49, 2013.

[13] H. G. Kang and S.-C. Jang, "Plant risk effect analysis focusing on digital I & C equipment failures," *J. Nucl. Sci. Technol.*, vol. 44, no. 4, 2007.

[14] H. G. Kang and T. Sung, "An analysis of safety-critical digital systems for risk-informed design," *Reliab. Eng. Syst. Saf.*, vol. 78, no. 3, pp. 307–314, 2002.