Allocation of Defense Resources against Cyber Attacks to Cyber-Physical Systems

Wei Wang

Department of Mechanical Engineering, City University of Hong Kong, Hong Kong, China. E-mail: wwang326@cityu.edu.hk

Francesco Di Maio

Department of Energy, Politecnico di Milano, Milano, Italy. E-mail: francesco.dimaio@polimi.it

Enrico Zio

Department of Energy, Politecnico di Milano, Milano, Italy
MINES ParisTech / PSL Université Paris, Centre de Recherche sur les Risques et les Crises (CRC), Sophia Antipolis, France
Eminent Scholar, Department of Nuclear Engineering, Kyung Hee University, Republic of Korea.
Email: enrico.zio@polimi.it

Protecting Cyber-Physical Systems (CPSs) from cyber attacks requires properly allocating defense resources. These can be selected by defend-attack and Adversarial Risk Analysis (ARA) models, which search for the optimal allocation based on specific assumptions. In particular, the defend-attack model assumes that each player is fully aware of the preferences of the opponent, considering complete information, whereas the ARA model assumes incomplete information and subjective probability distributions of the defender utilities, improving the realism of the modelling but still lacking a proper management of the uncertainties of the results it provides. In this work, we complement the ARA model with a multi-criteria decision model based on Value-at-Risk (VaR) measures to support the defender in identifying the optimal defense portfolio among alternatives, considering budget constraints and accounting for the uncertainties which the ARA model is subjected to. For demonstration purposes, an application is carried out concerning the digital control system of the Advanced Lead-cooled Fast Reactor European Demonstrator (ALFRED).

Keywords: Cyber-Physical System; Cyber attacks; Weak knowledge; Adversarial Risk Analysis; Value-at-Risk; Multi-criteria decision making; Nuclear Power Plant.

1. Introduction

Cyber-Physical Systems (CPSs) are systems with tight combination of (and coordination between) the physical and cyber processes (Alur, 2015; Khaitan & McCalley, 2014; Lee, 2008). As such, they are vulnerable to failures due also to cyber breaches (Wang, Di Maio, & Zio, 2017, 2020; Zio, 2018), which can compromise the CPS functionality. Defensive resources against cyber attacks can be allocated by applying defend-attack models (Hausken & Levitin, 2012; Lallie, Debattista, & Bal, 2017) and Adversarial Risk Analysis (ARA) models (Banks, Aliaga, & Insua, 2015; Rios Insua, Rios, & Banks, 2009), with the objective of minimizing the impacts of the attacks on the CPS functionality and, hence, maximizing their reliability and survivability.

In a previous work, an ARA model has been built by the authors (Wang, Di Maio, & Zio, 2019). The model is based on incomplete information about utility functions, payoffs, and strategies, expressed by subjective probability distributions of the occurrence of the defend-attack outcomes and consequences.

In the present work, we complement the ARA model with a multi-criteria decision model to support the defender task of identifying the best defensive barriers under budget constraints and accounting for the uncertainties which the ARA model is subjected to. The multi-criteria decision model is based on Value-at-Risk (VaR) measures (Rockafellar & Uryasev, 2002) for trading off the level of risk and the uncertainty associated to the potential loss arising from the occurrence of the extreme-risk scenarios.

For demonstration purpose, an application is carried out concerning the digital Instrumentation and Control (I&C) system of the Advanced Lead-cooled Fast Reactor European Demonstrator (ALFRED) (Ponciroli, Cammi,

Proceedings of the 30th European Safety and Reliability Conference and the 15th Probabilistic Safety Assessment and Management Conference
Edited by Piero Baraldi, Francesco Di Maio and Enrico Zio
Published by Research Publishing, Singapore.
ISBN: 978-981-14-8593-0; doi:10.3850/978-981-14-8593-0

Bona, Lorenzi, & Luzzi, 2015). In this system, multiple failures can be induced by external cyber attacks, and the defender needs to choose a defense strategy (Wang et al., 2019).

The remainder of the paper is organized as follows. Section 2 presents the ALFRED digital I&C system and the ARA model of the attacker-defender behavior. In Section 3, a multi-criteria decision model based on the VaR measures is built, complementing the ARA model, to identify a best compromise solution of defensive resource allocation. Results on the ALFRED case study are presented and discussed in Section 4, and conclusions are drawn in Section 5.

2. ALFRED and ARA Model

ALFRED is a concept of a small-size (300 MW) pool-type fast reactor, cooled by molten lead. At full power nominal conditions, the dynamics of the ALFRED primary and secondary cooling systems is controlled by a multi-loop PI (Proportional and Integral) control scheme (see Figure 1), i.e., a decentralized control scheme (Ponciroli et al., 2015). Both feedback and feedforward digital control schemes are adopted (see Figure 1 yellow shadowed part). The PI-based feedback control configuration employs four SISO (Single Input Single Output) control loops independent of each other.

2.1. Uncertain cyber attacks

We consider a pool of A=15 cyber attack strategies (of 4 types, as listed in Table 1, namely, a_1 (attacks to different sensor databases); a_2 (attacks to commands of different actuators); a_3 (attacks to changes of PI gain values); a_4 (attacks to changes of set point values of controlled variables)). These strategies are constrained by the resources available to launch a single attack to hit a single CPS component.

Table 1. Cyber attack strategies

Attack type (a_j)	Attack target, ($a_{j,y}$)
(a_1) sensor databases	($a_{1,1}$) Steam Generator (SG) outlet temperature (T_{steam})
	($a_{1,2}$) SG pressure (p_{SG})
	($a_{1,3}$) Coolant SG outlet temperature ($T_{L,cold}$)
	($a_{1,4}$) Thermal power (P_{Th})
(a_2) commands of actuators	($a_{2,1}$) Height of control rods (h_{CR})
	($a_{2,2}$) Feedwater mass flow rate (G_{water})
	($a_{2,3}$) Turbine admission valve coefficient (kv)
(a_3) changes of PI gain values	($a_{3,1}$) PI_1
	($a_{3,2}$) PI_2
	($a_{3,3}$) PI_3
	($a_{3,4}$) PI_4
(a_4) changes of set point values	($a_{4,1}$) T_{steam} set point ($T_{steam,set}$)
	($a_{4,2}$) p_{SG} set point ($p_{SG,set}$)
	($a_{4,3}$) $T_{L,cold}$ set point ($T_{L,cold,set}$)
	($a_{4,4}$) P_{Th} set point ($P_{Th,set}$)

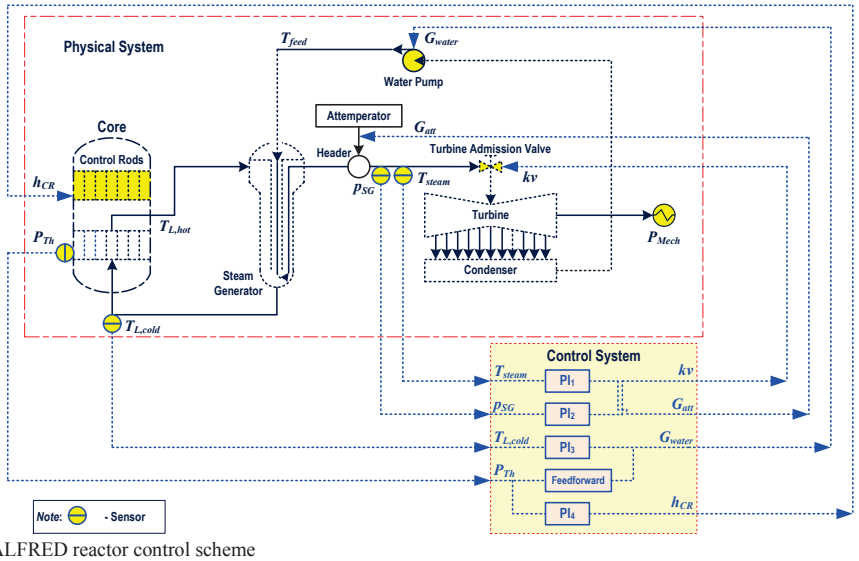

Figure 1. ALFRED reactor control scheme

2.2. *Defensive resources*

The digital I&C system features numerous hardware and software elements interfacing the monitoring and control system with the physical process, for controlling it and, at the same time, protecting it from cyber attacks. Defensive resources are deployed with the aim of (d_1) preventing cyber attacks and (d_2) recovering from a successful cyber attack. Table 2 lists the defensive resources considered, with their relevance to the different cyber attacks (Column 3), and their minimum and maximum deployable quantity (Column 4) as assessed by expert judgment.

Considering a maximum defense budget B_M equal to 2,000 k€ (for sake of illustration), a set of $\Re$ = 4834 alternative defense portfolios $d^r = \{d_1^r, d_2^r\} = \{n_{1,1}^r, n_{1,2}^r, n_{1,3}^r, n_{1,4}^r, n_{2,1}^r, n_{2,2}^r, n_{2,3}^r\} \in \Re$ are yielded with $n_{1,1}^r$ = 0, 1, $n_{1,2}^r$ = 0, 1, 2, 3, 4, $n_{1,3}^r$ = 1, 2, 3, 4, $n_{1,4}^r$ = 0, 1, 2, 3, $n_{2,1}^r$ = 0, 1, 2, 3, $n_{2,2}^r$ = 0, 1, 2 and $n_{2,3}^r$ = 0, 1, 2, with considering the mean values of the annual costs of the defensive resources taken from (Wang et al., 2019). The resulting $\Re$ = 4834 deployable portfolios are hereafter referred to by the rule of successively increasing one countermeasure at a time, beginning with $n_{2,3}^r$ and ending with $n_{1,1}^r$, and, thus, lead to the permutations $d^1 = \{0,0,1,0,0,0,0\}$, $d^2 = \{0,0,1,0,0,0,1\}$, …, and $d^{4834} = \{1,4,4,3,2,2,2\}$.

Table 2. Defensive resources with properties

Defense type, d_i	Countermeasures, $x_{i,k}$	Relevance	Min.-Max., $n_{i,k}$
(d_1) Prevention	($x_{1,1}$) firewall	High	0-1
	($x_{1,2}$) Intrusion Detection Systems (IDSs)	Moderate	0-4
	($x_{1,3}$) operators	Moderate	1-4
	($x_{1,4}$) security software	Moderate	0-3
(d_2) Recovery	($x_{2,1}$) mainframe computers	High	0-3
	($x_{2,2}$) database servers	High	0-2
	($x_{2,3}$) security engineers	High	0-2

2.3. *The ARA model*

The defender objective is to choose the best defense strategy $d^r = \{d_1^r, d_2^r\}$ among $\Re$ = 4834 available, for protecting the digital I&C system from an attack to one single element $a_{j,y}$ among the A = 15 possible ones, given limited resources. Both defender and attacker want to maximize their expected utilities, which are used to characterize the defender and attacker uncertain behaviors in the decision analyses. It is assumed that the defender has no information about the attacker's actions and the attacker has no knowledge about the specific configuration of the digital I&C system. Using the ARA model to describe the attacker-defender behavior, different combinations of defense and attack strategies $(d^r, a_{j,y}) = (d_1^r, d_2^r, a_{j,y})$ lead to uncertain outcomes and consequences, with uncertain costs for both the defender and the attacker (Hausken & Levitin, 2012; Wang et al., 2019).

An ARA model has been built by the authors in a previous work (Wang et al., 2019). The general idea is shown in Figure 2: the defender seeks for the optimal allocation of resources for the defensive barriers d^*, according to a subjective expected utility model that considers that the defender only knows his/her own beliefs of costs, utilities, and consequences of his/her decisions, and only speculates about those of the attacker. For example, $\pi_D(a_{j,y}|d^r)$, is the defender (uncertain) estimation of the probability of occurrence of attack $a_{j,y}$, launched against the system that is defended with resources d^r.

Consider the outcomes $s = \{s_1, s_2\}$ of the game generated in Section 2.3, where $s_1 = s_1(d_1^r, a_{j,y})$ defines the successful prevention of d_1^r to an attack $a_{j,y}$, and $s_2 = s_2(d_2^r|s_1)$ the successful recovery of d_2^r in case of successful attack (i.e., $s_1 = 1$), the optimal defensive strategy d^* can be obtained by the defender

expected utility $\Psi_D(d)$ maximization with reference to each defensive portfolio d^r:

$$d^* = \underset{d^r \in \Re}{\operatorname{argmax}}\, \Psi_D(d^r) \quad (1)$$

where $\Psi_D(d^r)$ is defined as in:

$$\Psi_D(d^r) = \sum_{a_{j,y} \in A} \left[\sum_{s \in \{0,1\}} \pi_D(a_{j,y}|d^r) \cdot p_D(s) \cdot u_D \right] \quad (2)$$

where u_D is the utility describing the defender cost of consequence and $p_D(s)$ denotes the probability of the defender outcome (i.e., s).

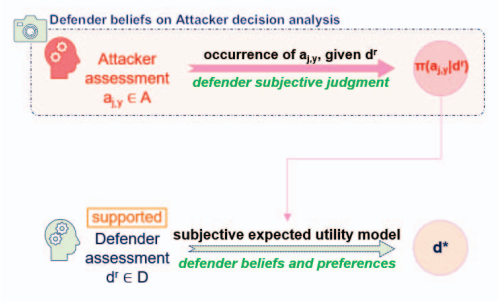

Figure 2. ARA framework (Wang et al., 2019)

For further details on the ARA model, the interested the reader may refer to (Wang et al., 2019).

3. Multi-Criteria Decision Model for Defensive Resource Allocation

A multi-criteria decision model is here proposed to identify the best compromise solution of the defensive resource allocation, complementing the results of the defend-attack ARA model of Section 2.3.

Notice that the weak knowledge of the defender, modelled by the subjective probability distributions of the uncertain parameters of the ARA model may reduce the persuasiveness of the optimal strategy d^* (Eq. (1)). To deal with this, in this section we propose a VaR-based multi-criteria decision model to provide the defender with a best compromise solution among many portfolio alternatives in $\Re$, for trading off the level of risk and the uncertainty associated to the potential loss arising from the occurrence of the extreme risky scenarios in the decision portfolio against possible cyber attacks.

A loss function $L_D(d^r)$ is here defined as the opposite value of $\Psi_D(d^r)$, expressed in Eq. (3), indicating the loss that may come from a defensive decision:

$$L_D(d^r) = -\Psi_D(d^r) \quad (3)$$

A Monte Carlo (MC) simulation with a total of N runs embedded into the ARA model of Eq. (2) can generate N values of $\Psi_D(d^r)$, thus, of $L_D(d^r)$ with respect to each d^r, to be fundamental for the estimation of the risk and the uncertainty that d^r may be subjected to.

Value-at-Risk (VaR) measures can be used to quantify and control the level of risk, that is the extent and occurrence ratio of the potential losses in the decision portfolios (Dabbagh & Sheikh-El-Eslami, 2015; Mena, Hennebel, Li, & Zio, 2016; Rockafellar & Uryasev, 2002). As shown in Figure 3, given a discrete approximation of the probability function of $L_D(d^r)$ obtained from the resultant N values, $VaR_\alpha[L_D(d^r)]$ indicates the smallest value of the loss $L_D(d^r)$, for which the probability that $L_D(d^r)$ does not exceed a predefined threshold value (e.g., infinity) is larger than or equal to an α percentile, as expressed:

$$VaR_\alpha[L_D(d^r)] = \inf\{z: F_{L_D(d^r)}(z) > \alpha\} \quad (4)$$

whereas, a conditional VaR measure, i.e., $CVaR_\alpha[L_D(d^r)]$, defines the expected value of $L_D(d^r)$ that is larger beyond or equal to $VaR_\alpha(L_D(d))$:

$$CVaR_\alpha[L_D(d^r)] = \mathbb{E}\big[L_D(d^r)|L_D(d) \geq VaR_\alpha(L_D(d^r))\big] \quad (5)$$

$CVaR_\alpha[L_D(d^r)]$ indicates the extent of the loss $L_D(d^r)$ originating from the occurrence of the extreme risky scenarios, quantifying the level of risk associated to the potential loss in the decision portfolio d^r against all possible cyber attacks.

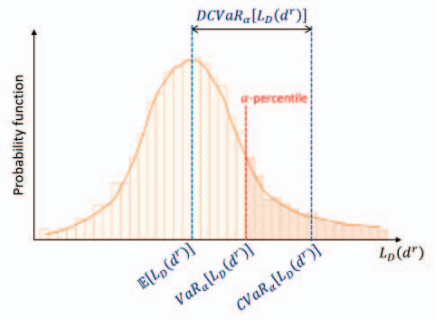

Figure 3. The VaR measures

A non-symmetric deviation CVaR measure, $DCVaR_\alpha[L_D(d^r)]$, is defined to take account of the uncertainty associated to the loss $L_D(d^r)$, which exceeds its expected value in the decision portfolio d^r against all possible cyber attacks:

$$DCVaR_\alpha[L_D(d^r)] = CVaR_\alpha[L_D(d^r)] - \mathbb{E}[L_D(d^r)] \quad (6)$$

A multi-criteria decision model is developed to identify the best compromise solution of the defensive resource allocation $\tilde{d}$, by minimizing both the CVaR and the DCVaR values as expressed in Eq. (7), where $CVaR_\alpha[L_D(d^r)]$ and $DCVaR_\alpha[L_D(d^r)]$ for each d^r are calculated according to the N values of $L_D(d^r)$ resulted from the MC simulation.

$$\tilde{d} = \underset{d^r \in \Re}{\operatorname{argmin}}\{CVaR_\alpha[L_D(d^r)], DCVaR_\alpha[L_D(d^r)]\} \quad (7)$$

4. Results

A total of $N = 1000$ runs of the MC simulation are performed to operationalize the ARA model with respect to the ALFRED case study, resulting in the calculation of $CVaR_{95}[L_D(d^r)]$ and $DCVaR_{95}[L_D(d^r)]$ for the defensive portfolios d^r among the $\Re$ = 4834 available, given the assumption of 95th percentile.

Figure 4(a) shows the results of $CVaR_{95}[L_D(d^r)]$ and highlights the six portfolios with smallest CVaR values, i.e., $d^{4807} = \{1,4,4,2,2,2,2\}$, $d^{4372} = \{1,3,4,3,2,2,2\}$, $d^{4834} = \{1,4,4,3,2,2,2\}$ (which is the defender best response $d^*_{Nash}(a_{j,y})$ of Nash equilibrium solution obtained from a classical defend-attack model in (Wang et al., 2019)), $d^{4749} = \{1,4,4,0,2,2,2\}$, $d^{4722} = \{1,4,3,3,2,2,2\}$ and $d^{4779} = \{1,4,4,1,2,2,2\}$ (which is the optimal defense decision d^* obtained from the ARA model in (Wang et al., 2019)). These portfolios are allocated with the most amount of countermeasures deployable, such that the extent of the probability of occurrence of the extreme risky scenarios of loss becomes relatively low, thus, resulting in a relative low level of risk.

The results of $DCVaR_{95}[L_D(d^r)]$ are reported in Figure 4(b), which also highlights the six portfolios with smallest DCVaR values, namely, $d^{37} = \{0,0,1,1,0,0,0\}$, $d^{2159} = \{0,4,2,1,0,0,0\}$, $d^{4} = \{0,0,1,0,0,1,0\}$, $d^{680} = \{0,1,2,1,0,1,0\}$, $d^{515} = \{0,1,1,0,0,2,0\}$ and $d^{212} = \{0,0,2,2,1,0,0\}$. These portfolios are allocated with few countermeasures deployable, such that the uncertainty associated to the loss exceeding its expected value becomes relatively low, probably due to the cost saving of their annual budget.

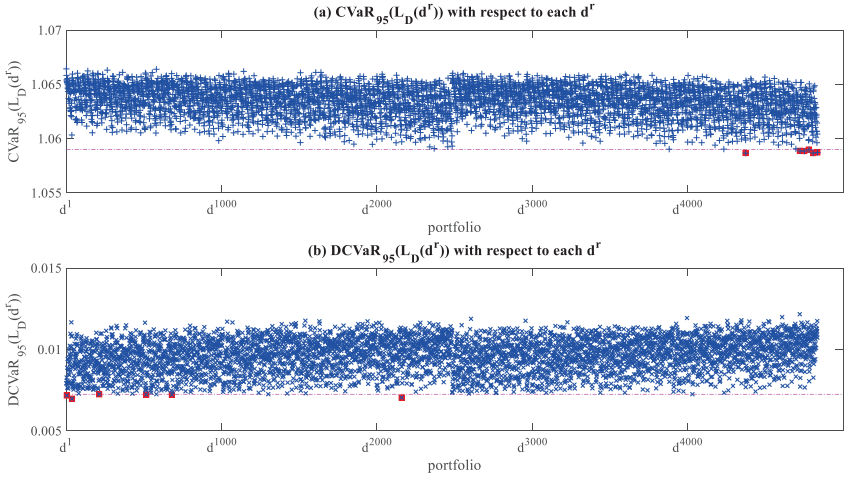

Figure 4. Results of (a) $CVaR_{95}[L_D(d^r)]$ and (b) $DCVaR_{95}[L_D(d^r)]$ with respect to each d^r

Figure 5 shows the mapping of all the defensive portfolios d^r, where the horizontal axis sorts d^r by ascending $CVaR_{95}[L_D(d^r)]$ and the vertical axis sorts d^r by ascending

$DCVaR_{95}[L_D(d^r)]$. The map is divided into 16 zones, in which the boundaries define that a portfolio d^r drops in a zone of $z_{\delta,\theta}$, where δ (= 1, 2, 3, or 4) represents the CVaR value of d^r ranks in the δ out of 4 ascending part and θ (= 1, 2, 3, or 4) means the DCVaR value ranks in the θ out of 4 ascending part. Thus, a smaller δ represents a lower level of risk of d^r once deployed, and a smaller θ represents a lower level of uncertainty associated to the loss that the deployed d^r may occur. The criteria lead to the expectance of the defense decision $\tilde{d}$ when dropping into the zone of $z_{1,1}$.

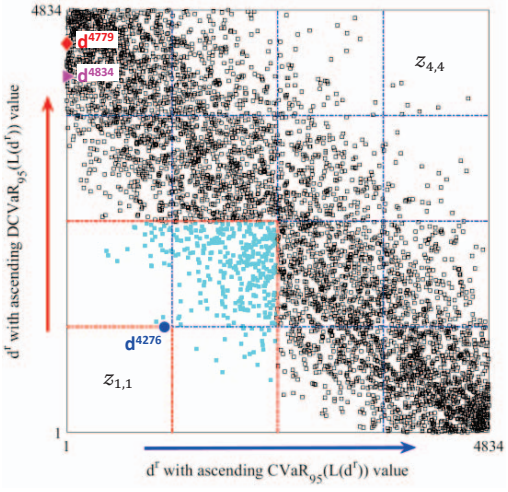

Figure 5. The mapping of d^r obtained from the ascending sorts of $CVaR_{95}[L_D(d^r)]$ and $DCVaR_{95}[L_D(d^r)]$

It is shown in Figure 5 that the only portfolio $\tilde{d} = d^{4276} = \{1,3,4,0,1,2,0\}$ (marked as dot) falls into the zone of $z_{1,1}$, which is identified as the best compromise solution among $\Re$ = 4834 available portfolios, in case of the assignment of the subjective probability distributions of the uncertain parameters based on the defender weak knowledge, which may result in the unacceptable level of uncertainty of the loss that the so-called optimal results $d^{4834} = \{1,4,4,3,2,2,2\}$ (i.e., the defender best response $d^*_{Nash}(a_{j,y})$ of Nash equilibrium solution, marked as triangle in Figure 5) and $d^{4779} = \{1,4,4,1,2,2,2\}$ (i.e., the optimal defense decision d^* of the ARA model, marked as diamond in Figure 5) struggle for. By so doing, $\tilde{d}$ enables the trade-off that leverages the acceptable levels of risk with uncertainty, which the loss of the defensive resource allocation faces with. It is worth pointing out that compared to the very low level of risks of $d^{4834} = \{1,4,4,3,2,2,2\}$ and $d^{4779} = \{1,4,4,1,2,2,2\}$ (although very high level of uncertainty falling into the zone of $z_{1,4}$), the relatively high level of risk of $\tilde{d}$ arises from the fact that the efficacy of the defensive resource allocation is reduced while not considering any security software (moderate relevance) or security engineer (high relevance).

5. Conclusions

In this study, we have proposed a multi-criteria decision model to complement the defend-attack Adversarial Risk Analysis (ARA) model, in case that the subjective probability distributions of the uncertain parameters are set based on the defender weak knowledge, aimed at guiding the identification of the best compromise solution for the defensive resource allocation against cyber attacks to Cyber-Physical Systems (CPSs). Value-at-Risk (VaR) measure has been considered in the multi-criteria decision model, to quantify the levels of risk and of uncertainty associated to the potential losses arising from the occurrence of the extreme risky scenarios with respect to the cyber security decision.

For demonstration, we have taken the digital Instrumentation and Control (I&C) system of the Advanced Lead Fast Reactor European Demonstrator (ALFRED) as the object of the study, where a cyber defend-attack game is generated between a security defender choosing the defense strategy among alternatives and an attacker launching an uncertain cyber attack against a single element of the digital control system. The proposed model provided the best compromise portfolio that leverages the levels of risk with uncertainty that the defense decision is subjected to, which is fundamental to an effective protection design of CPSs against uncertain cyber attacks.

Acknowledgement

The first author gratefully acknowledges the financial supports for this research from City University of Hong Kong (No. 7200654).

References

Alur, R. (2015). *Principles of cyber-physical systems*: MIT Press.

Banks, D. L., Aliaga, J. M. R., & Insua, D. R. (2015). *Adversarial risk analysis*: Chapman and Hall/CRC.

Dabbagh, S. R., & Sheikh-El-Eslami, M. K. (2015). Risk-based profit allocation to DERs integrated with a virtual power plant using cooperative Game theory. *Electric Power Systems Research, 121*, 368-378.

Hausken, K., & Levitin, G. (2012). Review of systems defense and attack models. *International Journal of Performability Engineering, 8*(4), 355-366.

Khaitan, S. K., & McCalley, J. D. (2014). Design techniques and applications of cyberphysical systems: A survey. *IEEE Systems Journal, 9*(2), 350-365.

Lallie, H. S., Debattista, K., & Bal, J. (2017). An empirical evaluation of the effectiveness of attack graphs and fault trees in cyber-attack perception. *IEEE Transactions on Information Forensics and Security, 13*(5), 1110-1122.

Lee, E. A. (2008). *Cyber physical systems: Design challenges*. Paper presented at the 2008 11th IEEE International Symposium on Object and Component-Oriented Real-Time Distributed Computing (ISORC).

Mena, R., Hennebel, M., Li, Y.-F., & Zio, E. (2016). A multi-objective optimization framework for risk-controlled integration of renewable generation into electric power systems. *Energy, 106*, 712-727.

Ponciroli, R., Cammi, A., Bona, A. D., Lorenzi, S., & Luzzi, L. (2015). Development of the ALFRED reactor full power mode control system. *Progress in Nuclear Energy, 85*, 428-440.

Rios Insua, D., Rios, J., & Banks, D. (2009). Adversarial risk analysis. *Journal of the American Statistical Association, 104*(486), 841-854.

Rockafellar, R. T., & Uryasev, S. (2002). Conditional value-at-risk for general loss distributions. *Journal of banking & finance, 26*(7), 1443-1471.

Wang, W., Di Maio, F., & Zio, E. (2017). Three-Loop Monte Carlo Simulation Approach to Multi-State Physics Modeling for System Reliability Assessment. *Reliability Engineering & System Safety, 167*.

Wang, W., Di Maio, F., & Zio, E. (2019). Adversarial Risk Analysis to Allocate Optimal Defense Resources for Protecting Cyber–Physical Systems from Cyber Attacks. *Risk Analysis, 39*(12), 2766-2785.

Wang, W., Di Maio, F., & Zio, E. (2020). Considering the Human Operator Cognitive Process for the Interpretation of Diagnostic Outcomes Related to Component Failures and Cyber Security Attacks. *Reliability Engineering & System Safety*, 107007.

Zio, E. (2018). The future of risk assessment. *Reliability Engineering & System Safety, 177*, 176-190.