ADVERSARIAL RISK ANALYSIS TO ALLOCATE OPTIMAL DEFENSE RESOURCES FOR PROTECTING NUCLEAR POWER PLANTS FROM CYBER ATTACKS

Wei Wang[1], Francesco Di Maio[1], Enrico Zio[1,2]

[1]*Energy Department, Politecnico di Milano, Via La Masa 34, 20156 Milano, Italy*

[2]*Chair on System Science and the Energy Challenge, Fondation Electricite' de France (EDF), CentraleSupélec, Université Paris Saclay, 91190 Gif-sur-Yvette, France*

Abstract: Defenders have to enforce defense strategies by taking decisions on allocation of resources, to protect the integrity and survivability of CPSs from intentional and malicious cyber attacks. In this work, we propose an Adversarial Risk Analysis (ARA) approach to provide a novel one-sided (i.e.,) prescriptive support strategy for the defender to optimize the defensive resource allocation, based on a subjective expected utility model, in which the decisions of the adversaries are uncertain. This increases confidence in cyber security through robustness of CPS protection actions against uncertain malicious threats, compared with prescriptions provided by a classical defend-attack game-theoretical approach. We present the approach and the results of its application to a nuclear CPS, specifically, the digital Instrumentation and Control (I&C) system of the Advanced Lead-cooled Fast Reactor European Demonstrator (ALFRED)).

Keywords: Cyber-Physical System; Cyber Security; Adversarial Risk Analysis (ARA); Game Theory; Defend-Attack Model; Defense Strategy; Optimization; Nuclear Power Plant.

ABBREVIATIONS

ALFRED	Advanced Lead-cooled Fast Reactor European Demonstrator
ARA	Adversarial Risk Analysis
CPS	Cyber-Physical System
CR	Control Rod
HMI	Human-Machine Interface
I&C	Instrumentation and Control
IDS	Intrusion Detection System
IP	Internet Protocol
MC	Monte Carlo
NPP	Nuclear Power Plant
PI	Proportional and Integral
R&D	Research and Development
SISO	Single Input Single Output

NOMENCLATURE

P_{Th}	Thermal power
h_{CR}	Height of control rods
$T_{L,hot}$	Coolant core outlet temperature
$T_{L,cold}$	Coolant SG outlet temperature
Γ	Coolant mass flow rate
T_{feed}	Feedwater SG inlet temperature
T_{steam}	Steam SG outlet temperature
p_{SG}	SG pressure
G_{water}	Feedwater mass flow rate
G_{att}	Attemperator mass flow rate
kv	Turbine admission valve coefficient
P_{Mech}	Mechanical power
a_j	Cyber attack type
$a_{j,y}$	Cyber attack strategy, $a_{j,y} \in A$
c_{prep}	Attack preparation cost
c_{Aq}	Monetized attack consequence
d_i	Defense type, i.e., d_1 for prevention and d_2 for recovery
$x_{i,k}$	k-th countermeasure of the i-th defense type
$c_{i,k}$	Annual cost of $x_{i,k}$
$\gamma_{i,k}^{j}$	$x_{i,k}$ relevance with respect to a_j
$\phi_{s_1}^{j}$	Probability of attack success (probability of prevention failure)
$\phi_{s_2}^{j}$	Probability of recovery failure (if attack success)
B_M	Maximum defense budget
d^r	Defense portfolio, $d^r = \left\{d_1^r, d_2^r\right\} \in \Re$
$n_{i,k}^{r}$	Quantity of $x_{i,k}$ deployed in the d^r
$\vec{s}$	Outcome set, i.e., $\vec{s} = \{s_1\left(d_1^r, a_{j,y}\right), s_2\left(d_2^r \mid s_1\right)\}$

$s_1\left(d_1^r, a_{j,y}\right)$	Successful prevention of the d_1^r to an $a_{j,y}$
$s_2\left(d_2^r \mid s_1\right)$	Successful recovery of the d_2^r in case of successful attack (i.e., $s_1\left(d_1^r, a_{j,y}\right)=1$)
$\vec{p}$	Probability set of $\vec{s}$, i.e., $\vec{p}=\left\{p\left(s_1\left(d_1^r, a_{j,y}\right)\right), p\left(s_2\left(d_2^r \mid s_1\right)\right)\right\}$
$p\left(s_1\left(d_1^r, a_{j,y}\right)\right)$	Probability of the $s_1\left(d_1^r, a_{j,y}\right)$
$p\left(s_2\left(d_2^r \mid s_1\right)\right)$	Probability of the $s_2\left(d_2^r \mid s_1\right)$ in case of successful attack (i.e., $s_1\left(d_1^r, a_{j,y}\right)=1$)
$c_{Dl}\left(\vec{s} \mid d^r, a_{j,y}\right)$	Defender cost of the *l*-th consequence of $\vec{s}$
$p_l^{\alpha}\left(\vec{s} \mid a_{j,y}\right)$	Conditional probability of $\vec{s}$ to the occurrence of $a_{j,y}$ with the α-th effect of the *l*-th consequence
c_{Dl}^{α}	Defender cost of an attack with the α-th effect of the *l*-th consequence
c_{annual}^r	Defender annual cost
$c_D\left(\vec{s} \mid d^r, a_{j,y}\right)$	Defender total cost
$u_D\left(\vec{s} \mid d^r, a_{j,y}\right)$	Defender utility
$c_A\left(\vec{s} \mid d^r, a_{j,y}\right)$	Attacker total cost
$u_A\left(\vec{s} \mid d^r, a_{j,y}\right)$	Attacker utility
$\psi_D\left(d^r\right)$	Defender expected utility
$\pi_D\left(a_{j,y} \mid d^r\right)$	Defender estimation of the probability of occurrence of any $a_{j,k}$ attack, given that the d^r is deployed
d^*	Defender optimal defense strategy
$\psi_A\left(a_{j,y} \mid d^r\right)$	Attacker expected utility of $a_{j,k}$ conditioned on the d^r
$a^*\left(d^r\right)$	Attacker optimal attack strategy, conditioned on the d^r
$\vartheta\left(a_{j,y} \mid d^r\right)$	Number of the $a_{j,k}$ being the optimal attack strategy at MC runs, conditioned on the d^r
$\hbar_D\left(d^r\right)$	Frequency of the d^r being the optimal portfolio at MC runs
$\varpi\left(d^r\right)$	Number of the d^r being the optimal portfolio at MC runs
$a_{Nash}^*\left(d^r\right)$	Attacker best response with respect to the d^r
$d_{Nash}^*\left(a_{j,y}\right)$	Defender best response with respect to the $a_{j,k}$

1. INTRODUCTION

Cyber-Physical Systems (CPSs) combine and coordinate physical processes with high automation level through interconnections via the cyber domain (Jazdi, 2014; Lee, 2008; Colombo et al., 2017; Wan et al., 2016). They are increasingly operated in aerospace, automotive, transportation, medical and health-care, and energy (Lee, 2008; Khaitan and McCalley, 2015; Bradley and Atkins, 2015). Specifically to nuclear energy, the introduction of digital Instrumentation and Control (I&C) systems allows Nuclear Power Plants (NPPs) to take advantage of CPSs (IAEA, 2009), for improved control and safety. However, while digitalization enhances smart systems, it can also bring new cyber risks. For example, in 2009, a malware manipulated the speed of centrifuges in a nuclear enrichment plant, causing them to spin out of control. This malware, known as Stuxnet (Langner, 2011), was introduced into a stand-alone network via flash-drives and, then, autonomously spread across the network. Attackers can pinpoint the weakest link of the system and target the most vulnerable components of a CPS to maximize the loss of system functionality (Levitin, 2007; Wang et al., 2017b).

Risk assessment of CPS must address both safety and security issues (Aven, 2009; Aven and Krohn, 2014; Zio, 2016; Zio, 2018; Kriaa et al., 2015; Piètre-Cambacédès and Bouissou, 2013; Zalewski et al., 2016; Wang et al., 2017a). Safety concerns stochastic components failures that can result in accidental scenarios leading the system towards unacceptable consequences. Security concerns malicious and intentional attacks that can impair both the physical and cyber parts of the system, and lead to unacceptable consequences. Developing a fully integrated risk assessment approach to safety and security is fundamental to address all possible failures and threats in a comprehensive and holistic way (Zio, 2018). While safety analyst relies on consolidated approaches to identify, analyze and take decisions to counteract hazards (Zio, 2016; Aven and Zio, 2011), cyber threats identification and analysis methods (including defend-attack models) are still under development.

The minimization of attacks impacts on CPS functionality and the maximization of CPS reliability and survivability are sought by defenders decisions on the allocation

of defensive resources (Bier et al., 2007; Levitin, 2007; Levitin and Hausken, 2009; Fang and Sansavini, 2017; Chen et al., 2018). A variety of defend-attack models have been proposed for this scope, focusing on the strategic interactions between defenders and attackers or/and the effectiveness of optimal defense resource allocations against adaptive cyber attacks. Graphical models (e.g., attack graphs (McQueen et al., 2006; Polatidis et al., 2018)) have been used to illustrate to a defender the proper security measures for defending the system. Potential system vulnerability paths that the attacker could exploit to gain access to a targeted cyber domain need to be identified and defended (Sheyner and Wing, 2003; McQueen et al., 2006; Shandilya et al., 2014; Ingols et al., 2006; Ge et al., 2018; Bi and Zhang, 2014). Mathematical models (e.g., *Copula*-based models (Hu et al., 2017), a trilevel *planner-attacker-defender* model based on min-max-min optimization (Fang and Sansavini, 2017)) generally rely on a game-theoretical analysis and apply it to many areas (such as economics, political science, psychology, biology, computer science, and so on (Roger, 1991; Kreps, 1990; Nisan et al., 2007)), with the goal of advising the defender on the optimal allocation of defensive resources against attackers (Chen et al., 2018; Xiang and Wang, 2017; Sun et al., 2017; Backhaus et al., 2013; Wang et al., 2017; Ezhei and Ladani, 2017; Fielder et al., 2016; Zhang J., et al., 2018; Ma et al., 2013).

However, all models mentioned above are developed from the viewpoint of a neutral opponent governing the attack/defense loss, under the strong assumptions of mutually consistent knowledge, rather than from the viewpoint of an intelligent adversary (attacker or defender) exploring the impacts of malicious (or self-interested) actions under uncertainty (Cox Jr, 2009; Rios Insua et al., 2009; Banks et al., 2015; Rothschild et al., 2012). Adversarial Risk Analysis (ARA) addresses this limitation by modeling and analyzing intelligent actors (attackers or defenders), for which the outcomes (or losses) in the game-theoretical model are uncertain (Rios Insua et al., 2009; Banks et al., 2015; Rios and Insua, 2012).

ARA has been applied to counter terrorisms, natural disasters, bidding and corporate competition. (Banks and Anderson, 2006) combined statistical risk analysis

with a zero-sum game with random payoffs to evaluate defense strategies that have been considered for the threat of smallpox. (Zhuang and Bier, 2007) applied game theory and highlighted the assumption of endogenous attacker effort to identify Nash equilibrium strategies in a defend-attack model of resource allocation for countering terrorism and natural disasters. Critical infrastructures have been the focus in (Cano et al., 2016a; 2016b) and in (Quijano et al., 2016; Insua et al., 2016), for devising security resource allocation plans from the attacks of intelligent adversaries in airports and railways, respectively. In the context of military combat modelling, the effects of military deceit and of gaining insights into aggregating longer chains of military events were modeled (Roponen and Salo, 2015). Recently, (Busby et al., 2017) analyzed the effects of cyber attacks to industrial control systems.

In this work, we propose an ARA model to advise the CPS defender, with his own beliefs and preferences, for identifying the optimal defense resource allocation that would minimize the system integrity loss when constrained by limited defense resources against (unknown and uncertain) cyber attacks. The proposed approach is illustrated, without loss of generality and for demonstration purpose, with respect to the prescriptive support it can provide to a defender within a defend-attack game, whose opponents resources and decisions are uncertain. The system considered potentially under attack is the digital I&C system of the Advanced Lead Fast Reactor European Demonstrator (ALFRED) (Alemberti et al., 2013; Ponciroli et al., 2014; Ponciroli et al., 2015).

The rest of the paper is organized as follows. Section 2 presents the main characteristics of the ALFRED with its digital I&C system, the cyber attacks it may suffer and the deployable defensive resources. Based on the proposed ARA framework, the cyber defend-attack model with respect to ALFRED is built in Section 3. In Section 4, the optimized resource allocation among the available portfolios of alternatives is provided; as a comparison, the Nash equilibrium optimal result of a classical game-theoretical analysis is also given. Conclusions are drawn in Section 5.

2. THE ADVANCED LEAD-COOLED FAST REACTOR EUROPEAN DEMONSTRATOR

We consider the protection of the ALFRED against potential cyber threats. The ALFRED reactor with its full power mode control scheme is briefly described in Section 2.1, whereas, cyber attacks and the deployable defensive resources are presented in Sections 2.2 and 2.3, respectively. The game originated between the defender and the attacker is described in Section 2.4.

2.1 The ALFRED and its digital I&C system

ALFRED is a small-size (300 MW) pool-type lead-cooled fast reactor, cooled by molten lead to ensure the favourable physical features and realize a simplified plant layout (Alemberti et al., 2013). In the ALFRED core, Control Rods (CRs) systems adjusting the heights of CRs h_{CR} have been foreseen for thermal power (P_{Th}) regulation and reactivity swing compensation during the cycle, and for scram purposes with the required reliability for a safe shutdown (Grasso et al., 2013).

At full power nominal conditions, the dynamics processing of the ALFRED primary and secondary cooling systems is controlled by a multi-loop PI (Proportional and Integral) control scheme (see Fig. 1), i.e., a decentralized control scheme, because of its simplicity of implementation and robustness to malfunctioning of the single control loops (Ponciroli et al., 2014; Ponciroli et al., 2015). Both feedback and feedforward digital control schemes are adopted for ALFRED (see Fig. 1 shadowed part). The PI-based feedback control configuration employs four SISO (Single Input Single Output) control loops independent of each other. The parameters specification of ALFRED at full power nominal conditions are reported in Table 1.

Table 1 ALFRED parameters values, at full power nominal conditions

Parameter	Parameter Description	Value	Unit
P_{Th}	Thermal power	$300 \cdot 10^6$	W
h_{CR}	Height of control rods	12.3	cm
$T_{L,hot}$	Coolant core outlet temperature	480	°C
$T_{L,cold}$	Coolant SG outlet temperature	400	°C
Γ	Coolant mass flow rate	25984	$kg \cdot s^{-1}$
T_{feed}	Feedwater SG inlet temperature	335	°C
T_{steam}	Steam SG outlet temperature	450	°C
p_{SG}	SG pressure	$180 \cdot 10^5$	Pa
G_{water}	Feedwater mass flow rate	192	$kg \cdot s^{-1}$
G_{att}	Attemperator mass flow rate	0.5	$kg \cdot s^{-1}$
kv	Turbine admission valve coefficient	1	-
P_{Mech}	Mechanical power	$146 \cdot 10^6$	W

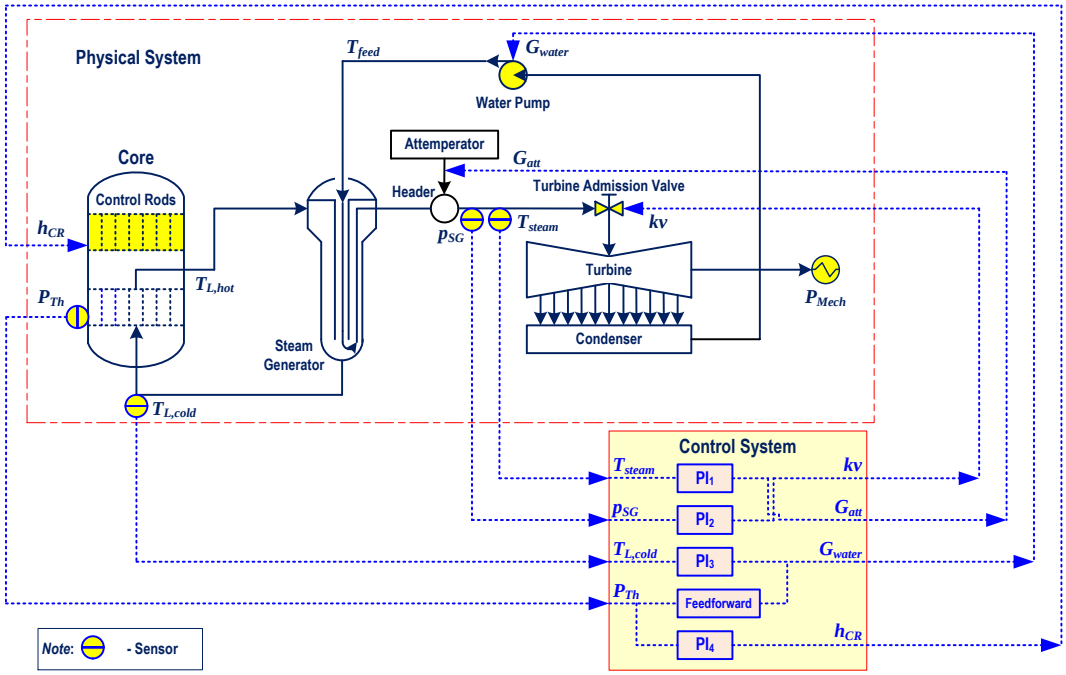

Fig. 1. ALFRED reactor control scheme

2.2 The cyber attacks

Besides components failures, CPS functionality can also be compromised by malicious attacks. Responses of the digital I&C system of ALFRED to 15 different cyber attack strategies aimed at altering sensors, actuators and PI regulators (i.e., PI gains and set point values) have been investigated in (Wang et al., 2017b). It is shown that cyber attacks to actuators challenge the most the entire system functionality, along with the attacks to the lead temperature sensor, whereas, functionality is negligibly

affected by attacks that alter the values of PI gains. This is ascribed to the PI controller capability of regulating the errors of controlled variables close to zero even if the (relatively small) gain values are changed to 3 or 4 orders of magnitude larger than the reference settings (Wang et al., 2017b). It is worth pointing out that the prioritization of cyber threats in terms of their impact on the ALFRED functionality, as proposed in (Wang et al., 2017b), are usually unknown (i.e., uncertain) to attackers, or not equally perceived by attackers and defenders, at least in reality.

In this study, a poor attacker cognitive awareness on cyber threats prioritization is assumed, resulting in a pool of A=15 different cyber attack strategies (of 4 types, as listed in Table 2, 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)) that the attacker can undergo, constrained by resources that allow him/her to launch a single attack to target a single CPS component.

An intentional attack can be launched either from an outsider or from an insider (e.g., a bribed operator) with probabilities ξ_{out} and ξ_{in} (hereafter taken equal to $\xi_{out} = 0.99$ and $\xi_{in} = 0.01$, respectively) with preparation cost (Mehetre et al., 2018; Noureddine et al., 2017):

$$c_{prep} = \begin{cases} c_{out}, & \text{if outsider attacker} \\ c_{in}, & \text{if insider attacker} \end{cases} \quad (1)$$

where, c_{out} and c_{in} are the front money for financing an outsider attacker and the bribery cost of an insider operator (hereafter assumed to be distributed as truncated normal distributions TN(5e1, 1e1) (k€) and TN(2e2, 5e1) (k€), respectively, according to the statistics listed in (Ponemon Institute, 2017)).

Attack consequences can be monetized in terms of attack loss due to c_{A1} (attacker arrest) and c_{A2} (cyber attacker remunerations of launching an attack), and attack revenues from c_{A3} (radiological effect), c_{A4} (public panic effect) and c_{A5} (media effect). The total attack cost becomes:

$$c_A = c_{prep} + \sum_{q=1}^{2} c_{Aq} - \sum_{q=3}^{5} c_{Aq} \quad (2)$$

Notice that:

(1) Loss c_{A1} (i.e., cost for an attacker arrest by a security personnel (e.g., police office)) is here estimated by:

$$c_{A1} = \xi_{arrest} \cdot c_{arrest} \quad (3)$$

where, ξ_{arrest} is the arrest probability hereafter assumed to be distributed as a Uniform distribution U(0.0, 0.5), and c_{arrest} is equal to 3e2 (k€) (Quijano et al., 2016; Viscusi and Aldy, 2003; Viscusi, 2009).

(2) The attacker remunerations c_{A2} are usually deliberated between the attacker and the employer before launching an attack, according to uncertain factors such as attacker experience, attack technical means, etc.; thus, c_{A2} is estimated to be several times larger than the front money, hereby distributed as a truncated normal distribution TN(1e3, 2e2) (k€).

(3) c_{A3}, c_{A4} and c_{A5} are the attacker revenues induced from the launched attack and mainly depend on the confrontation between the attack and any possible defensive countermeasures.

2.3 The defensive resources

A typical digital I&C system is a SCADA system that features numerous hardwares and softwares, interfacing the monitoring and control system with the physical process, aimed at controlling it and, at the same time, protecting it from cyber attacks, from which recovery is needed (in case of attack success) to maintain the system in normal operation conditions (Nazir et al., 2017; Abdo et al., 2018; Xiang et al., 2018).

Defensive resources are, therefore, aimed at: (d_1) preventing from cyber attacks and, (d_2) recovering when suffering a successful cyber attack.

Prevention can be enforced by (Nespoli et al., 2017; Xiang et al., 2018; Wang et al., 2017a; Yang et al., 2014):

- Firewall that prevents intrusions and blocks unauthorized or unwanted communications;
- Intrusion Detection Systems (IDSs) that identify common patterns of unwanted network access or malicious activities, and alert operators;
- Operators that monitor the process status through sophisticated Human-Machine Interfaces (HMIs) that embed IDSs distinguishing cyber attacks from stochastic component failures;
- Security software that prevents from operators unauthorized access and information leakage by password authentication, communication encryption, or/and access authorization;

Recovery from successful cyber attacks can rely on (Nespoli et al., 2017):

- Mainframe computers that allow the digital I&C system to run interrupted and provide correct commands to actuators, even under some types of cyber attacks (such as Internet Protocol (IP) spoofing);
- Database servers that store clean databases and can be used for recovery of data in case of some types of cyber attacks (such as false data injection);
- Security engineers that maintain the digital I&C system once exposed to cyber attack, to guarantee a secure network communication and service.

Table 3 lists the defensive resources considered, with their relevance to cope with the cyber attacks discussed in Section 2.2 (Column 3) and their minimum and maximum deployable quantity (Column 5), both assessed by expert judgment (Yang et al., 2014; Nazir et al., 2017; Xiang et al., 2018).

The annual costs of deployment of defensive resources (Column 5) are estimated on salaries (for operators and security engineers), software research and development (R&D) (for firewall and security software), and equipment costs (for IDSs, mainframe computers and database servers). In details, salaries correspond to annual base wages and pay incentives, whereas costs of software R&D and equipment are estimated as in Eqs. (4) and (5), respectively (De Roze and Nyman, 1978):

$$c_{i,k} = \frac{c_{i,k}^{R\&D}}{T_{NPP}} + c_{i,k}^{M}, \text{ if } c_{i,k} = c_{1,1}, c_{1,4} \quad (4)$$

$$c_{i,k} = \frac{c_{i,k}^{Buy}}{T_{equipment}} + c_{i,k}^{E}, \text{ if } c_{i,k} = c_{1,2}, c_{2,1}, c_{2,2} \quad (5)$$

where, $c_{i,k}$ is the annual cost of the *k*-th resource of the *i*-th defense type, $c_{i,k}^{R\&D}$ is the R&D cost, T_{NPP} is the lifetime of a ALFRED NPP, $c_{i,k}^{M}$ is its annual maintenance cost, $c_{i,k}^{Buy}$ is its purchase cost, amortized for its lifetime $T_{equipment}$ (without depreciation), and $c_{i,k}^{E}$ is the annual cost of electricity needed to run the resource.

Table 2 Cyber attack strategies: types and targets

Attack type (a_j)	Probability of attack success (probability of prevention failure), $\phi_{s_1}^j$	Probability of recovery failure (if attack success), $\phi_{s_2}^j$	Attack target, ($a_{j,y}$)			
(a_1) sensor databases	0.65	0.40	($a_{1,1}$) T_{steam}	($a_{1,2}$) p_{SG}	($a_{1,3}$) $T_{L,cold}$	($a_{1,4}$) P_{Th}
(a_2) commands of actuators	0.55	0.45	($a_{2,1}$) h_{CR}	($a_{2,2}$) G_{water}	($a_{2,3}$) kv	/
(a_3) changes of PI gain values	0.40	0.80	($a_{3,1}$) PI_1	($a_{3,2}$) PI_2	($a_{3,3}$) PI_3	($a_{3,4}$) PI_4
(a_4) changes of set point values	0.40	0.50	($a_{4,1}$) $T_{steam,set}$	($a_{4,2}$) $p_{SG,set}$	($a_{4,3}$) $T_{L,cold,set}$	($a_{4,4}$) $P_{Th,set}$

Table 3 Defensive resources with properties

Defense type, d_i	Countermeasures, $x_{i,k}$	Relevance	$x_{i,k}$ relevance with respect to (a_j), $\gamma_{i,k}^j$				Min.-Max., $n_{i,k}$	Annual cost distribution, $c_{i,k}$ (k€)
			(a_1)	(a_2)	(a_3)	(a_4)		
(d_1) Prevention	($x_{1,1}$) firewall	High	0.25	0.25	0.25	0.25	0-1	TN(80,20)
	($x_{1,2}$) Intrusion Detection Systems (IDSs)	Moderate	0.10	0.10	0.01	0.05	0-4	TN(15,2)
	($x_{1,3}$) operators	Moderate	0.10	0.10	0.01	0.05	1-4	Tri(35,50,60)
	($x_{1,4}$) security software	Moderate	0.06	0.06	0.10	0.10	0-3	TN(80,10)
(d_2) Recovery	($x_{2,1}$) mainframe computers	High	0.17	0.45	0.35	0.35	0-3	TN(520,2)
	($x_{2,2}$) database servers	High	0.25	0.25	0.05	0.15	0-2	TN(70,2)
	($x_{2,3}$) security engineers	High	0.50	0.35	0.25	0.25	0-2	Tri(90,100,110)

Notes: Tri(a,b,c) denotes a triangular distribution with lower limit a, upper limit c and mode b, and TN(μ,σ) denotes a normal distribution with mean value μ and standard deviation σ, truncated at zero.

Considering a maximum budget B_M generates a set of $\Re$ alternative defense portfolios $d^r = \{d_1^r, d_2^r\} = \{n_{1,1}^r, n_{1,2}^r, n_{1,3}^r, n_{1,4}^r, n_{2,1}^r, n_{2,2}^r, n_{2,3}^r\} \in \Re$ characterized by an annual cost c_{annual}^r, $r = 1, 2, \ldots, \Re$:

$$c_{annual}^r = c_{annual}^{d_1^r} + c_{annual}^{d_2^r} = \sum_i \sum_k n_{i,k}^r \cdot c_{i,k} \leq B_M \quad (6)$$

where, $c_{annual}^{d_1^r}$ and $c_{annual}^{d_2^r}$ are the annual costs of the d_1 and d_2 types of defensive resources of the portfolio d^r. Assuming a B_M equal to 2,000 k€ (for sake of illustration), Eq. (6) yields $\Re$ = 4834 alternative defensive resource allocations 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 $n_{2,3}^r = 0,1,2$, and $c_{i,k}$ are taken to be the mean values of the annual costs of the defensive resources. The resulting $\Re$ = 4834 deployable portfolios are hereafter referred to by the rule of sequentially increasing the values of $n_{2,3}^r, n_{2,2}^r, n_{2,1}^r, n_{1,4}^r, n_{1,3}^r, n_{1,2}^r$ and $n_{1,1}^r$, 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 $d^{4834} = \{1,4,4,3,2,2,2\}$.

2.4 The game

The defender needs to choose a defense strategy d^r from the $\Re$ available, to optimally protect the digital I&C system from an (unknown) attack strategy $a_{j,k}$ among the A that can threaten the system, originating a game between the defender and the attacker. Different combinations of defense and attack strategies $\left(\vec{d}^r, a_{j,y}\right) = \left(d_1^r, d_2^r, a_{j,y}\right)$ would result in different outcomes and consequences, with different costs for both the defender and the attacker.

Since the scope of the work is to prescriptively support the defender with an optimal resource allocation, outcomes and consequences generated from each $\left(d_1^r, d_2^r, a_{j,y}\right)$ are hereafter described only with focus on the defender decision making.

2.4.1 The outcomes probabilities

Each combination $\left(d_1^r, d_2^r, a_{j,y}\right)$ originates the outcome set $\vec{s} = \{s_1\left(d_1^r, a_{j,y}\right), s_2\left(d_2^r \mid s_1\right)\}$, where $s_1\left(d_1^r, a_{j,y}\right)$ defines the successful prevention of d_1^r to an attack $a_{j,y}$:

$$s_1(d_1^r, a_{j,y}) = \begin{cases} 1, & prevention\ failure\ (attack\ success) \\ 0, & prevention\ success\ (attack\ failure) \end{cases} \quad (7)$$

and $s_2\left(d_2^r \mid s_1\right)$ the successful recovery of d_2^r in case of successful attack (i.e., $s_1\left(d_1^r, a_{j,y}\right) = 1$):

$$s_2(d_2^r|s_1) = \begin{cases} 1, & recovery\ success \\ 0, & recovery\ failure \end{cases} \quad (8)$$

The outcome set comes with a probability set $\vec{p} = \left\{p\left(s_1\left(d_1^r, a_{j,y}\right)\right), p\left(s_2\left(d_2^r \mid s_1\right)\right)\right\}$, where $p\left(s_1\left(d_1^r, a_{j,y}\right)\right)$ defines the probability of the prevention outcome $s_1\left(d_1^r, a_{j,y}\right)$, and $p\left(s_2\left(d_2^r \mid s_1\right)\right)$ the probability of the recovery outcome $s_2\left(d_2^r \mid s_1\right)$ in case of successful attack (i.e., $s_1\left(d_1^r, a_{j,y}\right) = 1$). As proposed in (Quijano et al., 2016), the values of $p\left(s_1\left(d_1^r, a_{j,y}\right)\right)$ and $p\left(s_2\left(d_2^r \mid s_1\right)\right)$ are calculated as in Eqs. (9) and (10), respectively,

$$p\left(s_1\left(d_1^r, a_{j,y}\right)\right) = \begin{cases} \phi_{s_1}^j \cdot \exp(-\sum_k n_{1,k}^r \cdot \gamma_{1,k}^j); & s_1\left(d_1^r, a_{j,y}\right) = 1 \\ 1 - \phi_{s_1}^j \cdot \exp(-\sum_k n_{1,k}^r \cdot \gamma_{1,k}^j); & s_1\left(d_1^r, a_{j,y}\right) = 0 \end{cases} \quad (9)$$

$$p\left(s_2\left(d_2^r \mid s_1\right)\right) = \begin{cases} 1 - \phi_{s_2}^j \cdot \exp(-\sum_k n_{2,k}^r \cdot \gamma_{2,k}^j); & s_2\left(d_2^r \mid s_1\right) = 1,\ s_1\left(d_1^r, a_{j,y}\right) = 1 \\ \phi_{s_2}^j \cdot \exp(-\sum_k n_{2,k}^r \cdot \gamma_{2,k}^j); & s_2\left(d_2^r \mid s_1\right) = 0,\ s_1\left(d_1^r, a_{j,y}\right) = 1 \\ 1; & s_1\left(d_1^r, a_{j,y}\right) = 0 \end{cases} \quad (10)$$

where, $\phi_{s_1}^j$ is the probability of prevention failure when the *j*-th type attack is occurring, $\phi_{s_2}^j$ is the probability of recovery failure when the *j*-th type attack is successful (see Table 2), $\gamma_{1,k}^j$ is the estimated relevance of the *k*-th prevention countermeasure in decreasing the attack success probability, and $\gamma_{2,k}^j$ is the estimated relevance parameter of the *k*-th recovery measure in increasing the recovery success probability (see Table

3).

Being all these parameters estimated by the defender on his personal judgment, Eqs. (9) and (10) are the defender opinion on the outcomes probabilities, i.e., $p_D\left(s_1\left(d_1^r, a_{j,y}\right)\right)$ and $p_D\left(s_2\left(d_2^r \mid s_1\right)\right)$. However, the defender ignores the attacker assumptions on the probabilities of the outcomes, i.e., $p_A\left(s_2 \mid d_2^r, s_1\right)$ and $p_A\left(s_1 \mid d_1^r, a_{j,y}\right)$, and can only speculate assuming them to be distributed as normal distributions with $p_D\left(s_2 \mid d_2^r, s_1\right)$ and $p_D\left(s_1 \mid d_1^r, a_{j,y}\right)$ as mean values, and Eqs. (11) and (12) as standard deviations (Quijano et al., 2016),

$$\sigma_A\left(s_2 \mid d_2^r, s_1\right) = \min\left(p_D\left(s_2 \mid d_2^r, s_1\right), 0.05\right) \tag{11}$$

$$\sigma_A\left(s_1 \mid d_1^r, a_{j,y}\right) = \min\left(p_D\left(s_1 \mid d_1^r, a_{j,y}\right), 0.05\right) \tag{12}$$

2.4.2 The attack consequences

Consequences of attacks are monetized in terms of economic loss (i.e., for c_{D1} system integrity loss and c_{D2} decrease of P_{Mech}) and compensation for post-attack impact (i.e., c_{D3} radiological effects, c_{D4} public panic and chaos, and c_{D5} media impact) (see Table 4) (Zou, 2017; Wurm et al., 2017).

Table 4 Consequences of attacks

Consequences	Description
(c_{D1}) System integrity loss	CPS recovery and protection improvement
(c_{D2}) Decrease of P_{Mech}	Business interruption
(c_{D3}) Radiological effects	Compensation for radiation pollution
(c_{D4}) Public panic and chaos	Social network reconstruction
(c_{D5}) Media impact	Public relation management

For simplicity, and in line with (Wang et al., 2017b; Cano et al., 2016a), the costs of the l-th consequence $c_{Dl}\left(\vec{s} \mid d^r, a_{j,y}\right)$, l=1, 2, …, 5, that depend on $\left(d^r, a_{j,y}\right)$ and on the outcomes $\vec{s}$, is calculated according to the law of total probability (Modarres, 2016):

$$c_{Dl}\left(\vec{s} \mid d^r, a_{j,y}\right) = \sum_{\alpha} p_l^{\alpha}\left(\vec{s} \mid a_{j,y}\right) \cdot c_{Dl}^{\alpha} \tag{13}$$

where, c_{Dl}^{α} is assumed to be the cost of a negligible (α=N), medium (α=M) and severe

(α=*S*) attack (listed in Table 5), $p_l^{\alpha}\left(\vec{s} \mid a_{j,y}\right)$ is the conditional probability of the outcome $\vec{s}$ to the occurrence of $a_{j,y}$ with the α-th effect of the *l*-th consequence, and:

$$\sum_{\alpha} p_l^{\alpha}\left(\vec{s} \mid a_{j,y}\right)=1 \tag{14}$$

The defender assesses $p_l^{\alpha}\left(\vec{s} \mid a_{j,y}\right)$ according to the effects an attack $a_{j,y}$ may have on the system functionality. For this scope, in this work, we rely on the safety margins estimates of the ALFRED under different cyber attacks (listed in Tables 9, 11 and 13 in (Wang et al., 2017b)). In general terms, the smaller the safety margin to cyber attack, the more probable a *l*-th consequence with severe (α=*S*) effect. In Table 6, the defender assumptions for $p_l^{\alpha}\left(\vec{s} \mid a_{j,y}\right)$ (for system integrity loss (I), decrease of P_{Mech} (II), radiological effects (III), public panic and chaos (IV) and media impact (V)) are listed.

Table 5 Defender's assessment of base costs of the consequences

Consequences	Cost of each level (k€)		
	Negligible	Medium	Severe
	c_{Dl}^{N}	c_{Dl}^{M}	c_{Dl}^{S}
(c_{D1}) System integrity loss	1e1	4e2	1e3
(c_{D2}) Decrease of P_{Mech}	91.11	91.11*24	91.11*1e2
(c_{D3}) Radiological effect	0	1e4	1e6
(c_{D4}) Public panic and chaos	1e1	1e3	1e4
(c_{D5}) Media impact	1e1	1e3	1e4

Notes: 91.11 (k€/h) is an estimate of the economics profit per hour for a 300MW nuclear reactor.

Table 6 Defender assessment of the probabilities of occurrence of each consequence level

(I) System integrity loss

Probabilities	Attack strategies														
	$a_{1,1}$	$a_{1,2}$	$a_{1,3}$	$a_{1,4}$	$a_{2,1}$	$a_{2,2}$	$a_{2,3}$	$a_{3,1}$	$a_{3,2}$	$a_{3,3}$	$a_{3,4}$	$a_{4,1}$	$a_{4,2}$	$a_{4,3}$	$a_{4,4}$
$p_1^N\left(s_1=0\mid a_{j,y}\right)$	1.00	0.80	0.75	0.70	0.75	0.65	0.75	1.00	1.00	1.00	1.00	1.00	0.85	0.80	0.85
$p_1^M\left(s_1=0\mid a_{j,y}\right)$	0.00	0.17	0.20	0.25	0.20	0.30	0.22	0.00	0.00	0.00	0.00	0.00	0.15	0.20	0.15
$p_1^S\left(s_1=0\mid a_{j,y}\right)$	0.00	0.03	0.05	0.05	0.05	0.05	0.03	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
$p_1^N\left(s_2=1,s_1=1\mid a_{j,y}\right)$	1.00	0.55	0.40	0.35	0.50	0.30	0.50	1.00	1.00	0.90	1.00	1.00	0.65	0.40	0.65
$p_1^M\left(s_2=1,s_1=1\mid a_{j,y}\right)$	0.00	0.40	0.50	0.55	0.50	0.50	0.50	0.00	0.00	0.10	0.00	0.00	0.35	0.50	0.35
$p_1^S\left(s_2=1,s_1=1\mid a_{j,y}\right)$	0.00	0.05	0.10	0.10	0.00	0.20	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.10	0.00
$p_1^N\left(s_2=0,s_1=1\mid a_{j,y}\right)$	1.00	0.99	0.99	0.99	0.99	0.99	0.99	1.00	1.00	1.00	1.00	1.00	0.99	0.99	0.99
$p_1^M\left(s_2=0,s_1=1\mid a_{j,y}\right)$	0.00	0.01	0.01	0.01	0.01	0.01	0.01	0.00	0.00	0.00	0.00	0.00	0.01	0.01	0.01
$p_1^S\left(s_2=0,s_1=1\mid a_{j,y}\right)$	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

(II) Decrease of P_{Mech}

Probabilities	Attack strategies														
	$a_{1,1}$	$a_{1,2}$	$a_{1,3}$	$a_{1,4}$	$a_{2,1}$	$a_{2,2}$	$a_{2,3}$	$a_{3,1}$	$a_{3,2}$	$a_{3,3}$	$a_{3,4}$	$a_{4,1}$	$a_{4,2}$	$a_{4,3}$	$a_{4,4}$
$p_2^N\left(s_1=0\mid a_{j,y}\right)$	1.00	0.80	0.75	0.70	0.75	0.65	0.75	1.00	1.00	1.00	1.00	1.00	0.85	0.80	0.85
$p_2^M\left(s_1=0\mid a_{j,y}\right)$	0.00	0.17	0.20	0.25	0.20	0.30	0.22	0.00	0.00	0.00	0.00	0.00	0.15	0.20	0.15
$p_2^S\left(s_1=0\mid a_{j,y}\right)$	0.00	0.03	0.05	0.05	0.05	0.05	0.03	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
$p_2^N\left(s_2=1,s_1=1\mid a_{j,y}\right)$	1.00	0.55	0.40	0.35	0.50	0.30	0.50	1.00	1.00	0.90	1.00	1.00	0.65	0.40	0.65
$p_2^M\left(s_2=1,s_1=1\mid a_{j,y}\right)$	0.00	0.40	0.50	0.55	0.50	0.50	0.50	0.00	0.00	0.10	0.00	0.00	0.35	0.50	0.35
$p_2^S\left(s_2=1,s_1=1\mid a_{j,y}\right)$	0.00	0.05	0.10	0.10	0.00	0.20	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.10	0.00
$p_2^N\left(s_2=0,s_1=1\mid a_{j,y}\right)$	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
$p_2^M\left(s_2=0,s_1=1\mid a_{j,y}\right)$	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
$p_2^S\left(s_2=0,s_1=1\mid a_{j,y}\right)$	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

(III) Radiological effects

Probabilities	Attack strategies														
	$a_{1,1}$	$a_{1,2}$	$a_{1,3}$	$a_{1,4}$	$a_{2,1}$	$a_{2,2}$	$a_{2,3}$	$a_{3,1}$	$a_{3,2}$	$a_{3,3}$	$a_{3,4}$	$a_{4,1}$	$a_{4,2}$	$a_{4,3}$	$a_{4,4}$
$p_3^N\left(s_1=0\mid a_{j,y}\right)$	1.00	0.95	0.95	0.95	0.95	0.95	0.95	1.00	1.00	1.00	1.00	1.00	0.95	0.95	0.95
$p_3^M\left(s_1=0\mid a_{j,y}\right)$	0.00	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.00	0.00	0.00	0.00	0.05	0.05	0.05
$p_3^S\left(s_1=0\mid a_{j,y}\right)$	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
$p_3^N\left(s_2=1,s_1=1\mid a_{j,y}\right)$	1.00	0.75	0.60	0.80	0.50	0.50	0.50	1.00	1.00	1.00	1.00	1.00	0.95	0.60	0.95
$p_3^M\left(s_2=1,s_1=1\mid a_{j,y}\right)$	0.00	0.20	0.30	0.15	0.40	0.40	0.40	0.00	0.00	0.00	0.00	0.00	0.05	0.30	0.05
$p_3^S\left(s_2=1,s_1=1\mid a_{j,y}\right)$	0.00	0.05	0.10	0.05	0.10	0.10	0.10	0.00	0.00	0.00	0.00	0.00	0.00	0.10	0.00
$p_3^N\left(s_2=0,s_1=1\mid a_{j,y}\right)$	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
$p_3^M\left(s_2=0,s_1=1\mid a_{j,y}\right)$	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
$p_3^S\left(s_2=0,s_1=1\mid a_{j,y}\right)$	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

(IV) Public panic and chaos

Probabilities	Attack strategies														
	$a_{1,1}$	$a_{1,2}$	$a_{1,3}$	$a_{1,4}$	$a_{2,1}$	$a_{2,2}$	$a_{2,3}$	$a_{3,1}$	$a_{3,2}$	$a_{3,3}$	$a_{3,4}$	$a_{4,1}$	$a_{4,2}$	$a_{4,3}$	$a_{4,4}$
$p_4^N\left(s_1=0\mid a_{j,y}\right)$	1.00	0.90	0.85	0.90	0.70	0.75	0.70	1.00	1.00	1.00	1.00	1.00	0.95	0.90	0.90
$p_4^M\left(s_1=0\mid a_{j,y}\right)$	0.00	0.10	0.10	0.07	0.25	0.20	0.25	0.00	0.00	0.00	0.00	0.00	0.05	0.10	0.10
$p_4^S\left(s_1=0\mid a_{j,y}\right)$	0.00	0.00	0.05	0.03	0.05	0.05	0.05	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
$p_4^N\left(s_2=1,s_1=1\mid a_{j,y}\right)$	1.00	0.80	0.70	0.80	0.50	0.40	0.40	1.00	1.00	1.00	1.00	1.00	0.95	0.60	0.60
$p_4^M\left(s_2=1,s_1=1\mid a_{j,y}\right)$	0.00	0.15	0.25	0.17	0.40	0.50	0.40	0.00	0.00	0.00	0.00	0.00	0.05	0.20	0.20
$p_4^S\left(s_2=1,s_1=1\mid a_{j,y}\right)$	0.00	0.05	0.05	0.03	0.10	0.10	0.20	0.00	0.00	0.00	0.00	0.00	0.00	0.20	0.20
$p_4^N\left(s_2=0,s_1=1\mid a_{j,y}\right)$	1.00	0.98	0.98	0.98	0.95	0.95	0.95	1.00	1.00	1.00	1.00	1.00	0.98	0.95	0.95
$p_4^M\left(s_2=0,s_1=1\mid a_{j,y}\right)$	0.00	0.02	0.02	0.02	0.04	0.04	0.04	0.00	0.00	0.00	0.00	0.00	0.02	0.04	0.03
$p_4^S\left(s_2=0,s_1=1\mid a_{j,y}\right)$	0.00	0.00	0.00	0.00	0.01	0.01	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.02

(V) Media impact

Probabilities	Attack strategies														
	$a_{1,1}$	$a_{1,2}$	$a_{1,3}$	$a_{1,4}$	$a_{2,1}$	$a_{2,2}$	$a_{2,3}$	$a_{3,1}$	$a_{3,2}$	$a_{3,3}$	$a_{3,4}$	$a_{4,1}$	$a_{4,2}$	$a_{4,3}$	$a_{4,4}$
$p_5^N\left(s_1=0\mid a_{j,y}\right)$	0.99	0.85	0.60	0.60	0.50	0.50	0.60	0.98	0.98	0.98	0.98	0.95	0.95	0.80	0.70
$p_5^M\left(s_1=0\mid a_{j,y}\right)$	0.01	0.12	0.30	0.25	0.30	0.35	0.30	0.02	0.02	0.02	0.02	0.05	0.05	0.15	0.25
$p_5^S\left(s_1=0\mid a_{j,y}\right)$	0.00	0.03	0.10	0.05	0.20	0.15	0.10	0.00	0.00	0.00	0.00	0.00	0.00	0.05	0.05
$p_5^N\left(s_2=1,s_1=1\mid a_{j,y}\right)$	0.99	0.30	0.20	0.30	0.10	0.10	0.10	0.98	0.98	0.98	0.98	0.90	0.90	0.60	0.60
$p_5^M\left(s_2=1,s_1=1\mid a_{j,y}\right)$	0.01	0.65	0.70	0.65	0.70	0.70	0.70	0.02	0.02	0.02	0.02	0.10	0.10	0.30	0.30
$p_5^S\left(s_2=1,s_1=1\mid a_{j,y}\right)$	0.00	0.05	0.10	0.05	0.20	0.20	0.20	0.00	0.00	0.00	0.00	0.00	0.00	0.10	0.10
$p_5^N\left(s_2=0,s_1=1\mid a_{j,y}\right)$	1.00	0.99	0.99	0.99	0.95	0.95	0.95	1.00	1.00	1.00	1.00	1.00	0.98	0.95	0.95
$p_5^M\left(s_2=0,s_1=1\mid a_{j,y}\right)$	0.00	0.01	0.01	0.01	0.04	0.04	0.04	0.00	0.00	0.00	0.00	0.00	0.02	0.04	0.03
$p_5^S\left(s_2=0,s_1=1\mid a_{j,y}\right)$	0.00	0.00	0.00	0.00	0.01	0.01	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.02

In conclusion, with respect to a generic $\left(d^r, a_{j,y}\right)$ the total cost to be considered for decision-making consists in the defenses deployment cost c^r_{annual} of Eq. (6) plus the sum of all possible consequences costs $c_{Dl}\left(\vec{s} \mid d^r, a_{j,y}\right)$, l=1, 2, 3, 4, 5:

$$c_D\left(\vec{s} \mid d^r, a_{j,y}\right) = c^r_{annual} + \sum_{l=1}^{5} c_{Dl}\left(\vec{s} \mid d^r, a_{j,y}\right) \tag{15}$$

The attacker must sustain the attack impact (i.e., c_{D3}, c_{D4} and c_{D5}), that has to be justified in light of the speculated costs resulting from the game, i.e., c_{A3}, c_{A4} and c_{A5}, by:

$$c_{Aq}\left(\vec{s} \mid d^r, a_{j,y}\right) = \sum_{\alpha} p^{\alpha}_q\left(\vec{s} \mid a_{j,y}\right) \cdot c^{\alpha}_{Aq}, \quad q = 3,4,5 \tag{16}$$

where c^{α}_{Aq} is the cost of a negligible (α=N), medium (α=M) and severe (α=S) attack (see Table 7 their personal distributions) and $p^{\alpha}_q\left(\vec{s} \mid a_{j,y}\right)$ is the conditional probability of the outcome $\vec{s}$ to the occurrence of $a_{j,y}$ with the α-th effect of the q-th consequence. Thus, the attacker costs of Eq. (2) become:

$$c_A\left(\vec{s} \mid d^r, a_{j,y}\right) = c_{prep} + c_{A1} + c_{A2} + \sum_{q=3}^{5} c_{Aq}\left(\vec{s} \mid d^r, a_{j,y}\right) \tag{17}$$

Table 7 Assessment of the distributions of base costs of the consequences

Consequences (q)	Cost of each level c^{α}_{Aq} (k€)		
	Negligible	Medium	Severe
	c^{N}_{Aq}	c^{M}_{Aq}	c^{S}_{Aq}
(iii) Radiological effect	0	TN(1e4,5e3)	TN(1e6,5e5)
(iv) Panic effect	0	TN(1e3,5e2)	TN(1e4,5e3)
(v) Media effect	0	TN(1e4,5e3)	TN(1e5,5e4)

In decision analysis, it is common to map the cost into a utility function that measures the decision maker preference on alternatives with uncertain outcomes (Bernoulli, 2011; Von Neumann and Morgenstern, 2007). The decision maker aims at optimizing his/her portfolio by maximizing his/her own utility function (Grechuk and Zabarankin, 2016; Bricha and Nourelfath, 2013). In our case, the defender may use the exponential utility $u_D\left(\vec{s} \mid d^r, a_{j,y}\right)$ of Eq. (18), that is a risk averse function that lowers

the uncertainty of consequences (i.e., costs) assuming constant absolute risk with the coefficient of risk aversion $\Lambda_D = -u_D''\left(\vec{s} \mid d^p, a_{j,y}\right) / u_D'\left(\vec{s} \mid d^p, a_{j,y}\right) < 0$ (Hershey and Schoemaker, 1985; Banks et al., 2015):

$$u_D\left(\vec{s} \mid d^r, a_{j,y}\right) = -\exp\left(k_D \cdot c_D\left(\vec{s} \mid d^r, a_{j,y}\right)\right) \tag{18}$$

where, k_D (here taken distributed as U(1e-5, 2e-5)) is defined according to $k_D = u_D''\left(\vec{s} \mid d^p, a_{j,y}\right) / u_D'\left(\vec{s} \mid d^p, a_{j,y}\right) > 0$, whose absolute value is constant with respect to costs and larger than 0 (Pratt, J.W., 1964; Cox and Sadiraj, 2006).

Relying on the concept of exponential utility with attacker's risk proneness attitude and coefficient of risk proneness $\Lambda_A = -u_A''\left(\vec{s} \mid d^r, a_{j,y}\right) / u_A'\left(\vec{s} \mid d^r, a_{j,y}\right) > 0$ (Hershey and Schoemaker, 1985; Banks et al., 2015; Pratt, J.W., 1964; Cox and Sadiraj, 2006), the defender can assess the attacker's utility $u_A\left(\vec{s} \mid d^r, a_{j,y}\right)$ as:

$$u_A\left(\vec{s} \mid d^r, a_{j,y}\right) = \exp\left(k_A \cdot c_A\left(\vec{s} \mid d^r, a_{j,y}\right)\right) \tag{19}$$

where, k_A is estimated based on the absolute risk proneness constant with respect to costs and $k_A = u_A''\left(\vec{s} \mid d^r, a_{j,y}\right) / u_A'\left(\vec{s} \mid d^r, a_{j,y}\right) < 0$, judged to be distributed from a uniform distribution U(-3.0e-5, 0) with the aid of experts.

3. THE DEFEND-ATTACK MODEL

An ARA model (Rios Insua et al., 2009; Banks et al., 2015; Rios and Insua, 2012) is here tailored to a problem of cyber security assessment of the digital I&C system of the ALFRED, for supporting the defender to allocate the optimal defenses under uncertain adversarial strategies and consequences of attacks. The resulting ARA model is also compared with a Nash equilibrium optimal solution of a classical game-theoretical analysis (Osborne and Rubinstein, 1994; Zhuang and Bier, 2007; Zhuang and Bier, 2011), where uncertainties are neglected.

3.1 The ARA model

ARA builds on game theory and statistical risk analysis, in order to advise one

player/agent on some uncertain adversarial decision situations against the other(s) (Rios Insua et al., 2009; Banks et al., 2015). ARA can realize a more realistic game, thanks to the weakened common knowledge assumption that the player/agent can only know his/her own beliefs of costs, utilities and consequences of the game, and only speculate those of the other(s) via statistical risk analysis.

In particular, a defender is given advice on the optimal defense portfolio against cyber attacks, when only acquainted with subjective (partial) knowledge on attacker decisions.

Considering the outcomes of the game $\vec{s} = \{s_1\left(d_1^r, a_{j,y}\right), s_2\left(d_2^r \mid s_1\right)\}$ in the decision making, the defender seeks for the optimal resource allocation $d^* = \left\{d_1^*, d_2^*\right\}$ that is expected to optimally prevent the digital I&C system from unknown cyber attacks and, at the same time, minimize the system functionality loss in case of successful cyber attack. The d^* is obtained by maximizing the defender expected utility $\psi_D\left(d^r\right)$:

$$d^* = \arg\max_{d^r \in \Re} \psi_D\left(d^r\right) \tag{20}$$

where $\psi_D\left(d^r\right)$ is defined as in Eq. (21):

$$\psi_D\left(d^r\right) = \sum_{a_{j,y} \in A}\left[\sum_{s_2 \in \{0,1\}}\sum_{s_1 \in \{0,1\}} \pi_D\left(a_{j,y} \mid d^r\right) \cdot p_D\left(s_2 \mid d_2^r, s_1\right) \cdot p_D\left(s_1 \mid d_1^r, a_{j,y}\right) \cdot u_D\left(\vec{s} \mid d^r, a_{j,y}\right)\right] \tag{21}$$

where $u_D\left(\vec{s} \mid d^r, a_{j,y}\right)$ is the defender utility of possible consequences costs, estimated by Eq. (16), $\vec{p}_D = \left\{p_D\left(s_1\left(d_1^r, a_{j,y}\right)\right), p_D\left(s_2\left(d_2^r \mid s_1\right)\right)\right\}$ defines the defender assumptions on the probabilities of the outcomes (i.e., $s_1\left(d_1^r, a_{j,y}\right)$ and $s_2\left(d_2^r \mid s_1\right)$), obtained according to Eqs. (9) and (10), and $\pi_D\left(a_{j,y} \mid d^r\right)$ is the defender estimation of the probability of occurrence of any $a_{j,k}$ attack, given that the defense resources d^r are deployed.

To cope with the uncertainty on the type of attack (unknown to the defender) the Monte Carlo (MC) approach sketched in Fig. 2 is used for (a) estimating the $\pi_D\left(a_{j,y} \mid d^r\right)$ that is fundamental for (b) estimating the defender optimal defense

strategy d^*.

(a) Estimation of $\pi_D\left(a_{j,y} \mid d^r\right)$

The shadowed loop of Fig. 2 (left) allows to mimic N_m different attacker decisions, and propagate the defender uncertainty on these decisions, with respect to a specific deployable d^r (from $d^1 = \{0,0,1,0,0,0,0\}$ to $d^{4834} = \{1,4,4,3,2,2,2\}$). At the m-th run, $m = 1, 2, \ldots, N_m$:

(a1) For each combination $\left(d_1^r, d_2^r, a_{j,y}\right)$ given a d^r, sample the values of c_{prep}, c_{Aq} and k_A from the defender subjective distributions in Section 2, to calculate the attacker consequences of costs $c_A\left(\vec{s} \mid d^r, a_{j,y}\right)$ of Eq. (17) and the corresponding utilities $u_A\left(\vec{s} \mid d^r, a_{j,y}\right)$ of Eq. (19);

(a2) After sampling $p_A\left(s_1 \mid d_1^r, a_{j,y}\right)$ and $p_A\left(s_2 \mid d_2^r, s_1\right)$ from the defender subjective distributions in Section 2.4, calculate the attacker expected utility of $a_{j,k}$ conditioned on the d^r, $\psi_A^m\left(a_{j,y} \mid d^r\right)$, by:

$$\psi_A^m\left(a_{j,y} \mid d^r\right) = \sum_{s_2 \in \{0,1\}} \sum_{s_1 \in \{0,1\}} p_A\left(s_2 \mid d_2^r, s_1\right) \cdot p_A\left(s_1 \mid d_1^r, a_{j,y}\right) \cdot u_A\left(\vec{s} \mid d^r, a_{j,y}\right) \tag{22}$$

(a3) Find the optimal attack strategy $a^{*,m}\left(d^r\right)$, with respect to the d^r:

$$a^{*,m}\left(d^r\right) = \arg\max_{a_{j,y} \in A} \psi_A^m\left(a_{j,y} \mid d^r\right) \tag{23}$$

(a4) Run N_m = 1000 time steps (a1) to (a3), to calculate $\vartheta\left(a_{j,y} \mid d^r\right)$ the number of $a_{j,k}$ being the optimal attack strategy at all the N_m runs, given the d^r;

(a5) Estimate $\pi_D\left(a_{j,y} \mid d^r\right)$ by:

$$\pi_D\left(a_{j,y} \mid d^r\right) = \frac{\vartheta\left(a_{j,y} \mid d^r\right)}{N_m} \tag{24}$$

(b) Estimation of d^* by a MC simulation

At the v-th run, $v = 1, 2, \ldots, N_v$, of the MC simulation of Fig. 2 (right),

(b1) For each one of the set of $\Re$ (=4834) defense portfolios, d^r, take the values

of $\pi_D\left(a_{j,y} \mid d^r\right)$ (see (a)), with respect to each type of attacks $a_{j,y}$.

(b2) For each combination $\left(d_1^r, d_2^r, a_{j,y}\right)$ given the d^r, sample the values of c_{annual}^r, $c_{Dl}\left(\vec{s} \mid d^r, a_{j,y}\right)$, and k_D from the defender subjective distributions, respectively; taking the values of $\pi_D\left(a_{j,y} \mid d^r\right)$ of (a) and $\vec{p}_D = \left\{p_D\left(s_1\left(d_1^r, a_{j,y}\right)\right), p_D\left(s_2\left(d_2^r \mid s_1\right)\right)\right\}$ of Eqs. (9) and (10), calculate the defender expected utility $\psi_D^v\left(d^r\right)$ by Eq. (21);

(b3) After calculating $\psi_D^v\left(d^r\right)$ for all the portfolios d^r at the v-th run, find the optimal one by Eq. (20) that is equivalent to:

$$d^{*,v} = \arg\max_{d^r \in \Re} \psi_D^v\left(d^r\right) \tag{25}$$

(b4) Run $N_v = 1000$ times steps (b1) to (b3) to build the empirical $\hbar_D\left(d^r\right)$, which is the frequency of d^r being the optimal portfolio in all N_v runs;

(b5) Obtain the defender optimal resource allocation d^* that is:

$$d^* = \arg\max_{d^r \in \Re} \hbar_D\left(d^r\right) = \arg\max_{d^r \in \Re} \frac{\varpi\left(d^r\right)}{N_v} \tag{26}$$

where, $\varpi\left(d^r\right)$ is the number of times the d^r is the optimal portfolio in all N_v runs.

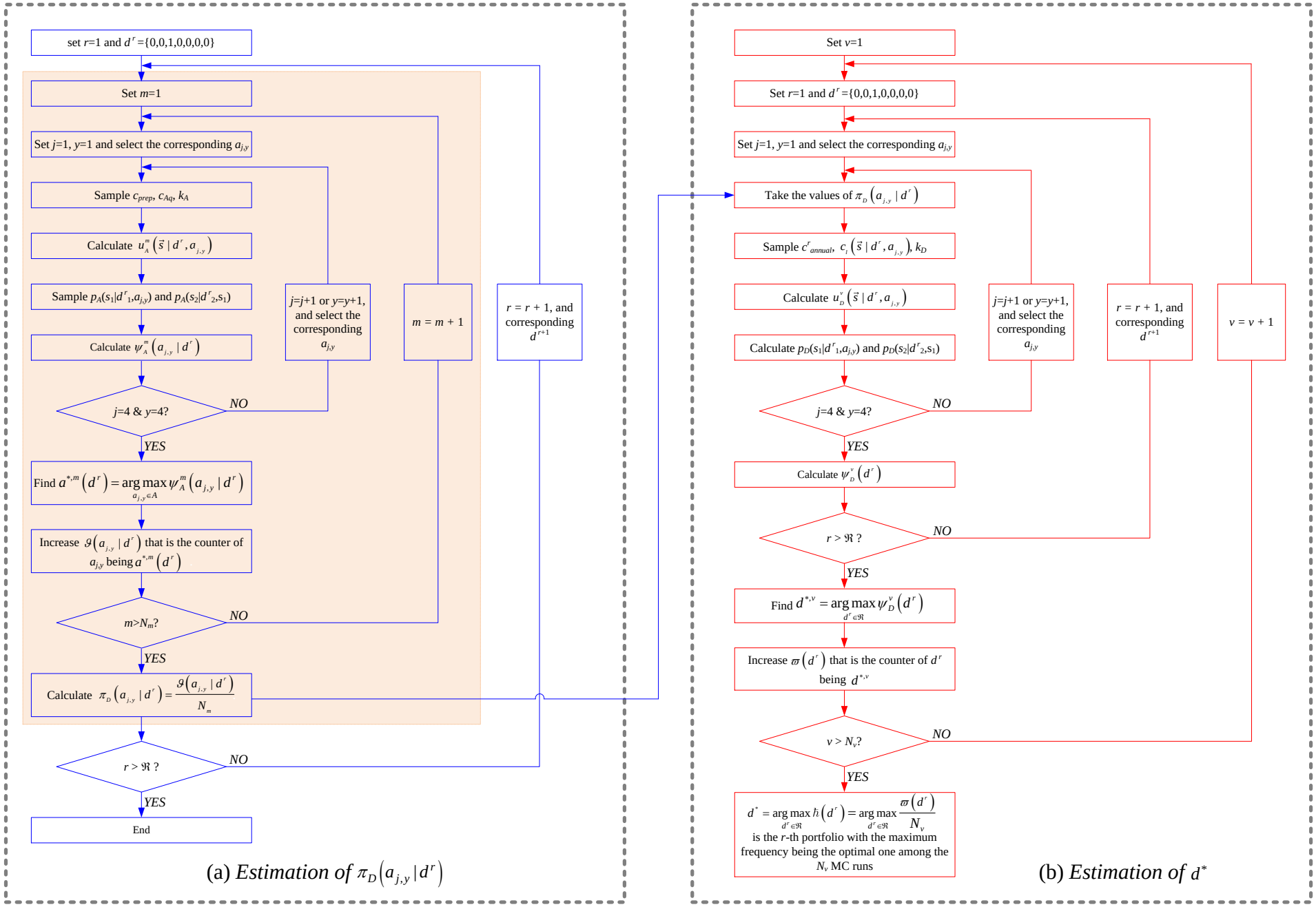

Fig. 2. The flowchart of the ARA approach for obtaining the optimal defense allocation

3.2 The classical defend-attack model (Nash equilibrium)

In most applications of traditional game theory (Zhuang and Bier, 2007; Hausken and Levitin, 2009; Zhang J., et al., 2018; Zhang C., et al., 2018; Moayedi et al., 2012; Piccinelli et al., 2017), the attacker and the defender are assumed to share common knowledge regarding utility functions and probabilities of outcomes. Such assumption allows combining defender and attacker decision analysis into a coupled (balanced) model.

As proposed in (Zhuang and Bier, 2007), taking either player strategies and beliefs as given in the coupled defend-attack model, decision analysis allows reaching one player best response with respect to each of the opponent strategies and seeking an intersection point, namely, a Nash equilibrium, $\left(d^*_{Nash}, a^*_{Nash}\right)$, that satisfies:

$$\psi_D\left(d^*_{Nash}, a^*_{Nash}\right) = \max_{d^r \in \Re} \psi_D\left(d^r, a^*_{Nash}\right) \quad \& \quad \psi_A\left(d^*_{Nash}, a^*_{Nash}\right) = \max_{a_{j,y} \in A} \psi_A\left(d^*_{Nash}, a_{j,y}\right) \qquad (27)$$

where, $a^*_{Nash}\left(d^r\right)$ is the attacker best response with respect to a defender decision d^r and $d^*_{Nash}\left(a_{j,y}\right)$ is the defender best response with respect to an attacker strategy $a_{j,y}$, and they are obtained by Eqs. (28) and (29), respectively:

$$a^*_{Nash}\left(d^r\right) = \max_{a_{j,y} \in A} \psi_A\left(a_{j,y} \mid d^r\right) \qquad (28)$$

$$d^*_{Nash}\left(a_{j,y}\right) = \max_{d^r \in \Re} \psi_D\left(d^r \mid a_{j,y}\right) \qquad (29)$$

The Nash equilibrium $\left(d^*_{Nash}, a^*_{Nash}\right)$ is commonly obtained as a combinatorial solution at which the defender and the attacker find a balanced strategy with the other, whereas, neither the defender nor the attacker can benefit from changing strategy with the other keeping the strategy unchanged (Rios Insua et al., 2009; Zhuang and Bier, 2007; Osborne and Rubinstein, 1994).

4. RESULTS

4.1 Optimal defense allocation by ARA

In ARA assessment, defender's beliefs on the launching of an attack $a_{j,y}$ given a defense portfolio d^r, $\pi_D\left(a_{j,y} \mid d^r\right)$, can be estimated by MC simulation as in Fig. 2(a).

On this basis, the defender optimal defense portfolio d^* is assesses by MC simulation as in Fig. 2(b), for taking into account the defender uncertainty on his/her predictive judgment on the countermeasure annual costs, the monetized consequences after attacks and the probabilities of outcomes.

As illustrative example, Fig. 3 shows one run of the N_v estimates of the defender expected utilities of d^r (dots): the optimal defense portfolio is estimated as $d^{*,v} = \{1,3,4,2,2,2,2\}$ (diamond) with the (absolute) lowest value of expected utilities $\psi_D\left(d^{*,v}\right)$ equal to 1.0753 (i.e., the defender reaches the setting of lowest expected investment against the uncertain attacks, with an attitude of risk aversion) and the countermeasures annual costs equal to 1,865 k€.

It can be seen that many other portfolios reach expected utilities close to $\psi_D\left(d^*\right)$. In Table 8, the top five defense portfolios with highest utility values are listed. As a matter of fact, even though the utility estimates are similar, countermeasures portfolios change much, supporting the need of a robust approach (as that described in Section 3.1(b)) to provide the optimal result with the needed confidence.

Table 8 The optimal defense portfolios with annual costs

d^r	$x_{1,1}$ $n^r_{1,1}$	$x_{1,2}$ $n^r_{1,2}$	$x_{1,3}$ $n^r_{1,3}$	$x_{1,4}$ $n^r_{1,4}$	$x_{2,1}$ $n^r_{2,1}$	$x_{2,2}$ $n^r_{2,2}$	$x_{2,3}$ $n^r_{2,3}$	$\psi_D\left(d^r\right)$	c^r_{annual} (k€)
d^{4345}	1	3	4	2	2	2	2	-1.0753	1,865
d^{4834}	1	4	4	3	2	2	2	-1.0773	1,960
d^{4779}	1	4	4	1	2	2	2	-1.0780	1,800
d^{4260}	1	3	3	3	2	2	2	-1.0846	1,895
d^{1997}	0	3	4	3	2	2	2	-1.0843	1,865

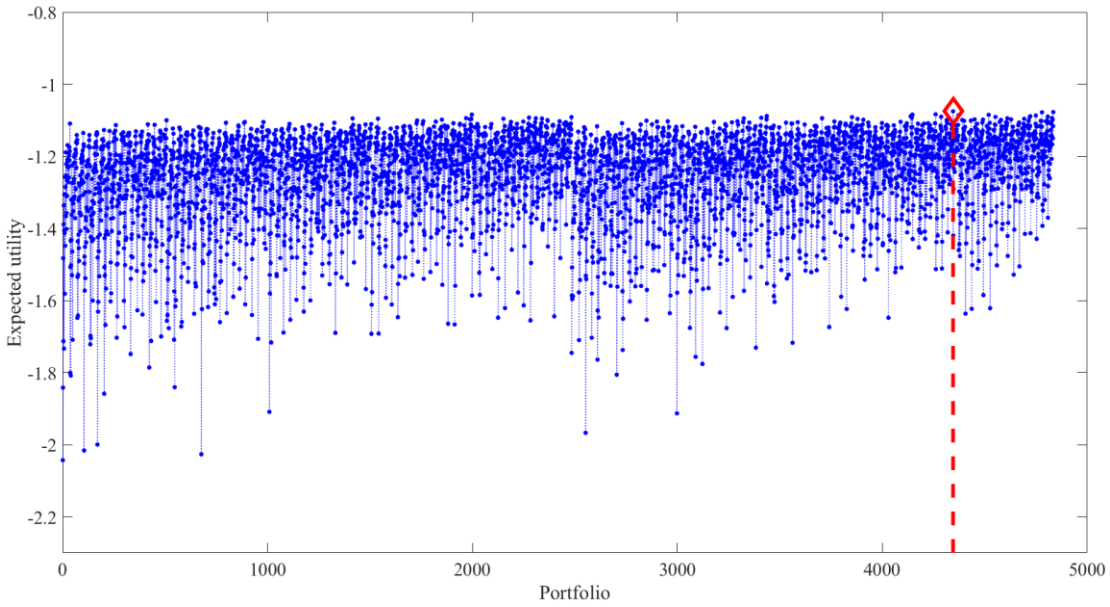

Fig. 3. The defender's expected utilities with respect to each portfolio

In line with the proposed approach, therefore, the N_v runs of MC simulation lead to $d^* = d^{4779} = \{1,4,4,1,2,2,2\}$ (diamond in Fig. 4) with the largest value of $\hbar_D(d^*)$ equal to 7e-3 for the optimal defense portfolio, that is taken as confidence measure for the result provided, leveraging the robustness of the protection actions on the ALFRED digital I&C system with uncertain malicious threats characteristics.

It is worth mentioning that $d^{4342} = \{1,3,4,2,2,1,2\}$, $d^{4749} = \{1,4,4,0,2,2,2\}$ and $d^{4345} = \{1,3,4,2,2,2,2\}$ turn out to be sub-optimal portfolios since they are estimated as $d^{*,v}$ among the N_v runs for 6, 5 and 5 times, as listed in Table 9, respectively. This suggests that the development and maintenance of a security software is usually time-consuming but may impair the CPS security level (if properly designed) (for example, the security analyst would be more likely to select d^{4779} (equipped with $n_{1,4}^r$=1 security software ($x_{1,4}$)) but not d^{4749} (without security software (i.e., $n_{1,4}^r$=0) for defense resource allocation), whereas, operators devoted to real-time monitoring of physical processing are more likely prone to human errors (for example, it is impossible to recruit only one operator in NPPs, as shown in d^{3998} (i.e., $n_{1,3}^r$=1)).

Table 9 The optimal defense portfolios with annual costs

d^r	$x_{1,1}$ $n^r_{1,1}$	$x_{1,2}$ $n^r_{1,2}$	$x_{1,3}$ $n^r_{1,3}$	$x_{1,4}$ $n^r_{1,4}$	$x_{2,1}$ $n^r_{2,1}$	$x_{2,2}$ $n^r_{2,2}$	$x_{2,3}$ $n^r_{2,3}$	$\varpi_D(d^r)$	c^r_{annual} (k€)
d^{4779}	1	4	4	1	2	2	2	7	1,800
d^{4342}	1	3	4	2	2	1	2	6	1,795
d^{4749}	1	4	4	0	2	2	2	5	1,720
d^{4345}	1	3	4	2	2	2	2	5	1,865
d^{3998}	1	3	1	2	2	2	2	4	1,715

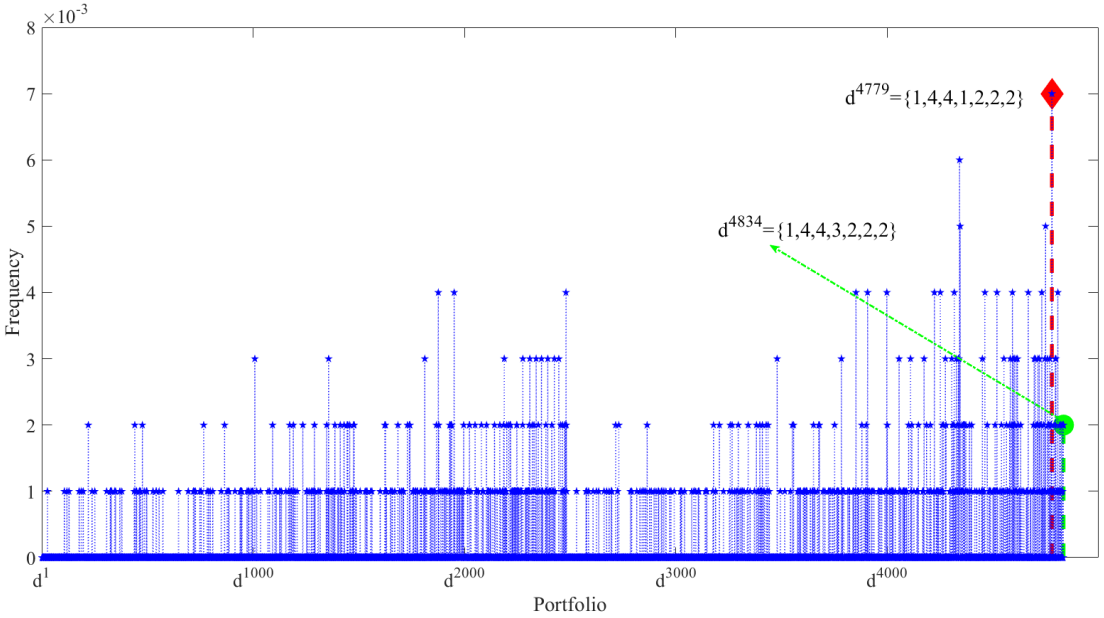

Fig. 4. Optimal defense portfolio from ARA assessment

4.2 Nash equilibrium solution from the classical defend-attack model

In the model of Section 3.2, the attacker beliefs, i.e., probabilities of outcomes, costs, risk aversion coefficient, are known to the defender, and assumed to be mean values from the corresponding distributions mentioned in Section 2.4.

Fig. 5 shows the Nash equilibrium solution $\left(d^*_{Nash}, a^*_{Nash}\right)$ obtained by finding the intersection node between the defender best responses with respect to $a_{j,y}$ and the attacker best responses with respect to d^r. On one hand, the attacker calculates his/her expected utility of each attack decision $a_{j,y}$ (of the pool of A) $\psi_A\left(a_{j,y} \mid d^r\right)$ and obtains a best response $a^*_{Nash}\left(d^r\right)$ with respect to a defender strategy d^r (out of $\Re = 4834$ portfolios). The solutions of Eq. (28) turn out to be the constant attacker best response

$a_{2,1}$ (attack to control rod actuator) with respect to different defense strategies (see dot line in Fig. 5). Whereas, on the other hand, the defender calculates his/her expected utility of each d^r (out of $\Re = 4834$ portfolios) $\psi_D\left(d^r \mid a_{j,y}\right)$ and obtains a best response $d^*_{Nash}\left(a_{j,y}\right)$ with respect to an attack decision $a_{j,y}$ (of A). The solutions of Eq. (29) vary with the attack decisions (see stars in Fig. 5). Notably, the attacker best responses $a^*_{Nash}\left(d^r\right)$ and the defender best responses $d^*_{Nash}\left(a_{j,y}\right)$ intersect at the point of $\left(d^{4834}, a_{2,1}\right)$, and Nash equilibrium is $\left(d^*_{Nash}, a^*_{Nash}\right)$ (see diamond in Fig. 5), where $d^{4834} = \{1,4,4,3,2,2,2\}$ is equipped with all deployable defensive resources under the restriction of B_M equal to 2,000 k€.

It must be noticed that the Nash equilibrium solution $d^*_{Nash} = d^{4834} = \{1,4,4,3,2,2,2\}$, obtained from the classical model that assumes that the defender and the attacker share common knowledge, differs from $d^* = d^{4779} = \{1,4,4,1,2,2,2\}$ by two sets of security softwares (i.e., $n^{4834}_{1,4} - n^{4779}_{1,4} = 2$).

Even if surprisingly marginal, there is indeed a fundamental difference between the two solutions: the Nash equilibrium solution $\left(d^{4834}, a_{2,1}\right)$ (shown with a circle in Fig. 4) in practice assumes the maximum quantity of defense resources to be installed with the maximum allowed budget B_M, whereas, the optimal decision of the ARA d^* (i.e., d^{4779} highlighted in diamond in Fig. 4) reaches the one-sided prescriptive optimal decision against all possible uncertain cyber attacks without reaching the maximum budget. Moreover, as shown in Fig. 4, the allocation strategy d^{4834} gives a value of $\hbar_D\left(d^*\right)$ equal to 2e-3 and, therefore, less effective in protecting the CPS from the uncertain attacks than d^{4779}.

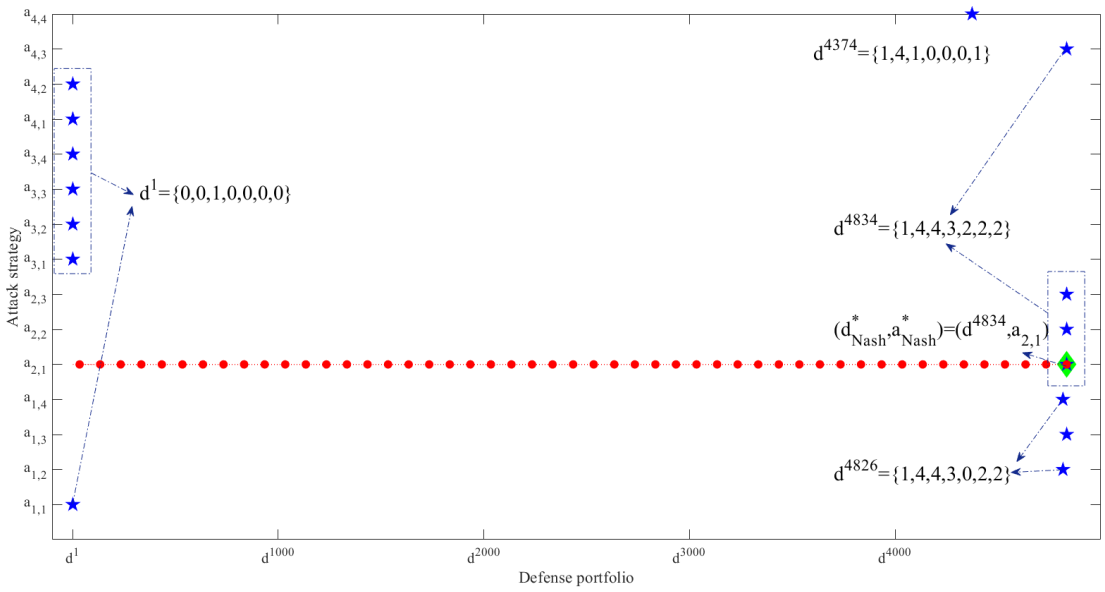

Fig. 5. Estimation of Nash equilibrium solution from the classical defend-attack model

5. CONCLUSIONS

In this study, we have proposed an Adversarial Risk Analysis (ARA) approach for analyzing decisions between intelligent adversaries provided a novel one-sided (i.e., defender) prescriptive support strategy for optimizing the defensive resource allocations based on a subjective expected utility model.

A Monte Carlo (MC) approach has been embedded into the ARA model for treating uncertainties in the decisions of the adversaries, for improving confidence in obtaining the optimal defense resource allocation, leveraging robustness of protection actions on the Cyber-Physical System (CPS) with uncertain malicious threats.

For demonstration, we have illustrated the proposed ARA framework to a cyber defend-attack game in the digital I&C system of the Advanced Lead Fast Reactor European Demonstrator (ALFRED). With respect to the prescriptive support, the ARA framework advised the defender the optimal portfolio of defense resource allocation, minimizing the system integrity loss against uncertain cyber attacks. The result has also been compared with the Nash equilibrium solution from a classical defend-attack model, in which the attacker and the defender share common knowledge regarding utility functions and probabilities of the outcomes of the game, showing that a stable status (Nash equilibrium) can be reached between the defender and the attacker, as a two-

sided prescriptively balanced strategy profile.

Future work will concern a sensitivity analysis aimed at identifying the most relevant uncertainties affecting the defend-attack model and its results, with the scope of limiting the subjectivity and/or conservatism of expert judgment in the assessment of the uncertainties considered.

REFERENCES

Abdo, H., Kaouk, M., Flaus, J.M. and Masse, F., 2018. A safety/security risk analysis approach of Industrial Control Systems: A cyber bowtie–combining new version of attack tree with bowtie analysis. Computers & Security, 72, pp.175-195.

Alemberti, A., Frogheri, M., Mansani, L., 2013. The Lead fast reactor demonstrator (ALFRED) and ELFR design. In: Proceedings of the International Conference on Fast Reactors and Related Fuel Cycles: Safe Technologies and Sustainable Scenarios (FR 13), Paris, France, March 4-7, 2013.

Aven, T., 2009. Identification of safety and security critical systems and activities. Reliability Engineering & System Safety, 94(2), pp.404-411.

Aven, T. and Krohn, B.S., 2014. A new perspective on how to understand, assess and manage risk and the unforeseen. Reliability Engineering & System Safety, 121, pp.1-10.

Aven, T. and Zio, E., 2011. Some considerations on the treatment of uncertainties in risk assessment for practical decision making. Reliability Engineering & System Safety, 96(1), pp.64-74.

Backhaus, S., Bent, R., Bono, J., Lee, R., Tracey, B., Wolpert, D., Xie, D. and Yildiz, Y., 2013. Cyber-physical security: A game theory model of humans interacting over control systems. IEEE Transactions on Smart Grid, 4(4), pp.2320-2327.

Banks, D.L., Aliaga, J.M.R. and Insua, D.R., 2015. Adversarial risk analysis. CRC Press.

Banks, D.L. and Anderson, S., 2006. Combining game theory and risk analysis in counterterrorism: A smallpox example. Statistical methods in counterterrorism,

pp.9-22.

Bernoulli, D., 2011. Exposition of a new theory on the measurement of risk. In The Kelly Capital Growth Investment Criterion: Theory and Practice (pp. 11-24).

Bi, S. and Zhang, Y.J., 2014. Graphical methods for defense against false-data injection attacks on power system state estimation. IEEE Transactions on Smart Grid, 5(3), pp.1216-1227.

Bier, V., Oliveros, S. and Samuelson, L., 2007. Choosing what to protect: Strategic defensive allocation against an unknown attacker. Journal of Public Economic Theory, 9(4), pp.563-587.

Bradley, J.M. and Atkins, E.M., 2015. Optimization and control of cyber-physical vehicle systems. Sensors, 15(9), pp.23020-23049.

Bricha, N. and Nourelfath, M., 2013. Critical supply network protection against intentional attacks: A game-theoretical model. Reliability Engineering & System Safety, 119, pp.1-10.

Busby, J.S., Green, B. and Hutchison, D., 2017. Analysis of Affordance, Time, and Adaptation in the Assessment of Industrial Control System Cybersecurity Risk. Risk Analysis.

Cano, J., Insua, D.R., Tedeschi, A. and Turhan, U., 2016a. Security economics: an adversarial risk analysis approach to airport protection. Annals of Operations Research, 245(1-2), pp.359-378.

Cano, J., Pollini, A., Falciani, L. and Turhan, U., 2016b. Modeling current and emerging threats in the airport domain through adversarial risk analysis. Journal of Risk Research, 19(7), pp.894-912.

Chen, D., Xu, M. and Shi, W., 2018. Defending a cyber system with early warning mechanism. Reliability Engineering & System Safety, 169, pp.224-234.

Colombo, A.W., Karnouskos, S., Kaynak, O., Shi, Y. and Yin, S., 2017. Industrial Cyberphysical Systems: A Backbone of the Fourth Industrial Revolution. IEEE Industrial Electronics Magazine, 11(1), pp.6-16.

Cox Jr, L.A.T., 2009. Game theory and risk analysis. Risk Analysis, 29(8), pp.1062-

1068.

Cox, J.C. and Sadiraj, V., 2006. Small-and large-stakes risk aversion: Implications of concavity calibration for decision theory. Games and Economic Behavior, 56(1), pp.45-60.

De Roze, B.C. and Nyman, T.H., 1978. The software life cycle—A management and technological challenge in the Department of Defense. IEEE Transactions on Software Engineering, (4), pp.309-318.

Ezhei, M. and Ladani, B.T., 2017. Information sharing vs. privacy: A game theoretic analysis. Expert Systems with Applications, 88, pp.327-337.

Fang, Y. and Sansavini, G., 2017. Optimizing power system investments and resilience against attacks. Reliability Engineering & System Safety, 159, pp.161-173.

Fielder, A., Panaousis, E., Malacaria, P., Hankin, C. and Smeraldi, F., 2016. Decision support approaches for cyber security investment. Decision Support Systems, 86, pp.13-23.

Ge, M., Hong, J.B., Yusuf, S.E. and Kim, D.S., 2018. Proactive defense mechanisms for the software-defined Internet of Things with non-patchable vulnerabilities. Future Generation Computer Systems, 78, pp.568-582.

Grasso, G., Petrovich, C., Mikityuk, K., Mattioli, D., Manni, F. and Gugiu, D., 2013, March. Demonstrating the effectiveness of the European LFR concept: the ALFRED core design. In Proc. of the IAEA International Conference on Fast Reactors and Related Fuel Cycles: Safe Technologies and Sustainable Scenarios.

Grechuk, B. and Zabarankin, M., 2016. Inverse portfolio problem with coherent risk measures. European Journal of Operational Research, 249(2), pp.740-750.

Hausken, K. and Levitin, G., 2009. Minmax defense strategy for complex multi-state systems. Reliability Engineering & System Safety, 94(2), pp.577-587.

Hershey, J.C. and Schoemaker, P.J., 1985. Probability versus certainty equivalence methods in utility measurement: Are they equivalent?. Management Science, 31(10), pp.1213-1231.

Hu, X., Xu, M., Xu, S. and Zhao, P., 2017. Multiple cyber attacks against a target with

observation errors and dependent outcomes: Characterization and optimization. Reliability Engineering & System Safety, 159, pp.119-133.

IAEA, 2009. Implementing Digital Instrumentation and Control Systems in the modernization of Nuclear Power Plants. Technical Report NP-T-1.4. IAEA.

Ingols, K., Lippmann, R. and Piwowarski, K., 2006, December. Practical attack graph generation for network defense. In Computer Security Applications Conference, 2006. ACSAC'06. 22nd Annual (pp. 121-130). IEEE.

Insua, D.R., Cano, J., Pellot, M. and Ortega, R., 2016. Multithreat multisite protection: A security case study. European Journal of Operational Research, 252(3), pp.888-899.

Jazdi, N., 2014, May. Cyber physical systems in the context of Industry 4.0. In Automation, Quality and Testing, Robotics, 2014 IEEE International Conference on (pp. 1-4). IEEE.

Khaitan, S.K. and McCalley, J.D., 2015. Design techniques and applications of cyberphysical systems: A survey. IEEE Systems Journal, 9(2), pp.350-365.

Kreps, D.M., 1990. Game theory and economic modelling. Oxford University Press.

Kriaa, S., Pietre-Cambacedes, L., Bouissou, M. and Halgand, Y., 2015. A survey of approaches combining safety and security for industrial control systems. Reliability Engineering & System Safety, 139, pp.156-178.

Langner, R., 2011. Stuxnet: Dissecting a cyberwarfare weapon. IEEE Security & Privacy, 9(3), pp.49-51.

Lee, E.A., 2008, May. Cyber physical systems: Design challenges. In Object Oriented Real-Time Distributed Computing (ISORC), 2008 11th IEEE International Symposium on (pp. 363-369). IEEE.

Levitin, G., 2007. Optimal defense strategy against intentional attacks. IEEE Transactions on Reliability, 56(1), pp.148-157.

Levitin, G. and Hausken, K., 2009. Parallel systems under two sequential attacks. Reliability Engineering & System Safety, 94(3), pp.763-772.

Ma, C.Y., Yau, D.K., Lou, X. and Rao, N.S., 2013. Markov game analysis for attack-

defense of power networks under possible misinformation. IEEE Transactions on Power Systems, 28(2), pp.1676-1686.

McQueen, M.A., Boyer, W.F., Flynn, M.A. and Beitel, G.A., 2006, January. Quantitative cyber risk reduction estimation methodology for a small SCADA control system. In System Sciences, 2006. HICSS'06. Proceedings of the 39th Annual Hawaii International Conference on (Vol. 9, pp. 226-226). IEEE.

Mehetre, D.C., Roslin, S.E. and Wagh, S.J., 2018. Detection and prevention of black hole and selective forwarding attack in clustered WSN with Active Trust. Cluster Computing, pp.1-16.

Moayedi, B.Z. and Azgomi, M.A., 2012. A game theoretic framework for evaluation of the impacts of hackers diversity on security measures. Reliability Engineering & System Safety, 99, pp.45-54.

Modarres, Mohammad. Risk analysis in engineering: techniques, tools, and trends. CRC press, 2016.

Nazir, S., Patel, S. and Patel, D., 2017. Assessing and augmenting SCADA cyber security: A survey of techniques. Computers & Security, 70, pp.436-454.

Nespoli, P., Papamartzivanos, D., Mármol, F.G. and Kambourakis, G., 2017. Optimal countermeasures selection against cyber attacks: A comprehensive survey on reaction frameworks. IEEE Communications Surveys & Tutorials.

Nisan, N., Roughgarden, T., Tardos, E. and Vazirani, V.V. eds., 2007. Algorithmic game theory (Vol. 1). Cambridge: Cambridge University Press.

Noureddine, M.A., Marturano, A., Keefe, K., Bashir, M. and Sanders, W.H., 2017, January. Accounting for the Human User in Predictive Security Models. In Dependable Computing (PRDC), 2017 IEEE 22nd Pacific Rim International Symposium on (pp. 329-338). IEEE.

Osborne, M.J. and Rubinstein, A., 1994. A course in game theory. MIT press.

Piccinelli, R., Sansavini, G., Lucchetti, R. and Zio, E., 2017. A General Framework for the Assessment of Power System Vulnerability to Malicious Attacks. Risk Analysis, 37(11), pp.2182-2190.

Piètre-Cambacédès, L. and Bouissou, M., 2013. Cross-fertilization between safety and security engineering. Reliability Engineering & System Safety, 110, pp.110-126.

Polatidis, N., Pavlidis, M. and Mouratidis, H., 2018. Cyber-attack path discovery in a dynamic supply chain maritime risk management system. Computer Standards & Interfaces, 56, pp.74-82.

Ponciroli, R., Bigoni, A., Cammi, A., Lorenzi, S. and Luzzi, L., 2014. Object-oriented modelling and simulation for the ALFRED dynamics. Progress in Nuclear Energy, 71, pp.15-29.

Ponciroli, R., Cammi, A., Della Bona, A., Lorenzi, S. and Luzzi, L., 2015. Development of the ALFRED reactor full power mode control system. Progress in Nuclear Energy, 85, pp.428-440.

Ponemon Institute, 2017. Cost of cyber crime study: insights on the security investments that make a difference. Research report.

Pratt, J.W., 1964. Risk aversion in the small and in the large. Econometrica 32, 122–136.

Quijano, E.G., Insua, D.R. and Cano, J., 2016. Critical networked infrastructure protection from adversaries. Reliability Engineering & System Safety.

Rios, J. and Insua, D.R., 2012. Adversarial risk analysis for counterterrorism modeling. Risk analysis, 32(5), pp.894-915.

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

Roger, B.M., 1991. Game theory: analysis of conflict.

Roponen, J. and Salo, A., 2015. Adversarial Risk Analysis for Enhancing Combat Simulation Models. Journal of Military Studies, 6(2), pp.82-103.

Rothschild, C., McLay, L. and Guikema, S., 2012. Adversarial risk analysis with incomplete information: A level‐k approach. Risk Analysis, 32(7), pp.1219-1231.

Shandilya, V., Simmons, C.B. and Shiva, S., 2014. Use of attack graphs in security systems. Journal of Computer Networks and Communications, 2014.

Sheyner, O. and Wing, J., 2003, November. Tools for generating and analyzing attack graphs. In International Symposium on Formal Methods for Components and Objects (pp. 344-371). Springer, Berlin, Heidelberg.

Sun, H., Peng, C., Yang, T., Zhang, H. and He, W., 2017. Resilient control of networked control systems with stochastic denial of service attacks. Neurocomputing, 270, pp.170-177.

Viscusi, W.K., 2009. Valuing risks of death from terrorism and natural disasters. Journal of Risk and Uncertainty, 38(3), pp.191-213.

Viscusi, W.K. and Aldy, J.E., 2003. The value of a statistical life: a critical review of market estimates throughout the world. Journal of risk and uncertainty, 27(1), pp.5-76.

Von Neumann, J. and Morgenstern, O., 2007. Theory of games and economic behavior (commemorative edition). Princeton university press.

Wan, J., Tang, S., Shu, Z., Li, D., Wang, S., Imran, M. and Vasilakos, A.V., 2016. Software-defined industrial internet of things in the context of industry 4.0. IEEE Sensors Journal, 16(20), pp.7373-7380.

Wang, W., Di Maio, F., Zio, E., 2017a. A Non-Parametric Cumulative Sum Approach for Online Diagnostics of Cyber Attacks to Nuclear Power Plants. Resilience of Cyber-Physical Systems: From Risk Modelling to Threat Counteraction, accepted.

Wang, W., Cammi, A., Di Maio, F., Lorenzi, S. and Zio, E., 2018. A Monte Carlo-based exploration framework for identifying components vulnerable to cyber threats in nuclear power plants. Reliability Engineering & System Safety, 175, pp.24-37.

Wang, C., Hou, Y. and Ten, C.W., 2017. Determination of Nash equilibrium based on plausible attack-defense dynamics. IEEE Transactions on Power Systems, 32(5), pp.3670-3680.

Wurm, J., Jin, Y., Liu, Y., Hu, S., Heffner, K., Rahman, F. and Tehranipoor, M., 2017. Introduction to cyber-physical system security: A cross-layer perspective. IEEE Transactions on Multi-Scale Computing Systems, 3(3), pp.215-227.

Xiang, Y. and Wang, L., 2017. A game-theoretic study of load redistribution attack and

defense in power systems. Electric Power Systems Research, 151, pp.12-25.

Xiang, Y., Wang, L. and Zhang, Y., 2018. Adequacy evaluation of electric power grids considering substation cyber vulnerabilities. International Journal of Electrical Power & Energy Systems, 96, pp.368-379.

Yang, Q., Yang, J., Yu, W., An, D., Zhang, N. and Zhao, W., 2014. On false data-injection attacks against power system state estimation: Modeling and countermeasures. IEEE Transactions on Parallel and Distributed Systems, 25(3), pp.717-729.

Zalewski, J., Buckley, I.A., Czejdo, B., Drager, S., Kornecki, A.J. and Subramanian, N., 2016. A Framework for Measuring Security as a System Property in Cyberphysical Systems. Information, 7(2), p.33.

Zhang, C., Ramirez-Marquez, J.E. and Li, Q., 2018. Locating and protecting facilities from intentional attacks using secrecy. Reliability Engineering & System Safety, 169, pp.51-62.

Zhang, J., Zhuang, J. and Jose, V.R.R., 2018. The role of risk preferences in a multi-target defender-attacker resource allocation game. Reliability Engineering & System Safety, 169, pp.95-104.

Zhuang, J. and Bier, V.M., 2007. Balancing terrorism and natural disasters—Defensive strategy with endogenous attacker effort. Operations Research, 55(5), pp.976-991.

Zhuang, J. and Bier, V.M., 2011. SECRECY AND DECEPTION AT EQUILIBRIUM, WITH APPLICATIONS TO ANTI‐TERRORISM RESOURCE ALLOCATION. Defence and Peace Economics, 22(1), pp.43-61.

Zio, E., 2016. Challenges in the vulnerability and risk analysis of critical infrastructures. Reliability Engineering & System Safety, 152, pp.137-150.

Zio, E., 2018. The Future of Risk Assessment. Reliability Engineering & System Safety.

Zou, L.L., 2017, December. Risk Analysis of Cyber Security in Nuclear Power Plant. In Nuclear Power Plants: Innovative Technologies for Instrumentation and Control Systems: The Second International Symposium on Software Reliability, Industrial Safety, Cyber Security and Physical Protection of Nuclear Power Plant

(Vol. 455, p. 139). Springer.