An Adversarial Risk Analysis Framework for Cybersecurity

Rios Insua, David; Couce-Vieira, Aitor; Rubio, Jose A.; Pieters, Wolter; Labunets, Katsiaryna; G. Rasines, Daniel

DOI
10.1111/risa.13331

Publication date
2019

Document Version
Final published version

Published in
Risk Analysis

Citation (APA)
Rios Insua, D., Couce-Vieira, A., Rubio, J. A., Pieters, W., Labunets, K., & G. Rasines, D. (2019). An Adversarial Risk Analysis Framework for Cybersecurity. *Risk Analysis*, *41*(1), 16-36. https://doi.org/10.1111/risa.13331

 DOI: 10.1111/risa.13331

An Adversarial Risk Analysis Framework for Cybersecurity

David Rios Insua,[1,*] Aitor Couce-Vieira,[1] Jose A. Rubio,[2] Wolter Pieters,[3] Katsiaryna Labunets,[3] and Daniel G. Rasines[4]

Risk analysis is an essential methodology for cybersecurity as it allows organizations to deal with cyber threats potentially affecting them, prioritize the defense of their assets, and decide what security controls should be implemented. Many risk analysis methods are present in cybersecurity models, compliance frameworks, and international standards. However, most of them employ risk matrices, which suffer shortcomings that may lead to suboptimal resource allocations. We propose a comprehensive framework for cybersecurity risk analysis, covering the presence of both intentional and nonintentional threats and the use of insurance as part of the security portfolio. A simplified case study illustrates the proposed framework, serving as template for more complex problems.

KEY WORDS: Adversarial risk analysis; cybersecurity; cyber insurance; resource allocation; risk analysis

1. INTRODUCTION

At present, all kinds of organizations are being critically impacted by cyber threats (Anderson, 2008; Andress & Winterfeld, 2013). Risk analysis is a fundamental methodology to help manage such issues (Cooke & Bedford, 2001). With it, organizations can assess the risks affecting their assets and what security controls they should implement to reduce the likelihood of such threats and/or their possible impacts should they happen.

Numerous frameworks support cybersecurity risk management, including ISO 27005 (International Organization for Standardization [ISO], 2011), CRAMM (Central Communication and Telecommunication Agency [CCTA], 2003), MAGERIT (Ministerio de Hacienda y Administraciones Pblicas [MINHAP], 2012), EBIOS (Agence Nationale de la Scurit des Systems d'Information [ANSSI], 1995), SP 800-30 (National Institute of Standards and Technology [NIST], 2012), and CORAS (Lund, Solhaug, & Stlen, 2010). Similarly, several compliance and control assessment frameworks, like ISO 27001 (ISO, 2013), Common Criteria (The Common Criteria Recognition Agreement Members [CCRA], 2009), and CCM (Cloud Security Alliance [CSA], 2016), provide guidance on the implementation of cybersecurity best practices. They have many virtues, particularly their extensive catalogues of threats, assets, and controls, and provide detailed guidelines for the implementation of countermeasures to protect digital assets. However, much remains to be done regarding risk analysis from a methodological point of view. Indeed, a detailed study of the main approaches to cybersecurity risk management reveals that it often relies on risk matrices, with shortcomings well documented in Cox (2008) and Thomas, Bratvold, and Bickel (2014): compared to more stringent methods, the qualitative ratings in risk matrices (likelihood,

[1]Instituto de Ciencias Matematicas, Consejo Superior de Investigaciones Cientficas, Madrid, Spain.
[2]Analysis, Security and Systems Group, Complutense University of Madrid, Madrid, Spain.
[3]Faculty of Technology, Policy and Management, Delft University of Technology, Delft, The Netherlands.
[4]Department of Mathematics, Imperial College, London, UK.
[*]Address correspondence to David Rios Insua, Instituto de Ciencias Matematicas, C. Nicolas Cabrera, num. 13-15, Campus de Cantoblanco, UAM. 28049, Madrid, Spain; david.rios@icmat.es.

 0272-4332/19/0100-0001$22.00/1

severity, and risk) are more prone to ambiguity and subjective interpretation and, very importantly for our application area, they systematically assign the same rating to risks that are very different risks qualitatively, potentially inducing suboptimal cybersecurity resource allocations. Hubbard and Seiersen (2016) and Allodi and Massacci (2017) provide additional critical views on the use of risk matrices in cybersecurity. Moreover, with few exceptions, like IS1 (National Technical Authority for Information Assurance [HMG], 2012), those methodologies do not explicitly take into account the intentionality of certain threats. This is in contrast with the relevance that organizations like the Information Security Forum (ISF, 2016) start to give to such threats, receiving the name adversarial in contrast to more standard ones defined as accidental or environmental. Thus, ICT owners may obtain unsatisfactory results in relation to the prioritization of cyber risks and the measures they should implement. In this context, a complementary way for dealing with cyber risks through risk transfer is emerging: cyber insurance products, of very different natures and not in every country, have been introduced in recent years by companies like AXA, Generali, and Allianz. However, cyber insurance has yet to take off (Marotta, Martinelli, Nanni, Orlando, & Yautsiukhin, 2017).

We aim at developing methods to support decisions in relation to cybersecurity resource allocation, including the adoption of cyber insurance. Fielder, Panaousis, Malacaria, Hankin, and Smeraldi (2016) review and introduce various approaches to such problems, focusing on optimization and game-theory models and their combination. Schilling and Werners (2016) describe a combinatorial optimization approach for optimal selection of IT security safeguards, with no consideration of risk or adversarial aspects. Cavusoglu, Raghunathan, and Yue (2018) and Rao et al. (2015) provide game-theoretic models for cybersecurity resource allocation under common knowledge assumptions that might not be realistic in our context. Thus, we propose an alternative framework for cybersecurity risk analysis, combining optimization with an adversarial risk analysis (ARA) approach to deal with adversarial agents; we emphasize adversarial aspects for a better prediction of threats as well as include cyber insurance within the cybersecurity portfolio. Section 2 presents our framework, supported by a simplified case study in Section 3, which may serve as a template for more complex problems. We conclude with a brief discussion. An appendix compares our approach with a standard game-theoretic one in a stylized example.

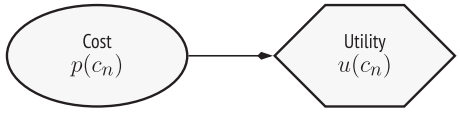

Fig. 1. Basic ID for system performance evaluation. c_n indicates costs associated with system operation over the relevant planning period; the utility function $u(c_n)$ accounts for risk attitudes. Note that in IDs, oval nodes represent uncertainties modeled with a probability distribution $p(\dots)$, and utility nodes represent preferences modeled with an utility function $u(\dots)$.

2. A CYBERSECURITY ADVERSARIAL RISK ANALYSIS FRAMEWORK

We introduce our integrated risk analysis approach to facilitate cybersecurity resource allocation. Our aim is to improve current cybersecurity frameworks, introducing schemes that incorporate all relevant parameters, including decisionmakers' preferences and risk attitudes (Clemen & Reilly, 2013) and the intentionality of adversaries. Moreover, we consider decisions concerning cyber insurance adoption to complement other risk management alternatives through risk transfer. We present the framework stepwise, analyzing the elements involved progressively. We describe the models (Banks, Rios, & Rios Insua, 2015) through influence diagrams (ID) and bi-agent influence diagrams (BAID) detailing the relevant elements: assets, threats, security controls, and impacts. At each step, we provide a brief description of the diagrams introduced and a generic mathematical formulation.

2.1. System Performance Evaluation

Fig. 1 describes the starting outline for a cyber system under study. c_n designates the costs associated with its operation over the relevant period; they are typically uncertain and modeled with a probability distribution $p(c_n)$. We introduce a utility function $u(c_n)$ over costs to account for risk attitudes (Ortega, Radovic, & Rios Insua, 2018). We evaluate system performance under normal conditions, that is, in absence of relevant incidents, through its associated expected utility $\psi_n = \int u(c_n)\, p(c_n)\ dc_n$ (French & Ros Insua, 2000). This scheme can be sophisticated in several directions. For example, there could be several performance functions, leading to a multiattribute problem, as reflected in the case in

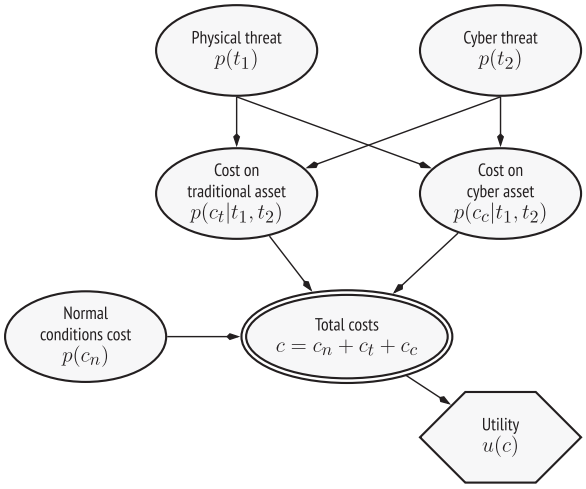

Fig. 2. Cybersecurity risk assessment. The threats (two in this example, t_1 and t_2) might impact on the organization's assets, causing costs (two in this example, c_t and c_c). These costs, and those under normal conditions c_n, are aggregated to determine the total costs c and evaluated through the utility function $u(c)$. Recall that in IDs, double-lined ovals represent deterministic aspects.

Section 3. A typical example in cybersecurity is to consider attributes concerning information availability, integrity, and confidentiality (Mowbray, 2013).

2.2. Cybersecurity Risk Assessment

Based on Fig. 1, we consider the cybersecurity risk assessment problem in Fig. 2. In general, we include m threats $t_1, \ldots, t_m$; some of them could be physical (e.g., a fire) and others cyber (e.g., a DDoS attack[5]). Their occurrence is random variables. We also include l types of assets; some of them could be traditional (e.g., facilities) and others could be cyber (e.g., information systems). Impacts on them will be, respectively, designated $c_i, i = 1, \ldots, l$ and are typically uncertain. If the impacts are conditionally independent given the threats, the corresponding model would be of the form $p(c_1|t_1, \ldots, t_m) \ldots p(c_l|t_1, \ldots, t_m)\, p(t_1, \ldots, t_m)$, where $p(t_1, \ldots, t_m)$ describes the probability of the threats happening,[6] and $p(c_i|t_1, \ldots, t_m)$ describes the probability of impact on the ith asset, given the occurrence of various threats. We aggregate costs additively at the total cost node c. Then, the expected utility would be:

$$\psi_r = \int \ldots \int u\left(c_n + \sum_{i=1}^{l} c_i\right) p(c_n)\, p(c_1|t_1, \ldots, t_m) \ldots p(c_l|t_1, \ldots, t_m)\, p(t_1, \ldots, t_m)\, dt_m \ldots dt_1\, dc_l \ldots dc_1\, dc_n.$$

We have assumed that consequences are additive, but we could have a generic utility $u(c_n, c_1, \ldots, c_l)$. Finally, we evaluate the loss in expected utility $\psi_n - \psi_r$. Alternatively, we could compare the difference in the corresponding certain equivalents (French, 1986). When such difference is sufficiently large, incidents are expected to harm the system significantly and we should manage such risks. Note that we could incorporate several utility nodes to describe multiple stakeholders' preferences.

2.3. Risk Mitigation in Cybersecurity Risk Management

As a next step, we add security controls. We introduce a portfolio of them to reduce the likelihood of threats and/or their impact. Examples include firewalls, employee training, or making regular backups. For simplicity, in Fig. 3, we assume that all controls have influence over all events and impacts. It will not always be so: a fire detector makes less harmful, but not less likely, a fire; resource accounting mechanisms (Mirkovic & Reiher, 2004) managing access based on user privileges make a successful DDoS attack less likely, but usually not less harmful. Node e describes the portfolio of controls, whose cost we model through the distribution $p(c_e|e)$. Controls might have influence on threat likelihoods $p(t_i|e)$, $i = 1, \ldots, m$, as well as on asset impact likelihoods $p(c_i|t_1, \ldots, t_m, e)$. We aggregate all costs through the total cost node c, under appropriate additivity assumptions. In this case, the organization's expected utility when we implement portfolio e is:

$$\psi(e) = \int \ldots \int u\left(c_n + c_e + \sum_{i=1}^{l} c_i\right) p(c_n) \times p(c_e|e)\, p(c_1|t_1, \ldots, t_m, e) \ldots p(c_l|t_1, \ldots, t_m, e) \times p(t_1, \ldots, t_m|e)\, dt_m \ldots dt_1\, dc_l \ldots dc_1\, dc_c\, dc_n.$$

We would then look for the maximum expected utility portfolio by solving $\psi_e^* = \max_{e \in E} \psi(e)$, where E is the set of feasible portfolios, which should

[5] A distributed denial of service (DDoS) is a network attack consisting of a high number of infected computers flooding with network traffic a victim computer or network device, rendering it inaccessible.

[6] Depending on the problem, we could have further decompositions. For example, in a case like that in Fig. 2 with independent threats, we would have $p(t_1, \ldots, t_m) = \prod_{i=1}^{m} p(t_i)$.

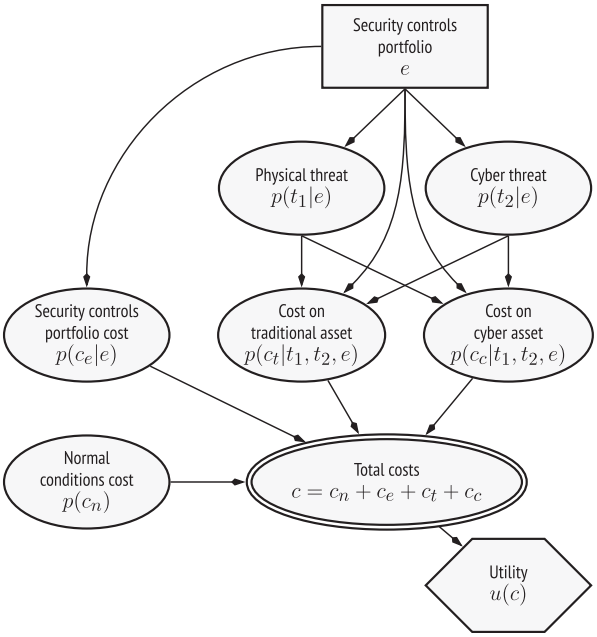

Fig. 3. Cybersecurity risk management. We add to Fig. 2 the security controls portfolio e (and its cost c_e) that the organization can implement to mitigate the threats or their impacts. Recall that rectangle nodes represent decisions.

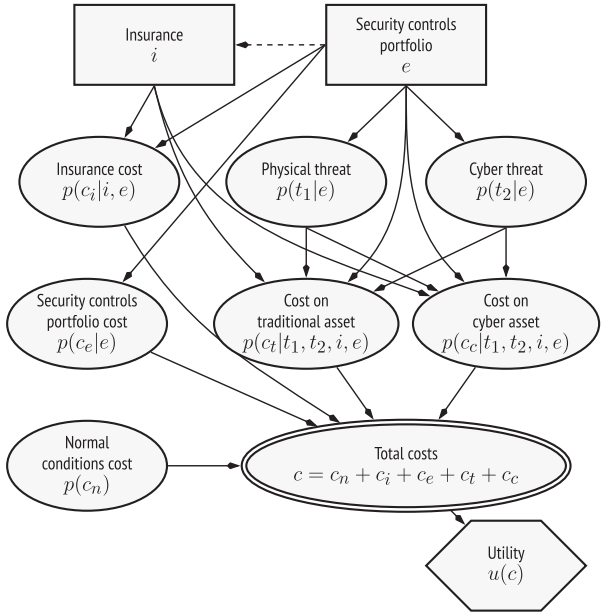

Fig. 4. Cyber insurance for cybersecurity risk management. We add to Fig. 3 the insurance i (and its cost c_i) to which the organization can subscribe to mitigate the impacts that the threats can cause.

satisfy incumbent constraints like economic (e.g., not exceeding a budget), legal (e.g., complying with data protection laws), logistic, or physical.

2.4. Risk Transfer in Cybersecurity Risk Management: Cyber Insurance

As a relevant element of increasing interest, we introduce the possibility of acquiring a cyber insurance product. Its cost will typically depend on the implemented portfolio of controls, as in Fig. 4: the better such a portfolio is, the lower the insurance premium would be. This cost will also depend on the assets to be protected. We could include the insurance within the portfolio of controls; however, it is convenient to represent it separately, since premiums will usually depend on the controls deployed. The decision node i describes the cyber insurance adopted, with entailed costs c_i with probability $p(c_i|i,e)$, although they will usually be deterministic. In addition, insurance and security controls will affect impacts, modeled through $p(c_j|t_1,\ldots,t_m,e,i)$, $j=1,\ldots,l$. The total cost node c aggregates the costs. The expected utility when we implement portfolio e together with insurance i is:

$$\begin{aligned}\psi(e,i) = \int\ldots\int u\left(c_n + c_e + c_i + \sum_{j=1}^{l} c_j\right) p(c_n) \\ \times\ p(c_i|i,e)p(c_e|e) \times p(c_1|t_1,\ldots,t_m,e,i)\ldots \\ \times\ p(c_l|t_1,\ldots,t_m,e,i)\ p(t_1,\ldots,t_m|e)\ dt_m\ldots \\ \times\ dt_1 dc_l\ldots dc_1\ dc_i\ dc_e\ dc_n.\end{aligned}$$

We seek the maximum expected utility portfolio of security controls and insurance by solving $\psi^*_{e,i} = \max_{e\in E, i\in I}\psi(e,i)$, where I represents the catalogue of insurance products available. The pair (e,i) could be further restricted jointly, for example, by compliance requirements or common budget constraints.

2.5. Adversarial Risk Analysis in Cybersecurity

As discussed, intentionality is a key factor when analyzing cyber threats. As an example, the ISF (2016) specifies a group of several adversarial threats within its catalogue. We use ARA (Banks et al., 2015) to model the intentions and strategic behavior of adversaries in the cybersecurity domain; see Merrick and Parnell (2011) for a comparison of various methods modeling adversaries in risk

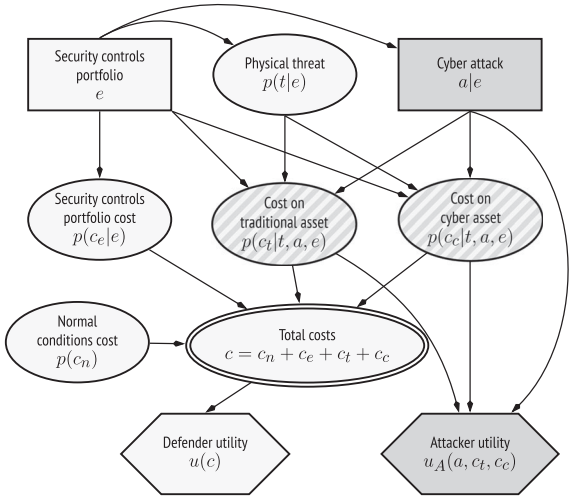

Fig. 5. Adversarial risk analysis in cybersecurity: defense–attack problem. We modify Fig. 3 by transforming the cyber threat into an adversarial one: an attacker is deciding whether to attack the organization (a) based on his own evaluation, $u(a, c_t, c_c)$, of the harm caused to the organization, and the cost of performing the attack. Lighter nodes refer to issues concerning solely the Defender; darker nodes refer to issues relevant only for the Attacker; nodes with stripped background affect both agents. Arcs have the same interpretation as in Shachter (1986).

management. Under ARA, the attacker has his own utility function and seeks to maximize the effectiveness of his attack. This paradigm is applicable to multiple types of strategic interactions between attackers and defenders. Two of them are specially relevant in cybersecurity.

2.5.1. *Defense–Attack Model*

The original examples, Figs. 2 and 3, evolve into Fig. 5, modeling an adversarial case through a BAID with a Defender and an Attacker. The unintentional threat remains modeled through a probabilistic node, whereas we model the adversarial threat through a decision node for the Attacker, who needs to decide whether to launch an attack to his benefit. For simplicity, in the diagram we model the physical threat t_1 as unintentional and the cyber threat a as adversarial, although adversarial physical threats and unintentional cyber threats could be relevant in certain cases, as exemplified in the case study. Also for simplicity, we only consider one attacker and one attack, but the ideas extend to multiple attacks by one attacker or to multiple attackers.

We present a sequential defense–attack template model for cybersecurity. For the Defender problem,

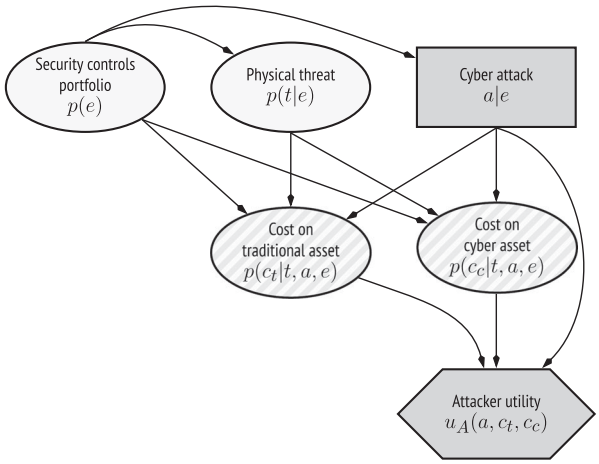

Fig. 6. Attacker problem in the defense–attack model.

this converts the Attacker's decision nodes into chance ones and eliminates the Attacker's nodes not affecting it, as well as the corresponding utility node. For the Attacker, where we assume here that there is only one Attacker responsible for the adversarial threat a independent of the other threats, given the portfolio e. Fig. 3 essentially presents the Defender problem and we covered its resolution in Section 2.3. The cyber attack is described probabilistically[7] through $p(a|e)$, which represents the probability that the Defender assigns to cyber threat a materializing, had portfolio e been adopted. However, given the strategic nature of this problem, rather than using a standard probability elicitation approach (Dias, Morton, & Quigley, 2018), we greatly facilitate and improve the assessment of the required distribution if we analyze the Attacker decision about which attack to perform, as argued in Rios, Insua, Banks, Rios, and Ortega (2019). Under the ARA paradigm, the Defender should analyze the Attacker strategic problem in Fig. 6.

Specifically, given portfolio e, and assuming that the Attacker maximizes expected utility, the Defender would compute, for each attack a, the expected utility for the Attacker:

$$\psi_A(a|e) = \iiint u_A(a, c_1, \ldots, c_l)\, p_A(c_1|t_1, \ldots, t_m, a, e) \times \ldots p_A(c_l|t_1, \ldots, t_m, a, e)\, p_A(t_1, \ldots, t_m|e) \times dt_m \ldots dt_1\, dc_c\, dc_t,$$

[7]We are assuming that given e, a is conditionally independent of $(t_1, \ldots, t_m)$.

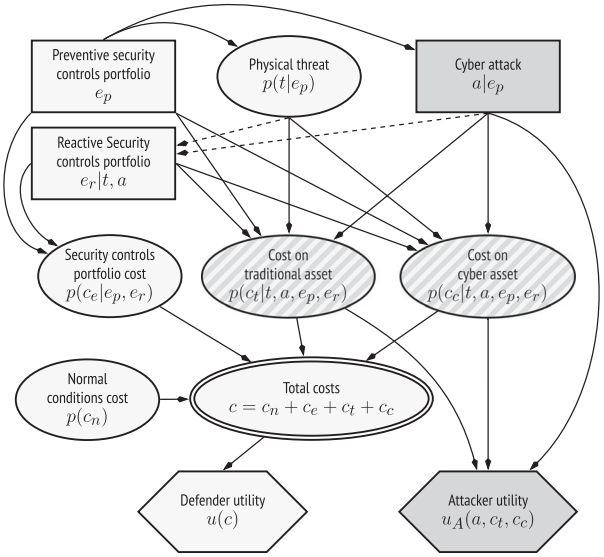

Fig. 7. Adversarial risk analysis in cybersecurity: defense–attack–defense problem.

where u_A and p_A designate, respectively, the utility and probabilities of the Attacker. The Defender must then find the attack solving:

$$\max_{a \in A} \psi_A(a|e),$$

where A is the set of attack options. However, the Defender will not typically know u_A and p_A. Suppose we are capable of modeling her uncertainty about them with random probabilities P_A and a random utility function U_A (Banks et al., 2015). Then, the optimal random attack, given e, is:

$$\begin{aligned} A^*(e) = \arg\max_{a \in A} \iiint U_A(a, c_1, \ldots, c_l)\, P_A(c_1|t_1, \ldots, \\ \times\, t_m, a, e) \ldots P_A(c_l|t_1, \ldots, t_m, a, e)\, P_A(t_1, \ldots, \\ \times\, t_m|e) dt_m \ldots dt_1\, dc_c\, dc_t. \end{aligned}$$

Finally, the distribution over attacks we were looking for satisfies $p(a|e) = P(A^*(e) = a)$, assuming that the attack set is discrete (e.g., attack options). Similarly, if the attack space is continuous (e.g., attack efforts), the probability becomes a density function. We can estimate such attack distribution through Monte Carlo (MC) simulation as in Algorithm 1 (see the Appendix), where we designate the distribution of random utilities and probabilities through $F = (U_A(a, c_1, \ldots, c_l),\, P_A(c_1|t_1, \ldots, t_m, a, e), \ldots, P_A(c_l|t_1, \ldots, t_m, a, e),\, P_A(t_1, \ldots, t_m|e))$.

2.5.2. *Defense–Attack–Defense Model*

Cybersecurity risk management also comprises reactive measures that can be put in place to counter an attack, should it happen. Therefore, we split the security portfolio into two groups: preventive e_p and reactive $e_r|t_1, \ldots, t_m, a$ security controls, as in Fig. 7. This corresponds to our sequential defense–attack–defense template model, in which the first move is by the Defender (preventive portfolio e_p), the second one is by the Attacker (attack after observing preventive controls, $a|e_p$), and the third one is by the Defender (reactive portfolio $e_r|t_1, \ldots, t_m, a$). We solve the Defender problem much as we did in Section 2.3, reflecting changes caused by splitting the security control node. Specifically, the expected utility when portfolio $e = (e_p, e_r)$ is implemented is:

$$\begin{aligned} \psi(e) = \int \ldots \int u\left(c_n + c_e + \sum_{i=1}^{l} c_i\right) p(c_n)\, p(c_e|e_p, e_r) \\ \times\, p(c_l|t_1, \ldots, t_m, a, e_p, e_r) \ldots p(c_1|t_1, \ldots, t_m, \\ \times\, a, e_p, e_r) p(t_1, \ldots, t_m|e_p)\, p(a|e_p)\, da\, dt_m \ldots \\ \times\, dt_1\, dc_l \ldots dc_1\, dc_t\, dc_e\, dc_n. \end{aligned}$$

We would then look for the maximum expected utility portfolio:

$$(e_p^*, e_r^*) = \arg\max_{(e_p, e_r) \in E_p \times E_r} \psi(e_p, e_r),$$

where E_p and E_r, respectively, define constraints for preventive and reactive portfolios, some of which could be joint.

The above represents a global view of the sequential problem, although we solve this kind of two-stage problems sequentially, as in He and Zhuang (2017). We would solve the Attacker problem providing $p(a|e_p)$ in a similar fashion as in Section 2.5.1.

3. A CASE STUDY TEMPLATE

We illustrate our cybersecurity risk analysis framework with a defense–attack case study, which can serve as a template for more complex problems. For confidentiality reasons, we have simplified the number of relevant issues and masked the data conveniently. This simplification will also allow us to better illustrate key modeling concepts and the overall scheme. Moreover, we include uncertain phenomena in which data are abundant and others in which they are not and, thus, we shall need to rely on expert judgment for its quantification (Dias et al.,

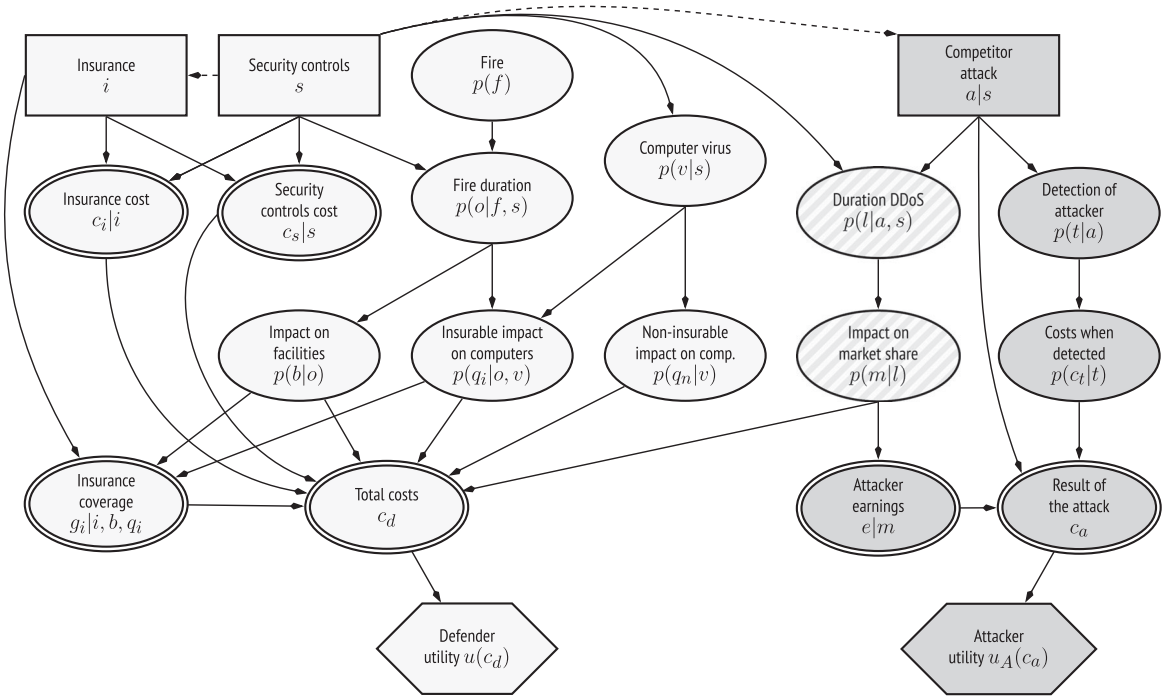

Fig. 8. Case study as a BAID.

2018). The Defender is an SME[8] with 60 people and 90 computers. A cyber attack might affect its online services. Prices and rates in Euros refer to Spain, where the incumbent organization is located.

In essence, we first structure the problem, identifying assets, threats, and security controls. The latter may have implementation costs in exchange for reducing the threat likelihoods and/or possible impacts. Subsequently, we assess the impacts that may have an effect on asset values to find the optimal risk management portfolio. Since we include adversarial threats, we consider the Attacker decision problem. In this case, there is a single potential Attacker that contemplates a DDoS attack with the objective of disrupting the Defender services, causing an operational disruption and reputational damage and the consequent loss of customers, besides incurring contractual penalties potentially affecting its continuity. Then, we simulate from this problem to obtain the attack probabilities, which feed back into the Defender problem to obtain the optimal defense. We focus on finding the optimal security portfolio and insurance product for the company, in the sense of maximizing expected utility. Other formulations are discussed in Section 3.5. We consider a one-year planning horizon.

[8]Small or medium-size enterprise.

3.1. Problem Structuring

We structure the problem through the BAID in Fig. 8 and describe its components next.

3.1.1. Assets

We first identify the Defender assets at risk. We could obtain them from catalogues like those of the methodologies mentioned in Section 1. Here we consider: *Facilities*, the offices potentially affected by threats; *Computer equipment*, the data center and workstations of the organization; *Market share*. Other assets not considered in this case include, for example, the company's development software, its business information, its mobile devices, or the staff.

3.1.2. Nonintentional Threats

We consider threats over the identified assets deemed relevant and having nonintentional character. This may include threats traditionally insurable as well as new ones potentially cyber

insurable. We model each threat with a probabilistic node associated with the Defender problem. We extract two threats from the MAGERIT (MINHAP, 2012) catalogue: fire and computer virus. A *fire* may affect facilities and computers; we do not contemplate impact on market share, as the organization has a backup system; we assume that a fire can occur only by accident, not by sabotage. The *computer virus* is aimed at disrupting normal operations of computer systems; we consider this threat nonintentional, as most viruses propagate automatically: their occurrence tends to be random from the Defender perspective. Other nonintentional threats, not considered here, could be water damage, power outages, or employee errors.

Table I. Insurance Product Features, Some of Them Referring to Cyber Impacts

Product	Coverage
No insurance	None.
Traditional insurance	80% of hired capital in buildings and contents, firefighters, and movement of furniture.
Cyber insurance	80% of cyber expenses related to: confidential data violation, investigation and legal costs, losses caused by threats and extortion, removal of computer viruses, measures related to data protection procedures, and computer fraud.
Comprehensive insurance	All of the above.

3.1.3. Intentional Threats

This category may include both cyber and physical threats. Again, we could use catalogues from, for example, ISF (2016). We should first identify the attackers. We then integrate the attack options available to each attacker within a single decision node. In our case, we just consider one competitor, reflected in the *competitor attack* node. He may attempt a DDoS to undermine the availability of the Defender site, compromising its customer services. For this, he must decide whether to launch the attack and the number of attempts. Other intentional attacks, not modeled here, could include launching an advanced persistent threat, instigating the misbehaviour of insiders, or the use of bombs.

3.1.4. Uncertainties Affecting Threats

We consider now those uncertainties affecting the Defender's assets. We model each of them with a probabilistic node. In our case, these will be the *duration of the DDoS* attack, which will depend on the number of attacks and security controls deployed, and the *fire duration*, which can be reduced with an anti-fire system. Other related uncertainties could come, for example, from a more detailed modeling of the virus (e.g., infection probability given the operating system) or the eventual propagation of the fire to adjacent buildings.

3.1.5. Attacker Uncertainties

We model the uncertainties that the Attacker might find relevant and that only affect him with probabilistic nodes (in his own color). In our case, we consider only the *detection of the Attacker*: if detected, his reputation would suffer and he might face legal prosecution. Other attacker uncertainties that might be included are the effectiveness of the DDoS platform or the number of customers affected by the DDoS.

3.1.6. Relevant Security Controls

We identify security controls relevant to counter the threats. For this, we may use listings from the above-mentioned methodologies. We associate a Defender decision node with the security controls. In our case, we consider an *anti-fire system* to detect a fire, facilitating early mitigation; a *firewall* to protect the network from malicious traffic; the implementation of *risk mitigation procedures* for cybersecurity and fire protection; and a *cloud-based DDoS protection*, diverting DDoS traffic to an absorbing cloud-based site. Other measures, not included here, could be a system resource management policy, a cryptographic data protocol, or a wiring protection.

3.1.7. Insurance

We also consider the possibility of purchasing insurance to transfer risk with the corresponding Defender decision node. The premium will depend on the protected assets and contextual factors such as location, company type and, quite importantly, the implemented controls. Table I displays the contemplated insurance products.

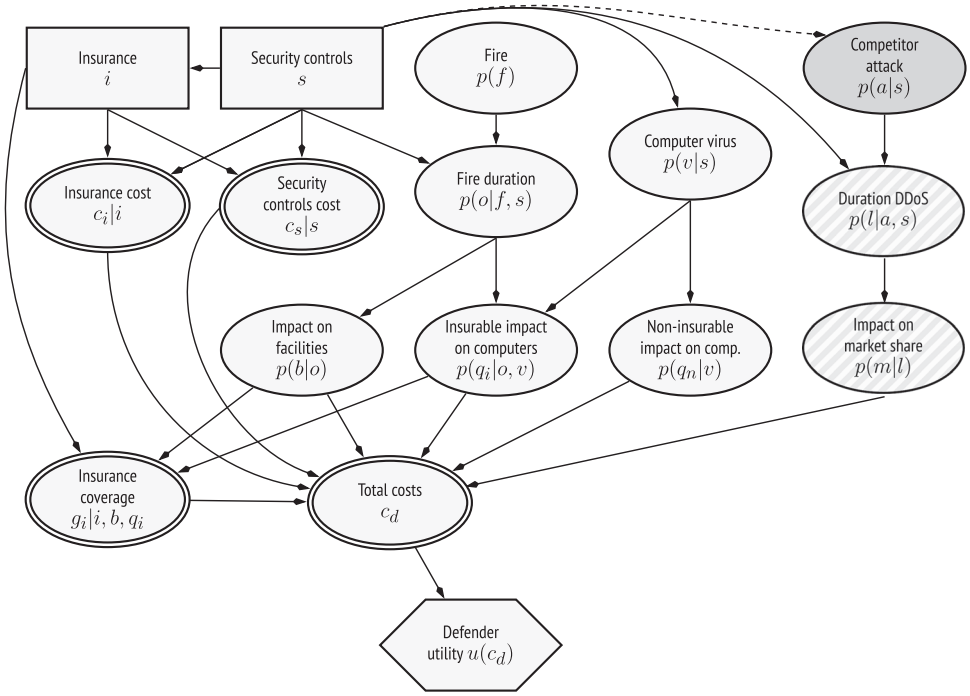

Fig. 9. Defender problem.

3.1.8. Impacts on Defender

Having identified the threats, we present their relevant impacts on the Defender's assets. We model each of them with a probabilistic node. We consider: *Impact on facilities*, the monetary losses caused by fire in or on them; *Impact on computers*, the monetary losses caused by fire or viruses split into insurable and noninsurable ones to assess the possible insurance coverage; *Impact on market share*. We also consider the impacts associated with safeguards as deterministic nodes: *cost of security controls*, *cost of insurance*, and *insurance coverage*. Finally, a deterministic *total costs* node aggregates the Defender's consequences to establish the final impact in the Defender problem. We could include other types of impacts such as the corporate image or the staff safety, although we do not do so here.

3.1.9. Impacts on Attacker

We consider the following impacts: *Attacker earnings* from increased market share, transferred from those lost by the defender; *Costs when detected*, covering possible sanctions by the regulator and legal costs, as well as loss of customers and reputation, if detected. The final *Results of attack* combines all previous impacts, as well as the costs of undertaking the attack. We model the *Costs when detected* as a probabilistic node. The remaining ones are deterministic.

3.1.10. Preferences

Value nodes describe how the corresponding agent evaluates consequences. We include one value node for each of the participating agents: the *Utility of Defender* node models the Defender preferences and risk attitudes over the total costs; the *Utility of Attacker* node models those of the Attacker.

3.1.11. Defender and Attacker Problems

Figs. 9 and 10, respectively, represent the Defender and Attacker problems derived from the strategic problem in Fig. 8. We use both diagrams to guide judgment elicitation.

3.2. Assessing the Defender's Nonstrategic Beliefs and Preferences

We now provide the quantitative assessment of the Defender beliefs and preferences not requiring strategic analysis. Some of them will be based on data and expert judgment, others just on expert

Table II. Cost of Individual Security Controls

Security Control	Cost			
Anti-fire system	€ 1,500			
Firewall	€ 2,250			
Risk mitigation procedures	€ 2,000			
Cloud-based DDoS protection	2 gbps	5 gbps	10 gbps	1,000 gbps
	€ 2,400	€ 3,600	€ 4,800	€ 12,000

Table III. Insurance Product Cost

	Security Controls			
Prod.	None	Anti-Fire	Firewall or DDoS prot.	Proc.
None	€ 0	€ 0	€ 0	€ 0
Trad.	€ 500	€ 300	€ 500	€ 500
Cyber	€ 300	€ 300	€ 200	€ 250
Compr.	€ 700	€ 500	€ 600	€ 650

Table IV. Industrial Fire Data in Vitoria (2005–2009)

Year	Buildings	Fires
2005	1,220	32
2006	1,266	29
2007	1,320	30
2008	1,347	28
2009	1,314	28

judgment due to the typical lack of data in cybersecurity environments (Hubbard & Seiersen, 2016). As a consequence, we populate most nodes in the model. Section 3.3 treats nodes that require strategic analysis. Finally, Section 3.4 analyzes the Defender problem to find the optimal controls and insurance. When incumbent, we provide the pertinent utility u, random utility U_A, probability p, random probability P_A, or deterministic model at the corresponding node.

3.2.1. *Economic Value of Defender Assets*

We consider the following values for the assets at risk: *Facilities*, with a value of € 5,000,000, reflecting only acquisition costs; *Computer equipment*, with a value of € 200,000, under similar considerations; *Market share* is estimated at 50%, which, translated into next-year expected profits, is valued at € 1,500,000.

3.2.2. *Modeling Security Controls*

Security controls decision s. The security portfolios that the Defender could implement derive from the options in Section 3.1. For the DDoS protection, we have the choice of not implementing it or subscribing to a 2, 5, 10, or 1,000 gbps service. For the other security controls, the choice is binary. We thus have 40 portfolios that could be constrained by, for example, a budget, as in Section 3.5.

Cost of security controls $c_s|s$. Table II provides them, from which we derive those of the portfolios.

3.2.3. *Modeling the Insurance Product*

Insurance decision i. This refers to the insurance product that the Defender could purchase (Table III) once the controls have been selected.

Insurance cost, $c_i|i$. It depends on the controls implemented by the organization (Table III).

Insurance coverage $g_i|i, b, q_i$, as reflected in Table I.

3.2.4. *Modeling the Fire Risk*

Likelihood, $p(f)$. This node provides the annual probability of suffering a fire. We use data from Vitoria (DSC de Vitoria, 2009), concerning fire interventions in industrial buildings (Table IV). As the fire rate remains fairly stable over time, we estimate such probability with a beta-binomial model with beta prior $\beta e(1/2, 1/2)$. The posterior would be:

$$f|\text{data} \sim \beta e(1/2 + \sum_{i=1}^{5} x_i,$$

$$1/2 + \sum_{i=1}^{5}(n_i - x_i)) \equiv \beta e(147.5, 6320.5),$$

where x_i is the number of fires affecting industrial buildings and n_i the number of buildings in the ith year, $i = 1, \ldots, 5$. As the posterior variance is small, such distribution can be reasonably summarized through its posterior expectation, $\hat{p} = 0.022$. The number f of fires can be approximated with a Poisson $\mathcal{P}(0.022)$ distribution. However, we consider only the probability that one fire occurs, since $Pr(f > 1) = 0.00024$. Thus, $f \sim \min[1, \mathcal{P}(0.022)]$.

Duration, $p(o|f, s)$. It is a major fire impact determinant (Bagchi, Sprintson, & Singh, 2013): the longer the fire, the more damaging it will be. Fig. 11 presents the histogram of industrial fire durations,

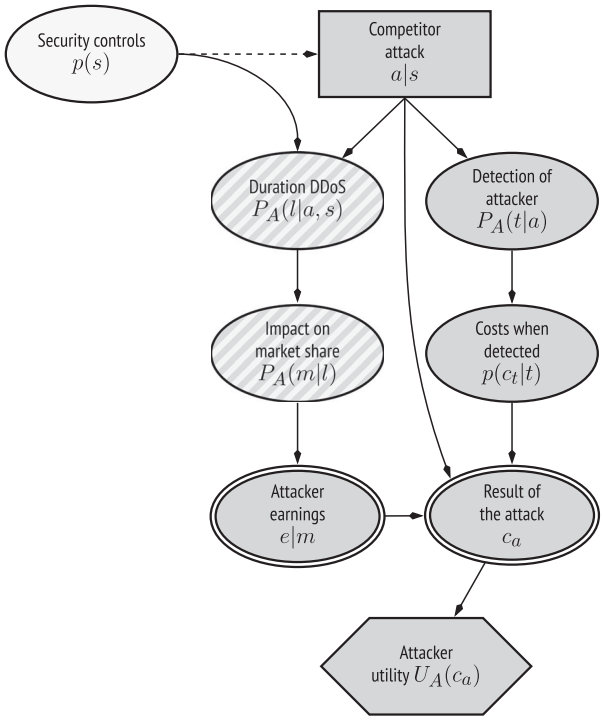

Fig. 10. Attacker problem.

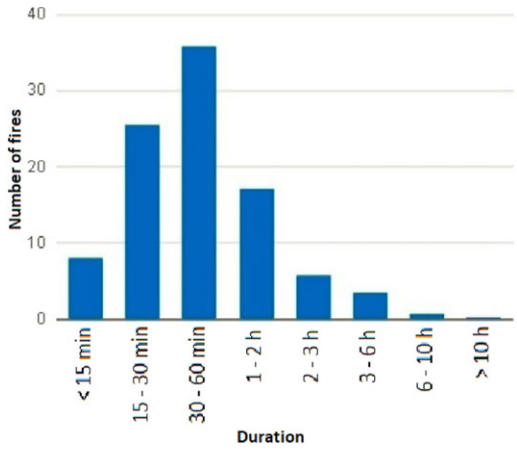

Fig. 11. Industrial fire duration histogram. Vitoria, Spain (2005–2009).

with mode [30,60] minutes. Adapting Wiper, Rios Insua, and Ruggeri (2001), we model the fire duration o with a gamma $\Gamma(\text{shape} = \gamma, \text{scale} = \gamma/\mu)$ distribution. We assume a noninformative, but proper, exponential prior for $\gamma \sim \mathcal{E}(0.01)$ and inverse gamma for $\mu \sim \text{Inv-}\Gamma(1,1)$. No expression for the posterior distribution is available, but we can introduce a Markov chain MC scheme to sample μ and γ from the data. Based on it, we estimate that $\text{E}(\gamma|\text{data}) \approx 0.85$ and $\text{E}(\mu|\text{data}) \approx 78$.

The only proposed control that may have an effect on fire duration is the anti-fire system. Using expert judgment (Dias et al., 2018), we determine its threshold duration under the proposed system with, respectively, suggested minimum, modal, and maximum durations of 1, 10, and 60 minutes. To mitigate expert overconfidence (Galway, 2007), we consider a triangular distribution with quantiles 0.05 at 1 and 0.95 at 60 minutes, resulting in a triangular distribution $\text{Tri}(0.8, 63, 10)$, which models o if there is a fire ($f = 1$) and the portfolio s contains the anti-fire system. On the other hand, $o \sim \Gamma(0.85, 0.0109)$ if the portfolio does not contain the anti-fire system.

Impact. We assume that the amount lost is linearly related to the fire duration. After consulting with experts, we consider that a fire lasting 120 minutes would degrade the facilities by 100% in the absence of controls. To simplify, we assume that the effect of fire duration is linear. Additionally, the impact on computer equipment derives from the percentage of facility degradation caused by fire. Assuming that computers are evenly distributed through the premises, a fire lasting 120 minutes would also degrade computer equipment by 100%. This impact is potentially insurable and will be modeled in Section 3.2.7.

3.2.5. *Modeling the Computer Virus Risk*

Likelihood, $p(v|s)$. This node provides the number v of virus infections during a year. The distribution of the number of infected computers in a month follows a binomial distribution $\mathcal{B}(h, q)$, with q the probability that a computer gets infected and h the number of computers. Various statistics suggest that the rate of virus infections worldwide is 33% (Panda Security, 2015), so we estimate $\hat{q} = 0.33$. The organization has 90 computers, which we assume have the same security controls and are equally likely to be infected. Since the analysis is for 12 months, we use $h = 12 \cdot 90 = 1,080$. Additionally, we consider the effect of our controls: if a firewall is implemented, the probability that a computer gets infected reduces to $\hat{q} = 0.005$, not completely eliminating the threat, even if this includes continuous updating based on the latest virus signatures; if the mitigation procedures are implemented, the infection probability reduces by 50%, with firewall or not, as this control entails improvements in the organization such as imposing safety

Table V. Number v of Annual Virus Infections

Sec. Controls in s	Distribution
Firewall and proc.	$v \sim \mathcal{B}(1080, 0.0025)$
Firewall	$v \sim \mathcal{B}(1080, 0.005)$
Procedure	$v \sim \mathcal{B}(1080, 0.1666)$
Otherwise	$v \sim \mathcal{B}(1080, 0.33)$

requirements on acquired systems. The number v of infections is, therefore, modeled as in Table V.

Impact. Viruses may impact the integrity and availability of computers, leading to information corruption or unavailability. Impacts on confidentiality are variable, as they depend on the stolen information. The average daily cost of these infections was estimated at € 2.683 (Solutionary, 2013), although this may vary depending on the monetary value of the information and services that the victim systems support. Bigger losses come from sophisticated campaigns (e.g., as with WannaCry) or targeted malware that, under our paradigm, we would better model as an adversarial threat. In our case, repairing a computer infected by a virus requires € 31, for two technician hours. Insurance options cover the removal of computer viruses. Therefore, we cover this impact within the insurable aspects in Section 3.2.7.

Additionally, most viruses entail performance reduction in aspects such as initialization of operating systems. Although small, this causes time losses to the user. We assume that most (70%) of the work time of the organization is in front of a computer and that it would take, on average, 40 hour to detect the problem. We therefore assume that when a computer is infected, 28 hour of its usage are affected by the virus. We model the time loss w with a uniform $\mathcal{U}(0, 0.05)$ distribution that represents that the percentage of lost time caused by a virus is between 0% and 5%. The average hourly cost of the employees is € 20/hour. Therefore, for each virus infection, the cost would be $20 \times 28 \times w$. Our insurance options do not cover this loss and, thus, we model it within the noninsurable aspects in Section 3.2.7.

3.2.6. *Modeling the DDoS Threat*

We consider now the nonstrategic aspects of the DDoS threat.

Duration, $p(l|a, s)$. The duration l in hours of a successful DDoS attack will depend on the intensity of the attacking campaign, how well-crafted the attack is, and the security controls implemented. An emerging type of control is cloud-based systems absorbing traffic when a site becomes a victim of a DDoS. If no control is deployed, it would be virtually impossible to block such attack. Based on Securelist (2016) and Verisign (2017), the average attack lasts four hours, averaging 1 gbps, with peaks of 10 gbps. We model l_j, the length of the jth individual DDoS attack, as a $\Gamma(4, 1)$. This duration is conditional on whether the attack actually saturates the target, which depends on the capacity of the DDoS platform minus the absorption by the cloud-based system. We assume that the Attacker uses a professional platform capable of 5 gbps attacks, modeled through a $\Gamma(5, 1)$ distribution. We then subtract the traffic s_{gbps} absorbed by the protection system to determine whether the attack is successful (its traffic overflows the protection system). Since the campaign might take a attacks, the output of this node is $l = \sum_j^a l_j$, with $l_j \sim \Gamma(4, 1)$ if $\Gamma(5, 1) - s_{\text{gbps}} > 0$, and $l_j = 0$, otherwise.

Impact. A DDoS attack might cause a reputational loss that would affect the organization's market share. We assume that all market share is lost at a linear rate until all value is gone, say, after five to eight days of unavailability: in the fastest case, the loss rate r would be $0.5/120 = 0.00417$ per hour, whereas in the slowest one it would be 0.0026. We model r as a $\mathcal{U}(0.0026, 0.00417)$.

3.2.7. *Modeling Impacts on the Assets*

We recollect here the impacts on the assets.

Impact on facilities, $p(b|o)$. The monetary loss b due to degradation of facilities through fire is $b \sim 5,000,000 \times \min(1, \frac{o}{120})$, following Section 3.2.4.

Insurable impacts on computers, $p(q_i|o, v)$. We model the monetary losses q_i due to degradation of computers covered by insurance, either caused by fire, Section 3.2.4, or through repairing computers infected with viruses, Section 3.2.5, as $q_i \sim 31v + 200,000 \times \min(1, \frac{o}{120})$.

Noninsurable impacts on computers, $p(q_n|v)$. The monetary losses q_m caused by degradation of computers due to the lost time caused by viruses are not covered by insurance. Following Section 3.2.5, we model $q_n \sim 560w \times v$.

Impact on market share, $p(m|l)$. The monetary loss m due to a reduced market share, following Section 3.2.6, is $m \sim \min[1,500,000, 3,000,000 \times l \times r]$.

Total Defender costs, $c_d|g_i, c_i, c_s, m, b, q_i, q_n$. The costs c_d suffered by the Defender are $c_d = m + b + q_i + q_n + c_s + c_i - g_i$, where c_s is the security controls cost, c_i that of insurance, g_i the insurance coverage (which reduces losses), and m, b, q_i, and q_n are the impacts on assets previously described.

3.2.8. Defender Utility, $u(c_d)$

The organization is constant risk averse over costs. Its utility function is strategically equivalent to $u(c_d) = a - b\exp(k(c_d))$. We calibrate it with three costs: worst, best, and an intermediate one. The worst reasonable loss is based on the sum of all costs and impacts (except that due to the computer virus) € 6,755,300. Computer virus impacts do not have an upper limit; based on simulations, it is reasonable to assume that they would not exceed € 50,000. Giving an additional margin, we assume that such maximum is $7,000,000$. The best loss is 0. For the intermediate cost $c_d^* = 2,660,000$, we find its probability equivalent α so that $u(c_d^*) = \alpha$ (Ortega et al., 2018); based on information provided by the company, $u(c_d^*) \simeq .5$. We rescale the costs to the (0,1) range through $1 - \frac{c_d}{7,000,000}$. Then, the utility function is $u(c_d) = \frac{1}{e-1}[\exp(1 - \frac{c_d}{7,000,000}) - 1]$.

3.3. Assessing the Attacker's Random Beliefs and Preferences

In the Defender problem, the competitor attack is described through a probabilistic node modeling the number of attacks launched by the Attacker given the security controls that are implemented. To obtain the corresponding probabilities, we model the Attacker problem based on Fig. 10. Its solution would provide the Attacker's optimal action. However, as argued in Section 2.5, we model our uncertainty about his preferences and beliefs through random utilities and probabilities to find the random optimal attack; for this, we simulate from it to forecast his actions and obtain the required probability distribution.

Defender's security controls. This node is probabilistic for the Attacker. However, we assume that he may observe through network exploration tools whether the Defender has implemented relevant controls.

Competitor attack decision, $a|s$. This decision node models how many attacks (between 0, doing nothing, and 30) the DDoS campaign will make. Attackers usually give up once the attack has been mitigated and move on to the next target or try other disruption methods. However, when the attack is targeted, the Attacker might continue the campaign for several days, causing an extensive impact.

Duration of the DDoS, $P_A(l|a, s)$. We base our estimate on that of the Defender (Section 3.2.6). We model the length of the jth individual DDoS attack as a random gamma distribution $\Gamma_{\text{length}}(\upsilon, \upsilon/\mu)$ with $\upsilon \sim \mathcal{U}(3.6, 4.8)$ and $\upsilon/\mu \sim \mathcal{U}(0.8, 1.2)$, adding some uncertainty around its average duration (between three and six hours) and dispersion. Similarly, we model the attack gbps through a random gamma distribution $\Gamma_{\text{gbps}}(\omega, \omega/\eta)$ with $\omega \sim \mathcal{U}(4.8, 5.6)$ and $\omega/\eta \sim U(0.8, 1.2)$. Next, we subtract s_{gbps} from Γ_{gbps} to determine whether the DDoS is successful. As in Section 3.2.6, we use $l = \sum_j^a l_j$, with $l_j \sim \Gamma_{\text{length}}$ if $\Gamma_{\text{gbps}} - s_{\text{gbps}} > 0$, and $l_j = 0$ otherwise.

Impact on market share, $P_A(m|l)$. We base our estimate on that of the Defender (Section 3.2.7), adding some uncertainty. The market share value and percentage are not affected by uncertainty, as this information is available to both agents. However, we model uncertainty in the market loss rate: the fastest one (five days in the Defender problem) is between four and six days in the Attacker problem and the slowest one (eight for Defender) is between seven and nine. Therefore, the random distribution describing the market loss is $m \sim \min[1,500,000, 3,000,000 \times l \times R]$ with $R \sim \mathcal{U}(\alpha, \beta)$, $\alpha \sim \mathcal{U}(0.0021, 0.0031)$, and $\beta \sim \mathcal{U}(0.00367, 0.00467)$.

Attacker earnings, $e|m$. Being the sole competitor, we assume that the Attacker gain e in terms of market share is $e = m$. The random uncertainty in earnings derives from the randomness of the preceding nodes.

Attacker Detection, $P_A(t|a)$. This represents the chance of the Attacker being detected. Detection probabilities are estimated via expert judgment at 0.2%, should the Attacker attempt a DDoS attack. Should there be a attacks, the detection has a binomial distribution $\mathcal{B}(a, 0.002)$. To add some uncertainty, we model the detection probability for each attack through a $\beta e(2, 998)$.[9] Then, we model the Attacker's detection t through a random binomial distribution that outputs *detected* if $\mathcal{B}(a, \phi) > 0$ with $\phi \sim \beta e(2, 998)$, and *not detected*, otherwise.

Cost for Attacker when detected, $p_A(c_t|t)$. As a competitor, if the Attacker is detected, he would face a serious discredit, together with compensation

[9]Its mean is 0.002.

Table VI. Conditional Probability Table for Random Optimal Attacks

	Number of Attempts															
DDoS Prot. System	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1,000 gbps	1.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
10 gbps	0.000	0.001	0.003	0.003	0.004	0.005	0.012	0.012	0.015	0.013	0.017	0.024	0.024	0.022	0.030	0.035
5 gbps	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.001	0.001	0.002
2 gbps	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
None	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000

	Number of Attempts														
DDoS Prot. System	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30
1,000 gbps	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
10 gbps	0.026	0.041	0.025	0.044	0.042	0.053	0.050	0.048	0.047	0.060	0.050	0.059	0.065	0.081	0.089
5 gbps	0.008	0.006	0.012	0.017	0.007	0.028	0.031	0.055	0.070	0.061	0.096	0.117	0.143	0.141	0.203
2 gbps	0.000	0.000	0.002	0.001	0.002	0.013	0.013	0.020	0.034	0.069	0.091	0.112	0.144	0.223	0.276
None	0.000	0.000	0.003	0.001	0.004	0.008	0.010	0.022	0.042	0.058	0.081	0.105	0.173	0.246	0.247

and legal costs as well as criminal responsibilities. We use this cost decomposition: € 550,000 of *expected reputational costs*, due to the communication actions required to preserve credibility; € 30,000 of *expected legal costs*; € 350,000 of *expected civil indemnities and regulatory penalties*; € 1,500,000 of *expected suspension costs*, related to losses derived from prohibition to operate for some time. We add some uncertainty, modeling the cost as a normal distribution with mean 2,430,000 and *SD* 400,000, that is, $c_t|t \sim \mathcal{N}(2,430,000, 400,000)$.

Result of attack, $c_a|e, c_t, a$. This combines the Attacker earnings and costs if detected, and those of undertaking the attacks. We consider that using a botnet to launch the DDoS attack would cost on average around € 33 per hour (Incapsula, 2015) (€ 792 for a day). Therefore, $c_a = e - c_t - 792a$.

Attacker's random utility, $U_A(c_a)$. We assume that the Attacker is risk prone, with a utility function strategically equivalent to $u_A(c_a) = (c'_a)^{k_a}$, where $k > 1$, c'_a are the costs c_a normalized to $[0, 1]$, and k_a the risk proneness parameter. We induce the random utility considering that k_a follows a $\mathcal{U}(8, 10)$ distribution.

3.3.1. *Simulating the Attacker Problem*

Summarizing the earlier assessments, the *distribution of random utilities and probabilities in the Attacker problem* is $F = (U_A(c_a), p_A(c_t|t), P_A(t|a), P_A(m|l), P_A(l|a, s))$. We calculate the *random optimal attack*, given the security controls s implemented through:

$$A^*(s) = \arg\max_a \int \dots \int U_A(c_a)\, p_A(c_t|t)\, P_A(t|a)\, P_A \times (m|l)\, P_A(l|a, s)\, dl\, dm\, dt\, dc_t.$$

To approximate it, we use an MC approach as in Algorithm 1 (Appendix A) with K=20,000, which we have implemented in R. For each size s of the DDoS protection system, we assess the distribution of the random optimal attack. Table VI displays the attack probabilities, conditional on the protection implemented. For instance, if the security portfolio does not contain a DDoS protection system ($s = 0$, none), an attack seems certain, and its duration would be between 18 and 30 attacks, 29 and 30 being the most likely attack sizes. We thus create the probability distribution $p(a|s)$. We have now fully specified the Defender problem and are ready to solve it.

3.4. Solving the Defender Problem

Summarizing earlier assessments about the Defender problem, we have that the involved distributions are $G = (p(m|l), p(q_n|v), p(q_i|o, v), p(b|o), p(l|a, s), p(a|s), p(v|s), p(o|f, s), p(f))$. The Defender's expected utility when the security portfolio s is implemented together with insurance

Table VII. Expected Utility for Three Best and Worst Combinations of Controls and Insurance

Anti-Fire	Firewall	Procedure	DDoS Protection	Insurance	Expected Utility
Anti-fire	Firewall	No procedure	1,000 gbps	Comprehensive	0.9954
Anti-fire	Firewall	No procedure	1,000 gbps	Traditional	0.9950
Anti-fire	Firewall	Procedure	1,000 gbps	Comprehensive	0.9949
...	...	...	...	...	...
No anti-fire	No firewall	No procedure	No protection	Cyber	0.8246
No anti-fire	No firewall	Procedure	No protection	No insurance	0.8246
No anti-fire	No firewall	No procedure	No protection	No insurance	0.8242

i is:

$$\psi(s,i) = \int \dots \int u(c_d)\, p(m|l)\, p(q_n|v)\, p(q_i|o,v) \times p(b|o)\, p(l|a,s)\, p(a|s)\, p(v|s)\, p(o|f,s)\, p(f) \times df\, do\, dv\, da\, dl\, db\, dq_i\, dq_n\, dm.$$

The *optimal resource allocation* is the maximum expected utility pair $(s^*, i^*) = \arg\max_{s,i} \psi(s,i)$. We use Algorithm 2 (Appendix A) to approximate the portfolio together with the optimal portfolio. We have implemented it in R with an MC sample size of K=20,000 and the results are summarized in Table VII. The best portfolio consists of *a* 1,000 *gbps cloud-based DDoS protection system, a firewall, an anti-fire system*, and *the comprehensive insurance*. Besides the ranking of countermeasures, we can obtain additional information from the simulation. For instance, the best portfolios tend to include a firewall, a 1,000 gbps DDoS protection with no risk management procedure. The best portfolios also include insurance, either traditional or comprehensive.

3.5. Further Analysis

The previous ARA model can be used to perform other relevant analysis, as we briefly discuss.

3.5.1. *Sensitivity Analysis*

We can assess the robustness of the previous solution by checking whether variations in the inputs to the model alter the optimal solution. This is especially important in a case like ours with small differences in expected utility among top alternatives and many inputs being purely judgmental. The approach would require the implementation of additional algorithms for sensitivity analysis that indicate whether small deviations in input parameters may lead to a large effect in the model outcome (Rios, 1990). As an example, the optimal portfolio in Table VII will remain as such until we sufficiently reduce the value of $p(f)$, specifically $f \sim \min[1, \mathcal{P}(0.0088)]$. If $p(f)$ is further reduced, the optimal portfolio will contain the same security controls and insurance as the optimal, except for the anti-fire system.

Additionally, sensitivity analysis can be used to explore the maximum cyber insurance price that the Defender would be willing to pay. This may be used, *inter alia*, to price insurance products.

3.5.2. *Introducing Constraints*

As mentioned, we may introduce constraints over the security portfolios. For example, we could add to the problem a budget limit of, say, € 8,000. Then, our problem would involve only those portfolios satisfying that constraint. In such case, the optimal portfolio would consist of the firewall, the 10 gbps DDoS protection system, and the comprehensive insurance, with a cost of € 7,650. Another example could refer to constraints on compulsory security controls, as certain insurance policies might demand their implementation before a policy is issued.

3.5.3. *Return on Security Investment*

Our formulation focused on choosing the best security portfolio. An additional aspect that could be addressed is calculating the return on security investment (ROSI) to assess the cost effectiveness of a cybersecurity budget (ENISA, 2012; Schatz & Bashroush, 2017). Calculating the optimal solution over a range of budgets (e.g., from € 5,000 to € 25,000) generates a function that, for a given budget, provides the optimal solution and its expected utility to explore the return on risk mitigation investments. Additionally, we could find the optimal increase in the portfolio so as to attain a

certain expected utility level or satisfy a certain risk appetite level.

3.5.4. Comparison with a Game-Theoretic Approach

Appendix C provides a comparison between our framework and a standard game-theoretic solution in a simplified cybersecurity example. The basic conclusions would be, first, that both approaches rely on different assumptions and, consequently, lead to different solutions; that the game-theoretic approach requires more stringent common knowledge assumptions that might not hold in cybersecurity; given that, we may view the ARA approach as more robust. Additionally, the proposed framework may be more adaptable to realistic cybersecurity scenarios with several potential attackers and several accidental and environmental threats as it more duly apportions various sources of uncertainty, as discussed in Merrick and Parnell (2011).

4. DISCUSSION

Current cybersecurity risk analysis frameworks provide relevant knowledge bases for understanding cyber threats, security policies, and impacts on business assets with dependencies on the IT infrastructure. However, most of such frameworks provide risk analysis methods that are not sufficiently formalized, nor comprehensive enough. Indeed, most of them suggest risk matrices as their main analytic basis, which provide a fast but frequently rudimentary study of threats. Hence, we have presented a formal framework supporting all steps relevant to undertake a comprehensive cybersecurity risk analysis. It implies structuring the cybersecurity problem as a decision model based on a multiagent influence diagram. It enables the assessment of beliefs and preferences of the organization regarding cybersecurity risks as well as the security portfolio and insurance it can implement to treat such risks. It takes into account, in addition to nonintentional threats, the strategic behavior of adversarial threats with ARA. We model the intentional factors through the decision problems of the attackers. The case introduced is a simplification of a real example but serves as template for complex problems. Among other things, we had to rely on expert judgment to assess the uncertainty nodes for which we lacked data. From the decision support point of view, ARA enables the calculation of optimal cybersecurity resource allocations, facilitating the selection of security and insurance portfolios. Furthermore, it also enables sensitivity analysis to evaluate whether the optimal portfolio remains as such, in case different elements affecting risk change.

Future work involves the application of this paradigm to study other cybersecurity adversarial problems, including granting a cyber insurance product and cyber reinsurance issues. The problem proposed here refers to strategic/tactical decisions; it would be interesting to develop dynamic schemes integrating strategic and operational decisions. Similarly, we shall address the development of parametric cyber insurance schemes in order to obtain premiums that reflect better risk management. We shall also pursue optimization algorithms beyond enumeration to reduce the computational burden.

When compared with standard approaches in cybersecurity, our paradigm provides a more comprehensive method, leading to a more detailed modeling of risk problems, yet, no doubt, more demanding in terms of analysis. We believe though that at many organizations, especially in critical infrastructures and sectors, the stakes at play are so high that this additional work should be worth the effort. Therefore, another relevant activity would be the development of a software environment that supports the implementation of our cybersecurity framework based on the R routines elaborated.

ACKNOWLEDGMENTS

This work is supported by the E.U. Horizon 2020 project 740920 CYBECO (Supporting Cyberinsurance from a Behavioural Choice Perspective). The work of DRI was supported by the Spanish Ministry of Economy and Innovation programs MTM2014-56949-C3-1-R, MTM2015-72907-EXP, MTM2017-86875-C3-1-R, the ESF-COST Action IS1304 on Expert Judgement, and the AXA-ICMAT Chair on Adversarial Risk Analysis.

APPENDIX A: ALGORITHMS

Algorithm 1. Estimating distribution over attacks (defense–attack).

For each *defense* e

  For $i = 1, \ldots, K$

    Generate

$$\left(U_A^i(t_2, c_t, c_c), P_A^i(c_t|t_1, t_2, e), P_A^i(c_c|t_1, t_2, e), P_A^i(t_1|e)\right) \sim F$$

    Compute

$$a^{*i} = \arg\max_a \iiint U_A^i(a, c_t, c_c)\, P_A^i(c_t|t_1, a, e)\, P_A^i(c_c|t_1, a, e)\, P_A^i(t_1|e) dt_1\, dc_c\, dc_t$$

  end

  Approximate

$$\hat{p}_A(a|e) = \frac{\#\{a^{*i} = a\}}{K}$$

end

Algorithm 2. Approximation of Defender's optimal portfolio.

$\psi(s, i) = 0$

For each (s, i)

  For $j = 1, \ldots, K$

    Generate

$$\left(m^j, q_n^j, q_i^j, b^j, l^j, a^j, v^j, o^j, f^j\right) \sim G$$

    Compute

$$c_s^j|s,\ c_i^j|i,\ g_i^j|i, b^j, q_i^j$$

    Compute

$$c_d^j = m^j + b^j + q_i^j + q_n^j + c_s^j + c_i^j - g_i^j$$

    Compute

$$\psi(s, i) = \psi(s, i) + \frac{u(c_d^j)}{K}$$

  end

end

Compute

$$(\hat{s}^*, \hat{i}^*) = \arg\max_{s,i} \psi(s, i)$$

APPENDIX B: NOTATIONS

Cybersecurity ARA framework notation

$p(\cdot)$	probability distribution
$u(\cdot)$	utility function
$P_A(\cdot)$	random probability distribution (attacker problem)
$U_A(\cdot)$	random utility function (attacker problem)
c_n	cost of normal system performance
ψ_n	expected utility under normal conditions
$t_1, \ldots, t_m$	threats
$c_1, \ldots, c_l$	costs of impacts on the assets
c	total costs
ψ_r	expected utility considering threats
e	security controls portfolio
c_e	security controls portfolio cost
$\psi(e)$	expected utility when portfolio e is implemented
ψ_e^*	expected utility of optimal portfolio
i	insurance
c_i	insurance cost
$\psi(e,i)$	expected utility when portfolio e and insurance i are implemented
$\psi_{e,i}^*$	expected utility of optimal portfolio and insurance
$u_A(\cdot)$	attacker utility function
$\psi_A(\cdot)$	expected utility for attacker
$A^*(e)$	optimal random attack given security portfolio e
e_p	preventive security controls portfolio
e_r	reactive security controls portfolio

Case study template notation

i	insurance
c_i	insurance cost
g_i	insurance coverage
s	security controls portfolio
c_s	security controls portfolio cost
f	fire probability
o	fire duration
v	number of computer virus infections
q	probability that a computer gets infected
w	percentage of time loss caused by computer virus
b	impact on facilities
q_i	insurable impact on computers
q_n	noninsurable impact on computers
c_d	total costs for defender
$u(c_d)$	defender utility
a	competitor attack
l	duration of DDoS
l_j	lenght of jth DDoS attack
r	market share loss ratio
m	impact on market share
e	attacker earnings
t	detection of attacker
c_t	cost when detected
c_a	result of attack
$u_A(c_a)$	attacker utility
$U_A(c_a)$	attacker random utility
$A^*(s)$	optimal random attack given security portfolio s
$\psi(s,i)$	expected utility when portfolio s and insurance i are implemented
(s^*, i^*)	optimal security portfolio s and insurance i
$\beta e(\cdot)$	beta distribution
$\mathcal{P}(\cdot)$	Poisson distribution
$\Gamma(\cdot)$	gamma distribution
$Tri(\cdot)$	triangular distribution
$\mathcal{U}(\cdot)$	uniform distribution
$\mathcal{B}(\cdot)$	binomial distribution
$\mathcal{N}(\cdot)$	normal distribution
$\mathcal{E}(\cdot)$	exponential distribution

APPENDIX C: COMPARISON WITH A GAME-THEORETIC APPROACH

This appendix compares our ARA framework with a standard game-theoretic (GT) approach by analyzing a simple example with both methods. We consider a defend–attack problem, in which a defender D has to decide (d) among three connecting options between two data centers in a campus shared with other institutions: using the campus network with encryption and other protection measures (d_1); using it without additional protection (d_2); or the most expensive, installing a dedicated line between the data centers (d_3). The danger resides in a potential targeted attacker A, insider to the campus, who decides whether to attack the defender's connection (a_1) or not (a_0). The result of the attack (r) leads to consequences related to data exfiltration, expressed as costs, for both the defender (c_D) and the attacker (c_A). They evaluate these consequences through utility functions (u_D and u_A) that incorporate their risk attitude. Fig. C1 represents the problem as an ID and Table C1 details the problem for various relevant defense–attack combinations.

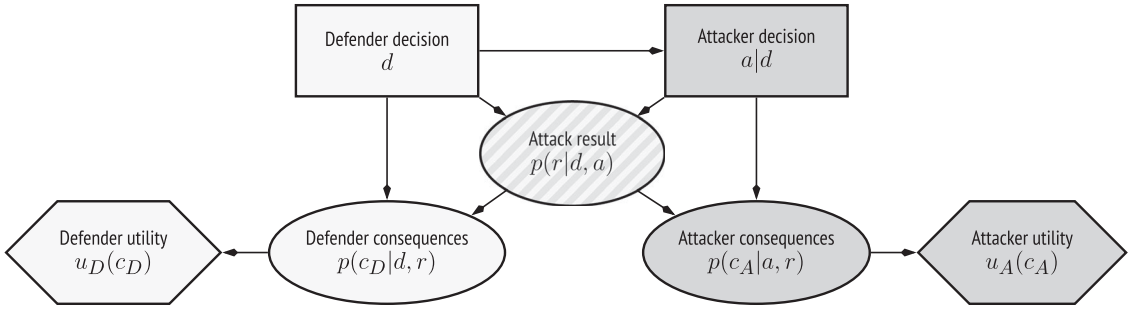

Fig. C1. Influence diagram representing the connecting problem.

Table C1. Defender and Attacker Elements

Defender Decision d	Attacker Decision a	Attack Result r	Defender Consequences c_D	Attacker Consequences c_A	Defender Utility $u(c_D)$	Attacker Utility $u(c_A)$
d_1	a_1	r_1	$s + kr_1$	$l - gr_1$	$1 - e^{\lambda(s+kr_1)}$	$e^{\mu(l+gr_1)} - 1$
	a_0	0	s	0	$1 - e^{\lambda s}$	0
d_2	a_1	r_2	kr_2	$l - gr_2$	$1 - e^{\lambda kr_2}$	$e^{\mu(l+gr_2)} - 1$
	a_0	0	0	0	0	0
d_3	a_1	–	–	–	–	–
	a_0	–	h	0	$1 - e^{\lambda h}$	0

Note: r_1 (r_2) is the attack result, in terms of fraction of data compromised, in case the defender uses the campus network with (without) protection d_1 (d_2); h is the cost of installing a new line between the data centers; s is the cost of taking the extra protection when using the campus network; k is the defender's cost relative to the fraction of data compromised; l is the attacker cost of executing the attack; g is the attacker's gain relative to the fraction of data extracted from the defender; λ is the defender risk aversion coefficient; μ is the attacker risk proneness coefficient.

Common ingredients to both approaches refer to the assessment of the defender elements. Suppose that we have $h = 100,000$, $s = 25,000$, and $k = 300,000$; her risk aversion coefficient is $\lambda = 3 \cdot 10^{-5}$; the attack result r_1, given the protection, follows a beta distribution $r_1 \sim \beta e(0.6, 1.4)$ (mean 0.3), whereas the attack result r_2, given the lack of protection, follows a beta distribution $r_2 \sim \beta e(0.36, 0.24)$ (mean 0.7).

Game-theoretic approach. Under common knowledge, we assume that the defender knows that the attacker's parameters are: $l = 12,000$; $g = 33{,}000$; $\mu = 1.8 \times 10^{-5}$; r_1 follows a beta distribution $\beta e(2.4, 6.7)$ (mean 0.2637); and r_2 follows beta distribution $\beta e(6.5, 4)$ (mean 0.619).

We first compute the attacker's best response to the defender choice d, which is $a^*(d) = \arg\max_a \psi_A(a, d)$, where $\psi_A(a, d) = \int\int u_A(c_A) p_A(c_A|a, r) p_A(r|d, a) \mathrm{d}r \, \mathrm{d}c_A$ is the attacker's expected utility. Knowing $a^*(d)$, we compute the defender's optimal decision from the game-theoretic perspective $d^*_{GT} = \arg\max_d \psi_D(a^*(d), d)$, where $\psi_D(a, d)$ is the defender's expected utility, defined in a similar fashion to that of the attacker. In our case, we have $a^*(d_1) = a_0$, $a^*(d_2) = a_1$, and $a^*(d_3) = a_0$, that is, attacking is the best decision for the attacker only when the defender uses the campus network without protection. We then compute the respective expected utilities as $\max(\psi_D(a^*(d_1), d_1), \psi_D(a^*(d_2), d_2), \psi_D(a^*(d_3), d_3))$ to find d^*_{GT}. In our case, $(-1.117, -19.086, -2960.141)$ and, thus, $d^*_{GT} = d_1$, using the campus network with the protection measures.

Adversarial risk analysis approach. Without common knowledge, we model the defender's beliefs about the attacker's judgment with random probabilities $P_A(\cdot)$ and random utilities $U_A(\cdot)$. Suppose that $l \sim \mathcal{U}(10,000, 20,000)$; $g \sim \mathcal{U}(10,000, 50,000)$; $\mu \sim \mathcal{U}(1 \times 10^{-5}, 2 \times 10^{-5})$; r_1 follows the random beta distribution $\beta e(\mathcal{U}(2, 4), \mathcal{U}(6, 8))$; and, similarly, r_2 follows $r_2 \sim \beta e(\mathcal{U}(5, 7), \mathcal{U}(3, 5))$.

We calculate the random optimal attack $A^*(d)$, given the defender's choice d, which is obtained through $\arg\max_a \int\int U_A(c_A) P_A(c_A|a, r) P_A(r|d, a) \mathrm{d}r \, \mathrm{d}a$. This leads to estimates $\hat{p}(a_1|d_1) = 0.180$, $\hat{p}(a_1|d_2) = 0.567$, $\hat{p}(a_1|d_3) = 0$, and the corresponding complementary probabilities for a_0. Knowing this, we calculate the defender's

expected utility, when d is her choice, as $\psi(d) = \iint u_D(c_D) p_D(c_D|a, r) p_D(r|d, a) \hat{p}(a|d) \mathrm{d}c_D \mathrm{d}r \mathrm{d}a$. The decisions' expected utilities are, respectively, $(-110.99, -1, 753.933, -19.086)$ and, thus, the ARA optimal defense is $d^*_{ARA} = d_3$, installing a dedicated line, which is different from the game-theoretic solution d_{GT}.

Comments. A first difference between the approaches is that GT assumes common knowledge. In our example, this entails that the defender knows the attacker's probability distributions and utility function. Alternatively, ARA does not assume such knowledge, but the defender needs to model her beliefs over the attacker judgments through random probability distributions and a random utility function. Consequently, a second difference is that GT informs the defender problem with the optimal decision of the attacker, whereas ARA provides a probability distribution of the attacker decision. Observe that the ARA approach may be seen as a way to induce robustness in the GT approach when we are not sure about the attacker assessments.

Similar comments hold for cases in which a game under a partial information approach is considered as common knowledge over the types prior and is required to approximate Bayes–Nash equilibria. See Rothschild, McLay and Guikema (2012) for additional discussion.

REFERENCES

Agence Nationale de la Sécurité des Systems d'Information (ANSSI). (1995). *Expression des Besoins et Identification des Objectifs de Sécurité*. Paris, France: Author.

Allodi, L., & Massacci, F. (2017). Security events and vulnerability data for cybersecurity risk estimation. *Risk Analysis*, *37*, 1606–1627.

Anderson, R. (2008). *Security engineering*. Hoboken, NJ: Wiley.

Andress, J. & Winterfeld, S. (2013). *Cyber warfare: Techniques, tactics and tools for security practitioners*. Amsterdam, The Netherlands: Elsevier.

Bagchi, A., Sprintson, A. & Singh, C. (2013). Modeling the impact of fire spread on an electrical distribution network. *Electric Power Systems Research*, *100*, 15–24.

Banks, D., Ros, J., & Ros Insua, D. (2015). *Adversarial risk analysis*. Boca Raton, FL: CRC Press.

Cavusoglu, H., Raghunathan, S., & Yue, W. (2018). Decision-theoretic and game-theoretic approaches to security investment. *Journal of Management Information Systems*, *5*(2), 281–304.

Central Communication and Telecommunication Agency (CCTA). (2003). *Risk analysis and management method*. London, U.K.: Author.

Clemen, R. T., & Reilly, T. (2013). *Making hard decisions with decision tools*. Boston, MA: Cengage Learning.

Cloud Security Alliance (CSA). (2016). *Cloud controls matrix*. Seattle, WA: Author.

Cooke, R., & Bedford, T. (2001). *Probabilistic risk analysis: Foundations and methods*. Cambridge, UK: Cambridge University Press.

Cox, L. A. (2008). What's wrong with risk matrices? *Risk Analysis*, *28*(2), 497–512.

Departamento de Seguridad Ciudadana (DSC de Victoria). (2009). *Memoria 2009 del Servicio de Prevencin Extincin de Incendios y Salvamentos*. Vitoria-Gasteiz, Spain: Author.

Dias, L. C., Morton, A., & Quigley, J. (2018). *Elicitation: State of the art and science*. Cham, Switzerland: Springer.

European Network and Information Security Agency (ENISA). (2012). *Introduction to return on security investment*. Heraklion, Greece: Author.

Fielder, A., Panaousis, E., Malacaria, P., Hankin, C., & Smeraldi, F. (2016). Decision support approaches for cyber security investment. *Decision Support Systems*, *86*, 13–23.

French, S. (1986). *Decision theory: An introduction to the mathematics of rationality*. New York, NY: Halsted Press.

French, S., & Ros Insua, D. (2000). *Statistical decision theory*. Hoboken, NJ: Wiley.

Galway, L. A. (2007). *Subjective probability distribution elicitation in cost risk analysis: A review*. Technical Report 410. Santa Monica, CA: Rand Corporation.

He, F., & Zhuang, J. (2017). Balancing pre-disaster preparedness and post-disaster relief. *European Journal of Operational Research*, *252*(1), 246–256.

Hubbard, D. W., & Seiersen, R. (2016). *How to measure anything in cybersecurity risk*. Hoboken, NJ: Wiley.

Incapsula. (2015). *Global DDoS threat landscape report: Attacks resemble advanced persistent threats*. Redwood Shores, CA: Author.

Information Security Forum (ISF). (2016). *Information risk assessment methodology 2*. London, UK: Author.

International Organization for Standardization (ISO). (2011). *ISO/IEC 27005. Information security risk management*. Geneva, Switzerland: Author.

International Organization for Standardization (ISO). (2013). *ISO/IEC 27001—Information security management systems—Requirements*. Geneva, Switzerland: Author.

Kaspersky Securelist, Russia. (2016). *DDoS attacks in Q4 2016*. Moscow, Russia: Author.

Lund, M. S., Solhaug, B., & Stlen, K. (2010). *Model-driven risk analysis: The CORAS approach*. Berlin: Springer.

Marotta, A., Martinelli, F., Nanni, S., Orlando, A., & Yautsiukhin, A. (2017). Cyber-insurance survey. *Computer Science Review*, *24*, 35–61.

Merrick, J., & Parnell, G. (2011). A comparative analysis of PRA and intelligent adversary methods for counterterrorism risk management. *Risk Analysis*, *31*, 1488–1510.

Ministerio de Hacienda y Administraciones Públicas (MINHAP). (2012). *Metodologa de Analisis y Gestin de Riesgos de los Sistemas de Informacin, Version 3*. Madrid, Spain: Author.

Mirkovic, J., & Reiher, P. (2004). A taxonomy of DDoS attack and DDoS defense mechanisms. *ACM SIGCOMM Computer Communication Review*, *34*, 39–45.

Mowbray, T. J. (2013). *Cybersecurity: Managing systems, conducting testing, and investigating intrusions*. Hoboken, NJ: Wiley.

National Institute of Standards and Technology (NIST). (2012). *NIST SP 800-30 Rev. 1—Guide for conducting risk assessments*. Gaithersburg, MD: Author.

National Technical Authority for Information Assurance (HMG). (2012). *HMG IA Standard Number 1*. London, UK: Author.

Ortega, J., Radovic, V., & Rios Insua, D. (2018). Utility elicitation. In L. Días, A. Morton, & J. Quigley (Eds.), *Elicitation: The science and art of structuring judgement*, (pp. 241–264). Berlin: Springer.

Panda Security. (2015). *Informe PandaLabs Q2 2015*. Bilbao, Spain: Author.

Rao, N. S. V., Poole, S. W., Ma, C. Y. T., He, F., Zhuang, J., & Yau, D. K. Y. (2015). Defense of cyber infrastructures against

cyber-physical attacks using game-theoretic models. *Risk Analysis*, *36*(4), 694–710.

Rios Insua, D. (1990). *Sensitivity analysis in multi-objective decision making*. Berlin: Springer.

Rios Insua, D., Banks, D., Rios, J., & Ortega, J. (2019). Adversarial risk analysis for structured expert judgement modelling. In French, Nane, Bedford, & Hanea (Eds.), *Expert judgement in risk and decision analysis*. Berlin: Springer.

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

Schatz, D., & Bashroush, R. (2017). Economic valuation for information security investment: A systematic literature review. *Information Systems Frontiers*, *19*(5), 1205–1228.

Schilling, A., & Werners, Br. (2016). Optimal selection of IT security safeguards from an existing knowledge base. *European Journal of Operational Research*, *248*(1), 318–327.

Shachter, R. D. (1986). Evaluating influence diagrams. *Operations Research 34*(6), 871–882.

Solutionary. (2013). *Global threat intelligence report*. United States.

The Common Criteria Recognition Agreement Members (CCRA). (2009). *Common criteria for information technology security evaluation, version 3.1 release 4*.

Thomas, P., Bratvold, R. B., & Bickel, J. E. (2014). *The risk of using risk matrices*. Paper presented at the Society of Petroleum Engineers Annual Technical Conference and Exhibition, New Orleans, LA.

Verisign. (2017). *Q1 2017 DDoS trends report*. Reston, VA: Author.

Wiper, M, Rios Insua, D., & Ruggeri, F. (2001). Mixtures of gamma distributions with applications. *Journal of Computational and Graphical Statistics*, *10*, 440–454.