An Adversarial Risk Analysis Framework for Cybersecurity

[1], D. Rios Insua[1], A. Couce-Vieira[1], J.A. Rubio[2], W. Pieters[3], K. Labunets[3], and D. G. Rasines[4]

[1]Instituto de Ciencias Matemáticas, Consejo Superior de Investigaciones Científicas, Madrid, Spain, david.rios@icmat.es, aitor.couce@icmat.es

[2]Analysis, Security and Systems Group, Complutense University of Madrid, Spain, joseantonio.rubio@fdi.ucm.es

[3]Safety & Security Science Group, Delft University of Technology, Delft, The Netherlands, w.pieters@tudelft.nl, k.labunets@tudelft.nl

[4]Department of Mathematics, Imperial College, London, UK, daniel.garcia-rasines16@imperial.ac.uk

Abstract

Cyber threats affect all kinds of organisations. Risk analysis is an essential methodology for cybersecurity as it allows organisations to deal with the cyber threats potentially affecting them, prioritise the defence 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 adversarial and non-intentional threats and the use of insurance as part of the security portfolio. A case study illustrating the proposed framework is presented, serving as template for more complex cases.

Keywords: cybersecurity, risk analysis, adversarial risk analysis, cyber insurance, resource allocation

1 INTRODUCTION

At present, all kinds of organisations are being critically impacted by cyber threats [3, 4]. The Cyberspace is even described as a fifth military operational space in which movements by numerous countries are common [25]. Risk analysis is a fundamental methodology to help manage such issues [11]. With it, organizations can assess the risks affecting their assets and what security controls should be implemented to reduce the likelihood of such threats and/or their impacts, in case they are produced.

Numerous frameworks have been developed to screen cybersecurity risks and support resource allocation, including CRAMM [7], ISO 27005 [23], MAGERIT [29], EBIOS [1], SP 800-30 [32], or CORAS [26]. Similarly, several compliance and control assessment frameworks, like ISO 27001 [22], Common Criteria [10], or CCM [9] provide guidance on the implementation of cybersecurity best practices. These standards suggest detailed security controls to protect an organisation's assets against risks. They have virtues, particularly their extensive catalogues of threats, assets and security controls providing detailed guidelines for the implementation of countermeasures and the protection of digital assets. Even though, much remains to be done regarding cybersecurity risk analysis from a methodological point of view. Indeed, a detailed study of the main approaches to cybersecurity risk management and control assessment reveals that they often rely on risk matrices, with shortcomings well documented in Cox [12]: compared to more stringent methods, the qualitative ratings in risk matrices (likelihood, 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 quantitatively very different risks, potentially inducing suboptimal security resource allocations. Hubbard and Seiersen [20] and Allodi an Massacci [2] provide additional views on the use of risk matrices in cybersecurity. Moreover, with counted exceptions like IS1 [33], those methodologies do not explicitly take into account the intentionality of certain threats. Thus, ICT owners may obtain unsatisfactory results in relation with the proper prioritisation of 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 nature and not in every country, have been introduced in recent years by companies like AXA, Generali, Allianz, or Zurich. However, cyber insurance has yet to take off [27].

In this paper we propose a more rigorous framework for risk analysis in cybersecurity. We emphasise adversarial aspects for better prediction of threats as well as include cyber insurance. Sect. 2 presents our framework, supported by a case study in Sect. 3. We conclude with a brief discussion.

2 A CYBERSECURITY ADVERSARIAL RISK ANALYSIS FRAMEWORK

We introduce our integrated risk analysis approach to facilitate resource allocation decision-making regarding cybersecurity. Our aim is to improve current cyber risk analysis frameworks, introducing dynamic schemes that incorporate all relevant parameters, including decision-makers' preferences and risk attitudes [8] and the intentionality of adversaries. Moreover, we introduce decisions concerning cyber insurance adoption to complement other risk management decisions through risk transfer. Fielder et al. [17] review and introduce various approaches to cyber security investment, which cover optimisation and/or game theoretic elements, under strong common knowledge assumptions. Our framework combines optimisation with an adversarial

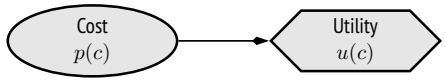

Figure 1: Basic influence diagram for performance evaluation.

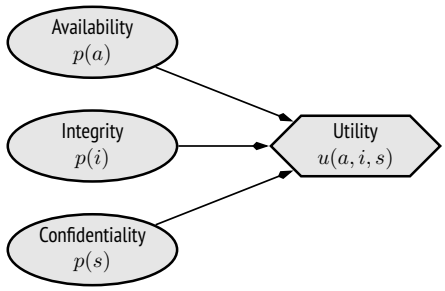

Figure 2: Cybersecurity attributes for performance evaluation.

risk analysis (ARA) approach to deal with adversarial agents.

We present the framework stepwise, analysing the elements involved progressively. We describe the models through influence diagrams (ID) and bi-agent influence diagrams (BAID) [6] detailing the relevant elements: assets, threats, security controls, costs and benefits. We provide a brief verbal description of the diagrams introduced and a generic mathematical formulation at each step.

2.1 System performance evaluation

Fig. 1 describes the starting outline for a system under study. Costs associated with system operation over the relevant planning period are indicated by c. Such costs are typically uncertain, modelled with a probability distribution $p(c)$. We introduce a utility function $u(c)$ [34] over costs to cater for risk attitudes. The evaluation of system performance under normal conditions, i.e. in absence of relevant cyber incidents, is based on its associated expected utility [18]

$$\psi_n = \int u(c)\, p(c)\ dc.$$

This basic scheme can be sophisticated in several directions. For example, there could be several performance functions. A typical case is to consider attributes concerning information availability (a), integrity (i) and confidentiality (s) [31], Fig. 2. These nodes could be, in turn, antecessors of the cost node. We use $p(a, i, s)$ as the distribution modelling uncertainty about system performance. If $u(a, i, s)$ represents the corresponding multi-attribute utility, the expected utility would be

$$\psi_n = \iiint u(a, i, s)\, p(a, i, s)\ da\, di\, ds.$$

We use $p(a, i, s)$ if interrelationships between such attributes are expected. If this were not so, e.g. in the case of independence, we would describe graphically the model as in Fig. 2, through

$$p(a, i, s) = p(a)\, p(i)\, p(s).$$

2.2 Cybersecurity risk assessment

Adopting the basic scheme in Fig. 1, on which we focus to simplify the exposition, we consider the problem of cybersecurity risk assessment in Fig. 3. For instance, consider a model with just two threats, one of them (t_1) physical (e.g., fire) and another one (t_2) cyber (e.g., DDoS attack[1]). Both t_1 and t_2 are random variables. We also consider two types of assets, one traditional (e.g., facilities) and the other cyber (e.g., computers). Impacts over these assets are, respectively, c_t and c_c and, typically, uncertain. If there is a relationship between them given either threat, the corresponding model would be of the form

$$p(c_t, c_c|t_1, t_2)\, p(t_1, t_2),$$

where $p(t_1, t_2)$ describes the probability of the threats happening, and $p(c_t, c_c|t_1, t_2)$ describes the probability over asset impacts, given the eventual occurrence of threats. Costs are added at the total cost node c, which aggregates those under normal circumstances with those due to the incidents. Then, the expected utility taking into account the threats and specific dependencies in Fig. 3 would be

$$\psi_r = \int \cdots \int \; u(c_n + c_t + c_c)\, p(c_n)\, p(c_t|t_1, t_2)\, p(c_c|t_1, t_2)\, p(t_1)\, p(t_2) \; dt_2\, dt_1\, dc_c\, dc_t\, dc_n.$$

We have assumed that consequences are additive, but we could have a generic utility $u(c_n, c_c, c_t)$. Finally, we evaluate the loss $\psi_n - \psi_r$ in expected utility considering the threats against that under normal conditions. When it is sufficiently large, incidents are expected to harm the system significantly and we should manage such risks.

The model can be extended to include a bigger number of threats and assets, as well as additional types of costs. Finally, several utility nodes could be incorporated to describe the preferences of multiple stakeholders.

2.3 Risk mitigation in cybersecurity risk management

The next step adds security controls to the model. We introduce a portfolio of them to reduce the likelihood of threats and/or their impact (Fig. 4). Examples of controls include firewalls, employee training, or making

[1] 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, making it inaccessible.

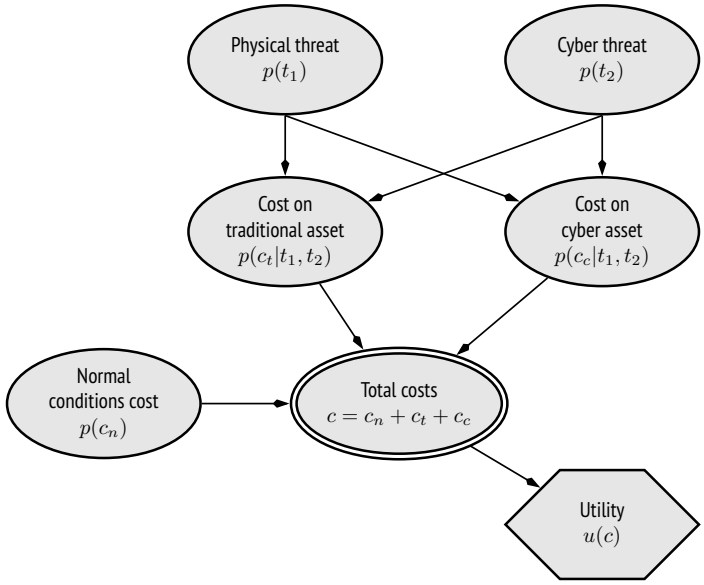

Figure 3: Risk assessment in cybersecurity.

regular backups. For simplicity, in Fig. 4 we assume that all controls 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 [30] managing access based on user privileges, make less likely a DDoS, 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 influence on threat likelihoods $p(t_1|e)$ and $p(t_2|e)$, as well as on asset impact likelihoods $p(c_t|t_1, t_2, e)$ and $p(c_c|t_1, t_2, e)$. All costs are aggregated in the total cost node c, under appropriate additivity assumptions.

In this case, the expected utility when portfolio e is implemented is

$$\psi(e) = \int ... \int \; u(c_n + c_e + c_t + c_c)\, p(c_n)\, p(c_e|e)\, p(c_t|t_1, t_2, e)\, p(c_c|t_1, t_2, e)\, p(t_1|e)\, p(t_2|e)\; dt_2\, dt_1\, dc_e\, dc_t\, dc_c\, dc_n.$$

We would then look for the maximum expected utility portfolio solving for

$$\psi_e^* = \max_{e \in E} \psi(e),$$

being E the set of feasible portfolios. Based on the available controls, we define portfolios that meet different constraints which may be economic (e.g., not exceeding a budget), legal (e.g., complying with data protection laws), logistic or physical.

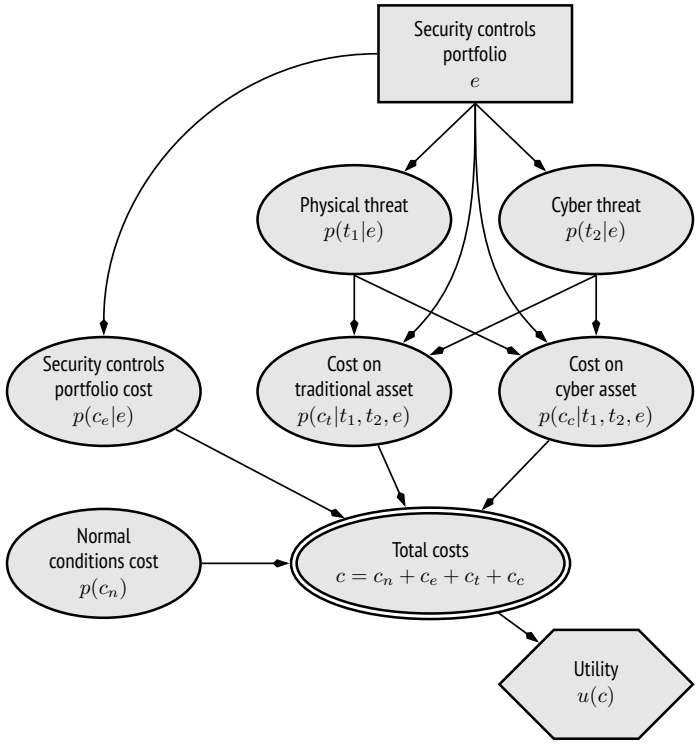

Figure 4: Risk assessment of cybersecurity controls.

2.4 Cyber insurance in cybersecurity risk management

As a relevant element of increasing interest, we introduce cyber insurance. Its costs will typically depend on the implemented portfolio of controls, as in Fig. 5: the better such portfolio is, the lower the premium will 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 them separately, since premiums will typically depend on the controls deployed. 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 typically affect impacts, modelled through $p(c_t|t_1, t_2, e, i)$ and $p(c_c|t_1, t_2, e, i)$. Costs are aggregated in the total cost node c. The expected utility when portfolio e is implemented together with insurance i is

$$\psi(e, i) = \int \cdots \int \; u(c_n + c_i + c_e + c_t + c_c)\, p(c_n)\, p(c_i|i, e)\, p(c_e|e)\, p(c_t|t_1, t_2, e, i)\, p(c_c|t_1, t_2, e, i) \times$$
$$\times \; p(t_1|e)\, p(t_2|e) \; dt_2\, dt_1\, dc_c\, dc_t\, dc_e\, dc_i\, dc_n.$$

We seek the maximum expected utility portfolio-insurance pair through

$$\max_{e \in E, i \in I} \psi(e, i),$$

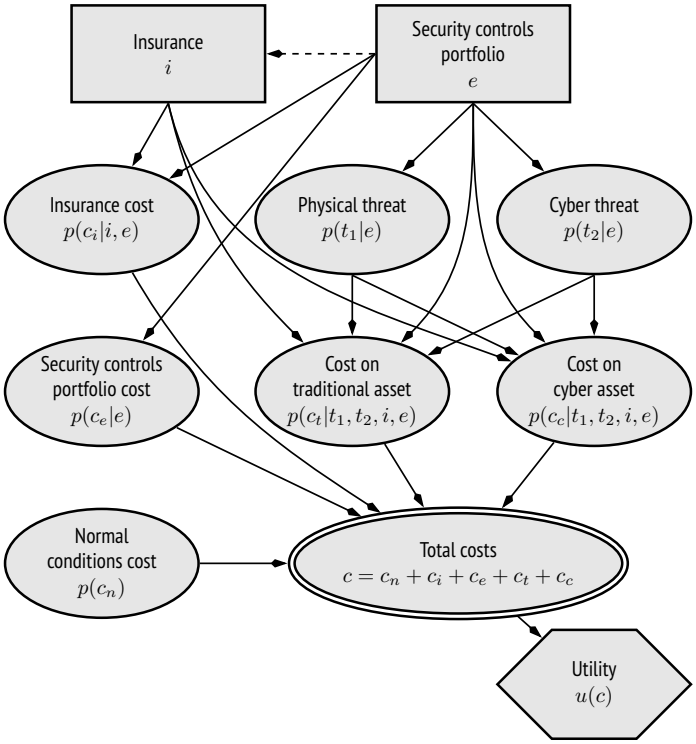

Figure 5: Cyber insurance in cybersecurity risk assessment.

where I represents the insurance catalogue. The pair (e, i) could be further restricted jointly, e.g., by a common budget constraint or legal requirements.

2.5 Adversarial risk analysis in cybersecurity

As discussed previously, intentionality is a key factor when analysing certain cyber threats. We shall use ARA [6] to model the intentions and strategic behaviour of adversarial cyber threats. Under ARA, the attacker has his own utility function u_A, seeking to maximise 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. First, the sequence *defence-attack*, in which the Defender deploys her security controls and the Attacker is able to observe them prior to attacking. Second, the sequence *defence-attack-defence*, in which the Defender deploys her preventive controls, then the Attacker observes them to decide his attack and, finally, the Defender recovers from the attack, should it be successful.

2.5.1 Defence-attack model

The original examples, Figs. 3 and 4 evolve into Fig. 6, modelling an adversarial case through a BAID with a Defender and an Attacker: physical threat t_1 remains unintentional whereas cyber threat t_2 becomes

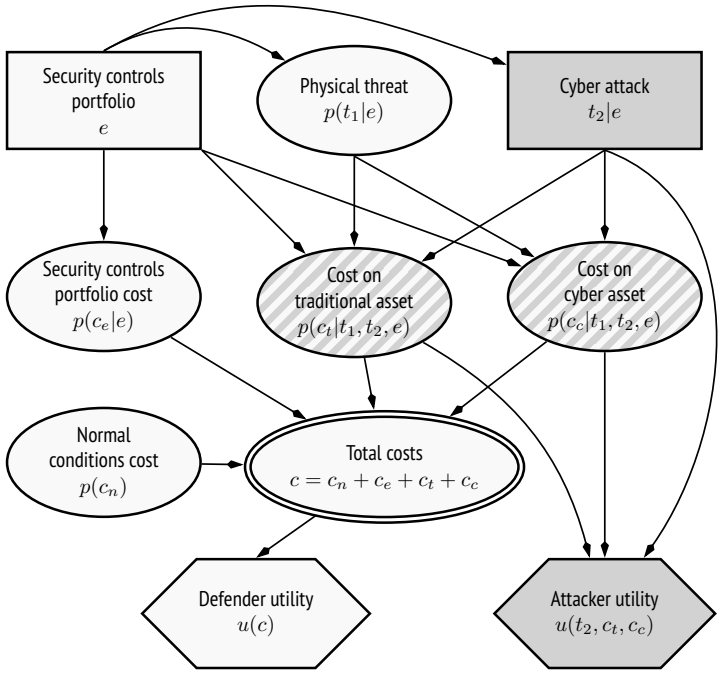

Figure 6: Adversarial risk analysis in cybersecurity: defence-attack problem.

adversarial through a decision node for the Attacker, who needs to decide whether or not to launch an attack to his benefit. It corresponds to a sequential defence-attack model [6]. The Defender problem is described in Fig. 4. Its resolution was covered in Sect. 2.3. There, the cyber attack is described probabilistically through $p(t_2|e)$, which represents the probability that the Defender assigns to cyber threat t_2 materialising, had portfolio e been adopted. However, the strategic nature of this problem, Fig. 6, requires the analysis of the Attacker decision about which attack to perform. Under the ARA defence-attack paradigm, the Defender should analyse the Attacker strategic problem in Fig. 7.

Specifically, given portfolio e, and assuming that the Attacker maximizes expected utility, the Defender would compute for each attack t_2, the expected utility for the Attacker

$$\psi_A(t_2|e) = \iiint u_A(t_2, c_t, c_c)\, p_A(c_t|t_1, t_2, e)\, p_A(c_c|t_1, t_2, e)\, p_A(t_1|e)\; dt_1\, dc_c\, dc_t,$$

where u_A and p_A are, respectively, the utilities and probabilities of the Attacker. The Defender must then find the attack maximising the Attacker's expected utility,

$$\max_{t_2 \in T_2} \psi_A(t_2|e),$$

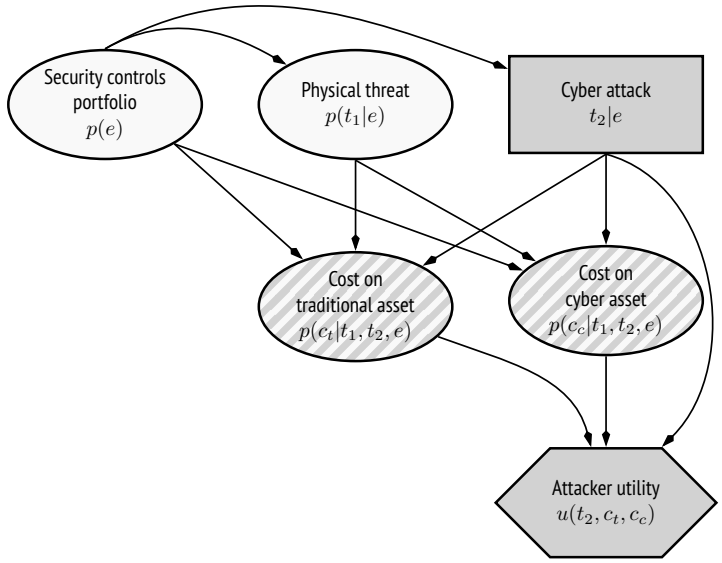

Figure 7: Attacker problem in the defence-attack model.

where T_2 is the attack set.

However, the Defender will not typically know u_A and p_A. Suppose we are capable of modelling her uncertainty about them with random probabilities P_A and a random utility function U_A [6]. Then, the optimal random attack given e is

$$T_2^*(e) = \arg\max_{t_2 \in T_2} \iiint U_A(t_2, c_t, c_c)\, P_A(c_t|t_1, t_2, e)\, P_A(c_c|t_1, t_2, e)\, P_A(t_1|e)\; dt_1\, dc_c\, dc_t.$$

Finally, the distribution over attacks that we were looking for satisfies

$$p(t_2|e) = P\big(T_2^*(e) = t_2\big),$$

assuming that T_2 is discrete (e.g., when referring to attack options), and, similarly, if they are continuous (e.g., when referring to attack efforts). Such distribution could be estimated through Monte Carlo (MC) simulation as in Algorithm 1 (Appendix), where the distribution of random utilities and probabilities is designated by

$$F = \Big(U_A(t_2, c_t, c_c), P_A(c_t|t_1, t_2, e), P_A(c_c|t_1, t_2, e), P_A(t_1|e)\Big)$$

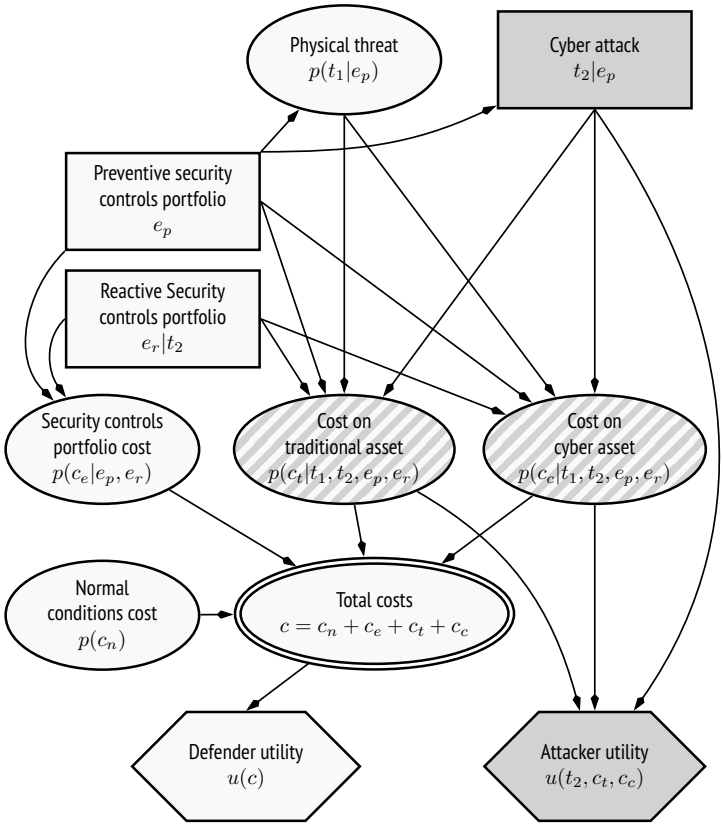

Figure 8: Adversarial risk analysis in cybersecurity: defence-attack-defence problem..

2.5.2 Defence-attack-defence model

As mentioned, 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 security controls e_p and reactive security controls $e_r|t_2$. This corresponds to a sequential defence-attack-defence model [6] in which the first move is by the Defender (preventive portfolio e_p), the second is by the Attacker (attack after observing preventive controls, $t_2|e_p$) and the third one is by the Defender (reactive portfolio $e_r|t_2$). The Defender problem is solved similarly to Sect. 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

$$\psi(e) = \int ... \int u(c_n + c_e + c_t + c_c)\, p(c_n)\, p(c_e|e_p, e_r)\, p(c_t|t_1, t_2, e_p, e_r)\, p(c_c|t_1, t_2, e_p, e_r) \times$$
$$\times\ p(t_1|e_p)\, p(t_2|e_p)\ dt_2\, dt_1\, dc_c\, dc_t\, dc_e\, dc_n.$$

We would then look for the maximum expected utility portfolio

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

where E_p and E_r, respectively, define constraints for preventive and reactive portfolios, some of which could be joint. The Attacker problem providing $p(t|e_p)$ would be solved in a similar fashion than in the defence-attack case.

3 A CASE STUDY TEMPLATE

We illustrate our framework for cybersecurity risk analysis with a defend-attack case study, which can serve as a template for more complex cases. The Defender is an SME dedicated to document management with 60 people and 90 computers. A cyber attack might affect, mainly, the online document management service. For confidentiality reasons, the number of relevant issues has been simplified and data conveniently masked. This simplification will allow us to better illustrate key modelling concepts and the overall scheme to follow for other case studies. Moreover, we include uncertain phenomena in which data are available and others in which it is not and, thus, we shall need to rely on expert judgement [14]. Prices and rates refer to Spain, where the incumbent SME is located.

In essence, we first structure the problem identifying assets, threats and security controls. The later may have implementation costs in exchange of reducing the threat likelihoods and/or eventual impacts. Subsequently, we assess the impacts that may have an effect on asset values to find the optimal risk management portfolio. Since adversarial threats are included, we also model the Attacker decision problem. Indeed, in this case there is a single potential Attacker which contemplates a DDoS attack with the objective of disrupting the SME services, causing an operational disruption and reputational damage with the consequent loss of customers, which would head to competitors, besides incurring in contractual penalties potentially affecting its continuity. Then, we simulate from this problem to obtain the attack probabilities, which feed back the Defender problem. In this way an optimal defence can be obtained. We consider a one-year planning horizon.

The problem we focus is on finding the optimal security portfolio and insurance product for the company, in the sense of maximising expected utility. Other formulations are discussed in Sect. 3.5.

3.1 Problem structuring

We structure the problem through the BAID in Fig. 9. Lighter nodes refer to issues concerning solely the Defender; darker nodes refer to issues relevant only for the Attacker; nodes with an stripped background

affect both. Should there be several attackers, we would use more background patterns or colours. Arcs have the same interpretation as in [38]. The only non-standard arc is that linking the security controls node and the attack node, meaning that the Attacker will implement his action once he identifies the controls adopted by the Defender.

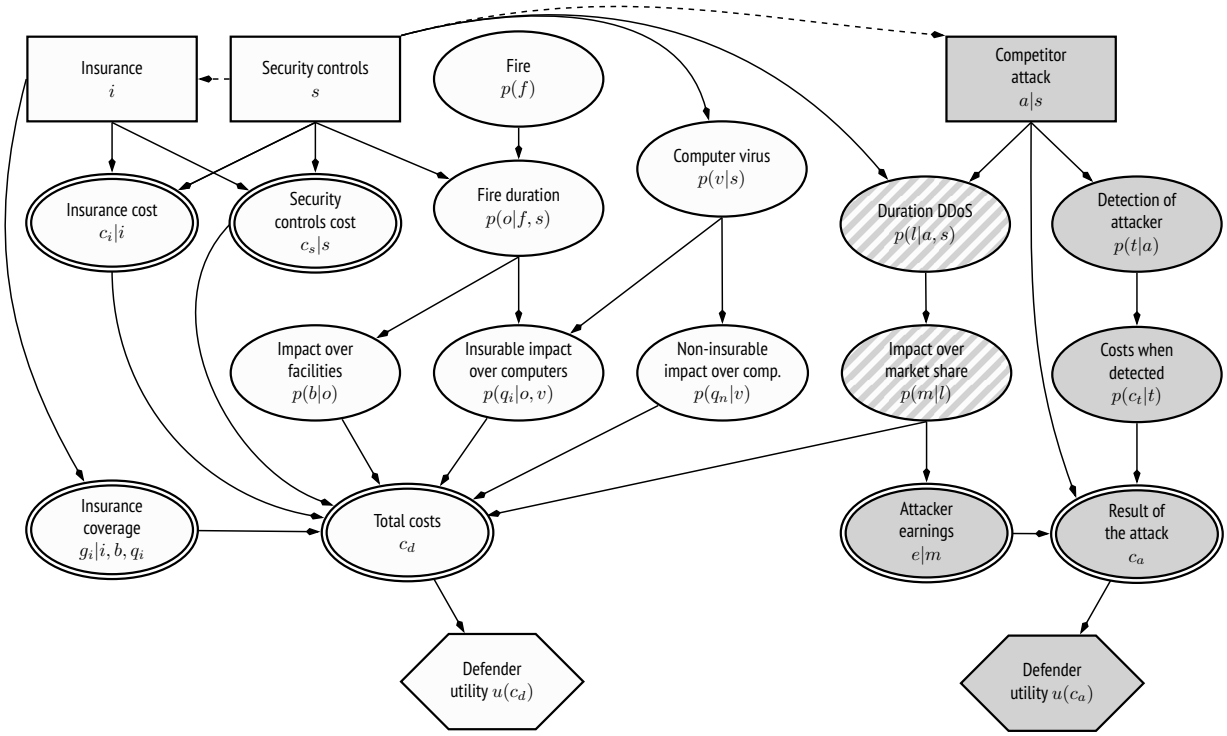

Figure 9: Case problem structure as a BAID.

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 the Introduction. Here we consider:

- *Facilities*: Offices potentially affected by threats. Without them, the organisation could not operate.
- *Computer equipment*: The data centre and workstations essential for this organisation. Should they be affected, costs could be substantial.
- *Market share*, directly impacting the company profits.

Other assets not considered in this case include the company's development software, business information, the mobile computing elements or the staff.

3.1.2 Non-intentional threats

We consider threats over the identified assets deemed relevant and having non-intentional character. This may include threats traditionally insurable as well as new ones potentially cyber insurable. We use a simplification of the catalogues in the methodologies in the introduction:

- *Fire*: It may affect facilities, as well as computers, which could even destroyed. No impact over market share is contemplated, as the organisation has a backup system. We assume that a fire can occur only by accident, not considering the possibility of sabotage.
- *Computer virus*: Aimed at disrupting normal operations of computer systems. We consider this threat non-intentional, as most viruses propagate ubiquitously: their occurrence tends to be random from the defender perspective. It may degrade computer performance.

We model each threat with a probabilistic node associated with the Defender problem. Other non-intentional threats, not considered here, could be water damage, power outages or employee errors.

3.1.3 Intentional threats

This category may include both cyber and physical threats. Again, we may use catalogues from, e.g., MAGERIT. We should identify the corresponding attackers, as well as their attack options available. In our case, we just consider a relevant attacker.

- *Competitor attack*: Our competitor may attempt a DDoS, to undermine the availability of the Defender's site, compromising her customer services. Should it happen, it would impact negatively the Defender's market share, damaging its reputation and, consequently, loosing customers to be gained by the Attacker. The decision is whether to launch the attack and the number of attempts.

We integrate attack options into a single decision node associated with the Attacker problem. Other intentional attacks, not modelled here, could include an abuse of access privileges, launching an advanced persistent threat, insiders or bombs.

3.1.4 Uncertainties affecting threats

We consider now those uncertainties affecting the Defender's assets.

- *Duration of DDoS*, will depend on the number of attacks and security controls deployed.
- *Fire duration*, which can be reduced with an anti-fire system.

Each one is modelled with a probabilistic node. Other related uncertainties could come, e.g., from a more detailed modelling of the virus (e.g., infection probability given the OS) or the fire propagation to adjacent buildings.

3.1.5 Attacker uncertainties

Additionally, we consider uncertainties that the Attacker might find relevant in his problem and affect only him.

- *Detection of Attacker.* If detected, his reputation would suffer and might face legal prosecution.

Each of them is modelled with a probabilistic node. Other attacker uncertainties include the number of customers affected by the DDoS or the performance of the attack platform.

3.1.6 Relevant security controls

We identify security controls relevant to counter the threats. We may use listings from the above mentioned methodologies. In our case we consider:

- *Anti-fire system.* It can detect a fire facilitating early mitigation.
- *Firewall.* It protects a network from malicious traffic.
- Implementation of *risk mitigation procedures* for cybersecurity and fire protection.
- *Cloud-based DDoS protection*, diverting DDoS traffic from the target to a cloud-based site absorbing malicious traffic.

We associate a Defender decision node with the security controls. 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. The premium will depend on the protected assets and contextual factors such as location, company type and, quite importantly, the implemented controls. Available insurance products are in Table 1.

We associate a Defender decision node with the insurance to be contracted. As its cost depends on the implemented controls, we include the corresponding decision node as a predecessor.

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

Table 1: Insurance product features.

3.1.8 Impacts for Defender

Having identified threats and assets, we present their potential impacts over the Defender's interests:

- *Impact over facilities*: Economic losses caused by fire over them.
- *Impact over computers*: Economic losses caused by fire or viruses over computers. We split them into insurable impacts and non-insurable ones. We need this split to calculate the eventual insurance coverage.
- *Impact over market share*: Costs due to market share lost.

We model each impact with a probabilistic node. We also consider the impacts associated with safeguards.

- *Cost of security controls* implemented by Defender.
- *Cost of insurance* acquired.
- *Insurance coverage.*

Finally, a node aggregates all Defender's consequences:

- *Total costs*: It summarises the above to establish the final monetary impact of the Defender problem.

The above cost nodes will be deterministic. Besides, we could also include other types such as corporate image or staff safety.

3.1.9 Impacts for Attacker

We consider the following impacts:

- *Attacker earnings* from increasing market share, transferred from that lost by the Defender.

- *Costs when detected*, covering eventual sanctions by the regulator, legal costs as well as loss of customers and reputation, if detected.
- The final *results of attack* combines all previous earnings and costs, as well as those of undertaking the attack, such as acquiring malicious tools or hiring hackers.

We model the costs when detected as a probabilistic node. The remaining nodes are deterministic.

3.1.10 Preferences

Value nodes describe how the corresponding agent evaluates consequences. We use the expected utility paradigm [18]. We, therefore, include these nodes:

- *Utility of Defender:* Models the Defender preferences and risk attitudes over the total costs.
- *Utility of Attacker:* It describes the Attacker preferences and risk attitudes.

We include a value node for each of the utility functions.

3.1.11 Defender and Attacker problems

Figs. 10 and 11 respectively represent the Defender and Attacker problems derived from the strategic problem in Fig. 9. For the Defender problem, this converts the Attacker's decision nodes into chance nodes and eliminates the Attacker's nodes not affecting the Defender problem, as well as the corresponding utility node. Similarly for the Attacker. We use both diagrams to guide judgement elicitation from the Defender.

3.2 Assessing the Defender's non-strategic beliefs and preferences

We provide now the quantitative assessment of the Defender beliefs and preferences not requiring strategic analysis. Some of them will be based on data and expert judgement, others just on expert judgement due to the lack of data typical in many cybersecurity environments. As a consequence, we populate most nodes in her problem. We incorporate in Sect. 3.3 those requiring strategic analysis. Finally, in Sect 3.4 we analyse 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:

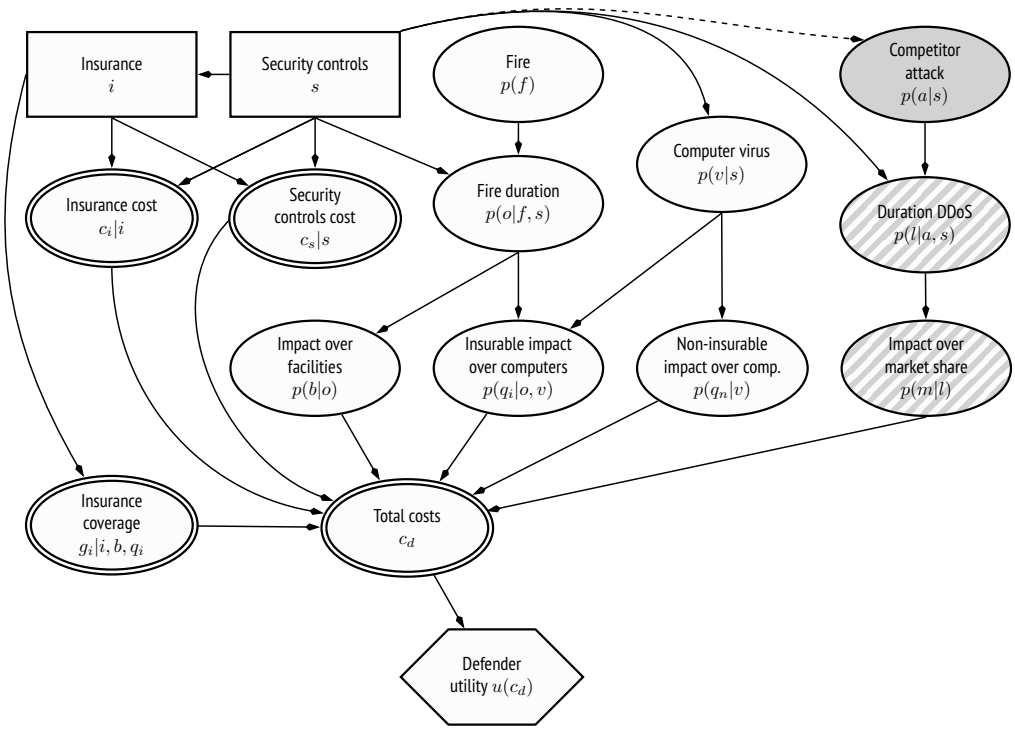

Figure 10: Defender problem.

- *Facilities*: Their value is 5,000,000€, reflecting only acquisition costs.
- *Computer equipment*: Valued at 200,000€, under similar considerations.
- *Market share*: Currently estimated at 50%. Translated into next year foreseen profits, we value it at 1,500,000€.

3.2.2 Modelling security controls

Security controls decision, s: The security portfolios that the Defender could implement derive from these options:

- Install an anti-fire system.
- Install a firewall to protect the infrastructure.
- Train employees on safety and cybersecurity procedures.
- Subscribe a cloud-based DDoS protection system with choice (2, 5, 10 or 1000) gbps.

We thus have 40 portfolios. These could be further constrained by, e.g., a budget, as in Sect. 3.5.

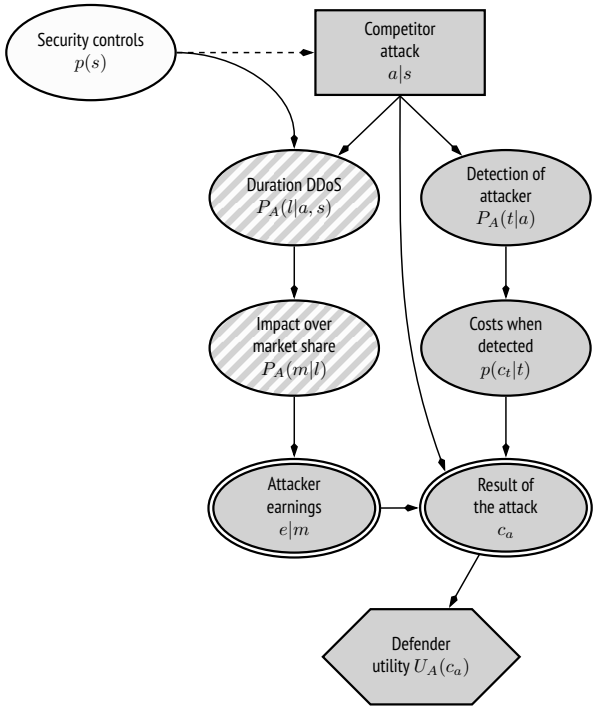

Figure 11: Attacker problem.

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

Table 2: Cost of individual security controls.

Cost of security controls, $c_s|s$: This node models the cost of implemented controls. Table 2 provides their costs, from which we derive the portfolio costs.

3.2.3 Modelling the insurance product

Insurance decision, i: This refers to the insurance product that the Defender could purchase (Table 3) once the controls have been selected.

Insurance cost, $c_i|i$: This models the insurance premiums. It depends on the controls implemented by the organisation (Table 3).

Insurance coverage, $g_i|i, b, q_i$: This node models g_i, the insurance product coverage reflected in Table 1. Traditional and comprehensive insurances cover 80% of burnt facilities and computer costs. The cyber and

Prod.	Security controls			
	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 3: Insurance product cost.

Year	Buildings	Fires
2005	1220	32
2006	1266	29
2007	1320	30
2008	1347	28
2009	1314	28

Table 4: Industrial fire data in Vitoria (2005-2009).

comprehensive insurances will cover 80% of the expenses related with virus removal.

3.2.4 Modelling fire risk

Fire likelihood, $p(f)$: This node provides the annual probability of suffering a fire in our facility. We use data from the Vitoria fire brigade [13], concerning interventions on industrial buildings (Table 4).

The fire rate remains fairly stable over the years. We estimate the probability that an organization suffers a fire in a year using a beta-binomial model with prior $\beta e(1/2, 1/2)$. The posterior would be

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

where x_i designates the number of fires affecting industrial buildings and n_i, that of such buildings in the $i-$th year, $i = 1, ..., 5$. Such distribution can be reasonably summarised through its posterior expectation, $\hat{p} = 0.022$, since the posterior variance is small. Its value is estimated from the probability that there are no fires, $p(0) = 1 - \hat{p} = 0.978$. The number of fires can be approximated with a Poisson $\mathcal{P}(0.022)$ distribution. However, we consider only the probability that one fire occurs, since probabilities beyond that are tiny ($p(f > 1) = 0.00024$. Thus, the number f of fires will follow

$$f \sim \min[1, \mathcal{P}(0.022)].$$

Fire duration, $p(o|f, s)$: It is a major impact determinant [5]: the longer the fire, the more damaging it will be. To study its duration, we employ the above Vitoria data. Fig. 12 presents the histogram of industrial fire durations, with modal duration between 30 minutes and one hour. Adopting the approach in [41], we

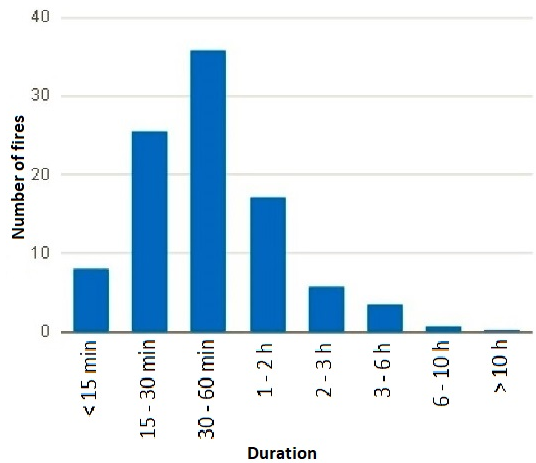

Figure 12: Industrial fire duration histogram. Vitoria, Spain (2005-2009).

model fire duration with a gamma $\Gamma(\text{shape} = \gamma, \text{scale} = \gamma/\mu)$ distribution. We assume a non-informative, but proper, prior $\gamma \sim Exp(0.01)$ and $\mu \sim \text{Inv-}\Gamma(1, 1)$. No analytical expression for the posterior distribution is available, but we can introduce a Markov chain MC scheme to sample from $\mu|data$ and $\gamma|data$, [41]. Based on this, we estimate that $\text{E}(\gamma|data) \approx 0.85$ and $\text{E}(\mu|data) \approx 78$.

The only security control among the proposed ones that may have an effect on fire duration is the anti-fire system, which enables faster fire detection. Using expert judgement [14], we determine its threshold duration under the proposed system with, respectively, suggested minimum, modal and maximum durations of 1, 10 and 60 min. To mitigate expert overconfidence [19], we consider a triangular distribution with quantiles 0.05 at 1 and 0.95 at 60 min, resulting in $Tri(0.8, 63, 10)$, which models the fire duration o if there is a fire ($f = 1$) and portfolio s contains the anti-fire system. On the other hand,

$$o \sim \Gamma(0.85, 0.01089)$$

if the portfolio does not contain the anti-fire system.

Fire impact: It models the impacts assuming that the fraction of affected assets is related with fire duration. After consulting with experts, we consider that a fire lasting 120 minutes would degrade the facilities by 100% in absence of controls. To simplify, we assume that the effect of fire duration is linear. This impact will be assessed in Sect. 3.2.7.

Additionally, the impact over 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

modelled in Sect. 3.2.7.

3.2.5 The computer virus risk

Computer virus likelihood, $p(v|s)$: This node provides the number of virus infections during a year in the organisation. The probability that a computer is infected in a month follows a binomial distribution $\mathcal{B}(m, q)$, where q represents the probability that a computer gets infected. For infection duration, we assume that the virus remains active until detected through appropriate controls. Then, it becomes eradicated by the system administrator. Various statistics suggest that the rate of virus infections worldwide is 33% [35], so we adopt $\hat{q} = 0.33$ as the probability that a computer is infected a certain month. The organisation has 90 computers, which we assume have the same security controls and are equally likely of being infected. Since the analysis is for 12 months, we use $m = 12 \cdot 90 = 1080$. Additionally, we consider the effect of our controls:

1. If a firewall is implemented, the probability that a computer gets infected is reduced to $\hat{p} = 0.005$, not completely eliminating the threat, even if this includes continuous updating based on latest virus signatures.

2. If maintenance is implemented, the infection probability gets reduced by 50%, with firewall or not, as this control entails improvements in the organisation such as imposing safety requirements on acquired systems.

The number v of infections is, therefore, modelled as in Table 5.

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	$v \sim \mathcal{B}(1080, 0.33)$

Table 5: Number of annual virus infections.

Computer virus impact: Viruses may affect the three information security dimensions.The impact on integrity and availability could lead to information corrupting or unavailability. Impacts over confidentiality are variable, as they depend on the stolen information. The average daily cost of these infections was estimated at € 2.683 [39], although this one varies according to the monetary value of the information and services the victim systems support. Bigger losses come from sophisticated campaigns (e.g., global ransomware like WannaCry) or targeted malware which, under our paradigm, should be better modelled as an adversarial threat. In our case, repairing a computer infected by a virus costs € 31 (two technician

hours). Insurance options potentially cover the removal of computer viruses. Therefore, this impact is modelled within the insurable aspects in Sect. 3.2.7.

Most computer viruses cause performance reduction in aspects such as initialisation of OS. Although small, this causes time losses to the user. We assume that most of the work time of the organisation is in front of a computer (70%), and that it would take, on average, five days (40 h of work) to detect the problem. We also assume that when a computer is infected, 28 hours of its usage are affected by the virus. We model the time loss as a uniform distribution $\mathcal{U}(0, 0.05)$, representing that the percentage of lost time caused by a virus is between 0 and 5 %. The hourly cost of the employee is € 20/hour. Therefore, for each virus infection, the cost would be $20 \times 28 \times \mathcal{U}(0, 0.05)$. Insurance options in node i do not cover this loss and, thus, we modelled it within the non-insurable aspects in Sect. 3.2.7.

3.2.6 Modelling the DDoS threat

We consider now non-strategic aspects of the DDoS threat. A model for the DDoS likelihood is in Sect. 3.3.

DDoS Duration, $p(l|a, s)$: This node models the duration l in hours of all successful DDoS attacks. Its length will depend on the intensity of the attacking campaign, how well crafted is the attack and the security controls implemented by the targeted organisation. Typical controls mitigating DDoS attacks involve configuring the digital system so that users and processes dedicate some resources for a certain period of time or distributing loads through a load balancer. An emerging alternative are cloud-based systems absorbing traffic from its customer site when they become victims of a DDoS. Otherwise, if no control is deployed, it would be virtually impossible to block such attack. Based on information in [24, 40], the average attack lasts 4 hours, averaging 1 gbps, with peaks of 10 gbps. We model l_j, the length of the j-th individual DDoS attack as a $\Gamma(4, 1)$, so that its average duration is 4 hours. This duration is conditional on whether the attack actually saturates the target, which depends on the capability of the DDoS platform minus the absorption of the cloud-based system. We assume that the Attacker uses a professional platform capable of 5 gbps attacks, modelled through a $\Gamma(5, 1)$ distribution. We then subtract the s_{gbps} absorbed by the protection system to determine whether the DDoS is successful, which happens when 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.

DDoS impact: The DDoS duration might cause a reputational loss that would affect the organisation market share. Recall that the current market share is 50% valued at € 1,500,000. To simplify, we assume that all market share is fully lost at a linear rate until lost in, say, 5-8 days of unavailability (120-192 hours of DDoS duration): in the fastest case the loss rate would be $0.5/120 = 0.00417$ per hour, whereas in the slowest one it would be 0.0026. We model this with a uniform distribution $\mathcal{U}(0.0026, 0.00417)$.

3.2.7 Modelling impacts over assets

We recall now the impacts over the assets.

Impact over facilities, $p(b|o)$: This node models monetary losses b due to the degradation of facilities by fire. Following Sect. 3.2.4, we model b through

$$b \sim 5000000 \times \min\left(1, \frac{o}{120}\right).$$

Insurable impacts over computers, $p(q_i|o, v)$: This models the monetary losses q_i due to degradation of computers to be covered by an insurance. This may be caused by fire, Sect. 3.2.4, and by repairing the computers infected with viruses, Sect. 3.2.5. We then model q_i through

$$q_i \sim 31v + 200000 \times \min\left(1, \frac{o}{120}\right).$$

Non-insurable impacts over computers, $p(q_n|v)$: This models the monetary losses q_m caused by degradation of computers not covered by insurance, due to the lost time caused by viruses over computer systems. Following Sect. 3.2.5, we model q_n through

$$q_n \sim 560v \times \mathcal{U}(0, 0.05).$$

Impact over market share, $p(m|l)$: This models the monetary value m of market share lost. Following Sect. 3.2.6, we use

$$m \sim \min[1500000, l \times \mathcal{U}(0.0026, 0.00417)].$$

Total costs for the Defender, $c_d|g_i, c_i, c_s, m, b, q_i, q_n$: This models the costs c_d suffered by the Defender through

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

being c_s the cost of security controls, c_i the cost of insurance, g_i the insurance coverage (which reduces losses) and m, b, q_i and q_n the impacts over assets earlier described.

3.2.8 Defender utility, $u(c_d)$:

The organisation is constant risk averse over costs. Its utility function is strategically equivalent to

$$u(c_d) = a - b\exp(k(-c_d)),$$

We adjust it calibrating the function with three costs: worst, best, and an intermediate one. The worst reasonable loss $\max c_d$ is based on the sum of all costs and impacts (except the computer virus one) which is equal to € 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 $\max c_d = 7000000$. The best loss is $\min c_d = 0$. For an intermediate cost c_d^*, we find its probability equivalent [34] α so that $u(c_d^*) = \alpha$. For instance, asking the company, we have $u(c_d^* = 2660000) \simeq .5$. Additionally, we rescale the costs to the (0,1) range through $1 - \frac{c_d}{7000000}$. Then, the utility function is

$$u(c_d) = \frac{1}{\mathrm{e} - 1}\left[\exp\left(1 - \frac{c_d}{7000000}\right) - 1\right].$$

3.3 Assessing the Attacker's random beliefs and preferences

In the Defender problem, the competitor attack is a probabilistic node modelling the number of attacks launched by the Attacker, given the security controls implemented by the Defender. We model the Attacker problem and simulate it to forecast its actions to obtain the probability distribution.

We must estimate the probability that the Attacker executes the DDoS, given the Defender controls implemented. For that, we consider his decision problem in Fig. 11. Its solution would give the Attacker's optimal action. However, as argued in Sect. 2.5, we model our uncertainty about his preferences and beliefs through random utilities and probabilities to find the random optimal attack.

3.3.1 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 against his attack.

3.3.2 Competitor attack decision: $a|s$

In the attacker problem, it is reflected in a decision node, modelling how many attacks (between 0 and 30) the DDoS campaign will consist of. Attackers usually give up once the attack has been mitigated and move onto the next target or try other disruption methods. However, when the sole objective is the victim, the Attacker might continue the campaign for several days, causing a pervasive impact. In our case, we assume that a DDoS platform would need a day to deploy their resources to launch a powerful and hidden DDoS.

3.3.3 Duration of the DDoS: $P_A(l|a, s)$

We base our estimation on that of the Defender (Sect. 3.2.6). The length of the j-th individual DDoS attack is modelled through a random gamma distribution $\Gamma_{\text{length}}(\upsilon, \upsilon/\mu)$ with $\upsilon \sim \mathcal{U}(3.6, 4.8)$ and $\upsilon/\mu \sim \mathcal{U}(0.8, 1.2)$, so that we add uncertainty about the average duration (between 3 and 6 hours) and the dispersion. Similarly, the attack gbps are modelled 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} to Γ_{gbps} to determine whether the DDoS is successful, which happens when its traffic overflows the protection system. As in Sect. 3.2.6, the number l of hours for which the site is unavailable during the campaign is modelled as

$$l = \sum_{j}^{a} l_j,$$

with $l_j \sim \Gamma_{\text{length}}$ if $\Gamma_{\text{gbps}} - s_{\text{gbps}} > 0$, and $l_j = 0$ otherwise.

3.3.4 Impact over market share: $P_A(m|l)$

We base our estimation on that of the Defender (Sect. 3.2.7), adding some uncertainty around such assessment. The market share value and percentage are not affected by the uncertainty, as this information is available to both agents. However, we model uncertainty in the market loss rate so that the fastest one (5 days in the Defender problem) is between 4 and 6 days in the Attacker problem and the slowest one (8 for Defender) is between 7 and 9. Therefore, the random distribution describing the market loss m is

$$m \sim \min\left[1500000, l \times \mathcal{U}(\alpha, \beta)\right]$$

with $\alpha \sim \mathcal{U}(0.0021, 0.0031)$ and $\beta \sim \mathcal{U}(0.00367, 0.00467)$.

3.3.5 Attacker earnings: $e|m$

This node models the Attacker gain e in terms of market share, derived from the DDoS duration. As the sole competitor, we assume that e corresponds to the share lost by the defender $e = m$. The random uncertainty in the earnings is derived from the randomness of the preceding nodes.

3.3.6 Detection of Attacker: $P_A(t|a)$

This node represents the chance of the Attacker being detected. In most cyber attacks, the attacker is not identified or prosecuted[2]. Detection probabilities are estimated via expert judgement at 0.2%, should the Attacker attempt a DDoS. He actually gambles his detection through a binomial distribution $\mathcal{B}(a, 0.002)$ where the number of trials is the number a of attacks and the detection probability is 0.002. To add some uncertainty, we model the detection probability for each attack through a beta distribution $\beta e(2, 998)$[3]. Therefore, the distribution determining the detection of the attacker t is modelled through a random binomial distribution that produces the output *detected* if $\mathcal{B}(a, \phi) > 0$ with $\phi \sim \beta e(2, 998)$, and *not detected*, otherwise.

3.3.7 Costs for Attacker when detected: $p_A(c_t|t)$

This node models the consequences associated with being detected when executing a DDoS. As a competitor, if the Attacker is disclosed, it would entail a serious discredit, together with compensation and legal costs besides incurring criminal responsibilities. To fix ideas, we use this cost decomposition:

- Expected reputational costs, due to the necessary communication actions to preserve credibility: € 550,000.
- Expected legal costs: € 30,000.
- Expected civil indemnities and regulatory penalties: € 350,000.
- Expected suspension costs, related with losses derived from prohibition to operate for some time: € 1,500,000.

To add uncertainty, we model the costs as a normal distribution with mean 2430000 and standard deviation 400000, i.e.,

$$c_t|t \sim \mathcal{N}(2430000, 400000).$$

[2]For instance, the FBI Internet Crime Compliant Center prosecuted two cases, and investigated 73, of nearly 298,728 complaints received in 2016 [16]

[3]Its mean is 0.002

3.3.8 Result of attack: $c_a|e, c_t, a$

This node combines the attacker earnings and costs if detected, as well as the cost of undertaking the attacks. To estimate these, we consider that using a botnet to launch the DDoS attack would cost on average around € 33 per hour [21]. Each attack would take one day, entailing costs of € 792. Therefore,

$$c_a = e - c_t - 792a.$$

3.3.9 Attacker's random utility: $U_A(c_a)$

We assume that the Attacker is risk prone, with utility function strategically equivalent to

$$u(c_a) = (c'_a)^{k_a}, \qquad k > 1,$$

where c'_a are the costs c_a normalised to $[0, 1]$, and k_a the risk seeking attitude of the attacker. To induce uncertainty, we assume k_a follows a $\mathcal{U}(8, 10)$ distribution. Therefore, the attacker random utility is

$$U_A(c_a) = (c'_a)^{K_a}$$

with $K_a \sim \mathcal{U}(8, 10)$.

3.3.10 Simulating the Attacker problem

Summarising the earlier assessments, the *distribution of random utilities and probabilities in the Attacker problem* is

$$F = \Big(U_A(c_a), p_A(c_t|t), P_A(t|a), P_A(m|l), P_A(l|a, s)\Big).$$

We calculate the *random optimal attack*, given the security controls s implemented, as

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

To approximate it, we may use an MC approach as in Algorithm 1 (see Appendix), which we implemented in R. For each s, the size of the DDoS protection system, we can assess the distribution of the random optimal attack. Table 6 displays the probabilities of the attacks, conditional on the protection implemented, with $K = 1000$. For instance, if the security portfolio contains no DDoS-protection system, an attack seems certain and its duration would be between 18 and 30 attacks (being 29 and 30 the most likely attack sizes). From this, we create the probability distribution $p(a|s)$, so that the Defender problem is fully specified and

ready to be solved.

DDoS	Number of attempts															
prot. system	**0**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**
1 tbps	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
2gbps	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

DDoS	Number of attempts														
prot. system	**16**	**17**	**18**	**19**	**20**	**21**	**22**	**23**	**24**	**25**	**26**	**27**	**28**	**29**	**30**
1 tbps	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
2gbps	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

Table 6: Conditional probability table for random optimal attacks.

3.4 Solution of the Defender problem

Summarising earlier assessments about the Defender problem, the corresponding probabilities are

$$G = \Big(p(m|l), p(q_n|v), p(q_i|o,v), p(b|o), p(l|a,s), p(a|s), p(v|s), p(o|f,s), p(f)\Big).$$

The expected utility when the security portfolio s is implemented together with insurance i is

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

We can calculate the *optimal allocation* as the maximum expected utility portfolio-insurance pair

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

We may use Algorithm 2 (see Appendix), to approximate the portfolio expected utilities and the optimal portfolio for the Defender. We have implemented it in R to calculate them (Table 7). Specifically, *the best portfolio* consists of:

- 1 tpbs cloud-based DDoS protection system.
- Firewall.
- Anti-fire system.

- Comprehensive insurance.

Anti-fire	Firewall	Procedure	DDoS protection	Insurance	Expected utility
anti-fire	firewall	no procedure	1 tbps	comprehensive	0.9954
anti-fire	firewall	no procedure	1 tbps	traditional	0.9950
no anti-fire	firewall	no procedure	1 tbps	comprehensive	0.9949
...	...	...	...	...	...
no anti-fire	no firewall	no procedure	no protection	no insurance	0.8246
no anti-fire	firewall	no procedure	no protection	cyber	0.8246
anti-fire	no firewall	no procedure	no protection	no insurance	0.8242

Table 7: Expected utility for 3 best and worst combinations of controls and insurance.

Besides the ranking of countermeasures, we can obtain additional information from the simulation. For instance, the best security controls contain a firewall, 1 tbps DDoS protection and no risk management procedure. Additionally, the best portfolios also includes 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

By introducing variations in the probabilities (e.g., probability of fire), we can evaluate the robustness of the previous solution by checking whether variations in the probabilities and parameters of the model alter the optimal solution or the relevance of different controls. This is specially relevant in a case like ours with little differences in expected utility among top alternatives and many inputs are purely of judgemental nature. The approach would require the implementation of additional algorithms for sensitivity analysis that indicate whether a small deviation in a parameter may lead to a large effect in the outcome of the model [36]. 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 to price insurance products, as well as for finding the best portfolio for different cybersecurity budgets.

3.5.2 Introducing constraints

As we mentioned, we may introduce constraints over the security portfolios. For example, we could add to the problem a budget limit of, say, €15,000. Then, our problem could involve only those portfolios satisfying such constraint. We can also consider constraints of insurance on security controls as in insurance policies there are some requirements regarding controls that the company should comply with to be insured. Other types of constraints could be dealt with similarly.

3.5.3 Return on security investment

Our formulation focused on choosing the best portfolio, but an additional aspect that could be addressed with our model is calculating the return on security investment (ROSI) to assess the cost effectiveness of a cybersecurity budget [15, 37]. Calculating the optimal solution over a range of budgets (e.g., from €5,0000 to €25,000) allows generating a function that, for a given budget, gives the optimal solution and 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 reach a certain risk appetite level.

4 DISCUSSION

Current cybersecurity risk analysis frameworks provide a thorough knowledge base for understanding cyber threats, security policies and impacts over assets that depend on the digital infrastructure. However, such frameworks provide risk analysis methods that are not sufficiently formalised, neither comprehensive enough. Most of them suggest risk matrices as their main analytic basis, which provide a fast but frequently rudimentary study of risks.

Hence, we present an ARA framework providing a formal method 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 diagrams. ARA enables the assessment of beliefs and preferences of the organisation regarding cybersecurity risks as well as the security portfolio and insurance they can implement to treat such risks. It takes into account, in addition to non-intentional threats, the strategic behaviour of adversarial threats. We model the intentional factor through the decision problems of the Attackers. The case introduced is a simplification of a real example but serves as a template for other cases. Among other things, we had to rely on expert judgement for the uncertainty nodes for which we lacked data.

From the decision-making 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 optimal, in case different elements affecting risk change. This may be used for insurance pricing.

Future work involves the application of this paradigm to study other cybersecurity adversarial problems. The proposed problem 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, aimed at supporting the obtention of premiums that, as complement of the implemented controls, facilitate more effective risk management. Another relevant activity would be

the development of a software environment that supports the implementation of the ARA framework for cybersecurity based on the R routines developed, as well as optimisation algorithms beyond enumeration, to reduce computational burden.

When compared with standard approaches in cybersecurity, our paradigm entails a more comprehensive method leading to a more detailed modelling of risk problems, yet more demanding in terms of analysis. We believe, though, that at many organisations, especially, critical infrastructures and sectors, the stakes at play are so high that the entailed additional work should be worth the effort.

REFERENCES

[1] Agence Nationale de la Sécurité des Systèmes d'Information (France). 2010. *Expression des Besoins et Identification des Objectifs de Sécurité.*

[2] Allodi, L., Massacci, F. 2017. "Security Events and Vulnerability Data for Cybersecurity Risk Estimation". *Risk Analysis*, Vol. 37, pp. 1606–1627.

[3] Anderson, R. 2008. *Security Engineering*, Wiley.

[4] Andress, J. and Winterfeld, S. 2013. *Cyber Warfare: Techniques, Tactics and Tools for Security Practitioners.* Elsevier.

[5] Bagchi, A., Sprintson, A. and Singh, C. 2013. "Modeling the Impact of Fire Spread on an Electrical Distribution Network". *Electric Power Systems Research*, Vol. 100, pp. 15–24.

[6] Banks, D., Ríos, J. and Ríos Insua, D. 2015. *Adversarial Risk Analysis.* Francis and Taylor.

[7] Central Communication and Telecommunication Agency (UK). 2003. *Risk Analysis and Management Method.*

[8] Clemen, R. T., Reilly, T. 2013. *Making Hard Decisions with Decision Tools.* Cengage Learning.

[9] Cloud Security Alliance. 2016. *Cloud Controls Matrix.*

[10] The Common Criteria Recognition Agreement Members. 2009. *Common Criteria for Information Technology Security Evaluation, Version 3.1 Release 4.*

[11] Cooke, R. and Bedford. T. 2001. *Probabilistic Risk Analysis: Foundations and Methods.* Cambridge University Press.

[12] Cox, L. A. 2008. "What's Wrong with Risk Matrices?". *Risk Analysis*, Vol. 28, No. 2, pp. 497–512.

[13] Departamento de Seguridad Ciudadana, Ayto. de Vitoria-Gasteiz (Spain). 2009. *Memoria 2009 del Servicio de Prevención Extinción de Incendios y Salvamentos*

[14] Dias, L.C., Morton, A. and Quigley, J. 2018. *Elicitation: State of the Art and Science.* Springer.

[15] European Network and Information Security Agency. 2012. *Introduction to Return on Security Investment.*

[16] Federal Bureau of Investigation, Internet Crime Compliant Center (USA). 2016. *2016 Internet Crime Report*

[17] Fielder, A., Panaousis, E., Malacaria, P., Hankin, C. and Smeraldi, F. 2016. "Decision Support Approaches for Cyber Security Investment". *Decision Support Systems*, Vol. 86, pp. 13-23.

[18] French, S. and Ríos Insua, D. 2000. *Statistical Decision Theory.* Wiley.

[19] Galway, L. A. 2007. *Subjective probability distribution elicitation in cost risk analysis: A review.* Tech. Rep. 410, Rand Corporation.

[20] Hubbard, D.W. and Seiersen, R. 2016. *How to Measure Anything in Cybersecurity Risk.* John Wiley & Sons.

[21] Incapsula (USA). 2015. *Global DDoS Threat Landscape Report: Attacks Resemble Advanced Persistent Threats.*

[22] International Organization for Standardization. 2013. *ISO/IEC 27001 – Information Security Management Systems - Requirements.*

[23] International Organization for Standardization. 2013. *ISO/IEC 27005. Information Security Risk Management.*

[24] Kaspersky, Securelist (Russia). 2016. *DDoS attacks in Q4 2016.*

[25] Leak Source. 2014. "CSEC Document Reveals Suspected France Intelligence Spyware "Babar"". [Retrieved 25/Sep/2017]

[26] Lund, M.S., Solhaug, B. and Stølen, K. 2010. *Model-driven risk analysis: the CORAS approach* Springer.

[27] Marotta, A., Martinelli, F., Nanni, S., Orlando, A. and Yautsiukhin, A. 2017. "Cyber-insurance survey". *Computer Science Review.* Vol. 24, pp 35–61

[28] Milke, J. A., Kodur, V. and Marrion, C. 2002. "An overview of fire protection in buildings". *Appendix A, World Trade Center Building Performance Study.* Federal Emergency Management Agency (USA).

[29] Ministerio de Hacienda y Administraciones Públicas (Spain). 2012. *Metodología de Análisis y Gestión de Riesgos de los Sistemas de Información, version 3.*

[30] Mirkovic, J. and Reiher, P. 2004. "A taxonomy of DDoS attack and DDoS defense mechanisms." *ACM SIGCOMM Computer Communication Review* Vol.34, pp 39–45.

[31] Mowbray, T. J. 2013. *Cybersecurity: Managing Systems, Conducting Testing, and Investigating Intrusions.* Wiley.

[32] National Institute of Standards and Technology (USA) *NIST SP 800-30 Rev. 1 – Guide for Conducting Risk Assessments.*

[33] National Technical Authority for Information Assurance (UK). 2012. *HMG IA Standard Number 1.*

[34] Ortega, J., Radovic, V. and Rios Insua, D. 2018. "Utility elicitation". In: Días, L., Morton, A. and Quigley, J., editors. *Handbook of judgement elicitation.* Springer.

[35] Panda Security (Spain). 2015. *Informe PandaLabs Q2 2015.*

[36] Rios Insua, D. 1990. *Sensitivity Analysis in Multi-objective Decision Making.* Springer.

[37] Schatz, D. and Bashroush, R. 2017. "Economic valuation for information security investment: a systematic literature review". *Information Systems Frontiers*, Vol. 19, No. 5, pp 1205–1228.

[38] Shachter, R.D. 1986. "Evaluating influence diagrams". *Operations Research* Vol. 34, No. 6, pp 871–882.

[39] Solutionary (US). 2013. *Global Threat Intelligence Report.*

[40] Verisign (USA). 2017. *Q1 2017 DDoS Trends Report*

[41] Wiper, M, Rios Insua, D. and Ruggeri F. 2001. "Mixtures of gamma distributions with applications." *Journal of Computational and Graphical Statistics* Vol. 10, pp. 440–454.

APPENDIX

Algorithm 1 Estimating distribution over attacks (defence-attack).

For each *defence* e

  For $i = 1, \ldots, K$

    Generate

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

    Compute

$$t_2^i = \arg\max_{t_2} \iiint U_A^i(t_2, c_t, c_c)\, P_A^i(c_t|t_1, t_2, e)\, P_A^i(c_c|t_1, t_2, e)\, P_A^i(t_1|e) dt_1\, dc_c\, dc_t$$

  end

  Approximate

$$\hat{p}_A(t_2|e) = \frac{\#\{t_2^i = t_2\}}{K}$$

end

Algorithm 2 Approximation of Defender's optimal portfolio.

$\psi(s, i) = 0$

For each (s, i)

  For $j = 1, \ldots, 1000$

    Generate

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

    Compute

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

    Compute

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

    Compute

$$\psi(s, i) = \psi(s, i) + \frac{u(c_d^j)}{1000}$$

  end

end

Approximate

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