CYBECO

Supporting Cyberinsurance from a Behavioural Choice Perspective

D3.1: Modelling framework for cyber risk management

Due date: M12

Abstract:

This deliverable presents the CYBECO modelling framework for cybersecurity risk management. We move beyond standard cybersecurity frameworks which do not take into account the intentionality of certain cyber threats and are essentially based on risk matrices. We provide models that overcome the ordinal scales used in risk matrices, properly consider intentionality of attackers, allow for repeated interactions between defenders and attackers, include behavioural elements in relation with risk aversion and, very importantly, reflect the decision of adopting cyber insurance. The core model (Task 3.1) refers to supporting an organization which needs to decide its optimal cybersecurity resource allocation, including what security controls and insurance product to acquire, if any. Yet we include two additional models referring to the reinsurance needs of an insurance company and to the decision of granting, or not, an insurance product to a customer. We provide a generic model (Task 3.2) that facilitates the elicitation of stakeholder objectives and attitudes towards risk to facilitate the implementation of the framework. We also describe several computational enhancements (Task 3.3) that would facilitate its treatment in complex large-scale settings. Finally, we include a description of how to implement the models in the CYBECO Toolbox. The core of the document describes in an accessible manner the above developments. We then include technical appendices referring to: i) the cyber insurance models developed; ii) the cybersecurity resource allocation model, including a template case study; iii) the general cybersecurity preference model; iv) the algorithmic approach with general defend-attack interactions;

v) the augmented probability simulation approach for large problems; vi) an outline of how the above may be implemented as software. The framework will be revised, enhanced and completed based on the experiments, the policy issues identified and the case studies proposed during the second year of the project.

Dissemination Level		
PU	Public	x
PP	Restricted to other programme participants (including the Commission Services)	
RE	Restricted to a group specified by the consortium (including the Commission Services)	
CO	Confidential, only for members of the consortium (including the Commission Services)	

Document Status

Document Title	Modelling framework for cyber risk management
Version	2.0
Work Package	3
Deliverable #	3.1
Prepared by	D. Rios Insua, A. Couce Vieira (CSIC)
Contributors	Wolter Pieters, Kate Labunets (TUDELFT), Jorge G. Ortega, Alberto Torres, Roi Naveiro (CSIC), Kreshnik Musaraj (AXA), Pam Briggs (UNN)
Checked by	TUDELFT
Approved by	
Date	
Confidentiality	PU

Document Change Log

Each change or set of changes made to this document will result in an increment to the version number of the document. This change log records the process and identifies for each version number of the document the modification(s) which caused the version number to be incremented.

Change Log	Version	Date
First draft	1.0	April 9, 2018
Second draft after peer-review	2.0	April 23, 2018

Table of Contents

List of Figures

1 Introduction

Cybersecurity. A defining feature of our society is its almost pervasive digitalization. Given that, all kinds of organizations, from corporations to governments to SMEs, may be critically impacted by cyber-attacks (Andress and Winterfeld, 2013). Indeed, the economic impact of cyber attacks is outstanding and, consequently, cybersecurity has become an issue of major importance, both technically and financially. Furthermore, attacks, espionage, insiders and breaches appear to increase in frequency, impact and sophistication (Lloyd's, 2017). For instance, the industry estimates that attacks costed as much as $ 450 billion globally in 2016, causing an impact over the global GDP (0.8% in 2014) of a similar magnitude to drug trade (0.9%) or international crime (1.2%) (McAfee, 2017).

Cybersecurity is emerging as one of the major global concerns (WEF, 2017). Although some experts criticize an excessive hype about the potential disruptive capability of large-scale cyber-attacks, cybersecurity is a truly relevant problem, due to the persistence, frequency and variety of threats. Such diversity may be classified according to their motivation, skill and constraints (Dantu et al., 2007), and their ability to exploit or create vulnerabilities on the targeted systems (DSB, 2013). Important cyber threat sources include the military units maintained by global powers; 'hacktivists'; insiders; and, profit-oriented cyber-criminal groups. When it comes to malware, they are usually developed with a goal-oriented behavior (Li et al., 2009) and, consequently, a sound approach is to treat them as adversarial actors and counter-attack them with behavioural approaches (Li et al., 2009). Furthermore, the concept of Advanced Persistent Threat has arisen as patiently orchestrated operations seeking to stay hide while they consolidate their path for executing their final objective.

Relevant cases (Command Five, 2011) include the 2007 Aurora attacks against Google to obtain confidential data about their algorithms and Chinese dissidents; the 2012 Shamoon attack that disabled 30.000 computers of Saudi Aramco (Brenner, 2013); and the 2013 credit card breach of 40 million customers of the US retailer Target (DeNardis, 2015); to name but a few. Attacks with physical consequences are also emerging, including the 2010 Stuxnet attack against an Iranian nuclear plant that disabled a fifth of its nuclear centrifuges (Brenner, 2013) or the attack on a German steelworks in 2014 that stopped their process (Lee et al., 2014). Another notorious trend over the last years have been the indiscriminate ransomware attacks such as the 2017 Wannacry case (Yaqoob et al., 2017) that affected thousands of large and small organizations across the globe for several hours.

Cyber risk analysis. Risk analysis emerges as a fundamental tool to help manage these problems (Cooke and Bedford, 2001). With it, organizations can assess the risks affecting their assets and what safeguards should be implemented to reduce the likelihood of such threats or their impacts, in case they are produced. Numerous frameworks have been developed to screen cybersecurity risks and support risk management resource allocation, including CRAMM (CCTA, 2003), ISO 27005 (ISO, 2011), SP 800-30 (NIST, 2012), or CORAS (Stolen, 2001). Similarly, diverse compliance and control assessment frameworks, like ISO 27001 (2013), Common Criteria (2012), or CCM (CSA, 2016) provide guidance on the implementation of cybersecurity best practices. These standards and frameworks cover detailed security measures suggested for protecting an organization's assets against the risks to which they are exposed. Although these proposals have virtues, particularly their extensive catalogues of threats and assets, much remains left to be done regarding risk analysis from a methodological point of view.

As an example, we sketch some ideas about MAGERIT (Min. Hacienda, 2012). This methodology provides a very detailed catalogue of assets, threats and impacts that save plenty of time to risk managers. One of its more relevant weaknesses is the use of qualitative methods and risk matrices for the analysis of risks. For instance, the treatment of the occurrence of threats is weak and based on a qualitative approach, as shown in Table 1a, where qualitative likelihoods are associated with arbitrary numerical values. As an example, the probability level high (labelled as H) is associated with the value 10 and with an incident that is considered frequent, which is assimilated to happening monthly. Impacts associated with threats are treated in a similarly ambiguous fashion, as shown in Table 1b. Again, qualitative values are used. For example, if the value of the asset is medium (M) and the degradation caused by the incident is around 10%, then it is considered that the impact is medium (M). Risks present a similar issue, as shown in Table 1c. E.g., if the value of the impact is high (H) and the probability of the threat is low (L), then it is considered that the risk is high (H). Ultimately, MAGERIT sheds ambiguous results by using risk matrices.

VH	100	Very frequent	Daily
H	10	Frequent	Daily
M	1	Normal	Monthly
L	1/10	Infrequent	Every few years
VL	1/100	Very infrequent	Every century

(a) Probabilities

Impact		**Degradation**		
		1%	**10%**	**100%**
Value	**VH**	M	H	VH
	H	L	M	H
	M	VL	L	M
	L	VL	VL	L
	VL	VL	VL	VL

(b) Impacts

Risk		**Probability**				
		VL	**L**	**M**	**H**	**VH**
Impact	**VH**	H	VH	VH	VH	VH
	H	M	H	H	VH	VH
	M	L	M	M	H	H
	L	VL	L	L	M	M
	VL	VL	VL	VL	L	L

(c) Risk analysis

Figure 1: Qualitative vision of threats in MAGERIT

As indicated in Cox (2008), these methods suffer from numerous shortcomings including a poor resolution to compare threats, the introduction of errors while assigning qualitative values; or, more importantly for our problem area, the induction of potentially suboptimal resource allocations. It is also important to stress the absence of the threat behaviour in the elicitation, which is a major component in the occurrence of targeted cyber threats. Indeed, the typical approach of security and criminal studies is evaluating three general aspects of a criminal: his capability to commit a crime, his motivation to do so, and the opportunities he can leverage to execute it. These three aspects should be taken into account in cybersecurity too. However, with counted exceptions like IS1 (NTAIA, 2012), standards do not explicitly take into account the intentionality of some of the cyber threats, a key factor to forecast what threats would target the system and what their strategic behaviour would be. Thus, ICT owners may obtain unsatisfactory results about the proper prioritization of risks and the measures that should be implemented. Another critical issue, unlike other risky domains, is that it is difficult to obtain and analyse data, since organizations are reluctant to disclose information about intrusion attempts or consequences of attacks (Balchanos, 2012), for reputational reasons.

Cyberinsurance. Regarding this, it is important to highlight how in recent years new cyber insurance products have been introduced, of very different nature and not in every country, by companies like AXA, Generali, Allianz, or Zurich. However, cyber insurance has yet to take off (Marotta et al., 2017; Low, 2017), in spite that organizations are increasingly aware of their dependence on new technologies and on how information is a critical asset that must be secured so as to not incur in loss of customers, reputational damage and sanctions by regulators. Obstacles for researching and developing cyber insurance (Marotta et al., 2017) include information asymmetry between agents that undermines trust, lack of data due to sensitivity concerns, and the difficulty of specifying rates of occurrence or damages.

Deliverable objectives. Within this context, in this deliverable we start by sketching three decision problems in cybersecurity economics around the concept of cyber insurance. The first one outlines a more rigorous framework for risk analysis in cybersecurity. It serves an organization to decide its best resource allocation strategy in terms of cybersecurity controls and cyber insurance. It also helps an insurance company to design their cyber products based on parametric variations. The second model serves an insurance company to decide their reinsurance portfolio. Finally, the third one supports also an insurance company in deciding whether to grant, or not, a given insurance product to a company. We describe all three models in terms of influence diagrams (ID) and bi-agent influence diagrams (BAID), see Ortega et al. (2017).

We emphasize the first model, which introduces an integrated cybersecurity risk analysis approach to facilitate decision-making regarding ICT systems security. Our goal is to improve current risk analysis frameworks, introducing dynamic schemes that incorporate all relevant parameters, including decision-makers' preferences and risk attitudes (Clemen and Reilly, 2013) and the intentionality of adversaries. Moreover, we introduce decisions concerning cyber insurance adoption to complement other risk management decisions. through risk transfer. We also introduce a template case study to facilitate implementation; however, it may entail a considerable work to extend it to large organizations. To facilitate and standardise its implementation, and with the aim of producing a tool to support such process, we propose a general cybersecurity preference model for an organization including the provision of generic trees of objectives for IT owners in a cybersecurity context; based on it, the provision of a generic multi-attribute utility function to assess the previous cybersecurity objectives, as well as the risk attitudes of the IT owners undertaking the risk management exercise; and the provision of forecasting models for such objectives.

In order to facilitate computations, we also introduce two computational enhancements. The first one refers to computing the Adversarial Risk Analysis (ARA) solutions with general interaction between a cyber-defender and a cyber-attacker. Our first model introduced the cyber security sequential defend-attack and sequential defend-attack-defend cases, but these could be more general. In his landmark paper, Shachter (1986) proposed extending the computation of optimal decision policies in influence diagrams to the multi-agent case as a fundamental problem. We thus consider general adversarial problems between two agents, allowing for complex interactions consisting of intermingled sequential and simultaneous movements, spanning across the corresponding planning period. Our aim is to support the defender in her decision making, for which we need to forecast the attacker's intentions. We assume that the attacker is an expected utility maximizer and we can predict his actions by finding his maximum expected utility policy. The uncertainty in our assessments about the attacker's probabilities and utilities propagates to his random optimal decision which provides the required attack forecast. We provide an ARA framework to solve general bi-agent adversarial problems using BAIDs ability to model complex interactions, taking advantage of the concept of strategic relevance (Koller and Milch, 2003), yet relaxing the common knowledge assumptions through the ARA methodology.

The second computational enhancement is in the realm of algorithmic game theory, see Nisan et al (2007), in that we aim at providing algorithms to approximate solutions to game theoretic problems, both in the standard and ARA approaches. We explore how augmented probability simulation (APS) may be used to compute game theoretic solutions. APS is a powerful simulation based methodology used to approximate optimal solutions in decision analytic problems, see Bielza et al (1999). We start by defining an augmented distribution proportional to the product of the utility and the original distribution and, then, come out with a method to simulate from the augmented distribution. The mode of the marginal in the decision of the augmented distribution coincides with the optimal decision. Note that Monte Carlo (MC) simulation at large, and APS in particular, is not frequently mentioned in the computational game theory literature, see Nisan et al (2007). Moreover, most of the emphasis in the ARA literature has been on foundational issues with little emphasis on computational challenges in complex problems. Therefore, we provide a complete outline of the role of APS for game theoretic computations.

Deliverable structure. The rest of D3.1 is structured as follows. Section 2 sketches three relevant models referring, respectively, to the reinsurance needs of an insurance company introducing cyber insurance products; the decision of an insurance company to grant or not a cyber insurance product to a potential client; the decision of a company about its security resource allocation, including an eventual cyber insurance product. This last one is the main and core model on which we have focused both methodologically and applied wise, including a simplified case that may be used as a template for more complex cases, to which we devote Section 3. To facilitate its application, we revise a generic cybersecurity preference model in Section 4, our computational enhancements in Section 5 and, finally, a sketch of how the framework can be implemented. We end up with a discussion and an outline on how the work will be continued and expanded in the second deliverable. The core of this document aims at describing in an accessible manner our developments, with more technical details in six appendices.

2 Cybersecurity Risk Analysis Framework

Risk management is a fundamental part of cybersecurity at any organisation. And a fundamental part of risk management is **risk analysis**[1] – the process of identifying, understanding and evaluating risks. The aim of risk analysis is to know what can happen, what are the possible consequences, how likely are they and whether these risks are adequately controlled or further action is required. From standards like ISO 31000, we can summarise the steps of risk analysis as follows:

1. Define the **scope** in terms of which systems to include in the risk analysis.
2. Identify the **assets** at risk, including the potential **impacts** and the values at risk. E.g., a customer database or an online store.
3. Identify the **threat agents**, i.e., elements, people and organisations that pose a threat to the system or the assets, including their motivation, capabilities and opportunities to cause harm. It is also important to identify the **threat actions**, i.e., malicious actions these threats can do and the system vulnerabilities that convert a threat into a risk to the system. E.g., an insider able to leak documents or a cyber criminal able to disrupt the ICT systems.
4. Combine the previous elements into a set of specific **risks**, in terms of a particular threat that could cause a particular impact. E.g., the risk of an insider leaking contracts or the risk of a criminal able to disrupt an online store.
5. Determine the **likelihood** and the **impact** of each risk so as to **rate** them. As mentioned, qualitative methods rate risk through a risk matrices. E.g., a risk with a low likelihood and high impact is rated as a high risk. Quantitative methods estimate a specific figure, range or distribution for the likelihood and the impact and, derived from these, for the risk. For a very basic example, a risk with a probability of 1% and an expected impact – if materialised – of € 100,000 produces an estimated risk of € 1,000.
6. Identify **risk treatment** options. This includes security controls that prevent, mitigate, detect or compensate such threats or their impacts in case they are produced. Another risk treatment option, key in CYBECO, is cyber insurance.
7. Determine the likelihood and impact of the risks, now considering the risk treatment options considered relevant.
8. **Evaluate the risks** and treatment options in terms of the stakeholder preferences and risk attitudes.

[1] Other authors and standards use the term *risk assessment.*

In the rest of the section, we describe three models. The procedure for building our models involve the identification and assessment of the previous elements, although some steps are done in a different order (e.g., we identify risk treatment options before calculating the likelihood and the impact of the risks). The first model represents the risk analysis model for an organisation. The other ones are modelled from the insurance company perspective: one for the reinsurance needs in the context of cyber insurance and, the other, for the decision of granting a cyber insurance product to a customer.

2.1 Cybersecurity risk analysis framework for an organisation

In this subsection, we present our **cybersecurity risk analysis framework** for an organisation. The framework pivots over a **cybersecurity resource allocation model** that represents the risk analysis problem mathematically.

The framework consists of the following steps:

1. Definition of the risk analysis scope.
2. Identification of risk components.
3. Problem structuring using our cybersecurity resource allocation model.
4. Problem solving.

In the rest of the subsection, we introduce the framework and the model. We provide further technical details in Section 2 of the paper ‘An Adversarial Risk Analysis Framework for Cybersecurity’ (Annex 1) and Section 2 of the paper ‘Some Decision Problems in Cyber Insurance Economics’ (Annex 2). We also provide examples based on the case study, further detailed in Section 3 of the paper ‘An Adversarial Risk Analysis Framework for Cybersecurity’ (Annex 1).

2.1.1 Definition of the risk analysis scope

As in other risk analysis frameworks, the first step of our framework is to define the scope of the risk analysis in terms of which systems to include in the risk analysis. It is also important to define the temporal scope of the undertaken process.

Example 1 – Risk analysis scope

An SME dedicated to document management and its online document management service, through which the SME provides its customer services. The temporal scope of the risk analysis is one year.

2.1.2 Identification of risk components

The next step is to identify the components of the risk analysis problem, given the scope previously defined. All of these elements could be identified with the support of catalogues like those of the methodologies mentioned in the introduction. We note, though, several differences of our approach as compared with the typical steps of risk analysis, described at the beginning of Section 2: we focus on identifying all components of the risk problem before analysing the interactions between them; we identify and model other relevant uncertainties affecting the threats or the assets, e.g. the fire duration uncertainty related with a fire threat; and we emphasise separating targeted threats from non-targeted ones, as we model them differently, since we consider strategic factors when analysing targeted threats.

Specifically, identification involves the following actions:

1. **Identify the organisation's assets at risk.**
2. **Identify non-targeted threats**, .i.e., those threats over the identified assets deemed relevant and having non-targeted character.
3. **Identify targeted threats**, i.e., those threats over the identified assets deemed relevant and having targeted character.
4. **Identify other uncertainties affecting risk that are only relevant to the organisation**. I.e., affecting the materialisation of non-targeted threats or
5. **Identify other uncertainties affecting risk that might be relevant to the targeted threats.**
6. **Identify security controls** to counter the threats and protect the assets.
7. **Identify cyber insurance products** to consider the possibility of transferring the risk.
8. **Identify impacts over the organisation's interests** considering the identified threats, assets and risk treatment options.
9. **Identify impacts over the interests of the targeted threats** as they might affect their actions against the organisation.
10. **Identify the preferences and risk attitudes of the organisation.**
11. **Identify the preferences and risk attitudes of the targeted threats** as they might affect their actions against the organisation.

Example 2 – Identification of the risk components

For a case like the scoped in example 1, we have that the risk components are the following:

Assets: Facilities, computer equipment and market share.

Non-targeted threats: Fire and computer virus.

Targeted threats: A DDoS[2] attack from a competitor.

Uncertainties of the organisation: Duration of the DDoS, Duration of fire.

Attacker uncertainties: Detection of attacker.

Security controls: Anti-fire system, firewall, risk mitigation procedures and a cloud-based DDoS protection system.

Insurance: Traditional insurance, cyber insurance and a comprehensive insurance including the previous two.

Impacts for the organisation: Impact over facilities, impact over computers and impact over market share. We also consider the costs of security controls and insurance, as well as the insurance coverage, should an incident happen.

Preferences: preferences of the organisation and preferences of the competitor.

2.1.3 Problem structuring

To facilitate understanding and communication of the problem, as well its assessment and evaluation, we structure the problem through influence diagrams (ID). In IDs, square nodes refer to decisions; oval nodes to uncertainties, modelled as random variables; double-oval nodes to deterministic variables; and hexagonal nodes to evaluations, modelled as utilities. An arrow directed to a decision node indicates that this decision is informed about the outcome of the parent node. An arrow directed to an uncertainty node indicates that this uncertainty is conditioned by the parent node outcome. An arrow to a utility node indicates that the outcome of the parent node is evaluated by the agent. Additionally, we can build IDs for multiple agents. We refer to them as multi-agent influence diagrams (MAID) or BAID in case of bi-agent diagrams. We use different colours when we refer to nodes owned by different agents and mixed colours when referring to nodes shared by several of the involved agents.

Figure 2 represents the basic model for cybersecurity resource allocation. There are two agents involved: the Defender (*D*, the organisation), who needs to decide its security resource allocation strategy, and the Attacker (*A*, a targeted threat), who aims at attacking the organisation to obtain some benefit.

[2] 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.

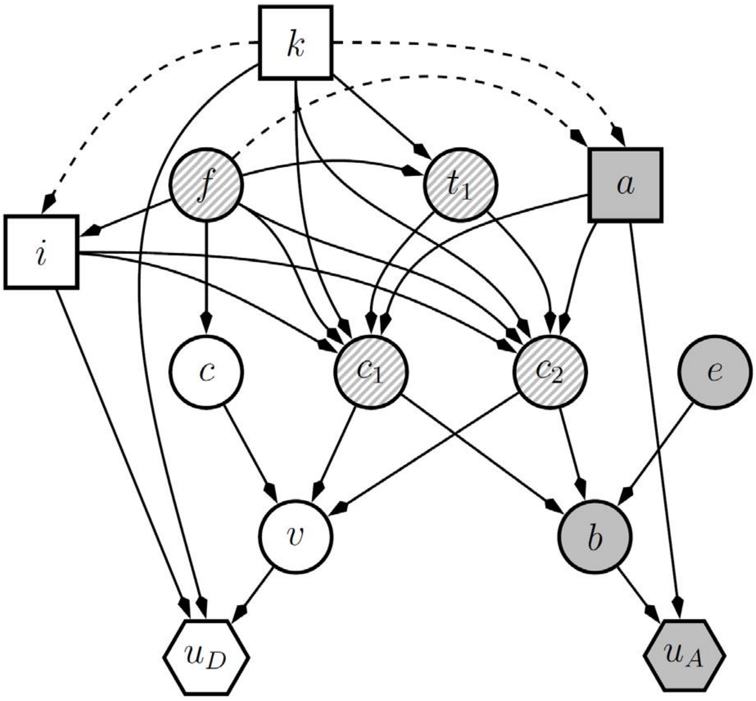


Figure 2: Cybersecurity resource allocation model. White nodes refer to the defender, grey nodes to the attacker and striped nodes are shared by both agents. Square nodes represent decisions, rounded nodes are uncertainties and hexagons represent the evaluation of the attacker (u_D) and the defender (u_A).

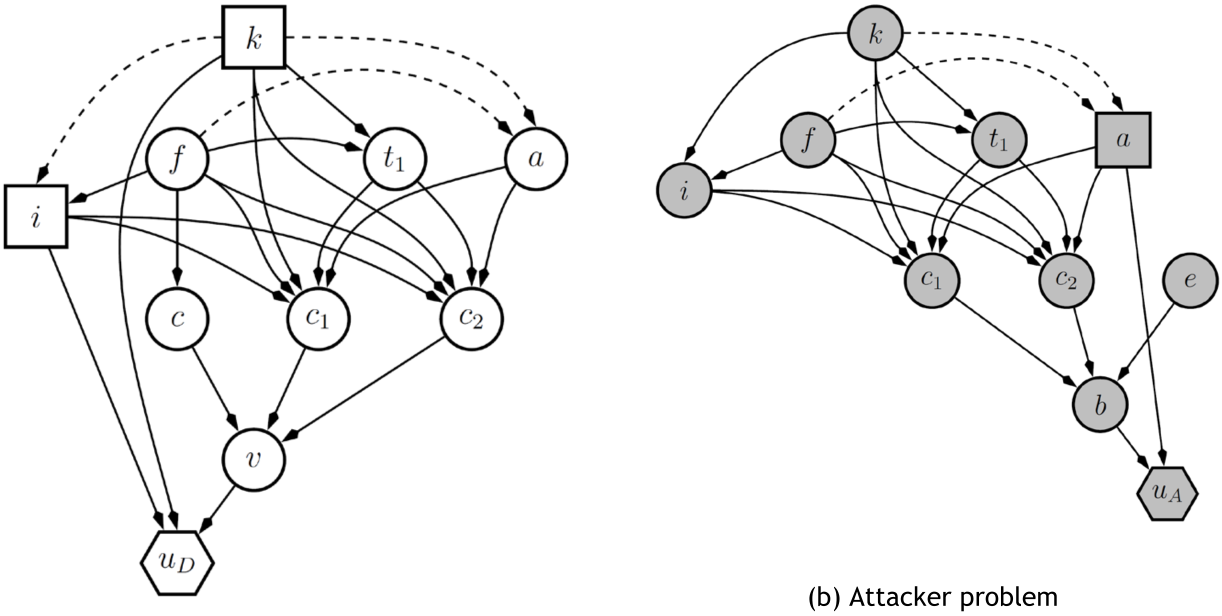
The organisation is characterised by certain security features *f* that might affect risk such as its security posture or its vulnerabilities. The organisation also has a performance measure *c*, associated to the normal operations of the system under study during the relevant planning period. This performance measure could be affected by the organisation security features. The system is exposed to a set of threats. Some of the threats are non-targeted(e.g., fire, energy blackout) or considered non-targeted(e.g., most computer viruses). We model this non-targeted threats as random variables as in t_1. Other threats are targeted (e.g., a DDoS attack or a bomb) and we model them as an agent decision as in *a*. The presence of these threats is affected by the security features *f* of the organisation, but also by the security control portfolio *k* implemented by the organisation (e.g., fire detectors, firewalls).

If the presence of these threats materialise to an actual incident, then the organisation could suffer a series of impacts over certain assets, represented in the model as c_1 and c_2. These impacts are also affected by the security controls implemented (they could mitigate the impacts) and the security features. Additionally, the adoption of cyber insurance, represented as *i*, could reduce the final economic impact of the incident through coverage. Note that we could include the cyber insurance within the portfolio of security controls. However, it is recommendable to separate them, since premiums will typically depend on the security control deployed.

We integrate the results under normal circumstances and the risk impacts with an evaluation function *v*, usually estimated in monetary terms. The utility u_D caters for the organisation preferences and risk attitudes towards the evaluation of the risk scenario and the costs of the security controls and cyber insurance.

However, the targeted threat is also deciding whether to attack or not, represented as *a* in our model. Threats (their potential actors) are capable of observing, to a certain degree, what defences and features are in place before their decision. Depending on the impacts caused to the organisation and other uncertainties particular to the attacker, represented as *e*, the attacker obtains the final result of his attack *b*. The utility u_A caters for the attacker preferences and risk attitudes, which depends also on the cost of his actions.

To facilitate problem understanding and strategic thinking towards problem solution, as we illustrate in the next subsection, the bi-agent problem represented in Figure 2 can be split in single-agent problems. Figure 3(a) represents the defender problem, in which the non-targeted threat *a* is now considered a random variable. Figure 3(b) represents the attacker problem, in which the organisation decisions *k* and *i* are now considered a random variable.



(a) Defender problem

(b) Attacker problem

Figure 3: Defender and attacker problem

Example 3 – problem structuring

For the risk components identified in Example 2, we have the problem structure represented in Figure 4.

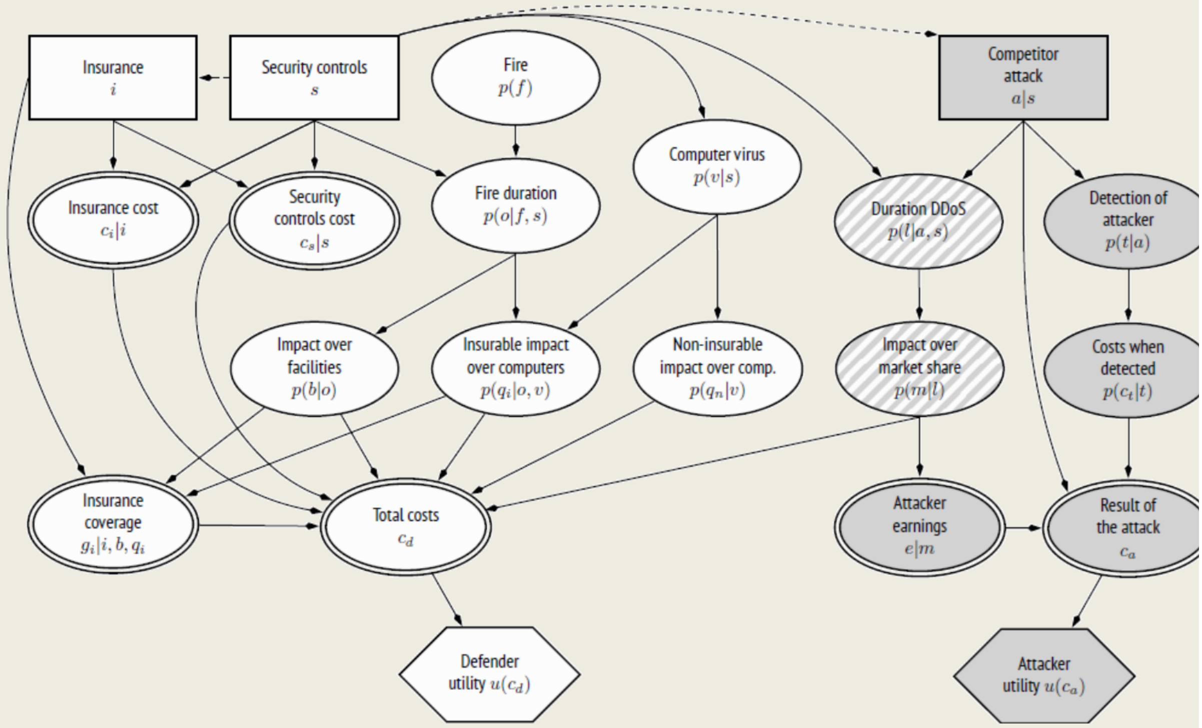

Insurance i
Security controls s
Fire $p(f)$
Competitor attack $a|s$
Computer virus $p(v|s)$
Insurance cost $c_i|i$
Security controls cost $c_s|s$
Fire duration $p(o|f,s)$
Duration DDoS $p(l|a,s)$
Detection of attacker $p(t|a)$
Impact over facilities $p(b|o)$
Insurable impact over computers $p(q_i|o,v)$
Non-insurable impact over comp. $p(q_n|v)$
Impact over market share $p(m|l)$
Costs when detected $p(c_t|t)$
Insurance coverage $g_i|i,b,q_i$
Total costs c_d
Attacker earnings $e|m$
Result of the attack c_a
Defender utility $u(c_d)$
Attacker utility $u(c_a)$

Figure 4: Problem structuring example

The defender (SME) faces three threats. The first one is a *fire*, a non-cyber and non-targeted threat. The second one is a *computer virus*, a cyber and non-targeted threat. The non-targeted threats are modelled as random variables through uncertainty nodes. The third threat is a *competitor DDoS attack*, cyber and intentional, aimed at benefiting from disrupting the market share of the SME. This targeted threat is modelled through an attacker decision node. Some of these threats are affected by the *security controls* implemented by the SME (decision node). These might reduce the number of virus infections and might be observed by the competitor before making the attack decision.

There are some uncertainties related to the threats that should be taken into account. When it comes to the fire, the *fire duration*, which is affected by the security controls (an anti-fire system detects a fire and, thus, reduces its duration). When it comes to the competitor attack, the *duration of the DDoS*, which is also affected by the security controls (the cloud-based DDoS protection). The attacker also has particular uncertainties. Namely, whether he is *detected* or not.

Most of these decisions and uncertainties involve costs modelled in their corresponding nodes. *Insurance costs* and *security controls costs* are deterministic variables as we now their price. The potential impacts of the threats, however, are uncertain and consequently modelled as random variables. We refer to the following:

- *Impacts over facilities*, should a fire happen.

- *Insurable impacts over computers* (caused by equipment destruction and repair), should a fire or a computer virus infection happen.

- *Non-insurable impacts over computers* (caused by productivity loss), should a computer virus infection happen.

- *Impact over market share* (caused by online store unavailability), should a successful DDoS in the online store happen.

- Additionally, the attacker also faces *costs when detected*.

The defender integrates all the impacts and costs affecting them into the *total costs* node, which also includes the *insurance coverage* (which depends on the insurance product acquired and the impacts that are insurable). The *defender utility* node evaluates the total costs against the SME preferences and risk attitudes.

On the attacker side, he integrates in the *result of the attack node* his *earnings* (derived from the market share gained from the SME loss of market share) with his costs (implementation of the attack and costs when detected). The *attacker utility* node evaluates the result of the attack against the competitor preferences and risk attitudes.

From the problem structure of Example 2, we have the defender problem represented in Figure 5 and the attacker problem represented in Figure 6.

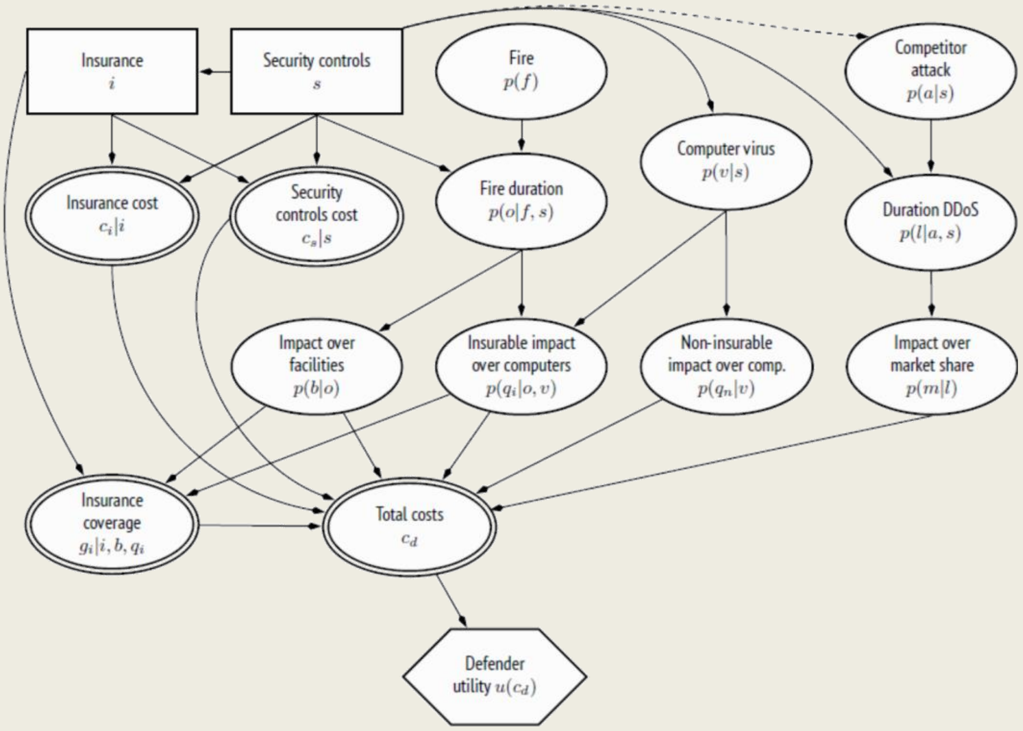


Figure 5: Defender problem example

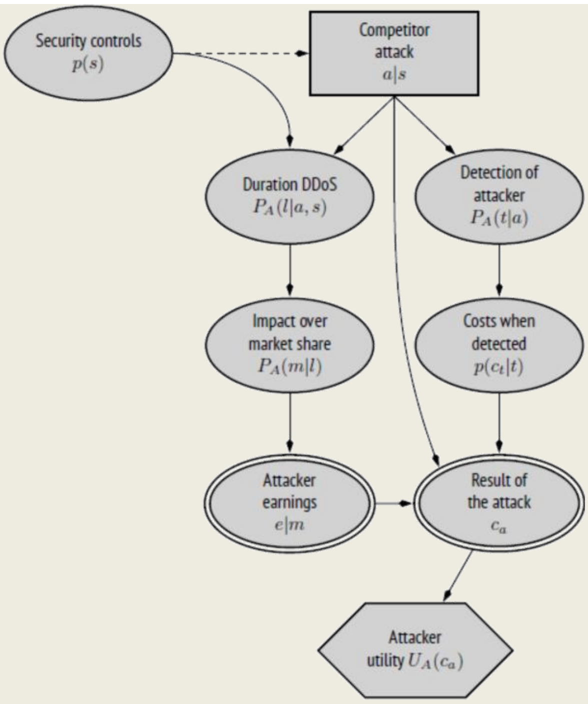


Figure 6: Attacker problem example

2.1.4 Problem solving

The organisation aims at finding the portfolio of security controls and insurance that maximises their expected utility, i.e., the best portfolio. On the other side, the attacker also aims at finding the action that maximises his expected utility.

Solution to the defender problem

To solve the defender problem represented in Figure 3(a), we need to assess:

- The probabilities of the threats happening, given the portfolio of security controls implemented as well as the organisation's features, which are $p(t_1|k,f)$ and $p(a|k,f)$.
- The impacts of the threats, should they happen, given the portfolio of security controls implemented and the insurance product adopted, which are $p(c_1|t_1,a,i,k,f)$ and $p(c_2|t_1,a,i,k,f)$.
- The performance of the system, which is $p(c|f)$.
- The evaluation function that integrates the normal conditions and the potential impacts of the risk, which is $v(c,c_1,c_2)$.
- The final evaluation of the utility, which includes $v()$ but also the costs of the security controls and insurance. The utility for the organisation is $u_D(v(c,c_1,c_2),c_i,c_k)$.

Once we have assessed these quantities, we would need to find the portfolio of security controls and insurance that maximises expected utility. Specifically, when portfolio *k* is implemented together with insurance *i*, the expected utility is

$$\psi(k,i) = \int \dots \int u_D(v(c,c_1,c_2),c_i,c_k)\; p(t_1|k,f)p(a|k,f)p(c_1|t_1,a,i,k,f) \times$$
$$\times\, p(c_2|t_1,a,i,k,f)p(c|f)\;\, dt_1\; da\; dc_1\; dc_2\; dc.$$

We seek, then, the maximum expected utility portfolio-insurance pair under the relevant restrictions, that is,

$$\max_{k\in K, i\in I} \psi(k,i),$$

where *K* represents the constraints over the security control portfolio and *I* the insurance catalogue. The pair *(k,i)* could be further restricted jointly, e.g., by a common budget constraint or certain legal or technical requirements.

In principle, all of the above elements may be modelled through statistical methods and expert judgement. However, the beliefs concerning the attacker decision, $p(a|k,f)$, entails an strategic element: as we explained earlier, the attacker intentionality is not taken into account in most existing methodologies. Specifically, it describes the probability that the organisation gives to receiving the attack *a* from the attacker, had the portfolio *k* been adopted when the features are *f*. We assess this aspect through ARA, considering the attacker problem represented in Figure 3(b) to build a statistical distribution of the optimal random attack for $p(a|k,f)$. Solving the attacker problem this way allows us to obtain a probability for the attack, taking into account the strategic motivations of the attacker.

Solution to the attacker problem

We have that, for a given portfolio *k* and features *f*, the optimal random attack is

$$\mathrm{A}^*(k,f) = \arg\max_{a\in A} \Psi_A(a|k,f) \int \dots \int U_A(a,b)\; P_A(t_1|k,f)P_A(c_1|t_1,a,i,k,f) \times$$
$$\times\, P_A(c_2|t_1,a,i,k,f)P_A(b|c_1,c_2,e)P_A(e)\;\, dt_1\; dc_1\; dc_2\; db\; de.$$

where U_A and P_A are, respectively, the random utility function and random probability distributions that model the organisation beliefs about the actual utility function u_A and probability distributions p_A of the attacker.

Based on that, we have that the distribution over the attacks we were looking for is

$$p(a|k,f) = P(A^*(k,f) = a)$$

Example 4 - Solutions of the defender and attacker problems

As we mentioned, we need to find the solution to the attacker problem before solving the defender problem. From the attacker representation depicted in Figure 6, we have that the random optimal attack *A* when the security portfolio *s* is implemented is

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

Based on that, we have that the distribution over the attacks we are looking is

$$p(a|s) = P(A^*(s) = a).$$

From the defender problem representation depicted in Figure 5, we have that the expected utility when the security portfolio *s* is implemented together with insurance *i* is

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

The optimal decision is the security control portfolio and insurance pair

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

Problem solving procedure

Having discussed how to solve analytically the problem for the organisation (defender problem) and non-targeted threats (attacker problem), we now highlight the practical procedure to solve the problem:

1. **Assess the organisation's non-strategic beliefs and preferences**. This step involves modelling quantitatively all the nodes of Figure 3(a) except *a*, through statistical models and with the support of data and expert judgement.

2. **Assess the random beliefs and preferences of the targeted threats**. We model and simulate the attacker problem of Figure 3(b) to forecast his actions and obtain the probability distribution that we will use to model *a* in the defender problem.

3. **Optimise the organisation's problem** now that all the strategic and non-strategic beliefs and preferences have been modelled. This involves the construction of algorithms (see Section 4) and its software implementation (Section 5).

Example 5 - Assessing a non-strategic belief of the organisation

A non-strategic belief of the organisation is the duration of the DDoS, represented through the uncertainty node *Duration DDoS* in Figure 4 and Figure 5. The basic description of this node is $p(l|a,s)$, meaning that the DDoS duration *l* is affected by the competitor decision to attack *a* and the security control *s* implemented by the organisation. We briefly describe how we model the duration *l* in hours of a DDoS campaign. Its length will depend on the intensity of such campaign (consisting of *a* DDoS attempts), how well-crafted is the attack and the security controls implemented by the SME. In our case, we consider an emerging alternative which are cloud-based systems absorbing traffic from customer sites when they become victims of a DDoS. Otherwise, if no control is deployed, it would be virtually impossible to block such attack. Based on experts and literature, we have that the average attack lasts 4 hours, averaging 1 gbps, with peaks of 10 gbps. We model *lj*, the length of the j-th individual DDoS attack with a distribution $\Gamma(4,1)$, 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 platform capable of 5 gbps attacks, modelled through a distribution $\Gamma(5,1)$. 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. 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 $\Gamma(5,1) - s_{gbps} > 0$, and $l_j = 0$ otherwise.

Solution (and other relevant results)

Implementing the previous procedure, we are able to calculate the **best security control and insurance portfolio**, and order all the portfolios from most to least preferred. Additionally, the model provides other relevant information from the risk analysis perspective:

- Overall probability of different events.
- Expected impacts given the different probabilities.

Further analysis are possible, we can use this model to perform other relevant assessments:

- Sensitivity analysis to evaluate the robustness of the solution against variations in the probabilities and parameters of the model.
- Introduction on constraints as budget limits or compliance with insurance policies or risk management policies.
- Calculation of the return on security investment to assess the cost effectiveness of a cybersecurity budget.
- Very importantly, parametrically design cyber insurance products.

Example 6 - Solution (and other relevant results)

Solving the defender problem, we find the best portfolio which in our case consists of:

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

- Comprehensive insurance.

Additionally, with the expected utility function we can rank the different combinations of security controls and insurance from the best to the worst, as in the below table.

Anti-fire decision	Firewall decision	Procedure decision	DDoS prot. decision	Insurance decision	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

With this model, we also have the probabilities of different events. For instance, from the probability distribution $p(a|s)$, we have that if the defender implements a 2 gbps cloud-based DDoS protection then the probability of DDoS attempts is as follows:

Number of attempts	0-17	18	19	20	21	22	23
Probability	0.00%	0.20%	0.10%	0.20%	1.30%	1.30%	2.00%

Number of attempts	24	25	26	27	28	29	30
Probability	3.40%	6.90%	9.10%	11.20%	14.40%	22.30%	27.60%

Another relevant information to model are expected values. For instance, the total costs.

Anti-fire decision	Firewall decision	Procedure decision	DDoS prot. decision	Insurance decision	Expected total costs
Anti-fire	Firewall	No procedure	1 tbps	Comprehensive	€ 20,173.93
Anti-fire	Firewall	No procedure	1 tbps	Traditional	€ 22,040.04
No anti-fire	Firewall	No procedure	1 tbps	Comprehensive	€ 23,023.71
…	…	…	…	…	…
No anti-fire	No firewall	No procedure	No protection	No insurance	€ 807,060.17
No anti-fire	Firewall	No procedure	No protection	Cyber	€ 811,410.23
Anti-fire	No firewall	No procedure	No protection	No insurance	€ 813,197.19

2.2 Cybersecurity risk analysis from the insurance company perspective

In this subsection, we briefly sketch two further risk analysis models from the insurance company perspective. The first one serves an insurance company to decide their reinsurance portfolio. The second one supports an insurance company in deciding whether to grant a given insurance product to a company. We provide further technical details in Section 3 and 4 of the paper 'Some Decision Problems in Cyber Insurance Economics' (Annex 2). As with the model in Section 2.1, we describe them in terms of IDs and BAIDs, as required.

2.2.1 Cyber reinsurance

Suppose that an insurance company has segmented the market in several sectors. For instance, standard SMEs, ICT based SMEs and large enterprises. In standard SMEs, ICT is just a support function and they rarely employ dedicated staff; in ICT SMEs, this technology is critical and core so they typically employ dedicated staff, possibly even focusing on cybersecurity; large enterprises maintain an important ICT infrastructure and usually have in-house ICT, security and information departments. Each of them would have their own specific threats, which we, respectively, summarise through t_1, t_2 and t_L. Moreover, there will typically be common threats which we summarise through *d*. This allows us to induce the potential accumulation effects that may hold in this application area. The effects of these threats in the insurance claims of each segment is established through s_1, s_2 and s_L. For assessing them, we would consider the size of each segment and aspects such as ICT systems, cybersecurity and financial resources, features, assets and threats at each segment, much as we did in the first model. Node s aggregates the effects s_1, s_2 and s_L over various segments, but is also compensated by the reinsurance decision *r*. This decision may consist on whether to reinsure and, if so, what portfolio of reinsurance products to acquire from reinsurance companies. This decision could be restricted by, say, financial, legal or compliance requirements.

Then, after building the probability and utility functions of the model, we would aim at maximising

$$\max_r \int \dots \int u\big(s(s_1, s_2, s_L, r)\big)\, p(d) \times \dots \times p(s_L, d, t_L)\ ds_L\, ds_2\, ds_1\, dt_L\, dt_2\, dt_1\, dt,$$

to find the optimal reinsurance decision of the insurance company.

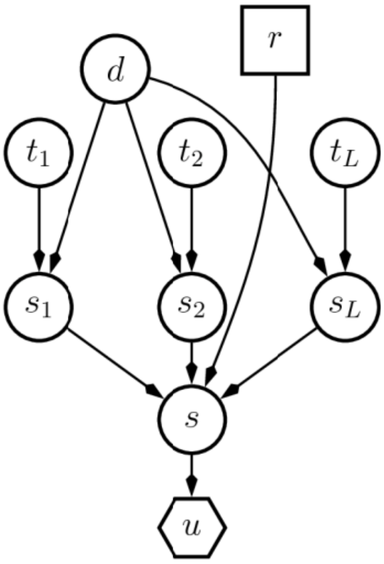


Figure 7: Cyber reinsurance

2.2.2 Granting insurance products

As represented in Figure 8, the insurance company needs to decide whether to grant or not an insurance product *i* to a customer which, in turn, faces threats, summarised in *t*. These threats determine the likelihoods and sizes of claims, as discussed in previous sections. However, the claim likelihood *c* is also affected by costumer decisions regarding cybersecurity compliance and care in terms of insurance liability *j*. This involves behaviours that could reduce cybersecurity effectiveness (e.g., adherence to security policy, security control maintenance, misuse) or, even worse, committing fraud. Should a claim happen, the insurer or a supporting cybersecurity auditor would typically perform a forensic investigation *d* on the claim, aimed at detecting fraud. The claim finally awarded to the insuree by the insurance company would depend on the initial claim and the result of the detection report r_d. Both the insurance company and the insuree would aim at maximising their respective utilities (u_I and u_J).

This is again an ARA problem, structurally resembling that in Section 2.1. Then, the process would go through two stages: the adversarial problem first (costumer), and the insurance company one, second. The decision faced by the insurance company is a standard decision analysis problem with the extra ingredient of having to forecast the customer decisions.

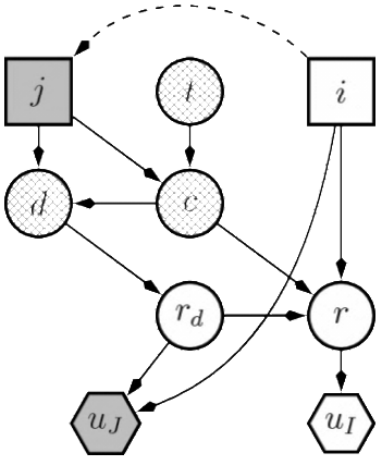


Figure 8: Insurance granting decision

3 Modelling preferences and uncertainty

Relevant aspects in our risk analysis approach is (a) forecasting the potential consequences of the different risk identified in the risk analysis and (b) building the evaluation of them regarding stakeholder preferences and risk attitudes. Specifically, the approach is as follows: First, assess the potential consequences of cybersecurity risks regarding a set of cybersecurity objectives; second, integrate these objectives in a multi-attribute utility function representative of the stakeholders' preferences and risk attitudes. Our approach is inspired by earlier work in counter-terrorism, homeland security and aviation safety management detailed in Keeney (2007), Keeney and von Winterfeldt (2011) and Rios et al (2017). Annex 3 provides technical details of the contents discussed in this section.

To facilitate the identification and assessment of objectives and preferences, we propose the following templates:

- A generic tree of potential cybersecurity objectives for ICT owners. We provide the attribute corresponding to each objective.

- A generic multi-attribute utility function to assess the previous cybersecurity objectives regarding ICT owners' preferences and risk attitudes.

- A forecasting model for such objectives.

3.1 Tree of Cybersecurity objectives

We propose a general cybersecurity preference model for the organisation undertaking a cybersecurity risk analysis. Specifically, this includes:

- The provision of a generic tree of objectives (the performance measures that we want to optimise) for ICT owners in a cybersecurity context. Ideally this could be shown to cybersecurity stakeholders who would pick from it the relevant objectives for their problem at hand. For each objective, we identify the corresponding attribute in which we assess it.

- Based on it, the provision of a generic multi-attribute utility function to assess the previous cybersecurity objectives as well as the risk attitudes of the organisation.

There are several requirements that the objectives in a decision-making problem should meet (Keeney and Gregory, 2005):

- Comprehensive: Objectives cover the whole range of relevant consequences.

- Measurable either objectively or subjectively.

- Relevant.

- Unambiguous, in the sense of having a clear relationship between consequences and their description using the objective.

- Understandable and clearly communicable.

We distinguish between **natural attributes** that provide a direct measure of the objective involved (e.g., repair costs in €), **proxy attributes** that have a relationship to the objective (e.g., website downtime – in an online store – regarding the income generation objective) and used when no natural attributes are suitable, and **constructed scales**.

Some cybersecurity frameworks provide catalogues of concepts analogue to our cybersecurity objectives, mostly those addressing business impact analysis in cybersecurity or business impact analysis in general including ETSI GS ISI 002 v1.2.1 (2015), ISO 22317 (2015b), OWASP business impacts (2017), OECD types of cyber losses (2017) the ENISA Information Package for SMEs (2007), the ENISA report on ICT business continuity management for SMEs (2010), and CYBECO deliverable on the definition of cyber insurance scenarios (CYBECO D4.2, 2018). In general, they depict a few general categories of impacts (legal and regulatory, productivity, financial, reputation and loss of customers) with some examples or subcategories. However they do not meet the requirements for objectives we mentioned earlier. Most of them provide a list of recurrent business impacts rather than a comprehensive list that encompasses less typical impacts (e.g., physical impacts). Similarly, they provide types of objectives that somehow overlap: most of the impacts affect monetary objectives and, thus, some categorisation among them is recommended. For instance, some costs affect specific assets (e.g., asset degradation or activity interruption) whereas others are more general (e.g., competitive advantage, reputation).

Besides the existing lists of cybersecurity impacts, the main conceptual influences on our final list come from asset management and law. First, asset management – e.g., ISO 55000 (ISO, 2014) for assets in general or ISO 19770 for ICT assets (2015a) – it helps to conceptualise the different status that an asset could attain is important, so that engineers could characterise how an asset affects a system or the organisation in terms, for instance, of reliability or predictability. Another conceptual influence comes from law, in particular, the distinction between damages on property (a.k.a. economic or pecuniary damages) and damage on persons (a.k.a. general or non-pecuniary damages). This facilitates the distinction between objectives that can be measured in monetary terms (directly or through estimation) and others that are of non-monetary nature and, thus, need special considerations when it comes to their evaluation (e.g., through the value of statistical life). It also helps on the distinction between the owners of the objectives (i.e., health and environmental damages are suffered always by third parties besides the monetary, legal or reputational consequences that these damages could cause to the organisation).

Based on such catalogues and the described approach, we have developed a generic tree of cybersecurity objectives for a generic organisation, which we summarise in Figure 9. The general categories are the following:

- **Minimize operational costs**. We refer here to the assets and activities that constitute the inventory and operations of an organisation. All of them measurable in monetary terms, i.e. the corresponding attribute would be euros.

- **Minimize income reduction**: We refer here to impacts that reduce the income obtained by the organisation. All of them are measurable in monetary terms. All of them measurable in monetary terms.

- **Minimize strategic costs**: These refer to impacts over intangible assets that apply to the organisation as a whole and in a long term manner. All of them measurable in monetary terms.

- **Minimize financial costs**: Minimize costs caused by the depreciation, abuse, unavailability or elimination of the financial assets of the organisation. All of them measurable in monetary terms.

- **Minimize compliance costs**. All of them measurable in monetary terms.

- **Minimize cybersecurity management costs**. All of them measurable in monetary terms.

- **Minimize institutional reputation impact**: We refer here to impacts over reputation that affect the trustworthiness of the organisation as an institution, rather than those more directly measurable in monetary terms that impact brand value or minimise income/service. These are not measurable in monetary terms and other types of attributes need to be developed.

- **Minimize impact to third parties**: An incident in the organisation might affect third-parties and, thus, the organisation objectives also involve minimising third-party objectives. Therefore, the objectives enumerated above are also applicable to such third parties. Additionally, we need to include additional objectives for non-organisational parties, people and nature. Some of them entail impacts which have been very rare, so far, in cybersecurity, including harm to people or environment. Cyber attacks with physical impact are rare, but the emergence of industrial systems and IoT brings these risks to the fore, e.g. Stuxnet. Examples of third parties are partner organisations, suppliers, competitors, customers and potential customers, staff, shareholders, other parties within range, regulators and government or society. Some are part of the organisation but suffer impacts or have objectives as independent parties, e.g. staff.

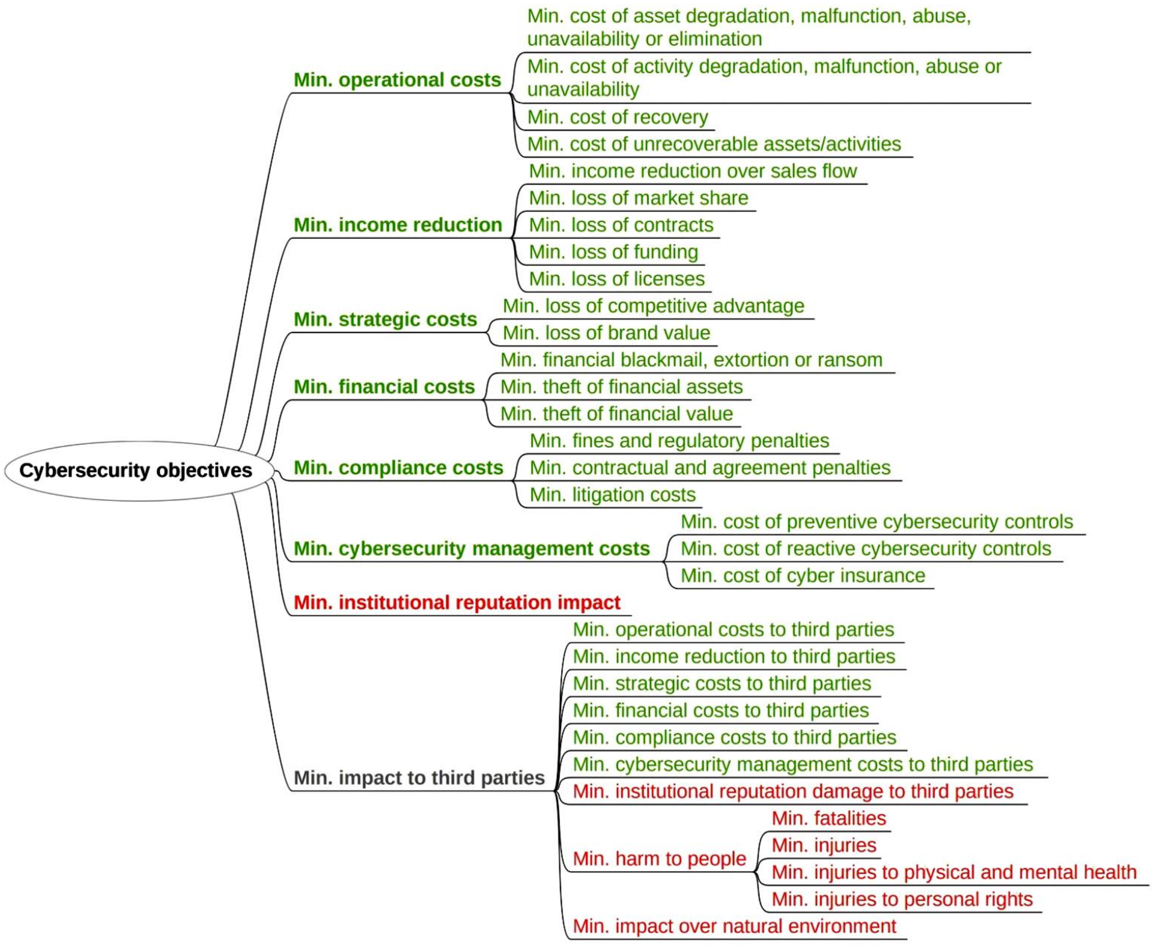

Cybersecurity objectives
Min. operational costs
Min. cost of asset degradation, malfunction, abuse, unavailability or elimination
Min. cost of activity degradation, malfunction, abuse or unavailability
Min. cost of recovery
Min. cost of unrecoverable assets/activities
Min. income reduction
Min. income reduction over sales flow
Min. loss of market share
Min. loss of contracts
Min. loss of funding
Min. loss of licenses
Min. strategic costs
Min. loss of competitive advantage
Min. loss of brand value
Min. financial costs
Min. financial blackmail, extortion or ransom
Min. theft of financial assets
Min. theft of financial value
Min. compliance costs
Min. fines and regulatory penalties
Min. contractual and agreement penalties
Min. litigation costs
Min. cybersecurity management costs
Min. cost of preventive cybersecurity controls
Min. cost of reactive cybersecurity controls
Min. cost of cyber insurance
Min. institutional reputation impact
Min. impact to third parties
Min. operational costs to third parties
Min. income reduction to third parties
Min. strategic costs to third parties
Min. financial costs to third parties
Min. compliance costs to third parties
Min. cybersecurity management costs to third parties
Min. institutional reputation damage to third parties
Min. harm to people
Min. fatalities
Min. injuries
Min. injuries to physical and mental health
Min. injuries to personal rights
Min. impact over natural environment

Figure 9: Cybersecurity objectives. Green (measurable in monetary terms), red (non-monetary).

3.2 Preference modelling

From the previous list of cybersecurity objectives, the incumbent stakeholders could choose the relevant ones. Then, we need a procedure to model stakeholder preferences and risk preferences. For this, we use the classic concepts of measurable multi-attribute value function (Dyer and Sarin, 1979) and relative risk aversion (Dyer and Sarin, 1982).

The basic approach consists of three steps:

Multi-attribute preferences

A practical assumption, feasible regarding our cybersecurity objectives, is *additive independence*, which implies that stakeholders attitude to risk on each of the attributes c_1, c_2, …, c_q does not depend on the other attributes. In this case the utility function is

$$u(c) = \sum_{i=1}^{q} k_i u_i(c_i),$$

where k_i is the **weight** for the attribute i. This weight is generated from the preference elicitation method, in which stakeholders select between a combination of attributes and another. For example, selecting between loosing € 10.000 or having the salaries of all the employees available on the Internet.

Risk attitude

There are three general attitudes to risk. A **risk prone** attitude prefers to take risks in order to have more gains. Mathematically, its utility function is convex. A **risk averse** attitude prefers to avoid risk in order to ensure her gains. Its utility function is concave. A **risk neutral** attitude has no strong preference between avoiding or taking the risk. Its utility function is linear. People might have different attitudes for different things (e.g., risk prone financially and risk averse in health) or even the same thing at different magnitudes (e.g., risk prone when risking less than € 100.000 and risk averse when risking bigger quantities). This last example is connected with the concept of **local risk aversion**, whose value is non-negative with risk aversion and non-positive with risk proneness.

Based on that, a simple but very useful form of utility function arises when the relative risk aversion is set to a constant, in which case we have **constant absolute risk aversion**, which can take the form

$$u(c) = 1 - \exp(-pc)$$

Note that utility functions may exhibit many shapes other than concave, convex or linear. Many empirical studies have suggested that individuals' utility for money passes through regions of convexity and concavity as the sums involved increase, with risk proneness changing to risk aversion. Furthermore, an individual's utility for money and her risk attitude is undoubtedly related to her total assets. Thus, in assessing utility, it is usual to integrate monetary outcomes into the final level of wealth.

Modelling risk attitude with multi-attribute preferences

The next step is to combine (1) the multi-attribute utilities for independent attributes with (2) constant absolute risk aversion utilities. If the attributes are utility-independent and compatible with an additive utility function, then the utility function must have one of the following forms:

$$u(c) = 1 - \exp\left(-\rho \sum u_i(c_i)\right), \quad \rho > 0.$$

$$u(c) = \sum u_i(c_i).$$

$$u(c) = 1 + \exp\left(\rho \sum u_i(c_i)\right), \quad \rho > 0\,;$$

This is the approach that we shall follow in CYBECO to facilitate modeling the Defender’s preferences, as it is not overly demanding cognitively and it is relatively general in its assumptions.

4 Algorithms

Our main model, introduced in Appendix 2 and detailed and illustrated in Appendix 1 with a case study, is based on adversarial risk analysis (ARA), essentially going through a phase of simulation to forecast the attacker actions and then use the corresponding probabilities to find the optimal defences. We described the defend-attack and the defend-attack-defend cases but these interactions between the cyber defender and the cyber attacker could be more general. In order to facilitate computations, we have introduced two computational enhancements.

The first one (Appendix 4) refers to computing the ARA solutions when there are generic interactions between a cyber-defender and a cyber-attacker. We provide an algorithm that facilitates the computations providing the steps required to undertake such task. In his landmark paper, Shachter (1986) proposed extending the computation of optimal decision policies in influence diagrams to the multi-agent case as a fundamental problem. We thus consider general adversarial problems between two agents, allowing for complex interactions consisting of intermingled sequential and simultaneous movements, spanning across the corresponding planning period. Our aim is to support the defender in her decision-making, for which we need to forecast the attacker's intentions. We assume that the attacker is an expected utility maximizer and we can predict his actions by finding his maximum expected utility policy. The uncertainty in our assessments about the attacker's probabilities and utilities propagates to his random optimal decision, which provides the required attack forecast. We provide an ARA framework to solve general bi-agent adversarial problems using BAIDs ability to model complex interactions, taking advantage of the concept of strategic relevance (Koller and Milch, 2003), yet relaxing the common knowledge assumptions through the ARA methodology.

The basic structures that we deal with essentially consist of coupled influence diagrams, one for the defender and one for the attacker, possibly with shared uncertainty nodes and some links between the attacker's and the defender's decision nodes. We designate them BAIDs. In them, as stated above, we observe several decision (square), chance (circle) and utility (hexagon) nodes, corresponding to the defender (white) and the attacker (grey) problems, respectively. Striped nodes represent common chance nodes, in the sense that such uncertainties are relevant in both agents' decision-making. However, they may entertain different probability models over such nodes, which will not be common knowledge. Besides, when an agent's action or a random event is observed prior to a decision, there is a dashed arrow pointing to the decision node of the observing agent. We shall not be able, though, to deal with all BAIDs conceivable as the mere junction of two IDs (one for the defender, one for the attacker) with several shared chance nodes and arrows linking their decision nodes. To ensure consistency between the informational structure and the ordering of the decision makers' analysis, we draw terminology from Shachter (1986), extending it to the bi-agent case. In our setting, a proper BAID will be an acyclic directed graph over decision, chance and utility nodes, where chance nodes can be shared by both agents, such that, from each decision maker's perspective, it is a proper ID. Essentially, we require that if two decisions are simultaneous, there is no directed path between them.

From an algorithmic point of view, given the sequentially of the problem, with alternation of simultaneous and sequential decisions, we shall solve it stepwise, scheduling optimization stages from the defender's problem and simulation stages from the attacker's problem, to get the required distributions in the defender's problem. We could think of tackling first all the simulation stages and then the optimization ones, but switching between both problems allows us to better apportion uncertainty around the defender's optimal decisions, as suggested in Merrick and Parnell (2011). In this manner, we can better assess the defender's distributions over the attacker's probabilities for her decisions, in line with Fermitisation strategies in structured expert judgement methodologies, see Tetlock and Gardner (2015). Then, essentially, the proposed approach solves as many D steps as possible with standard ID reduction operations, until some assessment from the attacker is required to solve another D step. Then, as few steps from problem A are solved, with ID reduction operations modified to take into account the uncertainty about the attacker's utilities and probabilities, until the required attacker's assessment is obtained. At this point we jump back to the defender's problem, and proceed as above until all the defender's decision nodes have been reduced. Deciding when to jump from problem the cyber defender problem to the cyber attacker problem, and backwards, is relatively simple to perform by hand, but can be messy from an algorithmic point of view. We facilitate this through the use of the relevance and component graphs, described in Koller and Milch (2003): using the topological ordering induced by the component graph, as well as the decision sequences for each agent, we may design a systematic approach to computing optimal cyber risk management strategies.

The second computational enhancement (Appendix 5) is in the realm of algorithmic game theory, Nisan et al (2007), in that we aim at providing algorithms to approximate solutions to game theoretic problems, within the ARA approach. The key contribution is to avoid the two step ARA procedure and accelerate computation. As a by-product, we also provide APS approaches to approximate standard Nash equilibria solutions. Therefore, we explore how augmented probability simulation (APS) may be used to compute game theoretic solutions. APS is a powerful simulation based methodology used to approximate optimal solutions in decision analytic problems, see Bielza et al (1999). We start by defining an augmented distribution proportional to the product of the utility and the original distribution and, then, come out with a method to simulate from the augmented distribution. The mode of the marginal in the decision of the augmented distribution coincides with the optimal decision. Note that Monte Carlo (MC) simulation at large, and APS in particular, is not frequently mentioned in the computational game theory literature. Moreover, most of the emphasis in the ARA literature has been on foundational issues with little emphasis on computational challenges in complex problems as we may have to face in cybersecurity risk management. Therefore, we provide a complete outline of the role of APS for game theoretic computations.

APS is based on treating the decision variables as random and converting the optimization problem into a simulation one in the joint space of both decision variables and random variables. Simulating from the augmented distribution of decision and states simultaneously solves for the expectation of the objective function and optimization problem: the marginal mode over the decision variable provides the optimal decision. The strategy is very general in that it can accommodate arbitrary probability models and non-negative utility functions.

We provide two APS approaches to approximating game theoretic solutions for the sequential defend-attack problem, under common knowledge. The first approach involves running one long chain on all decision variables and uncertain variables, whereas the second uses the idea of a nested APS framework similar to folding back a tree. If the game theoretic solution is not robust, we need to address the issue. One way forward is to perform an ARA approach. For this, we weaken the common knowledge assumption. Again, we introduce two approaches based on running a long chain and a nested approach.

5 Software implementation of the risk analysis models

We have implemented the case study in the paper ‘An Adversarial Risk Analysis Framework for Cybersecurity’ (Appendix 1) in the open source software R, as well as a version in which we may vary the involved parameters. Basically, it consists of the following elements:

1. Definition of R functions that model the different assessments over the organisation’s non-strategic beliefs and preferences (defender problem) as well as the random beliefs and preferences for the attacker (attacker problem). We also include the different decisions as variables.

2. Definition of the inputs and outputs of these R functions so that they reflect the conditionality expressed in the influence diagrams. This way, we are able to calculate the probabilities for the different events and values modelled in the R functions.

3. Implementation of the algorithms to calculate the random optimal decision of the attacker (solution of the attacker problem) and obtain the security control and insurance portfolio that maximises expected utility (solution of the defender problem). These require the probability calculations mentioned in item 2. Additionally, an important parameter in these algorithms is the number of simulation iterations.

All of this is coded in several R scripts that generate tables with the relevant risk analysis information of the case. Annex 6 provides a skeleton of the scripts and some examples of the code.

The CYBECO Toolbox of WP4 will provide a set of risk analysis templates based on these scripts. From the software implementation point of view, we have defined three ways of implementing the risk analysis cases. Specifically:

1. A risk analysis template that stores the analysis results in the Toolbox. In this case, we run the R scripts to generate tables that will be stored in the Toolbox, so that it is not necessary to run a simulation in R saving computational resources and time. The downside of this approach is that it is only useful for simple risk analysis or for consulting pre-calculated information.

2. A risk analysis template that performs simple calculations in the Toolbox. In this case, there is no interaction with the R scripts. The downside is that this approach is only useful for simple risk analysis.

3. A risk analysis template that interacts with R. In this case, the Toolbox provides input parameters to R, and R runs the risk analysis script for a specific number of simulation iterations. Once the simulation finishes, R provides its output to the Toolbox. This allows for a more granular risk analysis but may require computational resources and time.

6 The road ahead

This deliverable has presented the initial CYBECO modelling framework for cybersecurity risk management. We have been able to move beyond current cybersecurity risk management frameworks in the following directions:

- Unlike most of them, which are essentially based on risk matrices, we focus on detailed and careful analysis of likelihoods and multiple impacts of cyber threats, going beyond oversimplified ordinal models which may lead to inferior decisions.
- Unlike most of them, we are capable of taking into account the intentionality of some of the cyber threats, using the framework of adversarial risk analysis, combined with other risk analysis models.
- We incorporate various references to cyber insurance as a part of a cyber risk management strategy.
- We include behavioural aspects of cyber risk managers, including preferences and attitudes towards risk, through the utility functions included.
- We outline how to cope with eventual lack of data through structured expert judgement approaches.

Based on the use cases introduced in WP 4, we have presented graphically and analytically three models referring, respectively to,

1. the cyber insurance needs of an insurance company introducing cyber insurance products,
2. the decision of an insurance company to grant or not a cyber insurance product to a potential client, and
3. the decision of a company about its security resource allocation, including an eventual cyber insurance product.

This last one is the main and core model on which we have focused, both in methodology and applications, including a case that may be used as a template for more complex cases. Our approach, no doubt, entails more work than traditional cyber risk management approaches, however in many organisations the economical, environmental, political,… stakes at play are so large at that it such additional effort would be worth being implemented.

To facilitate application, we proceed in three directions. The first one refers to providing a generic cybersecurity preference model, based on identifying a set of generic cybersecurity objectives from a defender perspective, from which a risk manager may choose, as well as generic utility function which covers the above objectives (and caters for risk attitudes). The second one refers to developing the CYBECO Toolbox (Work package 5), for which we provide a high level description of the models to be implemented, the inputs required and the outputs produced. The third one sketches computational strategies that might alleviate the proposed initial computational scheme.

The core of this document has aimed at describing in an accessible manner the above developments, which are detailed in the enclosed technical appendices:

1. The cybersecurity resource allocation model, including a template case study.
2. The cyber insurance models developed.
3. The general cybersecurity preference model.
4. An algorithmic approach for general cyber defend-attack interactions.
5. An augmented probability simulation approach for large problems.
6. An outline on how the above may be implemented in the CYBECO toolbox.

The above framework will be revised, enhanced and completed on the second year of the project based on the behavioural experiments (Work package 6); the cyber insurance policy issues and gaps identified (Work package 7); the initial experiences with the CYBECO toolbox (Work package 5) and the case studies proposed (Work package 4). Work that we envision includes:

- A generic preference model for cyber attackers.
- An efficient robust computational approach to adversarial risk analysis that would benefit the CYBECO framework.
- Detailed analysis of models ii) and iii) with the corresponding template case studies.
- An analysis of all the case studies developed in WP 4.
- Suggestions for cyber insurance product design.

References

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

Balchanos, M.G. 2012. *A Probabilistic Technique for the Assessment of Complex Dynamic System Resilience*. Ph.D. Thesis, Georgia Institute of Technology (USA).

Bielza, C., Müller, P., and Rios Insua, D., 1999. "Decision Analysis by Augmented Probability Simulation". *Management Science*, Vol. 45, No 7, pp 995-1007.

Brenner, J.F. 2013. "Eyes Wide Shut: The Growing Threat of Cyber Attacks on Industrial Control Systems". *Bulletin of the Atomic Scientists*, Vol. 69, No. 5, pp. 15-20.

Cardenas, A., Amin, S., Sinopoli, B., Giani, A., Perrig, A. and Sastry, S. 2009. "Challenges for Securing Cyber Physical Systems". *Workshop on Future Directions in Cyber-Physical Systems Security*.

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

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

Cloud Security Alliance. 2016. *Cloud Controls Matrix*.

Command Five PTY LTD (Australia). 2011. *Advanced Persistent Threats: A Decade in Review*.

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

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

Dantu, R., Kolan, P., Akl, R. and Loper, K. 2007. "Classification of Attributes and Behavior in Risk Management Using Bayesian Networks". *IEEE Intelligence and Security Informatics 2007*, pp. 71-74.

Defense Science Board, DoD (USA). 2013. *Task Force Report: Resilient Military Systems and the Advanced Cyber Threat*.

DeNardis, L. 2015. "Five Destabilizing Trends in Internet Governance". *I/S: A Journal of Law and Policy for the Information Society*, Vol. 12, No. 1, pp. 113-133.

Dyer, J. and Sarin, R. 1979. "Group Preference Aggregation Rules Based on Strength of Preference." Management Science, Vol. 25, No. 9, pp. 822-832.

Dyer, J. and Sarin, R. 1982. "Relative Risk Aversion." Management Science, Vol. 28, No. 8, pp. 875-886.

European Telecommunications Standards Institute. 2015. *ETSI GS ISI 002 v1.2.1 - Information Security Indicators; Event Model, A Security Event Classification Model and Taxonomy*.

European Union Agency for Network and Information Security. 2007. *IT Business Continuity Management – An Approach for Small Medium Sized Organizations*.

European Union Agency for Network and Information Security. 2007. *Business and* .

Herley, C. and Florêncio, D. 2010. "Nobody Sells Gold for the Price of Silver: Dishonesty, Uncertainty and the Underground Economy". *Economics of Information Security and Privacy*, pp. 33-53. Springer.

International Organization for Standardization. 2013. *ISO/IEC 27005:2013 – Information Security Risk Management*.

International Organization for Standardization. 2013. *ISO/IEC 27001:2013 – Information Security Management Systems – Requirements*.

International Organization for Standardization. 2014. *ISO 55000:2014 – Asset Management – Overview, Principles and Terminology*.

International Organization for Standardization. 2014. *ISO/IEC 19770-1:2017 – Part 1: IT asset management systems -- Requirements*.

International Organization for Standardization. 2015. *ISO 22317: Societal Security – Business Continuity Management Systems – Guidelines for Business Impact Analysis*.

International Organization for Standardization. 2018. *ISO 31000:2018 – Risk Management – Guidelines*.

Keeney, R. 2007. "Modeling Values for Anti-terrorism Analysis." Risk Analysis, Vol. 27, No. 3, pp. 585–596.

Keeney, R. and Gregory, R. 2005. "Selecting attributes to measure the achievement of objectives." *Operations Research*, Vol. 53, No. 1, pp 1–11.

Keeney, R. and von Winterfeldt, D. 2011. "A Value Model for Evaluation Homeland Security Decisions." Risk Analysis, Vol. 31, No. 9, pp 1470–87.

Koller, D., and Milch, B. 2003. "Multi-agent influence diagrams for representing and solving games". *Games and Economic Behavior*, Vol. 45, No. 1, pp 181-221.

Lee, R. M., Assante, J. and Conway, T. 2014. ICS Defense Use Case Dec 301, 2014 - German Steel Mill Cyber Attack. SANS Institute (USA).

Li, Z., Liao, Q. and Striegel, A. (2009). "Botnet Economics: Uncertainty Matters". *Managing Information Risk and the Economics of Security*, pp. 245-267. Springer.

Lloyd's (UK). 2017. *Counting the Cost - Cyber Exposure Decoded*.

Low, P. 2017. "Insuring Against Cyber-Attacks". *Computer Fraud & Security*, Vol. 2017, No. 4, pp. 18-20.

Lund, M.S., Solhaug, B. and Stølen, K. 2010. *Model-driven Risk Analysis: the CORAS Approach*. Springer.

Marotta, A., Martinelli, F., Nanni, S., Orlando, A., and Yautsiukhin, A. 2017. "Cyber-Insurance Survey". *Computer Science Review*, Vol. 24, pp. 35-61.

McAfee (USA). 2014. *Net Losses: Estimating the Global Cost of Cybercrime*.

Merrick, J. and Parnell, G.S., "A Comparative Analysis of PRA and Intelligent Adversary Methods for Couterterrorism Risk Management." *Risk Analysis*, Vol. 31, No. 9, pp 1488-1510.

Ministerio de Hacienda y Administraciones Públicas (Spain). 2012. *Methodology of Analysis and Risk Management Information Systems, version 3*.

National Institute of Standards and Technology (USA). 2012. *NIST SP 800-30 Rev. – Guide for Conducting Risk Assessments.*

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

Nisan, N., Roughgarden, T., Tardos, E., and Vazirani, V.V. (Eds.). 2007. *Algorithmic Game Theory*, Vol. 1. Cambridge University Press.

Organization for Economic Cooperation and Development. 2017. *Enhancing the Role of Insurance in Cyber Risk Management*.

Ortega, J., Rios Insua, D., and Cano, J. 2017. "Adversarial Risk Analysis for Bi-agent Influence Diagrams". *XXXVI Congreso Nacional de Estadística e Investigación Operativa*.

Rios Insua, D., et al. 2017. "Forecasting and Assessing Consequences of Aviation Safety Occurrences." Unpublished.

Sastry, S., Cardenas, S., and Amin, A.A. 2008. "Research Challenges for the Security of Control Systems". *Proceedings of the 3rd Conference on Hot Topics in Security*, pp. 6:1-6:6.

Shachter, R.D. 1986. "Evaluating Influence Diagrams". *Operations Research*, Vol. 34, No. 6, pp 871-882.

Tetlock, P.E., and Gardner, D. 2015. *Superforecasting: The Art and Science of Predicition*. Broadway Books.

The Common Criteria Recognition Agreement Members. 2012. *Common Criteria for Information Technology Security Evaluation, Version 3.1 Release 4.*

The Open Web Application Security Project. "OWASP Risk Rating Methodology" https://www.owasp.org/index.php/OWASP_Risk_Rating_Methodology [Ret. 04/18]

World Economic Forum. 2017. *The Global Risks Report 2017.*

Yaqoob, I., Ahmed, E., Ur Rehman, M.H., Ahmed, A.I.A., Al-Garadi, M.A., Imran, M., and Guizani, M. 2017. "The Rise of Ransomware and Emerging Security Challenges in the Internet of Things". *Computer Networks*, [In Press].

Acronyms and Abbreviations

GDP	Gross Domestic Product
ICT	Information and Communication Technologies
IT	Information Technologies
ID	Influence diagram
BAID	Bi-agent Influence Diagram
MAID	Multi-agent Influence Diagram
ARA	Adversarial Risk Analysis
APS	Augmented Probability Simulation
MC	Monte Carlo Simulation
MCMC	Markov Chain Monte Carlo Simulation
WP	Work Package
SME	Small and Medium Enterprise
DDoS	Distributed Denial of Service

Annexes

We provide the following annexes:

- Annex 1: Paper ‘An Adversarial Risk Analysis Framework for Cybersecurity’
- Annex 2: Paper ‘Some Decision Problems in Cyber Insurance Economics’
- Annex 3: Paper ‘A generic preference model for cyber security defenders’
- Annex 4: Paper ‘Adversarial Risk Analysis for Bi-agent influence diagrams: An algorithmic approach’
- Annex 5: Paper ‘Augmented simulation methods for game theoretic problems’
- Annex 6: Skeleton and examples of the R Routines