Universidad
Rey Juan Carlos

TESIS DOCTORAL

Decision Models for Cybersecurity Risk Analysis

Autor:

Aitor Couce Vieira

Directores:

David Ríos Insua
Siv Hilde Houmb

Programa de Doctorado en Tecnologías de la Información y las Comunicaciones

Escuela Internacional de Doctorado

2019

Universidad
Rey Juan Carlos

Tesis Doctoral

Decision Models for Cybersecurity Risk Analysis

Aitor Couce Vieira

Universidad
Rey Juan Carlos

Decision Models for Risk Analysis in Cybersecurity

Aitor Couce Vieira

Universidad
Rey Juan Carlos

TESIS DOCTORAL

Decision Models for Cybersecurity Risk Analysis

Autor:

Aitor Couce Vieira

Directores:

David Ríos Insua
Siv Hilde Houmb

Programa de Doctorado en Tecnologías de la Información y las Comunicaciones

Escuela Internacional de Doctorado

2019

David Ríos Insua, Catedrático de Estadística e Investigación Operativa de la Universidad Rey Juan Carlos,

y **Siv Hilde Houmb**, Profesora Asociada II del Laboratorio Noruego de Seguridad de la Información de la Universidad Noruega de Ciencia y Tecnología,

directores de la Tesis Doctoral *Decision Models for Cybersecurity Risk Analysis*, realizada por el doctorando **Aitor Couce Vieira**,

hacen constar que esta Tesis Doctoral reúne los requisitos necesarios para su defensa y aprobación.

En / In MADRID, ESPAÑA ,
(lugar/place)

JUNE 9, 2019
(fecha/date)

David

David Ríos Insua
(firma/signature)

David Ríos Insua, Professor of Statistics and Operations Research at Rey Juan Carlos University,

and **Siv Hilde Houmb**, Associate Professor II at the Norwegian Information Security Laboratory of the Norwegian University of Science and Technology,

supervisors of the Doctoral Thesis *Decision Models for Cybersecurity Risk Analysis*, submitted by the PhD student **Aitor Couce Vieira**,

declare that this Doctoral Thesis fulfils the requirements needed for its presentation and approval.

En / In OSLO, NORWAY ,
(lugar/place)

JUNE 6, 2019
(fecha/date)

Siv Hilde Houmb

Siv Hilde Houmb
(firma/signature)

Acknowledgements

As the URJC's regulation on PhD studies declares, "the Thesis should capacitate the PhD student for working autonomously in the field of research, development and innovation ". That is, knowing how to create valid knowledge through logic and facts (research), how to put in practice such knowledge (development) and how to provide an original and relevant solution to a societal problem (innovation).

That was my main focus on this period I have gone through, 2012 to 2018: learning and gaining experience on these fields with my MSc in Decision Engineering, my work at Secure-NOK and ICMAT and my work on the PhD. This Thesis reflects, hence, the spirit of this personal period.

I am grateful to Siv and David, my Thesis supervisors, for the specific guidance and support. However, most of my recognition and admiration comes from working with them. At Secure-NOK with Siv from 2013-2016 and with David at ICMAT since 2017 to work in the CYBECO project. This experience has been insightful for what they told me but also for what I observed from them. Beyond the valuable time they dedicated to my dissertation, I would like to emphasise two aspects extremely precious to me. David's productivity as a researcher and Siv's holistic view of cybersecurity.

David's ability to produce models focused on the problem at hand, yet mathematically strong, helped me to raise the formal strength of my work. His productivity and experience in all aspects involved in research avoided me headaches and dead paths and allowed me to understand how to undertake solid scientific work.

Siv is an expert in cybersecurity in both the academic and business sides. This helped me to find a good conceptualisation of the field, but also in gaining confidence in the applicability of our approach. The work at Secure-NOK was not directly related to the PhD, yet it helped indirectly since we worked in many aspects of an innovative startup, from sales to research, going through risk analysis.

In summary, thanks to David for teaching me how to transfer and apply scientific knowledge to risk analysis and thanks to Siv for teaching me how to seek scientific answers to risk analysis problems.

In this context, I would also like to recognise all the people I worked or shared experiences during these years. My colleagues at Secure-NOK, the SPOR group at ICMAT and also most people involved in the EU research projects Seconomics and CYBECO.

I am also grateful to the institutions that supported my work directly or the work of my directors. Namely to the European Commission for the support of the Horizon 2020 project 740920 CYBECO on Supporting Cyber Insurance from a Behavioural Choice Perspective; To the Spanish Ministry of Economy and Competitiveness for supporting David with the MINECO MTM2014-56949-C3-1-R project; To AXA for supporting David with the AXA-ICMAT Chair in Adversarial Risk Analysis; To the European Union for supporting David with the COST IS1304 Action on Expert Judgement; And to the West Norway Regional Fund for supporting Siv – and me while working at Secure-NOK – with the project 245291 on a Cybersecurity Incident Response Framework.

O meu recoñecemento mais especial é para a miña familia, que estendo tamén ás miñas amizades. Estes anos que botei polo mundo adiante foron moi produtivos para gañar currículo, cartos e experiencia vital. Pero a costa de non botar tempo cos meus seres queridos e na miña terra. Inda máis agredido teño que estar por todo o esforzo e sacrificio por parte da miña familia para criar e educarme ata chegar aquí. Neste ultimo aspecto, no da educación, quero agradecer tamén a tódolos bos mestres que tiven ao longo da miña vida. É máis, quero reivindicar á educación pública, os valores que esta representa e a tódalas persoas que loitaron por traela, mantela e mellorala.

Sen este esforzo familiar e social esta Tese tampouco sería posible
e, por tanto, dedícovola a todos vos.

Resumen de la Tesis

Motivación

Nuestra sociedad está profundamente digitalizada. En algunas áreas, esto es un hecho desde hace décadas; e.g., los sistemas informáticos en empresas y administraciones públicas. En otros ámbitos la digitalización ésta emergiendo, como en el caso de los procesos industriales, que se ha acelerado en la última década.

Esto ha traído nuevas amenazas de ciberseguridad que podrían poner en riesgo procesos industriales, la seguridad de los trabajadores o el entorno medioambiental. De hecho las infraestructuras industriales ya han sido objetivo de sofisticados ciberataques, como Stuxnet o Shamoon, capaces de permanecer ocultos mientras realizan operaciones de sabotaje o espionaje. Otros paradigmas emergentes se enfrentan a riesgos ciber-físicos similares: e.g., el Internet de las Cosas, las ciudades inteligentes o los coches autónomos. En definitiva, la ciberseguridad en entornos expuestos a riesgos con consecuencias físicas es muy diferente a la ciberseguridad tradicional, centrada en la confidencialidad de la información y la privacidad.

Es más, nuestra vida moderna depende cada vez más de las tecnologías digitales. Actuamos y nos relacionamos en multitud de ciberespacios y, por tanto, nos exponemos a riesgos psicológicos y sociales que pueden ser explotados por terceros maliciosos, como las campañas de noticias falsas y trolls, el ciber-acoso o la exposición pública de datos personales.

Por tanto, es vital estudiar estos riesgos digitales para entender qué son y cómo nos afectan. Para esto existen multitud de métodos (e.g., matrices de riesgo, *bow-ties*) que, sin embargo, encontramos insuficientes a la hora de cubrir ciertos aspectos que consideramos relevantes como, por ejemplo, el estudio de amenazas adversarias o la existencia de objetivos de distinta naturaleza (e.g., monetarios, derechos personales).

Objetivos

Nuestro objetivo en esta Tesis es desarrollar *modelos de análisis de riesgos en ciberseguridad* que estudien aspectos no muy bien tratados por los métodos actuales más populares. Concretamente:

- Modelos que *analicen los riesgos durante incidentes*, que difieren de un análisis de riesgos típico en que el analista estudia un incidente particular que está ocurriendo o que podría ocurrir inmediatamente.
- Modelos que *analicen estratégicamente las amenazas adversarias*, ya que, en ciberseguridad, los análisis de riesgos típicos generalmente no tienen en cuenta el comportamiento o motivaciones de las amenazas inteligentes.
- Los riesgos digitales podrían causar impactos en la información, operativos, físicos o psicológicos. Esto requiere modelos que *faciliten la toma de decisiones con objetivos múltiples* de distinta naturaleza, valor e importancia para las partes involucradas.
- La *inclusión de la transferencia de riesgo*, en particular los ciber seguros, en los análisis de riesgos en ciberseguridad, como complemento a los controles de seguridad preventivos y reactivos.

Resultados

Después de introducir los temas tratados en la tesis, los siguientes dos capítulos se centran en el análisis de riesgos durante incidentes. El segundo capítulo presenta nuestro modelo general de análisis de riesgos durante incidentes (GIRA), que formaliza el proceso de dicho análisis mediante un diagrama de influencia. Primero, exponemos las consideraciones que se han de tener en cuenta a la hora de analizar los riesgos durante un incidente. Seguimos con una caracterización de los componentes básicos que constituyen un incidente y de las relaciones entre ellos. Partiendo de esta caracterización, introducimos GIRA y las particularidades de sus componentes: exposición a la amenaza, respuesta al incidente, materialización del incidente, consecuencias en los sistemas, impactos en los activos, objetivos en riesgo y evaluación del riesgo. Acompañamos GIRA con ejemplos. También presentamos, brevemente, la formalización matemática de GIRA y versiones adicionales: simplificada, para múltiples partes involucradas y para sucesos inminentes y futuros. GIRA se sitúa al mismo nivel de generalidad que los conceptos de riesgo e incidente establecidos en las normas ISO 31000 e ISO 22300.

El tercer capítulo presenta avances adicionales para GIRA y una adaptación para realizar un análisis rápido de riesgos en ciberseguridad. Presentamos un método simple de obtención cualitativa de probabilidades basado en la rareza del suceso (i.e., en función de si los diferentes sucesos en una cadena de sucesos son ciertos, posibles, raros o imposibles). Además, introducimos un mapa de categorías para entender las ramificaciones potenciales de los incidentes de ciberseguridad. Luego presentamos nuestro modelo para el análisis de riesgos durante incidentes de ciberseguridad (CSIRA) que es, básicamente, GIRA combinado con los previamente introducidos métodos de obtención y mapa de ramificaciones de incidentes de ciberseguridad. En la presentación de CSIRA también exponemos que para tomar la decisión sólo es necesario comparar los escenarios a los que conducen las distintas respuestas al incidente, sin la necesidad de obtener preferencias, típica en el uso de diagramas de influencia.

El resto de modelos se centran en el marco temporal típico de los análisis de riesgos, i.e., la vida útil de un sistema o un número de años específico.

Así, el cuarto capítulo presenta un modelo de asignación de recursos en ciberseguridad en una organización, incluyendo sus preferencias y actitudes frente al riesgo, la intencionalidad de las amenazas adversarias y las decisiones respecto a la adquisición de ciber seguros. La primera parte introduce diagramas de influencia, y su forma matemática, que describen diferentes modelos de análisis de riesgo. Empezando por una evaluación simple del rendimiento de un sistema, vamos añadiendo nuevos elementos al modelo: riesgo, mitigación del riesgo, transferencia del riesgo y análisis adversario. La segunda parte reproduce un ejemplo completo en el que detallamos todos los aspectos del estudio de riesgos: descripción de la estructura del problema de riesgos, estudio de las creencias de la organización sobre los elementos que afectan al riesgo, estudio de sus preferencias, modelización del problema del atacante para predecir sus acciones y cálculo de la mejor cartera de controles y seguro para la organización.

En el capítulo cinco describimos un arbol de objetivos para ciberseguridad. El propósito es facilitar una identificación exhaustiva de los objetivos de una organización que pueden ser afectados por ciber riesgos. En este contexto, es importante distinguir entre aquellos objetivos que pueden medirse en términos monetarios y aquellos que no pueden [o no deben] medirse en tales términos - por ejemplo daños a personas. También exploramos como medir esos objetivos no monetarios (e.g., reputación, derechos personales, daños medioambientales). Finalizamos detallando cómo usar este arbol de objetivos para construir una función de utilidad multi-atributo.

En el sexto capítulo presentamos varios modelos de análisis de riesgos en el contexto de los ciber seguros. También presentamos modelos para aseguradoras. En el primero, se decide qué reaseguro adquiere teniendo en cuenta los diferentes segmentos de compañías a las que está asegurando (e.g., PYMES, grandes empresas). En el segundo, la aseguradora decide si otorga o no un seguro a un cliente potencial.

Los resultados de la investigación se han materializado en cuatro artículos:

1. Rios Insua, D., Couce-Vieira, A., Rubio, J.A., Pieters, W., Labunets, K., y Rasines, D.G. "An Adversarial Risk Analysis Framework for Cybersecurity." En *Risk Analysis*.[1]
2. Couce-Vieira, A., Rios Insua, D., y Houmb, S.H. (2019) "GIRA: A General Model for Incident Risk Analysis." En *Journal of Risk Research*, Vol. 22, No. 2, pp. 191–208.[2]
3. Couce-Vieira, A., Houmb, S.H., y Rios Insua, D. (2018) "CSIRA: A Method for Analysing the Risk of Cybersecurity Incidents." En *Proc. of the 4th International Workshop on Graphical Models for Security*, LNCS Vol. 10744, pp. 57–74. Springer-Verlag.[3]
4. Rios Insua, D., Couce-Vieira, A., y Kreshnik, M. (2018) "Some Risk Analysis Problems in Cyber Insurance Economics." En *Estudios de Economía Aplicada*, Vol. 36-1, pp. 181–194.[4]

También hemos escrito informes técnicos para el proyecto de innovación europeo CYBECO, en que detallamos, en términos más generales, los contenidos tratados en los capítulos cuatro a seis. Específicamente, *D3.1: Modelling framework for cyber risk management 7*[5] and *D3.2: Improved modelling framework for cyber risk management*.

Conclusiones

La contribución de GIRA/CSIRA es un modelo de análisis de riesgos para situaciones de incidente con una representación matemática formal y fundamentado en una caracterización sintética, pero abarcadora, del concepto

[1] Publicado online 10/06/2019, `doi:10.1111/risa.13331`

[2] Publicado online 11/09/2017, `doi:10.1080/13669877.2017.1372509`

[3] `doi:10.1007/978-3-319-74860-3_4`

[4] `www.revista-eea.net/volumen.php?Id=99&vol=36&ref=1` [Rec. 30/05/2019]

[5] Disponible en `www.cybeco.eu/results` [Rec. 30/05/2019]

de incidente. Encontramos diferentes consideraciones que los métodos existentes no tratan de manera adecuada. Modelos como el *bow-tie* no cubren aspectos relacionados con el valor (activos, impactos, evaluación del riesgo) y las matrices de riesgo llevan a análisis sobresimplificados que cualifican el riesgo de manera inadecuada. GIRA/CSIRA tienen en cuenta la cadena de sucesos sistémicos y los aspectos de valor que componen un riesgo. Por otro lado, como diagramas de influencia, son compatibles con métodos cuantitativos de análisis de riesgos y con la obtención de preferencias para la construcción de funciones de utilidad. Sin embargo, también hemos desarrollado métodos para facilitar un análisis de riesgo rápido.

El marco para el análisis de riesgos en cibeseguridad contribuye al análisis de riesgos adversarios (ARA). Aplica modelos y aspectos existentes, pero integra el análisis adversario en la gestión de riesgos de ciberseguridad de la organización analizada, y aporta la descripción completa del procedimiento bajo el que se ha realizado y construido el estudio de riesgo. Este marco también incorpora los ciber seguros como componente del análisis del riesgo y sus particularidades (e.g., su dependencia de las medidas de seguridad implantadas, su influencia en el impacto monetario final de un ciberataque). Este es uno de los primeros capítulos de análisis de riesgos en ciberseguridad que incluye ciber seguros y, que sepamos, el primero que integra amenazas adversarias y ciber seguros. Este marco se complementa con el arbol de objetivos para ciberseguridad, que facilita el trabajo de identificar y medir los diferentes objetivos en riesgo en el contexto de la ciberseguridad. Aportamos una lista genérica pero comprehensiva de los objetivos a tener en cuenta a la hora de crear una organización cibersegura. También definimos la manera de medir estos objetivos e integrarlos en funciones de utilidad.

Este último marco ha sido parte del proyecto europeo de innovación CYBECO, bajo el programa H2020, centrado en el desarrollo de nuevas herramientas de análisis de riesgos en ciberseguridad y ciber seguros. Durante 2018 y 2019 avanzamos en el desarrollo de algoritmos para implementar una herramienta informática de análisis de riesgos basada en nuestro modelo y su metodología de construcción. Estos avances también pueden ser replicados o adaptados a GIRA/CSIRA, ya que son modelos más sencillos que representan problemas de riesgos más simples. En el futuro también sería de interés la realización de investigaciones descriptivas usando los modelos anteriores, por ejemplo la elaboración de estudios de riesgos en empresas prototípicas (e.g., PYMES), que podrían ser de interés para gobiernos o para compañías aseguradoras o de ciberseguridad.

Contents

Chapter 1

Introduction to the Thesis

1.1 Research motivation

Digitalisation is pervasive in our society. In some realms, this has been ubiquitous for a long time, e.g. the information systems at companies and public administration. In other domains, digitalisation is emerging. A paradigmatic case during the last decade has been the digitalisation of industry. This increased automation and connectivity has exposed industrial processes and facilities to cybersecurity risks, which can cause incidents with the equipment that could compromise operations, safety, or the environment. These systems are now the target of sophisticated cyber attacks, such as Stuxnet or Shamoon, capable of working in the background to conduct espionage or sabotage actions. An interruption of operations of a few hours could represent tens of thousands of Euros. Potentially, a manipulation of the equipment – even unintentional – might cause or facilitate an incident with safety or environmental consequences. Therefore, managing cybersecurity risks and incidents in these industrial environments exposed to physical consequences is fundamentally different from traditional cybersecurity, which focuses on information and privacy risks. Other emerging paradigms face similar cyber-physical risks, for instance, in relation with the Internet of Things, Smart Cities or autonomous cars. Additionally, our life is so dependent on digital technologies or cyberspaces that new psychological and social risks have come to the fore, such as the social impact of fake news and trolls, the psychological impact of cyber-bullying or the public exposition of intimate photos.

Our motivation is to bring innovative *cybersecurity risk analysis models* that address aspects not well covered by popular cybersecurity risk analysis methods. Specifically:

- Models that *address risk analysis during incidents*, which differ from traditional risk analysis in that the analyst studies a particular incident that is happening or could happen immediately.
- Models that *address the strategic analysis of adversarial threats*, since in traditional cybersecurity risk analysis these are usually studied without taking into account their behaviour or motivations.
- Digital risks – as described in the introductory paragraph – might lead to informational, operational, physical or psychological impacts and thus require models that *facilitate decision-making with multiple objectives* of different nature, value and importance for the involved stakeholders.
- The *inclusion of risk transfer*, insurance in particular, in cybersecurity risk analysis as a complement to protective and reactive measures.

The next two sections introduce the main topics of the Thesis. First, we introduce the general activity for which we build our model: *risk analysis* (Sect. 1.2). All activities and assets of an organisation involve risks and incidents, which – whatever their nature is – can be studied and managed in a common comprehensive way. Later, we introduce the domain in which we want to undertake risk analysis and, therefore, the domain in which the models are used. This domain is *cybersecurity* (Sect. 1.3). Indeed, we highlight on a specific domain, industrial cybersecurity, since it is paradigmatic of the full map of cybersecurity risks: informational, operational, physical and human. In further chapters we provide a more thorough review of the state of the art in the different topics addressed by this Thesis.

The last sections of this chapter present the research objectives (Sect. 1.4), methodology (Sect. 1.5) and an outline of our results (Sect. 1.6).

1.2 Risks and incidents

We introduce now the initial set of concepts for understanding risks, threats and incidents. The reason is that the high-level characterisation and study of risks and incidents are similar in most fields. This is manifested in the most widely used standards on these matters: the well established ISO[1] 31000 series on risk management [83] and the emerging ISO 22300 series on societal security [84] (which includes incidents in general terms).

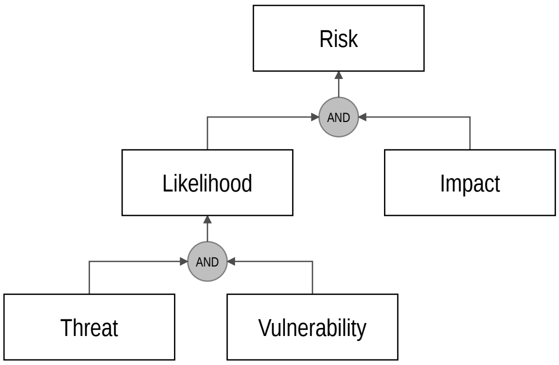
[1] International Organisation for Standardization

Risk and its components

Risk can be described as the possibility of an undesirable event resulting in a negative impact. A risk consists of the following two elements[2]: The *likelihood* of the undesirable event to happen and the *impact*, i.e., the damage caused by the undesirable event. If the event cannot happen, or if there are no undesirable consequences, there is no risk.

In addition, the likelihood of the event requires the existence of the following two elements: The *threat*, i.e., an element with the potential to induce damage; and the *vulnerability*, i.e., a weakness or condition of a system that, if exploited, could result in an undesirable consequence. If there is no threat or vulnerability, the undesirable event is not possible and therefore, there is no risk. Figure 1.1 depicts how risk components relate between them.

Fig. 1.1: Risk Components



An example of risk is the collision of an asteroid with the Earth. In this case, the threat is the asteroid. The vulnerabilities are (1) the Earth being in the trajectory of the asteroid and (2) the Earth's atmosphere being not capable of destroying the asteroid during its entry. The impact, in risk terms[3], consists of life destruction, civilisation destruction and climate alteration.

Another example of risk is the failure of the introduction of a product in a new market. In this case, there are multiple threats such as competing products; legal issues; or the customer's needs, tastes or purchasing power. The impact would be measured mostly in monetary losses.

[2] There are multiple definitions about risk components, although most of them are related. For a detailed compilation see Appendix M of [156], also available in Sect. 3 of [158].

[3] Technically, in risk terms, the physical impact of the asteroid on Earth is the potential incident.

Distinction between risks and hazards

In addition, it is important to differentiate between risks and *hazards*, which are situations that possess inherent and known dangers that are more predictable and easier to isolate than threats. There is no clear-cut distinction between them. Even though, we can difference them as follows:

- Hazards are situations for which it makes no sense to be in for a given activity (e.g., electrical current or molten material in a facility, cliff besides a road) and that can be avoided to a certain degree due to their easier isolation or predictability.
- On the other hand, risks are situations that, first, would be necessary to be in (e.g., severe weather in a town, work on heights), or, second, cannot be easily isolated and avoided (e.g., electrocution risk in high-voltage lines maintenance, security threats).

Security threats

When it comes to the characterisation of threats, it is important to distinguish between security threats and other threats [158] although, here too, there is no clear distinction between them. A *security threat* is any intentional or unwarranted action with the potential to cause loss or damage, with or without such intention, by exploiting the vulnerabilities of a system [151].

Other threats are any other circumstances or accidents – including acceptable actions – with the potential to cause loss or damage. They can be natural or technological, involving in both cases biological, chemical, mechanical, technological, environmental or physical agents. Safety practice usually deals with these threats and hazards.

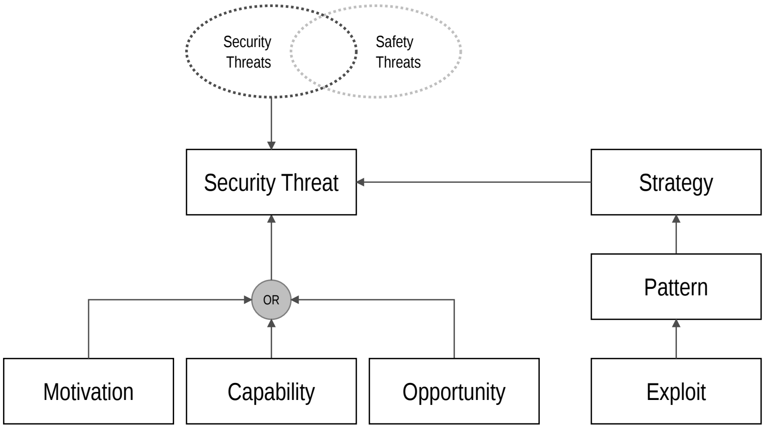
There is an overlap between security and safety, e.g., when the enforcement of safety policies becomes a security problem. For instance, driving a car drunk or at high speed are, originally, safety problems but they have been established as unwarranted actions and, thus, forbidden by a security agency – the police. An additional overlap is the exposition to safety and security risks in the decisions and actions we make, mitigating or amplifying vulnerabilities or threats, such as making business decisions focusing on cost instead of safety.

Regarding security threats, there is no widely standardised definitions of its components. The most common are capability, motivation and opportunity. This triad is a principle of criminal law and is widely used in fields such as criminology, security studies and cybersecurity. Indeed, this triad is used by behavioural analysts to understand behaviour [121]. Under this approach, a threat exists if it presents at least one of the following three factors: The threat *capability* representing the resources and skills necessary to become a threat, the threat *motivation* as the rationale and the intention to be a threat and the threat *opportunity* as the environmental elements or conditions that enable or facilitate the threat.

Even so, these aspects cover what we could call the profiling of the threat actor, i.e., what generates behaviour. Another important aspect is the behaviour itself, i.e., the threat actions. We can distinguish between their strategic behaviour, their tactical patterns and their exploits. First, the *strategic behaviour* involves the selection of their targets and goals and it is closely related with the threat motivations. Second, the *tactical patterns* refer to how they perform their actions to meet their goals. These patterns are usually characterised using methods such as *attack trees* or *kill chains*. Finally, *exploits* are the specific way the threat execute their actions in a system.

Figure 1.2 depicts how threat components relate between them.

Fig. 1.2: Threat Components



Incidents

An incident is a 'situation that might be, or could lead to, a disruption, loss, emergency or crisis' [84]. Similar concepts such as event, incidence, failure, emergency, or disaster are used to describe incidents of different intensity.

Incidents and risks are closely related. In a general sense, a risk is nothing but a prospective incident, and an incident is nothing but a materialised risk. Incidents are usually complex events that might consist of multiple events and lead to multiple types of consequences and new incidents. This nature forms the basis to the use of risk analysis techniques.

In the previous examples of risks, the incident of the first example would be the physical impact of the asteroid. In the second example, the incident would be a complex event consisting of each of the non-purchases made by potential costumers. However, from the organisation perspective the incident would be detected during the monthly sales report or any other similar information.

The case for incident risk analysis

Risk analysis techniques aim to identify what negative scenarios could happen and what the likelihood of such negative scenarios is. Generally, a risk analysis answers these questions in terms of a relatively long period (e.g., several years or the life of a project or installation). However, these questions are also relevant during incidents.

When an event associated with an incident is detected, incident handlers might be interested in identifying what events could happen next and what their likelihoods are. In the example of the asteroid collision risk, when a threatening asteroid is detected, the astronomers focus on identifying what events could happen next (e.g., the potential trajectories) and their likelihood (e.g., the probability of the different trajectories).

Uncertainty might exist also with the actual detected event: sometimes the event is rather an indication of the incident, which could be caused by any other harmless event. For instance, in a computer, an overuse of the fan (detected event) could be caused by a high use of the CPU (which can be caused by malware, or demanding applications) or environmental conditions (which can be caused by bad air circulation, moisture or room temperature).

The case for multi-objective risk analysis

When we make decisions, we have multiple and mostly incommensurable objectives [48]. Risk analysis is a decision activity, since one of its steps is to decide between different alternatives of risk treatments. Risks and their treatments might impact organisations in multiple ways and, therefore, a method is necessary for prioritising between treatments. But, ultimately, the consequences of risks are incomparable in different degrees. Sometimes this comparison can be solved in a systematic and reasonable way. For instance, the comparison between a consequence in reputation versus a loss of revenue is not straight-forward but, in the end, reputational consequences might be estimated as monetary costs. However, other times this comparison is challenging both analytically and ethically. For instance, when we face risks that involve loss of human life.

The purpose of this Thesis on this aspect is not to describe how to select between incommensurable objectives, but to facilitate the decision-making on this topic. As we discuss in the chapters, we found that popular risk analysis methods oversimplify this problem. For instance, the widely used risk matrices represent risk scenarios in a single impact dimension that synthesise all kinds of losses in a single scale metric. We think, supported on the literature on human decision-making, that this oversimplification might be counterproductive and, thus, it would be better to provide a risk analysis model with a more thorough way for comparing multiple objectives in a manner that better fits with user preferences on this decision problem and their limitations as decision-makers.

The case for adversarial risk analysis

In risk analysis, intelligent threats are usually modelled as phenomena that happens with a certain frequency. In social fields, from the Social Sciences literature to day-to-day business or criminal practice, social agents are generally studied and characterised as elements with purpose, capable of making decisions and with a behavioural pattern that reflects those decisions and purposes. However, most risk analysis involving intelligent agents (e.g., security risks such as the ones in cybersecurity) lack this characterisation and, instead, they usually model all threats in the same way: as events that happen with a certain frequency. Some approaches detail the chain of previous events that might trigger the main events but, at the end, such events are also modelled as frequencies. In the case of our Thesis, we will adapt our model to the advances in the field of Adversarial Risk Analysis [145], which provides a framework for analysing threats as intelligent adversaries with behaviour and objectives.

1.3 Cybersecurity

In this section we briefly present the domain in which our Thesis concerning risk analysis lies: cybersecurity. We introduce relevant concepts to understand it, including cyber threats and their strategic importance. In addition we also present the challenges that make industrial cybersecurity different from traditional cybersecurity, derived, basically, from the technical and functional differences between Operational Technology (OT), used in industrial processes, and traditional Information and Communication Technology (ICT or IT). This emphasis in industrial cybersecurity, with impacts beyond information, helps us to develop a more comprehensive risk analysis modelling approach capable of being applicable in any cybersecurity domain.

Cybersecurity is the protection of digital systems from malicious and unwarranted actions. In a broad sense, such systems consist of computers that process (i.e., create, access, modify and destroy) and store data, networks that communicates data between computers, input devices that enable computers to get non-digital input (e.g., keyboards or sensors) and output devices that enable computers to provide non-digital output (e.g., screens or actuators). It is also important to understand what these systems are (their technology, i.e., the technical characteristics of a digital system) and what these systems are for (their function, i.e., the ultimate purpose or activity for which a digital system is employed).

When it comes to digital systems, we can distinguish between several broad classes that differ technically and functionally. ICT represents the infrastructure that supports information and communication functionalities, as well as electronic services (e.g., e-government, e-commerce, e-banking). OT is an umbrella term for naming the technological infrastructure that supports automation, control and data acquisition functions on physical systems (e.g., industrial control systems or logistic automation). A related emerging technology is the Internet of Things (IoT), which refers to interconnected physical devices such as vehicles, homes, streets and even animals. IoT and OT are the foundation of more specific technologies such as smart grids, smart cities and intelligent transport.

The distinction between the previous groups is important for cybersecurity. Traditionally, cybersecurity focused on the challenges and requirements of ICT systems; namely, protecting the confidentiality of the data. However, technologies such as OT or IoT represent a change in the security paradigm, moving the focus towards the protection of the physical processes and assets.

At its core, cybersecurity follows the basic principles of the more general field of *information security*. These principles act as both attributes and goals, and cover the following famous triad: *Confidentiality*, i.e., the data shall not be made available or disclosed to unauthorised parties; *Integrity*, i.e., the data shall be accurate and complete over its entire life cycle; and, *Availability*, i.e., the data shall be available whenever is needed. In general, any cybersecurity objective, acceptance criteria or policy can be reduced to a combination of the previous triad. Although all three principles are important, most of the time they entail trade-offs between them [18].

Several guidelines have been developed to implement cybersecurity. The main standards in the field are the ISO/IEC[4] 27000 series on Information Security Management Systems [85] and standards within the NIST[5] SP–800 series of Computer Security, e.g. NIST SP 800-53 [128]. Both collections are also applicable in the OT domain, in addition to the ISA[6]/IEC 62443 series on Security for Industrial Automation and Control Systems [91].

Given the ubiquity of digital systems in today's economy, cybersecurity has become an issue of major importance - both technically and financially. Especially, if we take into account the increasing sophistication of attacks suffered by companies and public institutions worldwide. At present, all kinds of organisations are being critically impacted by cyber threats [7, 8]. The Cyberspace is even described as a fifth military operational space in which movements by numerous countries are common [139].

Cybersecurity risk analysis

Risk analysis is a fundamental methodology to help manage cybersecurity [33]. With it, organizations can assess the risks affecting their assets and what security controls should be implemented to reduce the likelihood of such threats anord/or their impacts, in case they are produced. It is a practice with huge uncertainties [92] and, unlike other risky domains, it is difficult to obtain and analyse data [11], since organisations are reluctant to disclose data about intrusion attempts or consequences of cybersecurity incidents [18, 11, 113].

Numerous frameworks have been developed to screen cybersecurity risks and support resource allocation, including CRAMM [25], ISO 27005 [87], MAGERIT [122], EBIOS [2], SP 800-30 [127], or CORAS [111]. Similarly, sev-

[4] International Electrotechnical Commission

[5] National Institute of Standards and Technology (United States)

[6] International Society for Automation

eral compliance and control assessment frameworks, like ISO 27001 [86], Common Criteria [160], or CCM [31] 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 [35]: 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 [76] and Allodi an Massacci [5] provide additional views on the use of risk matrices in cybersecurity. Moreover, with counted exceptions like IS1 [131], 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.

Cybersecurity threats

The 'cyber' environment is populated with all kinds of cyber threats, some of them representing a large and highly skilled global menace [22]. They are also of different nature such as military, intelligence services, business rivals, hacktivists, lucrative hackers, or terrorists [15] – as well as combinations of them. This diversity of menaces could be classified according to their attitude, skill and time constraints [3], or by their ability to exploit, discover or, even, create vulnerabilities on the system [36].

The most formidable threats are the units maintained by global powers (mainly the US, China and Russia), with resources and skills beyond those of other actors; although they are constrained by the possible military, economical and political repercussion of their attacks [36]. Other threat sources that are closely related with social institutions or movements are hacktivists, a wide profile that could cover from hackers trying to prove their ability to hackers closely related with terrorist organizations. Hacktivists are also becoming more skilled, obtaining more resources and focusing their attention on OT [124]. Insiders are the ‘most formidable enablers’ of cyber threats and, indeed, the major source [21], but they are also the easier to handle through, e.g., a sound cybersecurity program. Profit-oriented cyber criminal groups are now mature and professional organizations, with some of them employing dozens of hackers and managing large financial resources [22], carrying out a wide range of targeted and non-targeted attacks [21]. Malware is usually developed with a goal-oriented or profit-maximizing behaviour [137] and, consequently, a sound way to face them is treating them as adversarial actors and counter-attacking them with behavioural approaches [137]. Finally, a concept has arisen to name the most sophisticated menaces: Advanced Persistent Threats (APT) [32], which are patiently orchestrated operations seeking to stay hidden while they consolidate their path for executing their final objective. Anomalous and apparently innocuous events are practically the only indicators of the presence of an APT in a system.

Operational technology

OT deals with *industrial processes*, which are a series of chemical, physical, mechanical or electrical steps taken to manufacture, control or distribute physical objects (e.g., products, infrastructures, facilities).

Industrial processes are subdivided in three basic types: Discrete processes involve assembling individual products from individual pieces and their handling (e.g., a car factory); Continuous processes involve the control of flows of materials from a starting to a finish point (e.g., a water transmission system); and, finally, batch processes involve a series of steps to transform an input of raw materials or intermediate products into finished products (e.g., a paint factory). Most industrial processes are hybrid and combine discrete, continuous and batch steps.

The environment of the industrial process is important. On one hand, the process will typically have dependencies with other ones: upstream processes that provide input materials; downstream processes that represent the customers; support processes (e.g., energy, waste); and, finally, safety systems that protect the process. On the other hand, the location of industrial processes involves single site facilities, distributed facilities in several sites and transmission facilities that could span even different countries.

Therefore, we can define *operational technology* (OT) as the digital systems that monitor, control and automate the behaviour of industrial processes. OT is a generic term that encompasses more specific definitions like industrial control systems (ICS), industrial control and monitoring systems, or industrial automation and control systems (IACS).

OT improves the mechanisation (ability of doing a physical process with machinery) and enables the automation (ability of a process to operate without human assistance) of industrial processes. The implementation of OT in industrial processes revolutionised multiple sectors, increased productivity and quality, reduced costs and allowed new or enhanced products and functions. Its use is also controversial: automation, like most technologies, may eliminate more jobs than it creates. Given the current economic and political system, it directly contributes to the displacement of labour towards capital (i.e, the technology owners); and, as a consequence, contributes to the concentration of wealth, income and power in fewer people and organisations.

Currently, OT systems are rarely isolated and their connections with other systems provide advantages. First, the information generated and managed by the OT system is sent to other business systems of the organisation to support decision-making. Second, some operations related with the industrial process are performed in remote systems: support activities, higher-level control or oversight, remote phases of the industrial process, or other industrial processes.

Another consequence is the possibility of locating some activities related with the operation in remote control centres. In this context, there is a trade-off between three elements: human engagement of the site personnel, automation of the equipment and remote operations. OT facilitates the substitution of on-site personnel engagement by remotely-operated or automated control activities. This helps to reduce and move away personnel from the industrial process, and substitute handling tasks with less riskier supervisory tasks.

The technology, services and requisites of OT usually differ from typical ICT ones. This is reflected in the most elementary difference between ICT and OT: the *industrial process data*, which is the digital data that contains the physical properties of the industrial process. In the context of OT cybersecurity, it is important to differentiate the informational input and output devices (e.g., keyboards, screens, printers) from the operational input and output (e.g., field devices). The critical nature of industrial processes requires strict performance requirements to data communication within the ICS such as time-criticality, deterministic behaviour and minimization of delays and signal jitter.

In general, these systems are designed to support specific industrial purposes and can have a lifespan of 15 or even 20 years. Their resources are very constrained to their designed functions, and changes should be carefully tested and implemented. OT typically consist of the following main components: *field devices* that measure and control equipment within the industrial process, *field controllers* that handle and integrate field devices with the rest of the OT system, *central systems* that provide additional storage, monitoring, control and automation functions, *human-machine interfaces* (HMI) that enable the interaction with human operators and, finally, ICT components adapted to the operational environment to support corporate functions in the field.

Regarding communications, OT implements similar topologies to ICT systems, usually with far less components but with more redundant communication channels between them. The physical implementation of communications can consist of leased lines (analogue or digital), dedicated lines, wired media, power lines, Wi-Fi, or radio. The main components of OT systems use separated communication channels: *field protocols* for the communication between field devices and controllers and *command & control protocols* for the communication between field controllers and central systems. For these communications, OT systems use both ICT and OT communication protocols. Industrial protocols were traditionally developed to support specific or proprietary technology but, nowadays, most of them follow open standards and some of them adaptations of ICT protocols. Relevant industrial protocols are Modbus, DNP 3.0, ICCP, OCP, Fieldbus and Profibus. Relevant ICT protocols in OT are Ethernet, serial (RS-232 and RS-485), IPv4, IPv6 and TCP.

There are several architectures in OT depending of the needs and requirements of the process, facility and sectors. The main architectures are SCADA, DCS and PCS. A *Supervisory Control and Data Acquisition system* (SCADA) is a centralised system that monitors and controls other specialised controllers from a master terminal unit. These systems are very flexible in design and implementation. They are typical of transmission and distribution facilities such as the electrical grid or pipelines. In a *Distributed Control System* (DCS), the core controllers are distributed throughout a facility, so the monitoring and controlling functions are also distributed hierarchically. These systems provide more reliability and integration but they are less flexible than SCADA ones. They are typical at refineries or chemical plants. *Process Control Systems* (PCS) have a similar architecture than SCADA but perform many of the functions of distributed systems.

This differentiation of OT with respect to ICT, in both the technology and the function, is the root cause of the need for a different cybersecurity approach in OT [130]. Indeed, many security countermeasures that provide a sound protection in an ICT environment might be ineffective or even counterproductive in an OT system.

Industrial cybersecurity

Industrial cybersecurity for OT is different from traditional ICT cybersecurity in several ways: The importance of having full control over the equipment and the industrial process, the potential operational and physical consequences of incidents in the system and the need for making decisions in real time to handle incidents as quickly as possible to avoid escalation [130]. The base of these differences are a series of root causes, wrong assumptions and strenghts. Translated to the information security principles, this means that OT prioritises availability over integrity and, in turn, integrity over confidentiality (in standard ICT the priority is, in general, the opposite, namely confidentiality over integrity, and the last over availability).

OT is used in critical environments supporting demanding industrial processes, in which the continuous availability of the system and its predictable behaviour are critical. OT systems lifetime is in the order of 15, 20 or even 30 years; thus, the security of their design will get outdated but also vulnerable to new attack tools [15].

Nowadays, OT has become more standardised and use more pervasively ICT components, facilitating malicious access or the exploitation of such systems. The use of open standards and widely used protocols (e.g., TCP/IP), software (e.g., Windows, C and SQL) and hardware (e.g., layer 2/3 routers) enables most of their vulnerabilities in operational systems [18, 22, 176]. OT specific protocols are also vulnerable [176]. OT is a complex software prone to security vulnerabilities derived form bugs [22, 36], untested features [36] and problems with frequent patching and updating [22]. It is also hard to set alerts and detect penetrations due to weak or absent defence systems [113]. However, OT cybersecurity was traditionally relaxed due to two – now outdated – assumptions: isolation and the difficulty of attacking.

Concerning isolation, in the past, OT security was based on the assumption that systems were isolated, and the elements within, trusted. Without external connectivity, organizations established what was to be trusted and what was not, and implemented physical security measures to avoid untrusted elements penetrating the system. However, the convergence between OT, ICT and networks rendered those assumptions less valid if not dangerously wrong. OT is now connected to external networks such as the organization business networks or remote operational centres [18, 22, 67] and interacts with an always-changing number of elements. There is no guarantee that adversaries do not access the system nor gain rights and privileges, since external networks have multiple entry points and users [18]. Therefore, organizations should implement a variety of security measures and operate under the assumption that any part of the system might become untrusted.

The second assumption is that hackers would have difficulty in manipulating OT technology. Industrial systems use very specific and precise processes that require proficiency in both the process and hacking. This makes it difficult for a hacker to manipulate the industrial process in a specific manner. However, being able to disrupt communications or data no longer requires an effort bigger than that required for an ICT system.

On the other hand, OT presents strengths when it comes to cybersecurity. OT generates more predictable and repeatable traffic than ICT systems, since they have a very static design over time and their tasks are more specific. Most OT systems also employ simpler network dynamics than enterprise systems (e.g., fixed topology, stable user population). Therefore deviations from normal operations are easier to detect than in standard ICT, which could be critical in detecting anomalies or cybersecurity events. OT also employs limited processes and applications to those strictly required to operate the industrial process. Fewer applications and processes reduce vulnerabilities and risk of failures. The determinism of the software can also facilitate the detection of anomalies and attacks. This fact facilitates the implementation of Intrusion Detection Systems (IDS).

There is still a need for calibrating our understanding of the consequences of cyber attacks in OT, due to the complexity of cyber-physical systems. Consequences could lead to loss of production, the inability to control a facility, multi-million financial losses, or even impact stock prices [18]. However, one of the key problems for understanding the consequences of a cyber attack throughout OT, and probably the most difficult one, is that OT systems are also Cyber-Physical Systems (CPS) encompassing both computational and physical elements. Some of the areas of OT and ICT cybersecurity could be addressed with similar approaches (e.g., attackers, organizational cybersecurity, or vulnerabilities of hardware, software and networks), but such cyber-physical part is complex, and requires skills and capabilities different from Business Administration and Computer Security Engineering. Estimating the consequences is also difficult [14, 164] due to the lack of disclosure, difficulty of assigning consequences and costs, difficulty of valuing intangible or indirect costs, mis-estimating of costs, incommensurability of consequences, uncertainty, absence or ambiguity of information and multiple interests, criteria or biases.

Industrial cybersecurity threats

Cyber attacks are the continuation of physical attacks by digital means. They are less risky, cheaper, easier to replicate and coordinate, not constrained by distance [22, 21] and they could be oriented to cause high-impact consequences. Deception attacks manipulate information within the OT network trying to compromise the system and causing physical, operational, or information consequences [22, 67]. They include spoofing, Denial of Service, writing field controllers or central systems, control message modification, or even physical damage through altering the variables of a control system commanding some physical operation [21]. Some attackers are also capable of creating vulnerabilities inside an OT or ICT component during its development – what is called a supply-side attack [36]. In general, typical threats to sensor networks and conventional ICT systems are also threats to OT if the attacker has the means to exploit the additional vulnerabilities of an OT system [176]. These attacks can be carried out using sophisticated intrusion tools and methods such as dialling, scanning, traffic sniffing and password cracking [113], or zero-day exploits and rootkits [32]. It is also difficult to measure data related to attacks such as their rate and severity, or the cost of recovery [92].

All types of cyber attackers are becoming aware of the possibilities of OT and targeting these systems. From hacktivists to nation-states and criminals [18, 124]. Major incidents in critical systems are unlikely but possible: around 7% of cyber attacks to critical infrastructure penetrate into critical systems [126].

Relevant cases of actual cyber attacks include the 2010 Stuxnet cyber attack against an Iranian nuclear plant [15, 32], the 2012 Shamoon attack that disabled 30.000 computers of Saudi Aramco – the Saudi national oil and gas company [15, 32]; the attack on a German steelworks in 2014 [108], the attacks on a Brazilian electrical grid in 2007 and on an Australian sewage control [22]; or the attacks on a Venezuelan harbour's tanker loading systems 2007, on a Polish tram traffic system in 2008 and on a Russian natural gas pipeline regulator system in 2000 [26]. In addition, tests and experiments [13], such as the ones in the Idaho National Laboratory [15] or the US Department of Defense [36], also demonstrate how practicable penetrating OT and cause physical damage is. Sophisticated high-impact attacks on typical ICT systems also manifest state-of-the-art cyber operations, such as attacks for obtaining top secret information from governments and multinationals, compromising people, or gaining information about ICT vulnerabilities. For instance [32], the 2007 Operation Aurora attacks against Google, or the 2011 Dragonfly/Energetic Bear against European and US energy firms. Non targeted attacks could also be a problem, such as the Slammer worm infiltration in an US nuclear plant [18] or the case of a water controller sending spam [23].

The strategic relevance of cybersecurity

The digitalization of our economy and society has introduced big-scale cybersecurity problems. Furthermore, attacks, espionage, insiders and privacy breaches continue to increase in frequency, impact and sophistication [22]. The cybersecurity industry estimates an impact on the global economy for 2017 at more than $ 600 billion in annual cost, or around 0.8% of the global GDP (in comparison, drug trade represents 0.9% and international crime, 1.2%) [118]. Indeed, there are well-functioning "cyber black markets" [177, 74] that exchange attack tools and valuable information to perpetrate attacks. Those buoyant markets provide incentives for skilled people to develop new hacking products.

Political, economic and military leaders are raising their concerns on cyber attacks and the serious menace they pose to critical infrastructure, privacy and intellectual property. They are declaring, more and more, that cybersecurity risks are a geo-strategic menace that can lead to serious incidents in critical infrastructure or to the espionage of valuable assets. Governments, in coordination with companies and operators, are developing strategies and guidelines to improve critical infrastructure cybersecurity and prevent the increasing impact of attacks on the economy and society.

On the other hand, and although most of the attacks come from criminal organizations, governments are increasingly involved in attacks - directly and indirectly. They are actively developing highly sophisticated cyber offensive capabilities, sponsoring groups of hackers that target rival foreign countries and companies, intervening in software and hardware design to guarantee them back doors that can later be exploited for malicious intent and spreading out massive misinformation and propaganda through a combination of cyber attacks and social media manipulation [4].

State-led cyber attacks have become a serious concern that has raised privacy and human rights concerns: rival – and even friendly – countries attacking each other, high-developed countries being object of less developed countries that seek the exfiltration of industrial designs, or the massive surveillance of the Internet exerted by some countries. The major cyber powers are the US, Russia and China but most of the highly-developed countries (e.g., the UK, South Korea, Germany, Israel) as well as regional powers (e.g. Iran, North Korea) have prominent capabilities and strict measures for controlling its part of the Internet (including citizens) and their enemies. These conflicts are asymmetric in that attacking usually needs an amount of resources that represent a fraction of the ones needed by the defenders. In addition, the counter-attack effort needs that the counter-attacked country or organization were as vulnerable as the counter-attacker one, which is not generally the case (e.g., less rule of law, less dependence on IT infrastructures and an international policy focused on unilateral exertion of power rather than multilateral cooperation). In addition, identifying the attribution of attacks is difficult and therefore it is hard to point to a state or organization.

Cybersecurity has become a top priority in defence during this decade, although some experts criticize an excessive hype about the potential disruptive capability of big-scale cyber attacks. That criticism is true: the worst cybersecurity scenario would pale in comparison to a very limited nuclear or biological warfare scenario, or – when it comes to the digital infrastructure – to the potential disruption caused by a powerful solar storm such as the one in 1859. However, the point of prioritising cybersecurity is that the investment and efforts in this field are critical to shape the global economy, society and rule of law, as we are becoming more and more dependent on computers, the Internet and the opportunities they bring to us.

Protection of critical cyber infrastructures

The importance of securing the digital assets pervasive in critical infrastructures cannot be underestimated. A disruption in their operation or a cyber attack could lead to injury to workers or the public, loss of revenue or production, legal consequences, loss of public trust, physical damage to the facilities, or harm to the environment.

Indeed, regulators identify critical infrastructure as both physical and cyber. A *critical infrastructure* is an asset or system that is essential for the maintenance of health, safety, security, or the economic and social well-being of people. Its disruption or destruction would have a significant impact on society and, therefore, governments have defined the critical infrastructure they aim to protect and established regulations and protocols for their protection. The European Programme for Critical Infrastructure Protection

(EPCIP) [49] and the Directive on European Critical Infrastructures [52] establish activities and procedures for identifying critical infrastructures and improve their protection. The US Presidential Policy Directive 21 (PPD-21) Critical Infrastructure Security and Resilience [167] identifies sixteen critical infrastructures and key resource sectors that are strategic to the United States.

Regulatory compliance is challenging: organizations have to meet international, national and regional regulatory systems. In the case of cybersecurity, this requires looking at a myriad of rules that govern safety, privacy, security and cybersecurity consequences. This can involve dozens of regulations and frameworks, which are continuously evolving over time – especially in cybersecurity.

The European Union (EU) established the European Network and Information Security Agency (ENISA, 2004) to coordinate the cybersecurity efforts of the 28 member states and their computer emergency response teams (CERTs) and provide training and support in cybersecurity (including offensive capabilities through the European Defence Agency). In addition, the EU established, within Europol, its European Cybercrime Centre (EC3, 2004) for sharing cyber crime information between law enforcement agencies. The main cybersecurity policy in the EU is the European Cyber Security Strategy (2003) [50] that established the goals of (1) achieving cyber resilience, (2) reducing cyber crime, (3) developing cyber defence policy and capabilities, (4) developing the industrial and technological resources for cybersecurity and (5) establishing a coherent international cyberspace and promoting EU values. An important aspect is the European coordination effort, through which the Strategy seeks that each member state (1) designates a sole national agency as coordination for cybersecurity policy and operations, (2) has a sole CERT that would act as the operational centre in case of a major cyber incident and (3) ratifies the 2002 Budapest Convention as the baseline for combating cyber crime. The Strategy is complemented by several legislations and policies, of which the Network and Information Security Directive (NIS, 2014) [53] aims at operationalising the goals of the strategy by establishing minimum cybersecurity standards.

Cybersecurity lends itself to the cases for incident, multi-objective and adversarial risk analysis

Assessing risk is relevant during cybersecurity incidents. Cyber attacks or other cybersecurity incidents are usually comprised of a chain of steps. The earlier signs of one of these events are suspicious anomalies that could be also caused by legit user or system actions. Here, risk analysis focuses on identifying what could have caused the anomalous incidence and what potential ramifications of events might follow. For instance, a new connection within the network could be caused by the connection of a maintenance laptop or by an unauthorised party accessing the network. Additionally, if a specific attack or problem has been identified, then the risk analysis would focus on further identifying the consequences of the threat, how likely they are, or how the potential countermeasures would change the risk. An example could be the presence of malware in a computer or controller. The incident handlers should analyse aspects such as whether the malware is harmful for the actual business operation or the current industrial process, whether the malware can spread to other devices and the consequences of this spreading, or what the consequences of removing the malware or changing the device might be.

The cases for multiple risk objective comparison and adversarial modelling of threats is a good fit for cybersecurity for multiple reasons. One of them is the oversimplification of objectives is something more explicit in industrial cybersecurity than in traditional cybersecurity, as risks involve safety and environment consequences. Another one is that cybersecurity threats (even the malware programs themselves) would be better modelled as adversarial threats.

The case for including cyber insurance in risk analysis

Generally speaking, there are four types of options against a risk: accept the risk as it is, mitigate it either preventively or reactively, avoid it completely or transfer it to another party. Insurance belongs to the last category and is widely used in multiple domains: e.g., car, financial or health insurance. Usually, the focus in risk analysis is on modelling mitigating options but it is possible to model transfer/insurance options as well.

In recent years, insurance companies have been introducing cyber insurance products for small and large organisations [116]. It is expected that the number of products and adoption of cyber insurance will increase in the coming years. However, it still has to take off [110], specially in Europe. Insurance presents a series of particularities that needs to be dealt explicitly in the risk analysis models for both the insuree (e.g., how the insurance depends on the insuree security features) and the insurance company (e.g., designing cyber insurance products for different types of organisations).

1.4 Research objectives

This Thesis focuses on four general objectives as stated in Table 1.1. They are derived from the cases we discussed in the previous sections: risk analysis during incidents, adversarial risk analysis, multi-objective risk analysis and cyber insurance. We found that these aspects are not well covered by traditional cybersecurity risk analysis and, thus, our aim is to bring models that address them.

Table 1.1: Research goals

Goal	Description
RO1	Develop a risk analysis model for cybersecurity incidents
RO2	Integrate the analysis of adversarial threats into cybersecurity risk analysis
RO3	Integrate multi-objective decision-making into cybersecurity risk analysis
RO4	Integrate the analysis of risk transfer/insurance into cybersecurity risk analysis

1.5 Research methodology

The basis of our research starts with a thorough revision of the academic state of the art and current business practices in the different domains we address (e.g., risk analysis, cybersecurity). Then we proceed to develod decision models tailored to risk analysis and cybersecurity.

Decision models, like the ones we aim to bring, help people make decisions about problems. They are usually specified as models that help the selection of alternatives, and they are implemented as software or human procedures. The purpose is not to automate decisions but to support decision-making. Indeed, the construction of decision models focuses on the following aspects:

- Specification of how the model generates valuable and reliable information for users to make the decision, i.e., the internal logic. In the case of our

Thesis, this means that the model should provide valuable risk indicators for those managing risks.

- Specification of the type of information to be provided to the model, so that it generates reliable information, i.e., the external logic. In the case of our Thesis, this means that the model should be capable of integrating relevant information for risk analysis about the potential incidents and the affected system.
- Tailoring the model so users focus their attention on understanding the problem rather than the model, i.e., user friendliness. In the case of our Thesis, this involves two aspects. First, the model should facilitate the user in the insertion of data about the risk. Second, the model should provide accurate and easy to understand risk indicators and facilitate the comparison between response alternatives.
- Providing a general model of the decision problem so it is adaptable to multiple cases, i.e., to develop general conceptual models.
- Developing models that, to an extent, facilitate their implementation. Specifically, by detailing algorithms, calculations and assessment procedures that facilitate the reproduction of a risk problem in the models.

1.6 Outline of the results

The next two chapters focus on incident risk analysis. In chapter 2, we present our general incident risk analysis model (GIRA), which formalises the incident risk analysis process through an influence diagram. First, we discuss the considerations that should be taken into account regarding risk analysis when applied to incidents. As a basis for GIRA, we characterise the key elements of an incident and their relations. Then, we introduce GIRA and the particularities of its main components, accompanied with examples: threat exposure, incident response, incident materialisation, consequences in the systems, impacts on assets, risk objectives and risk evaluation. We also briefly discuss the mathematical representation of GIRA and additional extensions: simplified, for multiple agents and for immediate and non-immediate events.

In chapter 3, we present further advances for GIRA and a version adapted for a fast cybersecurity risk analysis (called CSIRA). First, we present a simple elicitation method based on the oddness of the event (i.e., on whether the different events in a chain of events are certain, possible, rare or impossible). Additionally, we introduce a category map for understanding the potential ramifications of cybersecurity incidents that might help when brainstorming about the risks of cybersecurity incidents. We then present CSIRA which is, basically, GIRA using the oddness elicitation method and the map for understanding the ramifications of cybersecurity incidents. In the presentation of CSIRA, we also discuss that decision makers only need to compare the scenarios of their different responses (without preference elicitation typical of influence diagrams).

The rest of the chapters focus on cybersecurity risk analysis for the traditional time frame of a number of years or the lifetime of an activity or system. Chapter 4 presents a comprehensive framework for cybersecurity risk analysis, covering adversarial and non-intentional threats and the use of insurance as part of risk management decisions. The first part is devoted to introducing influence diagrams that describe different risk analysis models and their mathematical formulation. Starting from a simple system performance evaluation we introduce, incrementally, new elements to the models (risk, risk mitigation, risk transfer and adversarial analysis). The second part reproduces a full example case in which we detail all the aspects of the assessment: The description and the structure of the risk problem, the assessment of the organisation beliefs about the elements affecting risk and their preferences, the modelling of the attacker problem to forecast his actions and the calculation of the best portfolio of security controls and insurance for the organisation.

In chapter 5, we describe a tree of cybersecurity objectives. Its purpose is to facilitate a comprehensive identification of the organisational objectives at risk. In this context, it is important to distinguish between those objectives that can be measured in monetary terms and those that cannot [or shouldn't], such as harm to people. We further explore how to measure those non-monetary objectives (e.g., reputation, personal rights, environmental damage). We conclude the chapter by detailing how to use these cybersecurity objectives and attributes with an utility function.

Finally, in chapter 6, we present several risk analysis models in the context of cyber insurance and for the insurance companies. In the first one, the insurance company has to decide what type of reinsurance product requires taking into account the different market segments the company is insuring (e.g., SMEs, large business). In the second, the company is deciding whether they grant or not an insurance product to a potential customer.

The Thesis concludes with a detailed discussions of our results in Chapter 7, where we also provide topics for future research.

The research has produced the following four academic publications, for the moment:

1. Rios Insua, D., Couce-Vieira, A., Rubio, J.A., Pieters, W., Labunets, K., and Rasines, D.G. "An Adversarial Risk Analysis Framework for Cybersecurity." In *Risk Analysis*.[7]

2. Couce-Vieira, A., Rios Insua, D., and Houmb, S.H. (2019) "GIRA: A General Model for Incident Risk Analysis." In *Journal of Risk Research*, Vol. 22, No. 2, pp. 191–208.[8]

3. Couce-Vieira, A., Houmb, S.H., and Rios Insua, D. (2018) "CSIRA: A Method for Analysing the Risk of Cybersecurity Incidents." In *Proc. of the 4th International Workshop on Graphical Models for Security*, LNCS Vol. 10744, pp. 57–74. Springer-Verlag.[9]

4. Rios Insua, D., Couce-Vieira, A., and Kreshnik, M. (2018) "Some Risk Analysis Problems in Cyber Insurance Economics." In *Estudios de Economía Aplicada*, Vol. 36-1, pp. 181–194.[10]

We also produced technical reports for the CYBECO project that detail, in more general terms, the contents addressed in chapters 4 to 6. Specifically, *D3.1: Modelling framework for cyber risk management*[11] and *D3.2: Improved modelling framework for cyber risk management*.

[7] Published online 10/Jun/2019, doi:10.1111/risa.13331
[8] Published online 11/Sep/2017, doi:10.1080/13669877.2017.1372509
[9] doi:10.1007/978-3-319-74860-3_4
[10] www.revista-eea.net/volumen.php?Id=99&vol=36&ref=1 [Acc. 30/05/2019]
[11] Available at www.cybeco.eu/results []Acc. 25/05/2019]

Chapter 2

GIRA: A general model for incident risk analysis

This chapter presents a general model (named GIRA), which formalises the incident risk analysis process through an influence diagram. Our aim is to provide a decision support model that generates reliable risk information and enhances incident risk evaluation. First, (Sect. 2.1) we introduce the existing methodologies suitable for incident risk analysis and some considerations regarding the analysis of risks during incidents. Next, we present incident risk analysis as a series of reasoning steps (Sect. 2.2) and briefly introduce influence diagrams (Sect. 2.3). Finally, we describe GIRA (Sect. 2.4 to 2.6) and discuss it (Sect. 2.7).

2.1 Introduction to incidents and risk analysis

Incidents and risks are closely related. In a general sense, a risk is nothing but a potential incident, and an incident is nothing but a materialised risk. Risk analysis, or risk assessment in ISO terms [83], aims at identifying what negative scenarios could happen, what are their likelihoods, how we can evaluate them, and what we can do to mitigate or stop them. Generally, it will address these questions for a relatively long horizon (e.g., several years or the life of a project or installation). However, these issues are also relevant during incidents, as they typically happen in a context of multiple uncertainties [9], complexity [153] and pressure, given the need to provide fast responses. When an event indicative of an upcoming incident is detected, incident handlers might be interested in identifying what events could happen next as well as their likelihood. Indeed, uncertainty might also exist within the identification itself. Incident risk analysis can be reduced to three types of approaches: upstream, downstream, and combined methods.

Upstream methods identify causing or enabling events, depicting imminent risks and possible prevention points. The most common methods are fault trees, attack trees, and probabilistic attack graphs. A *Fault Tree Analysis*, typical of safety risk analysis [47, 81, 30], decomposes a high-level problem or failure into specific events. *Threat Trees* [6] and *Attack Trees* [150] are the counterparts for analysing security threats and attacks [147]. Other relevant upstream methods include *Root Cause Analysis* [97] and *Probabilistic Attack Graphs* [154].

Downstream methods identify cascading consequences and aftermath impacts. Popular approaches are FMECA, HAZOP, and structured what-if techniques. *Failure Mode and Effects Analysis* (FMEA) and its extension *Failure Mode, Effects, and Criticality Analysis* (FMECA) [30, 39] identify the potential failures of particular system components. *Structured What If Techniques* (SWIFT) are structured brainstorming methods for quick risk identification [20]. Additional methodologies include *Hazard and Operability Studies* (HAZOP) [104], *Activity Hazard Analysis* [168], *Event Tree Analysis* [30], and *Reliability Block Diagrams* [30].

Finally, *combined methods* implement upstream and downstream analysis. The most widely used are risk matrices and bow-ties. A *Bow-Tie Analysis* [80, 82] combines a top-down tree for the events that cause an incident and a bottom-up tree for the events the incident triggers. It is a robust method that can integrate others such as attack or fault trees, or root cause analysis [158]. *Risk Matrices* assign an ordinal value to the likelihood and the severity of the risk, and then derive an ordinal risk rating from both values. They are easy to use and provide a simple and memorable way for communicating risks. However, Cox [35] identifies several limitations including a poor resolution that assigns same ratings to very different risks or high ratings to smaller risks, ambiguous inputs and outputs, and suboptimal resource allocation. Other interesting methods are *Failure-Attack-Countermeasure Analysis* (FACT), which combines security attack trees with safety fault trees [147], and CIA-ISM [138], which combines *Cross Impact Analysis* [68] with *Interpretive Structural Modelling* to generate a matrix that represents a cause-effect network between interrelated events.

The following considerations should be taken into account regarding the above risk analysis methods when referring to incidents.

1. *The little attention that many of the existing methods pay to risk evaluation*. Risk analysis consists of two activities. First, a description of the system's risks. Second, an evaluation of these risks from the stakeholders' perspective. Both differ in nature: *Risk description* (risk analysis in ISO terms) is an objective activity (i.e., what is?), whereas *risk evaluation* is a subjective and

normative activity (i.e., what ought be?), in which stakeholders evaluate risks with respect to their motivations, preferences, risk attitudes, and given target/shareholder values. Most of the methods concentrate on risk description, whereas risk evaluation and its subjective nature receive little attention – as for instance, in cybersecurity [28].

2. *The inadequacy of simple risk scoring*, as in risk matrices, which do not approximate the likelihood adequately. The main component of the traditional likelihood is the threat being present, but this is meaningless if we do not take into account that the likelihood might also depend on several triggering events and mitigating controls, and that the likelihood of the potential impacts also depends frequently on a series of cascading consequences [61].

3. *The difficulties of eliciting likelihoods*, which involves subjective estimation and information that may not be available nor measurable [158]. It is important to have a methodology that generates consistent, inexpensive, and contextualised metrics [94], and implemented in an automatic manner to increase the reliability of the data [154, 159, 27].

4. *The potential presence of adversarial threats*. Traditionally, threat characterisation is based on historical data or on an evaluation of how frequently the threat targets the system. This is mostly irrelevant during an incident, especially in security, as the system has already been targeted. Consequently, intelligent threats should be modelled as adversaries, as in *Adversarial Risk Analysis* [145], which deals with the defender-attacker interaction and how attackers behave to reach their objectives.

5. *The difficulty of comparing risks, given the trade-off between commensurability and comparability of objectives*. Espinoza [48] identifies as incommensurable risks those that 'cannot, or ought not, be accurately compared and can thus neither be weighed against their associated benefits nor be ranked along a single severity scale'. For instance, people often feel that it is unethical to assign monetary prices to risks imposed upon humans or the environment [114]. Another problem during risk evaluation is that it is recommendable to limit the number of objectives to be compared (e.g. two or three). The reason is minimising task [136], information [44], and choice [149] overload to facilitate decision-making. Those two problems together create the trade-off: A single objective (e.g., impact levels of risk matrix) is easy to interpret, but measuring all impacts in the same scale involves a high level of incommensurability. On the other hand, ranking risks in multiple severity scales reduces incommensurability but becomes more difficult to compare. Considering this trade-off, an example of a re-

duced number of objectives could be the triad safety, monetary, and ethical/legal compliance.

We have found that existing methods suitable for incident risk analysis are not comprehensive enough for overcoming the above considerations. Therefore, we propose, in the rest of the chapter, a model based on an influence diagram for generating relevant risk information during incidents.

2.2 Reasoning about the risk of incidents

This section presents incident risk analysis as a series of reasoning steps involving different participating elements. It also describes the use of expiration times to capture the dynamics of incidents in a simple manner. This characterisation of our universe of discourse is inherited by GIRA (Sect. 2.4).

An incident involves different systems that might be distinguished according to two general types. First, the *managed system* (MS) is the system that incident handlers are in charge of protection. Second, the *dependent systems* (DSs) are the other systems such that an incident in the MS could trigger additional incidents or consequences.

The incident may be depicted (Fig. 2.1) using the following steps:

Step 1. *Threat exposure*: A threat is present in the MS or its environment. A *threat* is any element, including risks from other systems, with the possibility of inducing an incident in, or through, the MS.

Step 2. *Incident materialisation*: The threat induces an incident enabled by the vulnerabilities of the MS. A *vulnerability* is any element of a system that makes possible an incident. Additionally, an incident might trigger new vulnerabilities or threats. For pracetical reasons, these would be modelled as new threats.

Step 3. *Incident response*: Incident handlers might be able to implement actions that stop or mitigate the incident.

Step 4. *Consequences in the MS*: The incident involves a series of undesired changes in the MS. These might escalate to incidents and consequences in the DSs. However, in this case, we propose to model them as impacts (from the incident handler perspective).

Step 5. *Impacts on assets*: The consequences affect assets contained or linked to the MS and DSs. A *stakeholder* is a party affected by the incident. An *asset* is any element affected by the incident and valuable to the involved stake-

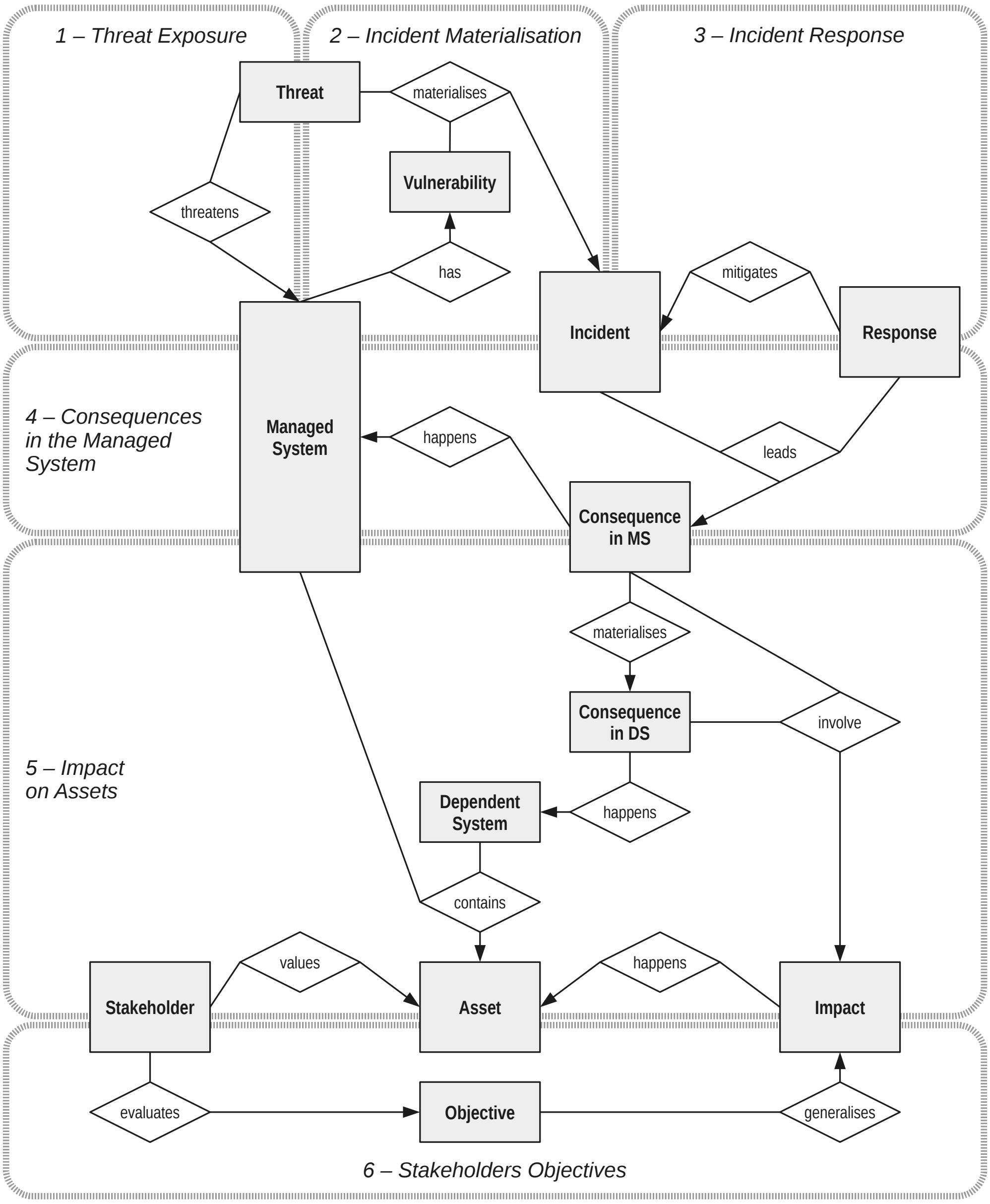


Fig. 2.1: Diagram with the different steps of an incident (dotted boxes), its components (rectangles), and their relations (diamonds and arrows). E.g., we can interpret that the "step 2 - incident materialisation" consists of "threats and vulnerabilities materialising in incidents" and "managed systems having vulnerabilities".

holders. The goal of the incident handlers is to protect the assets. The distinction between consequence and impact might be tricky but the following thumb rule could help: Consequences are relevant changes in the MS and in the domain of the incident handler[1], and impacts are relevant changes in the assets of the MS or DSs and from the perspective of all stakeholders (e.g., new risks, degradation of assets or operations, injuries).

Step 6. *Stakeholders objectives*: Reduced number of categories that generalise the multiple types of impacts, to facilitate stakeholders understanding and comparing the outcome of incidents.

2.2.1 The dynamics of incidents

A simple approach for capturing the temporal and dynamic factors of incidents consists of assigning an expiration time for each of the incident steps. Under this paradigm, analysts could harmonise likelihoods, computed through probability theory [19], to the time frame until the earliest expiration time. Specifically, the analysis would be valid until one of the following changes happens:

1. A change in the threat or the vulnerabilities of the MS would modify current conditions of threat exposure. This change might imply an additional change in the likelihood or the type of threat exposure, or even changes in the subsequent events (i.e., incident materialisation and further).
2. A change in the conditions of an asset would change the potential impacts the asset could suffer or the likelihood of these impacts.
3. The actual materialisation of the incident.
4. Expiration time of the current analysis set by the analyst.

Another dynamic factor is the distinction between immediate and eventual consequences which might be useful for incidents composed of phases (e.g., security kill chains or cascading incidents). *Immediate consequences* are those directly caused by the phase currently threatening the MS (i.e., phases still not materialised). *Eventual consequences* are those caused by the complete materialisation of the compound incident (e.g., an attacker capable of fully executing its kill chain). Immediate and eventual consequences do not affect the expiration of the analysis, since they always happen after the materialisation of the incident. However, it would be practical to carry out two

[1] For instance, in cybersecurity, following the McCumber Cube [119], we can express consequences as changes in the availability, integrity, or confidentiality of data.

analysis, one for the immediate consequences and another for the eventual ones. For instance, to determine whether a countermeasure should be implemented immediately.

2.3 Using influence diagrams to model risk analysis

Influence diagrams [75] provide a formal yet understandable description of decision problems under uncertainty. They can be generalised as Bayesian networks with utility nodes that allow probabilistic inference combined with multi-objective optimisation (typically, maximising expected utility). Modelling risk analysis through influence diagrams provides several advantages [61]. First, they integrate decision-making within risk analysis. Second, their mathematical foundations for quantitative analysis. Third, they facilitate understanding relevant cause-effect relations. And fourth, they are highly suitable for sensitivity analysis, parameter learning, and what-if analysis [132].

Influence diagrams are represented by nodes and arcs (see Fig. 2.2). A node consists of a disjoint and exhaustive set of elements that represent outcome events or states. A *decision node* (rectangle) represents a set of actions that decision-makers can take (i.e., 'what can we do?'). An *uncertainty node* (oval) represents a set of uncertain states relevant in the decision problem (i.e., 'what could happen?'). A *deterministic node* (double-lined oval) represents a set of certain states relevant in the decision problem (i.e., 'what would happen?'). Finally, a *value node* (hexagon) represents a set of preferences over the outcomes of a node (i.e., 'how we value what could happen?'). Arcs represent conditional relations between nodes (i.e., 'if this happens in the antecedent, then that happens in the consequent'). A *functional arc* indicates that a value node is a function of its antecedent nodes. A *conditional arc* indicates that an uncertain or deterministic node is probabilistically conditioned by its antecedent nodes. Finally, an *informational arc* indicates that a decision node is informed by the outcomes of its antecedent nodes.

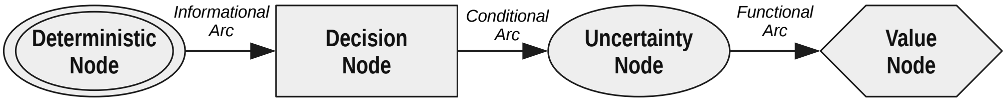


Fig. 2.2: Elements of influence diagrams

2.4 GIRA: A general model for incident risk analysis

In this section, we introduce the GIRA as an influence diagram that formalises the incident risk analysis process described in Sect. 2.2 as a decision support model designed to generate relevant risk indicators for an incident that the analysts believe has happened, is happening, or is going to happen at a specific point in time.

We build GIRA based on the following four assumptions.

1. *The threat is identifiable*. We assume analysts can detect and identify the presence of a threat. Relaxing this assumption would involve changes in likelihoods or the use of more generic threats (e.g., system malfunctions).
2. *The system faces a single threat at a time*. The model takes place in a context in which the system is facing a single threat, which could be a phase of a compound incident. The combination of this assumption with the previous one implies that analysts might be uncertain about which specific threat they are facing, but they know they are facing a single threat. If we relax this assumption, we can force a composition of incidents ordered them from the most to the least immediate.
3. *The model is compatible with quantitative and qualitative likelihoods*. There is some controversy between quantitative and qualitative risk analysis. Qualitative analysis are easier to implement, and they have a wider adoption. However, they are also ambiguous and prone to errors [34]. Quantitative measurement is more accurate and helps to reduce ambiguity. However, it still presents some shortages concerning empirical validation, and its adoption is low [171]. As an influence diagram, GIRA is quantitative in nature, so that the implementation of quantitative likelihoods is facilitated. In case analysts want to use qualitative risk, there should be a procedure for quantising their risk scoring method (i.e., using GIRA for semi-quantitative risk analysis).
4. *The model has a general value node*. Influence diagrams tend to operate with expected utilities. Utility is a subjective measure of preference and risk attitude that may model descriptive, normative, and prescriptive mechanisms of choice. Expected utility incorporates the uncertainty factor. The most important controversies of these theories are in the normative and descriptive realms. However, even as a prescriptive method, utilities might present some weakness due to elicitation biases or the violation of theoretical axioms, which are usually solvable but might require the support of a decision analyst. Therefore, the value node of GIRA represents how the stakeholders order their preferences, which should be specified

by a more detailed evaluation sub-model with a multi-criteria decision method [41] such as *Multi-Attribute Utility Theory* (MAUT) [102].

2.4.1 GIRA fundamental model

Fig. 2.3 depicts the *GIRA Fundamental Model*, through the influence diagram that captures the risk analysis process for the entire incident chain. Stacked nodes represent that, for certain node types, there could be several of them. For instance, the materialisation of an incident could lead to several simultaneous consequences and, thus, it is necessary to create one node per consequence type. Additionally, lighter nodes represent the risk description part of the incident risk analysis, whereas darker grey represents the risk evaluation part. Having said this, our model can be split into simplified, but interlinked, sub-models, as proposed in Sect. 2.6. The rest of Sect. 2.4 describes the particularities of all nodes.

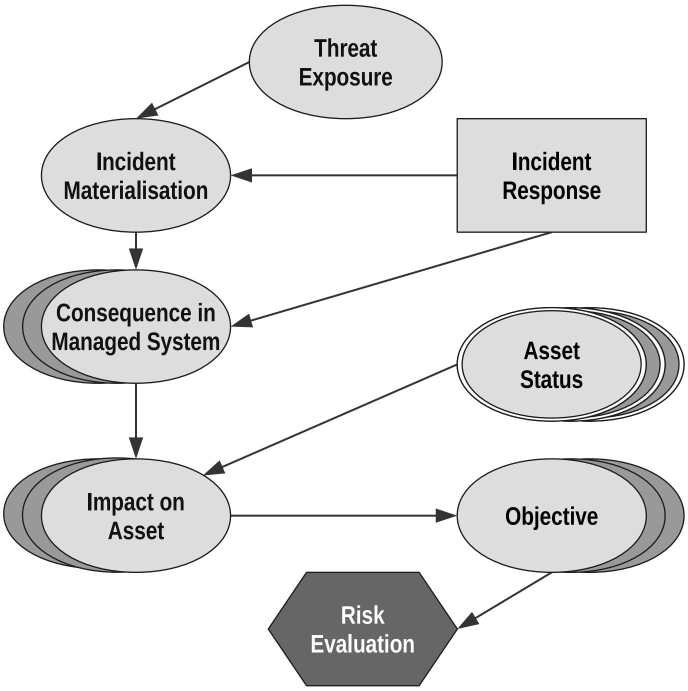


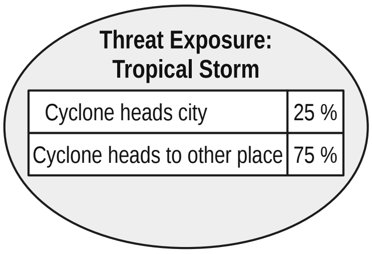
Fig. 2.3: GIRA Fundamental Model.

Threat exposure node

The *Threat Exposure* node (TE) contains the likelihood that the MS is exposed to a threat. The states of this node represent the candidate threats: the analysts might be uncertain amongst a set of potential candidates, including that no threat occurs at all. The likelihoods in this node reflect how likely it is that a candidate threat will actually happen.

A sample question for eliciting the states of the TE node could be: *What is the specific threat that the system is facing right now?* The question for eliciting the likelihoods of the TE node could be: *What is the likelihood that the system is currently being exposed to such a threat?* The most important elements for eliciting threat exposures are the threat characteristics (e.g., motivation, capability, and opportunity in case of security threats) and the vulnerabilities of the MS. Finally, it is important to set the expiration time for this elicitation (i.e., *for how long is the previous elicitation valid?*), which corresponds to the moment in which any of the states or likelihoods could change. Example 2.1 illustrates a TE node case.

Example 2.1. A person receives the news that a tropical storm is forming, and might become a cyclone and head to her city. Based on the news, the person estimates that there is 25% chance that there is a cyclone that hits her city (and 75% otherwise), as illustrated in Fig. 2.4. The expiration time of this node is the next update of the weather report (e.g., 6 hours), which might cause changes in the previous probabilities.



Cyclone heads city	25 %
Cyclone heads to other place	75 %

Fig. 2.4: TE node for Example 2.1.

Incident response node

The *Incident Response* node (IR) represents the actions that incident handlers could implement to avoid or mitigate the incidents that the threats of the TE node could induce in the MS. The states of this node represent disjoint actions. Therefore, this node must include all relevant possible combinations of actions, including doing nothing.

For instance, in the case of an increase in the terrorism threat level in an airport, the responses could be two, say, increasing security checks and increasing the presence of security personnel. However, considering the inaction case and that the two responses can be combined, this IR node would have four states.

A sample question for eliciting the IR node could be: *Assuming that such a threat is present, which are the responses we can implement to address the risk?* Finally, the expiration time for this node corresponds to the moment in which any of the responses could change (i.e., *for how long are the previous response actions valid?*).

Example 2.2. As seen in Fig. 2.5, the person has two options, stay at home or leave to her family house in a town safe from the cyclone. In this case, the expiration time is irrelevant, as the person can leave her home whenever she wants.

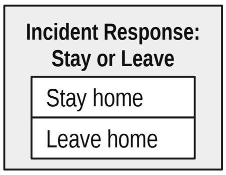


Fig. 2.5: IR node for Example 2.2. Its precedent node is displayed in white.

Incident materialisation node

The *Incident Materialisation* node (IM) provides the likelihood that the TE node threats materialise as an incident in the MS, taking into account the IR node response. The states of this node represent the potential incidents as materialisations of threats.

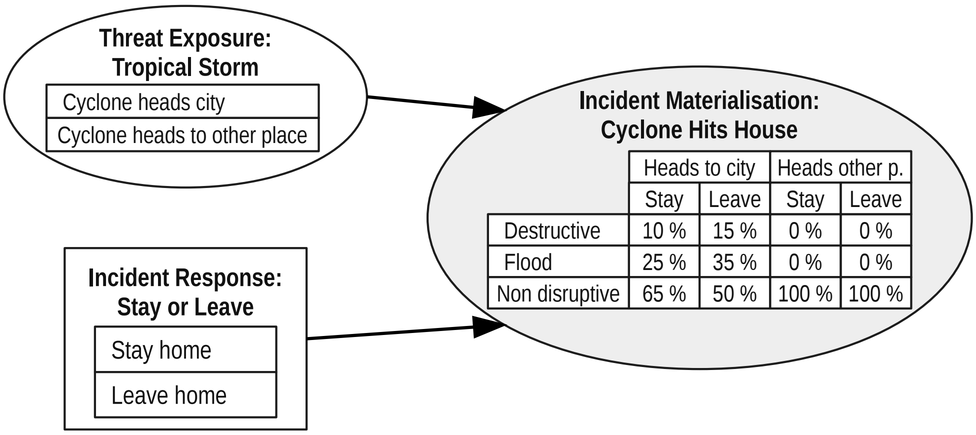
This node might have several states for two reasons. First, there could be several candidate threats in the TE node, each with its associated materialisations. In this case, the scenarios are disjoint because their parent threats would be disjoint (e.g., an incident materialisation 'theft' could not exist, should the threat exposure 'thief' not exist). Second, each candidate threat could have multiple materialisations. Then, this node must represent all relevant possible combinations of materialisations.

A case is when the threat scenario involves a behaviour such as a linear kill chain. It is easy to model each of the steps through disjoint states, in which state n represents that the threat materialised its n-th step. A second case is when the threat scenario involves actions that can be simultaneous. In this case, it will be necessary to make them disjoint. For example, a virus attack consists of 'step 1 – delivering the virus to system element', 'step 2A – exploiting that element' and, finally, 'step 2B – expanding to other nodes'. Steps 2A and 2B could be simultaneous. Therefore, it will be necessary to create four events: a state for step 1, a state for step 2A but no step 2B, a state for step 2B but no step 2A, and a state for simultaneous step 2A and 2B. More generally, if we have n potentially simultaneous events, we would have to create 2^n states. Alternatively, it is possible to model these events by adding threats in the TE node, create another IM node, or modelling them as consequence nodes.

The IM node is preceded by the TE node, which means that the threat exposure affects the likelihood of the incident materialisation. In addition, the IM node is preceded by the IR node, which means that the response selection of the incident handler could affect the likelihood of the incident materialisation.

The question for eliciting the states of the IM node could be: *Assuming that such a threat is present and we implement such a response, what are the incidents that the threat could materialise in the system?* The questions for eliciting the likelihoods of the IM node could be as follows: *Assuming that such a threat is present and we implement such a response, what is the likelihood that the threat materialises as such an incident?* Important elements related with eliciting incident materialisations are the threat characteristics and the vulnerabilities of the MS. Finally, the expiration time corresponds to the moment of time in which any of the states or likelihoods could change.

Example 2.3. The incident materialisation refers to the potential event of the cyclone hitting the person's house. The first state is that the cyclone destroys the house, partially or fully. The second is that the cyclone causes a flood. The third is the non-materialisation. In our example, a cyclone so strong that it could destroy a house also typically involves floods. Then, our node in Fig 2.6 has three disjoint states: the destructive scenario that also involves flooding, the simple flooding scenario, and the non-disruptive scenario. This node contains the likelihood of incident materialisation, given the presence of a threat and the response of the incident handlers (note that if she leaves the house, she would not be able to mitigate breaks or floods in the house). The expiration time is the supposed day the cyclone leaves the city (e.g., three days).



	Heads to city		Heads other p.	
	Stay	Leave	Stay	Leave
Destructive	10 %	15 %	0 %	0 %
Flood	25 %	35 %	0 %	0 %
Non disruptive	65 %	50 %	100 %	100 %

Fig. 2.6: IM node for Example 2.3. Its precedent nodes and their states are displayed in white.

Consequences in managed system nodes

The *Consequences in the Managed System* nodes (CO) provide the likelihood that an incident or its response cause further negative events in the MS. Consequence scenarios are not disjoint by nature; i.e., an incident materialisation can cause different types of consequences in the MS. In this case, there should be a CO node for each of the different potential consequences.

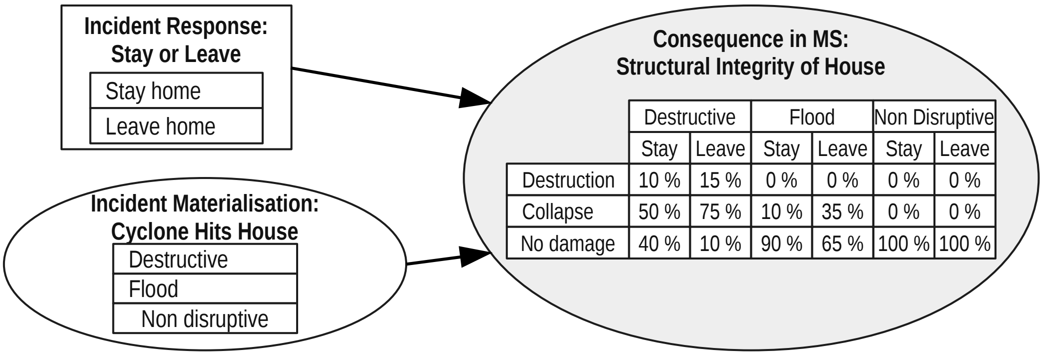
For instance, an incident reducing water supply in a summer camp might involve consequences on its showers, kitchen, or swimming pool. Each of them could have their CO node. The states of the CO nodes represent the severity levels for each consequence, which are disjoint in nature. For instance, the incident reducing water supply in the showers could have the following levels: fully operational, operational but with interruptions, unavailable.

CO nodes should include the interrelations amongst consequences implicitly in their likelihood, i.e, correlations shall be modelled with an external model that feeds GIRA. They might be interlinked with cause-effect relations amongst them. However, CO nodes in GIRA represent the occurrence likelihood for the particular consequence. Therefore, it is necessary to model multi-causality in an external model to find out the final likelihoods for each consequence.

CO nodes are preceded by the IM node, which means that the incident materialisation affects the likelihood of the consequence in the MS. In addition, CO nodes are preceded by the IR node, which means that the response selected by the incident handler could eventually mitigate negative consequences in the MS.

Questions for eliciting CO nodes could be: *Assuming (1) that such an incident materialises in the system, and (2) that such a response is implemented in the system, what would be the consequences in the system?* Questions for eliciting the states of each CO node could be: *What are the different levels such a consequence could have?* The question for eliciting the likelihoods of CO nodes could be: *Assuming that such an incident materialises in the system and that such a response is selected, what is the likelihood of this consequence level occurring?* Important elements for eliciting consequences are the vulnerabilities of the MS and, sometimes, threat characteristics. Finally, the expiration time corresponds to the moment in which any of the states or likelihoods could change.

Example 2.4. The cyclone hitting the house might cause consequences on the integrity of the house, its contents, or the person – if she stays. Each of them would be a CO node. For instance, the integrity of the house (Fig. 2.7) could involve no damage, collapse of building elements, and destruction of the house. Each of these levels is a state, as they are disjoint pairwise. This node provides the likelihood of the consequence level, given that the cyclone hits the house. Since the house would be the same during the studied period, the expiration time of this node is irrelevant.



	Destructive		Flood		Non Disruptive	
	Stay	Leave	Stay	Leave	Stay	Leave
Destruction	10 %	15 %	0 %	0 %	0 %	0 %
Collapse	50 %	75 %	10 %	35 %	0 %	0 %
No damage	40 %	10 %	90 %	65 %	100 %	100 %

Fig. 2.7: CO node for Example 2.4. Its precedent nodes and their states are displayed in white.

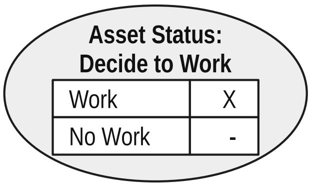
Asset status nodes

The *Asset Status* nodes (AS) apply to assets relevant when it comes to model the impacts over the stakeholders' interests. Different conditions in the assets might enable or increase the impact of an incident over them.

There could be multiple AS nodes, as each one would represent the status of a particular asset or set of assets. The states of the AS nodes represent each of the relevant operational status that an asset could have, i.e., asset status independent from the incident but relevant for the risk analysis (e.g., phase of a work, business situation, meteorological conditions). In addition, these types of nodes could be deterministic in case the analyst knows with certainty the current status and when it changes. Otherwise, nodes could be modelled as uncertainty nodes.

A sample question for eliciting the states of each AS node could be: *What are the status that enable such an impact on this specific asset?* The question for eliciting the likelihoods of AS nodes could be as follows: *What is the likelihood that this is the current status of the asset?* Finally, the expiration time corresponds to the moment in which any of the states or likelihoods could change.

Example 2.5. The person works as a freelance editor. Therefore, work is one of the assets (can be conceived as a DS) that might suffer an impact from the consequences of the cyclone in the house. The impact on her work depends on whether she is able to work, and whether she decides to work. The ability to work is determined by the consequences of the incident on her if she stays at home. However, the decision to work is modelled as an asset status (Fig. 2.8)[2]. The expiration time is also irrelevant here, since her work conditions would not change within the timespan considered.



Work	X
No Work	-

Fig. 2.8: AS node for Example 2.5.

[2] Although this node represents a decision, it is not modelled as a decision node, because we are not analysing that decision

Impact on assets nodes

The *Impact on Assets* nodes (IA) provide the likelihood that a consequence in the MS (CO nodes) leads to impacts on assets in the MS or DSs, taking into account the asset status from the AS nodes. Impacts refer to the valuable characteristics of the MS and the DSs. As explained in Sect. 2.2, we model consequences in the system that the incident handlers are responsible for or manage; and we model directly as impacts the consequences on relevant or dependent systems.

For example, in industrial cybersecurity, control systems are linked and overlapped with other systems, including information and communication systems, industrial operations, and physical assets. Therefore, in this example, a malfunctioning of a controller can destabilise the equipment, potentially leading to a physical accident or interrupting industrial operations. Analysing this chain of consequences might require an analysis of the DSs as well. In practice, it is better to model as impacts the consequences and losses in systems not managed by the incident handlers.

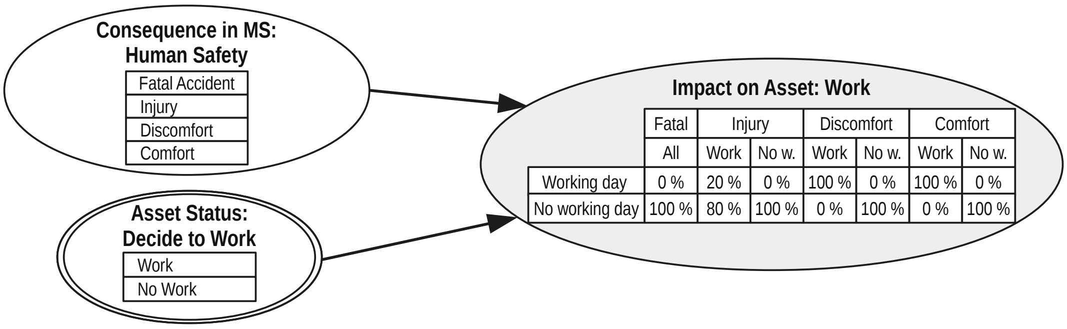
Impact scenarios are not disjoint by nature: a consequence might cause multiple impacts in the system. Hence, there could be multiple nodes of the IA type. Each of them would represent a different potential impact. In addition, the states of the IA nodes represent potential states of each impact. These states are disjoint in nature, typically, impact levels. For instance, the impact of a forest fire can have different levels depending on the amount of land burnt.

Additionally, there are two kinds of impacts based on the interaction of their antecedents. Impacts are *binary* if they only require one of its preceding consequences to exist (e.g., leaking a document). Impacts are *additive* if they increase their value when there are more preceding consequences causing them. For example, the more controllers malfunctioning in a control system, the more likely the equipment will fail. In addition, as consequences could be interlinked, it is important that the analyst does not identify the same impact twice; in general, this means ensuring that the impacts of a 'parent consequence' (e.g., empty hotel room when a customer cancels a reservation) do not duplicate the impacts of a 'child consequence' (e.g., empty suite). IA nodes should include the interrelations amongst consequences implicitly in their likelihood, i.e, correlations shall be modelled with an external model that feeds GIRA.

IA nodes are preceded by CO nodes, entailing that the likelihood of the consequence in the MS affects the likelihood of the impact on assets. In addition, IA nodes are preceded by AS nodes, so that the asset status also determines impact over assets.

A sample question for eliciting IA nodes could be: *Assuming that such a consequence happens in the MS and such a status happens in the asset, what are the impacts to the assets?* The question for eliciting the states of each IA node could be: *What are the different impact levels such impact type could have?* The question for eliciting the likelihoods of the IA node could be: *Assuming that such a consequence happens in the MS and such a status happens in the asset, what is the likelihood of such an impact level?* The most important elements related to eliciting impacts are those characteristics of the assets that make them valuable to the stakeholders. Finally, the expiration time is equivalent to the shorter amongst the CO or AS nodes expirations.

Example 2.6. The consequences of the cyclone hitting the house might lead to different impacts. For instance, loss of valuable items or the house, or human safety impacts. Another relevant impact could be that on work as explained in Example 2.5. Each of the impact types would be a node with different impact levels. For instance, a human condition node (Fig. 2.9) could have the following levels: correct, discomfort, injury, death. Each of these levels is a state, because they are disjoint (i.e., they escalate). This node provides the likelihoods for the impact on the working activity, given that the person had been harmed by the cyclone and her decision about working. The expiration time is inherited from its least-lasting precedent (CO node in Example 2.5), which was determined as irrelevant.



	Fatal	Injury		Discomfort		Comfort	
	All	Work	No w.	Work	No w.	Work	No w.
Working day	0 %	20 %	0 %	100 %	0 %	100 %	0 %
No working day	100 %	80 %	100 %	0 %	100 %	0 %	100 %

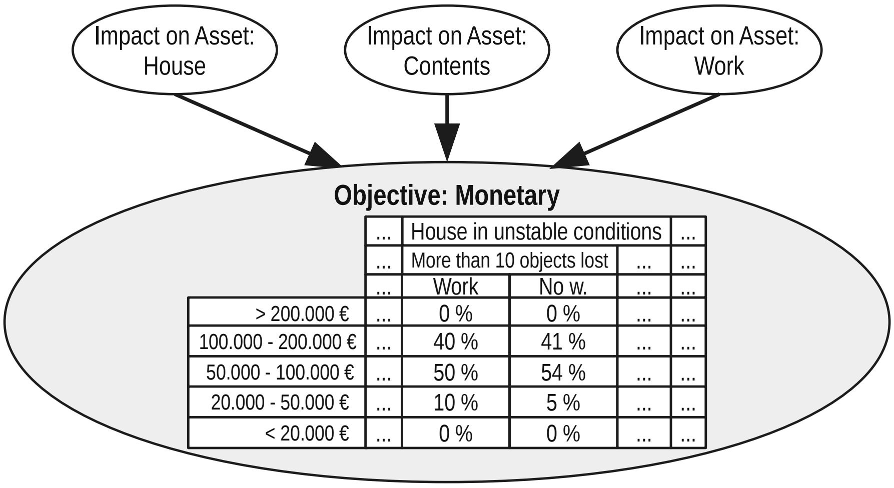
Fig. 2.9: IA node for Example 2.6. Its precedent nodes and their states are displayed in white.

Objective nodes

Objective nodes (OB) synthesise impact levels in a reduced number of objectives relevant to the stakeholders. The states of these nodes represent the potential states of each objective, typically, severity levels (e.g., in money or the severity of injuries).

A sample question for eliciting OB nodes could be: *Which objectives does such an impact affect?* A question for eliciting the states of each OB node could be: *What are the different objective levels of such an impact level?* The questions for eliciting the likelihoods of the OB nodes could be: (1) *Does this impact level translate automatically to this objective level?* If not, (2) *What is the likelihood that this impact level leads to such an objective level?* The most important elements related to eliciting objectives are, as in the case of impacts, those characteristics of assets that make them valuable to the stakeholders. The expiration time is also inherited from the predecessor nodes.

Example 2.7. The impacts of the cyclone on the house affect different objectives. For instance, human and monetary objectives. In our example, the Human Objectives node (Fig. 2.10) replicates the status of the IA node Human Condition. Therefore, it is a deterministic node. On the other hand, the Monetary Objectives node is an uncertainty node that assigns different likelihoods to different monetary loss levels. The expiration, inherited from the IA nodes, is irrelevant.



	...	House in unstable conditions			...
	...	More than 10 objects lost		...	...
	...	Work	No w.	...	...
> 200.000 €	...	0 %	0 %	...	...
100.000 - 200.000 €	...	40 %	41 %	...	...
50.000 - 100.000 €	...	50 %	54 %	...	...
20.000 - 50.000 €	...	10 %	5 %	...	...
< 20.000 €	...	0 %	0 %	...	...

Fig. 2.10: OB node for Example 2.7. Its precedent nodes and their states are displayed in white.

Risk description group of nodes

The combination of all the above nodes (2.4.1 to 2.4.1) represents the descriptive part of risk analysis. The use of an influence diagram enables the generation of the likelihoods for incident materialisation, consequences, impacts, and objectives. GIRA reduces the elements that the stakeholders need to compare to a limited number of objectives that synthesise likelihood and severity. This is equivalent to what risk matrices provide, but in a multi-objective version.

Therefore, GIRA can answer the following questions, among several other: *How likely is it that the threat materialises into this incident if we implement this response? And this consequence or consequence level? And this impact or impact level? And this objective or this objective level?* The expiration time for the overall risk description is the smallest amongst the different nodes.

Example 2.8. The model in Fig. 2.11 aggregates all the steps of the incident we modelled in Examples 2.1 to 2.7. Using this chain, GIRA calculates, for example, that the overall likelihood for the destruction of the house (IM node), given that the person stays home (IR node), is 2.5%. Another example could be that the overall likelihood for the person to die is 0.194% (Human OB node status), given that the cyclone heads to the city (TE node) and the person stays home (IR node). In case she decides to leave her home, it is 0% (note that we do not take into account other ways to die). The expiration time is the shortest amongst all nodes, in this case the TE node (6 hours).

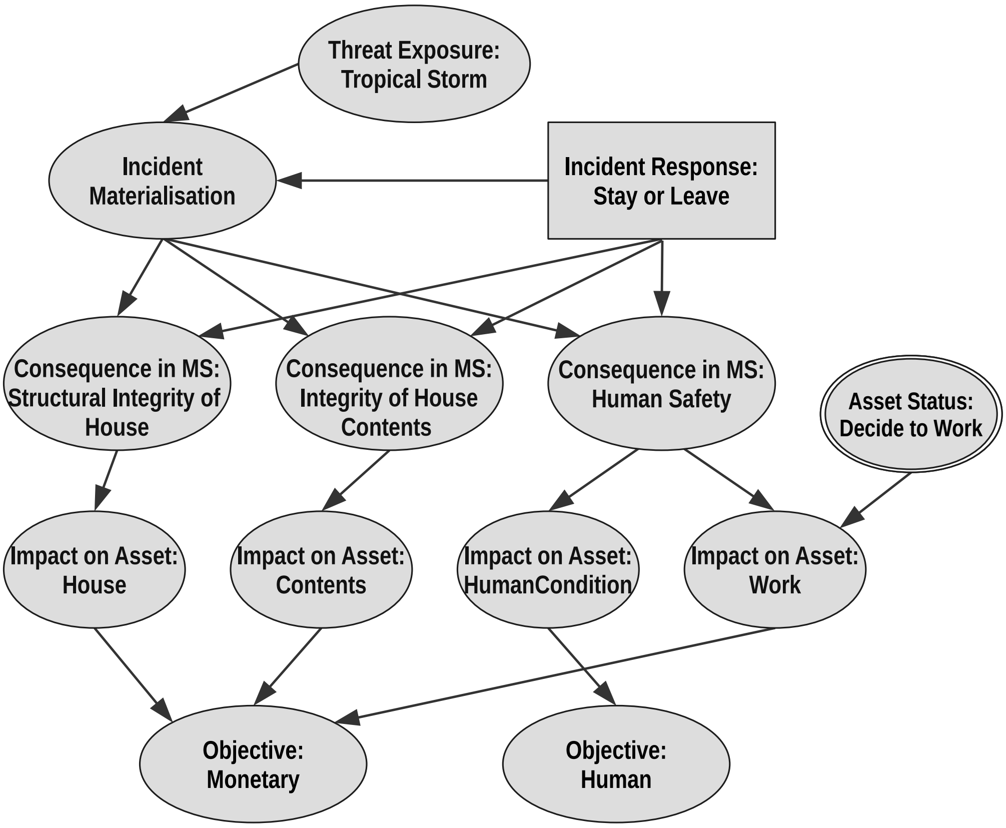
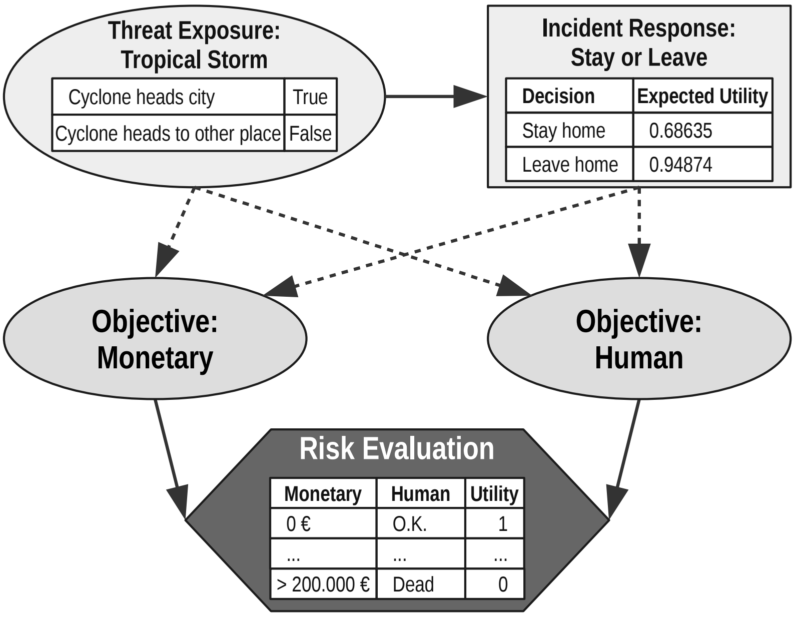

Threat Exposure: Tropical Storm
Incident Materialisation
Incident Response: Stay or Leave
Consequence in MS: Structural Integrity of House
Consequence in MS: Integrity of House Contents
Consequence in MS: Human Safety
Asset Status: Decide to Work
Impact on Asset: House
Impact on Asset: Contents
Impact on Asset: HumanCondition
Impact on Asset: Work
Objective: Monetary
Objective: Human

Fig. 2.11: Representation of the risk description nodes in Example 2.8.

Risk evaluation node

The *Risk Evaluation* node (EV) represents the stakeholders' evaluation of the risk scenario. The basic requirement for evaluation is ordering the different scenarios from the most preferred to the least. This node is general by assumption (Sect. 2.4), since the evaluation method is specific to the domain or the analyst. The EV node is designed for being fed by a sub-model that implements an evaluation methodology to order scenarios based on preferences and risk attitudes of stakeholders. For instance, MAUT, so that GIRA implements multi-objective optimisation] [102].

Example 2.9. As illustrated in Fig. 2.12, our protagonist orders her preferences from the least preferred scenario to the most preferred. In this case, she compares two objectives: human and monetary. The least preferred objective for her is dying without regard to how much money she loses; therefore, she assigns utility zero to this scenario. The preferred goal for her are a scenario without any impact on her health or wealth; therefore, she assigns a utility of one to this scenario. The rest of the scenarios have values in between. The next step is evaluating the utility of her potential responses to the incident. Recall that, under Utility Theory, the relevant issue is not the utility value itself but how the different choices are ordered [102].



Cyclone heads city	True
Cyclone heads to other place	False

Decision	Expected Utility
Stay home	0.68635
Leave home	0.94874

Monetary	Human	Utility
0 €	O.K.	1
...	...	...
> 200.000 €	Dead	0

Fig. 2.12: Representation of the EV node in Example 2.9, together with other relevant nodes. The TE node has a certain state (i.e., cyclone heading to the city). The IR node shows the expected utilities for each decision. Dashed lines represent indirect precedence of nodes (IM, CO, and IA nodes are not represented). The EV node displays the utility for the best scenario (i.e., no impact on the human's condition or monetary loss) and the worst.

2.5 Mathematical specification of GIRA

As an influence diagram, GIRA has a mathematical specification (Fig. 2.13). In this section we recapitulate the logic of the model and introduce its mathematical aspects.

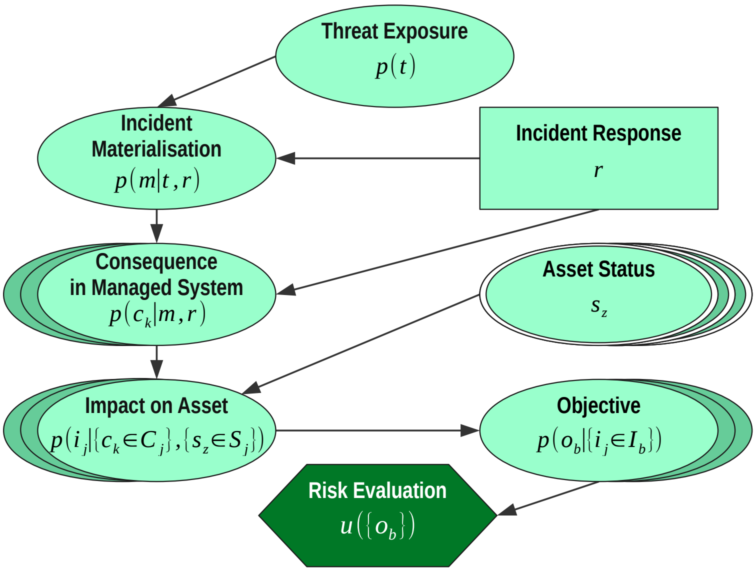


Fig. 2.13: GIRA with notation.

The *threat exposure* node represents the likelihood that a threat is present in, or targeting, the system that the incident handlers are in charge of protection (MS, the managed system). Mathematically, it is represented by the probability distribution $p(t)$. The *incident response* node represents the alternative actions that the incident handlers could implement to avoid or mitigate the incident. The variable representing these actions is r. The *incident materialisation* node represents the likelihood that the threat materialises as an incident in the MS, taking into account the response of incident handlers. This is the first conditional node, $p(m|t,r)$, which means that the probability of incident materialisation depends on the threat presence and the response. The *consequences in the managed system* nodes represent the likelihood that an incident or its response cause further negative events in the MS. Its distribution is modelled as $p(c_k|m,r)$. There could be multiple nodes of this type, so we define the set of consequence nodes as $\{c_k\} = \{c_1,\ldots,c_K\}$, being K the total number of consequences. An asset is any element affected by the incident and valuable to the stakeholders. The *impact on asset* nodes provide the likelihood that a consequence in the MS leads to impacts over the assets of the MS or other systems, or over any other stakeholders' interests. This node takes into account the current *asset status*, which might enable or escalate the impacts of the incident. An asset status is represented as s_z and the set of asset status nodes as $\{s_z\} = \{s_1,\ldots,s_Z\}$. An impact on asset node is represented as $p\Big(i_j|\{c_k : \exists\ c_k \to i_j\},\{s_z : \exists\ s_z \to i_j\}\Big)$, being

$\{c_k : \exists\, c_k \to i_j\}$ the set of consequence nodes parenting the j-th impact node[3] and, similarly, $\{s_z : \exists\, s_z \to i_j\}$ the asset status nodes parenting the j-th impact node. The set of impact on asset nodes is $\{i_j\} = \{i_1, \dots, i_J\}$. The *objective* nodes synthesise impacts in a reduced number of objectives to facilitate stakeholders understanding and comparing the outcome of the incident. An objective node is represented as $p\Big(o_b|\{i_j : \exists\, i_j \to o_b\}\Big)$, being $\{i_j : \exists\, i_j \to o_b\}$ impact on assets nodes parenting the b-th objective node. The set of objective nodes is $\{o_b\} = \{o_1, \dots, o_B\}$.

The combination of all the nodes, from threat exposure to objective nodes, represents *risk description*, which is modelled by the following equation:

$$\begin{aligned} p\Big(\{o_b\},\{i_j\},\{s_z\},\{c_k\},m,r,t\Big) = \\ = p(o_1,\dots,o_B,i_1,\dots,i_J,s_1,\dots,s_Z,c_1,\dots,c_K,m,r,t) = \\ = \left[\prod_{b=1}^{B} p\Big(o_b|\{i_j : \exists\, i_j \to o_b\}\Big)\right] \left[\prod_{j=1}^{J} p\Big(i_j|\{c_k : \exists\, c_k \to i_j\},\{s_z : \exists\, s_z \to i_j\}\Big)\right] \times \\ \times \left[\prod_{k=1}^{K} p(c_k|m,r)\right] p(m|t,r)\; p(t). \end{aligned}$$

Finally, the *risk evaluation* node represents the stakeholders' evaluation of the risk scenarios caused by the incident. It can be modelled, following the multi-attribute utility theory paradigm [102], as $u\Big(\{o_b\}\Big) = u(o_1, \dots, o_B)$. The actual risk evaluation is based on the expected utility when response r is implemented,

$$\psi(r) = \int \cdots \int \quad u\Big(\{o_b\}\Big) p\Big(\{o_b\},\{i_j\},\{c_k\},m,t\Big) \quad \mathrm{d}t \;\; \mathrm{d}m \;\; \mathrm{d}c_K \dots \mathrm{d}o_1.$$

From this equation, we can obtain the maximum expected utility response, by calculating $r^* : \max \psi(r)$.

Another aspect to consider is the time frame of the risk analysis. Specifically, the *expiration time* (x) of GIRA is the estimated moment of the earliest relevant change in any of the elements that participate in the incident (e.g., threat, system, assets). The expiration time could also be a specific time frame set by the analyst. The analysts should refer likelihoods to such time frame.

[3] More properly, the set of consequence nodes for which there exist an arc (directed edge as a graph) directed to the impact node i_j.

2.6 Additional GIRA models

GIRA can be derived into additional configurations, either for simplification or for extension.

The first simplified model is *GIRA for Simple Risk Description* (Fig. 2.14). Following the rules for simplifying influence diagrams [152], the likelihoods of CO and IA nodes (which can be modelled externally) can be simplified as likelihoods between the IM and OB nodes. The second simplified model is *GIRA for Simple Risk Evaluation* (Fig. 2.15). Following again the rules for simplifying influence diagrams, all uncertainty nodes can be modelled externally to synthesise them as OB nodes. This is the simplest version of GIRA.

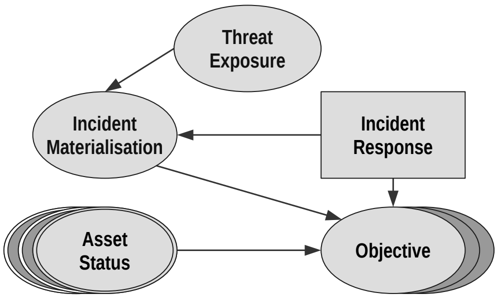


Fig. 2.14: GIRA for Simple Risk Description. CO and IA nodes are synthesised in the OB nodes with the support of an external model for analysing consequences and impacts.

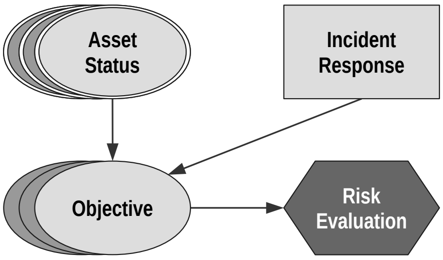


Fig. 2.15: GIRA for Simple Risk Evaluation. All uncertainty nodes are synthesised in the OB nodes with the support of an external model for analysing threat exposure, incident materialisation, consequences, and impacts.

When it comes to extension models, the first one is *GIRA for Multiple Stakeholders* (Fig. 2.16). The risk analysis might require that multiple stakeholders evaluate the risk. The solution is simply adding more IR nodes. This model is able to generate specific information for the stakeholders during negotiating processes, or for calculating social optimisation methods.

Another extension is *GIRA for Immediate and Eventual Consequences* (see Fig. 2.17). In Sect. 2.2.1, we discussed the distinction between immediate and eventual consequences of incidents. Modelling this would require, first, to establish that the time period we consider is the immediate one. The second step would be to classify the consequences between the immediate and the eventual groups of nodes. Some of the consequences would have to be split into both groups. The third step would be eliciting the likelihoods. This distinction between immediate and eventual consequences would also bifurcate the impacts and objectives between immediate and eventual nodes. This

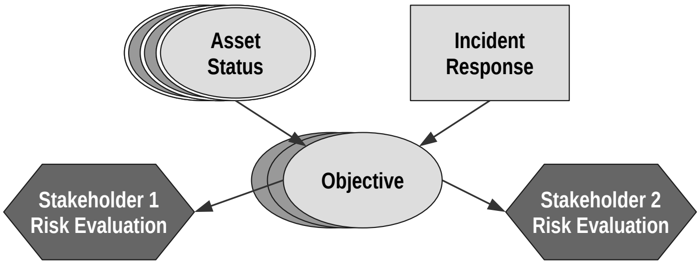


Fig. 2.16: GIRA for Multiple Stakeholders. This extension further adds EV nodes.

model might be useful for compound incidents in which the initial events are harmful or cause minor impacts and an early action could be counterproductive. For instance, removing a malware from a control system that is performing industrial operations could stop them and, thus, is usually better to do this at another moment.

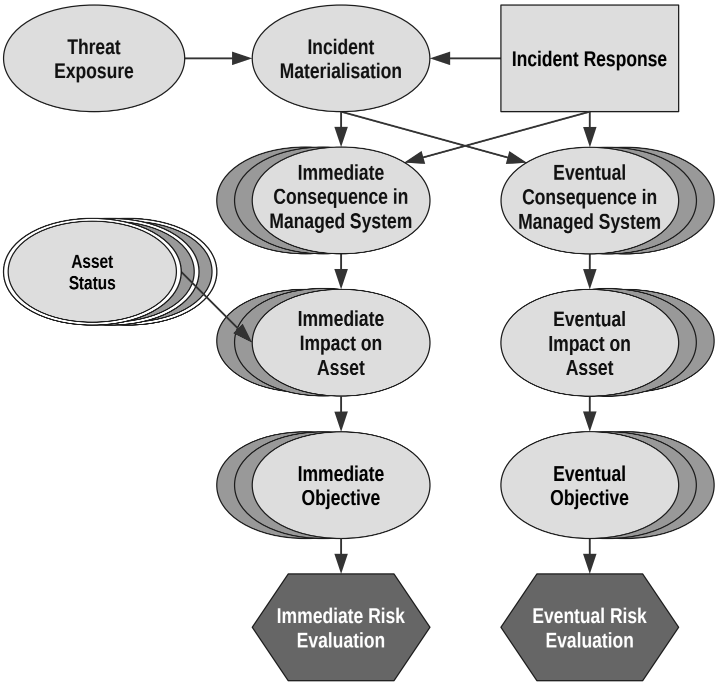


Fig. 2.17: GIRA for Immediate and Eventual Consequences. This extension bifurcates CO, IA, and OB nodes into two groups: immediate and eventual.

2.7 Discussion

We have proposed GIRA, a decision support model that formalises the incident risk analysis process through an influence diagram. Regarding our research objectives, the model of this chapter is the first step for developing a risk analysis model for cybersecurity incidents (RO1). Basically, we have found that the main aspects of the model can be generalised to incidents and risks in general. We also thought that a general and systematic discussion of incidents would be a useful contribution to thoroughly characterise the risk of incidents. The intention is to provide reliable risk indicators and a tool that overcomes various limitations that we have found in existing methods (Sect. 2.1). We discuss these contributions below.

Risk evaluation: Existing risk analysis methodologies provide a reliable analysis of incident risks from different points of view: upstream triggers (e.g., fault trees), downstream escalation (e.g., FMECA), and the combination of upstream and downstream approaches (e.g., bow-ties). However, these models do not cover risk evaluation. GIRA combines risk information from upstream and downstream analysis, but does also provide modelling of risk evaluation. Furthermore, GIRA generates risk information that serves as a complement to existing methods not covering the entire risk analysis process such as attack trees, FMECAs, and bow-ties.

Adversarial threats: Besides awareness on the inadequacy of frequencies for modelling adversaries, GIRA does not model explicitly adversarial threats. The simplest approach to overcome this limitation is using adversarial risk analysis [145] to feed the threat exposure or incident materialisation nodes.

Risk scoring and likelihood elicitation: Risk matrices cover the entire risk analysis process. However, their risk scoring method is oversimplified. This has advantages for brainstorming, structuring, and communicating risks. Yet, the use of an oversimplified scoring might lead to an ambiguous, and even meaningless, categorisation of likelihoods and impacts. GIRA, as an influence diagram, provides a graphical representation that facilitates the understanding of cause-effect relations in incident risk analysis. As a Bayesian network, it is suitable for sensitivity analysis, parameter learning, and what-if analysis. It is also possible to implement a qualitative analysis through the use of a semi-quantitative procedure for capturing the analysts' elicitation. In addition, GIRA elicits risk from the information of threats, responses, vulnerabilities, asset status, and the derived escalating events. Finally, the inclusion of expiration times makes GIRA actionable at real-time without entailing unnecessary complexity.

Risk comparison: The direct scoring of impacts and risks in one objective, as in risk matrices, obscures their incommensurability (e.g., monetary vs safety impacts). GIRA does not use a single objective but helps to reduce the elements that the stakeholders need to compare, by using objective nodes that synthesise risk likelihood and severity. This is equivalent to what risk matrices provide yet in a multi-objective fashion. Additionally, GIRA provides an evaluation node that supports the use of multicriteria decision and optimisation methods. As seen in Sect. 2.6, GIRA can be simplified to versions that resemble the communication simplicity of risk matrices without losing any of the analytical capabilities of the complete GIRA model.

Chapter 3

CSIRA: A method for analysing the risk of cybersecurity incidents

This chapter takes the general model GIRA (Chapter 2) and tailors it to cybersecurity (hereafter called CSIRA). Additionally, we provide supporting methods (Sect. 3.2) for simplifying the risk analysis: one for categorising the ramifications of cybersecurity incidents and a minimal method for eliciting likelihoods based on the oddness of events. Sect. 3.3 introduces CSIRA, supported by an example of its application in Sect. 3.4. Finally, Sect. 3.5 discusses CSIRA.

3.1 Brief on cybersecurity incidents

Cybersecurity incidents happen in a context of uncertainty in which incident responders have to analyse the potential uncertainties around the incident and the potential consequences in the system and on the assets. The earlier signs of one of these events are, typically, suspicious anomalies that could also be caused by legit actions by the system or users. Here, the analysis focuses on identifying what could have caused the anomalous event, and what events might follow. For instance, a new connection within a network could be caused by a maintenance laptop or an unauthorised party accessing the network. Additionally, if a specific attack or problem has been identified, then the analysis focuses on identifying the consequences of the threat, how likely they are or how the potential countermeasures would change the risk. For example, analysing the presence of malware in an industrial controller would deal with aspects such as whether it is harmful to the controller or the current industrial process, whether it can spread to other devices or what the consequences of removing the malware or changing the device are.

Analysing risk is also critical for dealing with cybersecurity incidents. However, there is no explicit method for analysing risk during cybersecurity incidents, relying on the general methods introduced in Sect. 2.1.

3.2 Base models

Our approach for building a cybersecurity incident risk analysis model is based on the following three models/methods:

1. The general model for incident risk analysis (GIRA, in Chapter 2).
2. A simple method for eliciting likelihoods based on the oddness of events.
3. A method for mapping the ramifications of cybersecurity incidents.

We describe the last two in the rest of the section.

3.2.1 Eliciting the likelihood based on the oddness of the event

The quality of risk analysis relies on how well it considers uncertainty [55]. This is achieved by using suitable and well-processed data, if available, or in the partial or complete support of expert knowledge [54] or other elicitation methods [141]. However, this information might not be available during the time frame of the incident, in which the analysts do not have access to data or experts.

Analysing the likelihood of events using a qualitative interpretation could be arbitrary, but a meaningful yet practical approach is basing this splitting on a qualitative interpretation of probability ranges: *certain* for $p(e) = 1$, *possible* for $p(e) = (t,1)$, *rare* for $p(e) = (0,t)$ and *impossible* for $p(e) = 0$. Any event x that clearly has a likelihood below the interpretative oddness threshold t is defined as rare, whereas the events with a likelihood around or above t are defined as possible. This simple method can be extended with several levels of *oddness*. Interpretatively, this means that rare would change to $p(e) = (t_2,t_1)$ and could be conceived as *rare (oddness 1)*, and that we could define a new *rarer than rare / rare (oddness 2)* event with $p(e) = (t_3,t_2)$. We can continue this process until a *rare (oddness i)* event, which might be useful for comparing the likelihoods of different events, although it would become more and more difficult to interpret in absolute terms.

Additionally, we can establish a rule for the likelihood of a chain of n events, based on the accumulated oddness, i.e.,

$$p(e_n|e_{n-1}|\dots|e_1) = (t_{l-1}, t_l) : l = \sum_i^n \mathrm{odd(e_i)},$$

being $\mathrm{odd}(e_i)$ the oddness of the event. Certain and possible events have an oddness of zero. Additionally, any chain with at least one impossible event is automatically impossible, and any chain with all of its events certain is automatically certain.

Following these rules we have that a chain of possible and certain events is possible, a chain with a rare event would be rare (one event with oddness 1), a chain with two rare events would be a rarer than rare event (two events with oddness 1), a chain with a rarer than rare event would be a rarer than rare event too (one event with oddness 2). For instance, in industrial cybersecurity, an analyst could interpret that the event of an attacker manipulating a controller is rare and that, given such a manipulation, the event of a controlled sabotage by the attacker is rare. Therefore, this chain of events would be elicited as 'rarer than rare event'.

Table 3.1 summarises these concepts. It also shows the numerical implementation in a Bayesian network like GIRA, which can take the qualitative likelihood as a numerical input to populate the probabilities of nodes and, vice versa, translate the overall probabilities calculated by the network into the qualitative interpretation again.These values are defined based on practical purposes. First, a probability range of 2 orders of magnitude, e.g. $(1 \times 10^{-2}, 1)$, allows us to model dozens of states. The differences among the magnitudes of the various probability ranges are established in a way so that a chain with a rare event will always have a lower probability than a chain without it. In the case of GIRA, we have a chain of 5 nodes and, taking into account that we use probability ranges of 2 orders of magnitude, the difference between probability ranges must be, at least, 10. This way, by multiplying the probabilities of the chain of events, we will get as output the overall probabilities, with their different orders of magnitude.

Qualitative likelihood	**Probabilistic interpretation**	**Numerical input to GIRA Bayesian network**	**Numerical output from GIRA Bayesian network**
Certain	1	1	1
Possible	$(t_1, 1)$	$(1 \times 10^{-2}, 1)$	$(1 \times 10^{-10}, 1)$
Rare (oddness 1)	(t_2, t_1)	$(1 \times 10^{-12}, 1 \times 10^{-10})$	$(1 \times 10^{-20}, 1 \times 10^{-10})$
Rarer than rare (oddness 2)	(t_3, t_2)	$(1 \times 10^{-22}, 1 \times 10^{-20})$	$(1 \times 10^{-30}, 1 \times 10^{-20})$
...	...	...	...
Impossible	0	0	0

Table 3.1: Table with the probabilistic interpretation of qualitative likelihoods.

3.2.2 Understanding potential ramifications of cybersecurity incidents

Multiple guidelines and taxonomies exist for identifying and categorising cybersecurity risks. We can distinguish two groups. One group at the technical level, the larger in the literature, deals with the categorisation of cyber attacks and their effects on digital systems. These guidelines might be useful for identifying elements related to threats, incidents, and system consequence. The other group deals with the impact that cybersecurity risks might have on assets, value or risk objectives. Examples of widely used methods are COBIT [93] or FAIR [161]. However, the majority of the categories for impacts and assets have a perspective that pivots on a business/organisational interpretation of assets and stakeholders. Although most risk management happens in organisational settings (e.g., business or public agencies), a more broad perspective is feasible when thinking about cybersecurity risk impacts, i.e., asset as something with value for somebody and stakeholder as somebody that might be affected by the incident.

A thorough categorisation model would require a combination of IT, OT, cyber-phisical and cyber-psychological risks, an analysis of their impact at microsocial and macrosocial level and an analysis of what new cyber risks would emerge in the future (e.g. what risks the pervasive use of virtual reality will bring and how they could become cybersecurity risks). There is no scientific or technical literature so comprehensive. However, a simplified model for quick elicitation may be established. Fig. 3.1 depicts a graphical model for categorising the potential ramifications of cybersecurity incidents. In the context of GIRA, this model might be helpful for identifying the consequences and impacts nodes.

The starting point is the managed system (MS), in which the analysed cybersecurity incident happens. The primordial risks of cybersecurity incidents are those involving the processing, storage and transmission of digital data. For example, ransomware, denial of service or man-in-the-middle attacks. These events could happen in the MS or other digital systems managed by the organisation dealing with the incident or third parties.

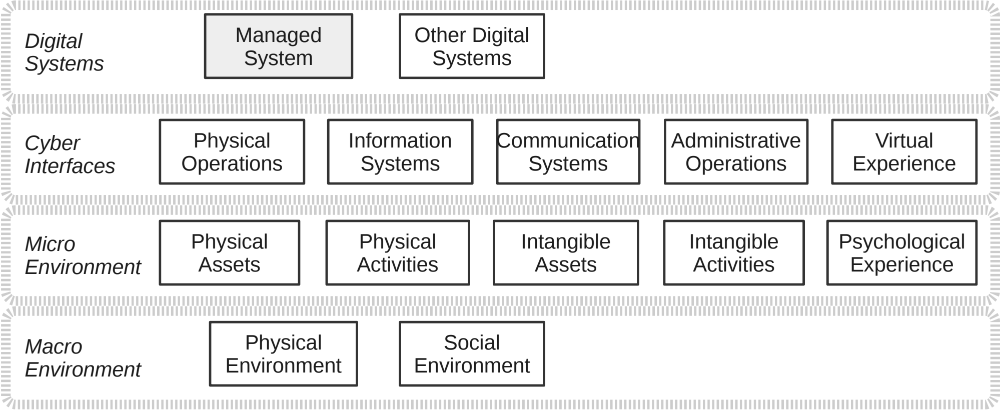


Fig. 3.1: Categories that classify the ramifications of cybersecurity risks

However, the importance of cyber risks resides, mostly, in the ramifications to other organisational or physical systems and assets that depend on, or can be affected by, the compromised digital systems. The most direct ramifications are the incidents grouped in the broad category of cyber interfaces. Physical operations refer to the interactions between physical reality and digital systems, such as input and output devices (e.g., keyboards, screens, printers, mouses, USB ports) or the actuators and sensors of industrial control systems. Examples of risks here involve unauthorised cyber-physical actions like the ones executed by Stuxnet [106] (manipulation of nuclear plant centrifuge speeds) or the malicious hijacking of laptop cameras. Information systems refer to the actual information contained in the digital systems (e.g., documents, pictures). An example risk in this case is the stealing of secret documents. Communication systems refer to the actual communication facilitated by the digital systems (e.g., chats, video conferences). Examples of risks here are the interference with a video conference or even the dissemination of false information through vulnerabilities in social networks (e.g., Twitter bots). Administrative operations refer to the affairs conducted with the digital systems (e.g., invoicing or buying online). An example risk in this area is the hijacking of an e-banking account. The virtual experience refers to the human experience in the reality created by the digital system (e.g., user experience in an application, human interaction in a social network). Examples of this type of risk are the exposure of personal information or sensitive images in social networks.

The indirect ramifications are categorised in a micro and a macro environment that refer to non-digital and non-cyber consequences. The micro-environment refers to risks at the particular or organisational level, as well as risks with organisations and people with a relatively direct relationship (e.g., customers and suppliers for a business, family and friends for a person). The first type of risks are in physical assets (e.g., machinery, personnel) and activities (e.g., manufacturing and transporting items). An example risk could be the sabotage by Stuxnet of the facility centrifuges (asset) and the enrichment of uranium (activity). Intangible assets refer to any characteristic or thing without physical presence. Example risks are the loss of secrets, reputation, compliance or money caused by a cyber attack. The psychological aspect refers to how cyber risks affect the human experience. Examples of these risks are the psychological problems derived from cyber-bulling or the exposure or personal data to the public. The macro-environment refers to the consequences at a social or ecosystem level. For instance, the political impact on Iran of Stuxnet, or the environmental and economic impact in the case a cyber attack facilitates an accident with contaminants or dangerous materials in an industrial facility.

3.3 CSIRA: Cybersecurity incident risk analysis

Now we introduce the *cybersecurity incident risk analysis model (CSIRA)*, which aims at providing a paradigm practicable as a quick risk analysis method during cybersecurity incidents. CSIRA combines GIRA, the oddness method for likelihood elicitation, the graphical model for brainstorming cybersecurity incident ramifications and a simplified method for risk evaluation based on comparing the outcomes of different incident responses.

First, CSIRA uses GIRA as the risk analysis model, so that a high-level but comprehensive method is applied to the cybersecurity incident assessment. As argued previously, risk matrices oversimplify many risks components and other methods are more focused on the technical side (e.g. bow-ties). It is feasible to combine the use of a more detailed technical model for the cyber attack (e.g., attack tree) and the consequences (e.g., fault tree) with the use of GIRA for the impact and objective analysis.

Second, CSIRA uses a simplified interpretation of likelihood (Sect. 3.2.1), so that the elicitation is quick but at least implementable numerically. The qualitative scale of risk matrices cannot be applied to a chain of events nor be interpreted easily as a probability range. We also assume that a quantitative or semi-quantitative elicitation is not feasible in real-time. If so, then it would also be feasible to directly use GIRA, with quantitative data or expert elicitation.

Third, CSIRA uses a simplified model for eliciting the ramifications of cybersecurity incidents (Sect. 3.2.2), so that all feasible types of incidents are thought about. This intends to facilitate brainstorming, based the contextual knowledge of the user undertaking the analysis. We think that this approach is more feasible and useful in real time than presenting a general catalogue of impacts.

Fourth, GIRA would need the elicitation of the preferences and risk attitudes of the stakeholders, following the standard process in influence diagram building. However, this would require time and support from experts. For CSIRA, we establish a faster alternative method, described in Sect. 3.4.4: Once the users build the risk description part, they could obtain the total likelihoods of the risk problem. From the decision-making perspective, the only comparison they have to make is how the responses to the incident, and inaction, affect risk objectives.

CSIRA does not contain any knowledge base or any process to build one. For that to be useful, it would be necessary with very tailored information adapted to the specific systems, assets and stakeholders of the organisation. Indeed, rather than the potential incorporation of cybersecurity knowledge, we would recommend the use of a collection of cybersecurity standards. The most relevant one in this case is the NIST Cybersecurity Framework [129], which provides (1) the most comprehensive structuring of the aspects that should be taken into account in cybersecurity management and (2) specific chapters that deal with these topics in other relevant collections of standards (e.g., NIST, ISO, COBIT). Nor do we provide any automatic reasoning mechanism besides the Bayesian calculation of likelihoods. Although automation would reduce human task load, it would also take decision-making from the users. Indeed, the intention is the opposite: providing a risk analysis model that explicitly relies, as much as possible, on human interpretation and decision-making.

3.4 An example cybersecurity risk analysis

This section introduces the steps for using CSIRA, supported by an example. Our intention is not to undertake a realistic risk analysis but to provide an example to show CSIRA. First, we cover risk description, which consists in three steps. The first step, in Sect. 3.4.1, is risk identification using the graphical model presented in Sect. 3.2.2 for identifying cybersecurity incident ramifications. The second step is risk elicitation (Sect. 3.4.2), using GIRA as the base risk model with the elicitation method presented in Sect. 3.2.1 to generate the likelihoods of different events. The final step of the risk description is risk calculation, using also the mentioned elicitation method. The outcome of risk description are the relevant risk scenarios for decision-making: the potential results of the different incident responses regarding their relevant risk objectives. The risk analysis finalises with the risk evaluation of Sect. 3.4.4.

The example case is applied to the industrial control systems (ICS) of an oil and gas drilling rig, as this facility is a paradigmatic case of the physical and organisational ramifications that a cybersecurity incident could have. The incident would be the presence of a wiper malware in the system in charge of drilling the well. This kind of malware is capable of erasing data in the operating system (OS) boot records or critical files. Interestingly [78] some of the most notorious wiper cyber attacks, like Shamoon and BlackEnergy, targeted the oil and gas industry. The human-machine interfaces (HMI) of industrial systems are typically installed on top of popular OS like Windows. Therefore, a disruption in the HMI caused by a wiper might affect, to some extent, the industrial operation that the HMI helps to control. This involves that incident handlers should think about the ramifications of the incident on industrial operations and assets.

3.4.1 *Risk description: identification*

Fig. 3.2 depicts the consequences and impacts of cybersecurity incidents, applying the method of Sect. 3.2.2 to our scenario of a wiper in a drilling rig. The managed system is the drilling ICS. The initial incident is the presence of the threat, i.e., the presence of the wiper malware in the ICS. The exposure to this threat could lead to the main incident, which is the execution of the wiper in the PC hosting the HMI software. The square represents the potential response of the incident handler. Given that a wiper could be a sophisticated tool, a full fresh re-installation of the HMI PCs would be a prudent response.

In case the wiper is successfully running in an HMI PC, the next consequence could be the disruption of the OS of the HMI PC. In addition, the incident response has also a consequence: a fresh installation of the HMI PCs would need to put the ICS under maintenance for 24 hours. The next step is to identify the ramifications that the disruption could have beyond the ICS. The first one is the disruption in the human-machine interface, i.e., the disruption of the interaction between operator, ICS and industrial operation. This could lead to a disruption of the drilling operations, which in turn might lead to incidents with equipment, the oil well or personnel. In addition, an incident involving the well integrity might lead to a spill involving hydrocarbons or other contaminants into the rig floor or the sea. An additional consequence, very relevant in oil platforms, is the loss of time, which can be caused by both the disruption in the drilling operations and the maintenance of the ICS (in the case of re-installing the HMI OS). However, one important element affects the disruption of the drilling operations: whether the platform is drilling or performing other activity.

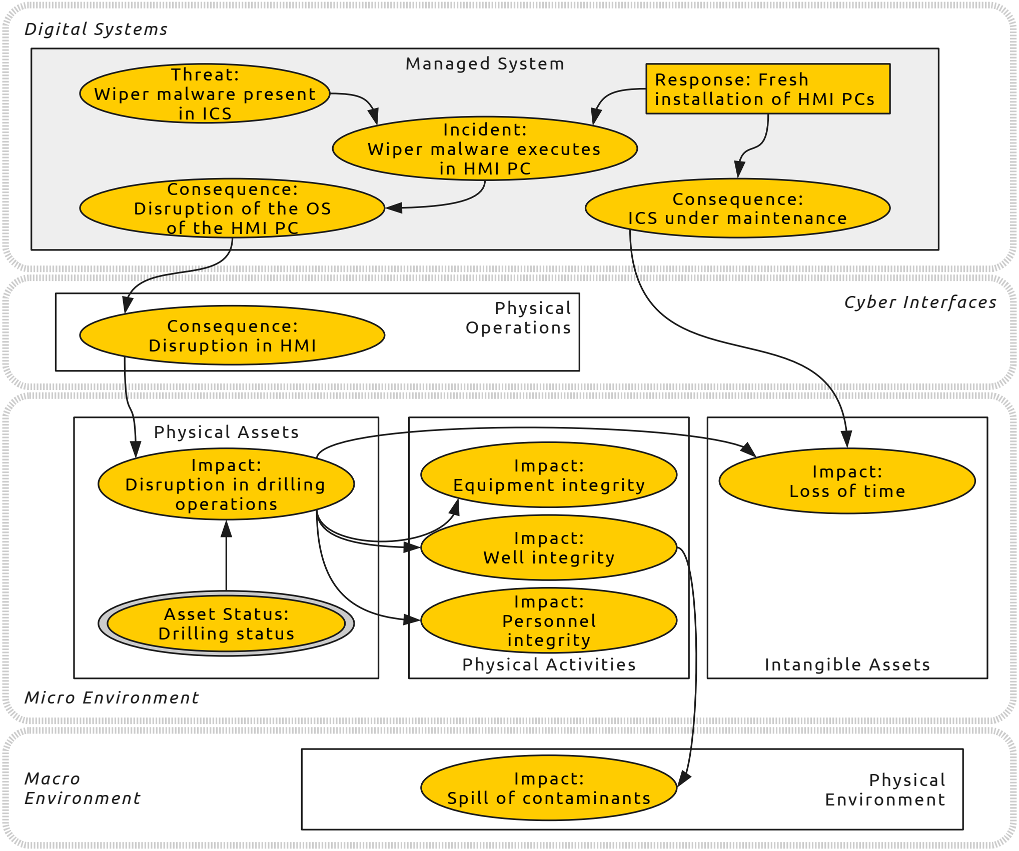

Digital Systems
Managed System
Threat: Wiper malware present in ICS
Response: Fresh installation of HMI PCs
Incident: Wiper malware executes in HMI PC
Consequence: Disruption of the OS of the HMI PC
Consequence: ICS under maintenance
Cyber Interfaces
Physical Operations
Consequence: Disruption in HMI
Physical Assets
Impact: Disruption in drilling operations
Asset Status: Drilling status
Impact: Equipment integrity
Impact: Well integrity
Impact: Personnel integrity
Physical Activities
Impact: Loss of time
Intangible Assets
Micro Environment
Macro Environment
Impact: Spill of contaminants
Physical Environment

Fig. 3.2: Graphical representation of potential risks of a wiper in a drilling rig. Rounded nodes represent uncertain events. Rectangles represent incident handler decisions. Double-rounded circles represent known states.

3.4.2 Risk description: elicitation

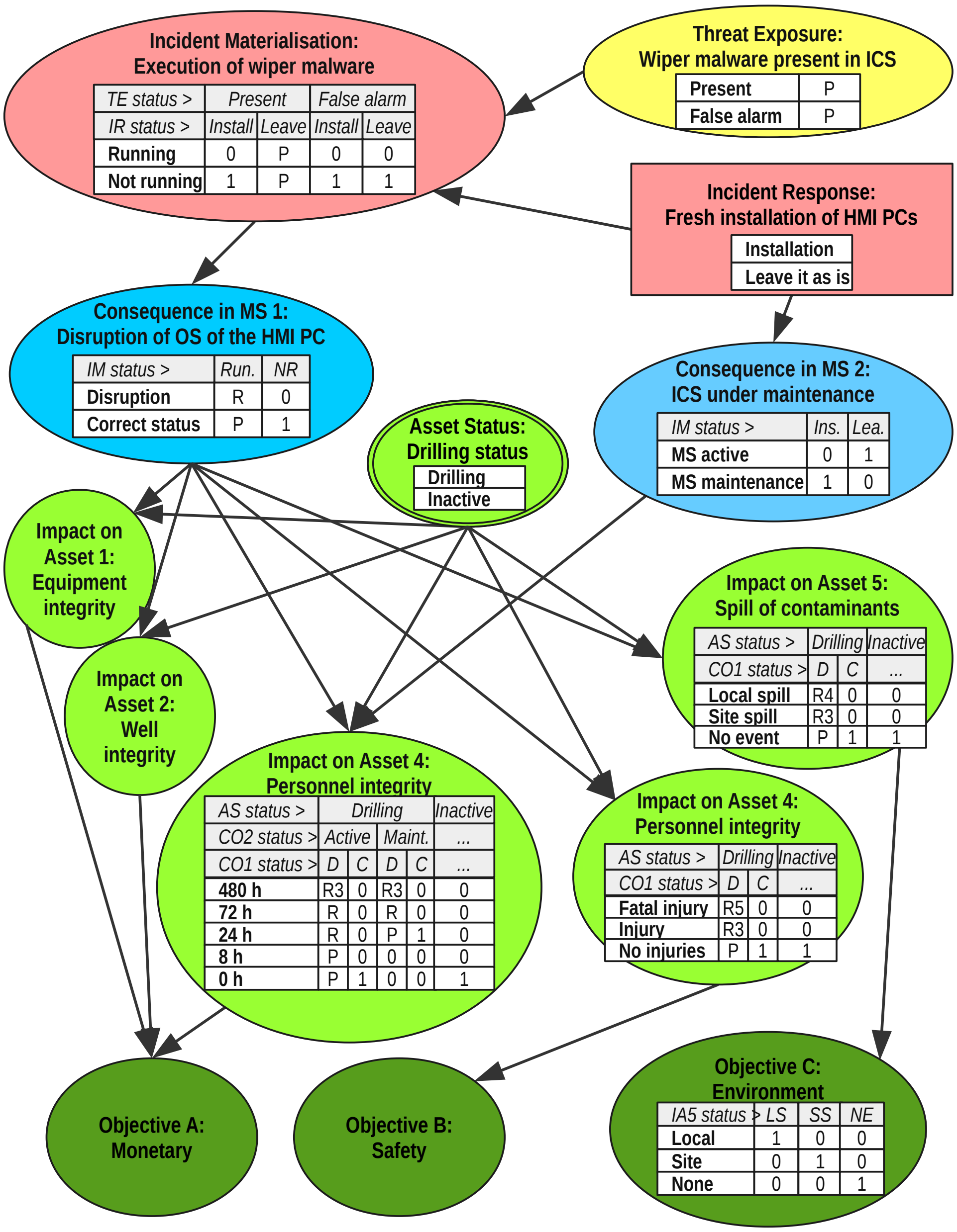
Fig. 3.3 illustrates the influence diagram of our example, using the likelihood elicitation of Sect. 3.2.1, and derived from the risks identified in Sect. 3.4.1.

The uppermost node is the *threat exposure*. It represents the uncertainty about the presence of the wiper. In this case, the analysts considered that the presence is possible (represented as P in the graph). Its complementary state (no presence of wiper) is also possible. Additionally, the *incident response* node represents the actions that the incident handler can take. In our case, the re-installation of the HMI OS with a fresh and updated version or the option of leaving the system as is.

The *incident materialisation* node represents the main incident: the execution of the wiper in the HMI PC. It has two uncertain states: whether the wiper runs in the PC or not. However, these events are conditioned by two factors. First, whether the wiper presence is a false alarm (threat exposure node). Second, whether the incident handlers re-install the HMI PCs. This is reflected in the likelihood assigned. If the wiper is present and the incident handlers leave the system as is, then it is possible that the wiper would run in the HMI PC. Otherwise, the wiper would not run (in the graph, 0 represents impossible and 1 represents certain).

There are two *consequence in the managed system* nodes. The first one represents the event of the wiper actually disrupting the OS of the HMI. In case the wiper is running in the HMI PC, then the likelihood of the HMI disruption is rare (as established earlier, rare (oddness 1), represented in the graph as R1) and the likelihood of its opposite is, thus, possible. In case the wiper is not running, then the certain event is the correct status. The second consequence node represents the event of putting the system under maintenance caused by the re-installation of the HMI PCs.

There are several *impact on asset* nodes. They represent most of the incident ramifications outside the managed system we identified in the previous section, except the disruption of drilling operations. The reason is that such disruption acts as an ‘intermediate‘ risk, i.e., its risks are reflected on other assets, like the integrity of the different assets, the loss of time or the spill of contaminants. These nodes are preceded by the asset status node informing whether the platform is drilling. In addition, the impact nodes should summarise the likelihood of the chain of events that do not happen in the MS but may lead to those impacts. This means that given a consequence in the MS and the status of some asset, they should reflect the likelihood of the different impact levels attainable. For instance, in case the impact 5



TE status >	*Present*		*False alarm*	
IR status >	*Install*	*Leave*	*Install*	*Leave*
Running	0	P	0	0
Not running	1	P	1	1

Present	P
False alarm	P

Installation
Leave it as is

IM status >	*Run.*	*NR*
Disruption	R	0
Correct status	P	1

Drilling
Inactive

IM status >	*Ins.*	*Lea.*
MS active	0	1
MS maintenance	1	0

AS status >	*Drilling*		*Inactive*
CO1 status >	*D*	*C*	*...*
Local spill	R4	0	0
Site spill	R3	0	0
No event	P	1	1

AS status >	*Drilling*				*Inactive*
CO2 status >	*Active*		*Maint.*		*...*
CO1 status >	*D*	*C*	*D*	*C*	*...*
480 h	R3	0	R3	0	0
72 h	R	0	R	0	0
24 h	R	0	P	1	0
8 h	P	0	0	0	0
0 h	P	1	0	0	1

AS status >	*Drilling*		*Inactive*
CO1 status >	*D*	*C*	*...*
Fatal injury	R5	0	0
Injury	R3	0	0
No injuries	P	1	1

IA5 status >	*LS*	*SS*	*NE*
Local	1	0	0
Site	0	1	0
None	0	0	1

Fig. 3.3: Influence diagram representing the risk analysis for the wiper incident in a drilling control system. When it comes to the likelihoods, a sure event is represented with 1, an impossible event with 0, a possible event with P, a rare event with R, a rarer than rare event with R2, and so on.

'spill of contaminants' we have that, given that the asset status is drilling and that the HMI PC has been disrupted, the likelihood of a local spill is rare (oddness 4), the likelihood of a site spill is rare (oddness 3) and the likelihood of the no spill event is possible.

It is necessary to analyse the chain of events to determine whether one event is clearly rarer than other, as in Sect. 3.2.1. For instance, the event of a fatal personnel injury is established as clearly rarer than a non-fatal injury and than a local spill. Then, we establish that the event of a local spill is clearly rarer than a site spill. Following this procedure, we assign the different oddness to different events.

A final aspect to take into account is the expiration time of this risk analysis. Most of the events described have no clear time boundary. However, one of the nodes of our example stands out as the compass of timely risk response: the asset status node. First, all of the relevant impacts happen when the platform is drilling. Second, the incident handlers are able to know whether the platform is drilling or not and when this status would change. For instance, drilling might be scheduled for turns lasting several hours in the upcoming weeks. As an example, the expiration time for the analysis could be 8 hours.

3.4.3 Risk description: calculation

Following the procedure for likelihood calculation in Sect 3.2.1, we can calculate the final conditional probabilities of the different nodes of the influence diagram. Fig. 3.4 displays the calculation for the case in which the incident response 'leave the MS as is' is selected and taking into account that the current asset status is 'drilling'.

The logic of the influence diagram allows us to disregard infeasible and impossible events. For instance, the stricken out text in grey cells highlights infeasible events (e.g., in the consequence 2 node, it is infeasible any event that is conditioned by the incident response event of 'installation') or impossible events (once again, in the consequence 2 node, the event of 'maintenance' is impossible, given that the incident response event is 'leave the system as is'). This kind of reasoning propagates through the diagram.

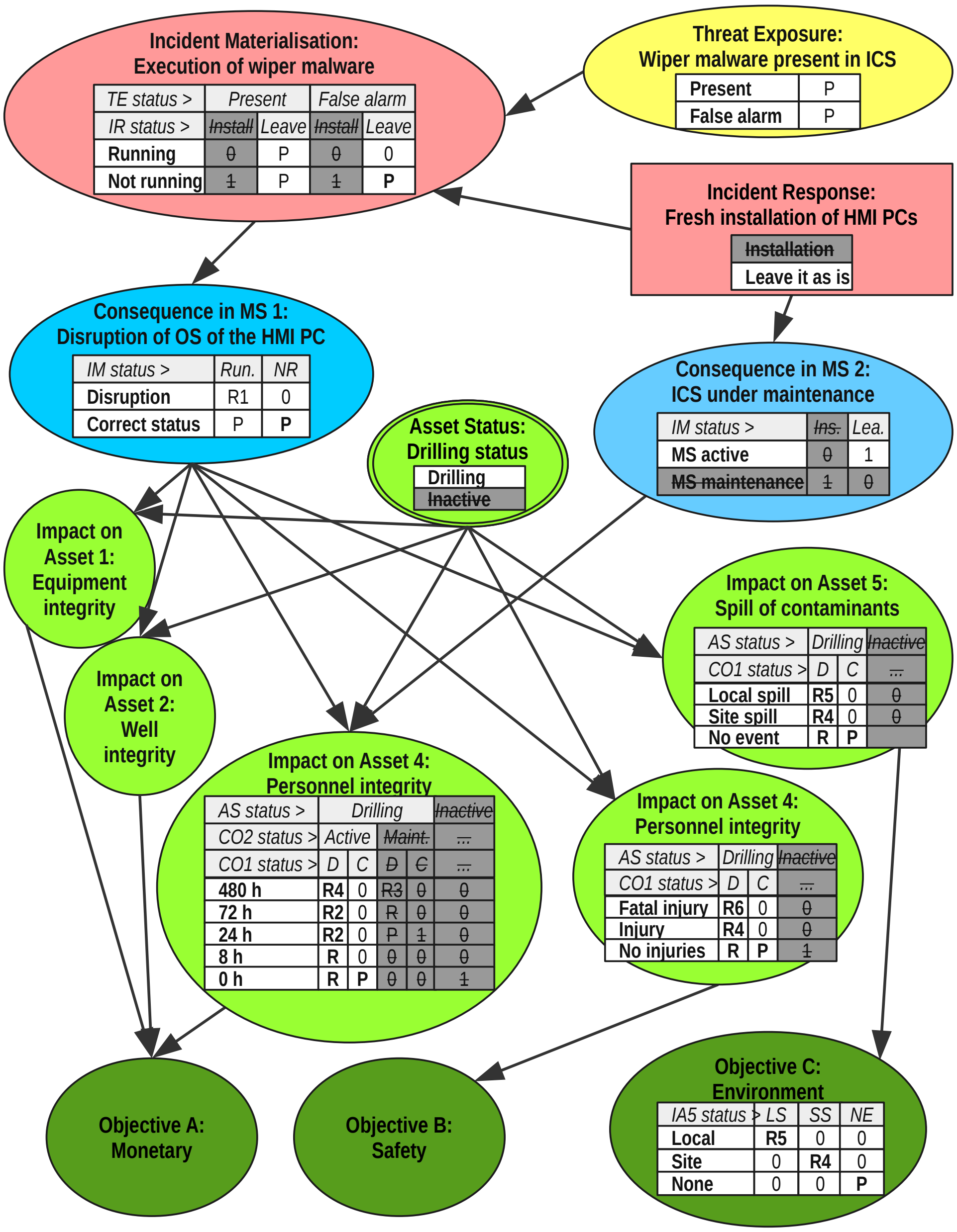

Incident Materialisation:
Execution of wiper malware
TE status > | Present | False alarm
IR status > | ~~Install~~ | Leave | ~~Install~~ | Leave
Running | ~~0~~ | P | ~~0~~ | 0
Not running | ~~1~~ | P | ~~1~~ | P
Threat Exposure:
Wiper malware present in ICS
Present | P
False alarm | P
Incident Response:
Fresh installation of HMI PCs
~~Installation~~
Leave it as is
Consequence in MS 1:
Disruption of OS of the HMI PC
IM status > | Run. | NR
Disruption | R1 | 0
Correct status | P | P
Asset Status:
Drilling status
Drilling
~~Inactive~~
Consequence in MS 2:
ICS under maintenance
IM status > | ~~Ins.~~ | Lea.
MS active | ~~0~~ | 1
~~MS maintenance~~ | ~~1~~ | ~~0~~
Impact on Asset 1:
Equipment integrity
Impact on Asset 2:
Well integrity
Impact on Asset 5:
Spill of contaminants
AS status > | Drilling | ~~Inactive~~
CO1 status > | D | C | ~~...~~
Local spill | R5 | 0 | ~~0~~
Site spill | R4 | 0 | ~~0~~
No event | R | P
Impact on Asset 4:
Personnel integrity
AS status > | Drilling | ~~Inactive~~
CO2 status > | Active | ~~Maint.~~ | ~~...~~
CO1 status > | D | C | ~~D~~ | ~~C~~ | ~~...~~
480 h | R4 | 0 | ~~R3~~ | ~~0~~ | ~~0~~
72 h | R2 | 0 | ~~R~~ | ~~0~~ | ~~0~~
24 h | R2 | 0 | ~~P~~ | ~~1~~ | ~~0~~
8 h | R | 0 | ~~0~~ | ~~0~~ | ~~0~~
0 h | R | P | ~~0~~ | ~~0~~ | ~~1~~
Impact on Asset 4:
Personnel integrity
AS status > | Drilling | ~~Inactive~~
CO1 status > | D | C | ~~...~~
Fatal injury | R6 | 0 | ~~0~~
Injury | R4 | 0 | ~~0~~
No injuries | R | P | ~~1~~
Objective A:
Monetary
Objective B:
Safety
Objective C:
Environment
IA5 status > | LS | SS | NE
Local | R5 | 0 | 0
Site | 0 | R4 | 0
None | 0 | 0 | P

Fig. 3.4: Influence diagram representing the total conditional likelihoods for the risk analysis problem. Grey cells with the text stricken out represent infeasible or impossible events. Likelihoods in bold highlight that the conditional likelihood differs from the marginal one in Fig. 3.3.

Additionally, the oddness method of likelihood propagation allows us to replicate conditional probability. For instance, in the incident materialisation node, the marginal likelihood of the event ‘wiper not running’, given the events ‘false alarm’ in the threat exposure node and ‘leave it as is’ in the incident response, is certain. However, its conditional probability is possible, since its materialisation is a chain of a possible event (‘false alarm’) and a certain event (‘wiper not running, given the false alarm and the leaving of the system as is’). This procedure propagates through the diagram. Additionally, when an event can happen through multiple event chains, then the likelihood of the likeliest one is selected. For example, in the impact on asset 5, the event ‘no spill event’ is rare if it comes from the chain with the consequence 1 event ‘disruption’, and it is possible if it comes from the chain with the consequence 1 event ‘correct status’. Since the event is, overall, at least possible, this is the likelihood passed to the child event ‘none’ in the objective C node.

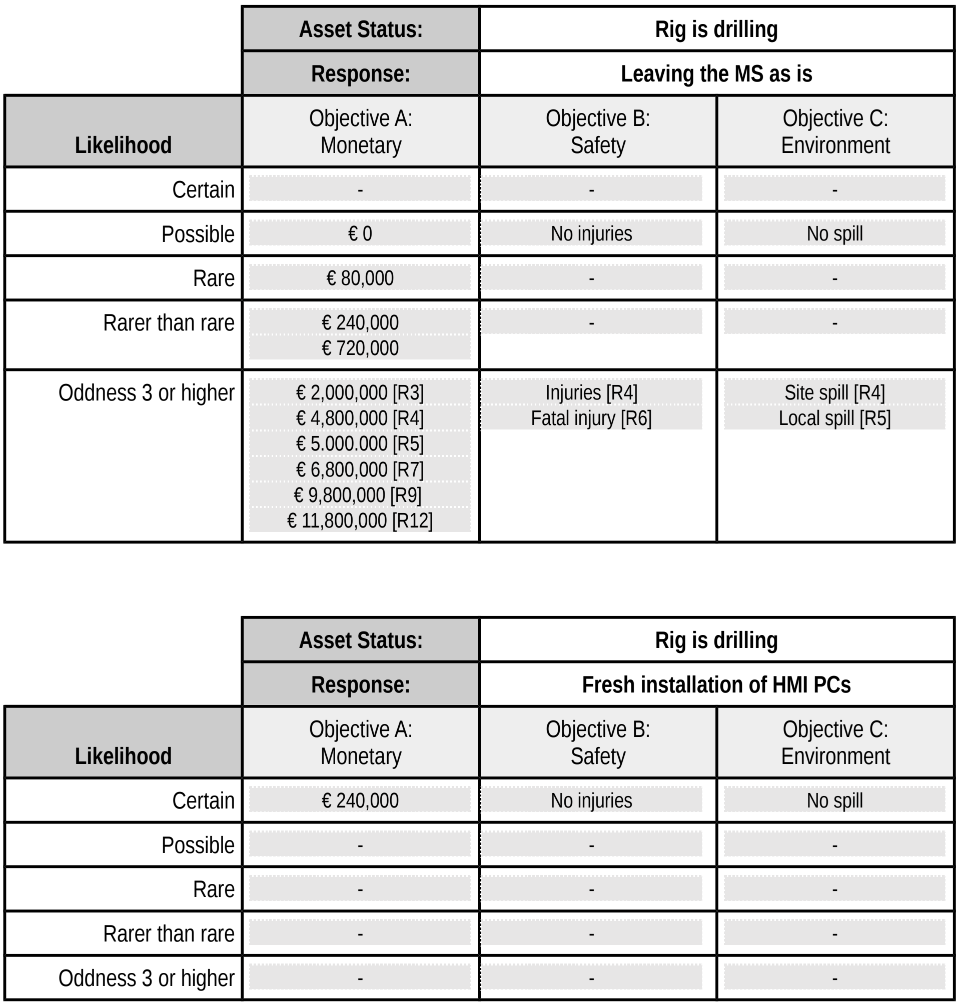
3.4.4 Risk evaluation

From an evaluative point of view risks and, specifically, impacts over value are incommensurable, i.e., they cannot, or ought not, be objectively evaluated in a single severity scale [48]. Therefore, a single scale, like the severity level of risk matrices, leads to a high level of incommensurability. On the other hand, it is recommendable to limit the number of elements to compare to facilitate decision-making. Multiple methods exist for evaluating risk, for instance, if the analyst has time and access to subject-matter experts, it is recommendable to use a method for preference and risk attitude elicitation, e.g. multi-attribute utility theory [140]. The rationality axioms make sense for generating a transparent and logical evaluation of the risk scenarios. Utility functions are flexible enough to represent multiple types of preference and risk attitudes and they offer strong analytical and mathematical properties. In addition, it is possible to avoid re-eliciting preferences as long as there are no changes in preferences.

The outcome of the risk description part is a set of scenarios representing how risk objectives could be affected by an incident, given the incident response. As depicted in Fig. 3.3, we created three objective nodes: monetary, safety and environment. The monetary node synthesises the cost that an incident in the assets might cause. On the other hand, the safety and environment nodes are practically direct translations of their precedent impact on asset nodes, as they have only one parent node.

As a decision problem, risk analysis is undertaken with the purpose of clarifying what are the best options to counter a risky situation. In our case, this involves that the main components to be evaluated are the potential responses of the incident handlers regarding risk objectives.

Tables in Fig. 3.5 display the relevant information that CSIRA presents to the stakeholders so that they are able to compare what different events regarding risk objectives, and their likelihood, might happen if they implement a response. In this case, the alternatives are either assuming a cost €240,000, caused by the lost time of maintaining the MS or face the rare event of losing €80,000, or the rarer than rare events of losing €240,000 or €720,000. If they disregard the even more rare events (oddness 3 or greater), then it seems a simple comparison between a certain lost of €240,000 and a loss three times greater but many more times less likely. However, should the stakeholders take into account the most rare events, then the comparison would become less clear.



	Asset Status:	Rig is drilling	
	Response:	Leaving the MS as is	
Likelihood	Objective A: Monetary	Objective B: Safety	Objective C: Environment
Certain	-	-	-
Possible	€ 0	No injuries	No spill
Rare	€ 80,000	-	-
Rarer than rare	€ 240,000 € 720,000	-	-
Oddness 3 or higher	€ 2,000,000 [R3] € 4,800,000 [R4] € 5.000.000 [R5] € 6,800,000 [R7] € 9,800,000 [R9] € 11,800,000 [R12]	Injuries [R4] Fatal injury [R6]	Site spill [R4] Local spill [R5]

	Asset Status:	Rig is drilling	
	Response:	Fresh installation of HMI PCs	
Likelihood	Objective A: Monetary	Objective B: Safety	Objective C: Environment
Certain	€ 240,000	No injuries	No spill
Possible	-	-	-
Rare	-	-	-
Rarer than rare	-	-	-
Oddness 3 or higher	-	-	-

Fig. 3.5: Tables representing the likelihood of different risk objectives when the incident handlers leave the wiper in the MS (upper table) and when they decide to do a fresh installation of the affected computers (lower table). Events with an oddness of 3 or higher contain their specific likelihood with squared brackets.

3.5 Discussion

We have presented CSIRA, a model for building a high-level cybersecurity incident risk analysis. CSIRA is based on an influence diagram that provides a more comprehensive risk analysis than risk matrices. Realising the fact that risk quantification is practically infeasible in real time, we have implemented an alternative qualitative method that is at least implementable in an influence diagram to follow the basic logic of probability. We have put a special emphasis on what stakeholders value (impact nodes), how to synthesize these impacts over value (objective nodes) and how do stakeholders evaluate potential responses with respect to these risk objectives (risk evaluation). These axiological aspects require, rather than plain business impact scales, decision analysis modelling, so that value aspects are better formalised [70].

We present our method as an alternative to risk matrices rather than to more technical methods like attack or failure trees. Namely for two reasons, matrices use a single severity scale to merge their different categories of impact, in contrast to our approach or a more granular identification of impacts and their synthesis in a reduced number of risk objectives. Additionally, our likelihood elicitation method is as simple as the risk matrices (and it shares its limitations) but is designed to follow probability axioms, so that it could be applied to chains of events.

Regarding our research objectives, CSIRA is our main proposal for a risk analysis model for cybersecurity incidents (RO1). The framework developed in Chapter 4 is more thorough and sophisticated and can be used to model the risk of cybersecurity incidents, however we think the simplifications developed in GIRA and CSIRA could be more useful in context that require a more immediate assessment. Additionally, the tree of cybersecurity objectives provides a more sophisticated way for multi-objective decision-making in cybersecurity (RO3), but the mapping presented in Sect. 3.2.2 helpful for a quick assessment.

Upcoming work shall focus on the implementation of CSIRA. The main aspect is its software implementation. R or Python offer an ideal framework for implementing GIRA or CSIRA with a thorough statistical modelling. Indeed, it can be done in a similar fashion to how we implemented in R the ARA model we introduce in the next chapter. Alternative, a JavaScript implementation would facilitate the creation of an small application to undertake a CSIRA analysis with the simplified method of elicitation, since it can be implemented with simple maths or logical reasoning. Besides the implementation of the influence diagram, that requires graph visualisation packages, it is also important to define a semantic model of CSIRA that captures the input from the users. Additionally, the elicitation method presented here would require a set of functions that transforms the user input (e.g., possible, oddness-1 rare event) into the marginal probabilities of the Bayesian nodes, and a set of functions that transforms the calculated probabilities into the 'oddness' language again. Future work after the implementation shall focus on test-based improvements of CSIRA and the construction of guidelines for its use.

Chapter 4

An adversarial risk analysis framework for cybersecurity

The previous chapters focused on risk analysis during incidents. This chapter focuses on the typical time frame of risk analysis: a planning period such as a year or the lifetime of a system or organisation. Here 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. 4.1 presents our framework, supported by a case study in Sect. 4.2. We conclude with a brief discussion.

4.1 A cybersecurity adversarial risk analysis framework

We introduce our integrated risk analysis approach to facilitate cybersecurity resource allocation. Our aim is to improve current cybersecurity frameworks, introducing schemes that incorporate all relevant parameters, including decision-makers' preferences and risk attitudes [29] and the intentionality of adversaries. Moreover, we consider decisions concerning cyber insurance adoption to complement other risk management alternatives through risk transfer. We present the framework stepwise, analysing the elements involved progressively. We describe the models [12] through influence diagrams (ID) and bi-agent influence diagrams (BAID) detailing the relevant elements: assets, threats, security controls and impacts. At each step, we provide a brief description of the diagrams introduced and a generic mathematical formulation.

4.1.1 System performance evaluation

Fig. 4.1 describes the starting outline for a cyber system under study. c_n designates the costs associated with its operation over the relevant period; they are typically uncertain and modelled with a probability distribution $p(c_n)$. We introduce a utility function $u(c_n)$ over costs to account for risk attitudes [134]. We evaluate system performance under normal conditions, i.e. in absence of relevant incidents, through its associated expected utility $\psi_n = \int u(c_n)\, p(c_n)\; dc_n$ [64]. This scheme can be sophisticated in several directions. For example, there could be several performance functions, leading to a multi-attribute problem, as reflected in the case in Sect. 4.2. A typical example in cybersecurity is to consider attributes concerning information availability, integrity and confidentiality [125].

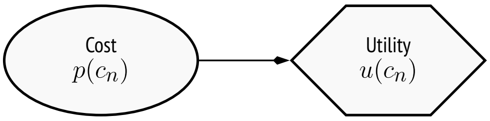


Fig. 4.1: Basic ID for system performance evaluation. c_n indicates costs associated with system operation over the relevant planning period; the utility function $u(c_n)$ accounts for risk attitudes. Note that in IDs, oval nodes represent uncertainties modelled with a probability distribution $p(\dots)$, and utility nodes represent preferences modelled with an utility function $u(\dots)$.

4.1.2 Cybersecurity risk assessment

Based on Fig. 4.1, we consider the cybersecurity risk assessment problem in Fig. 4.2. In general, we include m threats $t_1, \dots, t_m$; some of them could be physical (e.g., a fire) and others cyber (e.g., a DDoS attack[1]). Their occurrence are random variables. We also include l types of assets; some of them could be traditional (e.g., facilities) and others could be cyber (e.g., information systems). Impacts on them will be, respectively, designated $c_i, i = 1, \dots, l$ and are typically uncertain. If the impacts are conditionally independent given the threats, the corresponding model would be of the form $p(c_1|t_1, \dots, t_m) \dots p(c_l|t_1, \dots, t_m)\, p(t_1, \dots, t_m)$, where $p(t_1, \dots, t_m)$ describes the

[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, rendering it inaccessible.

probability of the threats happening[2], and $p(c_i|t_1,\ldots,t_m)$ describes the probability of impact on the i-th asset, given the occurrence of various threats. We aggregate costs additively at the total cost node c. Then, the expected utility would be

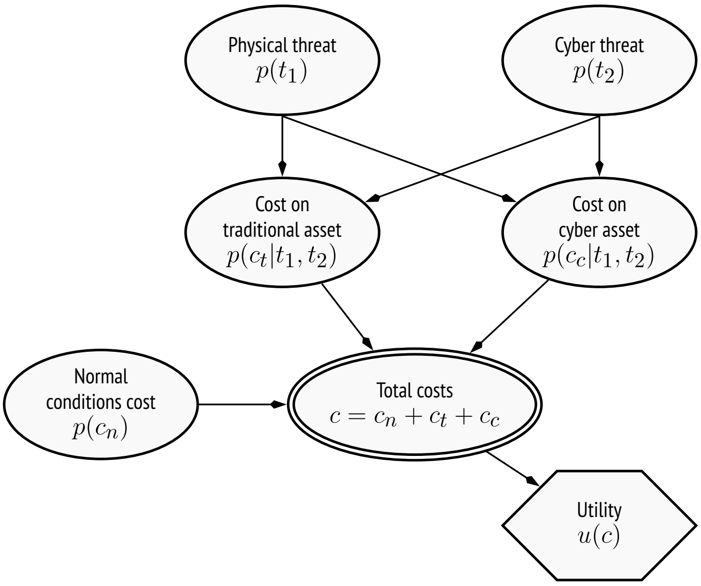


Fig. 4.2: Cybersecurity risk assessment. The threats (two in this example, t_1 and t_2) might impact on the organisation's assets, causing costs (two in this example, c_t and c_c). These costs, and those under normal conditions c_n, are aggregated to determine the total costs c and evaluated through the utility function $u(c)$. Recall that in IDs, double-lined ovals represent deterministic aspects.

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

We have assumed that consequences are additive, but we could have a generic utility $u(c_n, c_1, \ldots, c_l)$. Finally, we evaluate the loss in expected utility $\psi_n - \psi_r$. Alternatively, we could compare the difference in the corresponding certain equivalents [63]. When such difference is sufficiently large, incidents are expected to harm the system significantly and we should manage such risks. Note that we could incorporate several utility nodes to describe multiple stakeholders' preferences.

[2] Depending on the problem we could have further decompositions. For example, in a case like that in Fig. 4.2 with independent threats we would have $p(t_1,\ldots,t_m) = \prod_{i=1}^{m} p(t_i)$.

As a next step, we add security controls. We introduce a portfolio of them to reduce the likelihood of threats or their impact. Examples include firewalls, employee training, or making regular backups. For simplicity, in Fig. 4.3 we assume that all controls have influence over all events and impacts. It will not always be so: a fire detector makes less harmful, but not less likely, a fire; resource accounting mechanisms [123] managing access based on user privileges make a successful DDoS attack less likely, but usually not less harmful. Node e describes the portfolio of controls, whose cost we model through the distribution $p(c_e|e)$. Controls might have influence on threat likelihoods $p(t_i|e)$, $i = 1,\dots,m$, as well as on asset impact likelihoods $p(c_i|t_1,\dots,t_m,e)$. We aggregate all costs through the total cost node c, under appropriate additivity assumptions. In this case, the organisation's expected utility when we implement portfolio e is

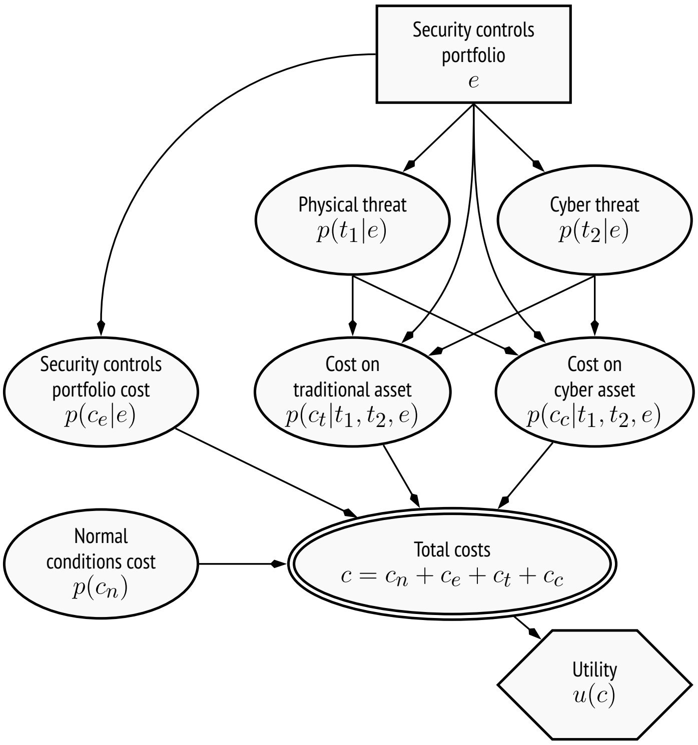

Security controls portfolio e
Physical threat $p(t_1|e)$
Cyber threat $p(t_2|e)$
Security controls portfolio cost $p(c_e|e)$
Cost on traditional asset $p(c_t|t_1,t_2,e)$
Cost on cyber asset $p(c_c|t_1,t_2,e)$
Normal conditions cost $p(c_n)$
Total costs $c = c_n + c_e + c_t + c_c$
Utility $u(c)$

Fig. 4.3: Cybersecurity risk management. We add to Fig. 4.2 the security controls portfolio e (and its cost c_e) that the organisation can implement to mitigate the threats or their impacts. Recall that rectangle nodes represent decisions.

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

We would then look for the maximum expected utility portfolio by solving $\psi_e^* = \max_{e\in E} \psi(e)$, where E is the set of feasible portfolios, which should satisfy incumbent constraints like economic (e.g., not exceeding a budget), legal (e.g., complying with data protection laws), logistic or physical.

4.1.3 Risk transfer in cybersecurity risk management: cyber insurance

As a relevant element of increasing interest, we introduce the possibility of acquiring a cyber insurance product. Its cost will typically depend on the implemented portfolio of controls, as in Fig. 4.4: the better such a portfolio is, the lower the insurance premium would be. This cost will also depend on the assets to be protected. We could include the insurance within the portfolio of controls; however, it is convenient to represent them separately, since premiums will usually depend on the controls deployed. The decision node i describes the cyber insurance adopted, with entailed costs c_i with probability $p(c_i|i,e)$, although they will usually be deterministic. In addition, insurance and security controls will affect impacts, modelled through $p(c_j|t_1,\ldots,t_m,e,i), j = 1,\ldots,l$. The total cost node c aggregates the costs. The expected utility when we implement portfolio e together with insurance i is

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

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

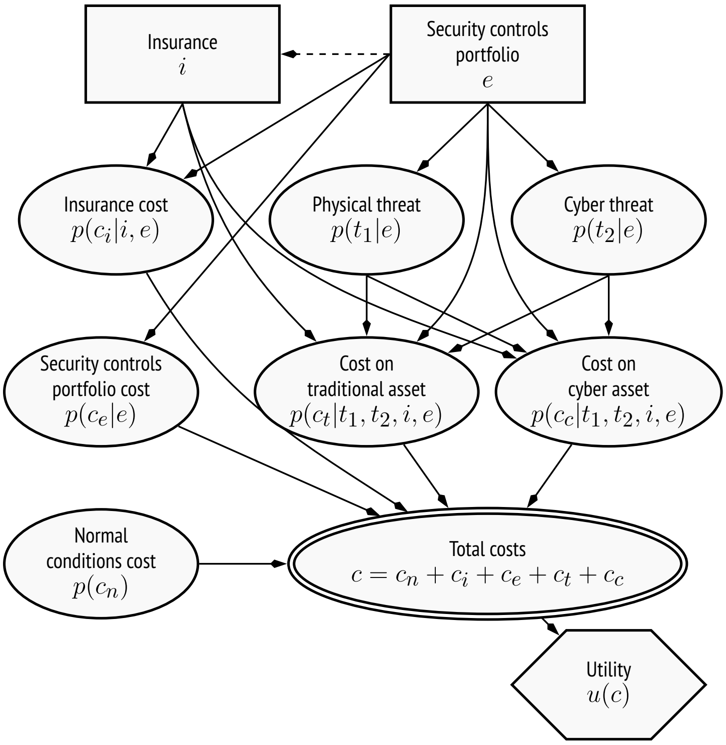


Fig. 4.4: Cyber insurance for cybersecurity risk management. We add to Fig. 4.3 the insurance i (and its cost c_i) that the organisation can subscribe to mitigate the impacts that the threats can cause.

4.1.4 *Adversarial risk analysis in cybersecurity*

As discussed, intentionality is a key factor when analysing cyber threats. As an example, the ISF [79] specifies a group of several adversarial threats within its catalogue. We use ARA [12] to model the intentions and strategic behaviour of adversaries in the cybersecurity domain, see Merrick and Parnell [120] for a comparison of various methods modelling adversaries in risk management. Under ARA, the attacker has his own utility function and seeks 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.

Defence-attack model

The original examples, Figs. 4.2 and 4.3, evolve into Fig. 4.5, modelling an adversarial case through a BAID with a Defender and an Attacker. The unintentional threat remains modelled through a probabilistic node, whereas we model the adversarial threat through a decision node for the Attacker, who needs to decide whether or not to launch an attack to his benefit. For simplicity, in the diagram we model the physical threat t_1 as unintentional and the cyber threat a as adversarial, although adversarial physical threats and unintentional cyber threats could be relevant in certain cases, as exemplified in the case study. Also for simplicity, we only consider one attacker and one attack, but the ideas extend to multiple attacks by one attacker or to multiple attackers.

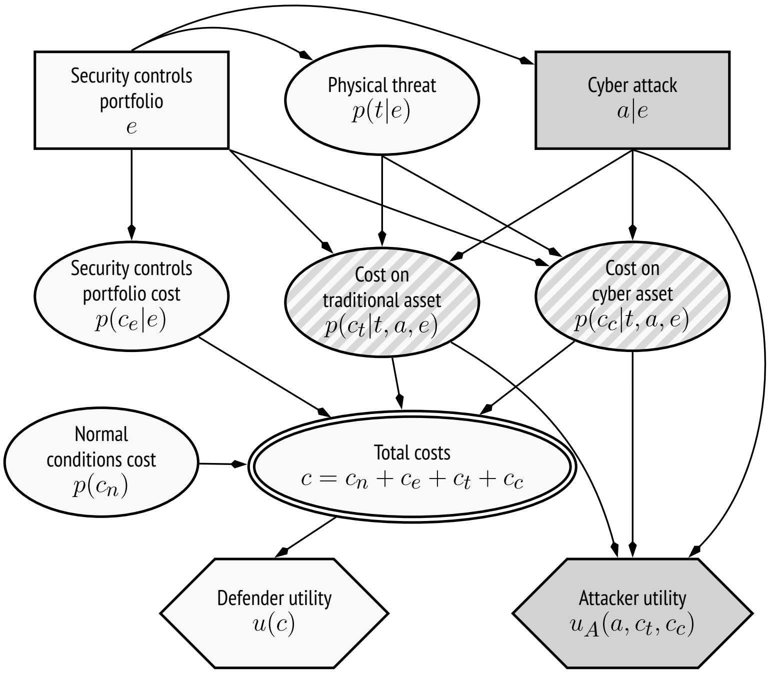


Fig. 4.5: Adversarial risk analysis in cybersecurity: defence-attack problem. We modify Fig. 4.3 by transforming the cyber threat into an adversarial one: an attacker is deciding whether to attack the organisation (a) based on his own evaluation, $u(a, c_t, c_c)$, of the harm caused to the organisation and the cost of performing the attack. Lighter nodes refer to issues concerning solely the Defender; darker nodes refer to issues relevant only for the Attacker; nodes with stripped background affect both agents. Arcs have the same interpretation as in [152].

We present a sequential defence-attack template model for cybersecurity. For the Defender problem, this converts the Attacker's decision nodes into chance ones and eliminates the Attacker's nodes not affecting it, as well as the corresponding utility node. Similarly for the Attacker, where we assume here that there is only one Attacker responsible of the adversarial threat a independent of the other threats, given the portfolio e. Fig. 4.3 essentially presents the Defender problem and we covered its resolution in Sect. 4.1.2. The cyber attack is described probabilistically[3] through $p(a|e)$, which represents the probability that the Defender assigns to cyber threat a materialising, had portfolio e been adopted. However, given the strategic nature of this problem, rather than using a standard probability elicitation approach [40], we greatly facilitate and improve the assessment of the required distribution if we analyse the Attacker decision about which attack to perform, as argued in Rios Insua, Banks, Rios and Ortega [143]. Under the ARA paradigm, the Defender should analyse the Attacker strategic problem in Fig. 4.6. Specifically, given portfolio e, and assuming that the Attacker maximises expected utility, the Defender would compute, for each attack a, the expected utility for the Attacker

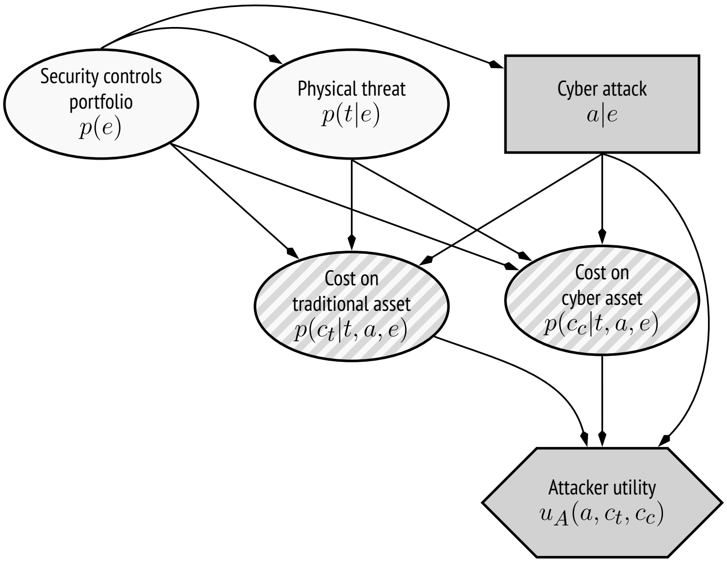


Fig. 4.6: Attacker problem in the defence-attack model.

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

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

where u_A and p_A designate, respectively, the utility and probabilities of the Attacker. The Defender must then find the attack solving $\max_{a \in A} \psi_A(a|e)$, where A is the set of attack options. However, the Defender will not typically know u_A and p_A. Suppose we are capable of modelling her uncertainty about them with random probabilities P_A and a random utility function U_A [12]. Then, the optimal random attack, given e, is

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

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

$$F = \Big(U_A(a, c_1, \ldots, c_l), P_A(c_1|t_1, \ldots, t_m, a, e), \ldots, P_A(c_l|t_1, \ldots, t_m, a, e), P_A(t_1, \ldots, t_m|e)\Big).$$

Defence-attack-defence model

Cybersecurity risk management also comprises reactive measures that can be put in place to counter an attack, should it happen. Therefore, we split the security portfolio into two groups: preventive e_p and reactive $e_r|t_1, \ldots, t_m, a$ security controls, as in Fig. 4.7. This corresponds to our sequential defence-attack-defence template model in which the first move is by the Defender (preventive portfolio e_p), the second one is by the Attacker (attack after observing preventive controls, $a|e_p$) and the third one is by the Defender (reactive portfolio $e_r|t_1, \ldots, t_m, a$). we solve the Defender problem much as we did in Sect. 4.1.2, reflecting changes caused by splitting the security control node. Specifically, the expected utility when portfolio $e = (e_p, e_r)$ is implemented is

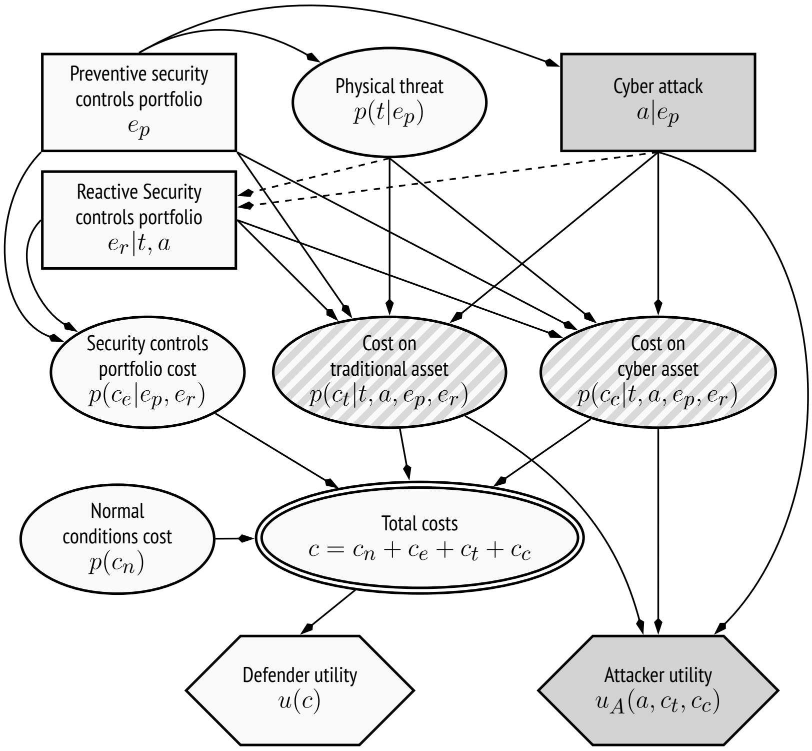

Preventive security controls portfolio e_p
Physical threat $p(t|e_p)$
Cyber attack $a|e_p$
Reactive Security controls portfolio $e_r|t,a$
Security controls portfolio cost $p(c_e|e_p,e_r)$
Cost on traditional asset $p(c_t|t,a,e_p,e_r)$
Cost on cyber asset $p(c_c|t,a,e_p,e_r)$
Normal conditions cost $p(c_n)$
Total costs $c = c_n + c_e + c_t + c_c$
Defender utility $u(c)$
Attacker utility $u_A(a,c_t,c_c)$

Fig. 4.7: Adversarial risk analysis in cybersecurity: defence-attack-defence problem.

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

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 above represents a global view of the sequential problem, although we solve this kind of two-stage problems sequentially, as in He and Zhuang [73]. We would solve the Attacker problem providing $p(a|e_p)$ in a similar fashion as in Section 4.1.4.

4.2 A case study template

We illustrate our cybersecurity risk analysis framework with a defence-attack case study, which can serve as a template for more complex problems. For confidentiality reasons, we have simplified the number of relevant issues and masked the data conveniently. This simplification will also allow us to better illustrate key modelling concepts and the overall scheme. Moreover, we include uncertain phenomena in which data are abundant and others in which it is not and, thus, we shall need to rely on expert judgment for its quantification [40]. The Defender is a SME[4] with 60 people and 90 computers. A cyber attack might affect its online services. Prices and rates in € refer to Spain, where the incumbent organization is located.

In essence, we first structure the problem identifying assets, threats and security controls. The latter may have implementation costs in exchange for reducing the threat likelihoods and/or possible impacts. Subsequently, we assess the impacts that may have an effect on asset values to find the optimal risk management portfolio. Since we include adversarial threats, we consider the Attacker decision problem. In this case there is a single potential Attacker which contemplates a DDoS attack with the objective of disrupting the Defender services, causing an operational disruption and reputational damage and the consequent loss of customers, besides incurring in contractual penalties potentially affecting its continuity. Then, we simulate from this problem to obtain the attack probabilities, which feed back into the Defender problem to obtain the optimal defence. We focus 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. 4.2.5. We consider a one-year planning horizon.

4.2.1 *Problem structuring*

We structure the problem through the BAID in Fig. 4.8 and describe its components next.

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*, the offices potentially affected by threats; *Computer equipment*, the data centre and workstations of the organisation; *Market share*. Other assets not considered in this case include, e.g., the com-

[4] Small and medium-size enterprise

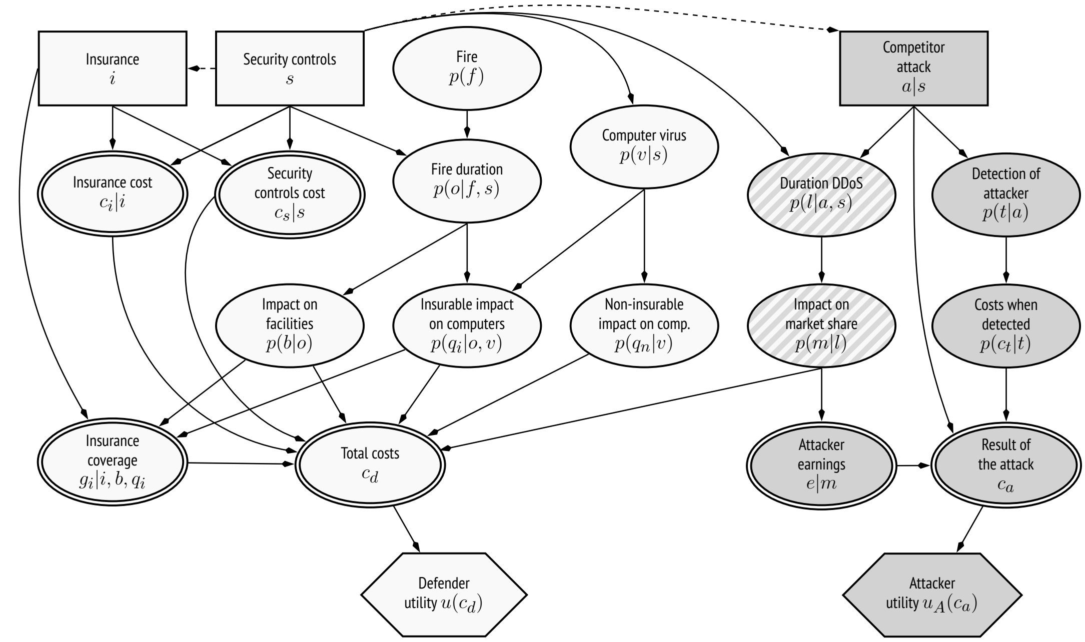

Insurance i
Security controls s
Fire $p(f)$
Competitor attack $a|s$
Computer virus $p(v|s)$
Insurance cost $c_i|i$
Security controls cost $c_s|s$
Fire duration $p(o|f,s)$
Duration DDoS $p(l|a,s)$
Detection of attacker $p(t|a)$
Impact on facilities $p(b|o)$
Insurable impact on computers $p(q_i|o,v)$
Non-insurable impact on comp. $p(q_n|v)$
Impact on 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 $u_A(c_a)$

Fig. 4.8: Case study as a BAID.

pany's development software, its business information, its mobile devices or the staff.

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 model each threat with a probabilistic node associated with the Defender problem. We extract two threats from the MAGERIT [122] catalogue: fire and computer virus. A *fire* may affect facilities and computers; we do not contemplate impact on market share, as the organisation has a backup system; we assume that a fire can occur only by accident, not by sabotage. The *computer virus* is aimed at disrupting normal operations of computer systems; we consider this threat non-intentional, as most viruses propagate automatically: their occurrence tends to be random from the Defender perspective. Other non-intentional threats, not considered here, could be water damage, power outages or employee errors.

Intentional threats. This category may include both cyber and physical threats. Again, we could use catalogues from, e.g., ISF [79]. We should first identify the attackers. We then integrate the attack options available to each attacker within a single decision node. In our case, we just consider one competitor, reflected in the *competitor attack* node. He may attempt a DDoS to undermine the availability of the Defender site, compromising their customer services. For this, he must decide whether to launch the attack and the number of attempts. Other intentional attacks, not modelled here, could include launching an advanced persistent threat, instigating the misbehaviour of insiders or the use of bombs.

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

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

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

Insurance. We also consider the possibility of purchasing insurance to transfer risk with the corresponding Defender decision node. The premium will depend on the protected assets and contextual factors such as location, company type and, quite importantly, the implemented controls. Table 4.1 displays the contemplated insurance products.

Impacts on Defender. Having identified the threats, we present their relevant impacts on the Defender's assets. We model each of them with a probabilistic node. We consider: *Impact on facilities*, the monetary losses caused by fire over them; *Impact on computers*, the monetary losses caused by fire or viruses split into insurable and non-insurable ones to assess the possible insurance coverage; *Impact on market share*. We also consider the impacts as-

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

Table 4.1: Insurance product features, some of them referring to cyber impacts.

sociated with safeguards as deterministic nodes: *cost of security controls*, *cost of insurance* and *insurance coverage*. Finally, a deterministic *total costs* node aggregates the Defender's consequences to establish the final impact in the Defender problem. Besides, we could include other types of impacts such as the corporate image or the staff safety, although we do not do it here.

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

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

Defender and Attacker problems. Figs. 4.9 and 4.10, respectively, represent the Defender and Attacker problems derived from the strategic problem in Fig. 4.8. We use both diagrams to guide judgment elicitation.

4.2.2 Assessing the Defender non-strategic beliefs and preferences

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

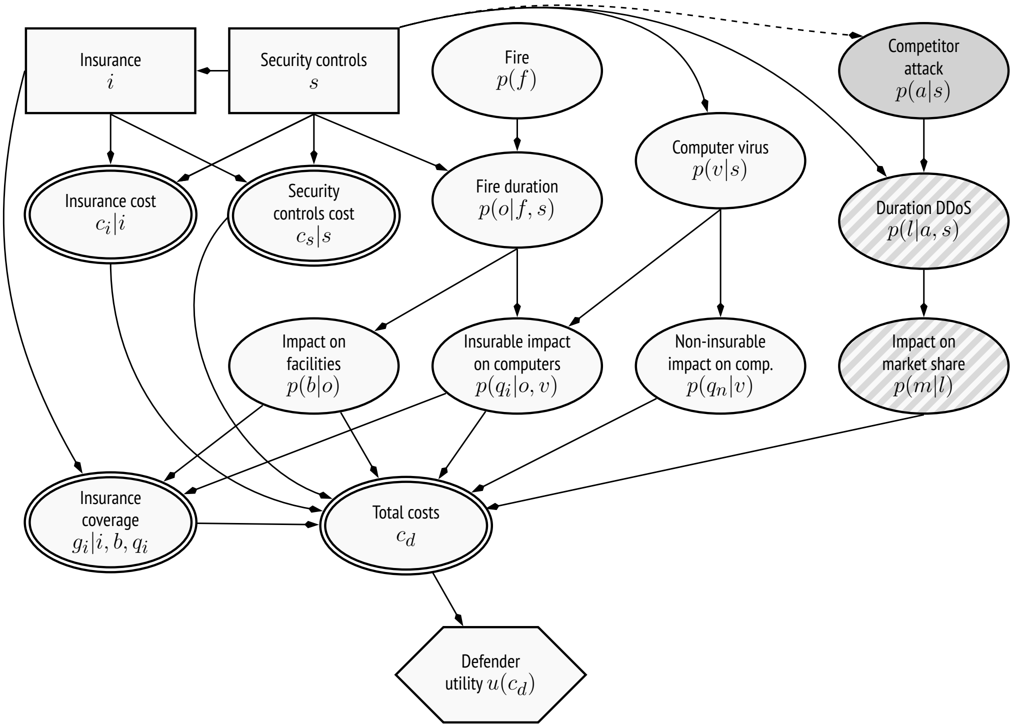


Fig. 4.9: Defender problem.

ical lack of data in cybersecurity environments [76]. As a consequence, we populate most nodes in the model. Sect. 4.2.3 treats nodes that require strategic analysis. Finally, Sect 4.2.4 analyses 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.

Economic value of Defender assets

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

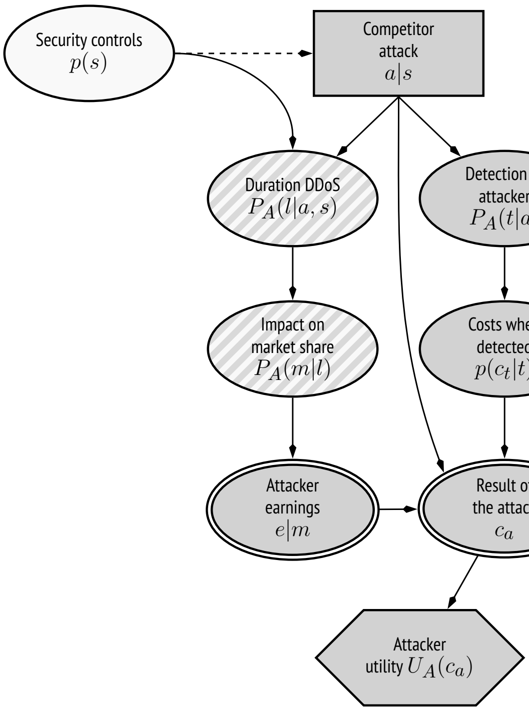

Security controls $p(s)$
Competitor attack $a|s$
Duration DDoS $P_A(l|a,s)$
Detection of attacker $P_A(t|a)$
Impact on market share $P_A(m|l)$
Costs when detected $p(c_t|t)$
Attacker earnings $e|m$
Result of the attack c_a
Attacker utility $U_A(c_a)$

Fig. 4.10: Attacker problem.

Modelling security controls

Security controls decision s. The security portfolios that the Defender could implement derive from the options in Sect. 4.2.1. For the DDoS protection we have the choice of not implementing it or subscribe to a 2, 5, 10 or 1000 gbps service. For the other security controls, the choice is binary. We thus have 40 portfolios which could be constrained by, e.g., a budget, as in Sect. 4.2.5.

Cost of security controls $c_s|s$. Table 4.2 provides them, from which we derive those of the portfolios.

<table>
<tr><th>Security control</th><th>Cost</th></tr>
<tr><td>Anti-fire system</td><td>€ 1,500</td></tr>
<tr><td>Firewall</td><td>€ 2,250</td></tr>
<tr><td>Risk mitigation procedures</td><td>€ 2,000</td></tr>
<tr><td>Cloud-based DDoS protection</td><td><table><tr><td>2 gbps</td><td>5 gbps</td><td>10 gbps</td><td>1000 gbps</td></tr><tr><td>€ 2,400</td><td>€ 3,600</td><td>€ 4,800</td><td>€ 12,000</td></tr></table></td></tr>
</table>

Table 4.2: Cost of individual security controls.

Modelling the insurance product

Insurance decision i. This refers to the insurance product that the Defender could purchase (Table 4.3) once the controls have been selected.

Insurance cost, $c_i|i$. It depends on the controls implemented by the organisation (Table 4.3).

Insurance coverage $g_i|i,b,q_i$, as reflected in Table 4.1.

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 4.3: Insurance product cost.

Modelling the fire risk

Likelihood, $p(f)$. This node provides the annual probability of suffering a fire. We use data from Vitoria [37], concerning fire interventions on industrial buildings (Table 4.4). As the fire rate remains fairly stable over time, we estimate such probability with a beta-binomial model with beta prior $\beta e(1/2,1/2)$. The posterior would be $f|\text{data} \sim \beta e\Big(1/2+\sum_{i=1}^{5} x_i, 1/2+\sum_{i=1}^{5}(n_i - x_i)\Big) \equiv \beta e(147.5, 6320.5)$, with x_i the number of fires affecting industrial buildings and n_i the number of buildings in the $i-$th year, $i = 1,\ldots,5$. As the posterior variance is small, such distribution can be reasonably summarised through its posterior expectation, $\hat{p} = 0.022$. The number f of fires can be approximated with a Poisson $\mathscr{P}(0.022)$ distribution. However, we consider only the probability that one fire occurs, since $Pr(f > 1) = 0.00024$. Thus, $f \sim \min[1, \mathscr{P}(0.022)]$.

Duration, $p(o|f,s)$. It is a major fire impact determinant [10]: the longer the fire, the more damaging it will be. Fig. 4.11 presents the histogram of industrial fire durations, with mode $[30,60]$ minutes. Adapting Wiper, Rios Insua and Ruggeri [173], we model the fire duration o with a gamma $\Gamma(\text{shape} = \gamma, \text{scale} = \gamma/\mu)$ distribution. We assume a non-informative, but proper, ex-

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

Table 4.4: Industrial fire data in Vitoria (2005-2009).

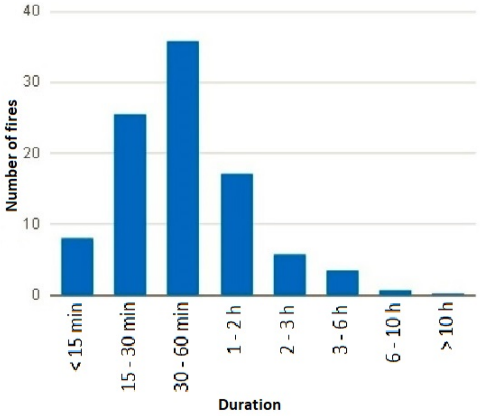


Fig. 4.11: Industrial fire duration histogram. Vitoria, Spain (2005-2009).

ponential prior for $\gamma \sim \mathscr{E}(0.01)$ and inverse gamma for $\mu \sim$ Inv-$\Gamma(1,1)$. No expression for the posterior distribution is available, but we can introduce a Markov chain MC scheme to sample μ and γ from the data. Based on it, we estimate that E$(\gamma|data) \approx 0.85$ and E$(\mu|data) \approx 78$.

The only proposed control that may have an effect over fire duration is the anti-fire system. Using expert judgment (Dias et al., [40]), 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 [66], we consider a triangular distribution with quantiles 0.05 at 1 and 0.95 at 60 min, resulting in a triangular distribution $Tri(0.8, 63, 10)$, which models o if there is a fire $(f = 1)$ and the portfolio s contains the anti-fire system. On the other hand, $o \sim \Gamma(0.85, 0.0109)$ if the portfolio does not contain the anti-fire system.

Impact. We assume that the amount lost is linearly related to the fire duration. After consulting with experts, we consider that a fire lasting 120 minutes would degrade the facilities by 100% in the absence of controls. To simplify, we assume that the effect of fire duration is linear. Additionally, the impact on computer equipment derives from the percentage of facility degradation caused by fire. Assuming that computers are evenly distributed through the premises, a fire lasting 120 minutes would also degrade computer equipment by 100%. This impact is potentially insurable and will be modelled in Sect. 4.2.2.

Modelling the computer virus risk

Likelihood, $p(v|s)$. This node provides the number v of virus infections during a year. The distribution of the number of infected computers in a month follows a binomial distribution $\mathscr{B}(h,q)$, with q the probability that a computer gets infected and h the number of computers. Various statistics suggest that the rate of virus infections worldwide is 33% [135], so we estimate $\hat{q} = 0.33$. The organisation has 90 computers, which we assume have the same security controls and are equally likely to be infected. Since the analysis is for 12 months, we use $h = 12 \cdot 90 = 1080$. Additionally, we consider the effect of our controls: if a firewall is implemented, the probability that a computer gets infected reduces to $\hat{q} = 0.005$, not completely eliminating the threat, even if this includes continuous updating based on the latest virus signatures; if the mitigation procedures are implemented, the infection probability reduces by 50%, with firewall or not, as this control entails improvements in the organisation such as imposing safety requirements on acquired systems. The number v of infections is, therefore, modelled as in Table 4.5.

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

Table 4.5: Number v of annual virus infections.

Impact. Viruses may impact the integrity and availability of computers, leading to information corruption or unavailability. Impacts on confidentiality are variable, as they depend on the stolen information. The average daily cost of these infections was estimated at € 2.683 [155], although this may vary depending on the monetary value of the information and services that the victim systems support. Bigger losses come from sophisticated campaigns (e.g., as with WannaCry) or targeted malware which, under our paradigm, we would better model as an adversarial threat. In our case, repairing a computer infected by a virus requires € 31, for two technician hours. Insurance options cover the removal of computer viruses. Therefore, we cover this impact within the insurable aspects in Sect. 4.2.2.

Besides, most viruses entail performance reduction in aspects such as initialisation of operating systems. Although small, this causes time losses to the user. We assume that most (70%) of the work time of the organisation is in front of a computer and that it would take, on average, 40 hour to detect the problem. We therefore assume that when a computer is infected, 28 hour of its usage are affected by the virus. We model the time loss w with a uniform $\mathscr{U}(0, 0.05)$ distribution which represents that the percentage of lost time caused by a virus is between 0 and 5 %. The average hourly cost of the employees is € 20/hour. Therefore, for each virus infection, the cost would be $20 \times 28 \times w$. Our insurance options do not cover this loss and, thus, we model it within the non-insurable aspects in Sect. 4.2.2.

Modelling the DDoS threat

We consider now the non-strategic aspects of the DDoS threat.

Duration, $p(l|a,s)$. The duration l in hours of a successful DDoS attack will depend on the intensity of the attacking campaign, how well crafted the attack is and the security controls implemented. An emerging type of controls are cloud-based systems absorbing traffic when a site becomes a victim of a DDoS. If no control is deployed, it would be virtually impossible to block such attack. Based on [98] and [172], 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 $\Gamma(4,1)$. This duration is conditional on whether the attack actually saturates the target, which depends on the capacity of the DDoS platform minus the absorption by the cloud-based system. We assume that the Attacker uses a professional platform capable of 5 gbps attacks, modelled through a $\Gamma(5,1)$ distribution. We then subtract the traffic s_{gbps} absorbed by the protection system to determine whether the attack is successful (its traffic overflows the protection system). Since the campaign might take a attacks, the output of this node is $l = \sum_j^a l_j$, with $l_j \sim \Gamma(4,1)$ if $\Gamma(5,1) - s_{\text{gbps}} > 0$, and $l_j = 0$, otherwise.

Impact. A DDoS attack might cause a reputational loss that would affect the organisation's market share. We assume that all market share is lost at a linear rate until all value is gone, say, after 5-8 days of unavailability: in the fastest case, the loss rate r would be $0.5/120 = 0.00417$ per hour, whereas in the slowest one it would be 0.0026. We model r as a $\mathscr{U}(0.0026, 0.00417)$.

Modelling impacts on the assets

We recollect here the impacts on the assets.

Impact on facilities, $p(b|o)$. The monetary loss b due to degradation of facilities through fire is $b \sim 5000000 \times \min\left(1, \frac{o}{120}\right)$, following Sect. 4.2.2.

Insurable impacts on computers, $p(q_i|o,v)$. We model the monetary losses q_i due to degradation of computers covered by an insurance, either caused by fire, Sect. 4.2.2, or through repairing computers infected with viruses, Sect. 4.2.2, as $q_i \sim 31v + 200000 \times \min\left(1, \frac{o}{120}\right)$.

Non-insurable impacts on computers, $p(q_n|v)$. The monetary losses q_m caused by degradation of computers due to the lost time caused by viruses are not covered by insurance. Following Sect. 4.2.2, we model $q_n \sim 560w \times v$.

Impact on market share, $p(m|l)$. The monetary loss m due to a reduced market share, following Sect. 4.2.2, is $m \sim \min[1500000, 3000000 \times l \times r]$.

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

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 calibrate it with three costs: worst, best, and an intermediate one. The worst reasonable loss is based on the sum of all costs and impacts (except that due to the computer virus) € 6,755,300. Computer virus impacts do not have an upper limit; based on simulations, it is reasonable to assume that they would not exceed € 50,000. Giving an additional margin, we assume that such maximum is 7000000. The best loss is 0. For the intermediate cost $c_d^* = 2660000$, we find its probability equivalent α so that $u(c_d^*) = \alpha$ (Ortega et al., [134]); based on information provided by the company, $u(c_d^*) \simeq .5$. We rescale the costs to the (0,1) range through $1 - \frac{c_d}{7000000}$. Then, the utility function is $u(c_d) = \frac{1}{\mathrm{e}-1}\left[\exp\left(1 - \frac{c_d}{7000000}\right) - 1\right]$.

4.2.3 Assessing the Attacker's random beliefs and preferences

In the Defender problem, the competitor attack is described through a probabilistic node modelling the number of attacks launched by the Attacker given the security controls that are implemented. To obtain the corresponding probabilities, we model the Attacker problem based on Fig. 4.10. Its solution would provide the Attacker's optimal action. However, as argued in Sect. 4.1.4, we model our uncertainty about his preferences and beliefs through random utilities and probabilities to find the random optimal attack; for this, we simulate from it to forecast its actions and obtain the required probability distribution.

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

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

Duration of the DDoS, $P_A(l|a,s)$. We base our estimate on that of the Defender (Sect. 4.2.2). We model the length of the j-th individual DDoS attack as a random gamma distribution $\Gamma_{\text{length}}(\upsilon, \upsilon/\mu)$ with $\upsilon \sim \mathscr{U}(3.6, 4.8)$ and $\upsilon/\mu \sim \mathscr{U}(0.8, 1.2)$, adding some uncertainty around its average duration (between 3 and 6 hours) and dispersion. Similarly, we model the attack gbps through a random gamma distribution $\Gamma_{\text{gbps}}(\omega, \omega/\eta)$ with $\omega \sim \mathscr{U}(4.8, 5.6)$ and $\omega/\eta \sim U(0.8, 1.2)$. Next, we subtract s_{gbps} from Γ_{gbps} to determine whether the DDoS is successful. As in Sect. 4.2.2, we use $l = \sum_j^a l_j$, with $l_j \sim \Gamma_{\text{length}}$ if $\Gamma_{\text{gbps}} - s_{\text{gbps}} > 0$, and $l_j = 0$ otherwise.

Impact on market share, $P_A(m|l)$. We base our estimate on that of the Defender (Sect. 4.2.2), adding some uncertainty. The market share value and percentage are not affected by uncertainty, as this information is available to both agents. However, we model uncertainty in the market loss rate: the fastest one (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 is $m \sim \min\left[1500000, 3000000 \times l \times R\right]$ with $R \sim \mathscr{U}(\alpha, \beta)$, $\alpha \sim \mathscr{U}(0.0021, 0.0031)$ and $\beta \sim \mathscr{U}(0.00367, 0.00467)$.

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

Attacker Detection, $P_A(t|a)$. This represents the chance of the Attacker being detected. Detection probabilities are estimated via expert judgment at 0.2%, should the Attacker attempt a DDoS attack. Should there be a attacks, the detection has a binomial distribution $\mathscr{B}(a, 0.002)$. To add some uncertainty, we model the detection probability for each attack through a $\beta e(2, 998)$[5]. Then, we model the attacker's detection t through a random binomial distribution that outputs *detected* if $\mathscr{B}(a, \phi) > 0$ with $\phi \sim \beta e(2, 998)$, and *not detected*, otherwise.

Cost for Attacker when detected, $p_A(c_t|t)$. As a competitor, if the Attacker is detected, he would face a serious discredit, together with compensation and legal costs as well as criminal responsibilities. We use this cost decomposition: € 550,000 of *expected reputational costs*, due to the communication actions required to preserve credibility; € 30,000 of *expected legal costs*; € 350,000 of *expected civil indemnities and regulatory penalties*; € 1,500,000 of *expected suspension costs*, related with losses derived from prohibition to operate for some time. We add some uncertainty modelling the cost as a normal distribution with mean 2430000 and standard deviation 400000, i.e., $c_t|t \sim \mathscr{N}(2430000, 400000)$.

Result of attack, $c_a|e, c_t, a$. This combines the attacker earnings and costs if detected, and those of undertaking the attacks. We consider that using a botnet to launch the DDoS attack would cost on average around € 33 per hour [77] (€ 792 for a day). Therefore, $c_a = e - c_t - 792a$.

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

Simulating the Attacker problem

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

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

[5] Its mean is 0.002

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

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

To approximate it, we use an MC approach as in Algorithm 1 (Appendix A) with $K = 20000$, which we have implemented in R. For each size s of the DDoS protection system, we assess the distribution of the random optimal attack. Table 4.6 displays the attack probabilities, conditional on the protection implemented. For instance, if the security portfolio does not contain a DDoS-protection system ($s = 0$, none), an attack seems certain, its duration would be between 18 and 30 attacks, being 29 and 30 the most likely attack sizes. We thus create the probability distribution $p(a|s)$. We have now fully specified the Defender problem and are ready to solve it.

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

```
For each defence e
    For i = 1, ..., K
        Generate
            $\left(U_A^i(t_2,c_t,c_c), P_A^i(c_t|t_1,t_2,e), P_A^i(c_c|t_1,t_2,e), P_A^i(t_1|e)\right) \sim F$
        Compute
            $a^{*i} = \arg\max_a \iiint U_A^i(a,c_t,c_c)\, P_A^i(c_t|t_1,a,e)\, P_A^i(c_c|t_1,a,e)\, P_A^i(t_1|e) dt_1\, dc_c\, dc_t$
    end
    Approximate
        $\hat{p}_A(a|e) = \frac{\#\{a^{*i} = a\}}{K}$
end
```

4.2.4 Solving the Defender problem

Summarising earlier assessments about the Defender problem, we have that the involved distributions are

DDoS	**Number of attempts**															
prot. system	**0**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**
1000 gbps	1.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
10 gbps	0.000	0.001	0.003	0.003	0.004	0.005	0.012	0.012	0.015	0.013	0.017	0.024	0.024	0.022	0.030	0.035
5 gbps	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.001	0.001	0.002
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**
1000 gbps	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
10 gbps	0.026	0.041	0.025	0.044	0.042	0.053	0.050	0.048	0.047	0.060	0.050	0.059	0.065	0.081	0.089
5 gbps	0.008	0.006	0.012	0.017	0.007	0.028	0.031	0.055	0.070	0.061	0.096	0.117	0.143	0.141	0.203
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 4.6: Conditional probability table for random optimal attacks.

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

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

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

The *optimal resource allocation* is the maximum expected utility pair $(s^*, i^*) = \arg\max_{s,i} \psi(s,i)$. We use Algorithm 2 to approximate the portfolio together with the optimal portfolio.

We have implemented it in R with an MC sample size of $K = 20000$ and results summarised in Table 4.7. The best portfolio consists of *a 1000 gbps cloud-based DDoS protection system, a firewall, an anti-fire system*, and *the comprehensive insurance*. Besides the ranking of countermeasures, we can obtain additional information from the simulation. For instance, the best portfolios tend to include a firewall, a 1000 gbps DDoS protection with no risk management procedure. The best portfolios also include insurance, either traditional or comprehensive.

4.2.5 *Further analysis*

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

Algorithm 2 Approximation of Defender′s optimal portfolio.

```
$\psi(s,i) = 0$
For each $(s,i)$
    For $j = 1,\dots,K$
        Generate
            $\left(m^j, q_n^j, q_i^j, b^j, l^j, a^j, v^j, o^j, f^j\right) \sim G$
        Compute
            $c_s^j|s,\ c_i^j|i,\ g_i^j|i, b^j, q_i^j$
        Compute
            $c_d^j = m^j + b^j + q_i^j + q_n^j + c_s^j + c_i^j - g_i^j$
        Compute
            $\psi(s,i) = \psi(s,i) + \frac{u(c_d^j)}{K}$
    end
end
Compute
    $(\hat{s}^*, \hat{i}^*) = \arg\max_{s,i} \psi(s,i)$
```

Anti-fire	Firewall	Procedure	DDoS protection	Insurance	Expected utility
anti-fire	firewall	no procedure	1000 gbps	comprehensive	0.9954
anti-fire	firewall	no procedure	1000 gbps	traditional	0.9950
anti-fire	firewall	procedure	1000 gbps	comprehensive	0.9949
...	...	...	...	...	...
no anti-fire	no firewall	no procedure	no protection	cyber	0.8246
no anti-fire	no firewall	procedure	no protection	no insurance	0.8246
no anti-fire	no firewall	no procedure	no protection	no insurance	0.8242

Table 4.7: Expected utility for 3 best and worst combinations of controls and insurance.

Sensitivity analysis

We can assess the robustness of the previous solution by checking whether variations in the inputs to the model alter the optimal solution. This is specially important in a case like ours with small differences in expected utility among top alternatives and many inputs being purely judgmental. The approach would require the implementation of additional algorithms for sensitivity analysis that indicate whether small deviations in input parameters may lead to a large effect in the model outcome [142]. As an example, the optimal portfolio in Table 4.7 will remain as such until we sufficiently reduce the value of $p(f)$, specifically $f \sim \min[1, \mathscr{P}(0.0088)]$. If $p(f)$ is further reduced, the optimal portfolio will contain the same security controls and insurance than the optimal, except for the anti-fire system. Additionally, sensitivity analysis can be used to explore the maximum cyber insurance price that the Defender would be willing to pay. This may be used, inter alia, to price insurance products.

Introducing constraints

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

Return on security investment

Our formulation focused on choosing the best security portfolio. An additional aspect that could be addressed is calculating the return on security investment (ROSI) to assess the cost effectiveness of a cybersecurity budget [58, 148]. Calculating the optimal solution over a range of budgets (e.g., from € 5,000 to € 25,000) generates a function that, for a given budget, provides the optimal solution and its expected utility to explore the return on risk mitigation investments. Additionally, we could find the optimal increase in the portfolio so as to attain a certain expected utility level or satisfy a certain risk appetite level.

Using the framework from the insurer perspective

The risk analysis model of Chapter 4 may also be used in a parametric manner to set cyber insurance prices and coverages as well as to segment the market, as we briefly outline. First, the insurance product prices were c_i, and their impact was reflected in the utility function $u_D\big(v(c, c_1, c_2), c_i\big)$.. Consequently, we could determine the optimal portfolio and insurance product $(k^*, i^*(c_i)|f)$ to make decisions about the optimal investment and insurance product, given the prices, for a company with features f. This would inform the pricing process: for a given profile f, we could determine the maximum prices that a customer would be willing to pay to acquire a certain insurance product. Moreover, we could define the set $F(i) = \{f : i^* = i\}$ which comprises all companies (as characterised by their features f) such that their optimal insurance product is i. This could constitute the basis to segment a cyber insurance market.

4.2.6 Comparison with a game-theoretic approach

This subsection compares our ARA framework with a standard game-theoretic (GT) approach by analysing a simple example with both methods.. The basic conclusions would be first that both approaches rely on different assumptions and, consequently, lead to different solutions; that the game-theoretic approach requires more stringent common knowledge assumptions that might not hold in cybersecurity; given that, we may view the ARA approach as more robust. Besides, the proposed framework may be more adaptable to realistic cybersecurity scenarios with several potential attackers and several accidental and environmental threats as it more duly apportions various sources of uncertainty, as discussed in Merrick and Parnell [120].

We consider a defend-attack problem in which a defender D has to decide (d) among three connecting options between two data centres in a campus shared with other institutions: using the campus network with encryption and other protection measures (d_1); using it without additional protection (d_2); or, the most expensive, installing a dedicated line between the data centres (d_3). The danger resides in a potential targeted attacker A, insider to the campus, who decides whether to attack the defender's connection (a_1) or not (a_0). The result of the attack (r) leads to consequences related to data exfiltration, expressed as costs, for both the defender (c_D) and the attacker (c_A). They evaluate these consequences through utility functions (u_D and u_A) that incorporate their risk attitude. Fig. 4.12 represents the problem as an ID and Table 4.8 details the problem for various relevant defence-attack combinations.

Common ingredients to both approaches refer to the assessment of the defender elements. Suppose that we have $h = 100000$, $s = 25000$, and $k = 300000$; her risk aversion coefficient is $\lambda = 3 \cdot 10^{-5}$; the attack result r_1, given the protection, follows a beta distribution $r_1 \sim \beta e(0.6, 1.4)$ (mean 0.3); whereas the attack result r_2, given the lack of protection, follows a beta distribution $r_2 \sim \beta e(0.36, 0.24)$ (mean 0.7).

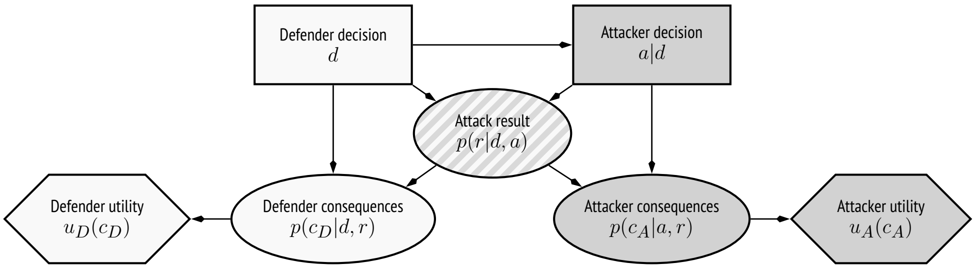


Fig. 4.12: Influence diagram representing the connecting problem.

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

Table 4.8: Defender and attacker elements: r_1 (r_2) is the attack result, in terms of fraction of data compromised, in case the defender uses the campus network with (without) protection d_1 (d_2); h is the cost of installing a new line between the data centres; s is the cost of taking the extra protection when using the campus network; k is the defender's cost relative to the fraction of data compromised; l is the attacker cost of executing the attack; g is the attacker's gain relative to the fraction of data extracted from the defender; λ is the defender risk aversion coefficient; μ is the attacker risk proneness coefficient.

Game-theoretic approach. Under common knowledge, we assume that the defender knows that the attacker's parameters are: $l = 12000$; $g = 33000$; $\mu = 1.8 \cdot 10^{-5}$; r_1 follows a beta distribution $\beta e(2.4, 6.7)$ (mean 0.2637); and r_2 follows beta distribution $\beta e(6.5, 4)$ (mean 0.619).

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

Adversarial Risk Analysis approach. Without common knowledge, we model the defender's beliefs about the attacker's judgment with random probabilities $P_A(\cdot)$ and random utilities $U_A(\cdot)$. Suppose that $l\sim\mathscr{U}(10000,20000)$; $g\sim\mathscr{U}(10000,50000)$; $\mu\sim\mathscr{U}(1\cdot 10^{-5},2\cdot 10^{-5})$; r_1 follows the random beta distribution $\beta e(\mathscr{U}(2,4),\mathscr{U}(6,8))$; and, similarly, r_2 follows $r_2\sim\beta e(\mathscr{U}(5,7),\mathscr{U}(3,5))$.

We calculate the random optimal attack $A^*(d)$, given the defender's choice d, which is obtained through $\arg\max_a \iint U_A(c_A)P_A(c_A|a,r)P_A(r|d,a)\mathrm{d}r\mathrm{d}a$. This leads to estimates $\hat{p}(a_1|d_1)=0.180$, $\hat{p}(a_1|d_2)=0.567$, $\hat{p}(a_1|d_3)=0$, and the corresponding complementary probabilities for a_0. Knowing this, the defender expected utility is $\psi(d)=\iint u_D(c_D)p_D(c_D|a,r)p_D(r|d,a)\hat{p}(a|d)\mathrm{d}c_D\mathrm{d}r\mathrm{d}a$, when d is her choice. The decisions' expected utilities are, respectively, $(-110.99,-1753.933,-19.086)$ and, thus, the ARA optimal defence is $d^*_{ARA}=d_3$, installing a dedicated line, which is different from the game-theoretic solution d_{GT}.

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

Similar comments hold for cases in which a game under partial information appproach is considered as common knowledge over the types prior is required to approximate Bayes-Nash equilibria. See Rothschild, McLay and Guikema [146] for additional discussion.

4.3 Discussion

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

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

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

Notation

Cybersecurity ARA framework notation

$p(\cdot)$	Probability distribution
$u(\cdot)$	Utility function
$P_A(\cdot)$	Random probability distribution (attacker problem)
$U_A(\cdot)$	Random utility function (attacker problem)
c_n	Cost of normal system performance
ψ_n	Expected utility under normal conditions
$t_1,\dots,t_m$	Threats
$c_1,\dots,c_l$	Costs of impacts on the assets
c	Total costs
ψ_r	Expected utility considering threats
e	Security controls portfolio
c_e	Security controls portfolio cost
$\psi(e)$	Expected utility when portfolio e is implemented
ψ_e^*	Expected utility of optimal portfolio
i	Insurance
c_i	Insurance cost
$\psi(e,i)$	Expected utility when portfolio e and insurance i are implemented
$\psi_{e,i}^*$	Expected utility of optimal portfolio and insurance
$u_A(\cdot)$	Attacker utility function
$\psi_A(\cdot)$	Expected utility for attacker
$A^*(e)$	Optimal random attack given security portfolio e
e_p	Preventive security controls portfolio
e_r	Reactive security controls portfolio

Case study template notation

i	Insurance
c_i	Insurance cost
g_i	Insurance coverage
s	Security controls portfolio
c_s	Security controls portfolio cost
f	Fire probability
o	Fire duration
v	Number of computer virus infections
q	Probability that a computer gets infected
w	Percentage of time loss caused by computer virus
b	Impact on facilities
q_i	Insurable impact on computers
q_n	Non-insurable impact on computers
c_d	Total costs for defender
$u(c_d)$	Defender utility
a	Competitor attack
l	Duration of DDoS
l_j	Lenght of j-th DDoS attack
r	Market share loss ratio
m	Impact on market share
e	Attacker earnings
t	Detection of attacker
c_t	Cost when detected
c_a	Result of attack
$u_A(c_a)$	Attacker utility
$U_A(c_a)$	Attacker random utility
$A^*(s)$	Optimal random attack given security portfolio s
$\psi(s,i)$	Expected utility when portfolio s and insurance i are implemented
(s^*,i^*)	Optimal security portfolio s and insurance i
$\beta e(\cdot)$	Beta distribution
$\mathscr{P}(\cdot)$	Poisson distribution
$\varGamma(\cdot)$	Gamma distribution
$Tri(\cdot)$	Triangular distribution
$\mathscr{U}(\cdot)$	Uniform distribution
$\mathscr{B}(\cdot)$	Binomial distribution
$\mathscr{N}(\cdot)$	Normal distribution
$\mathscr{E}(\cdot)$	Exponential distribution

Chapter 5

A tree of cybersecurity objectives

We have proposed a rigorous framework for risk management in cybersecurity in Chapter 4, which overcomes several of the defects in current standards by modelling cyber risks in more detail and including adversarial threats and insurance. The framework emphasises adversarial aspects for better prediction of threats, mitigates lack of data through structured expert judgement techniques and includes cyber insurance within the risk management portfolio. Although we included a template case study, its implementation may entail extensive work within large organisations. Towards the aim of facilitating a decision support system to aid in implementing our cybersecurity risk analysis framework, we propose here a generic tree of potential cybersecurity objectives for ICT owners. We describe potential attributes corresponding to each objective. Its purpose is to support the identification of all potential impacts of cybersecurity incidents in terms of relevant stakeholders' assets. Our approach is inspired by earlier work in counterterrorism, homeland security, aviation safety risk management and cybersecurity financial risk management [100, 103].

The rest of the chapter is structured as follows: First, we provide a generic objective tree for cybersecurity. Then, for each of the non-monetary objectives, we provide potential attributes that measure or estimate objective achievement. Then, we explain how to build an utility function once the objectives and their attributes are specified, including an illustrative example before concluding with a brief discussion.

5.1 Cybersecurity risk management objectives

Cybersecurity occurrences may entail very negative consequences in terms of costs, loss of reputation or even, in some cases, casualties. We track them through performance measures that we want to optimise, which we designate objectives. Through risk management, we aim at implementing treatments, possibly including a cyber insurance, to perform optimally with respect to such objectives, which will depend on the incumbent organisation. They will typically vary from state organisation to private ones and, among these, will differ for, say, a standard small enterprise, an Information and Communication Technology (ICT) based small enterprise, a medium enterprise or a large company. They may also vary in different countries and domains (e.g., air traffic management, healthcare, manufacturing). To each objective we associate, at least, one attribute with which to assess it.

We present here a generic list of objectives from which an organisation may choose when undertaking their cyber risk management process. The context of our problem is an organisation that aims at introducing a cyber risk management strategy, including possibly a cyberinsurance, to improve cybersecurity.

5.2 General concepts

Structuring objectives in trees can help a risk manager overcome the cognitive overload brought by the volume of information which needs to be integrated into the solution of large, complex issues as in cybersecurity risk management [17]. An analyst can work with risk managers to build such objective hierarchy or tree in several ways [99, 29], including a brainstorming process or a questionnaire to identify the relevant objectives. There are several requirements that these must meet if they are to be useful for decision support [101]: *comprehensive*, covering the whole range of relevant consequences for the incumbent organisation; *measurable*, either objectively or subjectively; *non-overlapping*, two objectives should not measure similar impacts; *relevant*, in the sense of being capable of distinguishing between alternatives; *unambiguous,* having a clear relationship between impacts and their description using the corresponding objective; *understandable,* the objective should be presented so that a reader reasonably familiar with risk or business can easily comprehend it.

The lowest tree nodes provide a series of dimensions, say q of them, which may be used to describe the consequences of alternatives, cybersecurity policies in our case, and uncertain scenarios. Each of these objective scales may be quantified in an *attribute*, allowing each consequence to be represented as a vector of attribute levels $c = (c_1, c_2, \dots, c_q)$. We distinguish three types of scales.

- A *natural attribute* gives a direct measure of the objective involved and is universally understood. An example of this, typical with cyber incidents to SMEs, are costs in relation with ICT support services that repair, reinstall or recover desktop computers and measured in EUROs.
- *Constructed attributes* are created for a specific decision context and are not universally understood. They are based on an artificially built ordinal scale, say 1 to 10. For example, in the case of image loss, level 1 could be associated with a case of minimal impact, e.g., a cyber attack with no loss of image even internally at the organisation. Level 10 could be associated with a maximum impact accident and a full compromise of the information assets of the organisation, like the exposure of thousands of private or personal information about customers, with appearance in global media. Henceforth, we would associate each of the levels with a qualitative description of severity with respect to image.
- *Proxy attributes* are used because of its perceived relationship with the objective, when no natural attributes are available and constructed scales are deemed too ambiguous. Variations in a proxy attribute are perceived to correlate with the issue of concern. For example, in online businesses the proxy attribute *website downtime* usually correlates with lost online sales.

5.3 Cybersecurity objectives

A popular approach to describing cybersecurity objectives is in terms of the information security attributes of confidentiality, integrity and availability [125]. However, such objectives may be difficult to interpret from a business perspective: they are useful for expressing security from an information security perspective, in which ICT systems are described in terms of sets of pieces of information that are stored, processed or transmitted. We can think of this as the technical perspective through which we can express cyber risk. Yet the business perspective focuses more on assets and activities relevant for the organisation and its stakeholders; this is even more relevant if we reflect the general principles introduced above: objectives should cover the consequences over relevant organisational assets and activities expressed in variables directly used, or understandable, in the language of the incumbent organisation.

Some cybersecurity frameworks provide catalogues of concepts analogue to our objectives, mostly those addressing business impact analysis in cybersecurity, including ETSI GS ISI 002 v1.2.1 [59], ISO 22317 [90][1], OWASP business impacts [162], the OECD cyber losses types [133], the ENISA Information Package for SMEs [56], the ENISA report on ICT business continuity management for SMEs [57], SABSA [163] or MAGERIT [122]. We also include the list of impacts identified in Hubbard and Seiersen [76] and the CYBECO deliverable on defining cyber insurance scenarios. 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 well the requirements earlier established. Most of them provide a list of recurrent or important business impacts rather than a comprehensive list encompassing less typical impacts (e.g., cyber attack physical impacts). Similarly, they provide types of objectives or impacts that somehow overlap: most of them affect monetary objectives and, thus, some categorisation among them is recommended. For instance, some costs affect specific assets (e.g., activity interruption), whereas others are more general (e.g., competitive advantage, reputation).

Of course, creating a comprehensive and non-overlapping set of objectives may have disadvantages, namely, the addition of more concepts. One example in business terms is that income generation is a clear and main objective for companies to make money through sales. However, companies have alternative means to earn money, which may be even more relevant in other types of organisations such as non-governmental organisations, including grants, investments or licenses. A second example refers to the emerging and potential impacts of cyber risks involving physical and psychological aspects. Thus, third-party impacts such as health and environment should be also taken into account.

As a consequence, our approach is to list objectives and impacts in cybersecurity and sort them in a hierarchy of objectives in a more comprehensive, measurable, non-overlapping, relevant, unambiguous and understandable manner. As mentioned, comprehensiveness and non-overlapping involve, mostly, careful addition of novel concepts. Relevance and understandability are more related with translating cybersecurity impacts from the confidentiality, integrity and availability realms to another one focused on assets, activities and stakeholders.

Besides the existing lists of cybersecurity impacts, the main influences on ours come from asset management and law. The first discipline, ISO 55000 [88] on asset management in general or ISO 19770 for ICT assets [89], helps us in conceptualising the different status that an asset could attain, so that engineers could characterise how an asset affects a system or the

[1] Standards in the ISO 22300 family are the continuation of BS 25999 [16], one of the most popular standards in business continuity management.

organisation in terms of reliability and predictability. The second influence comes from law, in particular, the distinction between damages on property (economic or pecuniary) and persons (general or non-pecuniary). This facilitates the distinction between objectives that can be measured or evaluated in monetary terms (directly or through estimation) and others that are non-monetary and, thus, need special consideration when it comes to their evaluation. It also helps with the distinction between the objectives' owners (e.g., health and environmental damages are suffered by third parties besides the monetary, legal or reputational consequences that such damages could cause to the organisation).

We thus have developed a generic tree of cybersecurity objectives for a generic organisation, summarised in Fig. 5.1 and aimed at reaching the properties mentioned in Section 5.2. When it comes to comprehensiveness, we have evaluated existing categories of impacts to, at least, have categories that cover them. Solving the overlapping problem would mean creating more abstract concepts. We think that this question should be actually addressed when performing the actual risk assessment. Should a risk involve impacts on several categories, it would be necessary to check that impacts included in one category are not included in different ones. We have also tried to bring more general terms for the objectives rather than more domain-specific (e.g., organisation instead of business). This may add a little more ambiguity and less understandability compared to domain-specific IT or business categories, but it provides a more comprehensive approach.

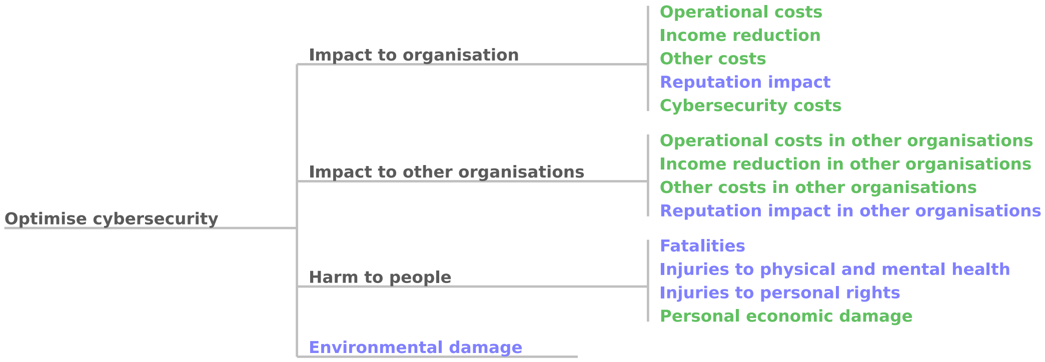

Optimise cybersecurity
Impact to organisation
Operational costs
Income reduction
Other costs
Reputation impact
Cybersecurity costs
Impact to other organisations
Operational costs in other organisations
Income reduction in other organisations
Other costs in other organisations
Reputation impact in other organisations
Harm to people
Fatalities
Injuries to physical and mental health
Injuries to personal rights
Personal economic damage
Environmental damage

Fig. 5.1: Cybersecurity objectives. Green, assessed in monetary terms; blue, not directly measurable in monetary terms (e.g., health, environmental); grey, with both types of subobjectives.

The rest of this section describes in some detail the rationale behind such objectives. Note that some of them refer to impacts that may last several years and, for instance, those measured in monetary terms should be dealt with net present values (NPV) [64]. We also provide, at the end of this chapter, a mapping of some of the previously mentioned catalogues to these cybersecurity objectives. As expressed in Figure 5.1, all objectives refer to minimisation, for example when mentioning *impact to the organisation* we understand *minimising impact to organisation*. Finally, unless explicitly mentioned, the objectives will be expressed in monetary terms.

5.3.1 Impact to the organisation

This objective consists of the following sub-objectives:

- **Operational costs**. We refer to those costs related with the assets and activities involved in the organisation's operations, i.e., the area responsible for producing goods or delivering services, the cost of degradation, malfunction, abuse, unavailability, elimination, recovery and unrecoverability of their assets and activities. We focus on assets such as software, ICT devices, documents, and equipment; and activities such as serving food, delivering merchandise, writing a report, or supporting administrative acts with citizens. All of these impacts can be represented with a monetary attribute. We include:
  - Degradation if the asset or activity performs its function in a less productive or more costly manner, e.g., a text processor running slower than normal as an asset degradation, or slower document production as an activity degradation.
  - Malfunction if the asset or activity has disturbances or a hazardous behaviour, e.g., a text processor producing errors when writing several pages.
  - Abuse if the asset or activity is maliciously manipulated, e.g., a malicious macro exfiltrating the document edited in the text processor.
  - Unavailability of the asset or activity, e.g., the employees cannot run the text processor.
  - Recovery as the actions and resources to restore an asset or activity to a normal situation. Note though that some assets might be unrecoverable (e.g. a piece of art) and this might have an operational impact (e.g., uninstallation of a text processor with several macros tailored to the business that cannot be reprogrammed because a programmer left the company).

- **Income reduction**: Impacts that reduce the income obtained by the organisation. In synthesis, minimizing loss of sales, contracts, market share, funding or licenses. In a business context, they typically involve marketing and commercial aspects related to sales. However, we also have to take into account that some income does not necessarily come from sales, e.g. in public and non-profit organisations. All of them can be assessed in monetary terms. We include:
  - Income reduction over sales flow, involving sales but also leads, quotes, post-sale and customer service.
  - Loss of market share, which can be expressed through the reduction over the sales flow. However, it might also be considered as an asset with an estimated economic value that can drop if market share is reduced.
  - In some cases, when the contracts are few but big, loss of contracts might be a more practical indicator than sales and market share.
  - Loss of funding not directly related with sales flow, e.g., through investments, grants or public funding.
  - Loss of licenses. It has a compliance origin but their loss could reduce income.
- **Other costs**: These refer to other impacts that affect an organisation. It includes some strategic, compliance and financial assets or potential costs. Although their identification or estimation might be difficult, all of them may translate into income (e.g., technological advantages) or costs (e.g., less advertisement for a well-known brand). All of them can be represented through monetary attributes. We include:
  - Loss of competitive advantage caused by leaked, spied, or publicly disclosed sensitive information, including intellectual property or commercial secrets. Although it could be correlated with income reduction or reputation impact, it is also considered an intangible but defined asset that can be estimated or sold.
  - Depreciation, abuse, unavailability or elimination of financial assets. Examples are changes in stock value, financial blackmail, extortion or ransom, theft of financial assets, including money or financial instruments.
  - Costs from non-compliance with contracts, regulations, standards or any other enforceable policy. Examples are fines and regulatory penalties, contractual and agreement penalties and litigation costs.

- **Reputation impact**: We refer 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, reduce income or operations or the activities towards recovering the reputation. In principle, these impacts cannot be represented with monetary attributes.
- **Cybersecurity costs**: It is practical to separate the costs related with managing cybersecurity, since this is the activity we aim to support in our decision-making model [144]. It covers the costs of preventive and reactive controls as well as the eventual cyber insurance. All of them can be represented in monetary terms.

5.3.2 Impact to other organisations

A cybersecurity incident in our organisation might affect other organisations and, thus, the organisation objectives also involve minimising damage to them. It replicates the objectives for our organisation except for minimisation of cybersecurity costs, since we are not supporting their cybersecurity decision-making. Therefore, it consists of the following sub-objectives: **Operational costs to other organisations**; **Income reduction to other organisations**; **Other costs to other organisations**; **Reputation impact in other organisations** (non-monetary).

5.3.3 Harm to people

A cybersecurity incident might also affect people such as employees, customers, or local communities. Therefore, the organisation objectives could also involve minimising harm to people. Some of the sub-objectives proposed in Figure 1 entail impacts which have been very rare, so far, in cybersecurity. For example, cyber attacks with physical impact are unusual but the emergence of industrial systems and smart infrastructures brings these risks to the fore, recall, e.g, Stuxnet. We include: **Fatalities** (non-monetary); **Physical and/or mental health injuries** (non-monetary); **Injuries to personal rights** (non-monetary), e.g., dignity or privacy; **Personal economic damage**.

5.3.4 Environmental damage

Similar to damage inflicted to people, the environment might be affected by cyber attacks against systems with physical operations. Here we model the impact over the natural environment as such (e.g., the costs of cleaning contamination are an impact to organisations or people).

5.4 Attributes for quantifying the non-monetary objectives

We have identified several objectives that were not measurable in monetary terms. We describe here how we may proceed for each of them:

1. We could start with the identification of the main scenarios that a cybersecurity incident could cause. These what-if scenarios should be comprehensive in terms of covering all feasible types of impacts, related to the objective that the relevant stakeholders, assets and activities of the organisation may suffer if attacked.
2. Once these scenarios are identified, they should be quantified for their use in the model, following the approach depicted in Sect. 5.2 based on natural, constructed or proxy attributes.

5.4.1 Impact on reputation

[76] discuss how to assess reputational damage in cybersecurity. The authors did not find strong evidence linking data breaches and stock prices of an attacked company, but observed that a relevant cyberattack cost is related to control the damage to limit reputation effects. As mentioned, this objective may impact brand value, reduce income or service or recovery costs from a public image perspective. However, reputation also encompasses aspects related with trustworthiness, legitimacy and image.

In the organisational theory literature, several authors apply an overall measure of reputation [62] whereas others use an attribute-specific measure [96, 69], since organisations may have multiple types of reputations. [24] provide four categories: *moral reputation* referring to how the organisation treats stakeholders; *procedural*, related to the extent the organisation follows legal and social norms; *performative*, referring to the capability of the organisation for performing their job; and, finally, *technical* related to the capability of the organisation for dealing with complex environments different from their business as usual status. We can use the same four categories with changes in names to facilitate understanding in the context of our model: moral, compliance, performative and adaptability reputations.

Common ways of measuring or building attributes for concepts like reputation are interviews with representatives of stakeholder groups or surveying a representative sample of such groups [169]. Indeed, measuring reputation is meaningful when it is done for specific groups of stakeholders [96, 69], relative to a competitor or a similar organisation [62] or past reputation performance [95].

If we proceed with the constructed-attribute approach, we should first identify the scenarios taking into account the previously mentioned components (e.g., what type of reputation? with respect to which stakeholders?). Once these scenarios are identified, they should be ordered from the most to the least preferred. Table 5.1 provides a simple example of different reputation situations for a particular organisation ranked from best to worst.

Rank	**Impact on reputation**
1	No impact
2	Loss of moral or compliance reputation in up to 10% of employees, customers or the general public.
3	Loss of performative or adaptability reputation in more than 25% of customers or general public.
4	Loss of moral or compliance reputation in up to 50% of employees, customers or the general public.
5	Loss of moral and compliance reputation in more than 50% of employees, customers or the general public.

Table 5.1: Example of reputational impact scenarios constructed scale.

Alternatively, if we proceed with a proxy attribute, we could focus on the salience of cybersecurity incidents in news, media and social networks or the cost of handling the reputation impact of the incident.

5.4.2 Harm to people: Fatalities and injuries to physical and mental health

Cybersecurity incidents may pose a physical risk and, thus, triggering incidents that may affect people's health. Indeed, they are a major concern in medical devices [65], industrial control systems [112] or self-driving vehicles

[157]. Additionally, mental health might be a relevant issue too, for instance, in relation to cyber bulling [170].

Our first sub-objective, minimising fatalities, could be assessed with a natural attribute such as the number of fatalities, as long as we do not distinguish between different types of victims. For the others, for example, the *WHO*[2] *International Classification of Diseases* [175] provides a list with all types of injuries, diseases and disorders and, together with the object or substance producing them, the place of occurrence and the activity when injured. These classifications provide thousands of events or injuries. However, in a real case, our assessment will be more straightforward. Usually, the physical risks of cybersecurity would be a new causing or facilitating event of an already existing safety risk that, most of the time, has been documented by the organisation through industrial or occupational assessments.

Risk analysis typically distinguish between major and minor injuries. We could use them as the two natural attributes. They are also suitable for a constructed-attribute approach. There are several methods that may help us to create an ordinal scale [72], such as the *Injury Severity Score* to assess the severity of injuries or the *Global Assessment of Functioning* or the *WHO Disability Assessment Schedule* [174] for physical or mental functioning. Table 5.2 provides an example of different levels of mental and physical impacts, based on some of the previous scoring systems but excluding those scores related to fatalities.

Rank	**Injuries to physical and mental health**
1	No injury, emergency or functional impairment.
2	Minor emergency that does not require medical intervention (NACA I); or minor injury (4 > ISS > 0); or absent or minimal psychological or physical symptoms, no more than everyday problems or concerns (GAF 81-90).
3	Slight to moderate non life-threatening emergency that requires medical intervention (NACA II and III); or moderate or serious injury (16 > ISS >= 4); or mild and moderate psychological or physical symptoms, causing slight to moderate impairment in social or occupational functioning (GAF 51-80).
4	Serious emergency that may be life-threatening and which requires medical care (NACA IV-VI); or severe to maximal (currently untreatable) injury (ISS >= 16); or serious psychological or physical symptoms or persistent danger causing serious to persistent inability in several areas of functioning including family, mood, relations, thinking or even danger of hurting self or others (GAF 1-50).

Table 5.2: Example of physical and mental impact scenarios constructed scale.

[2] World Health Organisation

Alternatively, we could use the number of people entering into hospital in relation with the cybersecurity event as a proxy attribute.

5.4.3 Harm to people: Injuries to personal rights

Cyber attacks may harm our dignity or privacy, accidentally or intentionally. Furthermore, large scale activities of governments or companies on the Internet have become a major issue on this topic, such as the US NSA surveillance [115], the Great firewall of China [107] or the scandal of Cambridge Analytica [105]. In this context, governments and international institutions are pushing for a more secure and governable cyberspace. Namely, the UN Human Rights Council has stated that "the same rights that people have offline must also be protected online" [165]. See also the recent GDPR (REF) in Europe.

These rights could be identified from national jurisprudence, but the collection of by UN provides an international and overreaching framework. For our purposes it might be useful the classification system in the *Universal Human Rights Index Database*[166] which covers the human rights recognised by UN under categories such as civil and political rights; economic, social and cultural rights; or rights to specific persons or groups.

The constructed-attribute approach may be the best one to operationalise this subobjective. However, the nature of these rights, hardly commensurable, and their relatively big number makes this task demanding. One approach could be creating a hierarchy inspired on [117] pyramid of needs. Most criticisms of this hierarchy focus on its last two categories (esteem and self-actualization); for instance, differences between individuals and societies on what constitutes esteem and self-actualization or even whether they consider the latter more basic than the former. Based on that, Table 5.3 provides an example of different impact levels over personal rights, using our modification of Maslow's pyramid.

Rank	**Injuries to personal rights**
1	No personal right violation
2	Violation of personal rights that may affect esteem and self-actualisation needs.
3	Violation of personal rights that may affect social belonging needs.
4	Violation of personal rights that may affect safety needs.
5	Violation of personal rights that may affect physiological needs, including safety needs that also affect physiological needs.

Table 5.3: Example of personal rights impact scenarios constructed scale.

Alternatively, we could use as a proxy attribute the number of legal actions against the organisation due to personal rights violations or the number of personal identifiable information records exposed.

5.4.4 Environmental damage

As in subsection 5.4.2, cybersecurity incidents may trigger incidents with environmental impact. Indeed, there is a high number of potential environmental risks. We have two relevant types of classifications: focused on the environmental impact of normal operations and on the environmental impact of incidents. For instance, the European eco-management and audit scheme (EMAS) [51] or the British environmental key performance indicators [38] provide suggestions to assess the environmental impact of normal activities such as land use, energy efficiency or emissions to air. In our case, these might be useful to identify impact scenarios in which the environmental performance of the organisation is disrupted by a cyber incident. Additionally, frameworks like the Irish [46] and British Common Incident Classification Scheme (CICS) [45] provide frameworks to identify environmental incidents such as the preservation of natural sites or habitats or contamination of water. They also provide severity scores that might be helpful for building a constructed scale for this objective. Note though that they include impacts that we classify in other sections such as human health or agricultural losses.

Based on the British frameworks [38], [45], we can build a constructed attribute for the environmental impacts. Table 5.4 provides a simple example of different environmental impacts based on these two frameworks.

Rank	Environmental damage
1	No environmental impact.
2	Disturbance in the environmental performance indicators of the organisation.
3	Limited environmental damage, corresponding to CICS category 3 incidents.
4	Significant environmental damage, corresponding to CICS category 2 incidents.
5	Major environmental damage, corresponding to CICS category 1 incidents.

Table 5.4: Example of environmental damage constructed scale.

Alternatively, the quantitative nature of environmental performance indicators might serve us to use them as proxy attributes. For example, we could employ the variation in percentage of the most affected environmental indicator.

5.4.5 *Summary*

Table 5.5 summarises the cyber security risk management objectives and attributes that we include in our study.

Objective	Natural attribute	Constr. attribute	Proxy attribute
Min. operational costs Min. income reduction Min. other costs Min. operational costs in other orgs. Min. income reduction in other orgs. Min. other costs in other orgs. Min. personal economic damage	Monetary units		
Min. reputation impact Min. reputation impact in other orgs.		Yes	Media salience Public relations cost
Min. fatalities	Num. fatalities		
Min. injuries to physical and mental health	Num. injured people	Yes	Num. of people in hospital
Min. injuries to personal rights		Yes	Num. of legal actions against the organisation Num. of personal identifiable information records exposed
Min. environmental damage		Yes	Percentage of variation in environmental indicator

Table 5.5: Summary of objectives and attributes.

5.5 Utility model

Section 5.1 identified a comprehensive list of cybersecurity objectives. From it, the incumbent organisation could choose the objectives relevant in its problem. Then, we need a procedure to model preferences over such impacts, as we do now through a utility function. We use the classic concepts of measurable multi-attribute value function [42] and relative risk aversion [43].

The approach that we adopt, as it is not overly demanding cognitively, is relatively general in its assumptions and is easy to assess in practice is as follows. Under sufficiently general conditions [143], the utility must have the following structure:

1. $u(c) = 1 - \exp(-\rho \sum v_i(c_i)), \quad \rho > 0.$
2. $u(c) = \sum v_i(c_i).$
3. $u(c) = 1 + \exp(\rho \sum v_i(c_i)), \quad \rho > 0.$

where ρ is the risk aversion coefficient and the v_i's are measurable value functions.

We discuss now how to assess the parameter ρ, facilitating scaling the utility to $[0,1]$. In our case, the relevant attributes may be viewed as costs, which are decreasing. We make $c = -d$, to make the attribute increasing. The minimum cost is 0 and suppose the maximum cost is c^*. The utility function has to be strategically equivalent to

$$u(d) = 1 - e^{-\rho d} = 1 - e^{\rho c}.$$

This means that its form should be

$$u(c) = a(1 - e^{\rho c}) + b.$$

We make

$$u(0) = 1, \; a(1 - e^{\rho 0}) + b = 1 \Rightarrow b = 1$$

$$u(c^*) = 0, \; a(1 - e^{\rho c^*}) + 1 = 0 \Rightarrow a(1 - e^{\rho c^*}) = -1.$$

We need one more judgement for a certain cost, which we fix at $c = c^*/2$. We use, for example, the probability equivalent method [60]. In order to acheve so, we ask the cyber risk manager to provide the probability p such that she finds equally interesting the lotteries

$$\begin{pmatrix} 1 \\ c \end{pmatrix} \sim \begin{pmatrix} 1-p & p \\ c^* & 0 \end{pmatrix}.$$

Then,

$$u(c) = (1-p)u(c^*) + pu(0) = p,$$

and we have the system

$$\begin{cases} a(1 - e^{\rho c}) = p - 1, \\ a(1 - e^{\rho c^*}) = -1, \end{cases}$$

from which

$$\frac{1-e^{\rho c}}{1-e^{\rho c^*}} = \frac{p-1}{-1}.$$

This leads to

$$e^{\rho c} + (p-1)e^{\rho c^*} - p = 0.$$

Taking $x = e^{\rho c}$, we have

$$(p-1)x^2 + x - p = 0,$$

whose solution is $x = \frac{1 \pm \sqrt{1-4p(1-p)}}{2(1-p)}$. We then make

$$\rho = \ln x / c,$$

and

$$a = \frac{1}{x^2-1}.$$

5.5.1 *Example of the utility model*

Assuming a risk averse organisation, then if we apply the utility function defined in the previous section, we use

$$u(m,r) = 1 - \exp\Big(-\rho\Big(v_m(m) + v_r(r)\Big)\Big)$$

where m is the monetary impact, r is the impact on personal rights and $v_m(m)$ and $v_r(r)$ are their corresponding value functions. To operationalise this function, we could use the quantitative attributes that measure such subobjectives, so that the utility function can be described as

$$u(m,r) = 1 - \exp\Big(\rho\Big(m + c_r r\Big)\Big)$$

The first one, *m,* is measured through a natural attribute (monetary units) that we shall express in €. Note that this also includes the security costs of security controls and insurance, since they are related to the objective *Min. cybersecurity costs*.

The second one, *r,* is measured with a proxy attribute (records exposed), associated with the parameter c_r. To elicit this parameter, we should provide an economic value to privacy. The legal costs of injuries to personal rights are part of the monetary costs. However, there is no solid estimations for the *value of privacy* [1]. Estimations based on British [109] and American [71] customers reveal that consumers' value of their personal information is up to £7.25 and \$44.62 respectively. Assuming that they are risk neutral and they assign a probability of less than one percent to a data exposure then taking the more conservative British figure (equivalent to €8.25), we shall use that at least they value their personal information at €825. Risk aversion would reduce this figure slightly, whereas the American figures or a lower perception of the likelihood would increase it (e.g., more than €4.000 with the American figures or €1.650 if we assume a probability of breach of less than 0,5 percent). Therefore, we use €825 as a conservative estimate of the economic value of privacy per record.

Then, the utility function that we shall be using is strategically equivalent to

$$u(m,r) = 1 - \exp\left(\rho\left(m + 825r\right)\right)$$

To adjust it, we determine the worst reasonable cost $c_* = m_* + 825r_*$, where m_* is the sum of the maximum cost of the impacts and the security budget and r_* is the maximum number of records that can be exfiltrated. Suppose that for a certain organisation, m_* is estimated at €2.000.000 and r_* is estimated at 5000, so that the worst cost is €6.125.000. We also determine the best cost, which is $c^* = 0$, for $m^* = r^* = 0$. Further suppose that, for $c_1 = \frac{1}{2}c_*$, we obtain $u(c_1) = 0.8$, through the probability equivalent method. We then obtain that the only valid root in 5.5 is $x = 4$, so that $a = 1/15 = 0.066$, $\rho = 4.5267 * 10^{-7}$ and $b = 1$, and the utility function is

$$u(m,r) = 0.066 * \left(1 - \exp\left(4.5267 * 10^{-7}\left(m + 825r\right)\right)\right) + 1.$$

5.6 Discussion

In earlier chapters, we have presented risk analysis models based on influence diagram and adversarial risk analysis, which provide a formal method supporting all relevant steps when undertaking a comprehensive cybersecurity risk analysis. To facilitate its implementation in case studies, as well as to develop a decision support system facilitating its implementation, we have introduced a generic objective tree for cybersecurity risk management from the Defender perspective, with the corresponding objectives and attributes, some ideas on the pertinent forecasting models and a generic preference model, illustrated with specific examples. From it, a cybersecurity risk manager could choose the relevant objectives to proceed in a risk analysis, formulate his preference model by responding a few simple questions and have an orientation on the forecasting models to be implemented, facilitating his analysis.

We consider that we provide a thorough but synthetic list of cybersecurity objectives. We find that most of the catalogues identify objectives from the perspective of an archetypical organisation (typically, business corporations but also public agencies). We tried to provide a more general version also valid from the perspective of other organisations and people.

Additionally, some of the objectives related with personal rights or physical impact are not typically included in these catalogues. The reason is that these are emerging or potential future risks. Even though, we consider that our physical world and our lives are becoming more and more digitalised, and digitalisation will be ubiquitous in our physical world (thus, the importance of physical risks to people and the environment) and in our behaviour and personal lives (thus, the importance of personal rights).

Mapping of existing catalogues to our cybersecurity objectives tree

Mapping of MAGERIT valuation criteria

Table 5.6: MAGERIT mapping to cybersecurity objectives tree

MAGERIT[122]	**Cybersecurity Objectives Tree**
Personal information	Injuries to personal rights (impacts to persons), Other costs (impacts to organisation due to non-compliance regarding personal information) and Operational costs (information asset degradation).
Legal obligations	Other costs
Security	Cybersecurity costs
Commercial or economic interests	Income reduction or other costs (if strategic)
Service interruption	Operational costs
Public order	For most organisations is Impact to other organisations. For those organisations responsible for public order it might be necessary to create a new cybersecurity objective of of non-monetary nature for evaluating the potential states of public order: *Max. public order*.
Operations	Operational costs
Administration and management	Operational costs
Loss of confidence (reputation)	Reputation impact
Prosecution of crimes and law enforcement	For most organisations is min. impact to other organisations. For those organisations responsible for these tasks it is related with Operational costs
Service recovery time	Operational costs
Classified information	As a characteristic of information assets, Operational costs

Mapping of SABSA high-level general business attributes

Table 5.7: SABSA mapping to cybersecurity objectives tree

SABSA[163]	**Cybersecurity Objectives Tree**
Financial - Accounted	Other costs
Financial - AML compliant	Other costs
Financial - Auditable	Other costs
Financial - Benefit-evaluated	Income reduction.
Financial - Cash-flow forecasted	Income reduction
Financial - Credit controlled	Other costs
Financial - Credit risk managed	Other costs
Financial - Investment returnable	Other costs
Financial - Liquidity risk managed	Other costs
Financial - Market risk managed	Other costs (understood as financial market risks)
Financial - Profitable	Income reduction
Financial - Reporting compliant	Other costs
Physical (all attributes)	Operational costs. Note that some characteristics are related to security/risk characteristics of the assets (access controlled, damage protected, defended, secure, theft protected).
Human (all sub-attributes)	Characteristics related to human capital, which could be classified as an asset. Therefore, the related objective is Operational Costs
Process (all sub-attributes)	Other costs
Strategic - Administered	Other costs
Strategic - Branded	Other costs
Strategic - Communicated	Other costs
Strategic - Competitive	Other costs
Strategic - Compliant	Other costs
Strategic - Financed	Other costs
Strategic - Goal oriented	Other costs
Strategic - Governed	Other costs
Strategic - Logistically managed	Operational costs
Strategic - Market penetrated	Income reduction
Strategic - Market positioned	Income reduction
Strategic - Reputable	Reputation impact.
Strategic - Supply chain managed	Operational costs
System (all attributes)	Operational costs. Note that some characteristics are related to security/risk characteristics of the assets (access controlled, incident managed, risk managed).

Chapter 6

Risk analysis models from the insurer prespective

In this chapter, we sketch two additional decision problems relevant in cybersecurity economics around the concept of cyber insurance. The first model serves an insurance company to decide their reinsurance portfolio. The second one supports also an insurance company in deciding whether to grant a given insurance product to a company. We describe all models in terms of influence diagrams and bi-agent influence diagrams.

6.1 Cyber reinsurance

We describe now another major problem for an insurance company referring to reinsurance, described in Fig. 6.1 through an ID. Suppose that an

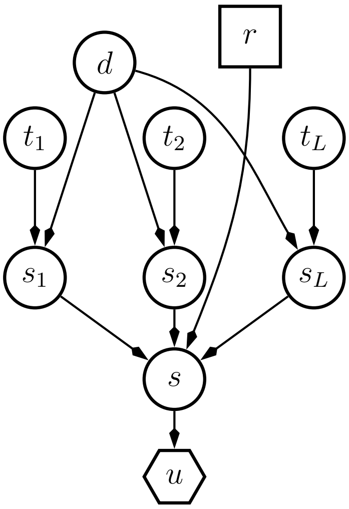


Fig. 6.1: Cybersecurity reinsurance model.

insurance company has segmented the market in several sectors, possibly as outlined in Section 2. To fix ideas, in the ID in Fig. 6.1 we have included three segments referring to standard SMEs S_1; ICT SMEs S_2; and, finally, large enterprises S_L. In standard SMEs, ICT is just a support function and they rarely employ dedicated staff. In ICT SMEs, this technology is critical and core; 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 effect 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. Nodes s_1, s_2 and s_L summarise all of this for each segment.

Node s aggregates the effects s_1, s_2 and s_L over various segments, but are also compensated by the reinsurance decision r, so that $s = g(s_1, s_2, s_L, r)$. The reinsurance decision could be restricted by, say, financial, legal or compliance requirements. It could actually refer to a portfolio referring to several reinsurers. Then, once we are capable of building $p(d)$, $p(t_i)$, $p(s_i|t_i, d)$ – with $i = 1, 2, L$ – and the utility function u, for the insurance company, we would aim at maximising

$$\max_r \int \cdots \int u\big(g(s_1, s_2, s_L, r)\big)\, p(d) \prod_i p(t_i) \prod_i p(s_i|t_i, d)\; ds_1\, ds_2\, ds_L\, dt_1\, dt_2\, dt_L,$$

to find the optimal reinsurance decision of the insurance company.

Again this model serves as a template in that it can be extended to include further details. The number of common and specific threats can be extended, for instance, to include the most common threats for each segment, information that could be derived, e.g., from the claims history or cybersecurity industry reports. Segments could be extended, too, for instance, differentiating by sector or country, or between medium and microbusiness (less than nine employees). Moreover, a dynamic model could be constructed, replicating the threats and effects nodes over several periods, typically years. This could be interesting as cyber insurance presents two very relevant dynamic aspects. First, cybersecurity is continuously evolving, with some types of attacks becoming more frequent or harmful for a number of years or some sectors suffering more attacks. Second, cyber insurance is an emerging market so the size of insuree segments could rapidly grow over time.

6.2 Granting insurance products

In our third and final model, we consider a problem relevant for an insurance company which refers to the decision of granting or not an insurance product to a potential customer. We describe the problem as a BAID in Fig. 6.2. There are two agents involved: the insurance company (I, white nodes) and the customer (J, grey nodes). Striped nodes are shared by both agents. The insurance company needs to decide whether to grant or not an insurance product (i) to the 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, worst case, committing fraud. Should a claim happen, the insurer or a supporting cybersecurity auditor would typically perform a forensic investigation on the claim, aimed at detecting fraud. The claim finally awarded to the insuree by the insurance company (r) 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).

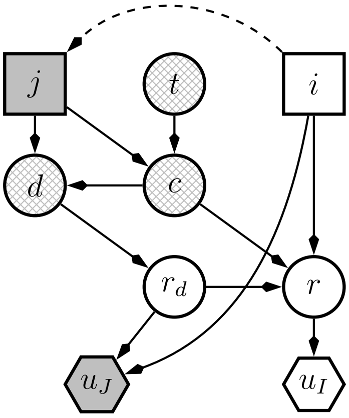


Fig. 6.2: Insurance granting decision.

This is again an ARA problem, structurally resembling that in Section 2. 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 whether the client decisions.

To do so, we consider the customer problem; as in the Attacker problem of Section 2, we model his decision as uncertain and use random utilities and probabilities to build the customer expected utility and find his random optimal decision which we use to estimate the desired fraud probabilities, that would be estimated through Monte Carlo simulation. This, in turn, feeds the expected utility of the insurance company to finally decide whether to grant or not the product. Finally, we seek the maximum expected utility decision for the insurance company.

Again, this model serves also as an extendible template. Insurer decisions could include alternative insurance products. Other behaviours of random nature (e.g., errors) or features of the customer could be added as uncertain nodes that precedes the claim node. Additionally, more companies could be added (replicating the grey and shaded nodes). The claim node could be bifurcated in different types of claims. Indeed, an adversarial threat could substitute or complement the random threat node to enable the assessment of the potential impact in claims of a specific cyber attack (this could be relevant during a surge of a type of cyber attack or during a notorious incident like WannaCry).

6.3 Discussion

We have presented decision making models in relation with cybersecurity and, specially, to help cyber insurance company to design its products. Once with them, we may formulate the cyber reinsurance problem, which allows a company to decide how to allocate its reinsurance portfolio. Finally, we have illustrated the insurance granting decision.

There are other relevant applied economics problems in the field. Specially relevant is the behaviour of agents in the cybersecurity arena. The effective implementation and maintenance of a cybersecurity program and culture is key for minimising risk and, thus, the mechanisms that incentivise adherence to such program or the economics of its implementation are relevant aspects to be studied. When it comes to threat agents, the study of the strategic interaction of adversarial threats could be further extended, as many hackers, more profit-oriented, face a choice problem when selecting their targets. A third interaction, on the protection side, is between governments establishing cybersecurity regulations and the organisations at risk, which could be enriched with the incorporation of cyber insurance companies. Other interesting interactions could be between threat sources (i.e., the agent that wants the attack) and threat perpetrators (i.e., the agent that undertakes the actual attack). Other cybersecurity economic problems, more present in the literature, could be the study of deep web markets related with cyber attacks, models for the economic impact analysis of cyber risks at a macroeconomic or market level or, less analysed, the socioeconomic conditions that incentivise becoming a hacker.

Chapter 7

Discussion and conclusions

7.1 Summary of the Thesis

The first two chapters focused on incident risk analysis. In the first one (Chapter 2) we presented our general incident risk analysis model (GIRA), which formalises the incident risk analysis process through an influence diagram. First, we discussed the considerations that should be taken into account regarding risk analysis when applied to incidents. As a basis for GIRA, we characterised the basic elements of incidents and their relations. Then, we introduced GIRA and the particularities of its main components, accompanied with examples: threat exposure, incident response, incident materialisation, consequences in the systems, impacts on assets, risk objectives and risk evaluation. We also introduced a formal mathematical version of GIRA and briefly discussed additional GIRA models: simplified, for multiple agents and for immediate and delayed events.

In the second one (Chapter 3) we presented further advances for GIRA and a version adapted for a fast cybersecurity risk analysis. We presented a simple elicitation method based on a qualitative interpretation of the likelihood of the event (i.e., based on whether the different events in a chain are certain, possible, rare or impossible) but with a mathematical representation of these qualitative interpretations. Additionally, we introduced a category map for understanding the potential ramifications of cybersecurity incidents that might help when brainstorming about the risks of cybersecurity incidents. We then presented our cybersecurity incident risk analysis model (CSIRA) which is, basically, GIRA using the previous elicitation method and the map for understanding the ramifications of cybersecurity incidents. In the presentation of CSIRA, we also discussed that decision makers only need to compare the scenarios of their different responses (without preference elicitation typical of influence diagrams).

The rest of the models address the traditional time period of risk analysis, e.g., the lifetime of a system or a period of several years (one in the case of Chapter 4).

Chapter 4 develops a cybersecurity resource allocation model that includes the preferences and risk attitudes of the organisation, the intentionality of adversaries and decisions concerning cyber insurance adoption. We presented a comprehensive framework for cybersecurity risk analysis, covering adversarial and non-intentional threats and the use of insurance as part of risk management decisions. The first part introduces influence diagrams that describe different risk analysis models and their mathematical formulation. Starting from a simple system performance evaluation we introduce, incrementally, new elements to the models (risk, risk mitigation, risk transfer and adversarial analysis). The second part presents a full example case in which we detail all aspects of the assessment: The description and the structure of the risk problem, the assessment of the organisation beliefs about the elements affecting risk and their preferences, the modelling of the attacker problem to forecast his actions and the calculation of the best portfolio of security controls and insurance for the organisation.

In Chapter 5, we described cybersecurity objectives with the purpose of facilitating a comprehensive identification of the organisational objectives at risk. We distinguished between those objectives that can be measured in monetary terms and those that cannot or shouldn't, such as physical or harm to people. We further explore how to measure those non-monetary objectives (e.g., reputation, personal rights, environmental damage). We concluded the chapter by detailing how to use this cybersecurity objectives and attributes with an utility function.

In Chapter 6 we presented models for insurance companies. In the first model, the insurance company is deciding what reinsurance product to acquire taking into account the different market segments that the company is insuring (e.g., SMEs, large business). In the second model, the insurance company is deciding whether they grant or not an insurance product to a potential customer.

The chapters that presents the GIRA and CSIRA models present a walktrough that provides the characteristics and an example of the different types of components of the model. Although in the examples not all nodes are described, it is possible to follow the chapter to replicate the example in Genie[1] or R[2]. In fact, the examples have been modelled with Genie. The

[1] Software for creating discrete influence diagrams and Bayesian networks through a graphical interface

[2] Statistical and data analysis programming language and environment

case presented in Chapter 4 provides a template on how to structure and perform a risk analysis under the presented framework. It details the modelling of all nodes and the calculations in the model. We provided the algorithms that estimate the distribution of attacks and the optimal security and insurance portfolio for the defending organisation. These algorithms have been implemented in R to simulate the problem and calculate the solutions. These software alghoritms are part of a software toolbox developed for a pan-European research and innovation project, the CYBECO Toolbox[3]

In the reminder of the section we discuss our contributions with respect to the objectives we established for the Thesis. This is followed by a discussion on future work.

7.2 Contributions to the research objectives

Development of a risk analysis model for cybersecurity incidents

The contribution of GIRA/CSIRA is a risk analysis model tailored to incident situations, following a synthetic but comprehensive characterisation of incidents and a formal mathematical representation. As discussed in Chapter 2, we found several shortcomings that existing methods do not address well. Namely, that models such as bow-ties do not cover aspects related to value (assets, impacts, risk evaluation) and risk matrices lead to oversimplified analysis.

Simple risk scoring, as in risk matrices, is too vague for defining scenarios beyond exploratory descriptions of risks. It does not approximate the likelihood adequately since it does not take into account chains of events. The origin of this issue is the characterisation of risks and incidents; thus, we started by developing an entity-relationship diagram to reason about the risks of incidents (Fig. 1, Chapter 2) to establish what elements compose incidents and their relations. It covers components relevant for both the technical (e.g., vulnerability or incident) and evaluative (e.g., asset or impact) characterisation of incidents.

GIRA/CSIRA are compatible with quantitative risk analysis. As influence diagrams, they are similar to the ones presented in Chapter 4. The main difference is the problem structure that defines the parent relations between nodes. Although every risk case will have its specific number of nodes, GIRA/CSIRA will have a reduced number of nodes, since we are analys-

[3] https://www.cybeco.eu/(retrieved28/05/2019).

ing a few potential incidents and not all the threats that a system might face over a long period. Therefore, if the analysts have time to perform a more in-depth risk analysis, we would recommend the use of GIRA/CSIRA replicating the assessment of beliefs and preferences done in Chapter 4.

However, we also include methods that facilitate a fast risk analysis. The first one is the method described in Chapter 3, which classify events as certain, possible, rare and impossible. Although qualitative in nature, it provides certain advantages against the qualitative method typical of risk analysis (e.g., low, medium, high). First, the concatenation of rare events creates several levels of likelihood (rare, rarer than rare, etc.). Second, we can define a threshold that has a probabilistic meaning. Even though, it is a limited method compared to that discussed in Chapter 4. The purpose was to design the most simple elicitation method viable for GIRA/CSIRA.

When it comes to risk evaluation, the recommended practice is to assess the preferences and risk attitudes of the involved stakeholders. However, this might not be possible during an incident (or secondary in a fast risk analysis). As a decision problem, the purpose is to clarify what are the best options to counter a risky situation. For complex risk scenarios, in which risk managers have multiple combinations of responses (e.g., the portfolio of Chapter 4), this could be an impractical and tedious task, and the best option is to switch to preference elicitation as we do in Chapters 4 and 5. However, in situations with a manageable number of alternatives, risk evaluation could be simplified to presenting the different scenarios to the stakeholders, in terms of their risk objectives and for each of the responses.

The dynamic aspects of risk are also taken into account. Although we do not provide dynamic models, we do emphasise the need to establish the likelihoods for a period of time (harmonised if possible) and an expiration time for the different assessments.

GIRA is at the same level of generality as the risk concepts in ISO 31000 or the incident concepts in IS0 22300. We explicitly discuss an example case outside cybersecurity to demonstrate the generality of the model. CSIRA is an adaptation of GIRA to cybersecurity incidents using the quick alternative methods for elicitation and scenario selection discussed above. This means that the adaptation was more focused on ease of implementation than on cybersecurity concepts. Indeed, the map of potential ramifications of cyber incidents can be thought of as an extension of the emphasis on multi-objective risk analysis and the concept of dependent systems discussed in GIRA.

Integration of adversarial risk analysis into cybersecurity

In the case of GIRA/CSIRA, we do not perform an explicit analysis of adversarial threats, but we discuss how to integrate such elements. In general, this information affects the threat presence node (i.e., the probability we assign that the detected or perceived incident has been caused by a specific threat actor) and the incident materialisation and consequence nodes (i.e., the probability we assign to an incident or consequence should a specific threat actor represent the threat).

The framework for cybersecurity risk analysis (Chapter 4 further advances the literature in the field of adversarial risk analysis (ARA). The modelling of the adversarial parts is based on recent developments in ARA. However, the contribution of this chapter is the integration of the adversarial analysis into an influence diagram that represents the cybersecurity risk management problem of an organisation, as seen in Sect. 2 of Chapter 4. Additionally, the case study provides a detailed assessment procedure, which illustrates how to use our model for building a risk analysis case. In this template, the risk analysis comprises adversarial threats modelled under the ARA paradigm, non-intentional threats, impacts or costs that the threats might cause, security and insurance portfolios that the organisation can implement to protect against the risks and any other relevant security features that might affect risk (e.g., compliance, behaviour).

Integration of risk transfer/insurance into cybersecurity risk analysis

The risk analysis framework in Chapter 4 incorporates cyber insurance as a component of the risk analysis, with its particularities (e.g., dependence on the security portfolio, how the insurance policy affects the final monetary impact of a cyber attack). This is one of the first chapters that integrates cyber insurance into a cybersecurity risk analysis model for an organisation and, to our knowledge, the first that integrates adversarial threats and cyber insurance (aspects that we found very relevant in cybersecurity risk analysis).

We also introduced models for insurance companies in Chapter 6. In the first one, the insurance company is deciding what reinsurance product acquires, taking into account the different market segments the company is insuring (e.g., SMEs, large business). In the second one, the insurance company is deciding whether they grant or not an insurance product to a potential customer.

Integration of multi-objective decision-making into cybersecurity risk analysis

As mentioned, many of the existing methods pay little attention to risk evaluation. Moreover, this adds to the difficulty of comparing risks, given the trade-off between commensurability and comparability (as discussed in Chapter 2): A single objective (e.g., risk matrices) is easy to evaluate, but using the same scale might present problems of incommensurability. On the contrary, evaluating multiple objectives is less incommensurable but becomes a more difficult task.

In the case of GIRA, we explicitly define the objectives as synthesisers of impacts over assets (Chapter 2). The risk evaluation is represented through a general value node agnostic to the preference elicitation method. In our Thesis, we used two elicitation methods: the simple scenario comparison of CSIRA, and the elicited utility function of Chapter 4. In the CSIRA chapter, we provide a conceptual map to classify the potential ramifications of cybersecurity incidents taking into account the informational, physical and psychological impacts as well as a broader stakeholder environment. These might serve to elicit non-commensurable objectives such as, for example, monetary, ethical/legal and human safety objectives.

When it comes to the Chapter 4 framework, we do not explicitly present a case with multi-objective decision-making (all impacts are measured in monetary terms). However, we briefly introduce in Sect. 2.1 a model in which the utility depends on three cybersecurity attributes (availability, integrity and confidentiality). This represents a multi-attribute utility theory problem. Additionally, the framework could be applied to cases with multiple attributes (e.g., monetary and reputation). In this case, the assessment of preferences should be done with an elicitation method that takes into account this plurality of attributes.

To facilitate the implementation of our previous models, and cybesecurity decision support in general, we propose a tree of cybersecurity objectives (chapter 5), with the corresponding attributes to measure them. We take special consideration when it comes to objectives that cannot be measured in monetary terms: objectives related to reputation, physical or mental harm to people, their rights or the natural environment. We provide a comprehensive tree to facilitate the identification of the objectives and clarify possible overlapping (e.g., between reputation and brand value). Although physical harm is currently a ”potential risk”, we consider that such kind of objectives must be included, due to the emergence of cyber-physical systems. The same for objectives related with personal rights. We provide attributes to integrate the non-monetary objectivies in multi-attribute utilities. We provide constructed tables with different levels of impact in these objectives based on existing evaluation frameworks for the different objectives (e.g., personal rights, occupational health). We also provide proxy attributes that correlate with the objectives.

7.3 Future work

Chapters 4 to 6 are part of the EU's H2020 research and innovation project CYBECO[4] to develop new tools for cybersecurity risk analysis. Finished in April 2019, we shall publish our advancements in the project in upcoming months. These include the following.

First, the development of a full decision support system. An initial algorithm in R has been developed for the case in Chapter 4. Additionally, the CYBECO project developed a software for risk analysis that contains the model presented in Chapter 4 but applied to a use case defined in the project, including the R alghoritm that calculate the optimal solution as wll as relevant indicators such as expected impacts or probability of events. Moreover, we are working on computational enhancements to facilitate the calculation of information such as sensitivity analysis or the computational implementation of large risk analysis models.

We also wrote a more detailed exposition of the risk analysis framework steps in the first of the CYBECO deliverables[5]. Basically, we describe core concepts in risk analysis, followed by the procedure for using our risk analysis models consisting of (1) definition of the risk analysis scope, (2) identification of the risk components like threats or assets, (3) problem structuring using our models and (4) problem solving using our algorithms. Important further work is developing materials for training people in the usage and understanding of our framework.

Additional future work in cybersecurity involves the role of compliance/regulation as a security control or objective and input from cyber insurance experts. It might include also the development of preference models for cyber attackers or suggestions for cyber insurance product design.

These advancements could also be replicated or adapted to GIRA/CSIRA, as they are simpler models representing simpler risk analysis problems. The first one might be adapting the R algorithms to GIRA/CSIRA. Further advancements could involve the development of a software environment for

[4] www.cybeco.eu

[5] *CYBECO D3.1 – Modelling framework for cybersecurity risk management*

GIRA/CSIRA, potentially based on Shiny, a platform for running web applications using R. Even a lightweight application based on JavaScript is possible. GIRA/CSIRA do not need sophisticated data analysis model, and its worth exploring its implementation as a javascript program in combination with html/css for the user interface - either as a web application or a desktop application through a framework such as Electron.

Beyond advancements in the refinement of the models and their software implementation, future work could also involve descriptive research about cybersecurity. Most risk analysis remain internal and confidential within the assessed organisations, especially in cybersecurity. However, there is sufficient information available in the academic and industrial literature for the development of publicly-available risk analysis that represent archetypical cases of cybersecurity problems. For instance, it is possible to identify the assets, threats, security controls and features of one or various archetypical organisation (e.g., SME, technological startup) and perform a risk analysis, with the support of cybersecurity and business experts that would provide valuable information to regulators, cybersecurity companies or insurance providers.

References

[1] Acquisti, A., Leslie, K.J., Loewenstein, G. 2013. “What is Privacy Worth ?.” In *The Journal of Legal Studies*, Vol.42, No. 2, pp. 249–274.

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

[3] Akl, P., Loper, R., Dantu, K., and Kolan, R. 2007. “Classification of Attributes and Behavior in Risk Management Using Bayesian Networks.” In *Proc. IEEE Intelligence and Security Informatics Conference*, pp. 71 – 74.

[4] Allcott, H. and Gentzkow, M. 2017. “Social Media and Fake News in the 2016 Election.” In *Journal of Economic Perspectives*, Vol. 31, No. 2, pp. 211 – 236.

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

[6] Amoroso, E.G. 1994. *Fundamentals of Computer Security Technology*. Prentice Hall.

[7] Anderson, R. 2008. *Security Engineering*. John Wiley & Sons.

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

[9] Aven, T., and Renn, O. 2009. “On Risk Defined as an Event Where the Outcome Is Uncertain.” In *Journal of Risk Research*, Vol. 12, No. 1, pp. 1–11.

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

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

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

[13] Borgonovo E., Tarantola S., Plischke E., and Morris, M.D. 2014. “Transformations and Invariance in the Sensitivity Analysis of Computer Experiments.” In *Journal of the Royal Statistical Society, Series B*, Vol. 76, No. 5, pp. 925 – 947.

[14] Brecht, T. and Nowey M. 2012. "A Closer Look at Information Security Costs." In *Workshop on the Economics of Information Security*.

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

[16] British Standards Institution. 2007. *BS 25999-2:2007 Specification for Business Continuity Management*.

[17] Brownlow, S. and Watson, S. 1987. "Structuring Multi-attribute Value Jierarchies." In *Journal of the Operational Research Society*, Vol. 38 No. 4, pp. 309–317.

[18] Byres, E. and Lowe, J. 2004. "The Myths and Facts behind Cyber Security Risks for Industrial Control System." In *Proc. of the VDE Congress*, pp. 116 – 121.

[19] Campbell, S. 2005. "Determining Overall Risk." In *Journal of Risk Research*, Vol. 8, No. 7-8, pp. 569–581.

[20] Card, A.J., Ward, J.R. and Clarkson, P.J. 2012. "Beyond FMEA: The Structured What-If Technique (SWIFT)." In *Journal of Healthcare Risk Management*, Vol. 31, No. 4, pp. 23–29.

[21] Cardenas, A.A., Amin, S., Sinopoli, B., Giani, A., Perrig, A. and Sastri, S. 2009. "Challenges for Securing Cyber Physical Systems." In *Workshop on Future Directions in Cyber-physical Systems Security*, Vol. 5.

[22] Cardenas, A.A., Amin, S., and Sastry, S. 2008. "Research Challenges for the Security of Control Systems." In *Proc. of the 3rd Conference on Hot Topics in Security*, pp. 6:1 – 6:6.

[23] Cardenas, A.A., Amin, S., Lin, Z.S., Huang, Y.L., Huang, C.Y, and Sastry, S. 2011. "Attacks Against Process Control Systems: Risk Assessment, Detection, and Response." In *Proc. of the 6th ACM Symposium on Information, Computer and Communications Security*, pp. 355 – 366.

[24] Carpenter, D.P. and Krause, G.A. 2011 "Reputation and Public Administration." In *Public Administration Review*, Vol. 72, No. 1, pp. 26–32.

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

[26] Centre for Strategic and International Studies (USA). "Significant Cyber Events since 2006."

[27] Chapman, C., and Ward, S. 2000. "Estimation and Evaluation of Uncertainty: a Minimalist First Pass Approach." In *International Journal of Project Management*, Vol. 18, No. 6, pp. 369–383.

[28] Cherdantseva, Y., Burnap, P., Blyth, A., Eden, P. Jones, K, Soulsby, H., and Stoddart, K. 2016. "A Review of Cyber Security Risk Assessment Methods for SCADA Systems." In *Computers & Security*, Vol., 56, pp. 1–27.

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

[30] Clemens, P.L., and Simmons, R.J. 1998. *System Safety and Risk Management: A Guide for Engineering Educators*. National Institute for Occupational Safety and Health.

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

[32] Command Five Pty Ltd (Australia). 2011. "Advanced Persistent Threats: A Decade in Review."

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

[34] Cox, L.A., Babayev, D. and Huber, W. 2005. "Some Limitations of Qualitative Risk Rating Systems." In *Risk Analysis*, Vol. 25, No. 3, pp. 651–662.

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

[36] Defense Science Board of the Department of Defense (USA). 2013. *Task Force Report: Resilient Military Systems and the Advanced Cyber Threat*.

[37] Departamento de Seguridad Ciudadana, Ayto. de Vitoria-Gasteiz, Spain. (009. *Memoria 2009 del Servicio de Prevención Extinción de Incendios y Salvamentos*.

[38] Department for Environment, Food and Rural Affairs (UK). 2006. *Environmental Key Performance Indicators Reporting Guidelines for UK Business*.

[39] Department of Defense (USA). 1980. *MIL-STD-1629A, Procedures for Performing a Failure Mode, Effect and Criticality Analysis*.

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

[41] Dodgson, J.S., Spackman, M. Pearman, A. and Phillips, L.D. 2009. *Multi-Criteria Analysis: A Manual*. Department for Communities and Local Government.

[42] Dyer, J. and Sarin, R. 1979. "Group Preference Aggregation Rules Based on Strength of Preference." In *Management Science*, Vol. 25, No. 9, pp. 822–832.

[43] Dyer, J. and Sarin, R. 1982. "Relative Risk Aversion." In *Management Science*, Vol. 28, No. 8, pp. 875–886.

[44] Edmunds, A., and Morris, A. 2000. "The Problem of Information Overload in Business Organisations: a Review of the Literature." In *International Journal of Information Management*—, Vol. 20, No. 1, pp. 17–28.

[45] Environment Agency (UK). 2006. *Incidents and their Classification: the Common Incident Classification Scheme (CICS), Version 12*.

[46] Environmental Protection Agency (Ireland). 2010. *Guidance to Licensees/COA Holders on the Notification, Management and Communication of Environmental Incidents*.

[47] Ericson, C.A. 2005. "Fault Tree Analysis." In *Hazard Analysis and Techniques for System Safety*. John Wiley & Sons.

[48] Espinoza, N. 2009. "Incommensurability: The Failure to Compare Risks." In *The Ethics of Technological Risk*, pp. 128–143.

[49] European Commission. 2006. *Communication from the Commission on a European Programme for Critical Infrastructure Protection.*

[50] European Commission. 2013 *Cybersecurity Strategy of the European Union: An Open, Safe and Secure Cyberspace.*

[51] European Commission. 2017. *Commission Decision (EU) 2017/2285 of 6 December 2017 Amending the user's guide setting out the steps needed to participate in EMAS, under Regulation (EC) No. 1221/2009 of the European Parliament and of the Council on the voluntary participation by organisations in a Community eco-management and audit scheme (EMAS).*

[52] European Council. 2008. *Council Directive 2008/114/EC of 8 December 2008 on the identification and designation of European critical infrastructures and the assessment of the need to improve their protection (Text with EEA relevance).*

[53] European Parliament and European Council. 2016. *Directive (EU) 2016/1148 of the European Parliament and of the Council of 6 July 2016 concerning measures for a high common level of security of network and information systems across the Union.*

[54] European Food Safety Authority. 2014. *Guidance on Expert Knowledge Elicitation in Food and Feed Safety Risk Assessment.*

[55] European Food Safety Authority. 2016. *Guidance on Uncertainty in EFSA Scientific Assessment.*

[56] European Network and Information Security Agency. 2007. *Information Package for SMEs.*

[57] European Network and Information Security Agency. 2010. *IT Business Continuity Management – An Approach for Small Medium Sized Organisations.*

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

[59] European Telecommunications Standards Institute. 2015. "Annex B1.8 – With what kind of impact." In *ESTI GS Information Security Indicators; Event Model, A Security Event Classification Model and Taxonomy.*

[60] Farquhar, P.H. 1984. "State of the Art - Utility Assessment Methods." In *Management Science*, Vol. 30, No. 11, pp. 1283 – 1300.

[61] Fenton, N. and Neil, M.. 2011. "The Use of Bayes and Causal Modelling in Decision Making." In *CEPIS Upgrade*, Vol. 12, No. 5, pp. 10–21.

[62] Fombrun, C.J. 2012. "The Building Blocks of Corporate Reputation: Definitions, Antecedents, Consequences." In *The Oxford Handbook of Corporate Reputation*. Oxford University Press.

[63] French, S. 1986. *Decision Theory: An Introduction to the Mathematics of Rationality.* Halsted Press.

[64] French, S. and Ríos Insua, D. 2000. *Statistical Decision Theory*. John Wiley & Sons.

[65] Fu, K., Blum, J. 2013. "Controlling for Cybersecurity Risks of Medical Device Software." In *Communications of the ACM*, Vol. 56, No, 10, pp. 35–37.

[66] Galway, L. A. 2007. *Subjective Probability Distribution Elicitation in Cost Risk Analysis: A Review*. Tech. Rep. 410, Rand Corporation.

[67] Gianni, A., Sastri, S., Johansson, K.H. and Sandberg, H. 2009. "The VIKING Project: an Initiative on Resilient Control of Power Networks." In *Proc. 2nd Int. Symp. on Resilient Control Systems*. pp. 31 – 35.

[68] Gordon, T.J., and Hayward. H. 1968. "Initial Experiments with the Cross Impact Matrix Method of Forecasting. " In *Futures*, Vol. 1, No. 2, pp. 100–116.

[69] Greenwood, R., Li, S.X., Parkish, R. and Deephouse, D.L. 2005. "Reputation, Diversification, and Organizational Explanations of Performance in Professional Service Firms." In *Organization Science*, Vol. 16, No. 6, pp. 661–673.

[70] Gregory, R., Failing, L. Harstone, M., Long, G. McDaniels, T. and Ohlson, D. 2012. *Structured Decision Making: A Practical Guide to Environmental Management Choices*. John Wiley & Sons.

[71] Hann, I.-H., Kai-Lung, H., Sang-Yong, T.L., Ivan, P.L.P. 2007. "Overcoming Information Privacy Concerns: An Information Processing Theory Approach." In *Journal of Management Information Systems*, Vol. 24, pp. 13–42.

[72] Hasler, R. M., Kehl, C., Exadaktylos, A. K., Albrecht, R., Dubler, S., Greif, R., anf Urwyler, N. 2012. "Accuracy of Prehospital Diagnosis and Triage of a Swiss Helicopter Emergency Medical Service. *Journal of Trauma and Acute Care Surgery*, Vol. 73, No. 3, pp. 709–715.

[73] He, F. and Zhuang, J. 2016. "Balancing Pre-disaster Preparedness and Post-disaster Relief." In *European Journal of Operational Research*, Vol. 252, No. 1, pp. 246–256.

[74] Herley, D. and Florencio, C. 2010. "Nobody Sells Gold for the Price of Silver: Dishonesty, Uncertainty and the Underground Economy." In *Workshop on the Economics of Information Security 2009*, pp. 33 – 53.

[75] Howard, R.A., and Matheson, J.E. 2005. "Influence Diagrams." In *Decision Analysis*, Vol. 2, No. 3, pp. 127–143.

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

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

[78] Industrial Control Systems Cyber Emergency Response Team . 2017. *Destructive Malware*. National Cybersecurity and Communications Integration Center (US).

[79] Information Security Forum. 2016. *Information Risk Assessment Methodology 2*.

[80] International Association of Drilling Contractors. 2015. *Health, Safety and Environment Case Guidelines for Mobile Offshore Drilling Units, Issue 3.6*.

[81] International Electrotechnical Commission. 2006. *IEC 61025:2006, Fault Tree Analysis (FTA)*.

[82] International Organisation for Standardization. 2000 *ISO 17776:2000, Petroleum and Natural Gas Industries – Offshore Production Installations – Guidelines on Tools and Techniques for Hazard Identification and Risk Assessment*.

[83] International Organisation for Standardization. 2009 *ISO 31000:2009, Risk Management – Principles and Guidelines*.

[84] International Organisation for Standardization. 2012. *ISO 22300:2012, Societal Security – Terminology*.

[85] International Organisation for Standardization. 2012. *ISO/IEC 27000:2012, Information Technology – Security Techniques – Information Security Management Systems*.

[86] International Organization for Standardization. 2013. *ISO/IEC 27001:2013, Information Security Management Systems - Requirements*.

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

[88] International Organisation for Standardization. 2014. *ISO 55000:2014, Asset Management – Overview, Principles and Terminology*.

[89] International Organisation for Standardization. 2015. *ISO 19770-5:2015, IT Asset Management – Overview and Vocabulary – Part 5*.

[90] International Organisation for Standardization. 2015. *ISO/TS 22317:2015, Societal Security – Business Continuity Management Systems – Guidelines for Business Impact Analysis*.

[91] International Society for Automation. 2017. *ANSI/ISA-62443.00.01-2007. Security for Industrial Automation and Control Systems Part 1: Terminology, Concepts, and Models*.

[92] Ionnidis, D., Anderson R., Fuloria, R., Moore, S., and Pym, T. 2010. "Security Economics and Critical National Infrastructure." In *Economics of Information Security and Privacy*, pp. 55 – 56. Springer.

[93] ISACA. 2012. *COBIT 5: A Business Framework for the Governance and Management of Enterprise IT*.

[94] Jaquith, A. 2007. *Security Metrics: Replacing Fear, Uncertainty, and Doubt*. Pearson Education.

[95] Jensen, M. and Roy, A. 2008. "Staging Exchange Partner Choices: When do Status and Reputation Matter?" In *Academy of Management Journal*, Vol. 51, No. 3, pp. 495–516.

[96] Jensen, M., Kim, H. and Kim, B.K. 2012. "Meeting Expectations: A Role-theoretic Perspective on Reputation." In *The Oxford Handbook of Corporate Reputation*. Oxford University Press.

[97] Julisch, K. 2003. "Clustering Intrusion Detection Alarms To Support Root Cause Analysis." In *ACM Transactions On Information And System Security*, Vol. 6, No. 4, pp. 443–471.

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

[99] Keeney, R. 1992. "On the Foundations of Prescriptive Decision Analysis." In *Utility theories: Measurements and Applications*. pp. 57–72. Ward Edwards

[100] Keeney, R. 2007. “Modeling Values for Anti-terrorism Analysis.” In *Risk Analysis*, Vol. 27, No. 3, pp. 585–596.

[101] Keeney, R. and Gregory, R. 2005. “Selecting Attributes to Measure the Achievement of Objectives. ” In *Operations Research*, Vol. 53, No. 1, pp. 1–11.

[102] Keeney, R., and Raiffa, H. 1993. *Decisions with Multiple Objectives*. Cambridge University Press.

[103] Keeney, R. and von Winterfeldt, D. 2011. “A Value Model for Evaluation Homeland Security Decisions.” In *Risk Analysis*, Vol. 31, No. 9, pp. 1470–87.

[104] Kletz, T. 1999. *HAZOP and HAZAN: Identifying and Assessing a Process Industry Hazards*. ICHemE.

[105] Kurtz, C., Semmann, M., and Schulz, W. 2018. “Towards a Framework for Information Privacy in Complex Service Ecosystems.” In *39th Int. Conf. on Information Systems*.

[106] Langner, R. 2011. “Stuxnet: Dissecting a Cyberwarfare Weapon.” In *IEEE Security & Privacy*, Vol. 9, No. 3, pp. 49–51.

[107] Lee, J. A., Liu, C. U. 2012. “Forbidden City Enclosed by the Great Firewall: The Law and Power of Internet Filtering in China.” In *Minnesota Journal of Law, Science & Technology*, Vol. 13, No. 1, pp. 125–151.

[108] Lee, R. M., Assante, J. and Conway, T. 2014. *ICS Defense Use Case Dec 301, 2014 – German Steel Mill Cyber Attack*. SANS Institute.

[109] London Economics (UK). 2017. *Research and Analysis to Quantify the Benefits Arising from Personal Data Rights under the GDPR – Report to the Department for Culture, Media & Sport*.

[110] Low, P. 2017. “Insuring Against Cyber-Attacks.” In *Computer Fraud & Security*, Vol. 2017, No. 4, pp. 18 – 20.

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

[112] Macaulay, T., Singer, B. L. 2016. *Cybersecurity for Industrial Control Systems: SCADA, DCS, PLC, HMI, and SIS*. Auerbach Publications.

[113] Manimaran, C.-C , Ten, G., and Liu, C.-W. 2008. “Vulnerability Assessment of Cybersecurity for SCADA Systems.” In *IEEE Transactions on Power Systems*, Vol. 23, No.4, pp. 1836 – 1846.

[114] Margolis, Howard. 1997. *Dealing with Risk: Why the Public and the Experts Disagree on Environmental Issues*. University of Chicago Press.

[115] Margulies, P. 2013. “The NSA in Global Perspective: Surveillance, Human Rights, and International Counterterrorism.” In *Fordham Law Review*, Vol. 82, No. 5, pp. 2137–2167.

[116] Marotta, A., Martinelli, F., Nanni, S., Orlando, A. and Yautsiukhin, A. 2017. “Cyber Insurance Survey.” In *Computer Science Review*, Vol. 24, pp. 35 – 61.

[117] Maslow, A. H. 1943. "A Theory of Human Motivation." In *Psychological Review*, Vol. 50, No. 4, 370–96.

[118] Mcafee (USA). 2018. *Economic Impact of Cybercrime – No Slowing Down*.

[119] McCumber, J. 1991. "Information Systems Security: A Comprehensive Model." In *Proc. of the 14th NIST-NCSC National Computer Security Conference, Washington DC (USA)*: 328–337.

[120] Merrick, J. and Parnell, G. 2011. "A Comparative Analysis of PRA and Intelligent Adversary Methods for Counterterrorism Risk Management." In *Risk Analysis*, Vol. 31, No. 9, pp. 1488–1510.

[121] Michie, S., van Stralen, M.M and West. R. 2011. "The Behaviour Change Wheel: A New Method for Characterising and Designing Behaviour Change Interventions." In *Implementation Science*, Vol. 6, No. 1, pp. 6 – 42.

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

[123] Mirkovic, J. and Reiher, P. 2004. "A Taxonomy of DDoS Attack and DDoS Defense Mechanisms." In *ACM SIGCOMM Computer Communication Review*, Vol. 34, pp. 39–45.

[124] Mo, Y., Kim, T.H.J., Brancik, K., Dickinson, D., Lee, H., Perrig, A., and Sinopoli, B. 2012. "Cyber-Physical Security of a Smart Grid Infrastructure." In *Proc. of the IEEE*, Vol. 100, No. 1, pp. 195 – 209.

[125] Mowbray, T. J. 2013. *Cybersecurity: Managing Systems, Conducting Testing, and Investigating Intrusions*. John Wiley & Sons.

[126] National Cybersecurity and Communications Integration Center / Industrial Control Systems Computer Emergency Readiness Team (USA). 2015. *NCCIC/ICS-CERT Year in Review*.

[127] National Institute of Standards and Technology (USA). 2012. *NIST SP 800-30 Rev. 1 – Guide for Conducting Risk Assessments*.

[128] National Institute of Standards and Technology (USA). 2013. *NIST SP 800-53 Rv. 4 – Security and Privacy Control for Federal Information Systems and Organizations*.

[129] National Institute of Standards and Technology (USA). 2014. *Framework for Improving Critical Infrastructure Cybersecurity*.

[130] National Institute of Standards and Technology (USA). 2015. *NIST SP 800-82 Rev. 2, Guide to Industrial Control Systems (ICS) Security*.

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

[132] Neil, M., Fenton, N. and Marquez, D. 2007 "Using Bayesian Networks and Simulation for Data Fusion and Risk Analysis." In *Computational Models of Risks to Infrastructure*, Vol. 13, pp. 204–215.

[133] Organisation for Economic Cooperation and Development. 2017. "Types of Cyber Incidents and Losses." In *Enhancing the Role of Insurance in Cyber Risk Management*.

[134] Ortega, J., Radovic, V. and Rios Insua, D. 2018. "Utility Elicitation." In *Handbook of Judgement Elicitation*. Springer.

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

[136] Payne, J.W. 1976. "Task Complexity and Contingent Processing in Decision Making: An Information Search and Protocol Analysis." In *Organizational Behavior and Human Performance*, Vol. 16, No. 2, pp. 366–387.

[137] Striegel, Q., Li, A., and Liao, Z. 2008. "Botnet Economics: Uncertainty Matters." In *Managing Information Risk and the Economics of Security*, pp. 245 – 267. Springer.

[138] Ramirez de la Huerga, M., Bañuls Silvera, V.A., and Turoff, M. 2015. "A CIA-ISM Scenario Approach for Analyzing Complex Cascading Effects in Operational Risk Management." In *Engineering Applications of Artificial Intelligence*, Vol. 46, pp. 289–302.

[139] Rascagnères, P. 2016 "Babar: Espionage Software Finally Found and Put Under the Microscope."

[140] Reichert, P., Langhans, S.D., Lienert, J. and Schuwirth, N. 2015. "The Conceptual Foundation of Environmental Decision Support." In *Journal of Environmental Management*, Vol. 154, pp. 316–332.

[141] Renooij, S. 2001. "Probability Elicitation for Belief Networks: Issues to Consider." In *The Knowledge Engineering Review*, Vol. 16, No. 3, pp. 255–269.

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

[143] Rios Insua, D., Banks, D., Rios, J. and Ortega, J. 2019. "Adversarial Risk Analysis for Structured Expert Judgement Modelling", in *Expert Judgement in Risk snd Decision Analysis*, Springer.

[144] Rios Insua, D., Couce-Vieira, A., Rubio, J.A., Pieters, W., Labunets, K. and Rasines, D.G. 2018. "An Adversarial Risk Analysis Framework for Cybersecurity." In *Risk Analysis* [accepted].

[145] Rios Insua, D., Rios, J. and Banks, D. 2009. "Adversarial Risk Analysis." In *Journal of the American Statistical Association*, Vol. 104, No. 486, pp. 841 – 854.

[146] Rothschild, C. McLay, L. and Guikema, S. 2012. Adversarial Risk Analysis with Incomplete Information: A Level-k Approach. *Risk Analysis*, *32*(7), 1219–1231.

[147] Sabaliauskaite, G., and Mathur, A.P. 2015. "Aligning Cyber-Physical System Safety and Security." In *Complex Systems Design & Management Asia*, pp. 41–53.

[148] Schatz, D. and Bashroush, R. 2017. "Economic Valuation for Information Security Investment: A Systematic Literature Review." In *Information Systems Frontiers*, Vol. 19, No. 5, pp. 1205–1228.

[149] Scheibehenne, B., Greifeneder, R. and Todd, P.M. 2010. "Can There Ever Be Too Many Options? A Meta-Analytic Review of Choice Overload." In *Journal of Consumer Research*, Vol. 37, No. 3, pp. 409–425.

[150] Schneier, B. 1999. "Attack Trees." In *Dr. Dobb's Journal*, Vol. 24, No. 12, pp. 21–29.

[151] Schneier, B. 2003. *Beyond Fear: Thinking Sensibly about Security in an Uncertain World*. Springer New York.

[152] Shachter, R.D. 1986. “Evaluating Influence Diagrams.” In *Operations Research*, Vol. 34, No. 6, pp. 871–882.

[153] Shultz, J.M., Garcia-Vera, M.P. Gesteira Santos, C., Sanz, J., Bibel, G., Schulman, C.,Bahouth, G., Dias Guichot, Y., Espinel, Z. and Rechkemmer, A. 2016. “Disaster Complexity and the Santiago de Compostela Train Derailment.” In *Disaster Health*, Vol. 3, No. 1, pp. 11–31.

[154] Singhal, A., and Ou, X. 2011. *Security Risk Analysis of Enterprise Networks Using Probabilistic Attack Graphs*. National Institute of Standards and Technology (USA).

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

[156] Standards Australia. 2006. *HB 167:2006, Security Risk Management*.

[157] Taeihagh, A., Lim, H. S. M. 2018. “Governing Autonomous Vehicles: Emerging Responses for Safety, Liability, Privacy, Cybersecurity, and Industry Risks.” In *Transport Reviews*, pp. 1–26.

[158] Talbot, J. and Jakeman, M. 2011. *Security Risk Management Body of Knowledge*. John Wiley & Sons.

[159] Taylor, C., Krings, A. and Alves-Foss, J. 2002. “Risk Analysis and Probabilistic Survivability Assessment (RAPSA): An Assessment Approach for Power Substation Hardening.” In *Proc. of the ACM Workshop on Scientific Aspects of Cyber Terrorism*.

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

[161] The Open Group. 2009. *Risk Taxonomy*.

[162] The Open Web Application Security Project Foundation. 2017. *The OWASP Risk Rating Methodology*.

[163] The SABSA Institute. 2009. *The SABSA White Paper*.

[164] Thomas, R.C., Atkiewicz, M., Florer, P., Widup, S. and Woodyard, M. 2013 “How Bad Is It? A Branching Activity Model to Estimate the Impact of Information Security Breaches.” In *Workshop on the Economics of Information Security*.

[165] United Nations Human Rights Council. 2015. *Resolution on The Promotion, Protection and Enjoyment of Human Rights on the Internet*

[166] United Nations Human Rights Council. 2016 *Universal Human Rights Index Database*.

[167] United States Government. 2015 *Presidential Policy Directive 21: National Preparedness*.

[168] US Army Corps of Engineers. 2014. *EM 385-1-1, Safety and Health Requirements Manual*.

[169] van Riel, C.B.M and Fombrun, C.J. 2007. *Essentials of Corporate Communication*. Routledge.

[170] Vandebosch, H., van Cleemput, K. 2008. "Defining cyberbullying: A Qualitative Research Into the Perceptions of Youngsters." In *CyberPsychology & Behavior*, Vol. 11, No. 4, pp. 499–503.

[171] Verendel, V. 2009. "Quantified Security Is a Weak Hypothesis." In *Proc. of the 2009 New Security Paradigms Workshop*, pp. 37–50.

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

[173] Wiper, M, Rios Insua, D. and Ruggeri F. 2001. "Mixtures of Gamma Distributions with Applications." In *Journal of Computational and Graphical Statistics*, Vol. 10, pp. 440–454.

[174] World Health Organization. 2010. *Measuring Health and Disability: Manual for WHO Disability Assessment Schedule, WHODAS 2.0*.

[175] World Health Organization. 2018. *International Statistical Classification of Diseases and Related Health Problems, 11th Revision*.

[176] Zhu, B. Joseph, A. and Sastry, S. 2011. "A Taxonomy of Cyber Attacks on SCADA Systems." In *Proc. of 2011 Int. Cf. on the Internet of Things and 4th Int. cf. on Cyber, Physical, and Social Comp.*, pp. 380 – 388.

[177] Zhuge, J., Holz, T. Song, C., Guo, J., Han, X. and Zou, W. 2009. "Studying Malicious Websites and the Underground Economy on the Chinese Web." In *Workshop on the Economics of Information Security 2008*, pp. 225 – 244.

Universidad
Rey Juan Carlos

Decision Models for Risk Analysis in Cybersecurity

Aitor Couce Vieira