Integrating Adversarial Risk Analysis in Cyber Resilience Assessment to Support SME Cybersecurity Decision-Making

Tuomas Rajala

School of Science

Thesis submitted for examination for the degree of Master of Science in Technology.

Espoo July 31, 2026

Thesis supervisor:

Docent Tuomas Raivio

Thesis advisors:

M.Sc. Lauri Pykälä

M.Sc. Sara Soikkeli

Aalto University
School of Science

AALTO UNIVERSITY
SCHOOL OF SCIENCE

ABSTRACT OF THE
MASTER'S THESIS

Author: Tuomas Rajala

Title: Integrating Adversarial Risk Analysis in Cyber Resilience Assessment to Support SME Cybersecurity Decision-Making

Date: July 31, 2026 Language: English Number of pages: 10+88

Department of Mathematics and Systems Analysis

Degree programme: Mathematics and Operations Research

Supervisor: Docent Tuomas Raivio

Advisors: M.Sc. Lauri Pykälä, M.Sc. Sara Soikkeli

Small and medium-sized enterprises (SMEs) face a rapidly expanding cyber threat landscape driven by geopolitical tensions, the proliferation of Crime-as-a-Service models, and the increased accessibility to offensive capabilities through artificial intelligence. Despite their growing exposure, SMEs frequently lack the resources and expertise to defend themselves effectively, and are increasingly targeted precisely because of this asymmetry as well as the asymmetric financial structures of common cyberattacks being inexpensive to enact while causing expensive destruction in poorly protected organizations. Existing approaches to cybersecurity decision-making in this context rely predominantly on traditional qualitative risk analysis, which lacks the analytical rigour required for complex adversarial environments, and on resilience frameworks that offer high-level design postulates without modelling the behaviour of rational threat actors.

This thesis proposes adversarial risk analysis (ARA) as a methodology to support cybersecurity decision-making and resilience planning for SMEs. ARA models cyber threats not as stochastic events but as decisions made by rational, utility-maximizing adversaries, using Bayesian inference to represent and update the defender's uncertainty about attacker behaviour. The application of ARA produces value along three dimensions: formalizing the attacker-defender interaction structures the organization's understanding of its own information environment; the resulting model yields either an optimal defence portfolio or the elimination of robustly dominated options; and it offers opportunities to update beliefs dynamically to reflect the contemporary threat landscape.

The methodology is applied through two case studies with SMEs in different fields targeted by attack vectors with industry-specific relevance. ARA is applied in these cases to derive Pareto-optimal defence portfolios balancing expected utility against cost, with solution robustness assessed via sensitivity analysis. The results are discussed in terms of their implications for resilience planning, the role of information structure in adversarial modelling, and the broader applicability of ARA in complex cyber interactions. Although appealing in concept, applying ARA to a specific case faces considerable challenges in parameter selection and computational complexity. Nevertheless, the results can be interpreted in the context of cyber resilience, and the modelling process itself surfaces important structural factors, such as the role of information asymmetry in the attacker-defender setup and uncertainty in cybersecurity.

Keywords: Small and Medium-Sized Enterprises(SMEs), Cyber Resilience, Decision-Making Under Uncertainty, Adversarial Risk Analysis, Game Theory, Cybersecurity

AALTO-YLIOPISTO
PERUSTIETEIDEN KORKEAKOULU

DIPLOMITYÖN
TIIVISTELMÄ

Tekijä: Tuomas Rajala

Työn nimi: Vastakkainasetuksellisen riskianalyysin integrointi osaksi kyberresilienssin arviointia pk-yritysten kyberturvallisuuden päätöksenteon tueksi

Päivämäärä: July 31, 2026 Kieli: Englanti Sivumäärä: 10+88

Matematiikan ja systeemianalyysin laitos

Koulutusohjelma: Matematiikka ja Operaatiotutkimus

Työn valvoja: Dosentti Tuomas Raivio

Työn ohjaajat: DI Lauri Pykälä, DI Sara Soikkeli

Pienet ja keskisuuret yritykset (pk-yritykset) kohtaavat nopeasti laajenevan kyberuhkien kentän, jota kiihdyttävät geopoliittiset jännitteet, Crime-as-a-Service-palvelumallien yleistyminen sekä tekoälyn myötä kasvanut hyökkäyskykyjen saavutettavuus. Kasvavasta altistumisesta huolimatta pk-yrityksiltä puuttuu usein tehokkaaseen puolustautumiseen tarvittavat resurssit ja osaaminen, minkä vuoksi niihin kohdistuu yhä enemmän hyökkäyksiä tarkalleen tämän epäsymmetrian sekä yleisten kyberhyökkäysten epäsymmetrisen kustannusrakenteen vuoksi: hyökkäyksen toteuttaminen on halpaa, kun taas huonosti suojatun organisaation kohtaamat vahingot ovat merkittäviä. Kyberturvallisuuteen liittyvä päätöksenteko nojaa perinteisesti kvalitatiiviseen riskianalyysiin, mikä ei huomio monimutkaisia rakenteita eritoten vastakkaisasetuksellisissa asetelmissa. Riskianalyysin komplementtina toimiva resilienssiajattelu ja sitä ympäröivät viitekehykset keskittyy taas häiriöistä palautumiseen ottamatta kantaa rationaalisten uhkatoimijoiden päätöksentekoon

Tämä diplomityö esittää vastakkainasetuksellisen riskianalyysin (adversarial risk analysis, ARA) metodologiana, joka tukee pk-yritysten kyberturvallisuuteen liittyvää päätöksentekoa ja resilienssisuunnittelua. ARA mallintaa kyberuhkia satunnaisten tapahtumien sijasta rationaalisten, hyötyään maksimoivien vastustajien tekeminä päätöksinä, hyödyntäen bayesilaista päättelyä puolustajan hyökkääjän käyttäytymiseen liittyvän epävarmuuden tulkinnassa ja päivittämisessä. ARA:n soveltamisen edut voidaan jakaa kolmeen osa-alueeseen: hyökkääjän ja puolustajan välisen vuorovaikutuksen formalisointi jäsentää organisaation ymmärrystä omasta informaatioympäristöstään; syntyvä malli tuottaa joko optimaalisen puolustusportfolion tai eliminoi selvästi dominoidut vaihtoehdot; ja se tarjoaa mahdollisuuden päivittää uskomuksia dynaamisesti ajankohtaisen uhkakentän mukaisesti.

Metodologiaa sovelletaan kahteen käytännön esimerkkiin, missä eri PK-yritykset kohtaavat heidän osaltaan relevantin hyökkäysskenaarion. Mallien lopputuloksena ARA tuottaa optimaaliset puolustusportfoliot suhteessa odotettuun suojauksen hyötyyn ja kustannukseen, eli Pareto-optimaaliset puolustukset. Lopulta, mallien vakautta arvioidaan herkkyysanalyysin avulla. Tuloksien vaikutuksia tulkitaan resilienssisuunnittelun ja informaation epäsymmetrian osalta kyberturvan viitekehyksessä. Lisäksi ARA:n laajempaa sovellettavuutta tutkitaan suhteessa monimutkaisempiin interaktioihin, joiden seuraksena ARA todetaan olevan realistinen päätösanalyysin pohjautuva metodologia, joka kuitenkin kohtaa mittavia haasteita parametrien valinnassa ja laskennallisessa kompleksisuudessa. Tästä huolimatta tuloksia voidaan tulkita kyberresilienssin viitekehyksessä, ja itse mallinnusprosessi tuo esiin tärkeitä rakenteellisia tekijöitä, kuten informaatioepäsymmetrian roolin hyökkääjän ja puolustajan asetelmassa sekä epävarmuuden sietäminen kyberturvallisuudessa.

Avainsanat: Pienet ja keskisuuret yritykset(pk-yritykset), Kyber resilienssi, päätösanalyysi, vastakkainasetuksellinen riskianalyysi, peliteoria, kyberturva

Preface

The completion of this thesis has been one of the pivotal challenges of my studies, and I feel like this marks the end of my five-year passage at Aalto. I would like to thank Prof. Kimmo Kaski for the opportunity to write this thesis, the folks at Cyberwatch for their industry-specific knowledge, M.Sc. Lauri Pykälä for his counsel, and finally, Docent Tuomas Raivio for his expert guidance throughout the whole thesis process from start to finish.

I would also like to thank my family, friends and my girlfriend for their role in supporting me during my studies and making life wonderful.

Otaniemi,July 31, 2026

Contents

Abbreviations and Symbols

Abbreviations

AI	Artificial Intelligence
APS	Augmented Probability Simulation
APT	Advanced Persistent Threat
ARA	Adversarial Risk Analysis
BAID	Bi-Agent Influence Diagram
BCA	Business Critical Asset
BNE	Bayesian Nash Equilibrium
CER	Critical Entities Resilience
CIA	Confidentiality, Integrity and Availability
CK	Common Knowledge
CMU	Crisis Management Unit
CRA	Cyber Resilience Act
CSF	Cybersecurity Framework
DDoS	Distributed-Denial-of-Service
DiD	Defence in Depth
EDR	Endpoint Detection and Response
FSD	First-order Stochastic Dominance
GDPR	General Data Protection Regulation
ICS	Industrial Control Systems
IMM	Immutable Backups
LLM	Large Language Model
MC	Monte Carlo
MCDA	Multi-Criteria Decision Analysis
MCMC	Markov Chain Monte Carlo
MFA	Multi-Factor Authentication
NE	Nash Equilibrium
NIST	National Institute of Standards and Technology
OT	Operational Technology
PRA	Probabilistic Risk Analysis
SME	Small and Medium-sized Enterprise
TTP	Tactics, Techniques and Procedures
WTP	Willingness-to-Pay

General symbols and notation

G	defender's model of outcomes and opponent behaviour
F	defender's uncertainty over attacker type
$\mathcal{A}$	attacker decision space
$\mathcal{D}$	defender decision space
c_d	total defender cost
c_a	attacker's cost/payoff
c'_a	normalized attacker cost/payoff
$u_d(c_d)$	defender utility
$u_A(c_a)$	attacker utility
$U_A(c_a)$	attacker random utility
$\psi(\cdot)$	expected utility
$\psi_D(s)$	expected defender utility given controls s
$\psi_A(a)$	expected attacker utility given a
k_a	attacker risk-proneness coefficient
c	risk-aversion coefficient
$\succ$	strict preference
Γ	gamma distribution
$\mathcal{N}$	normal distribution
$\mathcal{U}$	uniform distribution

Water utility case notation

B_0	baseline attack load
$c_{a,\max}$	attacker's cost upper bound
$c_{a,\min}$	attacker's cost lower bound
c_i	insurance premium cost
c_{op}	attacker's operating cost per attack
c_s	security protection cost
c_t	fine imposed on attacker if caught
$c_{t,\max}$	maximum detection cost bound
e	impact/gain realized by the attacker
i	insurance policy choice
K	number of Monte Carlo simulation iterations
l	total downtime caused by successful attacks
l_j	duration of the j-th individual attack
m	aggregate impact to the defender
m_{net}	net impact to defender after insurance coverage
N	number of volumetric attack spikes
r	loss rate per hour of downtime
$R_{\max}$	maximum possible damage
s	protection/defence threshold
$\mathcal{S}, \mathcal{I}$	protection threshold set / insurance policy set
S_{gbps}	attacker's modelled attack strength
s_{gbps}	defender's belief on attack strength
t	probability of attacker being caught
$\hat{p}_D(a\|s)$	empirical PMF of optimal attack count given defence s
$\hat{P}_D(a^\star \leq a\|s)$	empirical CDF of optimal attack count
α	gamma shape/rate parameters for attack spike sizes
γ_i	insurance coverage fraction for policy i
λ	poisson rate parameter for attack spike frequency
$\mu_1, \sigma_1^2, \mu_2, \sigma_2^2$	lognormal mixture parameters for loss-rate belief
$\succeq_{FSD}$	first-order stochastic dominance relation

Nursing home case notation

a	attacker's attack decision
η	attacker's belief distribution over defender's insurance status
C_{CMU}	crisis management unit fixed cost
$c_d(d_2)$	defender's cost function for reactive decision d_2
c_{det}	cost imposed on attacker if detected
c_{EDR}	EDR implementation cost
c_i	insurance premium cost
c_{IMM}	immutable backup implementation cost
c_{MFA}	MFA implementation cost
c_{op}	attacker's operational cost
c_s	total security controls cost
d_1	defender's proactive defence portfolio decision
d_2	defender's reactive decision
$\mathcal{D}_1$	proactive defence decision set
$\mathcal{D}_2$	reactive defence decision set
i	insurance tier choice
p_{crisis}	CMU successfully recovers
p_{EDR}	defender's belief on EDR gate success probability
P_{EDR}	attacker's belief on EDR gate success probability
p_{MFA}	shared probability MFA gate stops the attack
p_{rec}	defender's belief on backup gate probability
P_{rec}	attacker's belief on backup gate probability
$R^\star$	optimal ransom demand
$\mathcal{R}(s)$	ransom demand with respect to defence s
$R_{\max}$	maximum feasible damage to the defender
s	security control configuration
S_n	phase of interaction
v_a	attacker's estimated value of exfiltrated data
v_d	defender's estimated loss due to data loss
w_a	attacker's evaluation of defender's willingness-to-pay
α_k, β_k	beta distribution shape parameters
γ_{ransom}	insurance coverage proportion of ransom
γ_{recovery}	insurance coverage proportion of recovery cost
π_k	mixture component weight
Θ	outcome node of Bernoulli gate structure
Ψ_n	backward-induction utility at node n
ξ_n	uniform comparison variates

1 Introduction

1.1 Research Objectives

The cyber threat landscape has grown rapidly in recent years, driven by tensioning geopolitics, the increased popularity of crime-as-a-service attacks, and the rise of large language models[1]. This development is particularly relevant for small and medium-sized enterprises (SMEs), which often lack the resources and expertise to protect themselves effectively against these threats[2]. As a result, decision-makers face the challenge of optimally allocating limited resources to defend against cyber threats.

This problem leads to the research objective of this thesis: to support SME decision-making by modelling the complex space of cyber threats and adversaries' strategic decision-making. Instead of focusing on individual risks and their likelihoods and impacts, the aim is to prioritize the organization's ability to handle disruptions. This perspective, referred to as resilience, describes an organization's ability to remain functional before, during, and after a disruption[3, 4]. Several resilience frameworks rely largely on qualitative, often abstracted assessments of an enterprise's ability to handle disruptions, meaning the specific type of disruption is not explicitly defined[5]. Nor do these frameworks offer the target organization valuable steps to evaluate its existing information structure with respect to adversaries.

Being resilient against such intentional, rational adversaries calls for a methodology grounded in strategic decision-making. Adversarial Risk Analysis (ARA) is such a methodology, assessing attacker behaviour and selecting optimal defensive controls against an expected utility-maximizing attacker[6]. Its key premise is that threats do not arise as random events but originate from a rational adversary seeking to maximize their profit-to-effort ratio[7]. To capture this, ARA represents the defender's uncertainty about the attacker through subjective Bayesian distributions, refined as new evidence emerges. These distributions are then propagated through the attacker's own optimization problem, inducing a distribution over his likely actions and informing how defences should be enacted. In the literature, ARA has often been used to determine optimal defensive controls against a rational attacker across a variety of contexts[8, 9, 10, 11]. In this thesis, ARA will be used to determine optimal defence portfolios with respect to certain attack vectors.

Thus, the research questions of this thesis are as follows:

1. How does formalizing the attacker-defender interaction through adversarial risk analysis support cyber resilience planning under uncertainty for SMEs?
2. How can adversarial risk analysis be used to model attacker decision-making and structure the information environment of cyber threats facing SMEs?
3. Does the use of adversarial risk analysis improve upon traditional risk analysis methods in producing optimal defences?

To summarize the research objectives of this thesis, the goal is to aid the resilience-building and decision-making of SMEs through adversarial modelling of rational threats.

1.2 Background

As stated earlier, the recent developments in artificial intelligence and increasing geopolitical tensions have contributed to the rapid expansion of the cybersecurity threat landscape[1]. As digital technologies become more embedded in organizational operations, the need to assess cyber risks and respond robustly to disruptions has become increasingly important. At the same time, advances in artificial intelligence risk broadening access to offensive capabilities, enabling less technically skilled attackers to carry out increasingly sophisticated attacks, particularly through enhanced social engineering and phishing techniques[12]. Furthermore, the adoption of AI technologies within organizations can expand the potential attack surface, introducing new vulnerabilities into digital systems[13].

Cyber attacks are often motivated by financial gain, strategic espionage, or the intent to cause political disruption[14]. While financially motivated attackers often target organizations that offer the highest potential returns, large corporations typically allocate significant resources to cybersecurity and maintain specialized teams to manage digital threats[15]. As a result, small and medium-sized enterprises (SMEs), which often operate with limited cybersecurity resources, have increasingly become attractive targets for attackers[2]. This development is particularly concerning because many SMEs also operate within supply chains or sectors connected to critical infrastructure, such as energy production and water utilities[16]. These risks concerning SMEs become even more pronounced as the cost of Crime-as-a-Service trend downwards and the increased accessibility to such services can cause a surge in the volume of cyberattacks[17].

Consequently, SMEs face growing cyber risks while often relying on traditional qualitative risk analysis approaches such as risk matrices. These methods can lack sufficient analytical rigour in complex threat environments and may lead to reduced risk resolution and suboptimal resource allocation[18, 19]. In response to these limitations, the earlier discussed resilience has gained increasing attention in cybersecurity research. Rather than focusing solely on estimating the likelihood and impact of individual risks, resilience-based approaches emphasize an organization's ability to absorb disruptions, maintain operations, and recover from cyber incidents. This shift in perspective moves away from exhaustingly assessing each risk's respective impact and likelihood and instead focuses on the internal redundancies and robustness[20, 5]. The improvement and creation of these resilient qualities in the context of cyber threats requires the target organization to evaluate their information structure with respect to malicious actors, and actively model how such an actor would elicit information and how possible attack scenarios would play out.

While qualitative and quantitative risk analysis are well-established fields, resilience assessment remains relatively underdeveloped and less clearly defined. In literature, the operational resilience of an organization is often modelled using a performance curve, which visualizes the performance of a system or a company at different points of the disruption cycle[21]. Other methods of resilience evaluation include models such as the Linkov Matrix[22], the 4R framework[23], and the NIST Cybersecurity Framework 2.0[24], which often break down the resilience characteristics of an organization into smaller components, such as organizational and technical resilience. The inherent limitations of cyber resilience frameworks in literature is that they mostly offer high-level postulates for resilient systems design without necessarily providing industry-specific guidelines and frameworks around which to model the behaviour of rational actors.

As ARA is a relatively novel approach for evaluating attacker behaviour, its application in cybersecurity has been limited, and its use in resilience assessment is effectively unexplored. Existing literature on ARA currently focuses on attacker-defender games, where the focus is on choosing the optimal defence controls with respect to the utility-maximizing attacker, while this thesis aims to argue that ARA as an analysis tool is fundamentally an act of resilience planning. Specifically, propagating uncertainty through realistic information structures, as well as making explicit the incentive structures of attackers, causes ARA to function as a resilience-building paradigm that also supports the defender in their assessment of optimal defences.

1.3 The structure of the thesis

The structure of this thesis is as follows. Chapter 2 establishes the current adversarial landscape by reviewing common attack vectors and threat actor taxonomy such that the unique characteristics of common cyberattacks are understood. Chapter 3 then assesses this modern threat landscape in the context of SMEs and how cyber resilience arises as a complement to traditional risk analysis methods. The inherent deficiencies in modelling adversarial uncertainty in common cyber resilience frameworks is highlighted as a relevant research gap this thesis aims to answer. Chapter 4 presents the methodological core of this thesis by introducing ARA with its theoretical background, a literature review and an algorithmic breakdown of its core machinery. Chapter 5 then implements this choice in methodology within two case studies using qualitatively different attack scenarios. The chapter constructs the attack scenarios using postulates of ARA, formulates priors using industry literature and produces a set of optimal defences with respect to utility maximizing attackers along with sensitivity analysis. Chapter 6 includes discussion how formalizing the attacker-defender interaction supports resilience planning under uncertainty, how attack effectiveness is substantially determined by the elicited information structure, and what the broader applicability of ARA looks like in more complex adversarial settings. Chapter 7 concludes the thesis with summarizing key findings and proposing directions for future work.

2 Anatomy of Cyber Attacks and Cybersecurity

To fully understand the state of modern cybersecurity, it is vital to discuss ongoing trends in the industry from the point of view of organizations and attackers. With the ongoing proliferation of cybercrimes caused by developing LLMs and tense geopolitical situations, the incentive structures of organized cybercriminals are vital to assess with the goal of identifying points of improvement in an organization. This section will discuss common cyberattack structures and industry frameworks, while discussing the incentive structures of different types of cybercriminals.

2.1 Attack life cycle and structure

The structure of decision-making during a cyberattack has been researched extensively within the literature, and uncovering patterns in attacker incentives, motivations or active deterrents can uncover structures around their decision-making against which to defend. As cyberattacks are rarely individual decisions or incidents, but rather result from a continuum of sequential decisions, it is vital to discuss some notable frameworks that model such decision-making.

One of the earliest frameworks formulated to postulate the unfolding of cyber attacks into distinct phases is called the cyber kill chain developed by Lockheed Martin[25]. The cyber kill chain was formulated in response to the increasing prevalence of Advanced-Persistent-Threats(APTs), which are often state-sponsored, with abundant resources or sustained perseverance who can discover novel zero-day exploits in a system. These threats are particularly difficult to address due to their sophistication and discreet dwell-time prior to attack execution causing the need to evaluate the different phases of the attack. A visualization of the kill chain can be seen in Figure 1:

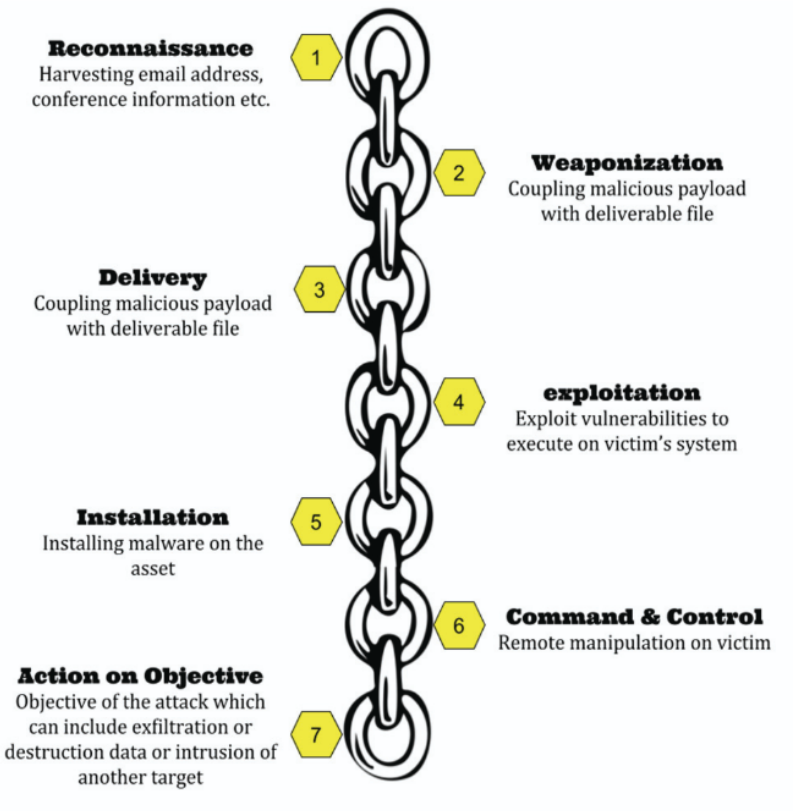

Figure 1: Cyber kill chain [26].

As these phases allow the defender to enact courses of action per phase, this can offer meaningful guidelines for resilience especially given data on novel attacks. As the kill chain offers a chronology of an attack and proposes that a disruption of the attack during any phase causes a deterministic stop to the attack, its strong time-dependence does not account for multiple attack campaigns occurring at once, nor does it consider the non-linearity of common cyberattacks especially zero-day exploits.

These limitations were addressed in the MITRE ATT&CK knowledge base[27], which lessens the focus on the temporal order of cyberattacks and focuses more on the decomposition of high-level objectives into smaller, more tangible methods deployed by the attacker. Fundamentally, the framework decomposes adversarial goals into tactics, each achieved through documented techniques and sub-techniques. In the literature, this ensemble of methods is referred to as tactics, techniques and procedures (TTPs). This decomposition of decision-making allows the defender to improve their strategic decision-making and selection of mitigation practices by evaluating the motivations behind a given attack technique. Figure 2 showcases the structure of the framework.

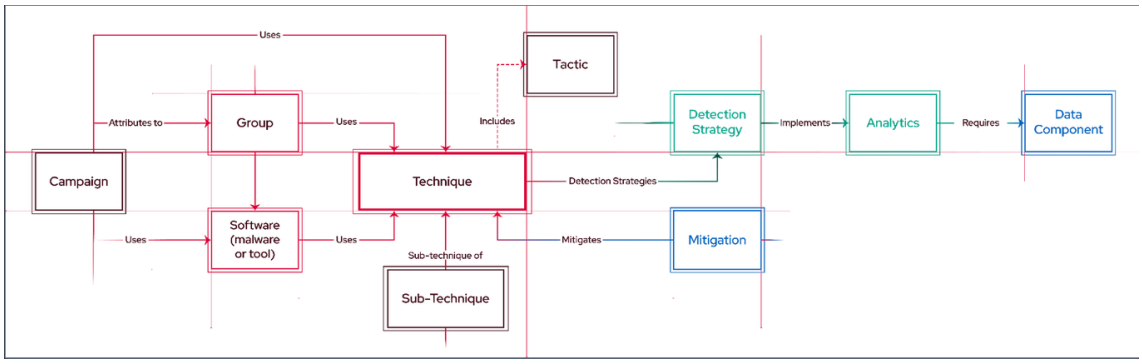

Figure 2: MITRE ATT&CK framework structure[28].

Within the context of this thesis, the sequential nature of attacker decision-making is central. A cyberattack perpetrated by a rational actor is modelled as being a campaign of sequential decisions iterated upon past attempts with the goal of maximizing perceived utility at each step. These decisions are made under uncertainty and fundamentally the success of the attacker's decision is determined by the level of information and perseverance at each step.

2.2 Attacker taxonomy

Considering that the attacker decision-making process is inherently dependent on their level of information, that level of information is also dependent on the attacker's incentives and resources at hand. This motivates our discussion on the dimensions determining the level of capability and motivation for threat actors, and how this categorization will be implemented within attacker modelling.

Fundamentally, attackers can be evaluated using the dimensions of skills and motivation[29], which in this thesis will be treated as the overarching axes along which attacker types are distinguished. These dimensions can offer insights into the level of sophistication and perseverance of the attacker, information that in turn informs what level of strategic decision-making can be expected from them. Cyberattackers in the literature are classified along these dimensions and referred to as novices, hacktivists, professionals and insider threats among many other terms[30]. The motivations of a cybercriminal differ drastically, as opportunistic attackers mostly focus on financial gain while state-sponsored, sophisticated attackers can focus on creating disruptions in infrastructures and causing societal degradation. An attacker of increased sophistication and purpose also makes decisions with respect to a long-term goal, while low-skill, low-motivation attackers such as script-kiddies often focus their efforts on deploying low-effort, high-volume attacks across potential victims with methods such as automated phishing campaigns, where sequential decision-making is mostly absent.

As opportunistic cybercriminals do not invest as much time in reconnaissance and attack preparations, their attacks are not conducted with respect to an accurate evaluation of victim defences, and such an attacker does not adapt in real-time to the

defences enacted. Opportunistic attackers therefore require no adversarial modelling, as their behaviour is not conditioned on the defender's posture and can instead be treated as a draw from Nature[31]. For this class of attacker, historical incident data, when available, provides a sufficient basis for likelihood estimation, allowing the defender to construct a probability distribution over attacker actions through standard statistical inference.

This thesis will therefore focus upon threat actors of sufficient resources and a level of strategic decision-making with respect to observable signals from the defender. This expectation of rational adversaries motivates our need for modelling adversarial interactions, and since an attacker with ample resources and sufficient perseverance is inherently impossible to completely defend against, a methodology that can offer optimized defences with respect to attacker behaviour can be of value to the defender. As this thesis is written from a European perspective and the ongoing Russian aggression continues to tighten the geopolitical situation, it has been shown in the literature that powerful authoritarian governments are more likely to deploy cyberattacks[32], thus justifying further the choice to focus on strategic and rational adversaries.

3 Cybersecurity and Cyber Resilience in SMEs

To understand the environment in which these attackers operate, it is necessary to cover the fundamentals of cybersecurity such as established best practices and the regulatory frameworks that mandate certain risk management procedures. It is also vital to point out the general issues faced by small and medium sized enterprises in not only meeting the mandated requirements but also remaining generally vigilant and proactive in light of the evolving cyberspace, especially considering regulations and directives apply to all enterprises with the caveat of SMEs having much smaller budgets to comply with. These central challenges therefore cause the emergence of an additional characteristic separate from traditional forms of risk management: resilience where the organization's focus shifts from external threats and risks to an inward assessment of its ability to handle disruptions, whatever form they may take. The central resilience frameworks from industry and literature are introduced with a concluding chapter discussing the fundamental role of uncertainty in cybersecurity and resilience planning.

3.1 Cybersecurity in SMEs

3.1.1 Fundamentals of cybersecurity

Cybersecurity is the ability to defend and protect the respective cyberspace from cyberattacks[33]. As the world grows increasingly digitalized, the number of possible attack vectors increases in kind, with cyberattacks increasingly directed at the cyberspace of individuals and enterprises alike. The increased digitalization also gives

way to the cyberspace being interconnected with the physical world, with many industries being fundamentally dependent on IT-infrastructure and digital control systems. This causes the general terms of cybersecurity to be not dissimilar to general security and safety, meaning that cyberattacks often have clear ramifications in the physical world causing societal harm if critical infrastructures experience disruptions. Notable examples of cyberattacks causing major disruptions on critical infrastructures include the 2015 cyberattack against the Ukrainian energy grid, the WannaCry ransomware attack and the Triton malware attack against a Saudi-Arabian petrochemical plant. Essentially, cybersecurity can be thought of as the first line of defence against critical infrastructure disruptions. Even more critically, as critical infrastructures are often interconnected, a failure in one system can propagate to others leading to a cascading failure[34].

Having established the operating environment in terms of cyber threats and attacker taxonomies, it is vital to discuss and bring forward general principles and good practices around cybersecurity that uphold protection against said threats. A foundational framework for cybersecurity is the CIA triad[35], which defines the objectives of security as safeguarding the confidentiality, integrity and availability of information and systems. Most defences or controls in cybersecurity can be attributed to protecting one or more of these properties, and consequently, most attacks against a system can be attributable to violating one or more of these central properties. In practice, most defences enacted in isolation are not sufficient in protection, since cyberspace spans across multiple layers of possible entry points causing a vast network of interconnected decisions. A defence principle known as defence-in-Depth(DiD) addresses this by operationalizing defences through controls layered across multiple domains. The fundamental principle of this practice is to ensure that a potential attacker must overcome defences not merely at a single point, but across multiple, overlapping layers of defence thereby forcing the expenditure of additional resources and effort[36]. This also motivates the principle of least privilege[37], which functions as a tenet of cybersecurity by limiting permissions and access rights. Figure 3 showcases the layered structure of DiD within which a Business Critical Asset(BCA) is protected.

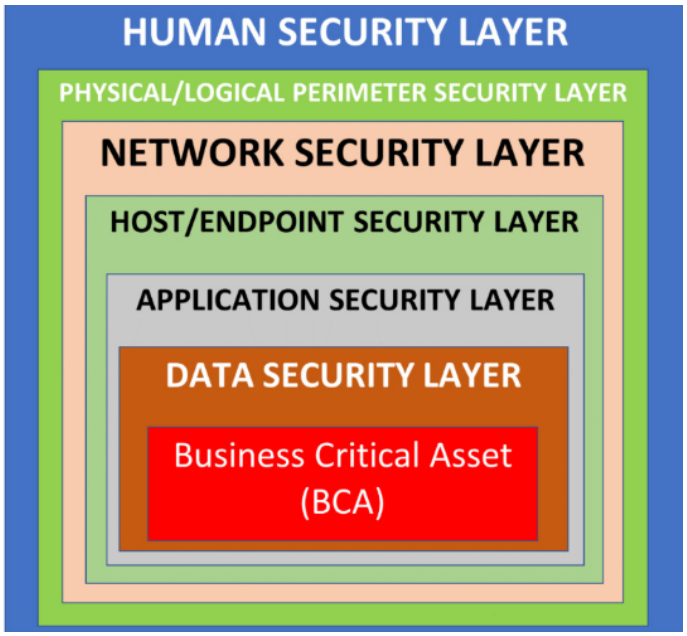

Figure 3: Defence-in-Depth visualization [38].

As in many other fields coping with risk, cybersecurity is concerned with managing risk rather than eliminating it, since elimination is neither economically feasible nor practically possible. Therefore, robust cybersecurity is to be done with the above mentioned principles under uncertainty and with respect to a budget. There exists a field of economics that specifically focuses upon qualitatively optimizing this level of spending in cybersecurity referred to as cyber risk economics, containing models such as the Gordon-Loeb model[39] or Factor Analysis of Information Risk[40]. These economic models lay the foundation for treating cybersecurity investment as an optimization problem, though they typically rely on qualitative or subjectively-assessed inputs, such as the probability of attack or the value of data, and do not explicitly account for the strategic decision-making of adversaries.

3.1.2 Regulatory context in the EU

To manage the risks brought upon the volatile cyberspace, many governing bodies have instilled regulations and directives to ensure enterprises apply enough protection against cyber threats and risk management procedures are unified. The most notable regulation in this context is the NIS2 directive, active in the EU, concerning over 160,000 entities covering 18 critical sectors. It ensures that member states and their respective entities under their directive enhance their cybersecurity capabilities by upholding certain risk management measures, sharing information and enforcing cybersecurity measures[41]. This is achieved by mandating each member state to adopt their own national implementations, which in Finland are implemented in the cybersecurity law"[42]. In the context of this thesis, NIS2 is relevant as it mandates a certain level of business continuity practices, incident reporting, and risk management procedures that are mandatory irrespective of the operating budget of the covered

critical sectors.

As NIS2 is fundamentally a cyber risk management and governance directive concerning critical sectors, the cyber resilience act(CRA) is a regulation more broad in its definition and inclusion of sectors[43]. The goal of the CRA is to ensure software and hardware built in the EU are designed, updated and maintained for safe usage throughout the product lifecycle. This regulation concerns manufacturers, product developers, third-party suppliers and distributors and is upheld by mandatory cyber risk assessment during product lifecycle and vulnerability handling.

The General Data Protection Regulation (GDPR)[44] effective in the EU also creates guidelines for sensitive data handling ensuring adequate risk management procedures, compliance with security best practices and mandatory notification procedures given a cyberattack. In the context of this thesis, GDPR in essence is an additional guideline or an enforcement on companies to uphold cybersecurity. Given a cyberattack resulting in the exfiltration of sensitive data, an enterprise that is found to have violated its obligations under the GDPR may be held liable and subject to administrative fines.

Finally, the Critical Entities Resilience(CER)[45] directive, effective in the EU, focuses on the resilience of critical entities with respect to hazards be that man-made, accidental, intentional or natural. Its key standpoint is preparing for resilience in light of any risk being realized, rather than focusing on specific threat types. This is achieved by requiring Member States to adopt a national approach to critical entity resilience, which encompasses periodic risk assessments, identification of critical entities, obligations to strengthen resilience across critical sectors, and incident notification procedures. Notably, CER complements NIS2 in formulating a more cohesive and unified approach to safeguarding critical infrastructures as CER focuses fundamentally on operational and physical resilience, while NIS2 addresses cybersecurity risk management procedures.

In the context of this thesis, an environment based on regulations and national implementations of directives creates a basis and a scaffolding around cyber procedure and risk management irrespective of budget or capabilities of a company, therefore creating a base layer of defence that can be built upon.

3.1.3 SME-specific challenges in cybersecurity

Since SMEs employ nearly two-thirds of the workforce in the EU[46], it is vital, especially in the context of this thesis, to discuss the unique and difficult challenges faced by SMEs in enacting robust cybersecurity practices. Junior et al.[2] researched the key challenges concerning SMEs in the modern era of cybersecurity, the contents of which will be discussed in relation to central cybersecurity practices such as CIA and defence-in-depth.

Generally, SMEs are commonly lacking in awareness in terms of cyber risks. This

is a key issue as a lack of awareness can create a culture of lacking vigilance therefore compromising one of the most important lines of defence in cybersecurity, human behaviour[47]. This aspect to lacking awareness and false perception of safety can propagate throughout the company from executives not prioritizing cybersecurity initiatives[48] to employees engaging in risky behaviour. This sentiment is supported by Wynn, Jr et al.[49] who stated that SMEs commonly experience a false sense of safety courtesy of the 'security by obscurity' fallacy as small companies expect not being targeted for their relatively small profits or profile within the industry. The lack of awareness is also manifested as overly optimistic risk assessments when it comes to the likelihood of a cyberattack[50] which is exacerbated by the self-serving bias in cybersecurity[51] where stakeholders systematically overestimate their cyber maturity. Chidukwani et al.[52] additionally discovered significant disparities between SMEs self-reported cybersecurity practices and evidence from expert assessments. In relation to cybersecurity standards, this general lack of awareness and lax behaviour undermines the over-arching theme of DiD as seen in Figure 3 as even a fully protected enterprise is susceptible to cyber attacks given employees of low vigilance and understanding of the risks involved while also fundamentally negating the elements of CIA.

Additionally, the resource pool of SMEs is often very constrained and the companies in question rarely allocate resources to cybersecurity proactively. This sentiment is supported by Tedeschi et al.[53] who determined that while larger corporations invest proactively to cybersecurity seeing it as a critical component of business continuity, SMEs tend to view cybersecurity as a cost causing variability in the amounts invested. This lack of investment manifests as outdated software, external or nonexistent IT support and expertise, lack of contingency planning and awareness training[54]. The cost of possible cyberattacks is also especially difficult for SMEs to evaluate resulting from the lack of expertise[55]. The result of this chronic underinvestment and resource constraints is that the focus is mostly on reactive controls while principles surrounding redundancies and diversification of controls are not prioritized.

Cyberattackers are increasingly targeting SMEs as this lack of resources and expertise causes the general cyber posture and resilience to be lower than larger corporations, which often have dedicated cybersecurity teams and a larger budget[15]. This threat space is exacerbated even further by cybercriminals who increasingly exploit the supply chains of large companies often comprised of SMEs as initial access points to infiltrate the larger, typically more secure organizations[56]. This becomes even more pertinent as the motivations of cybercriminals differ from monetary gain to politically motivated disruptions of critical infrastructures, as then the effects of the cyberattack cause societal harm as well as disrupt the normal operations of the company.

SMEs are often attractive targets for numerous types of cybercriminals, ranging from financially motivated script-kiddies to state-sponsored malevolent actors. For a

cybercriminal aiming to profit by encrypting a company's data and demanding a ransom, a small company holding sensitive data is a viable target as under GDPR, the leakage of such data can be subject to a fine[57], in addition to reputational damage. A state-sponsored adversary, by contrast, may prioritize disrupting critical infrastructure entirely, for instance by executing a DDoS attack against a power plant's control systems. Regardless of attacker motivation or risk preference, this range of scenarios illustrates the same underlying problem: the current security posture and attitude of many SMEs show a significant discrepancy when compared to the modern cyber threat landscape. This discrepancy stems from the fact that traditional layered defences and core information security principles, as conventionally understood, are difficult or unsustainable for SMEs to implement given persistent budget constraints and a lack of expertise. As a result, resilience emerges as a natural complement to traditional risk analysis.

3.2 Resilience

3.2.1 Definition of resilience

The definition of resilience depends largely on the context that it is evaluated within but the general notion and definition of resilience is defined by Martin-Breen and Anderies.[58] as "*Systems are dynamic, undergoing constant change. Resilience in these systems can be defined as maintaining system function in the event of a disturbance.*". In the context of cybersecurity, Hilger [59] describes cyber resilience as an emergent property of how organizations, technologies, and governance structures interact and evolve over time, while Björck et al.[60] and Linkov et al. [61] define it as the ability of the organization or the system to prepare, absorb, recover and adapt to adverse effects such as cyberattacks while remaining operational, minimizing losses and adapting in real time to novel threats. As stated by Kaplan et al.[62], resilience is both a state and an outcome and the process that leads to said state or outcome.

The motivation behind employing a supporting point of view in the form of resilience has risen from the need to supplement risk analysis and management as some fields, especially cybersecurity, has experienced a proliferation in the amount of risks causing qualitative risk management procedures to become unable to identify cascading failures or possibly causing sub-optimal resource allocation[18, 61]. This sentiment is supported by Aven et al.[63] who stated the following: *"The resilience field arose as a supplement to the traditional probabilistic risk assessment approach, which has strong limitations in analysing many types of real-life systems, particularly complex systems that are characterized by large uncertainties and the potential for surprises. By strengthening the resilience of the system, the safety is enhanced without the need to perform risk calculations."*. This distinction is captured by the Rumsfeld knowledge matrix (Figure 4), where unknown unknowns represent risks that cannot be anticipated in advance. In cybersecurity, zero-day vulnerabilities are commonly viewed as examples of unknown unknowns as defenders cannot prepare for the specific vulnerability before it is discovered, but they can prepare for the possibility of such

unforeseen threats through resilient security practices.

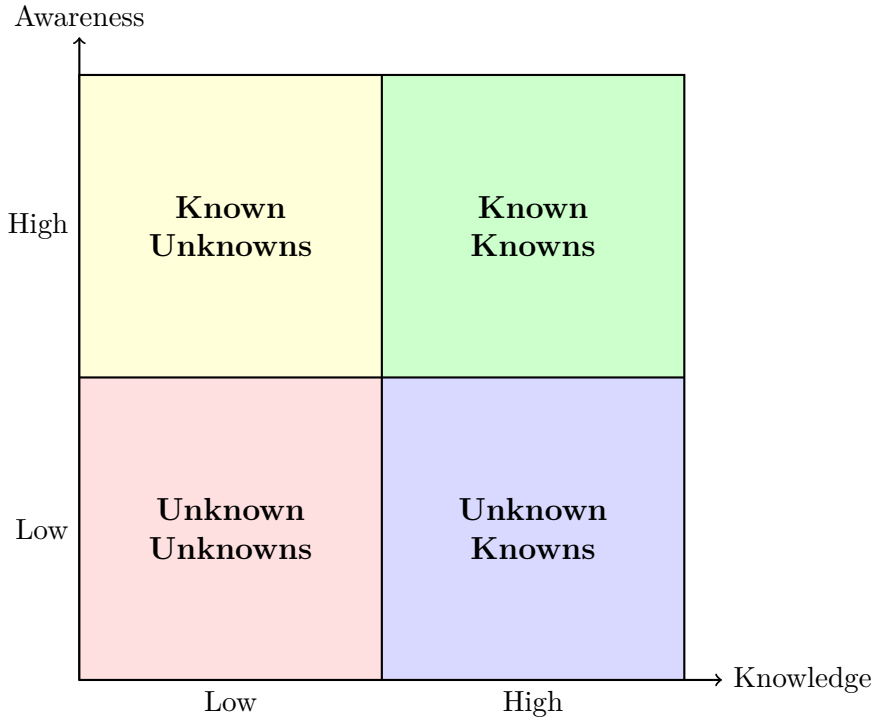

Figure 4: The Rumsfeld knowledge matrix[64].

Traditional definitions of resilience arose in the context of disaster loss reduction[4] thus it is also important to discuss the nuances of operational resilience in the context of SMEs. Fundamentally, resilience for SMEs centers on the enterprise's to restore its pre-disruption state while learning from the disruption itself[65]. In practice, these disruptions often refer to events such as major economic downturns, market volatility, or natural crises, among many others, depending on the operating environment of the SME in question or, as in this thesis, the adverse events are intentional cyberattacks. Given the unique characteristics and challenges of SMEs, as discussed in Section 3.1.3, their resilience extends beyond traditional definitions by including leadership engagement, organizational culture, and external cooperation [66].

3.2.2 General frameworks of cyber resilience

Given the fluctuating definitions of resilience, the real challenges arise when resilience is to be quantified for an organization or a system. This section introduces frameworks, standards and metrics for determining resilience of a system while identifying key areas of improvement.

A concrete example of an international standard for resilience is the ISO 22316 standard, which focuses on embedding resilience into an organization through leadership, culture, and operations. Rather than implementing a clear framework, it

focuses more on the dimensions that form resilience including leadership, situational awareness, and business continuity. The standard also defines that "*Organizations can only be more or less resilient; there is no absolute measure or definitive goal.*". Within the context of this thesis, this statement is vital as we are aiming to support resilience planning and create new dimensions in which to evaluate said resilience

The National Institute of Standards and Technology (NIST) is a U.S. federal agency that provides standards, unified measurements, and technical frameworks for diverse industries. They have defined a framework for cybersecurity: NIST CSF 2.0[24]. This represents not a descriptive approach to resilience but rather an operational view on what organizations and companies should implement across different stages of disruptions. Figure 5 showcases the overview of the framework.

Figure 5: NIST Cybersecurity Framework[24].

To complement the framework further, it also includes a tier-structure that illustrates the level of rigour the target organization has with respect to its cybersecurity risk governance and management ranging from partial to adaptive. In the context of this thesis, this tier structure centrally provides context on how cybersecurity risks are viewed and what processes are in place, the structure of which is shown in Figure 6.

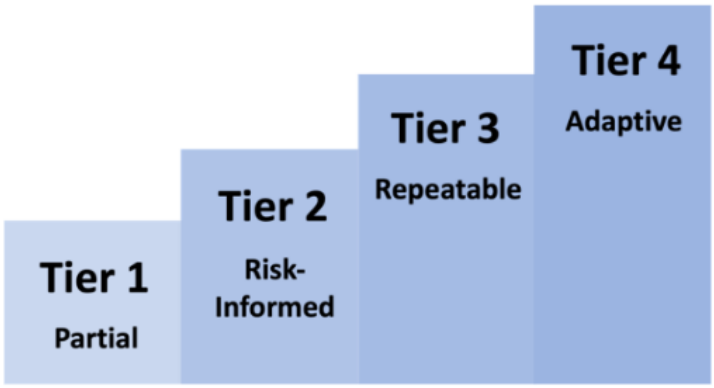

Figure 6: CSF tier structure for cybersecurity risk governance[24].

A notable cyber resilience framework that takes a holistic view of the system and guides the selection of resiliency measures is the Cyber Resiliency Engineering Framework (CREF), developed by MITRE to provide structured cyber resiliency goals, objectives, techniques, and implementation approaches [67]. It decomposes each high-level goal of resilient systems down to objectives and sub-objectives, making it easier to specify action points from high-level goals as visualized in Figure 7.

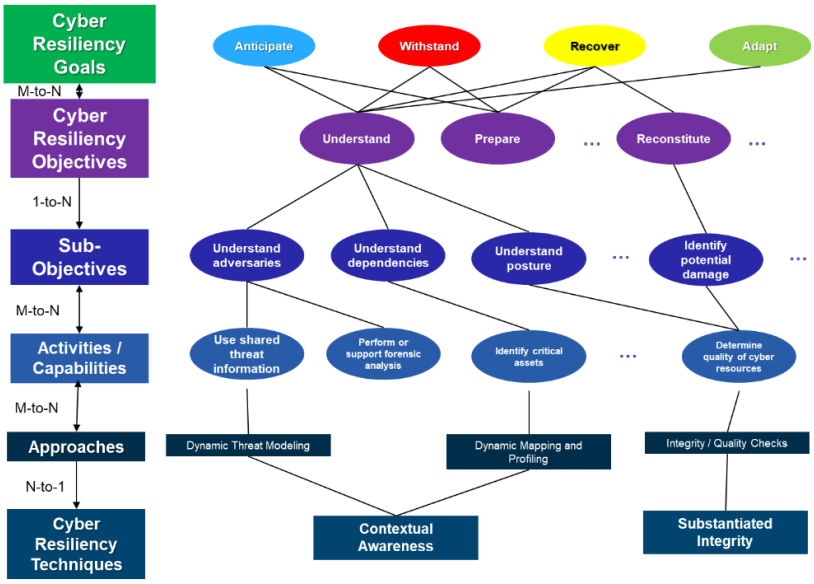

Figure 7: CREF Framework for cyber resilience.

A notable goal of resilience is mitigating damages caused by disruptions, which can be done effectively by containing said disruption, thereby preventing cascading failures in the system. A report published by ENISA in 2012[68] sought to define resilience as an end-to-end process concerning all the components in the infrastructure. ENISA defined a useful visualization for the possibility and event cycle of a cascading failure seen in Figure 8 referred to as the collapse ladder:

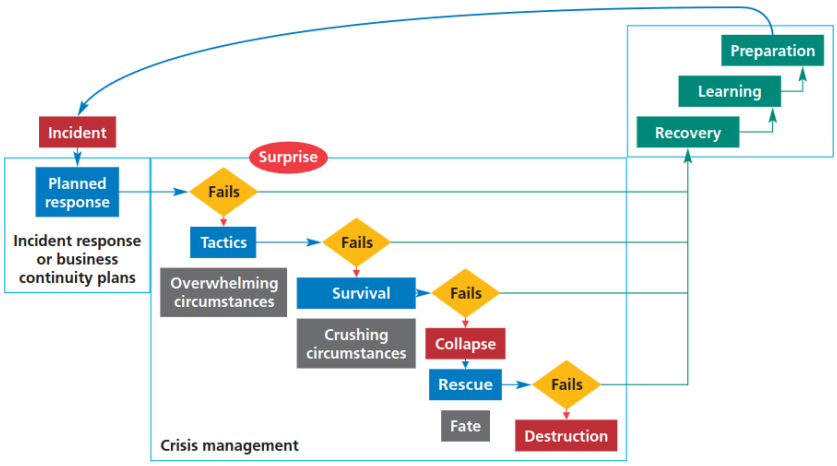

Figure 8: The collapse ladder[68].

3.2.3 Frameworks of cyber resilience in literature

One of the first frameworks formulated for quantifying resilience in literature was formed by Linkov et al. in 2013[22], and later refined by them to fit the context of cyber systems[5]. The framework was built on using two different dimensions related to crisis management and the dimensions of an organization. As stated before in the previous section, a system's resilience is dependent on its performance in four key stages: preparation, absorption, recovery, and adaptation[5, 69]. Linkov et al. also included in the framework lessons learned from military doctrines to divide the system into 4 domains, which should be able to communicate and decentralize decision making in the organization. The four domains are physical, information, cognitive, and social[5]. These two frameworks together form the Linkov matrix, which showcases each domain's risk mitigation acts with respect to each moment in the event cycle. Figure 9 showcases an example of Linkov's resilience matrix.

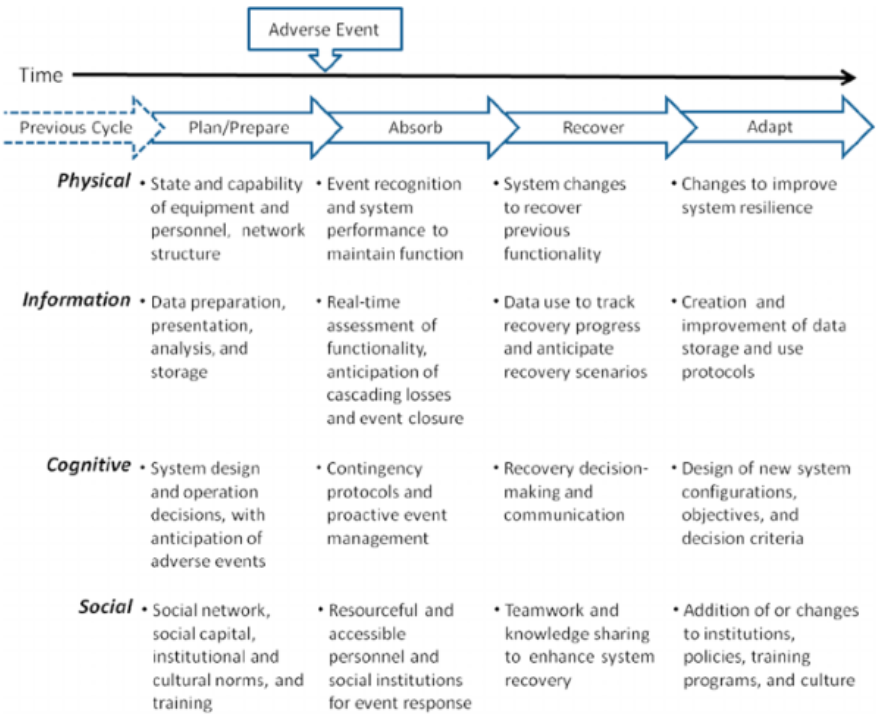

Figure 9: Linkov resilience matrix[22].

The 4R, as defined by Bruneau et al.[23, 4], is a framework is comprised of 4 elements:Robustness, Redundancy, Resourcefulness, and Rapidity. All of these are characteristics of the Linkov matrix and CSF, where the goal is to ensure the organization is robust, resourceful, and active during a crisis. The 4R framework is especially analogous to the defence-in-depth as defences allocated across layers creates redundancies in the system. In Figure 10 we see a graph from Bianchi et al., where the four different resilience metrics are highlighted within the disruption event cycle[21].

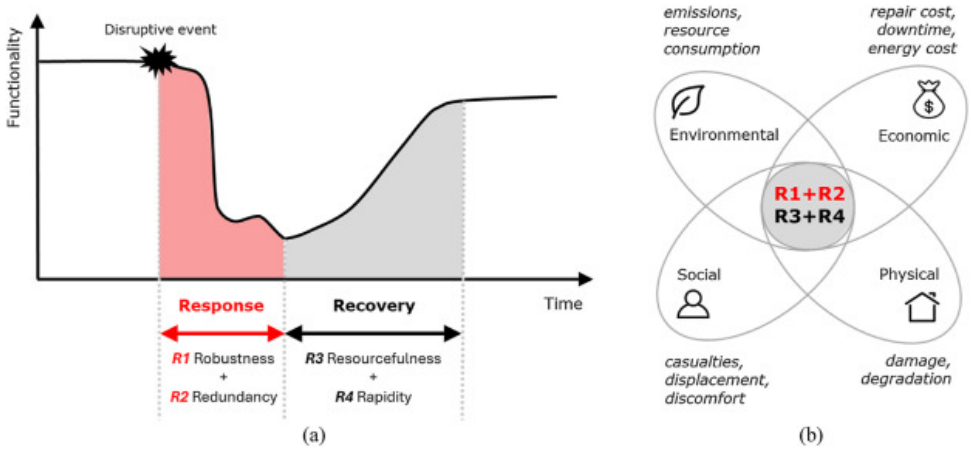

Figure 10: 4R framework within disruption event[21].

This framework is also closely tied to Linkov's matrix where we assess the differ-

ent parts of the disruption event cycle to determine where the organization is least active and resilient. In [4], they also added additional domains to the 4R framework to make issues around each metric more comprehensible. These dimensions are technical, organizational, social, and economic. As for the temporal structures of the framework, during disruption response period, the organization ideally has redundancies in place to ensure the system does not experience cascading failures resulting from one component disruption, and the organization is robust in acting quickly to mitigate damages. During recovery, certain goals should be prioritized to recover system functionality and resources should be allocated quickly towards said priorities. This framework does not account for improving the system after a disruption, as the Linkov matrix does. This is important to note, because modern resilience is fundamentally built on learning gained from crisis situations.

A modern contribution to the cyber resilience framework ensemble is the cyber resilience cube presented by Hilger[70] in which the two-dimensional nature of the Linkov matrix was extended by an additional dimension of system scale. This was motivated by real-world cyber scenarios that showcase failures of not just individual systems but cascading failures moving across system scale essentially ramping up in scale and significance. Figure 11 visualizes the dimensions of the cube.

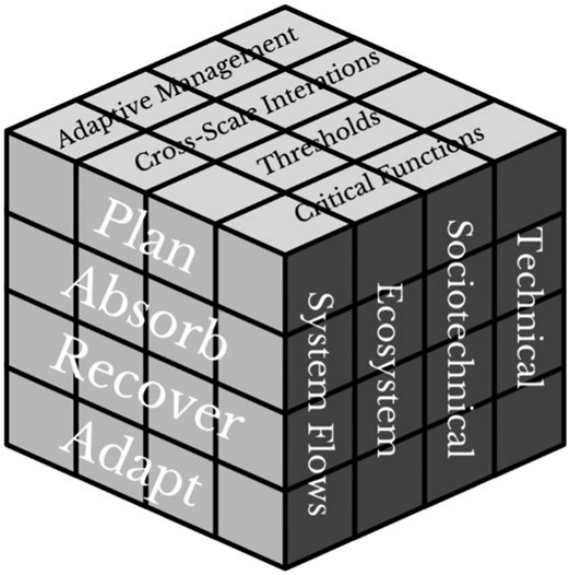

Figure 11: Cyber resilience cube[70]

Using this structure, a vague statement such as "considering the preparedness of the technical components of an system" can be decomposed along the system-scale dimension, meaning one can identify whether the system is resilient only at the technical or organizational level, or also as part of a larger ensemble.

3.3 Cyber resilience under uncertainty for SMEs

To further highlight this problem scenario, we introduce the core sources of uncertainty in cybersecurity, which, when compounded with the modern developments and exacerbations of cyberspace, create the need for additional methods of ensuring resilience.

As discussed, cybersecurity and cyber threats are constantly evolving, the volatility of which not taking into account the persistent underreporting in cybersecurity[71]. Organizations are rarely incentivised to disclose information regarding past breach data for fear of reputational risk, causing the information base regarding attack volumes and strategies to be skewed significantly. As well as this, the regulations surrounding cybersecurity are far from standardized internationally, causing inconsistent reporting. Even given perfect data on past incidents, i.e. epistemic uncertainty is negated, it is worthwhile to note that cybersecurity struggles with a lack of quantifiable metrics and measures[72]. This also gives rise to the untested defence paradox in cybersecurity as a control never breached is not an indication of perfect protection but rather a lack of a sufficiently sophisticated attacker[73]. These phenomena create persistent uncertainty in cybersecurity, which motivate the need for resilience.

In the context of resilience assessment, we see that most of the resilience frameworks either include high-level instructions and techniques to improve resilience without specifying a method of implementation, or some viable techniques are listed without considerations of their relative importance or scheme for prioritization. This is generally also noticeable in cybersecurity directives, which, as well as cyber-resilience frameworks, provide scaffolding around which to build cybersecurity, while also introducing complexity for SMEs leading to dependence on consultants and external IT providers. Such compliance-driven approaches also emphasize adherence and risk mitigation over proactivity, creating a rather static stance on cybersecurity [66]. Given that resilience is determined with respect to disruptions, it is vital to point out that in cybersecurity, resilience does not exist in isolation but within adversarial interactions that have to be inspected. Inspecting these interactions using the structures discussed in Section 2.1, identifying clear and actionable insights to improve resilience is a non-trivial task, as the parameter space of possible outcomes is vast, and the pool of resources from which investment decisions are made creates a multi-attribute problem of investment allocation [65], which must be optimized given that budget has been identified in the literature as the primary barrier to cyber resilience [66]. This problem scenario motivates the coming sections of this thesis, the hypothesis of which is shared by Zebrowski et al. [74] who highlighted the usage of Bayesian networks in resilience planning.

4 Adversarial Risk Analysis as a Tool in Adversarial Interactions

This section introduces the main methodology of this thesis that is adversarial risk analysis(ARA). The aim is to provide the adequate theoretical background on ARA through the mathematical formulations and algorithmic implementation along with a discussion on motivation related to the choice of methodology. The limitations of common security games using game theory are also explored while providing an extensive literature review on the field of ARA and it's different used cases across industries and literature while also highlighting it's use as an analysis tool for addressing the information structures in adversarial interactions.

4.1 Introduction to Adversarial Risk Analysis

Many decision-making situations are inherently strategic as the consequences of one's own decisions depend on how other actors respond to them. Such situations arise in domains including cybersecurity, defence, crime prevention, business, and international politics. In these settings, decision-makers must form beliefs not only about uncertainty in the environment but also about the objectives, information, and potential actions of other actors.

Adversarial Risk Analysis (ARA) provides a systematic framework for decision-making in such settings. The central idea of ARA is to treat the adversary's behaviour as a source of uncertainty. Rather than assuming that the adversary's strategy is known, the decision-maker constructs a Bayesian subjective probability distribution over the adversary's possible decisions, payoffs , and utilities, and selects their own course of action based on this uncertainty assumption. Generally, ARA belongs to the broader field of strategic decision-making. It combines decision-theoretic modelling with probabilistic reasoning to analyse situations in which multiple rational and goal-oriented actors influence one another's decisions. Unlike classical game theory, ARA models the adversary's decision-making as the decision-maker's subjective uncertainty rather than assuming a shared equilibrium solution, which is generally intractable under realistic assumptions of incomplete information. ARA also decomposes the uncertainties of the adversary's behaviour in to distinct levels, disambiguating the decision-making process and making explicit the level of information available to the analyst.

4.2 Notation and visualization of ARA

To introduce the paradigm of ARA, it's origins within Bayesian decision theory are to be highlighted as it's inceptive force. As discussed above, given uncertainty over attacker incentives and beliefs, the Bayesian approach is to assign a subjective probability distribution reflecting the current state of information[75]. Representing this kind of belief structure warrants a formalism native to Bayesian reasoning.

Influence diagrams, a variant of Bayesian networks, and a tenet of Bayesian decision theory[76], serve this purpose directly as they encode the probabilistic dependencies between decisions and outcomes, and the algorithmic procedure for determining the expected utility of a scenario. Influence diagrams are directed acyclic graphs that are constructed from three types of nodes with arcs connecting them: decision(rectangular), outcome(oval), and utility nodes(hexagonal)[77, 78, 79]. Directed arcs between the nodes are deployed to reflect internal dependencies within the decision-making process as filled lines reflect probabilistic dependence while dashed lines guides the availability of information at decision nodes. Figure 12 illustrates a simple influence diagram structure.

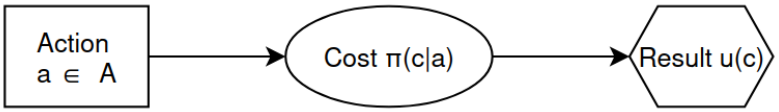

Figure 12: Simple influence diagram.

In this setting, the decision-maker chooses an action $a \in \mathcal{A}$, represented by the rectangle. This decision results in an uncertain cost c, modelled through the density $\pi(c|a)$, represented by the oval node. At the hexagonal node, the utility function $u(c)$ evaluates the cost c drawn from $\pi(c|a)$, yielding the expected utility of the interaction.

However, as the cost c is random rather than deterministic, and $\pi(c|a)$ may be a continuous density, the expected utility resulting from action a is defined as the probability-weighted average of utility over all possible cost outcomes. A utility-maximizing player then solves the following optimization problem to determine the action $a^\star \in \mathcal{A}$ that maximizes expected utility:

$$\psi(a) = \int u(c)\, \pi(c|a)\, dc,$$
$$a^\star = \arg\max_{a \in \mathcal{A}} \ \psi(a).$$

The utility function $u(c)$ determines how the decision-maker values costs associated with action a modelling the risk behaviour of the decision-maker[80]. A concave utility function indicates risk-aversion as a increase in profits is not valued as highly as a decrease in risk, whereas a convex utility function corresponds to a risk-seeking decision-maker as an increase in gain is valued higher than the prospect of a loss[81].

As ARA is fundamentally about the analysis of adversarial interactions consisting of two intervening agents, this leads to the development of bi-agent influence diagrams (BAID)[82]. As ARA assigns the role of analyst to one side of the influence diagram, the convention is to call this player the defender, whose opponent is often called the attacker. It is convention as well to highlight the decisions made by the defender

in white, the attacker in black and shared outcome nodes concerning both players in grey. The BAID in Figure 13 is a reflection of a simultaneous Defend-Attack interaction, in which both decision-makers choose their actions independently of each other at the same time, leading to a shared outcome determining the end result and thus utility of both players. An example of such an interaction is a first-price, sealed-bid auction as shown by Rios Insua et al.[7].

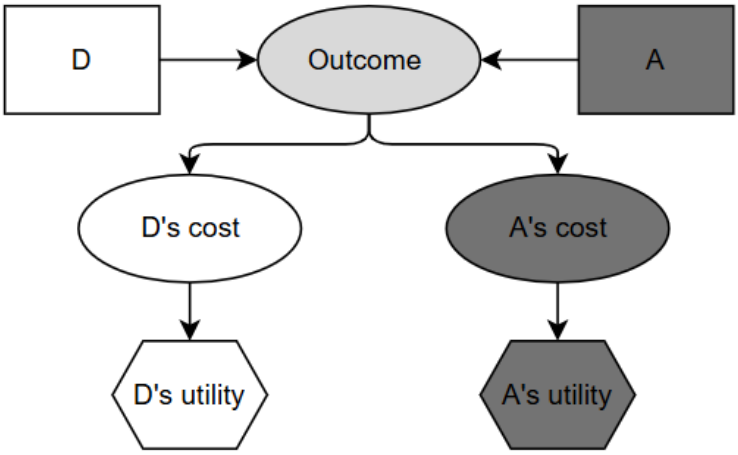

Figure 13: BAID, simultaneous Defend-Attack interaction.

In the first-price, sealed-bid auction, both players submit a single bid without observing the other's, and the highest bidder wins the item at the price of their own bid. This creates a trade-off for the defender as bidding higher increases the probability of winning against the attacker, but reduces the profit margin relative to their evaluation v_d of the item. The defender therefore aims to choose a bid d that balances this trade-off by maximizing her expected utility,

$$\max_{d \in \mathcal{D}} \; u_d(v_D - d)\, \mathbb{P}_d(d > A \mid d),$$

where v_D is the defender's valuation of the item and $\mathbb{P}_d(d > A \mid d)$ is the subjective probability assessment of outbidding the attacker given bid d.

Beyond simultaneous interactions, sequential interactions showcase a special relevance with respect to security applications. Figure 14 shows such a sequential Defend-Attack interaction where the defender moves first, and the attacker observes this move before responding. The dotted line showcases this dependency in which the attacker observes the defender's problem and responds accordingly.

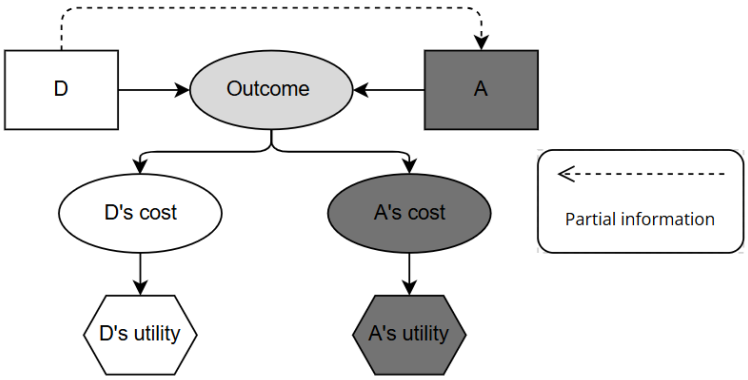

Figure 14: BAID of sequential Defend-Attack interaction.

The defender commits to a defence $d \in \mathcal{D}$, which informs the attacker's decision $a \in \mathcal{A}$, i.e. the attacker's decision is conditioned on the defender's decision $a \mid d$.

Fundamentally, irrespective of whether the interaction is simultaneous or sequential, the decisions made by the opponents jointly determine an outcome, visualized as the grey uncertainty node in Figures 13 and 14. The costs incurred by the opponents reflect the consequences, losses, or gains arising from this outcome, and may be monetary, material, or reputational depending on the particularities of the interaction as well as containing the costs incurred from decisions made prior to the outcome node. Critically, however, the goals of the opponents often differ and can be direct opposites.

4.3 Modelling and solution approaches

4.3.1 ARA as an influence diagram

Given the BAID in Figures 13 and 14, prior to determining this structure, the defender must analyse the key assets at risk exposed to an adversarial threat. In the context of this thesis, these assets are referred to as business critical assets that a potential attacker seeks to damage or use as leverage. Once the asset is identified, the critical attack scenario or method must be identified, which largely determines the cost structures and incentives of the attacker, and thereby those of the defender.

Once these have been identified and the initial BAID is constructed, the convention is to divide the influence diagram into two: one representing the interaction from the defender's point of view, and one representing the defender's model of the attacker's problem. This is done to reorganize the problem structure around a single analyst, the defender. Within the BAID, from the point of view of the players, the opponent's decision node, along with their probabilistic costs and utilities, is collapsed into a single uncertainty node depicting the uncertainty of how the opponent is going to act. Figure 15 contains the visualizations for these IDs from the point of view of the defender, and the defender's view of the attacker's model.

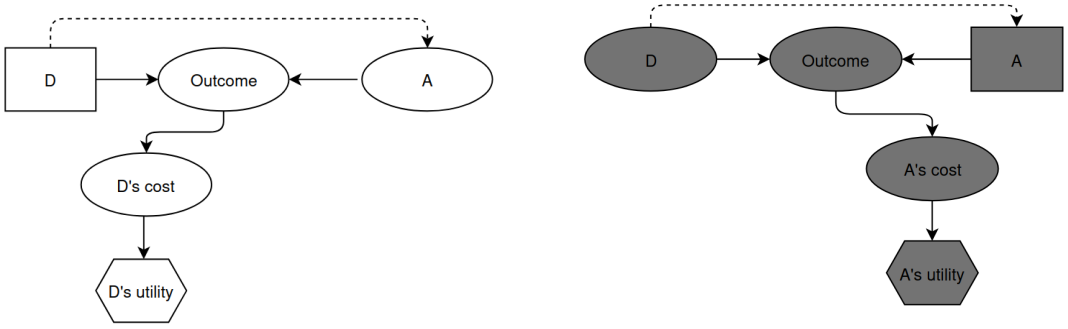

Figure 15: The ID's from the point of view of the defender and attacker[7].

This visualization functions as a basis upon which the analysis in ARA is based upon, reflecting that the ID of the attacker problem exists purely as an instrument to inform the defender's problem by constructing the attacker problem, which formulates the probabilistic assessment for distribution over attacks. The interactions in question can also be modelled using decision trees, but these suffer scalability issues especially for increasingly complicated interactions with several participants or a substantial decision space.

4.3.2 Populating the nodes of the influence diagram

In an interaction of two opponents, the structure of the problem is inherently such that first the analyst observes from the point of the defender, the information landscape and makes judgements based on this determining the attacker's likely action and then optimizing against this. The mathematical formulation of this structure is highlighted below with first assessing the subjective probability assessment of $\hat{p}_D(a|d))$ that constitutes the attacker's problem right of Figure 15.

$$\psi_A(a,d) = \int u_A(c)\pi_A(c|a,d)dc$$
$$\Rightarrow \hat{p}_D(a|d) = \mathbb{P}[a = \arg\max_{a\in\mathcal{A}} \psi_A(a,d)]$$

This subjective assessment of $\hat{p}_D(a|d)$ is used in the utility-maximizing decision of the defender as the attacker's influence diagram is now collapsed into a single probabilistic node, as seen in the left of Figure 15.

$$\psi_D(a,d) = \int u_D(c)\pi_D(c|a,d)dc$$
$$d^\star = \arg\max_{d\in\mathcal{D}} \int \hat{p}_D(a|d)\psi_D(a,d)da$$

Given the problem structure and the influence diagram of the opponent, the nodes have to be populated with uncertainty. For the defender branch of a BAID,

this reflects a standard task in information elicitation, as the defender can evaluate within reasonable bounds their own expected utility function $u_D(c)$ as well as their probability judgment over cost $\pi_D(c|a,d)$ conditioned on the actions of both players, which is denoted as $(u_D, \pi_D) \sim G$ reflecting the defender's non-strategic beliefs and aleatoric uncertainty.

For the attacker branch of the BAID, as the goal is the non-trivial task of computing $\hat{p}_D(a|d)$, i.e. the conditional probability of an attack given defences, how one populates the nodes preceding the expected utility to be maximized has a direct effect on the the computed subjective probability assessment. This part of the problem structure is motivated by the propagation of two distinct forms of uncertainty. Epistemic uncertainty, representing the defender's uncertainty over attacker type and denoted $(U_A, P_A) \sim F$ and concept uncertainty, representing the defender's uncertainty over the solution concept, or rationality, deployed by the attacker[31]. In Bayesian terms, $(U_A, P_A) \sim F$ constitutes a prior belief over attacker type, incentives and risk appetite.

4.3.3 Computational solutions

As discussed above, expected utility for either player is inherently a probability-weighted sum of the utilities of the possible outcomes. In interactions with large influence diagrams and deep structures, the expected utility is a function of several sources of cost preceding it. To account for this, the expected utility formulation integrating over the costs takes into account all sources of cost. This results in a multidimensional integral over the different cost outcomes, the integrand of which is generally intractable in closed form. This is mitigated via Monte Carlo (MC) simulation: at each iteration, a random sample is drawn from the distribution $(U_A, P_A) \sim F$, and the expected utility is approximated by averaging over these draws. This allows the epistemic uncertainty encoded in F to propagate through the model as the solution converges toward an approximation[7]. The general algorithmic implementation of estimating the distribution over attacks is seen in Algorithm 1, yielding $\hat{p}_D(a|d)$.

Algorithm 1 MC Approximation of Distribution over Attacks[7]

Place distribution F over (U_A, P_A) to express uncertainty about attacker type
for each defence $d \in \mathcal{D}$ **do**
  for $i = 1, \ldots, K$ **do**
    Draw $(U_A^i, P_A^i) \sim F$
    Approximate attacker expected utility for each $a \in \mathcal{A}$:

$$\psi_A^i(a,d) = \int U_A^i(c)\, P_A^i(c \mid a,d)\, dc \;\approx\; \frac{1}{M} \sum_{m=1}^{M} U_A^i(c^m), \quad c^m \sim P_A^i(\cdot \mid a,d)$$

    Record optimal attacker action: $a^{*i}(d) = \arg\max_{a \in A}\ \psi_A^i(a,d)$
  end for
  Return attacker decision distribution:

$$\hat{p}_D(a \mid d) \;\approx\; \frac{\#\{a^{*i}(d) = a\}}{K}$$

end for

Once $\hat{p}_D(a|d)$ has been obtained, it is combined with the defender's own information to compute the defence that maximizes expected utility under uncertainty. Even given perfect knowledge of the attacker's behaviour in this probabilistic form, persistent real-world stochasticity means the optimal defence must still be selected under aleatoric uncertainty of outcomes as seen in Algorithm 2.

Algorithm 2 MC Approximation of Optimal defence[7]

for each defence $d \in D$ **do**
  for each attacker action $a \in A$ **do**
    for $m = 1, \ldots, M$ **do**
      Draw consequence: $c_D^m \sim p_D(\cdot \mid d,a)$
    end for
    Approximate defender expected utility given (d,a):

$$\psi_D(a,d) = \int u_D(c)\, p_D(c \mid d,a)\, dc \;\approx\; \frac{1}{M} \sum_{m=1}^{M} u_D(c_D^m)$$

  end for
  Combine over attacker actions using $\hat{p}_D(a \mid d)$ ▷ from Algorithm 1

$$\hat{\psi}(d) \;=\; \int \psi_D(a,d)\hat{p}_D(a \mid d)da$$

end for
Return optimal defence:

$$d^* = \arg\max_{d \in D}\ \hat{\psi}(d)$$

To conclude, Banks et al.[6] describe the ARA process as comprising of three main stages: modelling system threats, simulating attacks, and adopting defensive strategies. The implementations of ARA in this thesis follow this chronology.

4.3.4 Multi-criteria Comparison of Defence Alternatives

Having obtained the optimal defence with respect to expected utility, a multi-criteria decision analysis (MCDA) problem arises once the attributes related to each defence option are additionally accounted for. Especially in problems with large decision spaces with multiple attributes associated with such decisions, maximizing expected utility alone, if other attributes are not included in the utility computation, can lead to solutions that perform poorly on other attributes relevant to the alternatives, which incentivises evaluating alternatives with respect to multiple attributes simultaneously. This motivates the use of Pareto-optimality, defined as:

$$P = \{x \in \mathcal{X} : \nexists\, x' \in \mathcal{X},\ F(x') \succ F(x)\}\,.$$

Fundamentally, Pareto-optimality is used to identify the non-dominated (Pareto-optimal) solutions from a set of possible alternatives in problems of multi-objective optimization and MCDA[83]. For a set of alternatives denoted $x \in \mathcal{X}$, each evaluated with respect to a vector of attributes $F(x) = (F_1(x), F_2(x))$, an alternative x is said to be dominated by x', denoted $F(x') \succ F(x)$, if x' is at least as good as x on every attribute and strictly better on at least one. A solution is Pareto-optimal, or non-dominated, if no such x' exists meaning no other alternative can improve one attribute without worsening another. The set of all Pareto-optimal solutions is referred to as the Pareto-optimal set, and its image in attribute space is the Pareto frontier: the boundary formed by the best attainable trade-offs between attributes, beyond which no alternative can simultaneously improve upon both. Alternatives lying strictly inside this boundary are dominated by at least one point on the frontier, and hence are referred to as the dominated solutions. Figure 16 shows an example frontier where $F_2(x)$ is to be minimized and $F_1(x)$ is to be maximized.

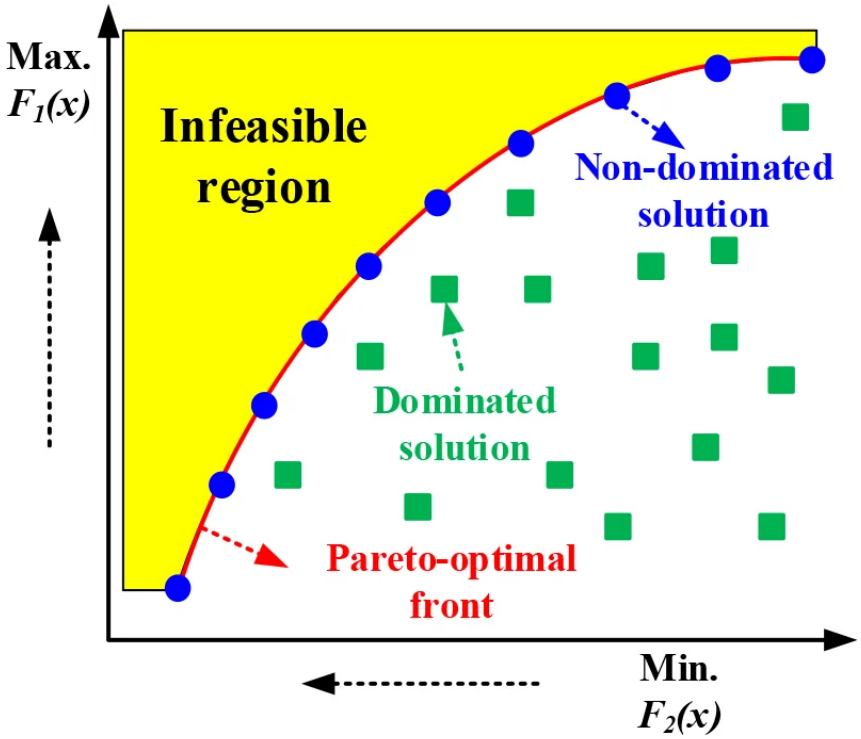

Figure 16: Max-Min Pareto frontier[84].

The attributes relevant to the alternatives in MCDA problems are context-dependent, and could include increased exposure, monetary cost of alternatives, increased complexity, the need for ad-hoc upkeep, level of availability, etc. For example, Roponen et al.[9] used Pareto-type dominance relations to disregard a large subset of dominated solutions by identifying the Pareto-optimal solutions to choose from, where the preferential statements guiding Pareto-optimality consisted of maximizing damage while minimizing exposure.

4.4 ARA, Game Theory and Risk Analysis

4.4.1 Justification for a Bayesian Approach to Interactions

ARA as a methodology arose from the field of Bayesian decision analysis that focuses upon how interactions can be modelled with increased realism under uncertainty[75]. The Bayesian approach is intrinsic to ARA in distinct ways. Within a single interaction, it governs how the analyst constructs the prior $(U_A, P_A) \sim F$ over the attacker's unknown type from available information forming an initial subjective probability estimate. This prior is not static as through repeated interactions and the gathering of information on attack behaviour, it is updated via Bayesian revision, allowing epistemic uncertainty to resolve progressively. In terms of concept uncertainty, i.e. the rationality exhibited by the attacker, rather than committing to a single solution concept, one can formulate a Bayesian mixture model over several candidate rationalities, weighting each by a subjective probability of its validity[31]. Through repeated interaction, these weights can as well be updated to

refine the rationality that governs the attacker's decision-making over time. Thus, the Bayesian aspect intrinsic to adversarial interactions comprises both making subjective probability assessments [85] and updating the inherent uncertainty of these assessments through Bayesian updating of the defender's beliefs about the opponent's unknowns [6].

4.4.2 ARA and Game Theory

As for game-theory, which is a study of strategic interactions[81, 85], the limitations when applied to general security games are worthy of mention in support of our argument for ARA. As mentioned above, the interactions are modelled objectively such that the utilities, payoffs and strategies of the players are common knowledge(CK). Games are solved by computing the Nash equilibrium (NE), which refers to the strategy the game converges to, where no player can improve their respective utility by a unilateral change of strategy[86]. Using the expected utility notation of 4.1, NE for pair $(d^\star, a^\star)$ is defined as[6]:

$$d^\star \in \arg\max_{d \in \mathcal{D}} \psi_D(d, a^\star),$$
$$a^\star \in \arg\max_{a \in \mathcal{A}} \psi_A(d^\star, a).$$

In real-life strategic interactions and security games, the assumptions required for the solution concept of the Nash equilibrium are often unrealistic and the solutions given can cause suboptimal choices in strategy if implemented directly. This sentiment is supported by Banks et al.[7] in their landmark paper on ARA where they determined in terms of classical game theory (in zero-sum games) with complete information that each player is in essence choosing the strategy of the worst-case scenario while also generally offering non-decomposable solutions in terms of decentralized decision-making as two-player game Nash equilibria rarely produce global optimal solutions for a collective of decision-makers. Bayesian games address the problem of incomplete information through the common prior assumption, wherein players are assumed to share identical beliefs over the distribution of opponent types [85]. The corresponding solution concept is the Bayesian Nash Equilibrium (BNE). Antos and Pfeffer [87] argue, however, that this assumption remains significantly removed from reality, as real-world adversarial interactions are characterised by fundamentally asymmetric and heterogeneous beliefs that a common prior cannot adequately capture. In addition to the Bayesian updating discussed in 4.4.1, BNE requires that belief updates be mutually consistent across players[88], thus increasing the conceptual complexity of updating belief when compared to the single-analyst point of view of ARA. Table 1 showcases a summary of the differences between game theory and ARA adapted from Gil and Arnua[89] and Banks et al. [7]:

Table 1: Comparison of Game Theory and ARA.

Requirement	**Standard Game Theory**	**ARA**
Opponents aim to maximise their utility	✓	✓
Uncertainty about the attacker's actions	×	✓
Incomplete information about the evaluation of objectives between opponents	×	✓
Simultaneous and sequential decisions	✓	✓
Applicable under sparse historical data	×	✓
Players have infinite logic and unbounded computation	✓	×
Information exposure evaluated	×	✓

4.4.3 ARA and Risk Analysis

As the game theoretic studies discussed above suffer limitations related to assumptions of common knowledge leading to worst-case scenario based equilibrium solutions, some notable challenges have also been identified in risk analysis as it pertains to risk being realized from an adversarial counterparty. Notably, central tools in qualitative risk analysis(key within the context of this thesis) suffer from simplicity and ambiguity. Cox[18] criticized the usage of risk matrices, noting their tendency to assign higher qualitative ratings to quantitatively smaller risks as well as offering poor resource allocation advice while in essence being dependent on subjective interpretations of risk severity and likelihood. These sentiments are shared by Simpson [19] in the context of cybersecurity who determined the role of uncertainty and unawareness to key in developing improved approaches to managing risks and understanding the landscape of risk.

In terms of probabilistic risk analysis(PRA) as a tool within quantitative risk analysis, the key simplification is the collapse of uncertainty over attacker behaviour to a single probabilistic assessment irrespective of defences of the defender or the capability or risk-appetite of the attacker. Such an analysis does not take into account the dynamic nature of the attacker, but rather implies a static evaluation, which is often referred to as Nature or a neutral opponent[7]. This means that fundamentally as choices from the side of the attacker are assumed to be made exogenously of defender choices and the observable signals on said choices, the attacker therefore neither has intelligence in the sense that they do not seek to maximize utility nor are they sentient in any other sense than a constant probabilistic assessment of some act being attempted. These claims are supported by Merrick and Parnell[90], who determined that traditional methods of probabilistic risk analysis can systematically underestimate risk to the defender and fail to show how risk shifts from the acts of the defender. Table 2 contains the central observations between ARA and traditional risk analysis inspired by Merrick and Parnell[90].

Table 2: Comparison of Risk Analysis and ARA.

Requirement	**Traditional Risk Analysis**	**ARA**
Explicit attacker decision model	×	✓
Defender-centric analysis	✓	✓
Models strategic adaptation	×	✓
Attack probability from historical data	✓	×
Attacker rationality explicitly stated	×	✓

4.5 Applications of ARA

Here we introduce the most notable papers in the context of adversarial risk analysis and introduce their computational solutions and structures of interaction. We choose to focus on two paper's of special relevance to this thesis as well as offering an overview on the applications and identifying the variety of contexts in which ARA is applied in.

In the context of this thesis, the most important and later point of reference is the study done by Rios Insua et al.[8], where the use of ARA was researched in the context of cybersecurity under unintentional and adversarial threats. The general problem scenario is concerned with a SME that comes into contact with a single attacker while also experiencing natural or unintentional threats. The SME, i.e. the defender has a risk mitigation portfolio that includes controls $s \in \mathcal{S}$ such as firewall protection, cloud-based Distributed-Denial-of-Service(DDoS) protection capacity and an anti-fire system, while $i \in \mathcal{I}$ is a binary choice in choosing a cyber insurance plan to cover the risks of the cyber threats. The Defend-Attack structure is inherently a Stackelberg game of a leader and a follower[91], where the defender acts first which the attacker reacts to as discussed earlier in chapter 4.2. This interaction is visualized in Figure 17:

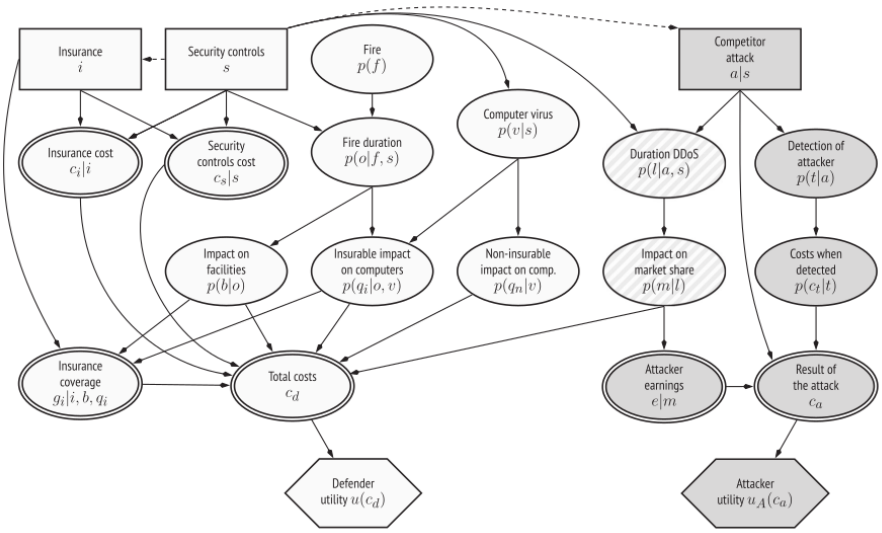

Figure 17: BAID of DDoS attack scenario[8].

In the context of counterterrorism, ARA was implemented by Rios et al.[10] and refined by Sevillano et al.[92]. These papers introduce the Defender-Attacker-Defender game, which showcases the sequential nature of decision-making in strategic interaction. In the context of cybersecurity, the sequential defender-attacker-defender game is the most applicable for attacks such as ransomware and Man-in-the-middle, where the attack consists of many phases for the attacker and different protective and reactive defences for the defender. Sevillano et al.[92] focus on a case of mitigating the risks of interacting with Somalian pirates while crossing the Suez channel, where the defender has a set of protective measures prior to the trip, and given an successful attack by the pirates, a set of reactive measures. This can be visualized by an influence diagram in Figure 18 and it's corresponding decision tree in Figure 19:

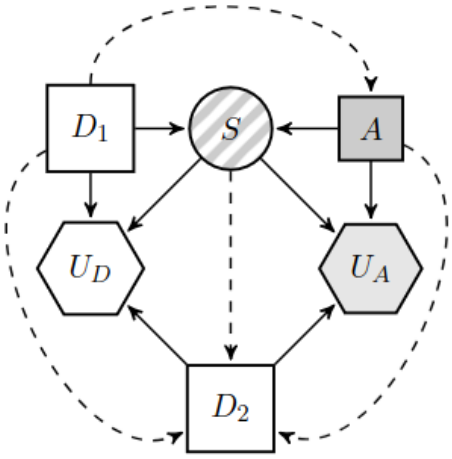

Figure 18: Influence diagram[6].

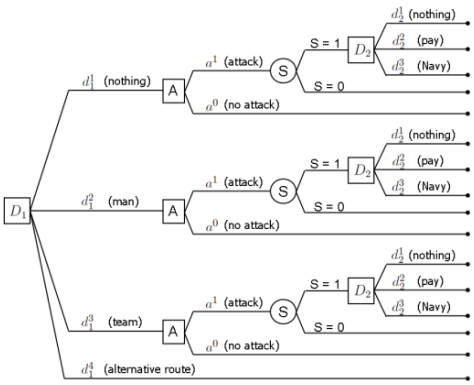

Figure 19: Decision tree[92].

Roponen et al.[9] researched ARA in the context of military planning with the goal of protection from unmanned aeriel vehicles while also introducing the paradigm of partial information to ARA. The focus is on a attacker-defender wherein the players lack information on each other with regards to the opponents utilities and probabilities conventionally elicited through $F \sim (U_A, P_A)$. The model in this game solves the optimal defence portfolio by applying first-degree stochastic dominance for the expected utilities given any set of utilities or conditional probabilities. This paradigm allows for the propagation of uncertainty in the ARA model when only general statements about preferences are known.

ARA has also been researched in the context of counterterrorist online surveillance[89], parole-board decision-making[11], multi-threat multi-site protection[93], urban security allocation problem[94], personalized pricing decisions in product launches[95, 96], insider threat modelling [97], critical infrastructure protection [98] and disinformation spread[99]. Table 3 consists of a summary of the relevant literature related to ARA.

Table 3: Summary of applications in literature.

Study	**Context**	**Results**
[8]	Cybersecurity	Optimal portfolio of controls
[10]	Counterterrorism	Optimal defences per stage
[9]	Military air surveillance	Optimal portfolio of anti-UAV drones
[89]	Online surveillance of nefarious actors	Threat detection
[11]	Parole-board decision-making	Optimal decision under uncertainty
[93]	Public transport fare evasion	Optimal allocation of defensive resources
[94]	Urban protection	Allocation of security across nodes
[95, 96]	Pricing decision	Optimal price under product launch
[97]	Cyber threat detection	defence optimization
[98]	Critical infrastructure protection	Optimal protection scheme
[99]	Disinformation	Combat disinformation campaign

5 Adversarial Risk Analysis Applied in Case Studies

In this chapter we apply ARA in the context of two case studies and example SMEs for which we will evaluate two industry-specific cyberattack scenarios, which also have variance in terms of interaction type. The goal is to construct the problem using postulates of ARA, produce a set of optimal defences and discuss how the information structure of the defender and the attacker affects the end result as well as how uncertainty is propagated in the models. The final section will also contain sensitivity analysis which focuses on examining the robustness of the models built.

5.1 Water Utility Case Study

5.1.1 Problem Structure

The first case concerns a medium-sized water utility plant, which governs the cities water treatment and distribution. The factory-based environment is mostly controlled using industrial control systems(ICS) or operational technology(OT) tools can facilitate control over the physical aspects of the plant and is mostly dependent on functioning control systems[100]. As a critical infrastructure provider, the adversaries in this case are assumed to be state-sponsored whose incentives are the destruction of infrastructure and the degradation of trust in public services in the city targeted. Therefore, we model the cyber threat concerning the water utility plant as a DDoS attack that aims to overwrite the OT systems of the plant causing damages, critically affecting the availability of their systems. The defender organization has to implement a protection scheme and a cyber insurance policy against the potential attacker aiming to optimize a choice in defence portfolio. The structure of this problem is inspired by Rios Insua et al.[8].

A DDoS attack is a cyberattack in which the victim's online services or remote control systems are overrun by a drastic influx of traffic caused by a nefarious actor. The goal is to overrun the victims server's capacity such that legitimate users cannot access the system or users or the victim companies OT-systems are blocked[101]. This can have serious repercussions on the target organization as a disrupted e-commerce site can cause a loss in the millions for a large retailer or a malfunctioning set of OT-systems can cause significant damages or cascading failures in critical infrastructures. In the modern cyberspace, DDoS attacks are often used as distractions while other attack vectors are played out in the system with the goal of overwhelming defences[102]. The severity of an attack is often characterised by targeted volume, packet rate, complexity and duration. In the context of this thesis, the attack severity will be measured using duration and volume where as the defender has a choice of implementing a threshold of traffic to be routed through a scrubbing center, removing non-legitimate users from the influx of user requests. The critical assets of the defender are therefore their OT-technology.

The choice of defence portfolio has to be done by the defender with respect to their evaluations of attacker behaviour, incentives and level of risk preference. In game theoretic terms, this game is a sequential interaction, where the defender aims to choose their optimal set of defences $(s^\star, i^\star) \in \mathcal{S}, \mathcal{I}$ that denote the threshold of scrubbing center volume and a choice of insurance policy. The attacker has the incentive of causing as much downtime as possible, and they aim to choose the optimal amount of attacks $a \in \mathcal{A}$ to implement maximum damage. The structure of the game is visualized by the BAID in Figure 20:

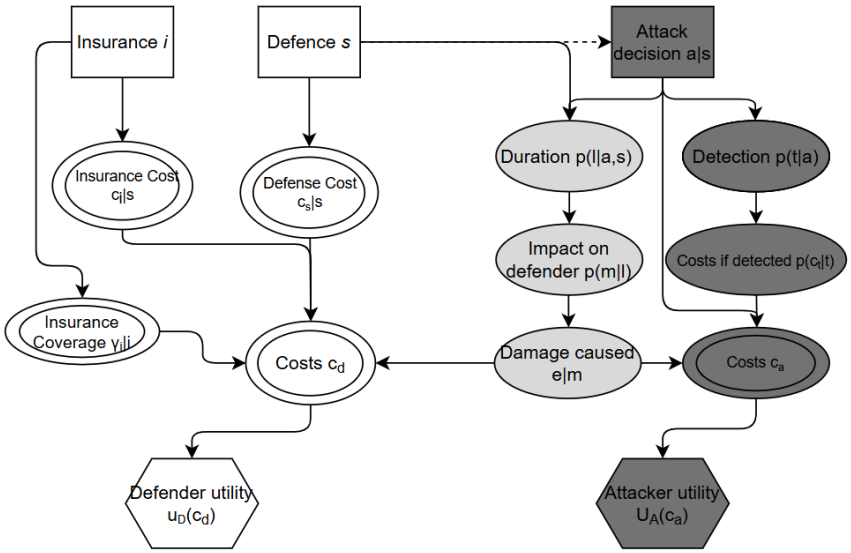

Figure 20: Influence diagram of the Defend-Attack scenario.

The dotted lines denote the defender's problem structure informing that of the attacker's as well as partial information. The defender first commits to a strategy, after which the attacker decides on the number of attacks conditioned on the perceived defence $a|s$. The cost total of the attacker is the deterministic node c_a, which is comprised of the operating cost per attack, cost of possible being caught and convicted as well as the damage caused to the defender that is perceived as gain. The defender has three deterministic cost sources which are the costs of the protection level and insurance policy, as well as the insurance coverage that lowers the total cost if applied with a singular probabilistic node of damage being uncertain for the defender. The following sections will introduce the influence diagrams of the opponents and populate the nodes of said influence diagrams.

5.1.2 Defender's Beliefs and Preferences

Herein we describe the beliefs and structures of the defender's decision-making that do not require any strategic evaluation of opponents. Figure 21 highlights the influence diagram of the defender.

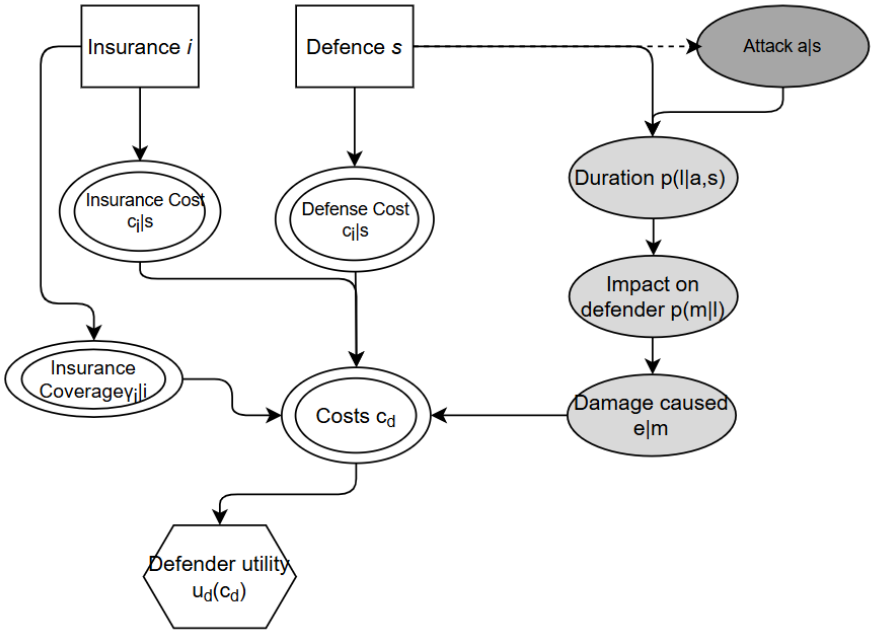

Figure 21: Influence diagram for the Defender problem.

To accurately evaluate the defender's uncertainty about a DDoS attack, the aggregation of data related to the average attack is required. According to the Digicert annual analysis of DDoS attacks for 2025[103], the average attack severity is 2.69 Gbps lasting approximately 29 minutes. This average attack severity is strongly skewed by the emergence of mega-volumetric attacks surpassing 1000 Gbps. In the Alerion DDoS threat landscape report 2025[104], the reported average DDoS attacks had a severity of 23 Gbps with an average duration of below 15 minutes. To accurately model the behaviour of the attacker in a DDoS attack, we will use the template given by Rios Insua et al.[8] and improve upon it to reflect the modern state of DDoS attacks.

Fundamentally, we model the decision-making of the attacker as a process trying to inflict maximum damage. The defender in this case is a small-sized critical infrastructure provider whose OT-systems are the main assets that are under attack. Given a destructive DDoS attack, their OT systems are damaged beyond repair requiring replacement at large cost. We assume that a sustained or repeated DDoS attack denies the defender's ability to safely and correctly operate their physical systems in their water utility plant, causing excessive damages[105]. As the defender lacks knowledge about the attacker's behaviour , we create realistic priors based upon data to sample from for the defender's damage model reflecting the non-strategic beliefs of the interaction.

The defender's belief about attack strength is modelled as being Gamma distributed as $s_{gbps} \sim \Gamma(9, 0.225)$ reflecting a aggregation of the current state of DDoS crime as well information privy to the SME on possible past incident data whereas individual attack duration is modelled to be distributed as $l_j \sim \Gamma(8, 0.052)$ approxi-

mating a mean of 25 minutes consistent with industry beliefs. An attack is successful if the sampled strength is larger than current defence, i.e. $s_{gbps} > s$ in which case the downtime of the defender is increased by $l_j \sim \Gamma(8, 0.052)$. Since the defender cannot be fully certain of the costs per hour of downtime, this is sampled from a distribution reflecting their private information on replacement costs and incident recovery. The defender's loss rate is modelled as $r \sim \mathcal{N}(40,000,\ 8{,}000)$

The total downtime caused by successful attacks is $l = \sum_j l_j$, and the gross impact is capped at $m \sim \min(R_{max},\ r \cdot l)$, where R_{max} denotes the maximum damage that can be caused, a value shared by the opponents in this interaction. The net impact to the defender is $m_{net} = m \cdot (1 - \gamma_i)$, where γ_i is the coverage fraction of the chosen insurance policy i.

The utility of the defender is therefore caused by the cumulative impact of the attacks negated by a possible choice in insurance policy as well as the costs of the security portfolios chosen. For DDoS protection threshold $s \in \mathcal{S}$ the defender has a choice from $s \in \{0, 2, 5, 10, 1000\}$ gbps and insurance $i \in \{\text{NoIns, cyber, comprehensive}\}$. Each defence has its own associated cost c_s, c_i and insurance policies define a parameter γ which determines the proportion of perceived damage covered. The utility function of the defender is as convention defines risk-averse. The utility function is of the form $u_d(c_d) = a - b \cdot \exp(c_d)$, calibrated such that the worst loss corresponds to a utility of zero while a successfully deterred attack results in a utility of one. This results in a utility function inspired by Rios Insua et al.[8]:

$$u_d(c_d) \;=\; \frac{1}{e^c - 1}\left[\exp\!\left(c\left(1 - \frac{c_d}{R_{\max}}\right)\right) - 1\right]$$

where c indicates the coefficient of risk aversion that will be used later in sensitivity analysis[92].

The total costs for the defender are denoted as $c_d \mid m,\ c_s,\ c_i,\ \gamma_i$, being conditional on the impact, cost of defence portfolio and insurance policy covering a proportion of m. The Tables 4 and 5 showcase the costs of the defences, which follow non-linear pricing conventions as well as the structure used by Rios Insua et al.[8].

Table 4: DDoS protection tier costs.

$s(gbps)$	0	2	5	10	1000
c_s (€)	0	2,400	3,600	4,800	12,000

Table 5: Insurance premium and coverage.

i	NoIns	Cyber	Comp
c_i (€)	0	200	600
γ	0	0.8	0.9

5.1.3 Attacker Strategic Behaviour and Cost Structure

We inscribe the defender's uncertainty of the attacker type through F, the main goal of which is to inflict maximum damage over the defender organization while minimizing their costs. The attacker's strategic decision to be made is the choice of attacking and the number of attacks capped at thirty maximum attacks $a \in \{0, 1, 2, \ldots, 30\}$. The succeeding probabilistic nodes resulting from the attacker's decision inflict their respective costs and formulate the utility of the attacker, the structure of which is shown in Figure 22.

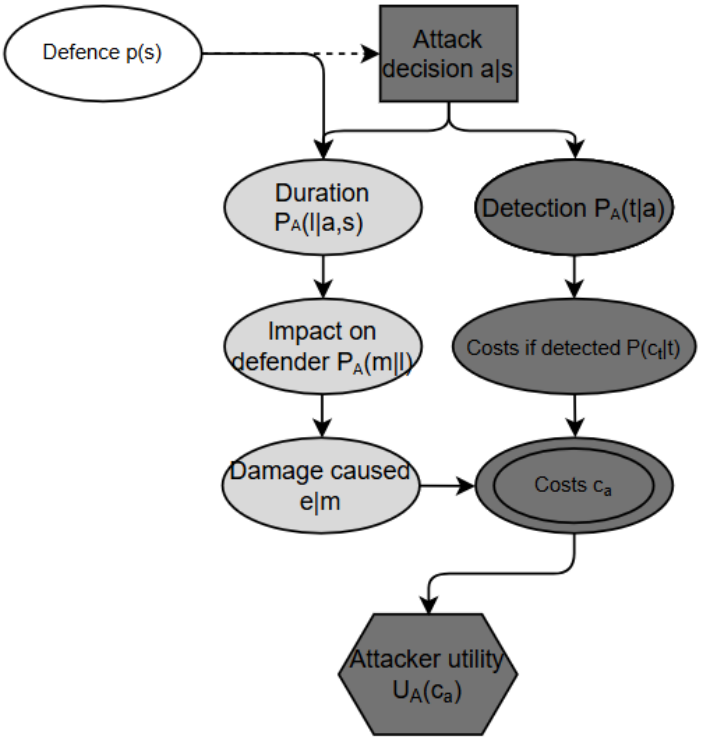

Figure 22: Influence diagram of the attacker problem.

As the DDoS attack strength has experienced increasing volatility, we choose to model the attack strength S_{gbps} as a compounding Poisson process[106] of the form $S_{gbps} \sim B_0 + \sum_{i=1}^{N} S_i,\ N \sim \text{Pois}(\lambda)$. This means that the severity of each attack is not a single sampled strength but a Gamma-distributed baseline load B and its sum with a compounding Poisson process simulating the stochastic nature of large spikes in volumetric traffic. The total attack volume is sampled as $B_0 \sim \Gamma(2,1)$, $S_i \sim \Gamma(\alpha,1),\ \alpha \sim \mathbb{U}(2.5,4)$, with a Poisson process depicting the amount of volumetric spikes $N \sim \text{Pois}(\lambda),\ \lambda = 2$. Attack success is again dependent on the defence threshold chosen, which we assume the attacker has access to through reconnaissance unlike the insurance policy which is private information to the defender. The total distribution of downtime caused by successful attacks is denoted $P_A(l \mid a, s)$.

Since the defender is uncertain of the attacker's knowledge, we assume the use of publically available data the loss rate caused by DDoS attacks and implement a distribution that matches the data given by CISA[107], denoted by r. The distribution takes the form

$$P_A(r) \ \sim\ 0.6\,\text{Lognormal}(\mu_1{=}9,\ \sigma_1^2{=}0.9)\ +\ 0.4\,\text{Lognormal}(\mu_2{=}11.2,\ \sigma_2^2{=}0.5),$$

which makes a binary assumption that all attacks are either contained or cause a cascading failure resulting in high losses. The impact caused by the attacker and

therefore incurred by the defender has the familiar definition $e = m \sim \min(R_{max},\, r \cdot l)$. Game-theoretically, this means that impact is a zero-sum variable since the perceived gain for the attacker equals the loss for the defender[108]. The constraint definition on m indicates the attacker aims to cause damage only up to a total system rebuild, indicating rationality.

The probability of getting caught and being charged for an attempted DDoS attack is increasingly low due to the sheer volume of mitigated attacks and the relatively high level of anonymity of the attackers. This is modelled as a Beta distribution centred at 0.2%,

$$P_A(t \mid a) \sim \text{Beta}(2, 998),$$

where the event occurs with probability sampled from $P_A(t \mid a)$, incurring cost c_t.The cost c_t is approximated as a €1.5 million fine with some uncertainty, yielding $P_A(c_t|t) \sim \mathcal{N}(1.5M,\, \sigma_{c_t}), \sigma_{c_t} = 0.2M$, with the maximum detection cost $c_{t,max}$ defined as the $4\sigma_{c_t}$ upper bound above the mean.

Finally, operating costs of a targeted DDoS attack are determined by the modern costs of a botnet rental, which are trending downwards now being around €7–20 per hour. We assume high-end costs and approximate the daily cost of operating a single attack to be $c_{op} =$ £480. These parameters define the attacker's utility function, in which the total cost c_a is normalized to the unit interval using its theoretical maximum and minimum values $c_{a.max}, c_{a,min}$.

$$\begin{aligned} c_a &= e - c_t - c_{op} \cdot a \\ c_{a,min} &= -c_t - c_{op} \cdot a_{max} \\ c_{a,max} &= R_{max} - a \\ U_A(c_a) &= \left(\frac{c_a - c_{a,min}}{c_{a,max} - c_{a,min}} \right)^{k_a} \\ k_a &\sim \mathcal{U}(8,\, 10). \end{aligned}$$

This corresponds to a convex utility function, the degree of which will be sampled at the simulation phase.

5.1.4 Results

As discussed in chapter 4, common ARA solutions are comprised of two phases. First we formulate the attacker problem, in which we sample the random utilities and conditional probabilities comprising their expected utility. We then find the attack decision that maximized their expected utility across the decision space, which is done via a MC simulation in which we elicit point estimates for the attacker problem for each iteration and average across the amount iterations K as shown in Algorithm 3. The random utilities and conditional probabilities form the prior assumptions and are denoted as:

$$\Big(U_A(c_a),\ P_A(c_t \mid t),\ P_A(t \mid a),\ P_A(m \mid l, r),\ P_A(l \mid a, s),\ P_A(r)\Big) \ \sim\ F$$

The defenders then evaluates the attacker to seek the attack that maximizes the expected utility by solving maximizing the expected utility:

$$a^* = \arg\max_{a \in \mathcal{A}} \int \cdots \int U_A(c_a)\, P_A(c_t \mid t)\, P_A(t \mid a)$$
$$P_A(m \mid l, r)\, P_A(l \mid a, s)\, P_A(r)\, dr\, dl\, dm\, dt\, dc_t.$$

The structure of this algorithm is seen in the appendix as Algorithm 3. The Figure 23 showcases the probability mass function for the number of attacks chosen $\hat{p}_D(a|s)$:

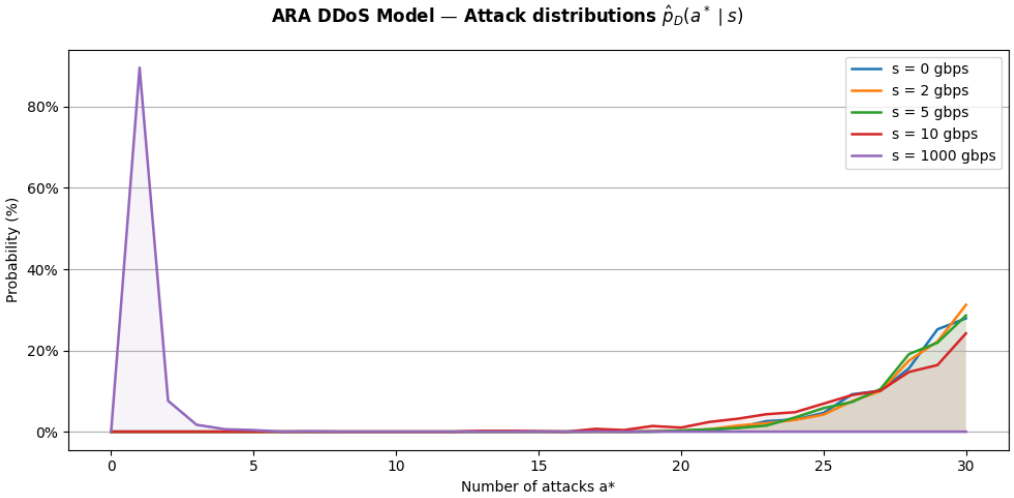

Figure 23: Attack distribution $\hat{p}_D(a^\star|s)$.

To highlight the effects of the different defences, we also plot the cumulative distribution function to visually identify dominant defences, which is shown in 24.

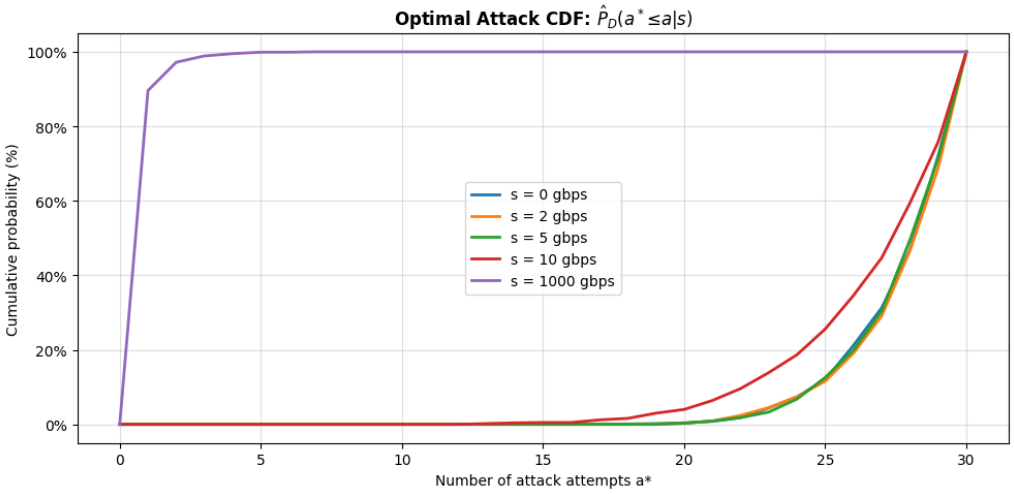

Figure 24: Cumulative distribution function$\hat{P}_D(a^\star \leq a|s)$.

Stochastic dominance can be used in determining preferred strategies from a set of CDF's[109]. Strict definition of first order stochastic dominance is $A \succeq_{FSD} B$ if $P[A \geq x] \geq P[B \geq x]$ meaning that a random variable A is at least as preferred as B. In the conventions of ARA, this is denoted as:

$$a \succeq_{FSD} a^{'} \Leftrightarrow \psi_A(a) \geq \psi_A(a^{'})$$

In the case of $\hat{P}_D(a^\star \leq a|s)$ from the point of view of the defender, there is a clear dominance relationship with respect to defence level as the attack amount is dispersed with respect to increasing defence.

Algorithm 4 thus takes the evaluation of attack probabilities and computes the optimal defences with respect to G with a risk-aversion coefficient $c = 1$. The distributions evaluated for the defender are:

$$G \sim (\hat{p}_D(a|s), p(l|a,s), p(m|l,r), p(r))$$

using which the defender optimizes their choice of $s, i \in \mathcal{S}, \mathcal{I}$ by solving :

$$(s^*,\, i^*) = \arg \max_{(s,i) \in \mathcal{S} \times \mathcal{I}} \psi(s,i),$$
$$\psi(s,i) = \iiiint u_D(c_d) p(m|l,r) p(l|a,s) \hat{p}_D(a|s) p(r) dr da dl dm$$

This leads to the expected utilities for all possible defence portfolios shown in Table 6.

Table 6: DDoS Attack; Portfolio ranking by expected utility.

rank	**defences**(s,i)	$\boldsymbol{\psi(s,i)}$
1	1000g+NoIns	0.9874
2	1000g+Cyber	0.9872
3	1000g+Comp	0.9868
⋮	⋮	⋮
13	0g+NoIns	0.5669
14	2g+NoIns	0.5657
15	5g+NoIns	0.5637

From Table 6, we identify of the optimal portfolio with respect to expected utility $(s^\star, i^\star)$ but considering the operating environment of SMEs of fundamental budget constraints, we include in the analysis the annual costs of each portfolio. This results in a MCDA problem of seeking to minimize cost while maximizing the expected utility[83], which is visualized in Figure 25 showcasing the dominated defences and the Pareto-frontier of non-dominated solutions on the upper-left boundary of the plot.

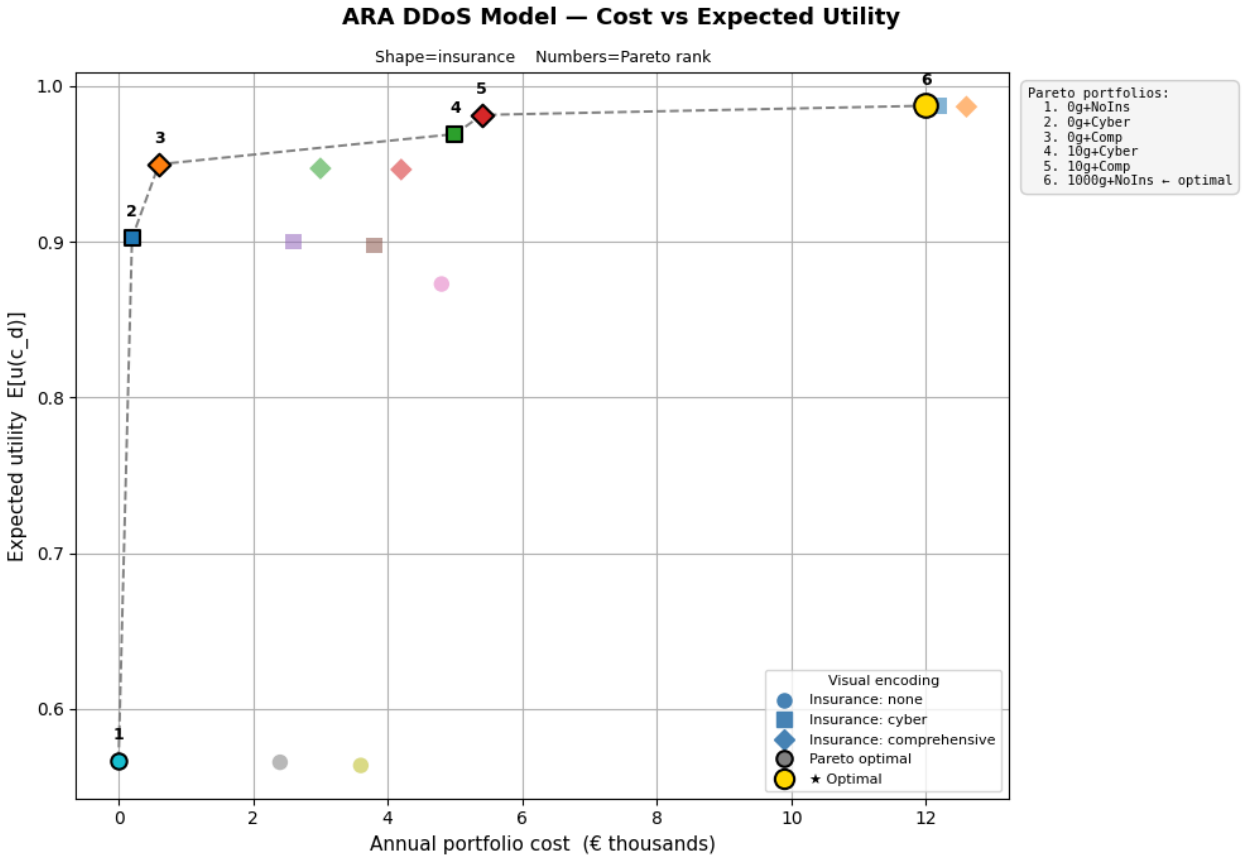

Figure 25: Pareto frontier of defence portfolios $(s,i) \in \mathcal{S}, \mathcal{I}$.

The optimal portfolio of defence in terms of expected utility(highlighted in yellow) is $(s^\star, i^\star) \rightarrow$ (1000gbps, NoIns), since the current uncertainty of attacker behaviour does not include a feasible possibility of an edge-case where the attack severity crosses the maximal defence. Therefore, the implementation of insurance is not a utility-increasing decision if the current model does not entertain the possibility of such an high-impact, low-probability attack occurring, and hence implementing insurance offers no additional gain.

5.2 Nursing Home Case Study

5.2.1 Problem Structure

Nursing homes and healthcare suffer from having access to patient-sensitive data while lacking in the amount of professionals focusing on cybersecurity. This nature of healthcare motivated our choice in focusing on a common strand of cyberattacks, ransomware attacks. In such attacks, the attacker aims to infiltrate the company's IT systems and steal their files demanding a ransom for the return of the files, critically compromising the tenet of confidentiality in information security. When compared to the DDoS attack, a ransomware attack involves many phases and defences affecting the probability of success for the attacker.

Referring to the cyber kill chain discussed in chapter 2.1, a cyberattack is fundamentally a series of decisions. For ransomware[110], the series of decisions typically progresses from gaining initial access, to lateral movement within the system, culminating in file encryption and a ransom demand. For the defender, strategic decision-making occurs at two points: prior to an attack, when choosing defences to prevent a successful breach, and after the attack, when deciding whether to pay the ransom or pursue an alternative course of action. The attacker, however has decisions to be made during the whole attack on the respective techniques chosen at different phases. This constitutes a multi-stage ARA model where each phase results in a utility maximizing technique. For the contents of this thesis, this analysis would be extensive which motivates our choice of interaction simplification. Using the structure formed by Sevillano et al.[92] as well as the multistage phase temporal evolution of failure states of Mancuso et al.[111], we formulate our ransomware model.

The game is essentially comprised of three phases meaning a Defender-Attack-Defender sequential interaction. We simplify the structure of the MITRE ATT&CK and the Cyber Kill chain to include three phases: initial access, lateral movement and exfiltration[112]. To combat these phases we choose as possible defences actions that among other possible actions within the set negate the probability of these phases succeeding. The key simplification along with this choice of phases is to collapse these phases to probabilistic Bernoulli gate structures, meaning that furthering along the kill chain requires passing all the gates to reach the point of successful data exfiltration and a basis for a ransom demand. Each gate has a distinct probability of successfully stopping the attack, causing certain costs being imposed on the attacker.

This leads to the strategic decision-making of the attacker being collapsed to a simple binary choice of attacking or not. The Figure 26 showcases the temporal nature of the interaction while the influence diagram in Figure 27, with the convention of dotted lines, showcases the nature of partial information:

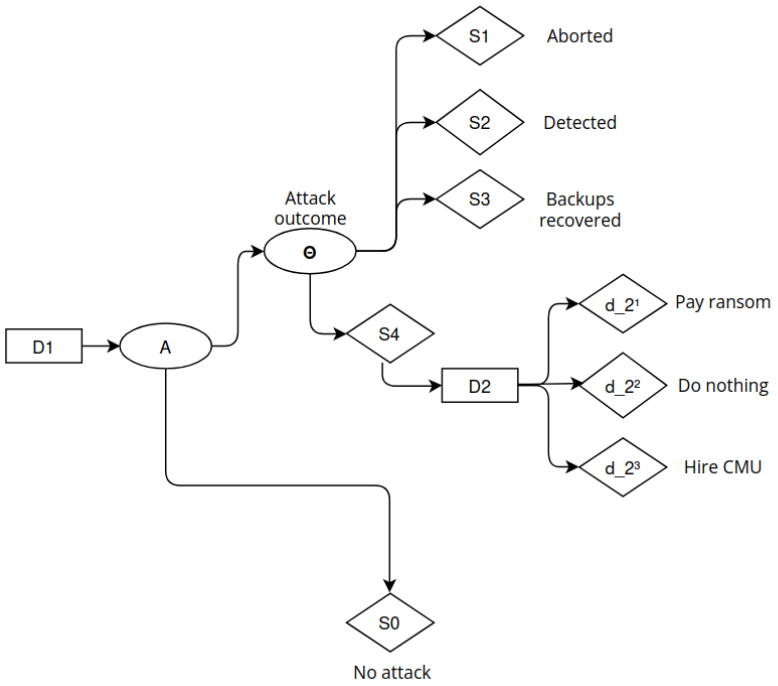

Figure 26: Structure of ransomware Defend-Attack-Defend game for the defender.

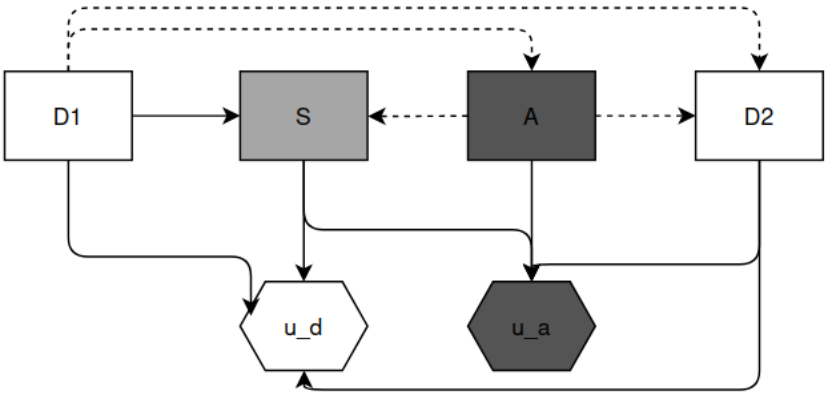

Figure 27: influence diagram.

Figure 28 showcases the structure of the outcome node at $\Theta(\cdot|A = 1)$, meaning the attack decision has been made.

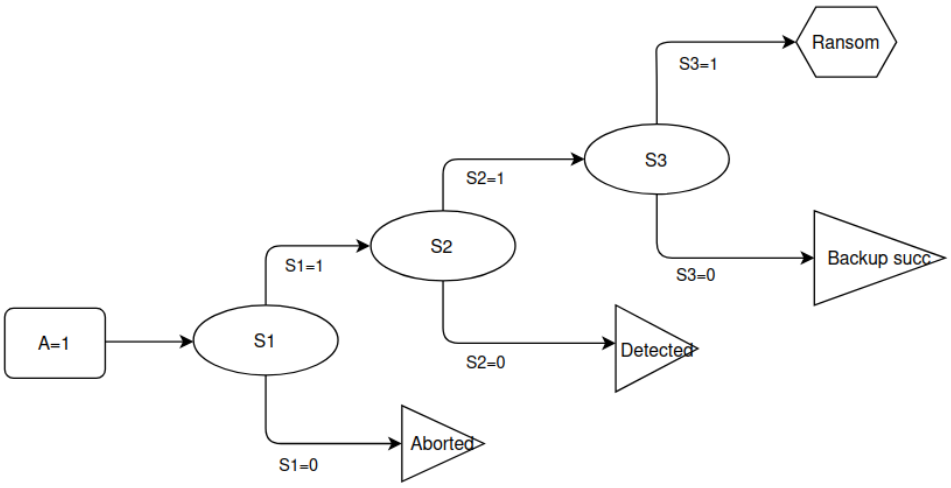

Figure 28: Structure of the outcome node $\Theta(\cdot|A = 1)$.

The defender moves first by deciding upon $d_1 \in \mathcal{D}_1 \Rightarrow (s, i)$, which is the defence portfolio, where each component effects the probability of a successful ransomware attack or insures the assets of the defender. Table 7 shows the alternatives for defences:

Table 7: Defence Configuration Options $s \in \mathcal{S}$.

Defences	**MFA**	**EDR**	**IMM**
Options	True/False	Low/Medium/High	True/False

Similar to the water utility problem, the defender also has a three-tiered set of options for insurance providing alternative protections and coverages in a successful attack yielding in total 36 different portfolios of defence. The attacker observes these defences, evaluates the perceived probability of success, and decides whether to attack.

The outcome node Θ maps each defence option onto its corresponding gate: the attack is aborted in the initial access phase by an implementation of Multi-Factor Authentication(MFA), detected via a robust Endpoint Detection and Response(EDR) defence, or the backups are recovered from immutable backups(IMM) after exfiltration, removing motivation to pay the ransom — the structure of which is highlighted in the decision tree in Figure 28. If all the gates prior to S_4 are cleared, the attack continues to the ransom phase, where the defender optimizes the expected utility $d_2 \in \mathcal{D}_{\in}$. The defender pays the ransom(d_2^1), does nothing and accepts the loss(d_2^2) or hires an external crisis management unit(CMU) to handle the disruption(d_2^3).

The key structure of this interaction is the temporal element, which requires backwards induction within the structure of the decision tree for the attacker. The attacker computes their assessment of an optimal ransom, using which moves backwards within the decision tree of Figure 28 to the origin of the attack decision, where

he uses expected utility to determine whether to attack, where the option is binary having the attack set as $a \in (0, 1)$. The defender then uses this information to evaluate the true outcome distribution of stages the attack proceeds to.

The two tier structure of defences are reflected in literature such as $\mathcal{D}_2$ is the reactive defence set and $\mathcal{D}_1$ is the proactive defence set. The defender inherently models the structure of this game from the point of view of wanting to put up an effective deterrent against a ransomware attack, or in other words wanting to enact a minimum viable deterrent such that potential attacker with their limited information on the defence level choose to attack some other potential victim.

5.2.2 Defender's Beliefs and Preferences

The key assets identified for the defender are their patient-sensitive data that if leaked could cause serious reputational losses as well as be subject to fines under regulations such as discussed in 3.1.2 with also causing extensive downtime in patient care as discussed by Spence et al.[113]. We will utilize their findings on the components that formulate the costs of a ransomware attack on a healthcare provider, which include data loss, business continuity, reputational losses and possible regulatory fines. These parameters formulate the loss incurred by the defender if the ransom is rejected and the attacker decides to leak the data. There is natural variance in this amount depending on if the data is simply encrypted indefinitely or also leaked but we simplify this by modelling total value of data as $v_d \sim \mathcal{N}(2M, 100k)$. This is a probabilistic estimate from the perspective of the defender and information not available to the attacker. The defence portfolios discussed above have the cost structures as well as insurance coverage schemes shown in Tables 8, 9 and 10:

Table 8: EDR tier costs.

EDR	Low	Medium	High
c_{EDR} (€)	0	3,000	8,000

Table 9: MFA and immutable backup costs.

Control	Off	On
MFA (c_{MFA}, €)	0	1,500
Immutable backup (c_{IMM}, €)	0	4,000

Table 10: Insurance premium and coverage.

i	NoIns	Cyber	Comp
c_i (€)	0	2,000	5,000
γ_{ransom}	0	0.80	0.90
γ_{recovery}	0	0.00	0.90

The probabilities governing these three gates are denoted $p_{\text{MFA}}, p_{\text{EDR}}, p_{\text{rec}}$ for the defender and $P_{\text{MFA}}, P_{\text{EDR}}, P_{\text{rec}}$ for the attacker, corresponding respectively to initial access prevented via MFA, detection of attacker behaviour via EDR, and recovery of exfiltrated data via immutable backups. Each of the three defences is additionally chosen to reflect a different combination of uncertainty and observability, illustrating how partial information can be explicitly stated.

The use of MFA has proven to stop a vast majority of initial access methods as stated by Myer et al.[114]. This being generally public knowledge we assume this to be shared information between the defender and the attacker yielding in the following distributions reflecting attack success.

$$p_{\text{MFA}} = P_{\text{MFA}} \sim \begin{cases} Beta(\alpha = 38.8, \beta = 1.2) & \text{MFA =False} \\ Beta(\alpha = 2, \beta = 18) & \text{MFA = True} \end{cases}$$

The lack of MFA implementation causes the gate to be near trivial with an expected probability of $\mathbb{E}[p_{\text{MFA}}|MFA = False] \approx 0.97$. If MFA is present, this yields a mean of $\mathbb{E}[p_{\text{MFA}}|MFA = True] \approx 0.1$ with wider variance reflecting the aleatoric uncertainty of human behaviour.

For EDR, we choose a set of distributions below highlighting the defender's level of information on the effectiveness of EDR on stopping lateral movement and detecting nefarious activity within the network:

$$p_{\text{EDR}}(s) \sim \begin{cases} \text{Beta}(\alpha = 32, \beta = 8) & \text{EDR} = \text{low} \\ \text{Beta}(\alpha = 13.5, \beta = 16.5) & \text{EDR} = \text{medium} \\ \text{Beta}(\alpha = 8.75, \beta = 26.25) & \text{EDR} = \text{high} \end{cases}$$

Immutable backups are one of the most important tools in negating ransomware demands, as they remove the key leverage the attacker holds by having readily available and regularly updated data that would otherwise be encrypted. Having non-immutable backups means that the defender's stored data can be altered, encrypted, or deleted by an external actor, leaving it highly exposed to tampering and, in the event of a breach, to exfiltration and manipulation alongside the primary systems. To negate this risk, immutable backups are designed such that they cannot be accessed or manipulated once written, ensuring the defender retains a clean, recoverable copy of their data even if the production environment is fully compromised[112], in essence supporting the integrity and availability of information.

However, the growing use of backups has led to the prevalence of double extortion [115], in which the attacker not only attempts encryption and exfiltrates the defender's data, but also threatens to release it publicly if the ransom is not paid. This model focuses on this dynamic, in which the presence of immutable backups lowers the success rate of a ransomware attack without eliminating it entirely, as well-maintained backups protect against the encryption threat but do not protect against double extortion. The presence of backups is known to the defender, along with the following distributions:

$$p_{\text{rec}}(s) \sim \begin{cases} \text{Beta}(\alpha = 5, \beta = 45) & \text{IMM} = \text{False} \\ \text{Beta}(\alpha = 40, \beta = 10) & \text{IMM} = \text{True} \end{cases}$$

If the all the gates pass and the game furthers along to the ransom phase at S_4, the defender has to evaluate their optimal reactive choice given a successful data exfiltration. This is denoted as $d_2 \in \mathcal{D}_2$ where the defender seeks to minimize cost $c_d(d_2)$. Each of the decisions has their own cost structures determining utility. The ransom demanded, denoted as $R^\star$ from the attacker is often a fraction of their assumption of the value of said data set, the details of which will be discussed in the section on attacker's parameters. The defender can also refuse the ransom demand and absorb the loss of the attack, which results in the aggregate loss of v_d.The third option is to hire an external CMU, which handles the data breach by actively trying to recover the breached data, returning system functionality and containing the attack severity. The cost of hiring the external CMU has a deterministic cost C_{CMU} of 50k, which is subsidized by γ_{recovery} which is the proportion of recovery costs that the chosen insurance covers. The expected result of hiring an CMU has a probabilistic success rate of $p_{\text{crisis}} \sim Beta(\alpha = 2, \beta = 6), \mathbb{E}(p_{\text{crisis}}) = 0.25$.

Below we summarize the definitions for the expected utility of each $d_2 \in \mathcal{D}_2$:

$$c_d(d_2) \;=\; \begin{cases} R^*(1-\gamma_{\text{ransom}}) & d_2^1 = \texttt{pay} \\ v_d & d_2^2 = \texttt{nothing} \\ C_{\text{CMU}}(1-\gamma_{\text{recovery}}) + (1-p_{\text{crisis}})v_d & d_2^3 = \texttt{CMU} \end{cases}$$

The defender's total costs in the game is defined as a sum of implemented security controls, insurance policy and expected cost after attacker decision, in which d_2 is chosen to minimize cost, i.e. $d_2^* = \arg\min_{d_2 \in \{\texttt{pay}, \texttt{nothing}, \texttt{CMU}\}} c_d(d_2)$. The total costs sum to $c_d = c_d(d_2^\star) + c_s + c_i$, which is fed into the defender's utility function:

$$u_d(c_d) \;=\; \frac{1}{e^c - 1}\left[\exp\left(c\left(1 - \frac{c_d}{R_{\max}}\right)\right) - 1\right]$$

R_{max} is the scales the utility to model the highest possible loss at $\mathcal{D}_2$, which is equal to the full fixed-cost defence portfolio $c_e + c_i$ plus the worst-case (3σ tail) outcome of the three alternatives.

5.2.3 Attacker Strategic Behaviour and Cost Structure

In a ransomware attack, the validity of the attacker's ransom demand is directly related to the existence of immutable backups. However, from the point of view of the attacker, the existence of such backups along with their update frequency and encryption practices are generally not observable outside the target organization. This information asymmetry is solved by modelling the probability of gate success through a mixed beta distribution model that includes along with a base rate backup implementation of the defender, the possibility of outlier cases. If an attacker receives no observable signals from the defender, or more accurately, the defender assumes the attacker type is such that he cannot observe this, a best guess for such an attacker would be to revert to industry-wide statistics on backup implementations. In Sophos 2025 Outlook on Ransomware[116], 54% of respondents(N=3400) used backups to recover their files after data encryption, meaning at S_3 in our interaction. We choose to assume the majority of defender's output no meaningful observable signals while a minority offer either signals on low maturity or expectation of high cybersecurity vigilance. This yields a distribution:

$$P_{\text{rec}} \sim \sum_{k=1}^{3} \pi_k Beta(\alpha_k, \beta_k)$$

The parameters of the distribution are shown in Table 11:

Table 11: Beta mixture model for recovery belief.

Type	π_k	α_k	β_k
Maturity signal found	0.10	4.5	1.5
Base rate	0.80	2.7	2.3
Weakness signal found	0.10	1.8	4.2

The goal of this choice in modelling is to highlight the lack of knowledge in relation to the existence of backups, and see how the model behaves given a defence-agnostic measure of attack decision.

The effect of EDR, in contrast, displays epistemic uncertainty on the effectiveness of the perceived defences. Herein we assume the attacker can observe the level of EDR as our ordinal scaling model, but he is more uncertain than the defender on the effectiveness of the control. Therefore, we choose the beta distributions below to highlight the effectiveness to the controls having a similar mean with the defender's evaluations of $p_{edr}(s)$ with a higher variance:

$$P_{\text{EDR}}(s) \sim \begin{cases} \text{Beta}(\alpha = 6.4, \beta = 1.6) & \text{EDR} = \text{low} \\ \text{Beta}(\alpha = 1.8, \beta = 2.2) & \text{EDR} = \text{medium} \\ \text{Beta}(\alpha = 1.25, \beta = 3.75) & \text{EDR} = \text{high} \end{cases}$$

In game-theoretic literature on ransomware attacks, a few notable observations are made, which are applied to this interaction. Firstly, motivate our structure by a sentiment discussed by Cartwrights et al[117], that agreeing to negotiations by the attacker generally lower the defender's perceived danger, in average lowering the defender's willingness to pay(WTP). Therefore, there is no negotiation in this interaction to model a vigilant and aggressive attacker. Secondly, the ransom demanded by the attacker is often a fraction of the attacker's estimate of the dataset's value to the defender, which is typically referred to as the defender's willingness-to-pay(WTP)[118]. In the studies referenced here, WTP is determined to be essential in optimizing the ransom demand as a high ransom remand with no incentive to pay it from the defender causes a low payment percentage amongst victims. Therefore, the attacker models the optimal ransom demand to be a function of the expected value of independent recovery as well as assumed WTP. Since the attacker cannot be certain of either of these, they are sampled from $w_A \sim \mathcal{N}(0.50, 0.10)$ and $v_a \sim \mathcal{N}(v_d, \sigma_v^2), \sigma_v = 100k$ to reflect the common payment amounts as well as an approximate evaluation of defender's v_d with some variance. To add additional variance to this optimal ransom demand, we include the possibility of the attacker assuming the existence of insurance, which can lead to the attacker assuming increased liquidity inflating the ransom demand[119]. This is modelled similarly to the existence of backups, as private knowledge to the defender, as a mixed beta model of type:

$$\eta \sim \sum_{k_i}^{2} \pi_{k_i} Beta(\alpha_{k_i}, \beta_{k_i})$$

Table 12: Beta mixture model for attacker's belief about defender insurance status (η).

Type	π_{k_i}	α_{k_i}	β_{k_i}
Signal found (broker/leak)	0.15	3.0	12.0
No signal (default)	0.85	1.0	19.0

The ransom demand is therefore computed as $R^\star = [\frac{w_A v_A}{1-\eta}]$ truncated on the interval $[0, v_a]$ ensuring the ransom demand does not go above the perceived value of independent recovery i.e. the monetary value of the files.

The costs imposed on the attacker if caught prior to ransom phase are $c_{op} \sim \mathcal{N}(5000, 1000)$ and $c_{det} \sim \mathcal{N}(50k, 8k)$, which denote costs of operations and getting caught and being imposed with cost c_{det} with a mean probability of 0.02, yielding $p_{caught} \sim Beta(2, 998)$ similar to DDoS prosecution rate. The cost of operations was inspired by Galinkin [120] whereas the cost of detection follows a case example of a fine structure coming from a case of an attempted ransomware attack.

And thus, the expected utility for the attacker is computed through backwards induction as shown in Figure 28 from successful ransomware attack to the decision node $\mathcal{A}$ shown in Figure 26, yielding $\mathbb{E}[c_a|s]$

$$\begin{cases} \Psi_3 & = (1 - P_{\text{rec}})\, R^\star \\ \Psi_2 & = P_{\text{EDR}}(s)\, \Psi_3 + (1 - P_{\text{EDR}}(s))(-p_{\text{caught}} \cdot c_{\text{detect}}) \\ \Psi_1 & = P_{\text{MFA}}(s)\, \Psi_2 \\ \mathbb{E}[c_{\text{a}} \mid s] & = \Psi_1 - c_{\text{op}} \end{cases}$$

The expected $c_a|s$ is then scaled as common between the best gain and worst loss and then raised to the power of k_a to reflect risk proneness:

$$c'_a = \frac{\mathbb{E}[c_a|s] - c_{a,min}}{c_{a,max} - c_{a,min}}$$
$$U_A(c_a) = (c'_a)^{k_a}, k_a \sim U(8, 10)$$

This form of expected utility informs the decision of the attacker at $a \in \mathcal{A}$ by comparing the the utility expected from attacking to not thus:

$$a = \begin{cases} 0 & U_A(c_a) \leq U_A(0) \\ 1 & otherwise \end{cases}$$

5.2.4 Results

The attack problem to be solved is the following where the attacker seeks to decide whether attacking is an expected positive-utility pursuit. The attacker's random

utilities and probabilities are drawn from each round.

$$(U_A(E[c_a]), P_{\text{MFA}}(s), P_{\text{edr}}(s), P_{\text{rec}}, p_{\text{caught}}, c_{\text{det}}, c_{\text{op}}, w_a, \eta, v_a) \sim F$$

The draws are thus used in determining the optimal ransom and utility at node S by comparing the utilities between attacking and not attacking, which is determined as $U_A(c_a) \geq U_A(0)$. This results in probability distribution $\hat{p}_D(a|s)$ for each level of defence along with a distribution $\mathcal{R}(s)$ denoting the distribution of ransom demands that is sampled from each successful trial.

The implementation of this algorithm can be seen in the appendix as Algorithm 5. We choose to visualize $\hat{p}_D(a|s)$ by choosing 3 different portfolios that reflect different defence regimes - weak, mid and full defence. This results in the attack probabilities conditioned on the defence options in Figure 29.

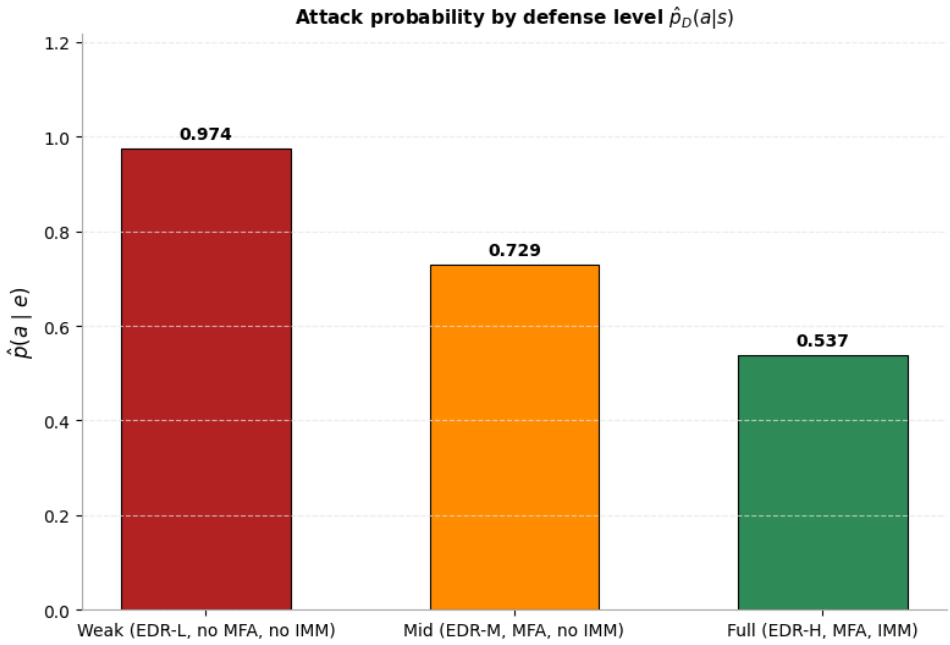

Figure 29: Attack probability $\hat{p}_D(a|s)$.

The defender problem then starts with evaluating $\hat{p}_D(a|s)$ with respect to aleatoric uncertainty and the defender's own evaluations of Bernoulli gate probabilities. The defender draws their own probabilities $(p_{\text{MFA}}(s), p_{\text{EDR}}(s), p_{\text{rec}}(s), \hat{R}(s), v_d, p_{\text{crisis}}) \sim G$ If the attacker decides upon attacking, the attack continues to phase $S = 1$ where the validity of attack success at each gate is now determined by defender's probabilities, p_{mfa} of which is common knowledge, but p_{edr}, p_{rec} are the actual probabilities of success. These differ from the attacker's evaluation as p_{edr} is assumed to be observable but with aleatoric uncertainty around the effectiveness of the control and p_{rec} reflecting the ability to recover independently via backups is pure epistemic uncertainty as the existence and update frequency of the backups is assumed to be private information to the defender. The outcome distribution per defence level is determined by Monte

Carlo simulating the outcome of of the attacker by the below Bernoulli trial structure using ransom draws $\xi_1, \xi_2, \xi_3 \sim U(0,1)$:

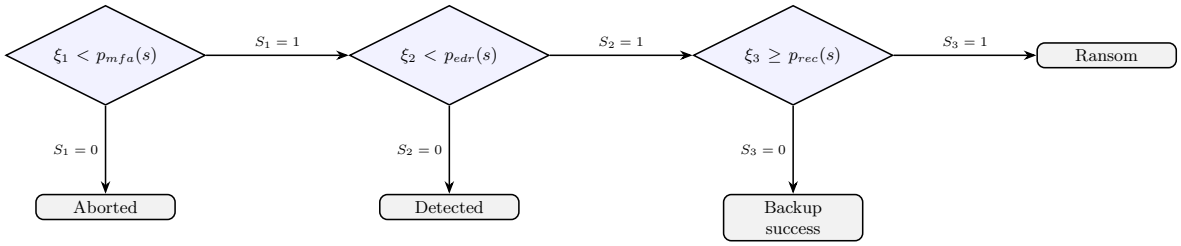

Figure 30: Structure for computing $\hat{p}(\theta|s)$.

This results in a probability distribution $\hat{p}(\theta|s)$, which informs the backwards induction of the defender if the ransom stage is reached. If this stage is reached, the defender chooses the reactive defence $d_2 \in \mathcal{D}_2$ that minimizes cost, after which the total costs for the defender determine utility:

$$c_d = c_s + c_i + c_{d_2}$$
$$u_D(c_d) = \frac{1}{e^c - 1}\left[\exp\left(c\left(1 - \frac{c_d}{R_{\max}}\right)\right) - 1\right]$$

Using $K = 20000$ for the number of Monte Carlo simulations with $c = 1$ for risk-aversion, this yields the below optimal and suboptimal choices for $(s,i) \in \mathcal{S}, \mathcal{I}$ in Table 13:

Table 13: Ransomware Attack; Portfolio ranking by expected utility.

rank	**defences**(s,i)	$\boldsymbol{\psi(s,i)}$
1	EDR-L+MFA+IMM+Cyber	0.9969
2	EDR-L+MFA+IMM+Comp	0.9962
3	EDR-M+MFA+IMM+NoIns	0.9962
⋮	⋮	⋮
34	EDR-H+noMFA+noIMM+NoIns	0.9396
35	EDR-M+noMFA+noIMM+NoIns	0.8886
36	EDR-L+noMFA+noIMM+NoIns	0.8098

Using the common Pareto-optimality condition, we plot the annual costs of the portfolios with respect to their expected utilities in Figure 31:

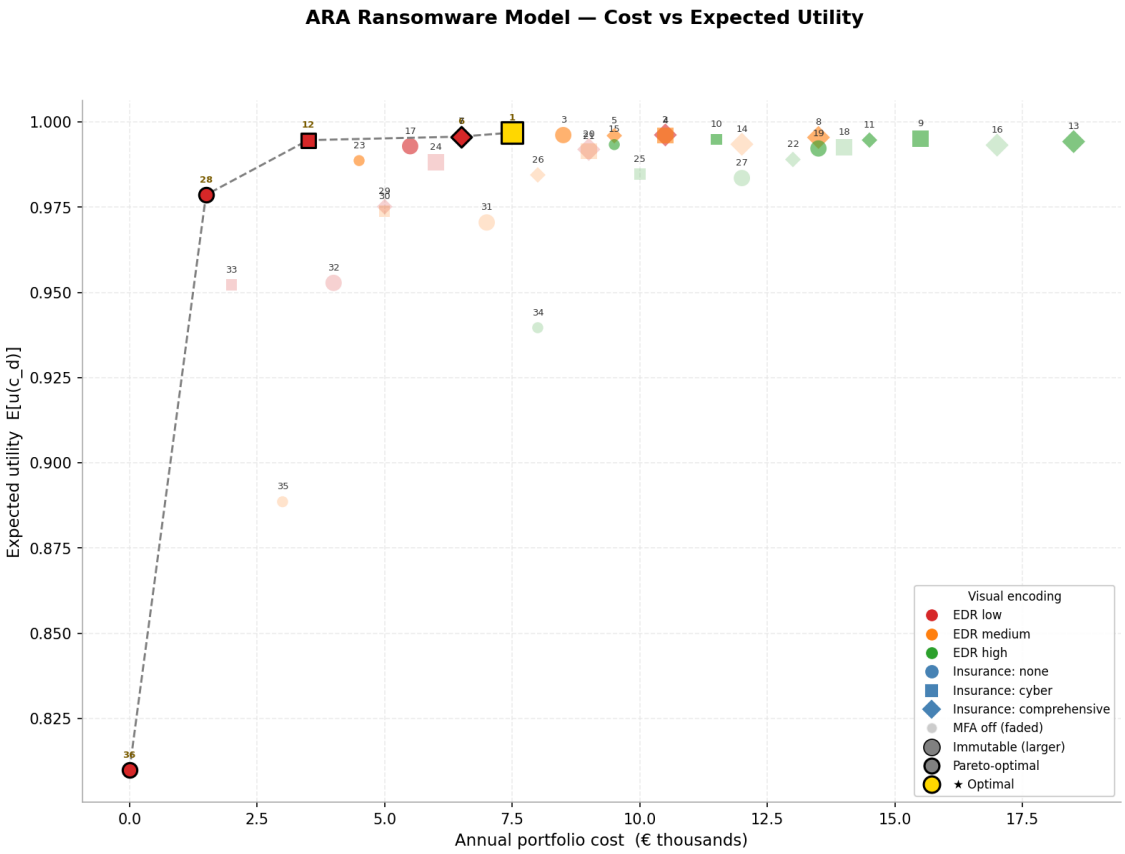

Figure 31: Pareto-optimal defences $(s, i) \in \mathcal{S}, \mathcal{I}$.

The effectiveness of the defences of the Pareto frontier are visualized in Figure 32 by the outcome distribution that shows how the different portfolios affect the success probabilities of the outcome node Θ after an attack decision, the structure of which is seen in Figure 30:

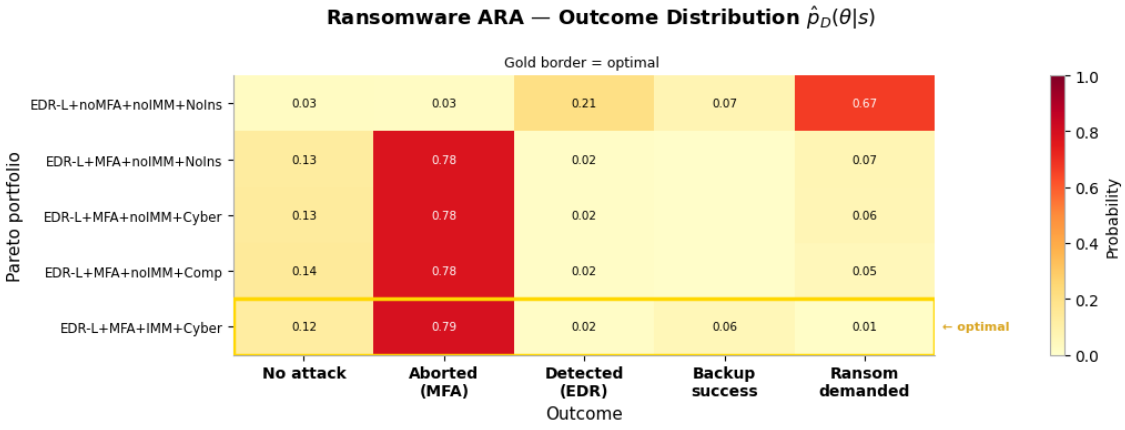

Figure 32: Outcome distribution $\hat{p}_D(\theta|s)$.

The results deem that the optimal defender portfolio with respect to expected

utility $\psi(s, i)$ and price is determined by focusing on mitigating initial access and data exfiltration by immutable backups as well as subsidizing the ransom cost by choosing cyber insurance without recovery cost coverage. These results indicate across the Pareto-optimal portfolios that mitigating initial access by implementing MFA is key in aborting the attack at S_1 where as detecting the attacker mid-attack via EDR is not valued as highly as an attack stopped prior to lateral movement is always preferred. The two-stage nature of this interaction means that the best defence is not an deterrent but rather a robust defence such that an attacker deciding to attack leads to the attack being aborted quickly without further escalation. The full-defence portfolios lower the probability of an attack, but the respective cost associated with a full defence is high, which is penalized by the utility function, thereby causing the preference for lower-cost defences. Additionally, from the 36 possible defence portfolios, 24 are deterministic at stage $\mathcal{D}_2$ as the optimal choice is to pay the ransom. Variance in this stage only occurs if no insurance is chosen which causes the optimal action at this stage to be largely dependent on variance caused by the draws for the different decisions with payment remaining the dominant strategy.

5.3 Sensitivity Analysis

To verify the robustness of our models, we introduce sensitivity analysis to the thesis to identify whether deviations in the inputs cause significant changes to the results, as for both models the variations in the expected utility of the top alternatives are marginal. As determined by Rios[121], sensitivity analysis is done to clarify to the decision maker what judgements should be evaluated more carefully. We check the effect of level of risk aversion on the Water Utility case study, and the choice of prior beliefs on the resulting optimal portfolio for the Nursing Home case.

For the DDoS model, we check how the level of risk aversion changes the defender's preference as the coefficient of risk aversion c is altered. Fundamentally, for a concave utility function of form $u(c_d) = a - b \cdot exp(c \cdot c_d)$, when c increases, higher losses are penalized more heavily. This is showcased in Figure 33 for a simple utility function spanning the loss space of interval $[c_{d,min}, c_{d,max}] \rightarrow [0, 100]$ with the coefficient of risk aversion spanning $c \rightarrow [0.1, 10]$:

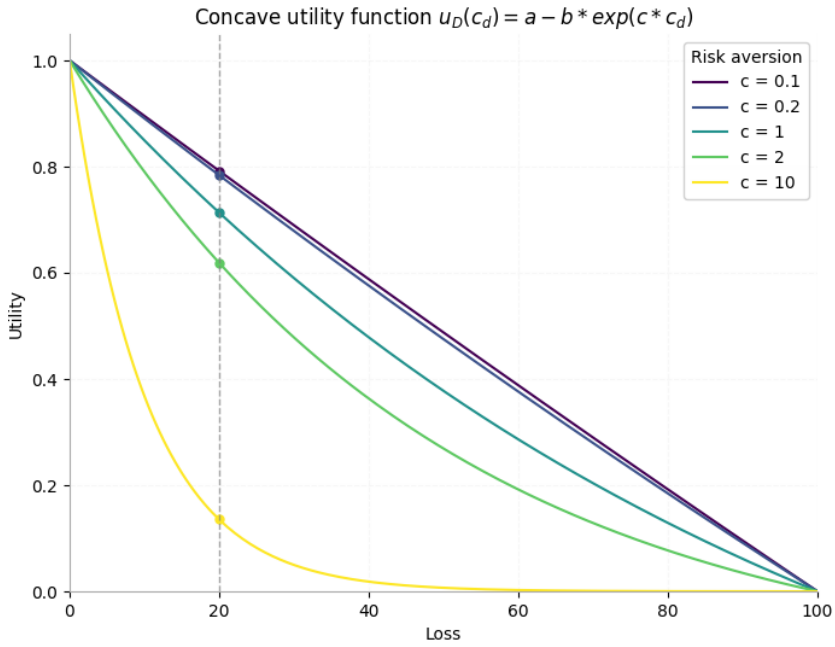

Figure 33: Concave utility function.

As c increases, the more quickly losses are penalized, meaning that such a player seeks to minimize exposure to risk. In ARA, a risk-averse player would therefore seek to pay more to lower their exposure to such risks.

We check the robustness of the DDoS model by iterating over $c \in [0.1, 0.2, 1, 2, 10, 30]$ as well as limiting the budget to 4.800€ to ensure the defender cannot choose the consistently high-utility strategy of $(1000gbps, NoIns)$ to introduce some variance in the results. This budget limitation was discussed by Rion Insua et al.[8] as a portfolio optimization in real-life applications often have tangible constraints. Table 14 showcases the new utility-maximizing portfolios as c is varied:

Table 14: Optimal portfolio and expected utility for different levels of risk aversion.

c	Optimal portfolio	$\psi(s, i)$
0.1	0g+Comp	0.966320
0.2	0g+Comp	0.964697
1.0	0g+Comp	0.950057
2.0	0g+Comp	0.928186
10.0	0g+Comp	0.727236
30.0	0g+Comp	0.390480

The new dominant strategy across risk aversion coefficients is $(0gbps, Comp)$, which means the defender in his aversion to risk values more a low-cost, insurance based protection that covers the costs resulting from a DDoS attack. This is a

tangible result displaying how a insurance policy $i \in \mathcal{I}$ covering a majority of losses while costing a fraction of the protections schemes $s \in \mathcal{S}$ is a utility-maximizing decision for any level of risk aversion. Further analysis could be done with new prices for the insurance policies, and as discussed by Rios Insua[8], the optimal price for a insurance policy could be computed by sensitivity analysis, as the risk premium to be paid can be found at the point where the defender is indifferent between two insurance options (or between purchasing insurance and forgoing it), holding the security controls portfolio fixed.

For the Ransomware case, we choose to analyse the effect of a parameter of common knowledge, $p_{mfa}(s)$. Especially as the existence of MFA in the system is originally assumed to lower the success rate to a mean of 10%. We introduce some increased epistemic uncertainty in the system by increasing the probability by 1% increments to see how the optimal portfolio changes with these perturbations. The optimal portfolio is stationary with inconsistent fluctuations until the probability is increased to 28% where the new optimal portfolio includes investment into a higher tier of EDR. The fluctuations in the expected utilities can be explained by MC noise as the differences amongst the top candidates are marginal(as seen in Table 13), resulting in an alternative optimal portfolio mid-sensitivity analysis. This is of note for future research as MC methods especially in the case of near identical utilities can cause inconsistencies in the results, while also referring to the Pareto-frontier for evaluations of non-dominated solutions provided that the monetary evaluations of the defences are realistic.

6 Discussion

To offer insights into resilience planning under uncertainty in cybersecurity, this chapter includes open discussion on the topic with implications given by the ARA implementation. Also the notable limitations of ARA are identified along with it's broad implications as an analysis tools as well as the practical implications of ARA in cybersecurity.

6.1 Interpretation and generalizability of results

The case studies researched in this thesis focused on attack types of high industry-specific relevance, so it is worthwhile to discuss how the results should be interpreted and how do they generalize.

The DDoS example concluded with the statement that the optimal portfolio is the largest defence option with no insurance policy required. The results indicate that choosing such a defence functions as an effective deterrent, as attackers on average attempt a few attacks before stopping or moving toward another potential victim. Given a budget-constrained SME as the defender, any of the Pareto-optimal solutions(disregarding the zero-protection portfolio) offer protection that, depending

on the risk-aversion of the defender, may be sufficient. However, as displayed by the sensitivity analysis in Chapter 5.3, given budget constraints, the defender's utility-maximizing portfolio becomes simply an insurance policy, meaning that a risk-averse defender, if not given the opportunity to enact full deterrence, would rather pay for an insurer to absorb the risk of an attack. This is sensible in the context of this case study, as the insurance is relatively inexpensive while covering the majority of the cost from inflicted downtime. The ransomware case produces a list of defences in line with observations from industry, determining that MFA usage and immutable backups are key in protection from ransomware attacks, alongside a choice of cyber insurance, resulting overall in a defence portfolio depicting defence-in-depth and distributed redundancy. The model additionally showed how different levels of observability can be explicitly stated (e.g., MFA being common knowledge, immutable backups private) and how this affects the results.

The results depict generalizable characteristics of good cybersecurity posture, namely that MFA usage is highly rated and that the best defence functions as a deterrent. This is also supported by common insurance mandates, as MFA is typically required from insurees — an argument reinforced by the ransomware case as the use of MFA is a utility-increasing act. However, the generalizability of these results beyond this qualitative conclusion is largely challenged by the simplification of the decision spaces. The DDoS example assumes the attacker has no say in the strength or duration of attacks attempted[122], but only in the actual number of attacks. The ransomware interaction showcases this same structure of simplification, as the choice given is binary, after which the attacker makes no further active choices, which is a conscious limitation worth noting, as in reality ransomware attacks follow a structure in which decisions are made sequentially [112, 123]. This is, however, modelled using the backward induction structure, which aims to describe how an attacker might evaluate the utility of attacking, i.e., using the available information to deem whether attacking is a positive-utility pursuit. The decision space limitation also applies to the defender, especially in the DDoS attack, since the choices available are discrete, without the option for dynamic optimization mid-interaction, as well as limiting the options available to just the strength of the attack, disregarding other definitions of severity in DDoS attacks. Beyond structural simplifications, the numerical inputs for the defence costs are in addition approximations. In addition, the Pareto-optimal solutions include the cost of respective defence portfolios both in the pure cost attribute evaluations and as part of the total cost in determining the attribute of expected utility. Thus, the generalizability of the results should be validated by disregarding the defence cost in the total cost calculation, to see whether this has a noticeable effect on the relations between the alternative defence portfolios. Given the simplified interaction structures, the universality of the results are to be verified with further research on more complicated decision spaces and alternative attributes for the Pareto-optimal defences.

6.2 ARA in resilience

6.2.1 ARA and disruption event cycle

As discussed in Section 3.2, the resilience of a system is analysed inherently with respect to a disruption. Especially within cybersecurity, where the possible threat scenarios and attack vectors are abundant, the task of assigning the limited resources to the correct threat scenarios and correctly allocating them within each threat scenario is a critical challenge to be solved. The attack scenarios of this thesis highlight ARA's use within certain scenarios while giving no weight to the prioritization of how prevalent a given attack is against the space of possible attacks. However, the results given by each model support resilience planning within scenarios as the results can guide the planning of effective deterrents.

If a defender deems that their current exposure leads them to be vulnerable to a certain attack, the Pareto-optimal defence portfolio results seen in Figures 25 and 31 fundamentally limit the set of feasible defence portfolios as the optimization task of minimizing cost while maximizing expected utility forms the Pareto-frontier of non-dominated solutions, i.e. the set of feasible defence portfolios for which no alternative portfolio achieves equal or greater expected utility at equal or lower cost. The defender can then assess based on their risk appetite which of the Pareto-optimal solutions to select, while, at the very least, limiting the set of defence portfolios to choose from easing the decision-making process. As the resilience of a defender with respect to a certain attack scenario is considered, the optimal defences also offer insights into effective deterrent design, fundamentally solidifying the prepare dimension of resilience. This is especially seen in the DDoS case, where a solution enacting an effective deterrent is found. This creates the possibility of optimizing for a minimal viable deterrent, i.e. the minimal defence that blocks most attacks. This is visualized in Figure 34 as we remove the coarse approximations for defence costs and plot the expected utilities of the defence portfolios for the ransomware case in ascending order, yielding $\psi(s, i)$ with the Pareto-optimal portfolios highlighted:

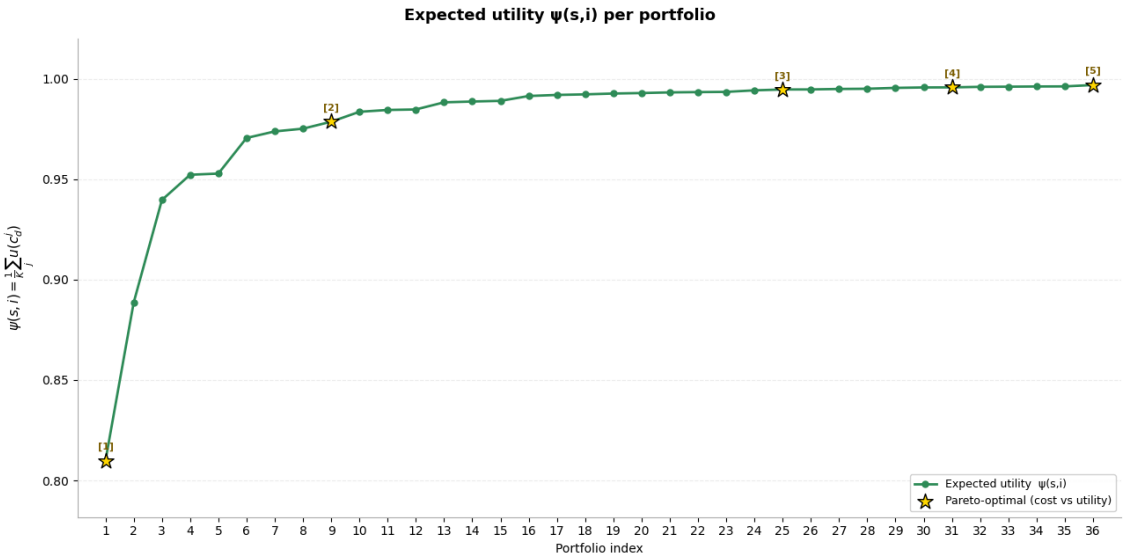

Figure 34: Expected utility as a resilience metric.

This can be thought of as a measure of resilience, and by disregarding the monetary dimension, from this plot in Figure 34 the point of diminishing returns can be determined leading to the identification of the most effective defences causing the largest increases in expected utility.

Reverting back to the disaster event cycle common in resilience frameworks, the sequential play presented in the ransomware case can offer insights into phases beyond preparation as the series of decisions made is effectively a division between proactive and reactive defences. This methodology could be expanded upon, but effectively modelling the best-practice proactive defences also models in two-stage games the optimal reactive defences ensuring system disruption is absorbed and recovery is swift. In the ransomware case, the common optimal reactive defence was to pay the illicit ransom, which was assumed in the game to return the encrypted files and essentially recover system functionality quickly. If expanded further, this could create a valuable argument for law enforcement to criminalize or regulate ransomware payments as this ARA model shows that despite defences, often payment is the least costly option after a successful attack, and therefore showcasing from a decision-analytic perspective the financial reality of cybercrime and the validity of the business model for the cybercriminals[117]. ARA also includes the possibility of improving upon the ability to adapt post attack as the parameters and prior beliefs of the model can be updated according to novel incident data making the system more robust.

Additionally, as the common disruption cycle seen only covers one event, a truly resilient system that recovers from disruptions and adapts while improving operational resilience as a result should improve its system functionality. This is comparable to the paradigm of ARA updating its prior beliefs and problem structure by novel attacks as information on attacker behaviour accumulates, i.e. referring

to ARA as a tool for analysing novel attacks and being thus more prepared for future events. In the language of Bayesian decision theory, the previous novel attack methods with their respective intricacies formulate the set of priors. The Figure 35 showcases a two-disruption cycle in light of an adverse event and how functionality improves post-disruption.

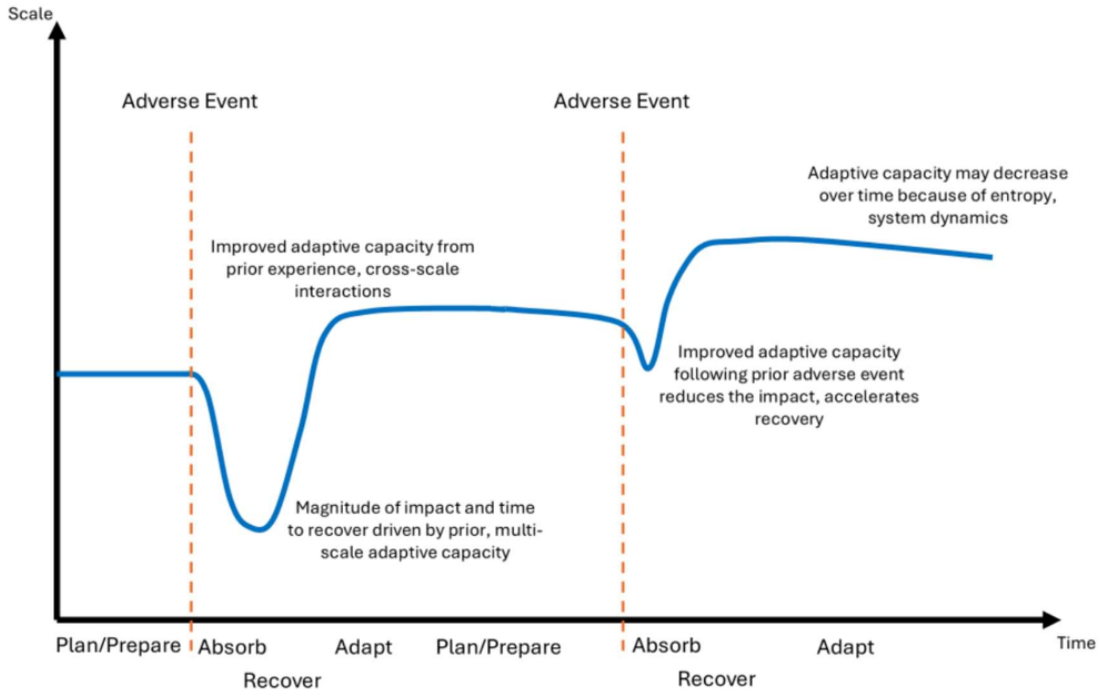

Figure 35: System functionality with multiple adverse events[59].

At their core, resilience frameworks such as the Linkov matrix [5] or the cyber resilience cube [70] prioritize breadth, where cybersecurity is treated as a vast system of interconnected nodes spanning disruption phases and domains. These frameworks offer tools for increasing the granularity with which resilience can be scoped and audited. However, similar to probabilistic risk analysis, treating the threat space as nature, or exogenously, does not account for the intentionality and reasoning of adversaries. Moreover, the descriptive nature of such frameworks does not focus upon the decision-analytic problem associated with enacting resilience, i.e. the investment decision to be made when optimizing for resilience. Therefore, the integration of ARA in cyber resilience frameworks and disruption event cycle is two-fold: expressing uncertainty over attacker behaviour and converting this uncertainty in to optimizable actions, where the highest utility is achieved through portfolios of distributed redundancies which is a key tenet of resilience[23, 4].

6.2.2 Decomposition of uncertainties in the disruption event cycle

One of the central arguments in the literature as discussed earlier in favour of ARA is its ability to partition uncertainty into different sources or dimensions. If we include the conventional assessment of resilience through its disruption life cycle, we could formalize a novel uncertainty-disruption matrix that considers how each level of uncertainty in ARA modelling considers each phase of the disruption cycle and and how it contributes to aggregate resilience by aiding the evaluation of these levels separately Table 15 showcases this matrix:

Table 15: Decomposition of Uncertainty within the Disruption Event Cycle[18, 99, 6].

Uncertainty Type	**Prepare**	**Absorb**	**Recover**	**Adapt**
Aleatory (irreducible randomness)	Stochastic exploit success rates; hardware failure distributions; timing variability of attacks	Random propagation of breach across systems; uncertain collateral damage scope	Residual randomness in recovery time; unpredictable cascading failures during restoration	Inherent variability in future threat landscape; irreducible uncertainty in long-run attack frequency
Epistemic (reducible knowledge gaps)	Unknown attacker capabilities; incomplete threat intelligence; unidentified vulnerabilities	Incomplete picture of attacker intent and lateral movement mid-incident	Uncertainty about whether all attack vectors have been closed; unknown persistence mechanisms	Bayesian updating of attacker type from observed incident; revised beliefs about capability set
Concept (unknown strategy space)	Attack vectors outside modelled threat space; novel TTPs absent from threat model	Attacker executing strategy outside defender's feasible set; zero-day in unmodelled component	Unanticipated recovery blockers from novel attack class; incomplete kill-chain assumptions	Expansion of feasible strategy set post-incident; influence diagram revision to include new action nodes

These sources of uncertainty with respect to the disruption event cycle allow for increased granularity and the decomposition of a general uncertainty into different types which all call for different actions and inform different dimensions of the decision-making process. Understanding how aleatoric uncertainty is exhibited while preparing for a disruption requires probabilistic assessment similar to quantitative risk analysis, whereas epistemic uncertainty stems from an endogenous attacker that optimizes its expected utility, while concept uncertainty stems from incomplete information regarding attacker rationality. As aleatoric uncertainty is irreducible, the decomposition of uncertainty as such highlights how certain aspects comprising resilience have to be accounted for even despite a theoretical perfect analysis of epistemic uncertainty, i.e. being prepared for unknown unknowns. This has to be managed by resilient system principles such as having redundancies in place or choosing a significant protection threshold since aleatoric uncertainty cannot be

negated by an increase in information. Epistemic uncertainty in principle is reducible and converges given more data on attacker behaviour, and therefore, information on past attempts is critical. Concept uncertainty is the most complicated to negate as the decision-making process for each rationality type needs to be re-evaluated. In this thesis, concept uncertainty was truncated to focus on one level of attacker rationality, but these claims persist nevertheless. ARA is an effective conduit for this matrix because it both distinguishes and propagates the three uncertainty types across the model, resulting in a probability distribution of likely decisions to be optimized against. As discussed in Section 6.1.1, ARA offers decision support across each phase of the disruption event cycle. The matrix structure introduced above extends that argument by decomposing ARA's contribution more precisely showcasing how each disruption phase is comprised of an aggregate sum of its sources of uncertainties.

6.3 Limitations

The main motivation in implementing ARA to cybersecurity was the intersection of two characteristics of cyber incidents: intentionality and incomplete information. This choice is commendable in principle, but given the ever increasing size of decision spaces for cybercriminals and the formulation of the cost structure diverging along with the decision space, some notable limitations of ARA were identified.

The solution concept of ARA assumes a certain problem structure as well as certain significant simplifications in terms of the solutions space. Conceptually this can be seen in the models built in this thesis as both exhibit certain simplifications to ensure the cost structures and influence diagrams remain tractable and the MC implementation does not require excessive computational resources, most of which were highlighted in 6.1. In addition to this, the use of subjective probability assessments during problem formulation (duration of DDoS attack, effectiveness of MFA etc.) has been contested in literature. Flake et al.[124] note that the use of subjective beliefs is often based on unjustified assumptions and incomplete information. Wang et al.[125] highlighted that the use of subjective beliefs in ARA fundamentally affects the validity of the resultant optimal defences, as these assessments depict a degree of belief in said occurrences.

As discussed in Section 4.4, while ARA offers increased realism in its propagation of uncertainty and traceability, the increased realism comes at a cost of increased computational and conceptual complexity[89]. Constructing the BAIDs in these interactions requires a thought process though useful in uncovering information about cyber exposure and the central assets at risk, also requires a lot of work in identifying the relations between actions on objectives, costs structures and the level of observability during the interaction. Figure 36 showcases an BAID for a cyber attack against a power grid system, showcasing the inherent complexity that arises when the decision space becomes larger and the interactions between uncertainties are interconnected:

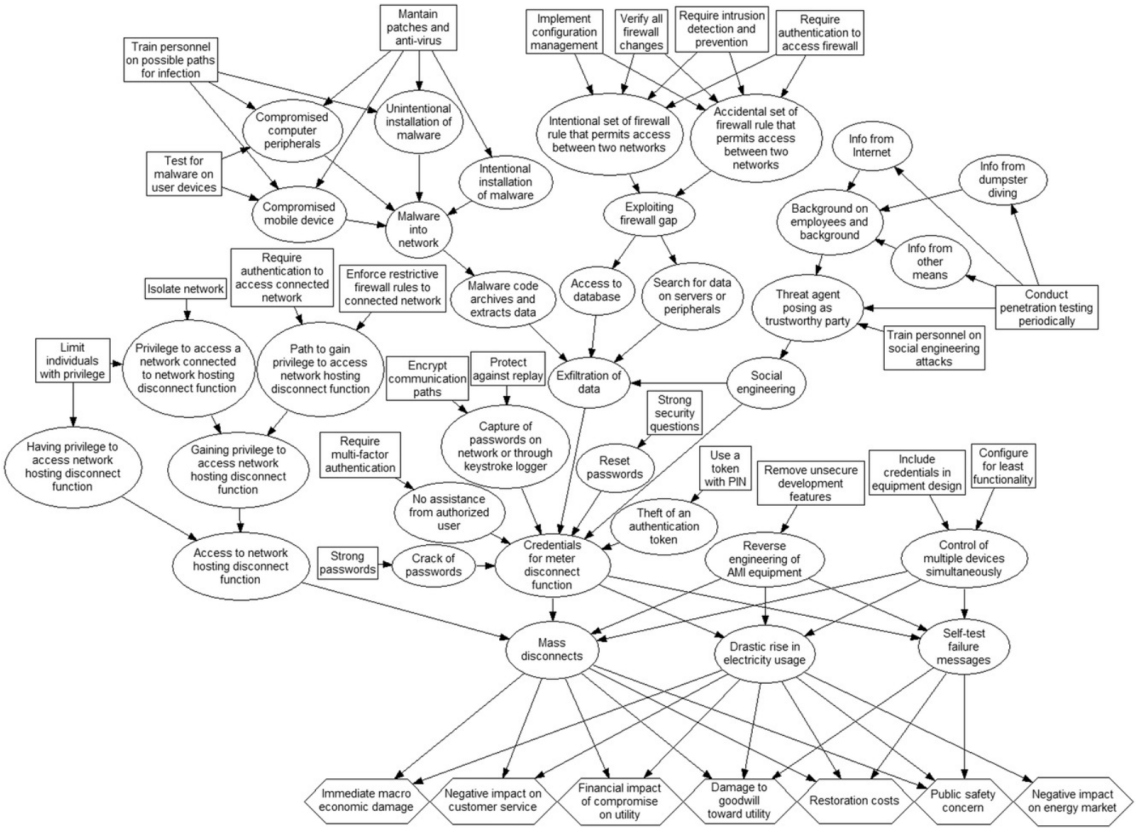

Figure 36: BAID for cyber threats against a power grid [74].

This influence diagram highlights the complex causal dependencies of decisions and outcomes of a complex system. Within the context of this thesis, this influence diagram functions as a point of comparison to reflect the complexity of real-life systems and how in reality, eliciting the information to populate the nodes of such an ID can in practice be exceedingly difficult.

As well as the additional work to be done given larger decision spaces, the theoretical implementation of concept uncertainty, while in principle being useful and resulting in more realistic solutions, the actual implementation of uncertainty over the attacker rationality requires extensive elicitation burdens. Firstly, each solution concept of the attacker (nonstrategic, level-k, Nash-seeking etc.) requires their own elicitation of uncertainty over attacker behaviour. This in and of itself is burdensome, but secondly, given true concept uncertainty meaning the defender propagates uncertainty not just over attack solution space, but also *which* solutions space the attacker exhibits. This is analogous to a Bayesian mixed-model in game theory, in that each attacker exhibiting a certain solution space has its own individual probabilities depicting most likely behaviour from a pool of adversaries[31]. In theory this is valuable, as the uncertainty-disruption matrix in Section 6.1.3 shows, but in practice eliciting the required information for each rationality paradigm as well as a prior probability over the rationality types makes the task non-trivial.

Thus, ARA's key limitation is not a flaw external to the method but a direct consequence of its founding motivation. It offers tractability and realism in its structure while also making the information structure of the defender explicit as well as supporting decision-making under uncertainty. This does however come at a cost of increased computation and elicitation as assumptions of complete information are not present nor does it allow for efficient computational solutions as of now as most literature relies upon brute-force enumeration through MC. Essentially, the more complex the interaction is, the more the model relies upon approximations, expert judgements and truncation choices that become more difficult to verify as the scale of the interaction is increased by a number of players, or dependencies between decisions.

6.4 Theoretical and practical implications of ARA

6.4.1 The role of information structure

As discussed in Chapters 4 and 5, the flow of information especially in sequential interactions is key in determining the behaviour of the attacker. For example, in the DDoS case study, a certain defence level was deemed optimal in that it effectively limited most attacks to only a handful of attempts, after which the attacker could be assumed to redirect their efforts towards a more viable target. Hence, the effect of information broadcast by a well-defended defenders warrants further discussion.

Information broadcast refers to how the behaviour of an attacker is altered if the defender willingly through certifications or compliance notices communicates a certain level of defence. Depending on the cost structure, i.e. the motivations and the type of the attacker, this could either function as an deterrent or an incentive. As the cost structures that determine the incentives of attackers in ARA are often formed from their conditional cost nodes and converted to a monetary form, an attacker seeking monetary gains could be deterred by this signal of cyber maturity and vigilance since it could cause the expected utility to decrease. An attacker who values prestige or notoriety may assign additional utility to successfully penetrating a well-defended organization and gaining information within said organization, with public statements of security maturity serving to further motivate such attacks. This is a well-documented phenomenon in cybercriminal communities[126], which could indicate that signaling only further incentives attacks despite defences. Using the language of game theory, this is a signaling game where the asymmetry of information between the players is a key determinant of the receiver's behaviour, and therefore, the outcome of the game[127]. Information is therefore deemed a strategic asset, and the effect that its broadcast has on the attacker space is to be understood.

6.4.2 Extending ARA to multiple defenders

If we assume that an attacker redirects their efforts toward a more viable target following an unsuccessful attempt, this motivates discussion on the implications and consequences of the single-analyst perspective intrinsic to ARA. In adversarial

interactions and in the cyber context of this thesis, this is especially pertinent, as using the language of game theory, ARA is a single-agent optimization framework meaning the defender optimizes their defences with respect to a single adversary. This does not take into account the existence of multiple adversaries, but perhaps more importantly, the existence of multiple defenders. In this thesis, this refers to other SMEs possibly in the same field or as a part of the same supply chain. As discussed, ARA can be used in the identification of optimal defences as well as concepts such as minimum viable deterrents. However, within an ensemble of different interlinked SMEs, the implementation of a minimal viable deterrent does not necessarily lower aggregate system level risk, but rather displaces the risk targeted at the observant. This systemic level cyber risk is modelled for example by Welburn and Strong[128]. The decision-making process of an attacker passing on the defender of minimal viable defence is showcased in the ransomware case study in which the attacker evaluates the expected utility of attacking. This can be generalized to N other defender's as highlighted by Sevillano et al.[92], which determines the probability of being attacked($A = a_1$) as:

$$p_D(A = a_1 \mid d_1) = \mathbb{P}(\psi_A(d_1, a_1) > \max\{U_A(a_0), \psi_A(a_2), \ldots, \psi_A(a_N)\}) .$$

Summarized, this is the mathematical formulation of being attacked($A = a_1$) in the presence of other viable victims($A \notin a_1$), where a single defender in the conventions of ARA essentially focuses on ensuring the expected utility for attacker pursuing you is lower than all other possible victims. In a budget-constrained SME environment, this behaviour is practically sensible, but extended across many independently optimizing organizations, it is formally captured by the concept of the Price of Anarchy [129]. In the context of this thesis, this refers to the gap between the socially optimal, cooperative outcome achievable across all defenders, and the NE each individual player reaches by optimizing in isolation. It is therefore worth noting that the locally optimal solutions ARA provides, while not being explicitly equilibrium states, single defender's maximizing their subjective utilities may have system-wide spillover effects not made explicit to the analyst[117].

6.4.3 Practical implications for cybersecurity

However, as analysing the cybersecurity of an SME, the system-wide implications as such are not of prioritized concern, as discussed in chapter 3.1.3, most SMEs suffer from implementing proactive defences and creating a robust culture around cybersecurity, let alone considering system-wide effects. Therefore, the practical implications of ARA and it's uses in support of decision-making in the SME are to be discussed.

The act of constructing an interaction via conventions of ARA requires the defender to commit to a threat model as well as making explicit their informational exposure. Committing to a threat model is an act of understanding your exposure

with respect to the possible attacker types that might seek to target your organization, whereas the informational structures in terms of observable signals and voluntary disclosures on cyber maturity guide the understanding of conditional on what information possible attacker operate under. After the analysis, the key results of this thesis and of assumed most interest to an organization are to Pareto-optimal defences with respect to expected utility and investment. The case studies of this thesis focuses on single attack types, but the generalizability across different attacks in the same influence diagram is practically viable though at the cost of increased elicitation requirements. As discussed in 5.3, budget limitations can be also included in the analysis and the decision-makers of the SME can use the results as a guiding post to support decision-making by identifying the most utility-increasing defences as well as communicating important financial concepts such as return on security investment[130], aiding the common misapprehension in SME cybersecurity where the value of proactive defence is not recognized.

Moreover extending the adversarial nature of the interaction, ARA allows for the explicit inclusion of non-strategic threats assessed using conventions of probabilistic risk analysis such as done by Rios Insua et al.[8], wherein certain threats were simplified to be the resulting from a purely random source, such as is convention in probabilistic risk analysis. This inclusion supports ARA as a method for decision making since effectively SMEs can identify the most critical attack scenarios in which one assumes rational and competent adversaries, and models these events as a adversarial interaction while including in the model purely probabilistically evaluated threats that could reflect for example automated phishing campaigns. Therefore, the SMEs decision-making includes both the analysis of a rational attacker under uncertainty and the assessment of random, non-strategic threats. In essence, this produces comparable results across threat types and distinct events eliminating the complex task of comparing qualitatively different results across impact and expected utility.

In light of Table 15, the uncertainties as it relates to cybersecurity warrant discussion on how addressing these uncertainties can be systematically characterized and, where possible, quantified. The NIST CSF framework discussed in Chapter 3.2 introduces a structure for ranking an organization's cybersecurity risk governance across four tiers, shown in Figure 6. However, the standard tier definitions primarily describe the formality, consistency, and reactivity of an organization's risk management practices, without explicitly distinguishing the type of uncertainty being managed or whether the threat is treated as a strategic, adaptive adversary rather than a stochastic, nature-like process. Table 16 below presents a speculative extension of the traditional CSF tier definitions by including with the conventional definition of the tier an attacker-aware, uncertainty dimension that makes explicit how the threat space is viewed and what uncertainties are addressed.

Table 16: CSF tier structure complemented with strategic uncertainty.

Tier	Standard CSF Definition	Strategic Uncertainty Posture
Partial	Ad hoc, reactive risk management; limited organizational awareness of cyber risk; little external collaboration.	No structured uncertainty treatment; threat implicitly static; low risk perception.
Risk Informed	Practices approved by management but not fully institutionalized org-wide; risk awareness exists at leadership level.	Aleatoric uncertainty via qualitative/quantitative risk analysis; adversary modelled as nature, not strategic agent.
Repeatable	Formally approved, org-wide policy; consistent, regularly updated risk methods; trained personnel; information sharing.	Bayesian updating of threat beliefs; adversary modelled as rational/strategic; epistemic uncertainty addressed.
Adaptive	Continuous improvement driven by lessons learned and predictive indicators; cybersecurity embedded in organizational culture; real-time collaboration and information sharing.	Aleatoric, epistemic, and concept uncertainty all addressed; adaptive optimization against an evolving threat-space

This table is illustrative in nature, meaning that the CSF tier structure is not necessarily the only pathway through which this novel dimension of strategic uncertainty could be incorporated. Rather, it highlights how making explicit what is known and unknown, and acting accordingly, can be embedded within the decision-making process that shapes an organization's overall approach to cybersecurity. The SMEs of this thesis, using the structure of Table 16, are of tier Repeatable, as the ARA model explicitly allows for Bayesian updating while also incorporating epistemic uncertainty regarding a rational attacker and the aleatoric uncertainty of outcome-conditional choices made by the opponent in the interaction. However, as discussed in Section 6.3.1, concept uncertainty was left as a conscious limitation.

In conclusion, the single-analyst framework of ARA in conjunction with the identified systemic and informational dynamics mean that locally rational ARA-derived defences may not translate to system-wide security nor should the informational choices of interconnected defenders be made in isolation. The practical interpretations and implications discussed here pertain to the single-attack-type analysis conducted in this thesis, though in practice this could be generalized across multiple,

qualitatively different attack scenarios within the same problem structure through extensive re-elicitation. However, given the budget constraints typically faced by SMEs, the adoption of ARA and the ability in SMEs to elicit realistic evaluations may limit the methodology's practical feasibility, thereby limiting its potential use cases to conceptual improvements in understanding informational structure, resilience assessment, and possible large-scale, systematic risk-displacement studies outside of the single analyst perspective.

7 Conclusions

7.1 Summary

This thesis seeks to improve SME cybersecurity practices in light of an accelerating cyber threat landscape that has widened the gap relative to the general vigilance and risk attitude of many SMEs. This landscape was recognized as requiring support from the perspective of resilience, the high-level principles of which were often deemed too obfuscated to directly aid SMEs in becoming more resilient to potential attacks. Motivated by this gap between principle and practice, Bayesian decision theory was identified as a natural formal language through which resilience could be operationalized, and Adversarial Risk Analysis specifically was identified as a novel, decision-analytical alternative to traditional risk analysis for resilience planning and assessment.

In light of these motivating factors, the use of ARA was highlighted via applications to case studies in which two SMEs were targeted by an attack of industry-specific relevance with the goal of using ARA to produce the optimal defences with respect to the level of protection and investment. These results were then interpreted for their relevance to resilience planning and generalizability, along with the implications and limitations of applying ARA in this context. To further make explicit the findings and limitations of this thesis, the research questions guiding this thesis are revisited below, followed by a concluding discussion for future work.

7.2 Research questions revisited

RQ1:*How does formalizing the attacker-defender interaction through adversarial risk analysis support cyber resilience planning under uncertainty for SMEs?*

ARA supports resilience planning since it forces explicit identification and precise characterization of the sources of uncertainty in the threat environment, and this clarity is itself what drives improved resilience. By representing the problem through influence diagrams, and an assumption of a rational, utility-maximizing attacker, the framework makes the systemic structure of the interaction visible rather than leaving it implicit, as is often the case in traditional approaches. In addition, expected utility serves as a meaningful resilience metric in this context, as the identification of the most utility increasing defences is an action of determining a minimal viable

defence, a goal of many budget-constrained SMEs. However, the case studies of this thesis do not directly aid resilience planning in general, as genuine preparedness requires resilience against multiple, heterogeneous attack types. While the conceptual improvements to resilience via ARA as discussed in Section 6.2 hold true across interactions, incorporating multiple attack types into the analysis would further improve the tangibility of the results in the context of resilience planning.

RQ2:*How can adversarial risk analysis be used to model attacker decision-making and structure the information environment of cyber threats facing SMEs?*

ARA provides a structured way of building an understanding of how attackers actually behave, by modelling their decision-making as the outcome of an endogenous, rational, utility-maximizing process. This formalization forces the analyst to make explicit the attacker's incentives, constraints, and sources of uncertainty, rather than leaving them as implicit assumptions embedded in the defender's own judgement. Central to this modelling is the treatment of observable signals: since informational asymmetry is intrinsic to cyberattacks, the informational structure available to the attacker becomes a direct input to the analysis, allowing the effect of shared or exposed information to be captured explicitly rather than left unaccounted for. In this sense, ARA can serve as a tool for reflecting on and communicating an SME's state of informational exposure. ARA could, in addition, conceptually convert an SME's exposure into return-on-security-investment terms, guiding decision-making by estimating the possible losses given the organization's current exposure and observable signals.

RQ3:*Does the use of adversarial risk analysis improve upon traditional risk analysis methods in producing optimal defences?*

Traditional risk analysis is largely adequate and forms the base layer of resilience alongside compliance-driven frameworks such as NIS2. However, genuine resilience benefits specifically from the explicit treatment of uncertainty that ARA provides. Considering the sources of uncertainties as well as evaluating effectively resource allocation while coping with such uncertainties supports *and* highlights the importance of resilient design principles such as redundancies and Defence-in-Depth. In other words, basic risk analysis is sufficient in justifying basic defences, a conclusion ARA itself confirms, since some controls (e.g., MFA) emerge as clearly dominant defences. Beyond that baseline, however, for potentially critical or adversarial scenarios, ARA is able to model the attacker-defender interaction explicitly and use that interaction to guide defence selection. In addition, the fundamental Bayesian nature of the paradigm functions in tandem with, and provides support for quantitative risk analysis as it produces an estimate of $\hat{p}_D(a|d)$ grounded in the behaviour of an rational, utility-maximizing attacker, which can serve as a baseline or point of comparison for further analysis. However, the practical adoption of ARA in industry remains an open question with it's relatively high level of elicitation burdens when compared to traditional risk analysis.

7.3 Future work

To address the limitations identified earlier and to conclude the thesis, some possible directions for future research were identified for the general methodology of ARA and it's implementation in the context of cybersecurity as well as the use of this union in resilience assessment and planning.

Firstly, the Defend-Attack-Defend structure of the ransomware attack interaction should be built upon to account for the decisions made during the attack, rather than simplifying the structure of the decision past the initial attack decision as a stochastic process. This could be done either by deploying the Cyber Kill Chain [131] high-level attack phase structure, or by using MITRE ATT&CK [27] to decompose certain tactics into techniques and sub-techniques. This could in principle create valuable insights into the dominant attack strategies, and thereby, the most utility-increasing defences, but only provided realistic assessments of priors and utilities of said techniques. As a logical next step, future research could begin by assessing how, within a certain tactic, the different techniques could be evaluated qualitatively, and what attributes formulate the cost structure for the attacker, and therefore guide decision making.

To complement this more complex problem scenario, research should be applied to the paradigm of partial information in ARA as researched by Roponen et al.[9]. Fundamentally, the key interpretation of this study is that one can identify non-dominated attack scenarios and defence options without eliciting the utilities and conditional probabilities of the players under certain assumptions such as increasing positive utility function. This paradigm structure could ease the complicated task of formulating the cost structures of players where the amount of decisions to be made is large and formulating the problem becomes difficult, thereby simplifying the feasible actions by dominance-pruning. This notion of partial information applied to cybersecurity was also pointed out in their study as weighting the probabilities of getting caught versus potential gains is doable given coarse approximations, but given a more complex interaction structure, this becomes infeasible to elicit or requires a tough-to-justify simplification showcasing the trade-off between traceability and complexity intrinsic to ARA. The Pareto-dominance used with respect to expected utility and monetary investment done in this thesis for the defender was done with the goal of aiding investment decisions, but future work in the context of large attacker decision spaces should focus upon the what attributes effect the decision-making of cybercriminals[132], and thereby simplifying the attacker decision space by conventions of Pareto-optimality.

If we assume that we have built an ARA model using the structure of MITRE ATT&CK and have successfully populated the nodes of the interaction with corresponding utilities and conditional probabilities, the joint decision space of such a game is quite large where brute force enumeration via Monte Carlo can cause excessive computational time. Such a more complex game structure of many phases

and decisions as motivated by MITRE ATT&CK could thus necessitate the implementation of more sophisticated computational solutions. This has been implemented in literature by Ekin et al.[133] and referred to as augmented probability simulation(APS) via Markov Chain Monte Carlo (MCMC), where instead of computing the expected utilities for each candidate decision the method converts the estimation and optimization into a single task, optimizing the compute process. A further avenue is to identify the conditions under which APS is preferable to dominance pruning, and vice versa. APS remains computationally tractable even for large decision spaces, but still requires the classical ARA elicitation of numerical attacker utilities and probabilities, whereas dominance pruning sidesteps that elicitation burden at the cost of depending on realistic, well-defined technique attributes being available to construct the partial order in the first place.

Given the Pareto-optimal solutions of the interactions discussed in this thesis, a further metric in validation could be cyber risk economics models, mainly Gordon-Loeb[39] or Factor Analysis of Information Risk[40] as introduced in Section 3.1.1. As the general statement of the Gordon-Loeb model guides an optimal fraction of the expected loss to be invested towards cybersecurity, the metric could fundamentally be used in the defence portfolio assessment phase as a point of comparison. Systematic divergence around the Gordon-Loeb bound could indicate the lack of strategic evaluation of the model or highlight miss-specified conditional probabilities or utilities, fundamentally assisting in sensitivity analysis. Overall, cyber risk economics models could simply be used as validation metrics and in support of the design and implementation phase of ARA.

Given an ARA model that is built well against a certain tactic or across different temporal phases of a cyberattack, future research should entail repeated play within such a structure to identify how robustly such models behave when given information on novel attacks, as Bayesian updating is inherently an act of resilience planning. As discussed in 6.3, ARA is fundamentally not of use when evaluating completely novel attacks, but under uncertainty it lends itself to being used as part of red-teaming simulations, i.e. simulated adversarial exercises designed to test system defences. [134, 135]. Concretely, it could be part of the pipeline in which the results of ARA are used in the red-team exercise, thereby informing blind spots and points of improvement identified by a rational, utility-maximizing attacker, which in turn enable repeatability and iterative improvement through Bayesian updating. A contemporary example of such a use-case by Sa'ad et al. [136], where ARA is applied as the mechanism that updates beliefs about the adversary's rationality type from observed behaviour , and uses that belief to guide risk-averse moving target defence reconfiguration decisions. In addition to this, the contemporary research on cyber resilience largely assumes a static adversary [137, 138]. By incorporating ARA and assuming adaptive and strategic adversarial behaviour, the realism of such models could feasibly be improved.

To conclude, as the adoption of ARA in industry remains a matter for future work,

and its direct use by SMEs may be infeasible given their limited expertise, applying ARA to larger-scale interactions involving multiple defenders — studying the effects of system-scale risk[128] — could be an insightful path for research into large-scale systems and the prevention of catastrophic failures, such as cascading failures across interconnected networks of enterprises. Such research could also support regulatory frameworks by uncovering how defence choices propagate across these networks, thereby guiding regulatory mandates.

A Algorithmic implementations of models

Algorithm 3 DDoS defence–Attack: Estimating $\hat{p}(a \mid s)$

1: **for each** protection level s **do**
2: **for** $i = 1, \ldots, K$ **do**
3: Sample from the defender's uncertainty over attacker type F:

$$\Big(U_A^i(c_a),\ P_A^i(c_t \mid t),\ P_A^i(t \mid A)P_A^i(m \mid l, r),\ P_A^i(l \mid a, s),\ P_A^i(r)\Big) \ \sim \ F$$

4: Solve for the attacker's best response:

$$a^{*i} = \arg\max_{a \in \mathcal{A}} \int \cdots \int U_A^i(c_a)\, P_A^i(c_t \mid t)\, P_A^i(t \mid a) \\ P_A^i(m \mid l, r)\, P_A^i(l \mid a, s)\, P_A^i(r)\, dr\, dl\, dm\, dt\, dc_t.$$

5: Approximate the attack distribution:

$$\hat{p}_D(a \mid s) \ = \ \frac{\#\{a^{*i} = a\}}{K}$$

Algorithm 4 DDoS defence–Attack: Defender's Optimal Portfolio

1: $\psi(s, i) = 0$
2: **for each** (s, i) **do**
3: **for** $j = 1, \ldots, K$ **do**
4: Generate distribution over attacks

$$\left(a^j,\ l^j,\ r^j,\ m^j\right) \sim G$$

5: Compute costs conditional on chosen defences

$$c_s^j \mid s, \quad c_i^j \mid i, \quad \gamma_i \mid i$$

6: Compute net loss negated by choice in insurance policy

$$m_{\text{net}}^j = m^j \cdot (1 - \gamma_i)$$

7: Compute defender cost

$$c_d^j = m_{\text{net}}^j + c_s^j + c_i^j$$

8: Compute expected utility

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

9: Compute

$$(\hat{s}^*,\ \hat{i}^*) = \arg \max_{(s,\, i) \in \mathcal{S} \times \mathcal{I}} \psi(s, i)$$

Algorithm 5 Ransomware Attack: Estimating $\hat{p}_D(a \mid s)$ and $\mathcal{R}(s)$

1: **for each** portfolio s **do**
2: **for** $i = 1, \ldots, K$ **do**
3: Sample from the attacker's belief distribution F:

$$\Big(U_A^i(E[c_a]),\ P_{\text{MFA}}^i(s),\ P_{\text{EDR}}^i(s),\ P_{\text{rec}}^i,\ p_{\text{caught}}^i,\ c_{\text{det}}^i,\ c_{\text{op}}^i,\ w_a^i,\ \eta^i, v_a^i\Big) \ \sim\ F$$

4: Solve for the optimal ransom demand and the attacker's expected net gain:

$$R^{*i} = \left[\frac{w_a^i v_a^i}{1-\eta^i}\right]_0^{v_a^i}, \qquad \begin{cases} \Psi_3^i = (1 - P_{\text{rec}}^i)\, R^{*i} \\ \Psi_2^i = P_{\text{EDR}}(s)^i\, \Psi_3^i + (1 - p_{\text{EDR}}^i)(-p_{\text{caught}}^i \cdot c_{\text{detect}}^i) \\ \Psi_1^i = P_{\text{MFA}}(s)^i\, \Psi_2^i \\ E[c_a^i] = \Psi_1^i - c_{\text{op}}^i \end{cases}$$

5: Decide whether to attack, comparing expected utility to inaction:

$$a^{*i} = \begin{cases} 0 & U_A\Big(E[c_a^i], k_a^i\Big) \le U_A\Big(0, k_a^i\Big) \\ 1 & \text{otherwise} \end{cases}$$

6: **if** $a^{*i} = 1$ **then**
7: store R^{*i} in $\mathcal{R}(s)$
8: Approximate the attack probability

$$\hat{p}_D(\text{a} \mid s) = \frac{\#\{a^{*i} = 1\}}{K}$$

Algorithm 6 Ransomware Attack: Defender's Optimal Portfolio

for each portfolio (s, i) **do**
  $\psi(s, i) = 0$
  for $j = 1, \ldots, K$ **do**
    Sample attack distribution over attacks $\hat{p}_D(a|s)$

$$a^j \sim \text{Bernoulli}\big(\hat{p}_D(\text{a}|s)\big)$$
$$\Big(p^j_{\text{MFA}}(s),\ p^j_{\text{EDR}}(s),\ p^j_{\text{rec}}(s),\ R^j,\ v^j_d,\ p^j_{\text{crisis}},\Big) \ \sim\ G$$

    if $a^j = 1$ **then**
      Draw $\xi^j_1, \xi^j_2, \xi^j_3 \sim \mathcal{U}(0, 1)$:

$$\theta^j = \begin{cases} \text{Aborted} & \xi^j_1 \geq p^j_{\text{MFA}}(s) \\ \text{Detected} & \xi^j_1 < p^j_{\text{MFA}}(s),\ \xi^j_2 \geq p^j_{\text{EDR}}(s) \\ \text{Backup Success} & \xi^j_1 < p^j_{\text{MFA}}(s),\ \xi^j_2 < p^j_{\text{EDR}}(s),\ \xi^j_3 < p^j_{\text{rec}}(s) \\ \text{Ransom} & \text{otherwise} \end{cases}$$

      if $\theta^j = \text{Ransom}$ **then**
        Solve for the defender's optimal reactive response:

$$d_2^{*j} = \underset{d_2 \in \{pay,\, nothing,\, CMU\}}{\text{argmin}} \begin{cases} R^j(1 - \gamma_R) & d_2^1 = pay \\ v^j_d & d_2^2 = nothing \\ C_{\text{CMU}}(1 - \gamma_{\text{rec}}) + (1 - p^j_{\text{crisis}})\, v^j_d & d_2^3 = CMU \end{cases}$$

$$c^j_d = c_{d^j_2} + c_s + c_i$$

      else
        $c^j_d = c_s + c_i$
    else
      $\theta^j = \text{No attack}, \quad c^j_d = c_s + c_i$
    $\psi(s, i) = \psi(s, i) + \dfrac{u(c^j_d)}{K}$
Compute

$$(s^\star, i^\star) = \underset{s,i \in \mathcal{S},\mathcal{I}}{\text{argmax}}\, \psi(s, i)$$

References

[1] W. E. Forum and Accenture, "Global cybersecurity outlook 2026," January 2026. Published 12 January 2026.

[2] C. R. Junior, I. Becker, and S. Johnson, "Unaware, unfunded and uneducated: A systematic review of sme cybersecurity," 2025.

[3] S. M. Alhidaifi, M. R. Asghar, and I. S. Ansari, "A survey on cyber resilience: Key strategies, research challenges, and future directions," *ACM Comput. Surv.*, vol. 56, Apr. 2024.

[4] K. Tierney and M. Bruneau, "Conceptualizing and measuring resilience: A key to disaster loss reduction," *TR News*, no. 250, 2007.

[5] I. Linkov, D. A. Eisenberg, K. Plourde, T. P. Seager, J. Allen, and A. Kott, "Resilience metrics for cyber systems," *Environment Systems and Decisions*, vol. 33, pp. 471–476, 11 2013.

[6] D. Banks, V. Gallego, R. Naveiro, and D. Ríos Insua, "Adversarial risk analysis: An overview," *WIREs Computational Statistics*, vol. 14, no. 1, p. e1530, 2022.

[7] D. R. Insua, J. Rios, and D. Banks, "Adversarial risk analysis," *Journal of the American Statistical Association*, vol. 104, no. 486, pp. 841–854, 2009.

[8] D. Rios Insua, A. Couce-Vieira, J. A. Rubio, W. Pieters, K. Labunets, and D. G. Rasines, "An adversarial risk analysis framework for cybersecurity," *Risk Analysis*, vol. 41, no. 1, pp. 16–36, 2021.

[9] J. Roponen, D. Ríos Insua, and A. Salo, "Adversarial risk analysis under partial information," *European Journal of Operational Research*, vol. 287, no. 1, pp. 306–316, 2020.

[10] J. Rios and D. Rios Insua, "Adversarial risk analysis for counterterrorism modeling," *Risk Analysis*, vol. 32, pp. 894–915, May 2012. Epub 2011 Dec 8.

[11] C. Joshi, C. Nel, J. Cano, and D. L. Polaschek, "Parole board decision-making using adversarial risk analysis," *The American Statistician*, vol. 78, no. 3, pp. 345–358, 2024.

[12] "ENISA Threat Landscape 2025 | ENISA," 11 2025.

[13] M. A. Ferrag, F. Alwahedi, A. Battah, B. Cherif, A. Mechri, and N. Tihanyi, "Generative ai and large language models for cyber security: All insights you need," 05 2024.

[14] T. Pseftelis and G. Chondrokoukis, "Cyber attack motivations: Connecting actors with event types," *Preprints*, May 2025.

[15] L. Šafár, M. Pekarčik, P. Morawiec, P. Rutecka, and M. Wieczorek-Kosmala, "Mapping cybersecurity in smes: The role of ownership and firm characteristics in the silesian region of poland," *Information*, vol. 16, no. 7, 2025.

[16] K. Rejman and M. Sihvonen, "Cybersecurity standards in critical infrastructure protection: A maturity model for finnish smes," in *Human Factors in Design, Engineering, and Computing* (T. Ahram, W. Karwowski, and J. Kalra, eds.), vol. 199 of *AHFE Open Access*, (USA), AHFE International, 2025. Proceedings of the AHFE (2025) International Conference.

[17] D. Makrushin, "The cost of launching a DDoS attack." Securelist by Kaspersky, Mar. 2017. Accessed: 2026-05-26.

[18] L. Anthony (Tony) Cox Jr, "What's wrong with risk matrices?," *Risk Analysis*, vol. 28, no. 2, pp. 497–512, 2008.

[19] A. Simpson, "Into the unknown: the need to reframe risk analysis," *Journal of Cybersecurity*, vol. 10, p. tyae022, 11 2024.

[20] I. Linkov, B. D. Trump, and W. Hynes, "Resilience strategies and approaches to contain systemic threats," 2019.

[21] S. Bianchi, M. Matteoni, K. Kim, A. M. Koniari, K. Koning, R. Costache, N. Ciobotaru, A. Luna-Navarro, Z. Peng, J. Ciurlanti, H. Shahriari, D. Mittal, E. Stavridou, A. Silva, M. Palmieri, S. Pampanin, F. Petrini, and M. Overend, "Resilience readiness levels for buildings: Establishing multi-hazard resilience metrics and rating systems," *International Journal of Disaster Risk Reduction*, vol. 128, p. 105746, 2025.

[22] D. Eisenberg, M. Bates, D. Chang, M. Convertino, J. Allen, S. Flynn, and T. Seager, "Measurable resilience for actionable policy," *Environmental science and technology*, vol. 47, 09 2013.

[23] M. Bruneau, S. Chang, R. Eguchi, G. Lee, T. O'Rourke, A. Reinhorn, M. Shinozuka, K. Tierney, W. Wallace, and D. Winterfeldt, "A framework to quantitatively assess and enhance the seismic resilience of communities," *Earthquake Spectra - EARTHQ SPECTRA*, vol. 19, 11 2003.

[24] National Institute of Standards and Technology, "The nist cybersecurity framework (csf) 2.0," NIST Cybersecurity Framework (CSWP 29) NIST.CSWP.29, National Institute of Standards and Technology, 2024. Available free of charge.

[25] E. M. Hutchins, M. J. Cloppert, and R. M. Amin, "Intelligence-driven computer network defense informed by analysis of adversary campaigns and intrusion kill chains," tech. rep., Lockheed Martin Corporation, 2011.

[26] P. N. Bahrami, A. Dehghantanha, T. Dargahi, R. M. Parizi, H. H. Javadi, *et al.*, "Cyber kill chain-based taxonomy of advanced persistent threat actors: Analogy of tactics, techniques, and procedures.," *Journal of information processing systems*, vol. 15, no. 4, p. 865, 2019.

[27] The MITRE Corporation, "MITRE ATT&CK®," 2026. Available: https://attack.mitre.org/. Accessed: 2026-07-06.

[28] S. Ö. Hacıoğlu, "Mitre att&ck® framework beginners guide," Feb. 2026.

[29] M. K. Rogers, "A two-dimensional circumplex approach to the development of a hacker taxonomy," *Digital Investigation*, vol. 3, no. 2, pp. 97–102, 2006.

[30] S. Chng, H. Y. Lu, A. Kumar, and D. Yau, "Hacker types, motivations and strategies: A comprehensive framework," *Computers in Human Behavior Reports*, vol. 5, p. 100167, 2022.

[31] D. Rios Insua, D. Banks, and J. Rios, "Modeling opponents in adversarial risk analysis," *Risk Analysis*, vol. 36, no. 4, pp. 742–755, 2016.

[32] L. Y. Hunter, C. D. Albert, and E. Garrett, "Factors that motivate state-sponsored cyberattacks," *The Cyber Defense Review*, vol. 6, no. 2, pp. 111–128, 2021.

[33] D. Craigen, N. Diakun-Thibault, and R. Purse, "Defining cybersecurity," *Technology Innovation Management Review*, vol. 4, pp. 13–21, 10/2014 2014.

[34] V. Palleti, S. Adepu, V. Mishra, and A. P. Mathur, "Cascading effects of cyber-attacks on interconnected critical infrastructure," *Cybersecurity*, vol. 4, no. 1, p. 8, 2021.

[35] J. Cawthra, M. Ekstrom, L. Lusty, J. Sexton, J. Sweetnam, and A. Townsend, "Data integrity: Detecting and responding to ransomware and other destructive events," Tech. Rep. 1800-26, National Institute of Standards and Technology, Dec. 2020.

[36] L. Cleghorn, "Network defense methodology: A comparison of defense in depth and defense in breadth," *Journal of Information Security*, vol. 4, no. 3, p. 144, 2013.

[37] F. B. Schneider, *Least Privilege and More*, pp. 253–258. New York, NY: Springer New York, 2004.

[38] S. Pawar and H. Palivela, "Need of paradigm shift in cybersecurity implementation for small and medium enterprises (smes)," *International Journal of Cybersecurity Intelligence & Cybercrime*, vol. 8, no. 1, p. 4, 2025.

[39] L. A. Gordon and M. P. Loeb, "The economics of information security investment," *ACM Trans. Inf. Syst. Secur.*, vol. 5, p. 438–457, Nov. 2002.

[40] FAIR Institute, "Factor analysis of information risk (FAIR) model: Standard artifact," Standard Artifact Version 3.0, FAIR Institute, Jan. 2025. Published January 15, 2025.

[41] European Parliament and Council of the European Union, "Directive (EU) 2022/2555 on measures for a high common level of cybersecurity across the union (NIS2 directive), article 21: Cybersecurity risk-management measures," Dec.

2022. ELI: http://data.europa.eu/eli/dir/2022/2555/2022-12-27. Accessed: 2026-04-23.

[42] "Kyberturvallisuuslaki, luku 1, 3 § [cybersecurity act, chapter 1, section 3]." Finlex, 2025. Finnish national transposition of Directive (EU) 2022/2555 (NIS2). Accessed: 2026-04-23.

[43] European Parliament and Council of the European Union, "Regulation (eu) 2024/2847 (cyber resilience act)," 2024. OJ L, 2024/2847, 20 November 2024.

[44] European Parliament and Council of the European Union, "Regulation (eu) 2016/679 (general data protection regulation)," 2016. OJ L 119, 4 May 2016; CELEX: 32016R0679.

[45] European Parliament and Council of the European Union, "Directive (eu) 2022/2557 (critical entities resilience directive)," 2022. OJ L 333, 27 December 2022; CELEX: 32022L2557.

[46] Eurostat, "Structural business statistics overview." Statistics Explained, European Commission, 2023. Data reference year: 2023. Accessed: 2026-04-23.

[47] K. Khadka and A. B. Ullah, "Human factors in cybersecurity: an interdisciplinary review and framework proposal," *International Journal of Information Security*, vol. 24, p. 119, 2025.

[48] K. A. Saban, S. Rau, and C. A. Wood, ""sme executives' perceptions and the information security preparedness model"," *Information and Computer Security*, vol. 29, no. 2, pp. 263–282, 2020.

[49] D. Wynn, Jr., W. D. Salisbury, and M. Winemiller, "Experiences and lessons learned at a small and medium-sized enterprise (sme) following two ransomware attacks," *MIS Quarterly Executive*, vol. 23, no. 4, 2024.

[50] M. Wilson, S. McDonald, D. Button, and K. McGarry, "It won't happen to me: Surveying sme attitudes to cyber-security," *Journal of Computer Information Systems*, vol. 63, no. 2, pp. 397–409, 2023.

[51] M. Cunningham, "Thinking about thinking: Exploring bias in cybersecurity with insights from cognitive science," tech. rep., Forcepoint, 2020. White paper.

[52] A. CHIDUKWANI, S. ZANDER, and P. KOUTSAKIS, "Beyond self-reporting: Uncovering the operational realities of sme cybersecurity through expert assessment," *Computers and Security*, vol. 164, p. 104839, 2026.

[53] S. Tedeschi, G. Marzi, M. Balzano, and G. Costa, "Managerial insights on investment strategy in cybersecurity: Findings from multi-country research," 2025.

[54] European Union Agency for Cybersecurity (ENISA), "Cybersecurity for SMEs: Challenges and recommendations," technical report, European Union Agency for Cybersecurity (ENISA), Athens, Greece, June 2021. Available at the EU Publications Office: https://op.europa.eu/en/publication-detail/-/publication/196d873f-00e2-11ec-8f47-01aa75ed71a1.

[55] M. F. Franco, F. Künzler, J. von der Assen, C. Feng, and B. Stiller, "Rcvar: An economic approach to estimate cyberattacks costs using data from industry reports," *Computers and Security*, vol. 139, p. 103737, 2024.

[56] J. K. Kwong and K. Pearlson, "How large companies can help small and medium-sized enterprise (SME) suppliers strengthen cybersecurity," *MIS Quarterly Executive*, vol. 23, no. 4, p. Article 4, 2024.

[57] European Parliament and Council of the European Union, "Regulation (EU) 2016/679 of the european parliament and of the council on the protection of natural persons with regard to the processing of personal data — article 83: General conditions for imposing administrative fines," 2016.

[58] P. Martin-Breen and J. M. Anderies, "Resilience: A literature review," report, The Institute of Development Studies and Partner Organisations, 2011.

[59] R. Hilger, *An Algorithmic Semantic Analysis of Cyber Security and Resilience Guidance Against Interdisciplinary Understanding of Resilience Concepts Across Time and Scale*. Doctor of philosophy (ph.d.), Colorado State University, Fort Collins, CO, 2025.

[60] F. Björck, M. Henkel, J. Stirna, and J. Zdravkovic, "Cyber resilience – fundamentals for a definition," in *New Contributions in Information Systems and Technologies* (A. Rocha, A. M. Correia, S. Costanzo, and L. P. Reis, eds.), (Cham), pp. 311–316, Springer International Publishing, 2015.

[61] I. Linkov and A. Kott, *Fundamental Concepts of Cyber Resilience: Introduction and Overview*, pp. 1–25. Cham: Springer International Publishing, 2019.

[62] H. Kaplan, *Toward an Understanding of Resilience*, pp. 17–83. 01 2002.

[63] T. Aven, "The call for a shift from risk to resilience: What does it mean?," *Risk Analysis*, vol. 39, no. 6, pp. 1196–1203, 2019.

[64] C. H. Stucky, J. A. Wymer, and M. L. Sipos, "The rumsfeld matrix: A strategic decision-making framework for nurse leaders to navigate uncertainty in volatile and dynamic health care environments," *Nurse Leader*, vol. 24, no. 2, p. 102561, 2026.

[65] G. H. Sagala and D. Őri, "Resilience and digital transformation: a small to medium enterprise resilience framework," *Management Decision*, vol. 64, pp. 126–161, 02 2026.

[66] A. Bahmanova and N. Lace, "Key factors shaping collaborative cyber resilience in smes through open innovation," *Journal of Open Innovation: Technology, Market, and Complexity*, vol. 12, no. 1, p. 100715, 2026.

[67] D. Bodeau, R. Graubart, W. Heinbockel, and E. Laderman, "Cyber resiliency engineering aid: The updated cyber resiliency engineering framework and guidance on applying cyber resiliency techniques," tech. rep., The MITRE Corporation, May 2015. MITRE Technical Report.

[68] Federal Financial Institutions Examination Council (FFIEC), "End-to-end resilience," tech. rep., FFIEC, Dec. 2012. Accessed: 2026-04-20.

[69] S. Cutter, J. Ahearn, B. Amadei, P. Crawford, E. Eide, G. Galloway, M. Goodchild, H. Kunreuther, M. Li-Vollmer, M. Schoch-Spana, S. Scrimshaw, E. Stanley, G. Whitney, and M. L. Zoback, "Disaster resilience: A national imperative," *Environment: Science and Policy for Sustainable Development*, vol. 55, pp. 25–29, 03 2013.

[70] R. P. Hilger, "From cybersecurity to cyber resilience: a systemic and scalable approach for improving resilience and adaptive capacity," *Journal of Cybersecurity*, vol. 12, p. tyag011, 01 2026.

[71] S. Sangari, E. Dallal, and M. Whitman, "Modeling under-reporting in cyber incidents," *Risks*, vol. 10, no. 11, 2022.

[72] National Cyber Security Centre, "The ncsc research problem book," 2023. Version 1.1, reviewed 10 July 2024.

[73] Y. Zhang and P. Malacaria, "Dealing with uncertainty in cybersecurity decision support," *Computers and Security*, vol. 148, p. 104153, 2025.

[74] P. Żebrowski, A. Couce-Vieira, and A. Mancuso, "A bayesian framework for the analysis and optimal mitigation of cyber threats to cyber-physical systems," *Risk Analysis*, vol. 42, no. 10, pp. 2275–2290, 2022.

[75] J. Kadane and P. Larkey, "Subjective probability and the theory of games," *Management Science*, vol. 28, pp. 113–120, 02 1982.

[76] J. Q. Smith, *Decision Influence Diagrams and Their Uses*, pp. 33–51. Dordrecht: Springer Netherlands, 1994.

[77] R. D. Shachter, "Evaluating influence diagrams," *Operations Research*, vol. 34, no. 6, pp. 871–882, 1986.

[78] R. A. Howard and J. E. Matheson, "Influence diagrams," *Decision Analysis*, vol. 2, no. 3, pp. 127–143, 2005.

[79] S. M. Olmsted, *On representing and solving decision problems*. Stanford University, 1984.

[80] H. U. Gerber and G. Pafum, "Utility functions," *North American Actuarial Journal*, vol. 2, no. 3, pp. 74–91, 1998.

[81] J. von Neumann, O. Morgenstern, and A. Rubinstein, *Theory of Games and Economic Behavior: 60th Anniversary Commemorative Edition*. Princeton University Press, 1944.

[82] D. Koller and B. Milch, "Multi-agent influence diagrams for representing and solving games," *Games and Economic Behavior*, vol. 45, no. 1, pp. 181–221, 2003. First World Congress of the Game Theory Society.

[83] A. V. Lotov and K. Miettinen, *Visualizing the Pareto Frontier*, pp. 213–243. Berlin, Heidelberg: Springer Berlin Heidelberg, 2008.

[84] R. Khezri and A. Mahmoudi, "Review on the state-of-the-art multi-objective optimisation of hybrid standalone/grid- connected energy systems," *IET Generation Transmission and Distribution*, vol. 14, p. 4285 – 4300, 09 2020.

[85] J. C. Harsanyi, "Games with incomplete information played by "bayesian" players, i–iii part i. the basic model," *Management science*, vol. 14, no. 3, pp. 159–182, 1967.

[86] J. Nash, "Non-cooperative games," *Annals of Mathematics*, vol. 54, no. 2, pp. 286–295, 1951.

[87] D. Antos and A. Pfeffer, "Representing bayesian games without a common prior," in *Proceedings of the 9th International Conference on Autonomous Agents and Multiagent Systems: volume 1-Volume 1*, pp. 1457–1458, 2010.

[88] R. Aumann and A. Brandenburger, "Epistemic conditions for nash equilibrium," *Econometrica*, vol. 63, no. 5, pp. 1161–1180, 1995.

[89] C. Gil and J. Parra-Arnau, "An adversarial-risk-analysis approach to counterterrorist online surveillance," *Sensors*, vol. 19, no. 3, 2019.

[90] J. Merrick and G. S. Parnell, "A comparative analysis of pra and intelligent adversary methods for counterterrorism risk management," *Risk Analysis*, vol. 31, no. 9, pp. 1488–1510, 2011.

[91] Y. Zhang and P. Malacaria, "Bayesian stackelberg games for cyber-security decision support," *Decision Support Systems*, vol. 148, p. 113599, 2021.

[92] J. Sevillano, D. Rios, and J. Rios, "Adversarial risk analysis: The somali pirates case," *Decision Analysis*, vol. 9, pp. 86–95, 06 2012.

[93] D. Ríos Insua, J. Cano, M. Pellot, and R. Ortega, "Multithreat multisite protection: A security case study," *European Journal of Operational Research*, vol. 252, no. 3, pp. 888–899, 2016.

[94] C. Gil, D. Rios Insua, and J. Rios, "Adversarial risk analysis for urban security resource allocation," *Risk Analysis*, vol. 36, no. 4, pp. 727–741, 2016.

[95] D. G. Rasines, R. Naveiro, D. R. Insua, and S. R. Santana, "Personalized pricing decisions through adversarial risk analysis," *International Transactions in Operational Research*, vol. 32, no. 4, pp. 2034–2060, 2025.

[96] P. G. Arce, S. Das, and D. R. Insua, "Supporting product launching decisions with adversarial risk analysis," 2025.

[97] C. Joshi, J. R. Aliaga, and D. R. Insua, "Insider threat modeling: An adversarial risk analysis approach," *IEEE Transactions on Information Forensics and Security*, vol. 16, pp. 1131–1142, 2021.

[98] J. González-Ortega, D. Ríos Insua, and J. Cano, "Adversarial risk analysis for bi-agent influence diagrams: An algorithmic approach," *European Journal of Operational Research*, vol. 273, no. 3, pp. 1085–1096, 2019.

[99] J. M. Camacho, R. Naveiro, and D. R. Insua, "Computational adversarial risk analysis for general security games," 2025.

[100] A. Hassanzadeh, A. Rasekh, S. Galelli, M. Aghashahi, R. Taormina, A. Ostfeld, and M. K. Banks, "A review of cybersecurity incidents in the water sector," *Journal of Environmental Engineering*, vol. 146, no. 5, p. 03120003, 2020.

[101] C. Douligeris and A. Mitrokotsa, "Ddos attacks and defense mechanisms: classification and state-of-the-art," *Computer Networks*, vol. 44, no. 5, pp. 643–666, 2004.

[102] B. Ricks, B. Thuraisingham, and P. Tague, "Lifting the smokescreen: Detecting underlying anomalies during a ddos attack," in *2018 IEEE International Conference on Intelligence and Security Informatics (ISI)*, pp. 130–135, 2018.

[103] DigiCert, "UltraDDoS Protect Annual Distributed Denial-of-Service Analysis: January–December 2025," annual report, DigiCert, Inc., 2025. Released as TLP:CLEAR.

[104] Arelion, "DDoS Threat Landscape Report 2025: Key Backbone Security Trends," annual report, Arelion AB, 2025. Data covers January–December 2024. Formerly Telia Carrier (AS1299).

[105] Cybersecurity for the Operational Technology Environment (CyOTE) Program, "Precursor analysis report: Cyber attack on thyssenkrupp blast furnace 2014," case study, Idaho National Laboratory, Dec. 2025. Accessed: 2026-07-27.

[106] S. Ross, *Stochastic Processes*. Wiley series in probability and mathematical statistics, Wiley, 1996.

[107] Cybersecurity and Infrastructure Security Agency (CISA), Office of the Chief Economist, "Cost of a Cyber Incident: Systematic Review and Cross-Validation," tech. rep., Cybersecurity and Infrastructure Security Agency (CISA), Oct. 2020.

[108] T. Raghavan, "Chapter 20 zero-sum two-person games," vol. 2 of *Handbook of Game Theory with Economic Applications*, pp. 735–768, Elsevier, 1994.

[109] H. Levy, "Stochastic dominance and expected utility: Survey and analysis," *Management science*, vol. 38, no. 4, pp. 555–593, 1992.

[110] Q. K. A. Mirza, M. Brown, O. Halling, L. Shand, and A. Alam, "Ransomware analysis using cyber kill chain," in *2021 8th International Conference on Future Internet of Things and Cloud (FiCloud)*, pp. 58–65, 2021.

[111] A. Mancuso, M. Compare, A. Salo, and E. Zio, "Portfolio optimization of safety measures for the prevention of time-dependent accident scenarios," *Reliability Engineering and System Safety*, vol. 190, p. 106500, 2019.

[112] P. K. Varshney and R. Kumar, "A governance-oriented cyber recovery framework for ransomware preparedness and resilience," *EDPACS*, vol. 0, no. 0, pp. 1–13, 2026.

[113] N. Spence, M. Niharika Bhardwaj, and D. P. Paul III, "Ransomware in healthcare facilities: a harbinger of the future?," *Perspectives in Health Information Management*, pp. 1–22, 2018.

[114] L. A. Meyer, S. Romero, G. Bertoli, T. Burt, A. Weinert, and J. L. Ferres, "How effective is multifactor authentication at deterring cyberattacks?," 2023.

[115] C. S. Anand and R. Shanker, "Advancing crypto ransomware with multi level extortion: A peril to critical infrastructure," in *2023 2nd International Conference for Innovation in Technology (INOCON)*, pp. 1–5, 2023.

[116] Sophos, "The state of ransomware 2025," June 2025.

[117] E. Cartwright, J. Hernandez Castro, and A. Cartwright, "To pay or not: game theoretic models of ransomware," *Journal of Cybersecurity*, vol. 5, no. 1, p. tyz009, 2019.

[118] T. Yin, A. Sarabi, and M. Liu, "Deterrence, backup, or insurance: game-theoretic modeling of ransomware," *Games*, vol. 14, no. 2, p. 20, 2023.

[119] A. Cartwright, E. Cartwright, J. MacColl, G. Mott, S. Turner, J. Sullivan, and J. R. C. Nurse, "How cyber insurance influences the ransomware payment decision: theory and evidence," *The Geneva Papers on Risk and Insurance - Issues and Practice*, vol. 48, no. 2, pp. 300–331, 2023.

[120] E. Galinkin, *Winning the Ransomware Lottery: A Game-Theoretic Approach to Preventing Ransomware Attacks*, p. 195–207. Springer International Publishing, 2021.

[121] D. R. Insua, *Sensitivity Analysis in Multi-objective Decision Making*, pp. 74–126. Berlin, Heidelberg: Springer Berlin Heidelberg, 1990.

[122] D. Makrushin, "Ddos economy and market analysis," 03 2017. Available: https://www.researchgate.net/publication/358139175_DDoS_Economy_and_Market_Analysis.

[123] D. S. Krishna Prasad and H. R. Prasanna Kumar, " A Systematic Study on Ransomware Attack: Types, Phases and Recent Variants ," in *2024 5th International Conference on Intelligent Communication Technologies and Virtual Mobile Networks (ICICV)*, (Los Alamitos, CA, USA), pp. 661–668, IEEE Computer Society, Mar. 2024.

[124] R. Flage, T. Aven, and C. L. Berner, "A comparison between a probability bounds analysis and a subjective probability approach to express epistemic uncertainties in a risk assessment context – a simple illustrative example," *Reliability Engineering and System Safety*, vol. 169, pp. 1–10, 2018.

[125] W. Wang, F. Di Maio, and E. Zio, "Adversarial risk analysis to allocate optimal defense resources for protecting cyber-physical systems from cyber attacks," *Risk Analysis*, vol. 39, pp. 2766–2785, Dec. 2019. Epub 2019 Jul 30.

[126] T. J. Holt, "subcultural evolution? examining the influence of on- and off-line experiences on deviant subcultures," *Deviant Behavior*, vol. 28, no. 2, pp. 171–198, 2007.

[127] C. Gao, Y. Wang, and X. Xiong, "A cyber deception defense method based on signal game to deal with network intrusion," *Security and Communication Networks*, vol. 2022, no. 1, p. 3949292, 2022.

[128] J. W. Welburn and A. M. Strong, "Systemic cyber risk and aggregate impacts," *Risk Analysis*, vol. 42, no. 8, pp. 1606–1622, 2022.

[129] P. Naghizadeh and M. Liu, "Closing the price of anarchy gap in the interdependent security game," 2014.

[130] European Network and Information Security Agency (ENISA), "Introduction to return on security investment," enisa report, European Network and Information Security Agency (ENISA), Dec. 2012. Available: https://www.enisa.europa.eu/sites/default/files/publications/Return%20On%20Security%20Investment.pdf. Accessed: 2026-07-06.

[131] T. Yadav and A. M. Rao, "Technical aspects of cyber kill chain," in *Security in Computing and Communications* (J. H. Abawajy, S. Mukherjea, S. M. Thampi, and A. Ruiz-Martínez, eds.), (Cham), pp. 438–452, Springer International Publishing, 2015.

[132] I. Kotenko, E. Fedorchenko, E. Novikova, and A. Jha, "Cyber attacker profiling for risk analysis based on machine learning," *Sensors*, vol. 23, no. 4, p. 2028, 2023.

[133] T. Ekin, R. Naveiro, D. Ríos Insua, and A. Torres-Barrán, "Augmented probability simulation methods for sequential games," *European Journal of Operational Research*, vol. 306, no. 1, pp. 418–430, 2023.

[134] S. Yulianto, B. Soewito, F. L. Gaol, and A. Kurniawan, "Enhancing cybersecurity resilience through advanced red-teaming exercises and mitre att&ck framework integration: A paradigm shift in cybersecurity assessment," *Cyber Security and Applications*, vol. 3, p. 100077, 2025.

[135] S. Armenia, M. Angelini, F. Nonino, G. Palombi, and M. F. Schlitzer, "A dynamic simulation approach to support the evaluation of cyber risks and security investments in smes," *Decision Support Systems*, vol. 147, p. 113580, 2021.

[136] U. Sa'ad, W. Na, N.-N. Dao, and S. Cho, "Adaptive risk analysis framework for network-level moving target defense under adversarial intelligence uncertainty," *Computers and Security*, vol. 166, p. 104890, 2026.

[137] H. Cho, J.-H. Sung, H.-J. Kang, J. Jang, and D. Shin, "Quantifying cyber resilience: A framework based on availability metrics and auc-based normalization," *Electronics*, vol. 14, no. 12, 2025.

[138] M. Luo, C. Tao, Y. Liu, S. Chen, and P. Chen, "An endogenous security-oriented framework for cyber resilience assessment in critical infrastructures," *Applied Sciences*, vol. 15, no. 15, 2025.