



# A system to calculate cyber-value-at-risk

Arnau Erola<sup>a,\*</sup>, Ioannis Agraftotis<sup>a</sup>, Jason R.C. Nurse<sup>b</sup>, Louise Axon<sup>a</sup>, Michael Goldsmith<sup>a</sup>, Sadie Creese<sup>a</sup>

<sup>a</sup> Department of Computer Science, University of Oxford, UK

<sup>b</sup> School of Computing and Institute of Cyber Security for Society, University of Kent, UK

## ARTICLE INFO

### Article history:

Received 26 January 2021

Revised 25 August 2021

Accepted 9 November 2021

Available online 14 November 2021

### Keywords:

Cybersecurity

Cyber-value-at-risk

Enterprise security

Risk controls

Cyber-harm

Online harm

Monte Carlo simulations

## ABSTRACT

In the face of increasing numbers of cyber-attacks, it is critical for organisations to understand the risk they are exposed to even after deploying security controls. This residual risk forms part of the ongoing operational environment, and must be understood and planned for if resilience is to be achieved. However, there is a lack of rigorous frameworks to help organisations reason about how their use of risk controls can change the nature of the potential losses they face, given an often changing threat landscape. To address this gap, we present a system that calculates Cyber-Value-at-Risk (CVaR) of an organisation. CVaR is a probabilistic density function for losses from cyber-incidents, for any given threats of interest and risk control practice. It can take account of varying effectiveness of controls, the consequences for risk propagation through infrastructures, and the cyber-harms that result. We demonstrate the utility of the system in a real case study by calculating the CVaR of an organisation that experienced a significant cyber-incident. We show that the system is able to produce predictions representative of the actual financial loss. The presented system can be used by insurers offering cyber products to better inform the calculation of insurance premiums, and by organisations to reason about the effects of using particular risk control setups on reducing their exposure to cyber-risk.

© 2021 The Authors. Published by Elsevier Ltd.

This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>)

## 1. Introduction

Cyber Value-at-Risk (CVaR) is designed to take account of the potential harm that can arise from cyber-threats, and the variable effectiveness of commonly-used risk controls (World Economic Forum and Deloitte, 2015). This is ultimately aimed at understanding the residual risk of organisations, the harms they may be exposed to in cyberspace, and the consequences of adopting risk controls. Obtaining an in-depth understanding of the risks that organisations face can lead to informed decisions on risk-control adoption, which can reduce the likelihood of a cyber threat occurring or improve the capability to mitigate different types of harm. Such information is critical to organisation's risk oversight and leadership functions as they plan for resilience and set risk appetite, and also to the cyber-security practitioners who need to understand the residual risk that can result from use of risk controls and a changing threat landscape. (Ipsos MORI, 2020).

In a similar vein, insurance companies that underwrite policies to help organisations share and transfer cyber-risk have a vested interest in an accurate estimation of CVaR. Unlike in the property insurance market where underwriting takes place based on coverage alone, underwriters of cyber policies examine evidence of security posture of organisations (European Insurance and Occupational Pensions Authority, EIOPA). The cornerstone for estimating premiums in cyber policies is, therefore, the effectiveness of the risk controls that organisations deploy. Here CVaR insights could inform the estimation of premiums, given the organisational behaviours around risk controls, the consequence they have for residual risk, and the potential losses organisations face as a result of being victim to cyber-attack.

The problem, however, is that risk controls typically viewed as necessary by the professional and expert community are generally not underpinned by any framework that facilitates rigorous reasoning, qualification or quantification of the benefits resulting from their deployment. This means that the real value of compliance, or variability of compliance, to risk-control standards is not well-reasoned or measurable in a scientific, unambiguous or verifiable sense.

\* Corresponding author.

E-mail addresses: [arnau.erola@cs.ox.ac.uk](mailto:arnau.erola@cs.ox.ac.uk) (A. Erola), [ioannis.agrafiotis@cs.ox.ac.uk](mailto:ioannis.agrafiotis@cs.ox.ac.uk) (I. Agraftotis), [j.r.c.nurse@kent.ac.uk](mailto:j.r.c.nurse@kent.ac.uk) (J.R.C. Nurse), [louise.axon@cs.ox.ac.uk](mailto:louise.axon@cs.ox.ac.uk) (L. Axon), [michael.goldsmith@cs.ox.ac.uk](mailto:michael.goldsmith@cs.ox.ac.uk) (M. Goldsmith), [sadie.creese@cs.ox.ac.uk](mailto:sadie.creese@cs.ox.ac.uk) (S. Creese).

## Nomenclature

$A_i \in \mathcal{A}$	Asset $A_i$ in the set of assets $\mathcal{A}$
$a_i$	Value of (replacement) asset $A_i$
$T_i \in \mathcal{T}$	Threat $T_i$ in the set of all threats $\mathcal{T}$
$t_i$	Probability of a threat $T_i$
$C_i \in \mathcal{C}$	Set of controls $C_i$ in the set of all controls $\mathcal{C}$
$c_i$	Residual risk of the controls in $C_i$ (probability of failure)
$c_{ij}$	Residual risk of the control $C_{ij} \in C_i$ (probability of failure of this specific control.) <sup>1</sup>
$H_i \subseteq \mathcal{H}$	Sub-graph $H_i$ in the graph of all harms $\mathcal{H}$
$H_{ij}$	Harm $H_{ij}$ in the harm sub-graph $H_i$
$h_i$	Set of probabilities for harms in $H_i$
$h_{ij}$	Probability of the harm $H_{ij} \in H_i$
$v_{ij}$	Cost of the harm $H_{ij} \in H_i$
$N$	Number of harms triggered in a Monte Carlo simulation run for a given harm graph $H_i$

Furthermore, gaining an accurate understanding of the CVaR of an organisation is challenging, due to the multitude of factors that may influence cyber-risk. Essential factors that underpin CVaR models are the assets of the organisation, the threat landscape, the risk controls that are in place, and the effect of these factors. Defining how the effectiveness of specific risk controls may reduce cyber-risk, as well as the likelihood and severity of certain threats occurring, is a relatively difficult task.

There is a clear need for reliable and verifiable methodologies by which the likely exposure to losses given the effect of deploying specific risk controls can be calculated. Companies are already using approaches to calculating value-at-risk (VaR) more generally; models for VaR have been published and their results reported outside the cyber domain (Hendricks, 1997). While the development of CVaR models has been considered, with the World Economic Forum (WEF) providing insights on the general components that CVaR models must consider (World Economic Forum and Deloitte, 2015), there has been no presentation of a fully-developed model, nor reporting of the results of a CVaR model.

In order to address this gap, we developed a model to calculate CVaR and applied it in a real case study to assess its viability. The aim of the CVaR model is to enable an understanding of the residual risk of organisations, the harms they may be exposed to in cyberspace and the consequences of adopting risk controls. The main contribution of this paper, therefore, is the CVaR model underpinning this reasoning. The validation case study being presented solely for illustrative purposes here, and not designed for re-use in other studies.

The remainder of this paper is structured as follows: in Section 2 we provide a review of the background and related literature, and highlight the importance of data schemas in risk-quantification models. Section 3 presents our conceptual model and the architecture of our CVaR system. The implementation of the CVaR tool based on the system architecture, alongside with the schema that guides the input of data, is presented in Section 4. Section 5 reports our application of the CVaR tool to a real case study and the results we obtained. In Section 6 we discuss limitations affecting the accurate calculation of CVaR and Section 7 concludes this paper by highlighting avenues for future research.

## 2. Related work

### 2.1. VaR and cyber VaR

The concept of Value at Risk (VaR) is well-established in the field of economics (Amaya et al., 2015; Cabedo and Moya, 2003; Duffie and Pan, 1997; Fantazzini, 2009; Guermat and Harris, 2002; Hendricks, 1997; Hoffman and Hammonds, 1994; Hull and White, 1998; Jorion, 2000; Linsmeier and Pearson, 2000; Rockafellar and Uryasev, 2000). Financial institutions use VaR as a benchmark to measure their trading portfolios and their exposure to market risk to the extent that regulators have included it as a quantitative measure for disclosing information (Linsmeier and Pearson, 2000). As a summative statistical measure, VaR provides an estimation of the maximum probable losses for a specific confidence interval. For a specific time framework  $t$  and a probability  $p$ , the VaR is the value that can be lost over time  $t$  with probability  $p$  (Duffie and Pan, 1997; Guermat and Harris, 2002).

To calculate the VaR of an asset, it is important to identify the core metrics or variables that affect the value of this asset. In finance for example, the metrics for assessing fluctuations in the value of the portfolio are identified by decomposing the portfolio into simpler concepts for which basic market risk factors are established (Linsmeier and Pearson, 2000). Once the datasets are identified, three methods are used to calculate VaR, namely historical simulations, delta-normal approach and Monte Carlo simulations (Linsmeier and Pearson, 2000).

Historical simulations rely on statistical distributions of historical data to identify past changes in the value of the asset. The delta-normal approach assumes that the risk factors have a multivariate normal distribution, therefore relying on computing correlations between these risk factors. Monte Carlo simulations are similar to historical simulations, with the difference being that the generated scenarios are based on distributions of the historical data and not on calculating  $N$  values based on the changes observed for every single data point over a period of  $N$  historical days.

In the field of finance, the risk factors for market changes are well established and the appropriate historical data is available. Therefore, efforts have focused on improving the statistical understanding of the historical data and identifying methods that capture changes best. Several parametric models are proposed to capture different situations based on assumptions about normality, serial independence, non-linearity of variables, skewness, kurtosis and volatility (Hendricks, 1997). Fantazzini explores how skewness and volatility in the historical data can lead to values that underestimate the VaR with the use of several Monte Carlo simulations (Fantazzini, 2009). Guermat and Harris (2002), examine how volatility in time-series data may be influenced when the assumption that kurtosis remains constant does not hold. They further provide Exponentially Weighted Moving Average (EWMA) and Generalised Autoregressive Conditional Heteroscedasticity (GARCH) models to address such scenarios. In a similar vein, Hull et al., propose a model where the user can choose between different probability distributions, accommodating cases where changes in daily values of historical data are not normally distributed (Hull and White, 1998).

In stark contrast to the field of finance, risk quantification in the cyber domain is in its infancy (Gordon et al., 2003). Despite the fact that data is gathered at exponential rates in terms of volume, variety and speed, traditionally cybersecurity risk assessments only consider the types of cyber attacks and the motives of the attackers (World Economic Forum and Deloitte, 2015). Note that most of the literature focuses on attacks (antagonistic threats), while we consider all threats (antagonistic and non-antagonistic). Other approaches utilise cyber-attack datasets, and apply methods from ac-

<sup>1</sup> Note that controls in harms and assets are usually considered in sets in our simulations, i.e., more than one control usually protects an asset or mitigates a specific harm.

tutorial science to reason about VaR (Eling and Wirfs, 2019). Attention has been drawn to the inference of losses from cyber events through examination of insurance pricing in cyber policies (Woods et al., 2019), and the lack of consensus on the security expenditure that could reduce such losses (Woods and Böhme, 2021).

These approaches remain oblivious to the assets that are targeted, yet according to the World Economic Forum (WEF) novel approaches should focus on the intersection of cyber attacks, assets at risk and motivations of attackers (World Economic Forum and Deloitte, 2015). VaR can effectively combine all these factors and successfully be used as “a proxy concept for risk exposure appeal[ing] to a wide range of industries and enterprises” (World Economic Forum and Deloitte, 2015), giving rise to the concept of Cyber VaR. WEF define Cyber VaR as a metric to estimate, given a cyber attack, the maximum amount of loss for an organisation over a period of time  $t$  with a probability  $p$ . Although this definition identifies CVaR as a single number, effectively capturing the worse case scenario, VaR has traditionally been represented as a distribution, which is useful because it allows us to consider a range rather than a single worst case by providing a full spectrum of possible losses. Henceforth, in this paper we consider CVaR as a distribution of losses which have likelihoods associated with them.

In the same WEF report, the authors deem that it is essential for a cyber-risk model to consider three key components: namely assets, profile of an attacker, and vulnerabilities. The dependencies between these components must also be considered, with the ultimate goal being a single distribution of losses. While the WEF provides insights on the general components that cyber-risk models must consider, the authors do not suggest how such a model should be designed.

Böhme et al., also argue for consideration of an additional risk factors, and suggest that controls should be considered in cyber-risk approaches (Böhme et al., 2018). Ruan (2017) concurs with this argument and emphasises the need to establish data schemas that will provide information on estimating digital assets and classifying cyber incidents. Pal et al. (2017), further argue that to provide such rich datasets the cybersecurity risk management community should seek synergies between security vendors, who hold information on price differentiating their clients, and insurance companies, who possess information about the security investments of such clients. Carfora et al. (2019) explore the peculiarities of cyber insurance, and exemplify these by using a database of publicly available breaches to calculate event distributions and estimate value at risk.

In Ipsos MORI (2020) a similar idea is followed to calculate the losses of a breach after it has occurred, which requires manual inputting of data, and differs from CVaR as it is not predicting losses. Findings of the study make emphasis on the importance to develop a tool to facilitate the tedious work of inputting data, and the subjectivity of data if there are no models to underpin it. Similarly, Uuganbayar et al. (2021) looks at which controls would benefit the most to the exposure risk, considering controls costs and effectiveness to be independent, and assuming cost of losses and risks are calculated beforehand.

In the field of Cybersecurity Risk Assessment (CRA), some proposals also consider assets, threats, and controls to calculate residual risk. Cherdantseva et al. (2016) provides a comprehensive overview of several risk assessment methods, some of them using attack trees that map the capabilities of the attacker to the risk faced by organisations. In Wang et al. (2020), the Factor Analysis of Information Risk<sup>2</sup> (FAIR) is extended to incorporate Bayesian networks in the quantification process allowing different distributions of losses. They also extend this approach to allow modelling the in-

teraction between attackers and defenders using process-oriented model and game theory. These approaches, however, do not fully capture the context of the risk assessment, such as the interdependencies between components of the system, such as assets, controls, and threats, like our approach does. Also, CRA methods are rarely validated against real case studies, which is another contribution of our research.

## 2.2. Scope of existing third party data available

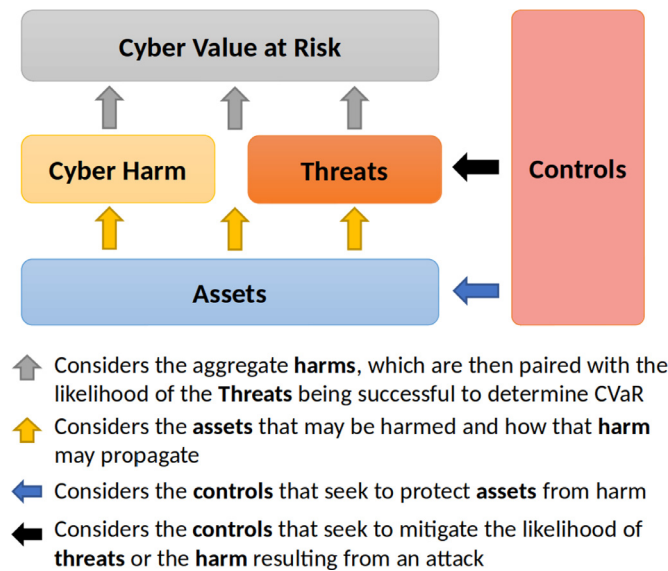
The main obstacles, identified in the literature, to designing models to estimate cyber risk are the challenges involved in identifying risk factors in cybersecurity, and the scarce presence of reliable and relevant data for these risk factors (Moore et al., 2019; World Economic Forum and Deloitte, 2015). Unlike the field of finance where data regarding risk factors is in abundance, in cybersecurity such data is scarce or does not exist yet. There have been endeavours to capture cyber incidents and their impact, such as the VERIS community's work (Veris, 2019) and the Advisen loss dataset (Advisen, 2019) (see also Section 4.2). There are, however, difficulties in capturing data due to delays between occurrences of the threats and their detection, and the complex dependencies between the event and types of risk factors (Schatz and Bashroush, 2018). Furthermore, data regarding assets and how these depend on IT systems, as well as data describing the effectiveness of controls is not captured by organisations in a systematic manner (PwC, 2019; Zurkus, 2018).

Our review of the related work in the field of data specifications for Value-at-Risk and insurance identified two data schemas, namely those from Risk Management Solutions (RMS) RMS and Cambridge Centre for Risk Studies (2016) and Applied Insurance Research (AIR) AIR (2016), which are similar in purpose to a data schema that can be used as input in our CVaR system. The fundamental difference is that both of these schemas only consider the perspective of an insurer, whereas our system focuses on understanding organisations in detail, by linking assets, controls, cyber harms and threats, thus providing a much richer capture. This capture will be crucial to the calculation of appropriate CVaR values and to enhanced modelling.

The AIR schema provides options for assets, considers the transference of data and presents quality criteria which are similar in nature to the notion of controls in our model. Specific emphasis is placed on the type of cyber insurance offered and a field is provided to link specific data to transfers. The RMS schema comprises six different fields, which mainly focus on describing the type of the company being insured, the insurance product and the potential loss it covers against. A field titled “cyber risk” attributes attempts to gather information regarding assets, with particular focus on data and revenue generated online, as well as generic information on controls/standards that companies apply/adhere to; the authors, however, do not provide sufficient information regarding the values this field may contain.

AIR and RMS fields revolve around assets, insurance policies and controls that are related to these policies. Neither captures descriptions of how assets are connected, how harms are linked to specific assets and, most importantly, which assets are protected by which controls, which threats target which assets and which harms can manifest when assets are targeted. If the links between these layers are not captured, all possible simulations are very abstract and provide inaccurate results. Thus, there is a need for a novel schema that will provide utility to existing applications built upon AIR and RMS, whilst also supporting the calculation of CVaR distributions. To address this gap, we proceed in defining a conceptual model for CVaR, a system architecture for a CVaR tool based on Monte Carlo simulations and a schema to incorporate data into the system.

<sup>2</sup> <https://www.fairinstitute.org/>



**Fig. 1.** A summary of the theoretical model displaying the association between assets, threats, harms, controls and CVaR.

### 3. A model for estimating cyber-value-at-risk

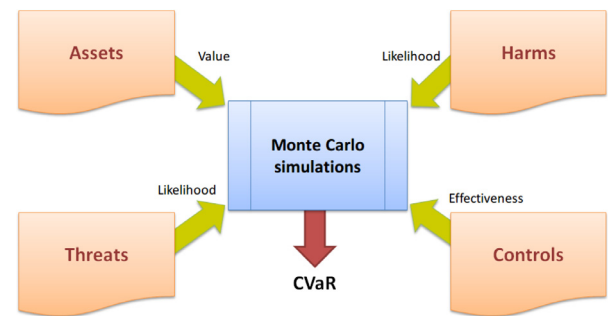
#### 3.1. Conceptual model for CVaR

Our conceptual model is presented in Fig. 1. This CVaR Model considers four different areas that are important in determining the fifth: CVaR, which is a probabilistic density function for losses from cyber incidents. These four areas are: assets, harms (types of losses from cyber threats, also known here as cyber harms), controls and threats. The model also provides links across and between its four areas. Organisations can declare how different assets (technical or otherwise) are linked together, their dependencies, what controls protect these assets, as well as what assets may give rise to what harms. Organisations can further define harms, express magnitude of losses and link harms to other harms, as well as note which controls reduce harm magnitudes. Controls can be mapped to assets or harms but can also be linked to other controls to capture dependencies between controls. Organisations can indicate which threat surfaces they intend to prevent from being realised, or mitigate the impact of realisation of, and with which controls.

Threats can be modelled based on their likelihood, the specific assets they target and the presence of threat actors with varying potential for introducing harms. At the CVaR level, the schema should allow for scenarios where more than one threat is present. Finally, and with a long-term view, the schema should capture different insurance products and link these to organisations, in an initial attempt to reason about systemic risk. Systemic risk occurs when catastrophic events target many organisations simultaneously. By capturing insurance products and linking these to organisations with specific assets, the schema allows us to consider scenarios where the same policy is triggered in multiple organisations in a short period of time. The model also provides the foundations for the design of a data schema that will be sufficiently expressive and rich to capture all of the datasets required for a CVaR calculation.

#### 3.2. System architecture

In this section we describe a support tool for CVaR calculations, its implementation and potential system configurations. Figure 2 presents an overview of the architecture of the system where



**Fig. 2.** Architecture overview.

the inputs are data files, which follow the schema described in Section 3, and the output is the CVaR distribution, which is calculated by running Monte Carlo simulations.

We opt for a sequential execution of Monte Carlo simulations throughout all levels of the model (assets, threats, controls and harms); the starting point is the assets of an organisation and the system moves gradually towards harms. This approach allows for granularity and can potentially indicate the events that lead to catastrophic scenarios. We first introduce the formalisation of these sequential calculations before discussing the values and distributions used.

##### 3.2.1. Formalisation of the system

Conceptually, the model consists of an unlabelled dependency graph over assets  $\mathcal{A}$  (such that the compromise of an asset triggers the compromise of those assets dependent upon it), and a propagation graph of harms  $\mathcal{H}$ . Over these is superimposed a set of edges (labelled with controls and residual risk probabilities, which denote the probability that a control will fail to deter or mitigate a harm), linking assets to the harms that they may provoke.<sup>3</sup> Finally, threats  $\mathcal{T}$  provide an entry point into the combined graph at asset nodes. One might consider CVaR in the face of one particular threat, or in the face of all conceivable threats.

We say CVaR for an organisation is effectively constructed by summing the CVaR for all assets, given all threats of interest.

Assets are assigned a replacement value, harms have a distribution of resulting costs, threats have a probability of success and controls a residual risk value associated with their likelihood of being effective. For a single asset and threat pair  $A_i, T_i$ , if the harm graph associated with an asset is a linear chain  $\langle H_{i_j}, h_{i_j}, v_{i_j} \mid j \in \{1, \dots, N\} \rangle$ , then the total losses associated with the asset's compromise in that particular threat, (occurring with probability  $t_i$ ) and residual risk  $c_i$  after the application of controls to that asset, (where each harm has its own probability of propagation  $h_{i_j}$  and residual risk  $c_i$  with potential loss  $v_{i_j}$ ), is its own replacement value  $a_i$  plus the sum of the harm costs down the chain, as far as the harm actually propagates. Thus, to estimate the CVaR for an asset of value  $a_i$  we need to combine information from all four concepts.

All probabilities have values between [0,1] therefore:

- $t_i \in [0, 1]$
- $c_{i_j} \in [0, 1]$
- $h_{i_j} \in [0, 1]$

Assets and harms are valued in a specific currency and get values greater than 0. For the case studies presented in Section 5 we use USD. Therefore:

<sup>3</sup> In principle there could be compromised assets as a result of harms, such as exfiltration of a password database. At present we model such scenarios in the propagation of harms and we do not attempt to link harms that have been triggered back to other assets.



- $a_i \geq 0$
- $v_{ij} \geq 0$

We obtain for every given asset  $A_i$  of value  $a_i$ , all threats  $T_i$  with likelihood of occurring  $t_i$  targeting that asset, the set of controls  $C_i$  with residual risk probabilities  $< c_i >$  that protect the asset, and a graph of harms  $H_i$  that may occur directly or indirectly if the threat is successful. These are noted within the tool using tuples of the form:  $<< A_i, a_i >, < T_i, t_i >, C_i, H_i >$ , where  $A_i$  is an asset with a replacement value  $a_i$ ;  $T_i$  is a specific threat which can occur with probability  $t_i$ ;  $C_i$  is a set of controls that are applied to asset  $A_i$ ; and  $H_i$  is the set of harms that can occur when asset  $A_i$  is targeted by threat  $T_i$ .

Our approach takes account of the variability of risk control effectiveness by abstracting away any environmental details (such as operational processes, configurations, use practices etc.) to a simple binary model whereby a control is either effective or ineffective. An effective control both prevents harm to the asset AND prevents onward harm propagation. An ineffective control does not protect the asset AND may allow onward harm propagation. Given the lack of empirical data, generating a more complex or subtle model felt too contrived at this time.

We extract each harm  $H_{ij}$  from a graph of harms that defines probabilities and values for each of the nodes and edges (i.e.,  $< H_{ij}, v_{ij}, h_{ij} >$ ), where  $H_{ij}$  is a type of harm,  $v_{ij}$  is the monetised value of this particular type of harm and  $h_{ij}$  is the probability of the corresponding harm being realised. The position of different types of harm in the graphs denotes the sequence of these types of harm as an event unfolds. It is important to note that the likelihood of a harm  $H_{ij}$  occurring is influenced by the presence of controls that may mitigate such harms.

A data schema was created for the purpose of capturing the necessary inputs, presented in Section 4. An example of such a tuple is:

```
((Email_system, $1000),
  ((Phishing, 0.0002), ((Phishing_software, 0.7))),
  ((Data_Breach, $20000, 0.02),
   (Notification, $10000, 0.6),
   (Monitoring, $25000, 0.6),
   (Regulator_Fines, $1000000, 0.001) ))
```

where a phishing attack targets an email account and can lead to a data breach. The full harm tree in this example for a data breach entails notification and monitoring costs, as well as regulatory fines.

---

**Algorithm 1** Single Monte Carlo iteration.

---

```
1: losses = 0
2: triggered = []
3: for each  $A_i \in \mathcal{A}$ 
4:   for all Tuples containing  $A_i$ 
5:     if Bernoulli( $t_i$ )
6:       ## the threat is triggered
7:       if protects( $C_i, T_i$ )
8:         ## the asset is protected
9:         continue
10:      if  $A_i \notin$  triggered
11:        losses +=  $a_i * \text{random}()$ 
12:        triggered.append( $a_i$ )
13:      losses += F_HARMS( $C_i, H_i$ )
14: return losses
```

---

latory fines.

The harm graph is typically non-linear, rendering analytic solutions difficult for a general directed-acyclic graph, as we need to take care to avoid double-counting while we sum across the

harms. To avoid this, we walk the graph marking the nodes that are reached by any propagation path that is triggered, and then again to sum the harm costs of the marked nodes, as described in Algorithm 2, which thus produces total CVaR for the organisation.

---

**Algorithm 2** Harm propagation algorithm.

---

```
1: function F_HARMS( $C_i, H_i$ )
2:   frontier ←  $H_i$ 
3:   total_harm = 0
4:   triggered = []
5:   while frontier do
6:     harm ← frontier.pop()
7:     if Bernoulli(harm.hi) then
8:       allfailed ← True
9:       for controls ∈  $C_i$  do
10:        if not Bernoulli(control.ci) then
11:          allfailed ← False
12:       if allfailed then
13:         ## the harm is triggered
14:         if not harm in triggered then
15:           total_harm += harm.v * random()
16:           triggered.append(harm)
17:           for successor ∈ harm.successors do
18:             frontier.append(successor)
19:   return total_harm
```

---

### 3.2.2. Data distributions for frequency of events

We run a number of Monte Carlo simulations which are executed sequentially for every asset in the tuples. The results of these simulations are combined to calculate the overall CVaR for a single run. This approach provides an explanation of how catastrophic events may occur for a given period of time.

More specifically, the first step of the Monte Carlo simulations entails determining if a threat occurs or not. We use the Advisen dataset (see Section 4.2) to retrieve the probabilities of the threats occurring, but this is a single number and not a distribution. Other work has used Poisson distributions (Wang and Franke, 2020) or Negative binomial distributions (Carfora et al., 2019) to determine frequency of random events. We are considering, however, Bernoulli distributions as we assume threats to be independent and we do not observe threats occurring more than once within a time period (a year) for the same organisation in the Advisen dataset. Bernoulli distributions are discrete distributions with two possible outcomes, 1 or 0, with a probability  $p$  and  $1 - p$ . This matches our requirements, as we obtain this probability from the Advisen dataset and we want to calculate the occurrence of threats in independent simulations. For example, if the probability of a threat occurring is 20% and the Bernoulli trial output is successful, the next step of the simulation is triggered.

If a threat is realised the system will move to the next level, which is the control level. Similarly to threats, Bernoulli distributions are selected to determine if the controls available can stop a threat or not. The control effectiveness is a probability distribution (in our case study we use qualitative descriptions of 'High', 'Medium' and 'Low') that will determine if the control will be effective in a specific run or not. In the case where the control is not effective, the first part of the system concludes and prepares the transition to the next level, which is the harm level. This is described in Algorithm 1, where the function *protects*(controls, threat) returns a *True* or *False* value, depending on whether any of the controls succeed in stopping the threat causing damage to the asset. The Function *Bernoulli*( $p$ ) returns *True* or *False* based on the probability  $p$ . To interpret the value  $p$ , if for example a control is 80% effective, the system produce a

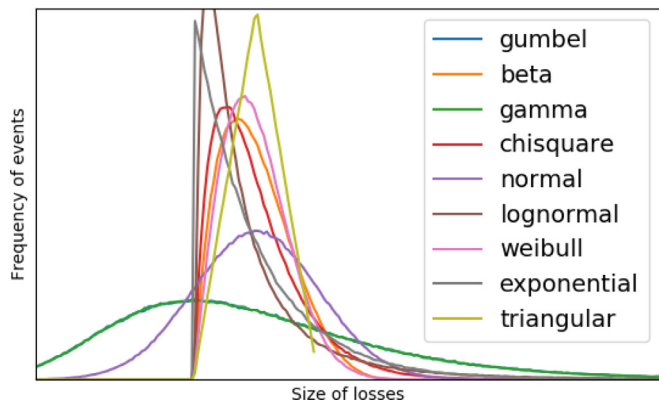


Fig. 3. Illustration of Monte Carlo simulations based on different probability distributions for choosing numbers randomly with  $\mu = 1$ .

random number between 0 and 100. If the number is greater than 80, the control is not effective in this specific run. Otherwise, the controls is effective. Therefore, the overall number of simulations, based on the big number theorem, will observe the given probability distribution for the control effectiveness as in approximately 80% of the runs as effective.

The system will move next to the harm level and run a Monte Carlo simulation with a Bernoulli distribution for the likelihood of the first harm in the graph occurring. We use Bernoulli distributions in this case as harms can only happen once for an asset and threat pair, thus having the same output than a Poisson distribution. If the harm is realised, the system will randomly select a value from a normal distribution with mean  $v_1$ , which is the value of the harm  $H_1$  and move to the next type of harm in the graph. The process is repeated and the run completes when the following step from the current state of the system is not triggered or when all harms in the graph are realised and their respective losses are estimated. In case a node is revisited the additional harm value will not be added to the overall loss.

In order to accommodate cases where the same type of harm may exist in more than one path of the harm graph, the system will keep a state of all nodes in the harm graph that are triggered. This approach ensures that the system does not add the loss from a single type of harm more than once. The overall CVaR of a run is the summation of the values that occur from the random selection of random numbers with mean values of the respective harms and assets. This part of the system is described in Algorithm 2, where *frontier* contains all nodes of the graph and the realisation of a harm (probability function) depends, amongst other things, on the effectiveness of controls that organisations deploy to mitigate such harms.

### 3.2.3. Distributions for loss sizes

For the Monte Carlo simulations and the random generated values, the tool supports a number of different probability distributions: namely Beta, Chi-square, Exponential, Gamma, Gumbel, Normal (Gaussian), Log-normal, Triangular, and Weibull. The tool's operator should select the distribution that better fits the losses to estimate. Depending on the selected distribution, the CVaR distribution can be heavy tailed (high kurtosis), thus producing larger losses. Figure 3 illustrates the shapes of different distributions with mean  $\mu = 1$  over 10,000,000 samples.

Finding a distribution that closely fits losses is crucial to accurate calculations, yet a difficult task. Previous work on data breaches has used lognormal or skew-normal distributions to calculate losses finding them adequate for this task (Carfora et al., 2019; Edwards et al., 2016; Farkas et al., 2020; Franke et al., 2014). Other works, however, use different distributions with equally

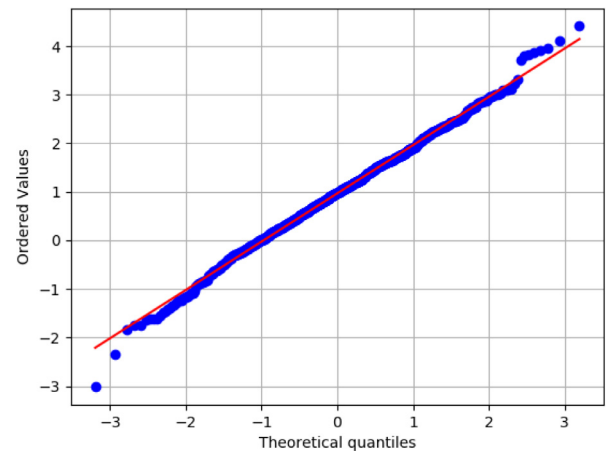


Fig. 4. Q-Q plot for 1,000 runs using a normal distribution

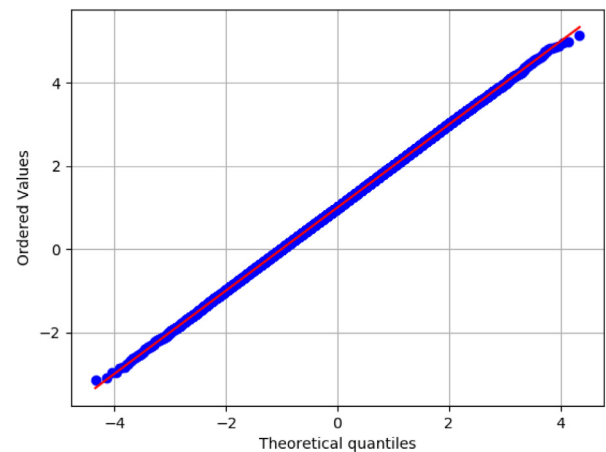


Fig. 5. Q-Q plot for 100,000 runs using a normal distribution

good results too. It is therefore clear that there is no distribution that fits all data (Woods and Böhme, 2021).

Nevertheless, our harms are not only related to data breaches, but to many other types of losses. We cannot therefore be certain that the data would fit any of the proposed distributions. Unfortunately, the data analysed from claim forms do not provide enough detail to calculate the loss distributions, so we opt to run the simulations with two different distributions: normal and lognormal. Our decision was informed by the fact that i) normal distributions are used to represent random variables whose distributions are not known (Azzalini and Valle, 1996) as well as they are more moderate when deviating from the mean compared to other distributions; ii) lognormal has been used in many studies achieving good results. This will allow us to compare the result differences just by using different distributions (see Section 6).

### 3.2.4. Estimating an optimal number of runs

It is important to identify a minimum number of Monte Carlo runs  $z$ . Given the same probability distribution, a high number of runs will provide a richer set of values to create the CVaR plot and it is more likely to produce extreme values (closely fitting the tail of CVaR) than a small number of runs. However, by increasing the number of runs, we increase linearly the time required to execute the simulation.

Figures 4 and 5 show the Q-Q plot for a normal distribution running the simulations 1,000 and 100,000 times, respectively. For 100,000 runs, results are closer to the distribution, especially in the tail. If we check the skewness and kurtosis of the data, we obtain

0.061 and 0.122 for 1,000 runs and 0.006 and 0.012 for 100,000 runs, which indicates that the data has better symmetry and is better normally tailed. Similar results are obtained for the other distributions. We therefore opt to run simulations with at least 100,000 runs per case study.

#### 4. Tool implementation

Based on the system architecture, we have designed a CVaR tool that outputs CVaR calculations. The tool has been implemented in Python. The *input files*, contain data about *assets*, *threats*, *harms* and *controls*, they use the standard *json* library and follow our schema described in the section below, ensuring that all functionalities described in the system architecture are adhered to.

The number of Monte Carlo simulations that the system runs is determined by the number of tuples  $k$ , corresponding to the set of assets  $\mathcal{A}$  and the set of threats  $T$  affecting them. The number of simulations  $z$  is configurable in the system, although a minimum of 100,000 is recommended. Therefore, the cost of running the system is  $k \times z$ , which is linear to both variables.

As an example, for the real scenario described in Section 5, the number of tuples  $k$  is 13. To execute  $z = 1,000,000$  runs the system needs 32 s on a standard laptop and 320 s when the number of runs is increased to  $z = 10,000,000$ . Further optimisation of the code can be achieved parallelising different runs, given that all runs for all tuples are independent.

##### 4.1. A schema for the CVaR tool

Guided by the conceptual model and the architecture of our CVaR system, we define a schema, whose components are needed to perform cyber-risk calculations. The schema guides how data will be used as input to our tool. It encompasses all elements which are present in AIR and RMS described in Section 2.2, and incorporates new fields to facilitate the needs of our system architecture as described in Section 3.2, resulting in a rich and expressive schema. This additional data, however, may be difficult to obtain from organisations. For instance, challenges may arise from the amount of information that organisations may be willing to disclose (to an insurer for example), as well from the availability of appropriate metrics to estimate the effectiveness of controls and the magnitude of harms. To address this issue, not all fields in our schema are required in our tool, which can be applied with the data that other schemas require and is data that insurance companies currently collect. The more information there is available from organisations however, the more accurate the results of the simulations of our tool will be.

We should note that for every field in our schema, we provide the accepted type of a value (e.g., a *string* as a series of letters, or a list of strings to be presented as *[string]*). Functions which link layers are written in capital letters (e.g., CONTAINS). The list of fields (or simply, items of data) presented in this report seeks to be exhaustive. However, the range of values which these fields can accept should be extended and refined to capture the different contexts in which organisations operate<sup>4</sup>

At the organisational level, the data which our schema seeks to gather about organisations is as follows:

```
'organisation': {
  name: string
  revenue: int
  currency: (3 letter convention)
  industry type: string
  industry code: (NAICS Industry codes)
  country code: (2-3 letters)
  country name: string
  area code: (CRESTA area code)
  public listed: (Yes/No)
  employees: int
  premium: int
  claims: int
}
```

The necessary fields for capturing organisational assets are:

```
'asset': {
  type: (machine.laptop,
        routine.finance,
        data.pii,...)
  id: string
  description: string
  cost: int
  number of records: int
  cost unit: int
  CONTAINS: {
    people: [People] (e.g. ['user1'])
    machines: [Machine]
    software: [Software]
    routines: [Routine]
    data: [Data]
  }
  PROTECTED_BY: [Controls]
  criticality: (High, Medium, Low)
  metrics for criticality: [string]
  value : (High, Medium, Low)
}
```

We use two functions to describe relationships across assets: CONTAINS and PROTECTED\_BY. CONTAINS can be either a physical or a logical relationship. For instance, Asset A (e.g., a set of customer records) is on Asset B (e.g., a computer server). An important point about this relationship is that it is (a) directional and (b) transitive. Therefore, in the example above, it should be noted that Asset A is within Asset B, as opposed to Asset B being on or a part of Asset A. Transitivity is relevant as if Asset C (e.g., data) is within Asset D (e.g., a server) and Asset D is within Asset E (e.g., a physical server room), then Asset C is also within Asset E. CONTAINS also implies that Asset A depends on Asset B, in the example above. That means that if Asset B (computer server) is compromised, Asset A is also compromised. PROTECTED\_BY describes which controls are applied to assets. For instance, IP data is protected by two-factor authentication and appropriate respective access controls. The fields *number of records* and *cost unit* are specific for when the asset is a dataset, and they are 0 otherwise. The fields *cost* and *cost unit* are integers representing the cost in the organisation's currency.

Throughout the schema, we use values High, Medium and Low to define magnitudes of certain relevance. Given the lack of objective ways to precisely measure such qualitative events, it is preferable to describe such events with qualitative values. Nonetheless, a numerical interpretation of the qualitative values can be introduced as well. For our calculations in the case study, the qualitative values of High, Medium and Low are quantified by being replaced with 0.8, 0.5, and 0.2 respectively.

Progressing from assets, the next component of the schema that we consider is harm.

<sup>4</sup> A full description of the schema, definitions and examples can be found in The Relative Effectiveness of widely used Risk Controls and the Real Value of Compliance.

```

'harm': {
  name: string
  type: string
  magnitude: (High, Medium, Low)
  first_order: {
    costs: int
    measures: (Data that evidences
      value assigned to costs)
  }
  TRIGGERS: [Harm] (e.g. ['harm',prop])
  AMPLIFIES: [Harm]
  MITIGATED_BY: [controls]
  source: [Asset, Harm]
  PROBABLE : (High, Medium, Low)
  data for estimating likelihood:
    [link to claims]
}

```

In this section of the schema we presented four functions, namely TRIGGERS, AMPLIFIES, MITIGATED\_BY and PROBABLE. TRIGGERS may cause another harm to be realised. For instance, loss of customers' data may lead to reputational damage. AMPLIFIES may add the loss from a type of harm, which has already been realised. For instance, regulator fines as a result of a cyber-attack lead to financial losses, and these financial losses may be further exacerbated due to the loss in customers. MITIGATED\_BY refers to a harmful situation being mitigated by a control in some fashion. This definition describes the relationship between a control and a harm over a specific asset. PROBABLE defines the probability of this harm's occurrence.

Our schema to define the threats faced by the organisation is as follows:

```

'threat': {
  name: string
  type: (DDoS, Cloud Service Provider
    Failure, Data Exfiltration,
    Financial Theft, Cyber
    Extortion,...)
  motivation: (High, Medium, Low)
  threat actor: (state power,
    activist groups,
    competitor,...)
  vulnerability: (CVE-XXXX, Oday,...)
  duration: (Hours, Days,...)
  TARGETS: {
    people: [People]
    machines: [Machine]
    software: [Software]
    routines: [Routine]
    data: [Data]
  }
  CAUSES: {
    harm: [], (e.g. ['harm',High])
    threat: [] (e.g. ['threat',Medium])
  }
  first_order: (Yes,vNo),
  ADDRESSED_BY: [Control]
  criticality: (High, Medium, Low)
  PROBABLE: (High, Medium, Low)
  data for estimating likelihood:
    [string]
}

```

Here we utilised three functions, namely TARGETS, CAUSES and ADDRESSED\_BY. TARGETS aims to capture the specific asset or set of assets which the threat would affect. For instance, phishing emails target the "IT admin" of the organisation (hence the "people" element). CAUSES associates a threat to a direct harm that can occur when an asset is targeted, or another threat that can occur after. ADDRESSED\_BY seeks to define the control or set of controls that have been put in place to prevent the threat from being suc-

cessful. For instance, ransomware is partially addressed by having a back-up storage facility.

The schema for risk controls that have been put in place by the organisation is below:

```

'control': {
  standard: (CIS-20, ISO 27001,
    NIST Cybersecurity
    Framework, Cyber
    Essentials,...)
  type: string
  id: string
  description: string
  MITIGATES: {
    APPLIED: {
      asset: {
        asset-name:{
          effectiveness: {
            residual risk origins:
              [string]
            residual risk:
              (High, Medium, Low)
            dependency:
              (Other controls that
                enhance the utility of
                this security control)
          }
        }
      }
    }
  }
  AGAINST: {
    threat/harm: {
      threat/harm-name:{
        effectiveness: {
          residual risk origins:
            [string]
          residual risk:
            (High, Medium, Low)
        }
      }
    }
  }
  baseline: [string]
  baseline value: (High, Medium, Low)
  value : (High, Medium, Low)
  automated: (Yes, No)
  reports: (Yes, No)
  metrics: [string]
  overall value: (High, Medium, Low)
}

```

MITIGATES describes how the control constrains harmful situations. For instance, Control A (e.g., back-up data) may reduce the impact of an attack aiming at reducing the availability of data. This definition describes the relationship between a control and a harm over a specific asset. AGAINST maps a control to a threat it aims to eliminate. APPLIED maps a control to the asset it is applied to.

## 4.2. Using the schema

In this section we describe the data sources we used to run our experiments. The general idea about operating our tool is that data input can be automated from external feeds, although these can be not free. In this way, the only data that would need to be introduced by the underwriter when calculating the risk would be the one related to the company's assets.

The assets that are present in the schema for modelling different scenarios vary depending on the organisation. Assets are drawn from the typology below, which can be extended to incorporate any further assets that an organisation possesses:



- Machine: local server, desktop, laptop, smart phone, USB device, external storage devices, cloud server, etc.
- Data: Personally Identifiable Information (PII), payment card information, Intellectual Property (IP), commercially confidential information, trade data, etc.
- Software: MacOS version Y, Windows version X, Ubuntu version X, Office, applications, middleware, cloud services, etc.
- People: Sales employee, PA, CEO, contractor, etc.
- Routine: Finance, Sales, Dispatch, Production, Online service, Research and Development, Human Resources, etc.
- Enterprise: Reputation, Knowledge, Profitability, Financial, stability, Policy, Culture, Governance, etc.

The more details on the assets of the company the more granular our system can be. The system is also able to perform the calculations with any number of assets, and therefore we can run simulations with only one asset if needed.

Threats can be specified manually but an initial list extracted from the Advisen (2019) dataset is proposed. The Advisen dataset provides a comprehensive list of insurance claims related to cyber-incidents. From this list, we can calculate the probability of a threat occurring over a fixed period of time (see Section 3.2.2 for conversion to a distribution). The types of threats it contains are as follows:

- Cyber Extortion
- Data Exfiltration
- IT Errors
- Identity Theft/Digital Breach
- Industrial Controls & Operations
- Network/Website Disruption
- Phishing, Spoofing, Social Engineering
- Privacy - Unauthorized Contract or Disclosure, Unauthorized Data Collection
- Skimming, Physical Tampering
- Denial of Service (DDoS)/System Disruption
- Other

To capture elements for the control fields we use as a guideline the SANS Critical Security Controls (SANS Top 20) SANS (2019) because they contain information on the dependencies of security controls as well as metrics on the effectiveness of these controls. We complement SANS Top 20 with other less technical and more procedural controls (common in the NIST Cybersecurity Framework National Institute of Standards and Technology (2019) and ISO 27001 International Organization for Standardization (2019)) such as: organisational culture, information-security programme, risk assessment or business-continuity plan. Reports from underwriters who write cyber policies can be used to link controls to assets, since before signing any cyber-policy organisations provide detailed description of their infrastructure and evidence on which controls are deployed to protect these assets. Claim reports can guide us when deciding which procedural/technical controls mitigate which types of harm, since there are specific details on how threats occurred, how organisations responded to these threats and why certain losses were realised. In some cases, controls are applied to the whole organisation, such as information security program, and not to specific assets, thus we allow this mapping in our tool.

Finally, we need to take account of the likelihood that an organisation's risk controls are effective in meeting the threats they face, which means deciding for each control application the likelihood of residual risk being high, medium or low. CVaR will then convert these to our working probability estimates as referred to above (0.8, 0.5 and 0.2 respectively). Any user of CVaR could chose their own basis upon which to set this. There is scarce data available to evaluate the effectiveness of controls and organisations

often rely on experts' opinions when deciding which controls to implement. There are also datasets such as Cyence (2019) (see Section 4.2) which recently have been created, providing a comparison of cybersecurity postures between organisations based on evidence gathered from outside the perimeter of an organisation's network.

Different harms originating from different types of threats should be considered in the schema. We have created a taxonomy of the most common harms observed from data bases of incidents (Agraftotis et al., 2018); we have extended this work by identifying patterns of harm propagation in real cases. By obtaining access to claim forms from AXIS Insurance Company containing actual claims data, we investigated how harms occur, how specific types propagate and trigger new harms and the frequency of these propagations. Finally, we used the same claim reports to estimate the value for each type of harm and to design the harm propagation trees (Axon et al., 2019). Examples of harm trees can be found in Figure 8. Given the sample sizes, we were only able to extract average value of harms and probabilities of harm propagations. In Section 3.2 we discuss how we use them.

## 5. Validating the tool in a real case study

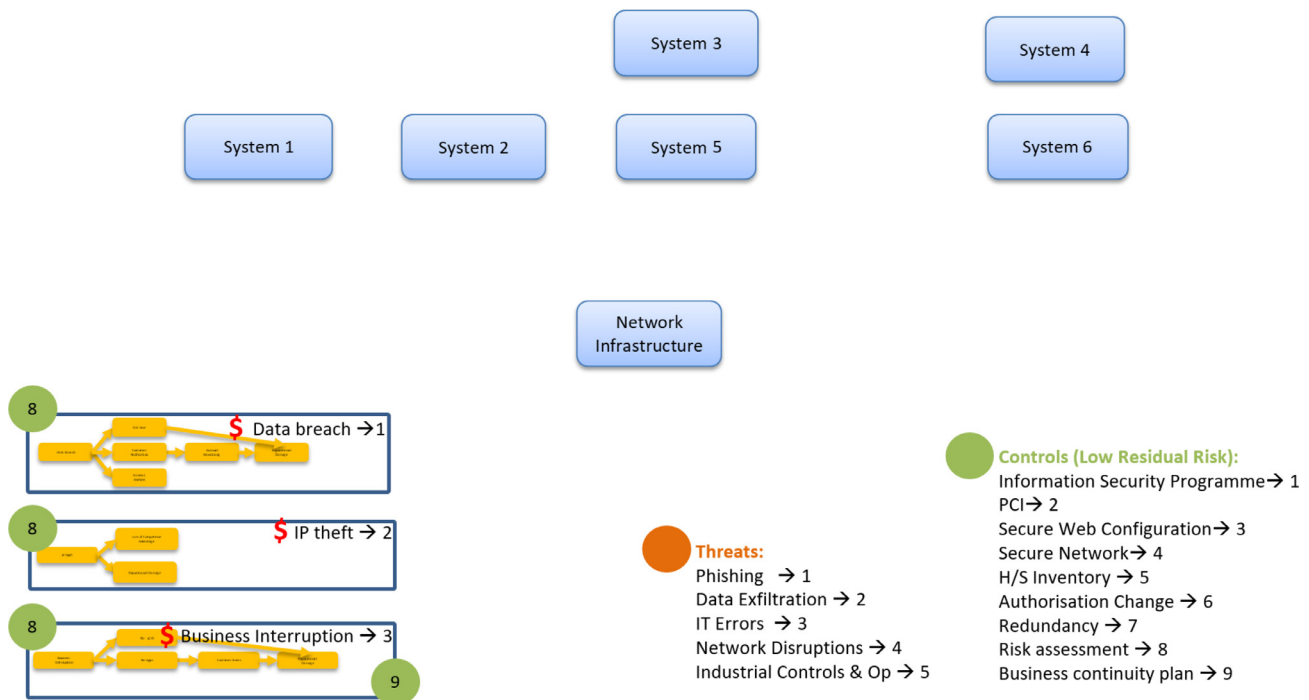
We now move to validating our tool using a real case study based on data provided by an insurance company. In collaboration with the AXIS Insurance company we examined datasets that they possess regarding claim forms, forensic reports, underwriter reports and information from tools that provide intelligence on effectiveness of controls and the threat landscape.

We decided to simulate a real case where an organisation filed a large claim. The rationale behind this decision was that for large claims detailed forensic reports are provided which may shed light on the specific sets of controls that contribute towards defending against cyber threats. Furthermore, more elaborate data exists from the underwriting screening process that can help to better capture controls, identify assets and estimate risk. Our aim is to validate the system by calculating CVaR based on information from the reports of underwriters and suggest a premium for this organisation. Our goal is not to approximate the premium that the large organisation paid but to simulate different situations and understand which controls are more important for defending against which threats. In this way, we can produce a powerful tool that underwriters can use to run multiple scenarios, create different threat landscapes and better understand the risk before suggesting a premium.

### 5.1. Description of the case study

We used the data from the screening process, as well as underwriters' notes, to understand the structure of the organisation, the value of the business processes, the systems that support these processes, and the controls that organisations have in place to protect assets and mitigate harms. We used the data from the claim forms to understand how events unfolded after the cyber incident occurred, and the data from Cyence (2019) to estimate the values for the effectiveness of the controls. Finally, we used the data from Advisen (2019) to identify the threat landscape and estimate the probabilities of these threats being realised. The output of these tools is a probability function for controls and threats respectively and these functions can be derived from other intelligence sources.

In order to construct the harm trees for the real case, we did not rely only on the claim form that the organisation filed, since it presents a single event that occurred from a specific threat. We used the harm trees presented in Agraftotis et al. (2018), and adjusted the harm values accordingly to reflect the values of the assets that the organisation held when the cyber-event happened.



**Fig. 6.** Overview of the real scenario based on data that the insurer holds. Orange circles represent threats and green circles controls that apply to specific systems. Controls 8 and 9, “Risk assessment” and “Business continuity plan”, are applied at organisational level, thus we apply them to the harm graphs to protect at operational level. Harm graphs at the bottom left corner are further represented in Fig. 8.

These harm trees were created by analysing cyber claim forms that AXIS Insurance company holds. They contain not only the sequence of events, but also variables that allow for better calculation of the values of these harms as well as averages of losses that occurred per harm type in all claim forms.

## 5.2. Data inputs and analysis of the case study

Below we present the CVaR analysis for the organisation which filed for a large claim. The candidate company suffered business interruption due to a malfunction of a network device that appeared to be a single point-of-failure. The failure of this piece of hardware occasioned a cascade failure of multiple systems and their normal operations were disrupted. The IT support team managed to contain the incident within the first hour but it took a few hours to bring back all the systems, and multiple days to recover and process stored transactions until they got all processes back to normal. The overall estimated losses are in the order of tens of millions of USD and are classified into business-interruption costs, customer loss and other expenses. For this scenario, we will anonymise all systems that the organisation possessed and refrain from providing values for assets and types of harm that may help identify the insured.

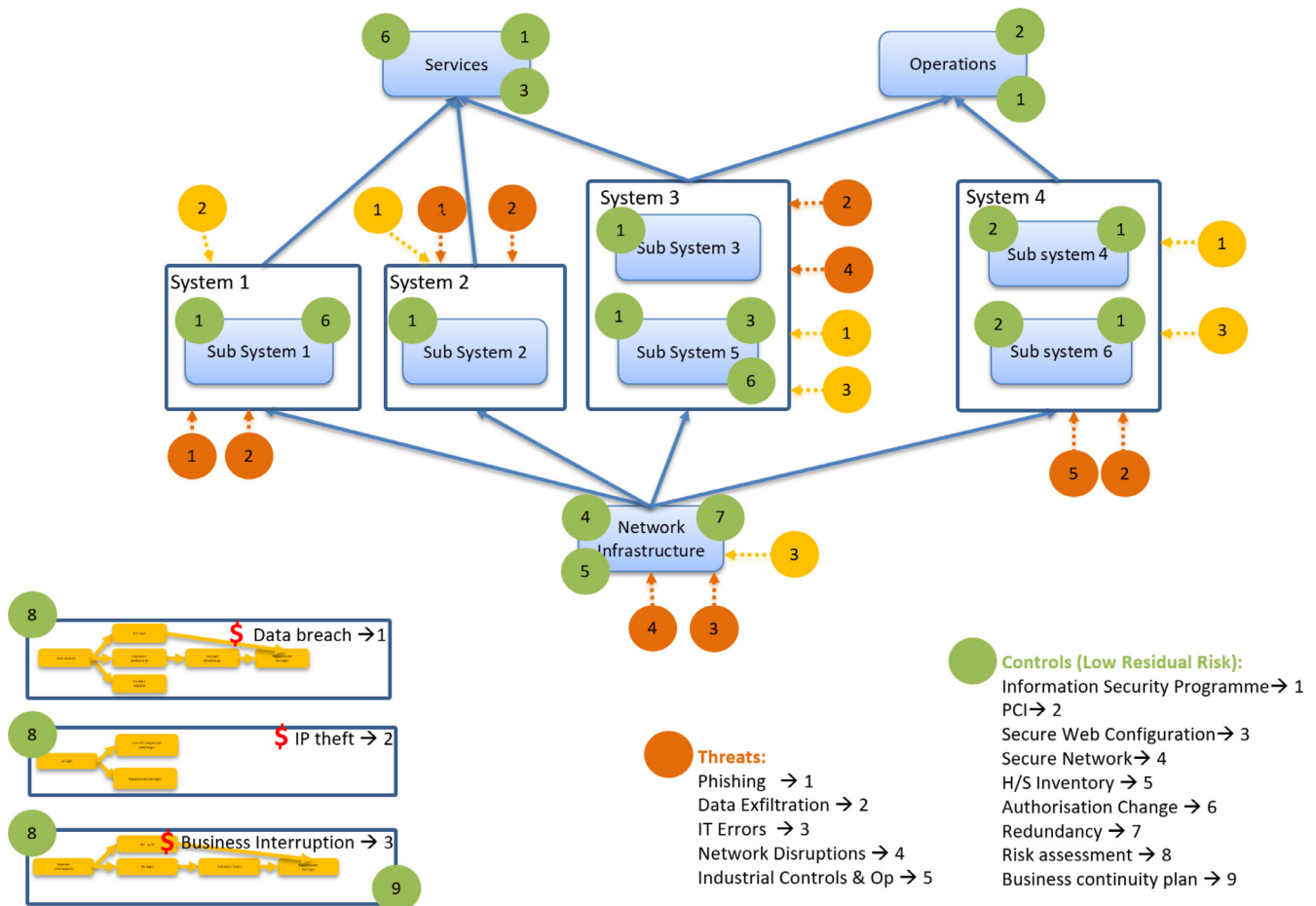
The first step was to formalise the scenario with the data schema presented in Section 3. Expressing the assets, harms, threats and controls in the format of our schema allowed us to convert the data into a JSON schema that can be used by the simulation software. We also had the opportunity to identify gaps where information required in the data schema was absent from all the data sources we had at our disposal. These gaps were mainly in the links between controls, assets and harms. Figure 6 shows how the data that the insurer holds for this organisation can be inserted into the model. This figure is created based on information from the underwriters’ report, the claims data, the Advisen dataset and information from the Cyence tool pertaining to the organisation’s effectiveness of controls (Cyence, 2019). Cyence

scores organisations for their performance in countering risk factors compared to their peer organisations. We map the Cyence risk factors to the controls the organisation in question is using; when Cyence determines them to rank higher than all peers we assume high control effectiveness; where organisations rank in the mid-range we presume medium effectiveness and low effectiveness otherwise.

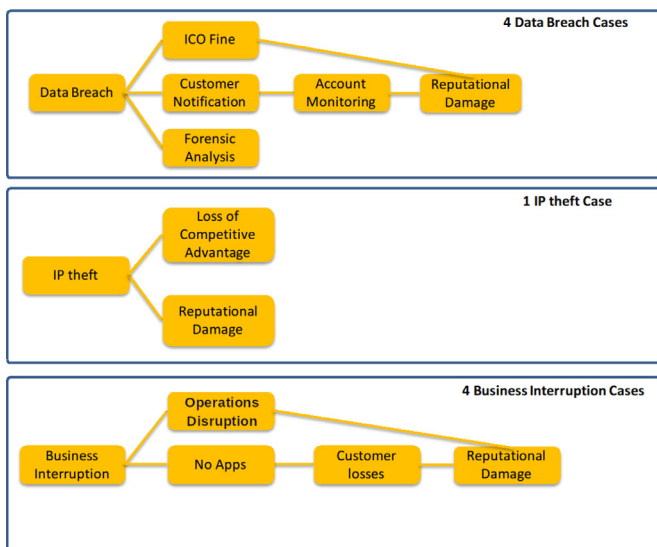
It is evident from Fig. 6 that for the tool to be more accurate in its simulations, certain assumptions need to be made. These assumptions revolve around the links between assets, harms, controls and threats. We also had to make assumptions about the system architecture of the company and the value of assets due to the high level of the overview of the assets given by the underwriters’ report. Finally, we assumed that the harmful situation described in the claim form is a worst case business interruption scenario. Our decision was informed by the fact that this was one of the largest claims filed by a company that operates in the specific sector in the last few years. Figure 7 illustrates the model we run in the simulations based on the aforementioned assumptions. A visual comparison between Figs. 6 and 7 reveals how much more data is required by the system to improve the accuracy of the model.

Regarding the harmful scenarios, we opted for three different types of harm. Our decision was informed by the assets which are present in the organisation and the evidence provided in the claim form. Figure 8 illustrates the harm trees and shows how harms may propagate for the three scenarios, namely data breach, IP theft and business interruption. In the model, we have three different instances of data breaches that can occur (loss of customer data with credit card details, loss of customer data without credit card details and loss of employee data), one instance of IP theft and two different instances of business interruption (one where the core process of the organisation is interrupted and one where it is not).

We mapped all the risk factors that Cyence provides information for, to the nine controls identified in the underwriter’s report. For the needs of this simulation we had to group the controls that the organisation had in place and map these to the risk cat-



**Fig. 7.** Overview of the real scenario with the additional data required for the model. Orange circles represent threats, green circles controls that apply to specific systems, and yellow circles represent which harms to expect when a specific asset is compromised. Controls 8 and 9, “Risk assessment” and “Business continuity plan”, are applied at organisational level, thus we apply them to the harm graphs to protect at operational level. Harm graphs at the bottom left corner are further represented in Fig. 8.



**Fig. 8.** Overview of the harm tree scenarios which are considered in our modelling based on the claim form.

egories that Cyence provided. Cyence defines scores based on how well organisations are performing when compared to their peers. Performing best amongst your peers, however, does not guarantee a good cybersecurity posture. It may well be that all organ-

isations in a specific sector have very poor configuration of controls. Therefore, the score obtained from Cyence is informative but may be misleading. Below we present the mapping between controls elicited in underwriter reports (left side of the arrow) and the Cyence risk factors (right side of the arrow):

1. Information security programme → Employee sentiment
2. PCI → Credit card exposure, Online payment present
3. Secure Web Configuration → Risky software/apps, Connected DB, Shared hosting, https misconfiguration, Website performance
4. Secure Network → Bad activity, SPF misconfiguration, Communication Device, DNS leakage, Network connectivity, Perimeter posture
5. Hardware/Software Inventory → Risky software/Apps, Technology exposure, Industrial control systems, Remediation rate, Outstanding vulnerabilities
6. Authorisation change → Dark web, Passwords, Compromised passwords, Leaked user accounts
7. Redundancy → CDN, Concurrent services
8. Risk assessment → Cybersecurity staff, Security breaches
9. Business continuity plan → Cybersecurity staff, Security breaches, Social media presence, Company stature

The majority of the risk factors from Cyence indicated that the organisation has very effective controls in place, with the exception of a “Business continuity plan” and “Risk assessment” which are medium.

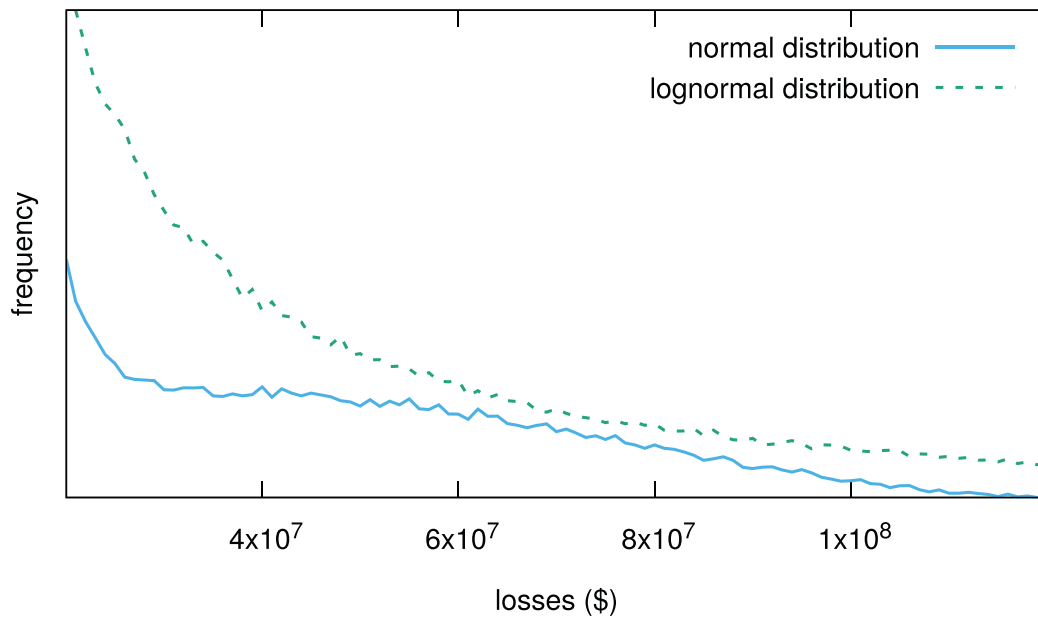


Fig. 9. CVaR estimation with control effectiveness High for controls 1-7 and Medium for 8-9, and different distributions to calculate losses.

Table 1

Estimated losses for different values of effectiveness of controls. In bold-italic, the simulation with Lognormal distribution using Cyence to inform effectiveness of controls. In italic, the result that better describes the real losses, which is when using the effectiveness of controls from Cyence and a normal distribution.

Distribution	1	2	3	4	5	6	7	8	9	95%	99%
<b>Normal</b>	Low	Low	Low	Low	Low	Low	Low	Low	Low	75M	115M
<b>Normal</b>	Medium	Medium	Medium	Medium	Medium	Medium	Medium	Medium	Medium	38M	88M
<b>Normal</b>	High	High	High	High	High	High	High	Medium	Medium	38M	87M
<b>Lognormal</b>	High	High	High	High	High	High	High	Medium	Medium	118M	378M
<b>Normal</b>	High	High	High	High	High	High	High	High	High	5M	1M

### 5.3. Analysis of results of the case study

We decided to run several instances of the model presented in Fig. 7 where we have linked controls to assets and harms and have specified which threats can target which assets. The Systems in Fig. 6 have been grouped into higher-order Systems in Fig. 7 as they are affected by the same threats, but they are still treated as independent Systems (assets) by our tool. In Fig. 9 we can see results of CVaR using normal and lognormal distributions for the losses. The tail of CVaR using lognormal distributions is much longer, producing bigger losses than the normal one. In fact, the 95 percentile for both simulations are 118 and 38 millions (Table 1), respectively, which is three times higher.

#### 5.3.1. Sensitivity of CVaR to control effectiveness

We further evaluate the sensitivity of CVaR to control effectiveness for this case study by experimenting with different configurations of the system. Table 1 shows the 95th and 99th percentiles for different values of the effectiveness of controls. As expected, the lower the effectiveness of controls, the higher the losses.

When comparing these results with the real claim, the losses occurred fall in the 99% percentile of the normal distribution. Given that it was an unusual big claim for the affected insurer, we believe our calculations match real losses. This does not mean that the affected company had no controls in place, but that they failed at some point, maybe as an effect of series of unfortunate events. The results obtained with the lognormal distribution are much larger (378 million for the 99th percentile) than the real

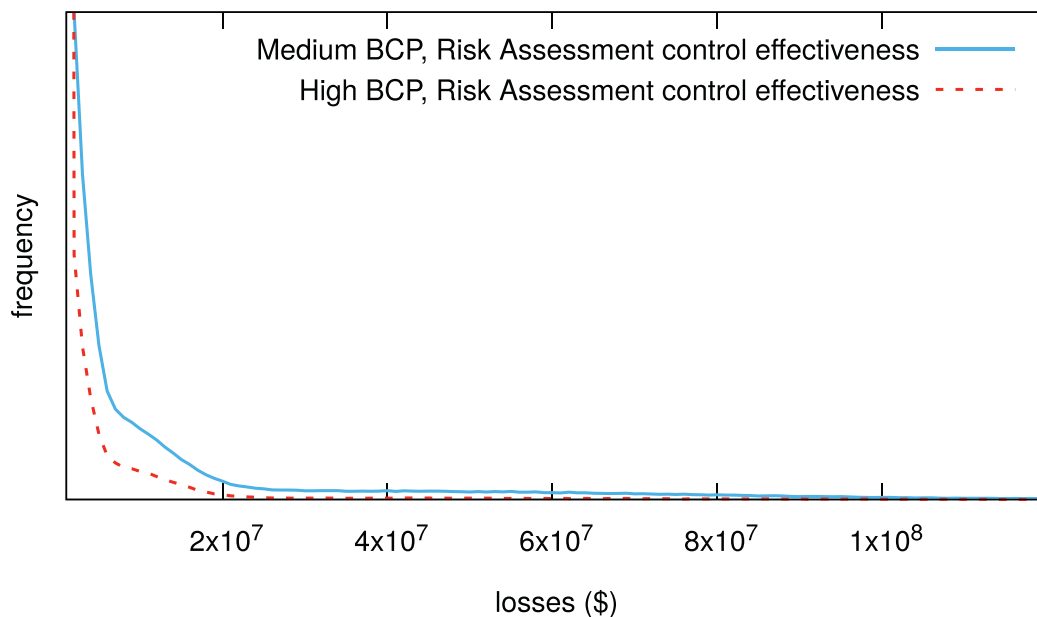
ones, thus diminishing the fitness of lognormal distribution for the CVaR calculations on this case study.

In Figure 10, we can also see that when 'Business Interruption' and 'Risk Assessment' controls have high effectiveness, the losses that occur are reduced drastically. These results indicate that the harm resulting from interruption of critical systems is disproportionate to the harms occurring when other systems are impacted or when data exfiltration occurs (even when there is an IP theft event). Given the strict regulatory framework under which this organisation operates, the lack of IP value (the company is not a manufacturer) and the need for effective and rapid customer service, these results are expected, indicating that 'Risk assessment' controls and 'Business continuity plan' are of paramount importance for this organisation.

## 6. Discussion on insights and gaps with calculating accurate CVaR

Applying the CVaR model to a real case study has validated the utility of such a tool. Our results indicate that the system is able to produce CVaR distributions representative of the actual loss values. Furthermore, we can also understand how residual risk changes in different scenarios and determine which controls have greater impact in protecting the organisation from threats and in mitigating harmful situations. The novelty of our model compared to other work (Eling and Wirfs, 2019; Pal et al., 2017; Woods et al., 2019) presented in Section 2, is that our approach is not limited to information from datasets that detail losses from cyber threats. Our





**Fig. 10.** CVaR estimation for control effectiveness High and Medium for Business continuity plan (BCP) and Risk assessment controls. All other controls have High effectiveness.

model provides an holistic view of the assets of an organization, the possible harms and how they can unfold, and which controls are deployed. The result is a more nuanced approximation of the risk that organisations have.

Such a tool can effectively be used by underwriters and risk practitioners to run multiple configurations with varying degrees of effectiveness of controls and different threat landscapes. These exercises will facilitate organisations to better capture the residual risk to which they are exposed and the range of potential losses that might arise. Gaining a deeper understanding of residual risk is fundamental to informing not only threat detection strategies, but also whether and to what degree cyber-insurance might be appropriate as a risk sharing strategy, and whether risk avoidance or transference (if at all possible) should be considered. Therefore, cyber-insurers have a powerful tool to run scenarios and calculate insurance premiums and organisations can visualise the impact of risk-control investments on their overall exposure to cyber-risk, as well as understanding general exposure.

There are some further contributions to the tool simulation that derive from our case study analysis, as it is worth making a distinction on how we utilise datasets acquired from different organisations. We believe that when processing data to determine critical assets and their value for organisations, as well as the effectiveness of controls, an organisation's perspective should be adopted (i.e., the focus is on the effectiveness of controls for the organisation whose CVaR is being estimated by the system and it is irrelevant to the effectiveness that other similar organisations have). On the other hand, when determining harm graphs and datasets for threats, aggregate data from organisations that function in the same sector should be considered (i.e., the probability of a ransomware being realised for an organisation is similar to the probability of other organisations with similar characteristics).

This exercise has also highlighted gaps in the availability of data that challenge the design of our CVaR model semantics and the analytical tool support. In our case study we had to assume the infrastructure of an organisation and how the controls apply to which assets, and link threats to specific assets. Availability of data for the links between the levels of our model is an important and necessary foundation for moving forward. Without in-

sights into organisational structures, we cannot practically experiment with datasets in order to determine the utility of the model, and the sensitivity of its outputs to data. This point is particularly worth considering, since data in this space is scarce, and therefore the ability to get useful results from coarse estimates would be a valuable feature. This analysis has also helped to complement existing research (Nurse et al., 2020) and to gain specific insights into what data must be collected by organisations and (cyber) insurance companies in order to progress the practice of risk management and predict the residual risk that organisations carry even after deploying controls.

It should be noted that our use of the Cyence and Advisen datasets is highly exploratory at this stage, and represents a creative step in our modelling which should be taken as a hypothesis as opposed to a firm proposal at this time. Datasets to enable estimation of probability distributions for threats and the effectiveness of controls are scarce and the security-practitioner community should focus their efforts on creating the appropriate datasets. To make progress in this direction, we will be exploring the potential for a pre-competitive dataset for insurance companies and an accompanying standardised format for claims. The purpose for this aspect of the discussions is to explore whether it might be possible to raise the quality and availability of data by encouraging sharing across insurers. This would not only serve to support CVaR data availability, but would likely have other commercial advantages as it could reduce costs in initial due-diligence, create a standardised approach that will facilitate better benchmarking across the sector, and still allow room for market-differentiating products to be developed based on the foundational data. Previous attempts to create this dataset, however have failed. Insurance companies usually see this data as a competitive advantage for them and as such are reticent to share it (Nurse et al., 2020).

## 7. Conclusions and next steps

We continue to face an increasing threat in cyberspace and whilst a range of controls can prevent and mitigate the harms resulting from cyber-threats, we are left with residual risks that may be realised. It is important that organisations can understand this

residual risk and the range of potential losses that might arise, since this can serve to inform not only threat detection strategies, but also alter decisions on which controls to deploy to mitigate risk and whether cyber-insurance might be appropriate as a risk sharing strategy. There is a lack of rigorous frameworks to help organisations and insurance companies to reason about the full range of losses that may result from a cyber-threat that can take account of both the use and effectiveness of risk-controls, and the potential for harm propagation across an enterprise. To address this gap, we presented a system that calculates the Cyber-Value-at-Risk (CVaR) of an organisation.

Our CVaR model takes a broad view of assets, how they are likely to be interdependent, the range of harms that might arise from a cyber-threat, the assets in scope for such a threat, the risk controls that are in use to protect the assets and their likely effectiveness, and how harm might propagate given the relationships between assets and effectiveness of risk controls. CVaR is then a probability distribution for a range of potential losses, which can either be calculated by considering all harms as in scope, or can be driven by specific threat intelligence and focused on a particular threat type.

We validated the effectiveness of the system in a real case study by calculating the CVaR for an organisation that experienced a significant cyber-incident. We showed that the system is able to produce CVaR distributions representative of the actual loss values. The results achieved the best fit using a normal distribution for losses as the final CVaR has a normal distribution shape. This could be a result of being a sum of different random variables, following the central limit theorem for sums of sets of random variables. This theorem states that if the sample size is large enough, the sum value forms a normal distribution (Wolfram, 2021). In this case, the sets are the harm values in each run.

We believe that the CVaR model development and associated analytics is now demonstrated as having utility in terms of its ability to capture organisational security postures at a level of abstraction appropriate for calculating and predicting the range of harms, and associated losses, that might arise from residual cyber-risk.

The next phase of our research will seek to improve the CVaR model's predictive capability and address the gaps that we identified in the availability of datasets required to calculate critical probability distributions for the accuracy of the tool. We will achieve this by focusing on improving the accuracy of the model in a range of dimensions, all of which will impact CVaR and therefore the predictive loss levels in any threat scenario. Our initial next steps will be as follows.

Capture the Flags (CTFs) exercises are competitions in which participants compete to complete a variety of computer-security puzzles designed to represent real-world hacking scenarios. Participants obtain *flags* as proof that they have completed a challenge. CTFs may consist of a variety of different challenges from a broad range of categories such as website security and forensics (Jeopardy-Style CTFs), or may involve teams playing against each other to attack or defend a network or server, taking flags from or planting flags on their opponent's machine/network (Attack-Defence CTFs).

The idea behind running a CTF-based experiment is that we can learn the difficulties of attackers in penetrating a system protected by different sets of risk controls (Holm et al., 2012; Moskal et al., 2018). Results on the difficulty of attacking a network protected by different sets of controls can then inform our CVaR model: the difficulty a control set creates for an attacker can be translated to a representation of that control set's effectiveness in the CVaR model.

As well as experimenting to assess the relative effectiveness of risk controls, we intend to test the sensitivity of the CVaR model to variations in the effectiveness values assigned to controls. This will

involve running the CVaR simulations while varying the presence of relevant risk controls, and the level of effectiveness assigned to the relevant controls, and observing the effect of this variation on the CVaR value.

If we can determine CVaR to be less sensitive to control performance (for specific sets of controls) then we do not need to consider deep experimentation into understanding control effectiveness for those controls, and might decide that a less granular set of estimates will suffice for our purposes. Whereas, if we determine CVaR to typically be highly sensitive to the performance of particular controls then not only will it be worthwhile developing a deeper understanding of the spectrum of performance results for CVaR, but it would also indicate justification for immediate attention to the practice of these controls by insured parties.

We anticipate that the CVaR model will vary in its sensitivity to the presence and effectiveness of the various controls, dependent on the type of control, the types of assets the control protects, and the types of threats being faced. There may be specific controls, for example, varying the presence and effectiveness value of which will have little effect on the CVaR, while for other types of control there may be a greater impact.

By analysing the extent to which the deployment of a set of relevant risk controls and their effectiveness affect the CVaR of an organisation according to our model, we hope to identify those controls of which effective deployment is critical to reducing CVaR. We can also verify whether this sensitivity of our CVaR model to control effectiveness matches the results of the control effectiveness experimentation.

## Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## CRedit authorship contribution statement

**Arnau Erola:** Conceptualization, Methodology, Formal analysis, Investigation, Writing – original draft. **Ioannis Agrafiotis:** Conceptualization, Methodology, Formal analysis, Investigation, Writing – original draft. **Jason R.C. Nurse:** Conceptualization, Methodology, Formal analysis, Investigation, Writing – original draft. **Louise Axon:** Conceptualization, Methodology, Formal analysis, Investigation, Writing – original draft. **Michael Goldsmith:** Conceptualization, Methodology, Formal analysis, Supervision. **Sadie Creese:** Conceptualization, Methodology, Formal analysis, Supervision.

## Acknowledgment

This research was sponsored by AXIS Insurance Company, whose support is gratefully acknowledged.

## References

- Advisen. 2019. URL: <https://www.advisenltd.com/data/cyber-loss-data/> [accessed 28/04/2021].
- Agrafiotis, I., Nurse, J.R.C., Goldsmith, M., Creese, S., Upton, D., 2018. A taxonomy of cyber-harms: defining the impacts of cyber-attacks and understanding how they propagate. *J. Cybersecur.* 4 (1). doi:10.1093/cybsec/tyy006.
- AIR. Verisk cyber exposure data standard and preparer's guide. 2016. URL: <http://www.air-worldwide.com/Documentation/Cyber-Exposure-Data-Standard/Index.html> [accessed 01/11/2020].
- Amaya, D., Christoffersen, P., Jacobs, K., Vasquez, A., 2015. Does realized skewness predict the cross-section of equity returns? *J. Financ. Econ.* 118 (1), 135–167.
- Axon, L., Erola, A., Agrafiotis, I., Goldsmith, M., Creese, S., 2019. Analysing cyber-insurance claims to design harm-propagation trees. In: 2019 International Conference On Cyber Situational Awareness, Data Analytics And Assessment (Cyber SA) doi:10.1109/CyberSA.2018.8551399.
- Azzalini, A., Valle, A.D., 1996. The multivariate skew-normal distribution. *Biometrika* 83 (4), 715–726.

- Böhme, R., Laube, S., Riek, M., 2018. A fundamental approach to cyber risk analysis. *Variance* 12 (2).
- Cabedo, J.D., Moya, I., 2003. Estimating oil price 'value at risk' using the historical simulation approach. *Energy Econ.* 25 (3), 239–253.
- Carfora, M.F., Martinelli, F., Mercaldo, F., Orlando, A., 2019. Cyber risk management: an actuarial point of view. *J. Oper. Risk* 14 (4), 77–103.
- Cherdantseva, Y., Burnap, P., Blyth, A., Eden, P., Jones, K., Soulsby, H., Stoddart, K., 2016. A review of cyber security risk assessment methods for SCADA systems. *Comput. Secur.* 56, 1–27.
- Cyence, 2019. URL: <https://www.guidewire.com/products/guidewire-cyence-risk-analytics> [accessed 01/11/2020].
- Duffie, D., Pan, J., 1997. An overview of value at risk. *J. Derivatives* 4 (3), 7–49.
- Edwards, B., Hofmeyr, S., Forrest, S., 2016. Hype and heavy tails: a closer look at data breaches. *J. Cybersec.* 2 (1), 3–14.
- Eling, M., Wirfs, J., 2019. What are the actual costs of cyber risk events? *Eur. J. Oper. Res.* 272 (3), 1109–1119.
- European Insurance and Occupational Pensions Authority (EIOPA), Understanding cyber insurance - a structured dialog with insurance companies. 2018. URL: <https://eiopa.europa.eu/Publications/Reports/EIOPAUnderstandingCyberinsurance.pdf> [accessed 01/11/2020].
- Fantazzini, D., 2009. The effects of misspecified marginals and copulas on computing the value at risk: a Monte Carlo study. *Comput. Stat. Data Anal.* 53 (6), 2168–2188.
- Farkas S., Lopez O., Thomas M., Cyber claim analysis through generalized Pareto regression trees with applications to insurance. Available at HAL: <https://halinriafr/hal-02118080/2020>.
- Franke, U., Holm, H., König, J., 2014. The distribution of time to recovery of enterprise IT services. *IEEE Trans. Reliab.* 63 (4), 858–867.
- Gordon, L.A., Loeb, M.P., Sohail, T., 2003. A framework for using insurance for cyber-risk management. *Commun. ACM* 46 (3), 81–85.
- Guermat, C., Harris, R.D.F., 2002. Forecasting value at risk allowing for time variation in the variance and kurtosis of portfolio returns. *Int. J. Forecast.* 18 (3), 409–419.
- Hendricks, D., 1997. Evaluation of value-at-risk models using historical data. *Econ. Policy Rev.* 2 (1).
- Hoffman, F.O., Hammonds, J.S., 1994. Propagation of uncertainty in risk assessments: the need to distinguish between uncertainty due to lack of knowledge and uncertainty due to variability. *Risk Anal.* 14 (5), 707–712.
- Holm, H., Ekstedt, M., Andersson, D., 2012. Empirical analysis of system-level vulnerability metrics through actual attacks. *IEEE Trans. Dependable Secure Comput.* 9 (6), 825–837.
- Hull, J., White, A., 1998. Value at risk when daily changes in market variables are not normally distributed. *J. Derivatives* 5, 9–19.
- International Organization for Standardization. Iso/iec 27000 family - information security management systems. 2019. URL: <https://www.iso.org/isoiec-27001-information-security.html> [accessed 01/11/2020].
- Ipsos MORI. Analysis of the full costs of cyber security breaches. 2020. URL: [https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment\\_data/file/901569/Analysis\\_of\\_the\\_full\\_cost\\_of\\_cyber\\_security\\_breaches.pdf](https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/901569/Analysis_of_the_full_cost_of_cyber_security_breaches.pdf) [accessed 15/11/2020].
- Jorion, P., 2000. *Value at Risk*. McGraw-Hill Professional Publishing.
- Linsmeier, T.J., Pearson, N.D., 2000. Value at risk. *Financ. Anal. J.* 56 (2), 47–67.
- Moore, T., Kenneally, E., Collett, M., Thapa, P., 2019. Valuing cybersecurity research datasets. In: 18th Workshop on the Economics of Information Security (WEIS).
- Moskal, S., Yang, S.J., Kuhl, M.E., 2018. Cyber threat assessment via attack scenario simulation using an integrated adversary and network modeling approach. *J. Defense Model. Simul.* 15 (1), 13–29.
- National Institute of Standards and Technology. Cybersecurity framework version 1.1. 2019. URL: <https://www.nist.gov/cyberframework/framework> [accessed 01/11/2020].
- Nurse, J.R.C., Axon, L., Erola, A., Agraftotis, I., Goldsmith, M., Creese, S., 2020. The data that drives cyber insurance: a study into the underwriting and claims processes. In: 2020 International Conference on Cyber Situational Awareness, Data Analytics and Assessment (CyberSA). IEEE, pp. 1–8.
- Pal, R., Golubchik, L., Psounis, K., Hui, P., 2017. Security pricing as enabler of cyber-insurance a first look at differentiated pricing markets. *IEEE Trans. Dependable Secure Comput.* 16 (2), 358–372.
- PwC. Cyber security insurance - how can insurers quantify the risk? 2019. URL: <https://www.pwc.co.uk/services/audit-assurance/actuarial/insights/cyber-security-insurance-how-can-insurers-quantify-risk.html#> [accessed 01/11/2020].
- RMS, Cambridge Centre for Risk Studies. Cyber insurance exposure data schema v1.0. 2016. URL: <https://www.jbs.cam.ac.uk/faculty-research/centres/centre-for-risk-studies/publications/space-and-technology/cyber-exposure-data-schema> [accessed 01/11/2020].
- Rockafellar, R.T., Uryasev, S., 2000. Optimization of conditional value-at-risk. *J. Risk* 2, 21–42.
- Ruan, K., 2017. Introducing cyberonomics: a unifying economic framework for measuring cyber risk. *Comput. Secur.* 65, 77–89.
- SANS. The CIS critical security controls for effective cyber defense. 2019. URL: <https://www.cisecurity.org/controls/> [accessed 01/11/2020].
- Schatz, D., Bashroush, R., 2018. Corporate information security investment decisions: a qualitative data analysis approach. *Int. J. Enterprise Inf. Syst. (IJEIS)* 14 (2), 1–20.
- Uuganbayar, G., Yautsiukhin, A., Martinelli, F., Massacci, F., 2021. Optimisation of cyber insurance coverage with selection of cost effective security controls. *Comput. Secur.* 101, 102121. doi:10.1016/j.cose.2020.102121. URL: <http://www.sciencedirect.com/science/article/pii/S0167404820303941>
- Veris. Veris community. 2019. URL: <http://veriscommunity.net/vcdb.html> [accessed 28/04/2021].
- Wang, J., Neil, M., Fenton, N., 2020. A Bayesian network approach for cybersecurity risk assessment implementing and extending the FAIR model. *Comput. Secur.* 89, 101659.
- Wang, S.S., Franke, U., 2020. Enterprise IT service downtime cost and risk transfer in a supply chain. *Oper. Manage. Res.* 1–15.
- Wolfram. Central limit theorem. 2021. URL: <https://mathworld.wolfram.com/CentralLimitTheorem.html> [accessed 15/04/2021].
- Woods, D., Moore, T., Simpson, A., 2019. The county fair cyber loss distribution: Drawing inferences from insurance prices. In: *Workshop on the Economics of Information Security*.
- Woods, D.W., Böhme, R., 2021. Systematization of knowledge: quantifying cyber risk. In: *IEEE Symposium on Security & Privacy*.
- World Economic Forum and Deloitte. Partnering for cyber resilience towards the quantification of cyber threats. 2015.
- Zurkus K. Lack of hardened benchmarks leads to poor cyber hygiene. 2018. URL: <https://www.infosecurity-magazine.com/news/lack-of-hardened-benchmarks-leads> [accessed 01/11/2020].



**Arnau Erola** is a Research Fellow at the Department of Computer Science of the University of Oxford, working on cyber insurance and better understanding the cyber-threat landscape. His research interests include, but are not limited to, enterprise security, defence systems and economics of cyber security. Dr Erola holds a Ph. D., M. Sc. and B.Sc. in Computer Science from the Universitat Rovira i Virgili (Tarragona). He is author of several international journal articles on online privacy, anonymity protocols and intrusion detection mechanisms.



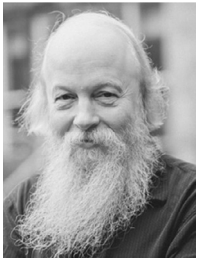
**Ioannis Agraftotis** is a Research Fellow at the Department of Computer Science and James Martin Fellow at the Global Cyber Security Capacity Centre, University of Oxford. His research interests include capacity building in cybersecurity, risk analysis and resilience in the cyber domain, cyber insurance, and anomaly detection techniques. Ioannis holds a PhD in Engineering (Warwick), a MSc in Analysis, Design and Management of Information Systems (LSE) and a BSc in Applied Informatics (University of Macedonia, Greece).



**Jason R.C. Nurse** is an Associate Professor in Cybersecurity at the University of Kent and a Visiting Academic at the University of Oxford. His research interests include organisational security and models for understanding harms resulting from cyber-attacks, and broader issues such as cybersecurity awareness and education. Dr Nurse holds a Ph.D. in Computer Science (Warwick, UK), and M.Sc. and B.Sc. degrees in Computing. He has authored numerous articles on various topics in cybersecurity, from both organisational and personal perspectives.



**Louise Axon** is a Research Associate in Cybersecurity at the University of Oxford. Her research interests include network-security monitoring and intrusion-detection approaches, cyber risk and insurance, security and privacy of distributed ledger technologies, and cybersecurity capacity building. She holds a DPhil in Cybersecurity (Oxford), an MSc in Mathematics of Cryptography and Communications (Royal Holloway) and a BA degree in Mathematics and Music (Cardiff).



**Michael Goldsmith** is Director of the Global Cyber Security Capacity Centre at the Oxford Martin School and a Senior Research Fellow in the Department of Computer Science and at Worcester College, University of Oxford. His research spans a wide range of topics within security, from the mathematical to the social. He received his DPhil in Computation from Oxford University three decades ago for work on support for specification logics, and has also worked in concurrency theory and formal verification through exhaustive state-exploration.



**Sadie Creese** is Professor of Cybersecurity in the Department of Computer Science at the University of Oxford. She was founding director of the Global Cyber Security Capacity Centre at the Oxford Martin School and a member of the Coordinating Committee for the Cyber Security Oxford network. She is engaged in a broad portfolio of cybersecurity research spanning situational awareness, visual analytics, risk propagation and communication, threat modelling and detection, network defence, dependability and resilience and privacy.