



REVIEW

# Modeling and management of cyber risk: a cross-disciplinary review

Rong He<sup>1</sup>, Zhuo Jin<sup>2</sup>  and Johnny Siu-Hang Li<sup>3</sup> 

<sup>1</sup>Department of Economics, The University of Melbourne, Melbourne, VIC, Australia; <sup>2</sup>Department of Actuarial Studies and Business Analytics, Macquarie University, Sydney, NSW, Australia; and <sup>3</sup>Department of Finance, The Chinese University of Hong Kong, Shatin, Hong Kong

**Corresponding author:** Johnny Siu-Hang Li; Email: [johnny.li@cuhk.edu.hk](mailto:johnny.li@cuhk.edu.hk)

(Received 15 December 2022; revised 05 November 2023; accepted 06 November 2023)

## Abstract

This paper provides a review of cyber risk research accomplished in different disciplines, with a primary goal to aid researchers in the field of insurance and actuarial science in identifying potential research gaps as well as leveraging useful models and techniques that have been considered in the literature. We highlight the recent advancements in cyber risk prediction, modeling, management, and insurance achieved in different domains including computer engineering, actuarial science, and business studies. The surveyed works are classified according to their respective modeling approaches, allowing readers to more easily compare the technical aspects of the surveyed works and spot out research gaps based on the research tools of their liking. We conclude this paper with a summary of possible research directions that are identified from the review.

**Keywords:** Cyber risk management; cyber risk prediction; cyber insurance; dependence modeling; machine learning

## 1. Introduction

### 1.1 Background

While the advancement in information technology brings a myriad of benefits to society, it introduces different forms of cyber risk, which may possibly lead to costly losses. In the academia, computer science researchers have been aware of the importance of security in the cyberspace since the rise of the digital era and have made a lot of technical contributions concerning risk detection (Moore *et al.*, 2006; Garcia-Teodoro *et al.*, 2009; Cárdenas *et al.*, 2011; Liu *et al.*, 2016b), security breach prediction (Zhan *et al.*, 2013; Liu *et al.*, 2015; Bakdash *et al.*, 2018), and computer system enhancement (Ballardie & Crowcroft, 1995; Jang-Jaccard & Nepal, 2014). In business studies, there is a cluster of research that investigates cyber risk in the enterprise risk management (ERM) context (Stoneburner *et al.*, 2002; Gordon *et al.*, 2003; Ögüt *et al.*, 2011; Paté-Cornell *et al.*, 2018).

In the field of insurance and actuarial science, researchers have studied cyber risk in terms of frequency, severity, and dependence modeling with a range of statistical techniques (Herath & Herath, 2011; Eling & Loperfido, 2017; Eling & Jung, 2018). However, relative to the depth and breath of the topic, cyber risk research in this field is still very much in its infantile stage. We believe that researchers in this field have the potential to accomplish a wide range of research goals concerning cyber risk, possibly by drawing on the recent developments in other domains.

This paper provides a review of cyber risk research accomplished in different disciplines, with a primary goal to aid researchers in the field of insurance and actuarial science in identifying potential research gaps as well as leveraging useful models and techniques that have been considered in the literature. Researchers in other domains may benefit from this paper through an understanding of the needs of the field of insurance and actuarial science with respect to cyber risk research. Such an understanding may facilitate cross-disciplinary research that offers innovative perspectives on cyber risk modeling and management.

Survey papers on cyber risk can be found primarily in the domains of computer science (Leau & Manickam, 2015; Ahmed & Zaman, 2017; Husák *et al.*, 2018; Sun *et al.*, 2018) and business studies (Eling & Schnell, 2016; Eling, 2020; Cremer *et al.*, 2022). Our work distinguishes from the existing survey papers in that it reviews cyber risk research performed in multiple domains to inspire cross-disciplinary research involving insurance and actuarial science. A wide range of topics, such as risk prediction, risk modeling, and risk management, are included. Our review encompasses publications from journals in various fields, including computer science and engineering, business and economics, risk management and insurance (RMI), and actuarial science, wherein computer science dominates in terms of publication outlets, followed by business and actuarial studies. Since cyber risk has been a long standing topic of interest in the world of computer science, contributions from this field are the most prominent in every aspect of cyber research, even in the actuarial-centric field of insurance. In recent years, a surge of works from actuarial journals can be observed, leading to a commensurate number of publication outlets with those from the computer science domain. The primary focus of these actuarial works revolves around mathematical and statistical modeling of cyber risk using more elaborate techniques than the classical approach.

Another distinguishing feature is that our review is written consistently in a “topic-oriented” and “model-based” manner. The existing works surveyed are sorted into different topics and further classified according to their respective modeling approaches, allowing readers to more easily compare the technical aspects of the surveyed works and identify research gaps based on the research tools of their liking. Upon reviewing, we identify the evolving trend in the research areas of interests over time. Publications in computer science and engineering adopt a technical perspective, focusing on the prediction of risk arrival and the development of risk assessment and mitigation strategies in the technical aspects. Meanwhile, RMI publications center on risk mitigation through operational controls. Risk modeling using more elaborate techniques than the standard distribution approach is a subject of interest to both the fields of actuarial science and computer science. Our findings reveal that machine learning prevails the field of cyber risk prediction in the past decade but finds less presence in the field of risk modeling. Nevertheless, the actuarial community sees the benefits of machine learning models, and there have been recent attempts employing the technique (Farkas *et al.*, 2021; Woods *et al.*, 2021). Based on these developments, we believe that we will see a growth in the actuarial literature on cyber risk modeling using machine learning techniques. Another burgeoning area of actuarial inquiry lies in the exploration of pricing principles that are more tailored to cyber insurance, a subject that also captivates computer scientists. In essence, the actuarial modeling framework can be enriched by embracing methodologies from other disciplines, and vice versa.

In the rest of this section, we discuss definitions of cyber risk, characteristics of cyber risk, and classification of cyber risk. We conclude this section with a description of the available data sources for cyber risk research.

### **1.2 Definition and characteristics**

There is not a unanimous definition of cyber risk. A widely accepted definition is the one provided by Cebula and Young (2010), which defines cyber risks as operational risks that may result in potential violation of confidentiality, availability, or integrity of information systems. Böhme and Kataria (2006) take a broader perspective by defining cyber risk as a risk leading to failure of

information systems. Böhme and Schwartz (2010) define cyber risk as a financial risk associated with network and computer incidents, emphasizing the importance of computer networks. Ögüt *et al.* (2011) regard cyber risk as an information security risk, highlighting the key element of computer interconnectivities. Stoneburner *et al.* (2002) also emphasize the cyber element in the definition. Biener *et al.* (2015) extend the definition provided by Cebula and Young (2010) to include a focus on the impact on information assets, while Mukhopadhyay *et al.* (2006) take into account malicious electronic events that lead to business disruptions and financial losses.

While cyber risk covers a broad spectrum of risks, some authors focus on parts of the spectrum. Hovav and D'Arcy (2004) describe a virus attack as a cyber event involving a malicious attack directed against a particular device or a network of devices. Hansman and Hunt (2005) define information gathering attacks as unauthorized information gatherings that may be used in further attacks. Moore *et al.* (2006) define Denial-of-Service attacks as attacks that consume the resources of a remote host or network which could otherwise be accessed by authorized parties. Jang-Jaccard and Nepal (2014) consider cyber risk at a national level and define cyber warfare as a nation's attempt to penetrate another nation's network with the intention to cause disruptions.

Although cyber risk has similar properties to operational and financial risks, it possesses some distinctive characteristics. Gordon *et al.* (2003) identify the uniqueness of cyber risk in terms of location, degree, and visibility compared to traditional business risks. Cyber risk is able to affect a wider range of individuals and organizations without being detected due to its mobile and intangible nature (Gordon *et al.*, 2003; Jang-Jaccard & Nepal, 2014). The low cost associated with initiating an attack also distinguishes cyber risk from other risks (Jang-Jaccard & Nepal, 2014). Biener *et al.* (2015) argue that cyber risk should satisfy three criteria: the risk has impact on critical technology or information asset, the risk involves a relevant actor in the cause of the accident, and a relevant outcome with respect to the victim's information inventory is present. The three criteria fundamentally separate cyber risk from traditional operational risks. Böhme (2005) pinpoints the interdependent nature of cyber risk, which means that insecure nodes not only jeopardize the security of their own systems, but also introduce risk to all other users. Also, a successful attack to one node can lead to victimization of other nodes, and even the entire network, especially if the security leak occurs in a monopolistic product (Böhme, 2005). Similarly, Anderson and Moore (2006) analogize cyber security externalities to traffic congestions, in a way that other internet users can be adversely affected by actions of the hosts of insecure nodes. Eisenbach *et al.* (2022) estimate that a cyber loss event in any of the five most active U.S. banks will result in a spillover to other banks, with an average impairment amounting to 38 percent of bank assets.

### 1.3 Classification

There exist different taxonomies of cyber risks. In the computer science and information technology literature, taxonomies of cyber risk typically focus on the technical properties. Ballardie and Crowcroft (1995) categorize cyber attacks as active or passive, with passive attacks being those that may result in an information release and active attacks being those that may lead to an information modification or denial of service. Howard (1997) presents a process-based taxonomy of computer attacks with five stages: attacks, tools, access, results, and objectives. Howard and Longstaff (1998) expand this proposed five-stage taxonomy into a network incident taxonomy that comprises seven stages: attackers, tool, vulnerability, action, target, unauthorized result, and objectives. Mirkovic and Reiher (2004) classify Denial-of-Service attacks by their degree of automation, exploited vulnerability, source address validity, attack rate dynamics, possibility of characterization, persistence of agent set, victim type, and impact on the victim.

The business and economics research community constructs taxonomies from a different perspective. Grzebiela (2002) segments cyber risks into four levels, namely, a technical risk level that includes fundamental system risks and malicious attacks, an individual risk level that describes risks associated with theft and abuse of private information, an economic risk level that describes

risks leading to economic losses, and a societal risk level that refers to risks that have an impact on a societal dimension. Eling and Schnell (2016) suggest that cyber risk can be classified according to the source, type of attack, and activity. Stoneburner *et al.* (2002) differentiate internet-related risks by sources including information misuse, unintentional errors and omissions, IT disruptions, and operational failures. Similarly, Cebula and Young (2010) classify cyber risk by sources including actions of people, systems and technology failures, failed internal processes, and external events. In the CyRiM report (Daffron *et al.*, 2019), cyber incidents are classified by their associated attack types, including data exfiltrations, contagious malware attacks, denial-of-service attacks, and information thefts. In addition, cyber incidents can be classified as criminal or non-criminal (Eling & Schnell, 2016), and the targets of cyber incidents may be included in the classification (Böhme *et al.*, 2019). In a recent publication, Awiszus *et al.* (2023) introduce a classification of cyber risks into idiosyncratic, systematic, and systemic risks, and review the corresponding actuarial modeling approaches and other more complex pricing techniques. Idiosyncratic risk is specific to each individual policyholder, systematic risk arises from exposure to common vulnerabilities, and systemic risk is caused by the contagion in interconnected systems.

#### 1.4 Available data sources

In cyber risk research, researchers face a relative dearth of data due to the fact that organizations are unwilling to reveal details of cyber incidents (Gordon *et al.*, 2003). Currently available databases include, but are not limited to, those mentioned below.

SAS OpRisk Global Data is the world's largest repository of publicly reported operational losses in consumer price index adjusted dollar amounts (Biener *et al.*, 2015). Access to the database is subscription based. The dataset records various key parameters related to an incident, including the time and type of the incident, characteristics such as geographic region, size, and industry classification of the affected entity, the magnitude of financial damage with the associated legal and regulatory obligations imposed on the victimized firm, and a detailed description of the event. The inclusive range of variables recorded enables users to identify the determinants of risk occurrence and explore the statistical properties of the resultant losses, which should be accounted for in capital consideration (Eling & Jung, 2022). Eling and Wirfs (2019) provide a list of keywords to identify the cyber incidents from the reported operational events.

Privacy Rights Clearinghouse (PRC) reports the number of personal records compromised in data breach incidents by breach type and industry. It primarily records breaches occurred within the United States. The database is freely available to the general public (Privacy Rights Clearinghouse, 2019) and has been widely used in empirical research to investigate the statistical and stochastic characteristics of cyber risk (Maillart & Sornette, 2010; Edwards *et al.*, 2016; Eling & Loperfido, 2017; Eling & Jung, 2018; Xu *et al.*, 2018; Farkas *et al.*, 2021).

The use of information about vulnerabilities, for example, honeypot data and the National Vulnerability Database, is prevalent in the field of information technology, with common use cases including intrusion detection and risk prediction (Zhan *et al.*, 2013; Chen *et al.*, 2015; Fang *et al.*, 2019; Zhang Wu *et al.*, 2023). Data collected from open-source honeypots, which are decoy systems or networks that gather information on network traffic generated by cybercriminals (Böhme & Kataria, 2006), often captures Internet Protocol (IP) addresses, packets sent or received, system-level details, and other network activities. The National Vulnerability Database is a free public database that stores information on known software flaws and their potential impacts, from sources including software developers and government agencies worldwide (National Institute of Standards and Technology, 2020). These types of data can be used to identify system vulnerabilities and analyze attack trends, which could offer valuable insights for risk assessment and technical controls.

On the actuarial front, the availability of actual insurance claims data plays an essential role in model development. However, only few databases exist for this purpose. The annual report published by NetDiligence and the Advisen cyber loss data are notable private databases that record

actual cyber claims globally. NetDiligence provides an overview of claims grouped by firm size, industry, and cause of loss (NetDiligence, 2022), while Advisen offers more detailed information on individual claims (Advisen, 2022). These claims data are not only crucial for developing insurance pricing models but also useful in the formulation of risk mitigation and risk transfer strategies as they are indicative of risk trends and the effectiveness of security measures. For a systemic review of data availability in the context of cyber research, we refer readers to Cremer *et al.* (2022).

Not only the research community recognizes the value of data collection, governments also notice the necessity of cyber risk incident reporting. In the European Union,<sup>1</sup> the General Data Protection Regulation came into force in 2018, requiring all entities operating physically or virtually within the European Economic Area<sup>2</sup> to notify their customers of a data breach within twenty-four hours of occurrence (Rustad & Koenig, 2019). Furthermore, the series of discussion papers published by the European Insurance and Occupational Pensions Authority (EIOPA) provide guidance on the methodological principles of insurance stress testing focusing on cyber risk and furnish a range of incident scenarios along with the relevant data required to assess the risk (EIOPA, 2022). This regulatory effort is believed to inspire and guide future collection of more precise information regarding cyber incidents. In the United States, all states have data breach notification laws in place, most of which cover social security numbers, driver's license information and credit card information (Voss & Houser, 2019). Australia has also made data breach notification to individuals and regulators mandatory for organizations from 2018 (Office of the Australian Information Commissioner, 2018). Australian regulators monitor cyber threats through the Australian Cyber Security Centre, which leads the Australian government's efforts to improve cyber security.<sup>3</sup> As governments are becoming more aware of the importance of cyber risk incident reporting, we expect to see an improvement in data availability in the future.

## 2. Risk prediction

Detection of malicious attacks is among the oldest and most studied subjects in cyber security, especially in the domain of computer science (Denning, 1987; Garcia-Teodoro *et al.*, 2009; Cárdenas *et al.*, 2011). However, reactive detection does not reduce the adverse impact of attacks as damage has already occurred upon detection. We have witnessed a shift in interest toward proactively predicting cyber risk, as it allows entities to take preemptive measures prior to the attack, fostering more effective risk assessment and mitigation practices. In particular, it allows organizations to identify potential risk factors and evaluate the efficiency of different controls in lowering the risk level. The designated tasks of risk prediction include attack projection, intention recognition, intrusion prediction, and network security situation forecasting (Husák *et al.*, 2018). In developing cyber prediction models, the true positive rate (TPR) is of the greatest interest, and at the same time the false positive rate (FPR) should be kept at a minimum since a falsely identified malware can incur massive costs (Sun *et al.*, 2018).

### 2.1 Existing surveys

Selected cyber risk prediction methods have been reviewed in a few survey papers. Ahmed and Zaman (2017) review various approaches for attack intention prediction including causal networks, path analysis, graphical analysis, and dynamic Bayesian network. Sun *et al.* (2018) survey

<sup>1</sup>The EU nations include Austria, Belgium, Bulgaria, Croatia, Republic of Cyprus, Czech Republic, Denmark, Estonia, Finland, France, Germany, Greece, Hungary, Ireland, Italy, Latvia, Lithuania, Luxembourg, Malta, Netherlands, Poland, Portugal, Romania, Slovakia, Slovenia, Spain, and Sweden (GOV.UK, 2020).

<sup>2</sup>The EEA includes all EU countries, Iceland, Liechtenstein, Norway, and Switzerland which is "not an EU or EEA member but is part of the single market" (GOV.UK, 2020).

<sup>3</sup><https://www.cyber.gov.au/acsc/report>

papers on data-driven cyber incident prediction schemes. The surveyed works are subsumed under six categories according to their utilized datasets: organization reports and datasets, network datasets, synthetic datasets, webpage data, social media data, and mixed-type data. Leau and Manickam (2015) group existing research on network security situation prediction into three categories according to the prediction method used, namely, machine learning, Markov models, and Grey theory. Husák *et al.* (2018) divide prediction methods into four categories: discrete models, continuous models, machine learning and data mining, and other approaches such as evolutionary computing and similarity-based approaches.

In this paper, we divide risk prediction approaches into three groups according to the theories and methods applied: graph models, time series models, and machine learning. This grouping is adopted, as time-series models and machine learning are two broad categories of modeling approaches that are often used in insurance and actuarial science. However, it is noteworthy that the vast majority of these publications come from journals in the disciplines of computer science and engineering, with scant contributions from statistics (Xu *et al.*, 2017), RMI (Zängerle & Schiereck, 2022), and social behaviors (Shu *et al.*, 2018).

## 2.2 Graph models

A graph model is a discrete model that depicts attack scenarios graphically and typically predicts the probability of a sequence of actions or attack path (Husák *et al.*, 2018). Attack graphs and some graphical models that are based on Bayesian networks or Markov models fall into the category of graph models.

### 2.2.1 Attack graphs

An attack graph predicts attacks by traversing all possible paths and then selecting the paths that lead to a successful system compromise (Husák *et al.*, 2018). GhasemiGol *et al.* (2016) introduce an uncertainty-aware attack graph method that accommodates the uncertainty of attack probabilities and uses additional information disregarded in traditional attack graphs, including current intrusion alerts, active responses, and network dependencies. The resulting forecasting attack graph comprises the lower and upper probabilities of attack, providing insights on the security level of the network being modeled. Polatidis *et al.* (2018) propose a recommendation system modified from the e-commerce recommender service for predicting cyber attacks. An attack graph illustrating all attack paths with specified conditions, such as attack location and length of the path, is generated. The attack graph is then combined with collaborative filtering, which is a system that provides personalized recommendation for unrated items based on user submitted ratings, to predict future attacks.

### 2.2.2 Bayesian networks

Bayesian networks are adopted in many cyber risk research papers. A Bayesian network is a directed acyclic graph where each node represents a variable with a set of states and edges represent the relationship between variables (Husák *et al.*, 2018). Bayesian networks are suitable for describing complex and evolving systems, given their capacity to incorporate new information as it becomes available. Additionally, the model is equipped to operate under incomplete and uncertain data adeptly.

Qin and Lee (2004) develop a graph-based approach that addresses attack correlation, plan recognition, and attack prediction. The authors apply a Bayesian-based correlation mechanism and statistical analysis to alert correlation and further correlate isolated alert sets by defining attack plan libraries made up of attack trees. The complexity of computing probabilistic inference on the causal network is reduced by implementing attack classes rather than specific attacks. A Bayesian

causal network is then built based on attack trees with each node representing a variable in a binary state and the directed edges representing the relationship among the nodes.

Okutan *et al.* (2017) suggest using a Bayesian network with unconventional signals (signals that are not directly related to the target entity but could be indicative of incoming cyber attacks) as indicative random variables and with different time granularities to forecast upcoming cyber incidents. The authors argue that the use of unconventional signals drawn from global events and social media improves the performance of predictions that are based on imbalanced data sets. A Bayesian classifier is built for each attack type considering all pairs of time granularities of signals and ground truth for an entity. The model outputs the probability that a particular entity will experience a cyber attack within the period of ground truth time granularity, with 63%–99% accuracy.

Huang *et al.* (2018) model cyber-to-physical attack propagation with a Bayesian network, from which the probabilities of sensors and actuators being compromised are inferred and fed into a stochastic hybrid system predictive algorithm. The algorithm describes a stochastic process on a hybrid state space containing both continuous and discrete states, which captures the coupling between continuous system dynamics and discrete attack behaviors, and further quantifies the physical impact of the attack.

A Bayesian network adds mathematical interpretations to the graphical illustration and deals with uncertain information effectively. However, a large set of data and complicated calculations are needed to construct such a network (Li *et al.*, 2020). Computing the conditional probability tables is resource-intensive, and the observed data may be inadequate to generate a tractable likelihood. In addition, the prediction results largely depend on the prior knowledge about probability distributions and conditional dependencies, and hence these inputs must be judiciously evaluated by experts. A limited number of public or private sources of empirical data can be used for the construction of Bayesian networks, such as the National Vulnerability Database, system logs, network topology, and incidents data (Chockalingam *et al.*, 2017). We anticipate an increase in the standardization of data and the extraction of useful information from diverse sources, given the increasing acceptance of legal requirements for breach reporting and the progress made in the field of machine learning.

### 2.2.3 Markov models

In a similar vein, Markov models complement the representation of attacks by adding mathematical reasoning to graphs. Contrary to Bayesian networks, Markov models do not rely on complete information as they can operate on sets with unobservable states and transitions, enabling attack predictions even when the system fails to detect some attack steps. Markov models include Markov chains and their variants such as hidden Markov models and Markov time-varying models.

Compared to a Markov chain, a hidden Markov model provides better adaptability as it dynamically updates the calculation for the state transition probability distribution (Li *et al.*, 2020). Ghafir *et al.* (2019) propose a probabilistic intrusion detection system based on a hidden Markov model for advanced persistent threat (APT) detection and prediction. The system undergoes an attack scenario reconstruction that sorts alerts into different APT scenarios, followed by the training of a hidden Markov model which generates prediction for the next stage of APT.

Some argue that hidden Markov models rely on overly simplified assumptions and the underlying time-independent assumption is often unrealistic (Li *et al.*, 2020). This limitation can be mitigated by a Markov time-varying model, another variation of Markov chains, which allows the states of the network and the transition probability matrix to vary with time. On the basis of a Markov time-varying model, Li *et al.* (2020) develop a cloud-fog-edge closed-loop feedback risk prediction model that is tailored to multitask compound attacks on electric power industrial control systems. The model encompasses a classification deep Boltzmann machine, an unsupervised feedback neural network composed of restricted connections of random binary variable units, in

which only nodes between different layers are connected. Given the classification results, Markov time-varying model updates the state transition probabilities and predicts the future status of the network.

Markov models are readily applicable and produce promising results, but the possible states and transition intensities are often difficult to define (Leau & Manickam, 2015). Estimating the states and transition intensities within a Markov model presents a challenge, as they are not directly observable and requires knowledge of system behavior. This challenge is amplified for more complex systems with more states and a large set of interdependent variables affecting state transitions. Similar to data scarcity encountered in constructing Bayesian networks, the development of Markov models utilizes system-specific characteristics and vulnerability data, whose collection is anticipated to become more standardized and readily accessible in the future.

### 2.3 Time-series models

Cyber incidents data are inherently stochastic and can be formulated as point processes (Daley *et al.*, 2003; Zhan *et al.*, 2013), suggesting that time-series models are suitable for predicting cyber risks. A time-series model allows for a straightforward examination of the time-series data and its temporal trends, making interpretation easier compared to other complex predictive models. Time-series techniques that are frequently used for modeling cyber risk include the families of integrated autoregressive moving average (ARIMA) models, fractionally autoregressive integrated moving average (FARIMA) models, and generalized autoregressive conditional heteroskedasticity (GARCH) models.

Park *et al.* (2012) use a moving average model (a special case of ARIMA) and linear regression to alert and predict worm propagation in a university campus network. The method estimates the remaining time to infection and identifies the branch point when the worm emerges.

Zhan *et al.* (2013) propose a statistical framework for predicting future attack rates from cyber attack data captured by honeypots. The framework incorporates long-range dependence (LRD) through a FARIMA model, which is demonstrated by the authors to produce better predictive power compared to processes that disregard LRD. Notably, The LRD-aware FARIMA process can predict network-level attack rates up to five hours in advance. Zhan *et al.* (2015) further improve the framework of Zhan *et al.* (2013) by considering the extreme value theory and gray-box time series theory. This extension is suited for predicting attack rates in the presence of extreme values and LRD.

Bakdash *et al.* (2018) claim that a time-series forecasting method featuring a Bayesian state-space model with a negative binomial distribution and a one-week lag has better predictive power over traditional time-series models due to the presence of overdispersion and bursts in the data. The Bayesian state-space model decomposes the process into states (observed number of attacks for a week) and observations (forecasts) with the transition between them defined. A Markov model of transition intensities is employed to identify and quantify bursts. Although the model produces reasonable forecasts for the majority of data, the prediction regarding bursts is unsatisfactory.

Fang *et al.* (2021) propose a framework to model and predict enterprise-level cyber risk utilizing sparse time-series data. The authors use a logit model to describe the probability of no breach in an entity at a specific time, an approach which inherently accommodates temporal trends and heterogeneities. They model the breach size using the peaks-over-threshold method with a generalized Pareto distribution for the upper tail. Temporal dependence is captured using a D-vine copula with pairwise dependence captured by the Frank copula. With the fitted model, the authors predict the probability of a breach in the entity within the next time interval and the corresponding breach size distribution. Simulation results and experiments based on the PRC data show the proposed approach has satisfactory prediction accuracy according to the ranked probability score and the uniform test.



Zängerle and Schiereck (2022) extend the framework of Fang *et al.* (2021) to include all types of cyber risk, and apply the extended framework to a novel dataset known as the Öffentliche Schadenfälle OpRisk (ÖffSchOR) database. The database records publicly disclosed losses over 100,000 Euros in the European financial sector and offers free access. Due to the geographical and industrial constraints, the dataset is even more sparse than the global PRC data at the enterprise level. The authors demonstrate that the framework generates valid forecasts according to the ranked probability score measure. Their results are in line with the findings of Fang *et al.* (2021).

While most of the existing research focuses on prediction methods, Chen *et al.* (2015) assess the predictability of attack frequency time series. They uncover the intrinsic spatiotemporal patterns of cyber attacks, challenging the prevalent view of attacker behavior as being completely random and unpredictable. By dividing spatiotemporal characteristics into deterministic and stochastic patterns, with the former implying predictability in attack frequency time series and the latter being analyzed by the flux-fluctuation relation, the authors successfully quantify the predictability of cyber attacks by the information entropy defined on the basis of the state transitions in the coarse-grained time series of attack frequencies. The attack patterns within a honeypot IP block are observed to be stably similar to each other, suggesting only a small number of sensors are required to capture the attack patterns in the whole cyberspace.

Another aspect of risk prediction is raised by Xu *et al.* (2017), who assess the effectiveness of cyber defense early-warning systems by considering a four-dimensional time series quantifying the number of attacks/victims with/without early warning mechanisms. The data used was obtained from the Center for Applied Internet Data Analysis at the University of California San Diego. The authors measure the effectiveness of an early warning mechanism by the reduction in the joint probability of certain numbers of attacks and victims post implementation. The required probabilities are calculated using Monte Carlo simulations and a copula-GARCH model, where the marginal distributions are captured by ARMA-GARCH models and the dependence is modeled by a copula. The results suggest the proposed copula-GARCH models outperforms the independence model and certain other copula models in terms of predicting the effectiveness of early warning mechanisms.

## 2.4 Machine learning

Machine learning has prevailed in many scientific fields, let alone the recent study of cyber risk borne by the development of computer science. It possesses the ability to learn from data and make predictions based on such learning. The quantity and quality of data used to train a machine learning algorithm are decisive factors in determining the accuracy and efficacy of the model. In situations where data is scarce, as is the case with cyber risk, researchers often resort to data mining to extract information from various sources, which can be integrated with existing datasets such as vulnerability databases and incidents reports, to create their own data for analysis.

Machine learning methods are characterized by their high fault tolerance, self-learning and organizing capability compared to traditional modeling approaches (Leau & Manickam, 2015). In studying cyber risk, they are often combined with other modeling methods to capture various attributes of the risk. Machine learning methods that are often used in cyber risk modeling include classification techniques, which aim to classify or predict discrete variables such as the risk class or attack type, and regression techniques, which aim to predict continuous values such as the probability of breach and the frequency of attacks in a given period. In our survey, all works implement classification techniques to predict a risk class (secure or risky) for a network system based on its input features, except for Fang *et al.* (2019) and Zhang Wu *et al.* (2023) that employ regression techniques to predict the attack rates.

#### 2.4.1 Random forest classifier

A random forest (RF) classifier is a classification method that randomly selects subsets from the input data to build trees and then outputs a probability of interest. It performs well with large and diverse feature sets, both numerical and categorical, and reduces overfitting as only a random subset is used each time (Bilge *et al.*, 2017).

Liu *et al.* (2015) propose a proactive forecasting methodology with a RF classifier by inspecting 258 externally observed features of organizations that may be indicators of network mismanagement, and malicious activity time series. Their model has decent long-term forecasting performance due to the use of mismanagement symptoms and long-term malicious behaviors in the feature sets, which are relatively stable and stand out as the most important categories in prediction. Testing on the incidents data provided by VERIS, Hackmageddon, and the Web Hacking Incidents Database yields a high level of accuracy, with a 90% TPR and a 10% FPR. Sarabi *et al.* (2016) extend the model of Liu *et al.* (2015) to include both cyber and non-cyber incidents that result in data breaching events. The precision level achieved by their model is commensurate with that of the prior work.

Bilge *et al.* (2017) develop a model with a RF classifier to predict the risk of infection months in advance, achieving a superb level of precision with a 96% TPR and a 5% FPR. The model quantifies infection risk with a risk score, defined as the posterior probability of infection given a vector of 89 features mined from enterprise customers of an antivirus company, which propagates to all user profiles according to similarity in features through an optimization framework that ensures similar profiles yield close risk scores and prevents clustering of risk scores around 0 and 1. The model also prevents evasion problems, in which invaders behave to avoid being detected by the machine learning system, since it collects information on benign usage rather than malicious behaviors before attacks. The authors find that in general, rarely updated machines, machines with higher usage over weekends and files signed by rarely known vendors tend to have higher probabilities of infection.

#### 2.4.2 Neural networks

The structure of a neural network can be described as a directed graph whose nodes, which correspond to neurons in the neural networks of a human brain, are joined by edges that link the outputs of some neurons to the inputs of other neurons. A standard neural network comprises input, hidden, and output layers that are partially or fully connected with each other (Shalev-Shwartz & Ben-David, 2014).

Subroto and Apriyana (2019) apply an artificial neural network that does not contain cycles to predict the likelihood of a vulnerability reported on Twitter being added to the Common Vulnerabilities and Exposures (CVE) database. They garner 1,000 most recent instances from Twitter and the CVE database and compare the performance of several statistical machine learning algorithms on the data. Their results are in favor of the artificial neural network, which could accurately predict up to 96.73% of the events.

In a recurrent neural network (RNN), the hidden layer completes the most important work. An RNN updates the hidden layers using inputs from the input layer and activations from the previous forward propagation, accommodating the temporal patterns exhibited by the input data (Fang *et al.*, 2019). Rhode *et al.* (2018) introduce an RNN model that utilizes machine activity metrics (including system CPU usage, user CPU use, packets sent/received, byte sent/received, memory use, swap use, total number of processes currently running, and maximum process ID assigned) as inputs to output a binary classification of the risk. The RNN model is able to detect a malicious ransomware attack within 5 seconds of its arrival with a 94% accuracy. As the average execution time of malware is around 5 minutes, the short detection time allows the user to predict an attack before its execution. One drawback of the model is that adversaries may be concealed if the malware plants long sleeps or benign behaviors at the start. Furthermore, the model lacks defensive power.

As an RNN only learns from the past, it fails to capture future information as it becomes available (Schuster & Paliwal, 1997). Instead of using a standard RNN, Fang *et al.* (2019) develop a bidirectional RNN with long short-term memory (BRNN-LSTM) to describe cyber attack time-series data captured by honeypots, with the objective of forecasting attack rates. The bidirectional RNN extends the standard unidirectional RNN by allowing learning from both the past and future. In a BRNN-LSTM framework, each hidden layer of BRNN is replaced with a LSTM cell, which stores the temporal state of the network. In general, the model can achieve a prediction error of less than 5%.

Similarly, Zhang Wu *et al.* (2023) present a deep learning framework for predicting the number of cyber attacks through an RNN model that captures the multidimensional dependence and accommodates LSTM. Additionally, the residuals are modeled using the peaks-over-threshold approach with a generalized Pareto distribution for the tail. A prediction of the mean and high quantiles of the number of attacks is produced using forward propagation. The authors perform a simulation study as well as an empirical study with the Amazon Web Services honeypot data to assess the feasibility and performance of the proposed framework. Prediction results evaluated with unconditional and conditional coverage tests suggest that the model enjoys not only accurate point estimates via deep learning but also good high quantile predictions via extreme value theory.

Sun *et al.* (2019) argue that the aforementioned neural networks may misclassify a risky class as a risk-free class when the ground truth dataset is imbalanced and small. They therefore propose a Siamese network-based deep learning classification framework to deal with imbalanced data in risk forecasting. The framework comprises a data processing phase, a model training phase, and a risk prediction phase. Inputs are arranged into pairs in the data processing phase and mapped into new sample spaces based on similarity calculations in the model training phase. A risk classification is then predicted using matrix operations. During the prediction phase, a test set is fed into the trained Siamese network and the corresponding transformed pairs, similarity distance, and probability of belonging to a risk category are calculated. Finally, a node is labeled as risky if the predicted probability is above a user-defined threshold.

#### 2.4.3 Logistic regression classifier

Logistic regression classifiers are linear classifiers that decide classification based on regions separated by hyperplanes (Denœux, 2019). They can be applied to various classification problems.

Canali *et al.* (2014) seek to predict how risk experience differs among different groups of users using a logistic regression classifier. Users are classified as safe, at risk, or uncertain, according to their browsing histories. The authors extract 74 attributes regarding user browsing behavior from the telemetry data collected by Symantec to build user profiles for each group and train the classifier. The extracted features are primarily related to the volume of online activity, active time windows, diversity, stability and popularity of websites visited, and the user's computer type. The authors find that frequent internet users are more likely to be at risk of encountering malicious websites. In addition, the proportion of time spent on browsing adult contents is positively related to the level of risk. With up to 87% accuracy, the research postulates the possibility for insurers to silently profile users based on information about HTTP requests and price policies accordingly.

Instead of browsing behavior, Shu *et al.* (2018) analyze sentiments extracted from Twitter data and cyber incidents reports provided by a financial company and a defense company to devise a model as part of a logistic regression predictor to forecast cyber attacks. The sentiment analysis method utilizes emotion signals, such as emotion words, emoticons, and punctuation, in an unsupervised manner and models the correlations among them. The information extracted is then used to train a logistic regression classifier, which outputs the predicted sentiment scores that can be used to forecast the probability of attacks. An analysis about the temporal variation of sentiments over time is also conducted to provide insights into the progression of ongoing cyber attack behaviors. The authors discover that the sentiment scores are strongly negative on days preceding the attack and scores tend to increase after the attack takes place.

Similarly, Sarkar *et al.* (2019) predict cyber incidents by leveraging the reply network structure of user interactions on the dark web hacker forums. The authors use 280 incidents records from Armstrong Corporation with attributes of event type and date of occurrence, and data from 53 forums on the dark web over a similar time period. Common vulnerabilities and exposures (CVE) mentioned in the posts from the 53 forums are extracted and assigned to different common platform enumeration groups considering the operating system platforms and application environment tags. A temporal network is built, in which a time series feature for every time point is generated. The system then undergoes feature curations for each time point, during which features regarding user/forum statistics, expert centric, and network centralities are mined. Experts refer to users who have a history of CVE mentions and whose posts have high impact in the forum. Following feature curation, the time-series data are utilized to produce a binary attack classification forecast at every given time point. To accommodate the longitudinal sparsity in high-dimensional data, a temporal feature selection window is specified. Features in the temporal window are used to predict the probability of an attack using a logistic regression with longitudinal ridge sparsity and group Lasso regularization.

## 2.5 Discussion

In this section, we review the literature on the prediction of cyber risk. We present the works under three broad categories, namely graph models, time series models, and machine learning, and further divide them into subcategories. Machine learning approaches constitute the most prominent segment of the literature surveyed surrounding cyber risk prediction. Machine learning in the context of cyber risk has been studied extensively especially in the past five years, with more than half of the cyber risk predictive algorithms developed being based on machine learning models.

Our review finds the most common use cases of these predictive mechanisms are to forecast the likelihood of encountering an attack, the time remaining to the next attack, and the number of attacks expected in a given time frame. Regrettably, prediction for the sizes of attacks is overlooked. Severity prediction can offer valuable insights for insurers with respect to pricing and capital considerations. Severity prediction hinges not only on features related to the underlying network system but also on the value of information stored in the system. The latter is likely confidential and difficult to estimate. It is possible to assimilate insights from severity models in the literature, as surveyed in Section 3, to supplement the current predictive framework.

More works from the perspective of attackers could provide a new angle to the subject. Previous studies that consider hackers' perspectives include those of Shu *et al.* (2018) who analyze the social media data related to a hacktivist group, and Sarkar *et al.* (2019) who analyze the mentions in dark web hacker forums. These studies establish a connection between the behaviors of potential attackers and the occurrence of attacks. However, the statistical relationship between features or actions of an organization and attacker behaviors is not explored. Understanding the link could help risk managers prevent attacks by avoiding operations that may provoke attackers. Achieving this understanding requires inputs from experts in social science and even first-hand data collected from attackers themselves.

Another avenue for future research is the development of metrics to evaluate prediction accuracy (Husák *et al.*, 2018). Since most models typically predict the binary risk class of a system with the input features, the results are often evaluated using the confusion matrix. In the presence of imbalanced data, common in cyber datasets (Sun *et al.*, 2019), the confusion matrix can be biased. The use of precision or recall values instead of the confusion matrix could alleviate the problem. Furthermore, a binary decision is often made by sorting the predicted value, such as the probability of breach or a risk score, through the use of a threshold value. Hence, the evaluation made by the confusion matrix also depends on the choice of the threshold, which could be selected through an optimization algorithm or set as a fixed value. These make prediction results hardly comparable

across different models. To overcome this limitation, more accurate and robust evaluation metrics are needed to ensure fair and reliable comparisons between different models.

### 3. Risk modeling

In addition to the prediction of risk arrival, accurate measurement of the resulting damage is another crucial step in effective risk management. By combining the ex-ante and the ex-post approaches to risk management, organizations are able to conduct scenario analyses and stress testing under adverse cyber conditions, so that they can assess the potential impact of various cyber threats more thoroughly. In the literature, certain statistical properties of cyber risk are identified. Böhme *et al.* (2019) pinpoint that cyber risk exhibits skewness and positive correlation. Bakdash *et al.* (2018) indicate that overdispersion exists in cyber loss frequency distributions, thereby suggesting the use of negative binomial instead of Poisson for frequency modeling purposes. Using data from the Open Security Foundation and PRC, Maillart and Sornette (2010) report a heavy tail for the number of identity losses per data breach incident. The property of heavy tailedness is supported by several other researchers (Edwards *et al.*, 2016; Wheatley *et al.*, 2016; Riek & Böhme, 2018; Sun *et al.*, 2021). As a counterargument, Eling and Loperfido (2017) demonstrate that data breach severity distributions may not be as heavily tailed as other operational risks. Their results echo that of Mukhopadhyay *et al.* (2006).

Various methods have been proposed to model the statistical attributes of cyber incidents. Reviewed in the rest of this section, these methods are divided into five categories: standard distributions, copula-based approaches, extreme value theory, stochastic processes, and epidemic models. These models are capable of addressing different facets of cyber risk. Standard distributions are utilized to model the frequency and severity of cyber losses separately, while the dependence between either frequency or severity can be accounted for using copula-based approaches. Extreme value theory is suitable for addressing the heavy tails in loss severity, while stochastic processes can effectively capture the temporal evolution in attack frequency and severity. Lastly, epidemic models are implemented to characterize the propagation of the risk. The works cited in this survey come from a diverse range of journals, primarily from the fields of computer science and actuarial science, while some are from statistics (Mishra & Pandey, 2014; Liu *et al.*, 2016a; Peng *et al.*, 2017, 2018) and natural science (Gil *et al.*, 2014; Antonio *et al.*, 2021).

#### 3.1 Standard distributions

A conventional approach to risk modeling is to model the frequency and severity separately using standard distributions. When there is sufficient information about the timing and sizes of incidents but a lack of information on the attributes of breached entities, standard distributions can prove valuable in describing the statistical properties of cyber risk. As previously mentioned, negative binomial distributions are often used for modeling the frequency of cyber incidents, while lognormal and gamma distributions are typically devised for modeling cyber loss severity. For instance, in developing a Bayesian generalized linear model for modeling data breach trends, Edwards *et al.* (2016) assume that the prior distribution of data breach frequency is a negative binomial with a skewness parameter that is gamma distributed and a location parameter that is specified as a function of time. The authors also recognize that losses in cyber incidents are typically heavy tailed. As such, in their Bayesian generalized linear modeling work, the prior distributions of malicious breach sizes and number of records lost in a negligent incident are assumed to be lognormal and log-skew-normal, respectively.

Similar distributional assumptions are adopted by Eling and Loperfido (2017) in their PRC data breach analysis that is based on a multidimensional scaling (MDS) and a multiple factor analysis for contingency tables (MFACT). The MDS is used to investigate differences between companies that experience data breaches and differences between types of breaches, whereas the MFACT

provides a joint analysis of multiple contingency tables containing cyber incident frequencies. The results reveal that both lognormal and log-skew-normal distributions provide a promising fit to the original (unscaled) data. However, the log-skew-normal distribution performs better in predicting insurance premiums.

Riek and Böhme (2018) study the characteristics of loss distributions for different types of cyber crimes using data concerning direct cyber losses and protection expenses in a victimization survey of adult internet users from six selected EU nations. The authors report that cyber loss distributions are typically skewed and zero-inflated, due to the presence of zero-losses and abundant small losses. To counter the issue of positive skewness and zero inflation, they consider the harmonized loss indicator (HLI) as an indicator for the unconditional losses inferred from the conditional losses. The HLI scales the distribution median by the probability of a loss, such that  $\text{HLI} = \hat{q}\rho_{50}$ , where  $\hat{q}$  is the empirical relative frequency of loss events and  $\rho_{50}$  is the median of the conditional distribution of loss severity given a loss has occurred. The total loss is then estimated as

$$\mathcal{L} = \sum_{i \in I} \hat{p}_i (M_i + \alpha T_i),$$

where  $I$  is the set of cybercrime types,  $\alpha$  is a conversion factor that converts time to monetary values, and  $\hat{p}_i$ ,  $M_i$ , and  $T_i$  represent the probability of being victimized, the monetary loss summarized by the HLI, and the time taken to deal with an incident for cybercrime type  $i$ , respectively. The authors notice deviations from the fitted lognormal distribution in the tails. While overestimation of small losses may be acceptable, underestimation in the upper tail may lead to dire consequences.

Woods *et al.* (2021) suggest aggregating individual parameterized distributions, including polynomial, lognormal, Pareto, Burr, gamma, and Weibull, into a “county fair cyber loss distribution.” They devise an iterative optimization process that is built on a particle swarm optimization of parameters of candidate distributions to infer loss distributions from insurance prices offered by each insurer. By applying the iterative optimization algorithm to 6,828 observed prices from regulatory filings of 26 insurers in California, the authors find gamma distributions are most suited for predicting individual insurance liability prices. The county fair cyber loss distribution is derived by averaging the optimal loss distributions across all insurers.

### 3.2 Copula-based approaches

Classical approaches to risk modeling using standard distributions rely on the assumption of risk independence. However, researchers have recognized the presence of interdependence among cyber risks (Biener *et al.*, 2015; Eling & Schnell, 2016; Marotta *et al.*, 2017), which should be adequately addressed in the mathematical modeling process. Copula-based approaches are useful in modeling cyber risk due to their ability to address nonlinear risk dependency, which could exist in frequency (Mukhopadhyay *et al.*, 2006, 2013; Bentley *et al.*, 2020), severity (Böhme & Kataria, 2006; Eling & Jung, 2018; Liu *et al.*, 2022), or attack rates (Peng *et al.*, 2018). A copula-based approach expresses the joint distribution of multiple random variables as a multivariate function of their marginal distributions (Aas *et al.*, 2009), without requiring more data compared to the standard distribution approach. For an  $n$ -dimensional random vector  $\vec{X} = (X_1, \dots, X_n)$  with marginal distribution functions  $F_1, \dots, F_n$ , Sklar’s theorem (Sklar, 1959) states that the multivariate distribution  $F$  for the random vector can be expressed in terms of a certain appropriate  $n$ -dimensional copula  $C$  as follows:

$$F(x_1, \dots, x_n) = C(F_1(x_1), \dots, F_n(x_n)).$$

Mukhopadhyay *et al.* (2006) propose a copula-aided Bayesian belief network for cyber vulnerability assessment and expected loss quantification. In this modeling approach, the loss frequency in each node is modeled by a normal distribution, while the dependence across nodes is described by

a copula. With the joint distribution, the conditional distribution for each node is derived and then fed into a Bayesian belief network, which outputs the frequency of a cyber incident. This modeling approach is revisited by Mukhopadhyay *et al.* (2013) who specify a Bayesian belief network with a multivariate Gaussian copula and normal marginals for risk assessment and quantification purposes.

Bentley *et al.* (2020) support the use of a Gaussian copula, but they suggest using negative binomial marginals. In addition to a copula-based multivariate distribution, the authors propose a mitigation model, modified from the economic model of Gordon and Loeb (2002) for optimal cyber security investment, to quantify the effect of risk mitigation strategies on the costs of cyber attacks. In the mitigation model, the probability of a successful attack given a certain mitigation measure is estimated by multiplying a baseline probability, which specifies the probability of a successful attack when there is no spending on mitigation, for the attack type in question by the scaling factor calibrated for the mitigation measure in question. The effect of risk mitigation spending on loss severity is modeled in a similar manner. This approach enables a separate modeling of loss frequency and severity and demonstrates the diminishing mitigating effects of increased security spending. However, as illustrated in the authors' numerical experiment, which involves analyzing tickets submitted by security engineers at an anonymous company, the modeling outcomes are notably sensitive to the underlying dependence assumption.

Böhme and Kataria (2006) propose using a  $t$ -copula, which may be more suitable for modeling the correlation for extreme events. In more detail, the authors use a twin-tier approach to describe correlations in data breaches and demonstrate that a firm's decision to purchase insurance is based on the intra-firm risk correlation while the global risk correlation influences the premium level set by insurers. The intra-firm risk correlation is characterized by a correlation measure under a beta-binomial assumption for the number of failed nodes in an incident, while the global correlation is modeled by a  $t$ -copula. An insurability analysis is conducted through simulation, and the correlation estimation is demonstrated through a numerical example based on honeypot data.

Herath and Herath (2007) use Archimedean copulas to model the dependence between the number of computers affected (denoted by  $X$ ) and the dollar value of losses (denoted by  $Y$ ) in a firm. Archimedean copulas enable asymmetric tail dependence, the degree of which can be adjusted through a single dependence parameter  $\theta$ . The Archimedean copulas considered by the authors include Clayton and Gumbel, which can be expressed as

$$C_{\theta}(u, v) = (u^{-\theta} + v^{-\theta} - 1)^{-\frac{1}{\theta}}$$

and

$$C_{\theta}(u, v) = \exp \left\{ - \left[ (-\log u)^{\theta} + (-\log v)^{\theta} \right]^{\frac{1}{\theta}} \right\},$$

respectively, where  $u = F_X^{-1}(x)$  and  $v = F_Y^{-1}(y)$ . Pricing using the proposed framework is illustrated through a numerical example with data from the International Computer Security Association (ICSA).

Eling and Jung (2018) propose another alternative, the R-vine pair copula construction (PCC), to model nonzero pair dependence between data breach risks in both cross-industry and cross-breach settings using the PRC datasets from 2005 to 2016. The R-vine PCC resolves the problem of multivariate dependence present in the Archimedean model and transforms a high-dimensional copula analysis to a bivariate analysis. It encompasses the process of tractable tree building: the first tree consists of random variables, the second tree is built on the basis of conditional variables estimated from the previous tree, and so forth. The R-vine model does not specify a particular structure, thereby offering more flexibility. To compare risk measurements and pricing of aggregate loss distributions across different loss dependence models, the authors employ a simulation study. The R-vine PCC produces empirical results that are consistent with those of Böhme

and Kataria (2006), which reveal a high internal correlation between hacking attacks and insider attacks.

The use of a R-vine PCC is supported by Peng *et al.* (2018). Peng *et al.* (2018) use a combination of a time-series process and a vine copula to model multivariate dependent cyber attacks. In more detail, the authors model the marginal attack rate (number of attacks per unit time) on a server over time with an ARMA-GARCH process, and accommodate the high-dimensional dependence across multivariate cyber attack time series with a truncated R-vine copula. The truncation of the R-vine copula is achieved such that all pair-copulas in trees higher than a certain order are set to be bivariate independent copulas. Parameters in the model are estimated using the inference function of margins method with maximum likelihood estimation. The findings of both simulation and an empirical study with honeypot data indicate ignoring dependence across the attack time series can underestimate the value-at-risk measure, leading to falsely optimistic assessments of cyber risk.

Building on Eling and Jung (2018), Liu *et al.* (2022) propose a Bayesian framework that selects the margins and copulas of different types of data breach losses simultaneously. The marginal distribution of loss severity is modeled by the generalized beta type II distribution, which nests a family of distributions commonly used to model loss sizes, including the lognormal, gamma, Weibull, and Pareto distributions. The vine copula tree structure and pairwise copulas are identified through Bayesian selection based on different posterior probabilities that serve different purposes of an analysis. Through experimentation with the PRC data and data collected by the U.S. Department of Health and Human Services, as well as a series of sensitivity tests, this study confirms the robustness of the Bayesian framework to prior distribution settings, and reveals the significance of incorporating parameter and model uncertainties.

### 3.3 Extreme value theory

Extreme tails of cyber losses reported by numerous studies can be addressed using standard distributions like lognormal and Pareto distributions (Edwards *et al.*, 2016; Eling & Loperfido, 2017; Eling & Wirfs, 2019). However, standard distributions inherently assume specific tail behaviors, which may not accurately reflect the actual data characteristics. In light of this, a more flexible tool for modeling extreme events is the Extreme Value Theory (EVT). Since Beirlant and Teugels (1992) discussed the relevance of EVT to modeling extreme insurance losses, EVT has flourished in the insurance world. The peaks-over-threshold (POT) technique from EVT is prevalent in modeling the severity of cyber losses. This technique assumes that exceedances (observations that lie below or above a specific threshold) follow a certain extreme value distribution. Early applications of EVT to cyber risk modeling include that of Maillart and Sornette (2010) who consider data from the Open Security Foundation and the PRC and report that a stable heavy-tail power-law distribution is suited for modeling the number of identity losses per data breach incident.

Following Maillart and Sornette (2010), Wheatley *et al.* (2016) apply EVT to modeling 619 data breach incidents recorded by PRC and the Open Security Foundation between 2007 and 2015 that involved more than 50,000 personal identity losses. Specifically, the authors use a Poisson generalized linear model with an identity link to regress the breach frequency per month, and observe that the occurrence of data breaches is relatively stable globally. With respect to the severity of data breaches, the POT method is implemented to determine if a maximum exists. To achieve this objective, the authors quantify the severity of large breaches by an extreme heavy-tail doubly-truncated Pareto, which is transformed to a doubly-truncated exponential by taking the natural log of the data for convenience. They posit a finite maximum for the size of breach  $X$ , which is formulated as  $v(t) = u - \beta(t)/\xi < \infty$ , where  $u$  is the threshold,  $\xi$  is the extreme value index, and  $\beta(t)$  is a time-dependent scale parameter. The authors observe a significant upward growth in the maximum of natural log of breach sizes, and therefore propose to further model the time-dependent scale parameter as  $\beta(t) = \beta_0 + \beta_1 \ln(t)$ , where  $\beta_0$  and  $\beta_1$  are the scale intercept and scale slope, respectively.



Eling and Wirfs (2019) study the actual monetary loss for all types of cyber risk using the SAS OpRisk global operational risk dataset and acknowledge the need to differentiate between daily cyber risks and extreme cyber risks. They use a dynamic EVT approach, which models the aggregate loss with frequency following a Poisson generalized linear model with a log-link function and severity following a POT with a generalized Pareto tail. The scale parameter of the generalized Pareto distribution is orthogonally transformed to ensure convergence. The authors notice an increasing trend in the number of attacks per year while the probability of extreme losses decreases over time, a result that appears to contradict that of Wheatley *et al.* (2016) who argue that the rate of large data breach events increases outside US.

Malavasi *et al.* (2022) implement EVT with an extension of the generalized linear model, called generalized additive models for location shape and scale (GAMLSS). The authors describe the risk frequency with a Poisson distribution and severity with the POT approach under the GAMLSS framework. To better identify the influential factors in the tail behavior, the GAMLSS identifies and addresses risk drivers in not only the mean and variance but also the higher moments of the frequency and severity processes. The authors further apply a rank-based ordinal regression to remove the distortion of the significance of covariates in GAMLSS that is caused by extreme cyber events. By fitting the models on the Advisen cyber loss data, the authors confirm that the modeling results are mainly driven by extreme losses, a conclusion that does not support the insurability of cyber risk.

Sun *et al.* (2021) study the hacking data breach records from the PRC dataset over 2010–2019. In this study, frequency is modeled with a hurdle Poisson model. Severity for each company is defined as the log-transformed average number of exposed records, and, similar to the proposal of Eling and Wirfs (2019), modeled with the POT approach that results in a mixed model with a generalized Pareto distribution for the tail and a non-parametric distribution for observations below the threshold. The joint density function of frequency and severity is then captured by a bivariate copula. The authors find the positive non-linear dependency between frequency and breach size is best captured by a Gumbel copula. Finally, conversion from the predicted loss that is based on the number of breached records to monetary impact is performed using the following linear formula adopted from Jacobs (2014) and Romanosky (2016):

$$\log(\text{impact}) = 7.68 + 0.76 \log(\text{records}).$$

The formula is derived by means of fitting the Ponemon cost of data breach reports data in 2013 and 2014, which might not be representative to the PRC data from 2010 to 2019, and the formula may be too simplistic as it only considers the number of breached records as the determinant. However, the linear relationship is used here to compare the rating performance of the proposed dependence model and the independence model directly, with no bearing on the actual losses.

The POT technique is combined with regression trees by Farkas *et al.* (2021) to classify and predict different parts of the loss distribution. The central part of the PRC data is fed into regression trees that iteratively split the observations into classes according to predefined partitioning rules set by different modeling objectives. From the maximal tree built, an optimal subtree is selected according to a pruning algorithm. The tail part is fed into generalized Pareto regression trees, which use the generalized Pareto loglikelihood as a split function. The regression tree approach allows the identification of characteristics that create heterogeneity in cyber events. Additionally, the dissemblance between the central regression tree and a tree obtained from the global set of observations implies the heaviness of the tail.

The POT technique relies on the choice of a threshold, whose optimal value is hard to identify. To mitigate this problem, Jung (2021) proposes to model maxima cyber loss data obtained from Cowbell Cyber with a generalized extreme value distribution that features the block maxima method. In this method, the probable maximum loss for data breach risk,  $\xi_p$ , based on the value-at-risk at the  $1 - p$  quantile satisfies

$$P(\tilde{M}_n \leq \xi_p) = 1 - p,$$

where  $\{\tilde{M}_n\}$  is the series of loss maxima that follows

$$P\left(\frac{\tilde{M}_n - b_n}{a_n} \leq x\right) \rightarrow G^\theta(x),$$

with  $n$  being the size of the time frame,  $b_n$  being the number of blocks,  $a_n > 0$  being a constant, and  $G^\theta$  being the limiting generalized extreme value distribution with an extremal index  $\theta$  that quantifies the degree of dependency in extremes. The maximum loss estimates obtained by Jung (2021) are significantly higher than those produced by Wheatley *et al.* (2016), a result which could indicate the model's potential to approximate a systemic risk event or a dragon king loss event.

### 3.4 Stochastic processes

The aforementioned methods are effective in describing the statistical features of cyber events, but they do not fully capture the temporal evolution of these events. To incorporate temporal dynamics and reveal dependencies and trends over time, stochastic processes are often devised to model cyber risk. Correlations between inter-arrival times and between breach sizes discovered by Xu *et al.* (2018) suggest that stochastic processes are better suited than standard distributions for the purpose of describing cyber events. In order to apply a stochastic process model, data must be organized into a time-series format. This is a straightforward task when dealing with established databases, as they typically record incident occurrence times.

Peng *et al.* (2017) propose using a marked point process to model the extreme value phenomenon relating to the time series of cyber attack rates collected from the network telescope and honeypots. In more detail, the authors model the magnitude of extreme attack rates with the POT technique and describe the arrival of extreme attack rates with an autoregressive conditional duration (ACD) approach, which is analogous to a GARCH model but accommodates additionally both a slow decay of autocorrelation and bursts of extreme value clusters. The marked point process has a conditional intensity that is dependent on time, past information about occurrence times and marks (extreme values), and functions of past durations modeled by an ACD or log-ACD that accommodates the correlated inter-exceedance times. They measure intense risk by value-at-risk, a metric that describes the extreme cyber attack rates that can lead to potentially catastrophic consequences under inadequate defense. It is demonstrated that the model has accurate in-sample fitting and out-of-sample prediction.

Xu *et al.* (2018) model the PRC data from 2005 to 2017 with stochastic processes along with copula and EVT methods. The authors use an autoregressive conditional mean point process for modeling inter-arrival times that indicate frequency and sizes of data-hacking breach attacks, and an ARMA-GARCH process for modeling the evolution of breach sizes. Innovations from the processes are modeled by a mixed extreme value distribution with a generalized Pareto distribution for exceedances and a normal distribution for other realizations. They discover an increasing trend in hacking frequency, but no significant change in breach sizes. The authors further use a Gumbel copula to account for the positive dependence between inter-arrival times and breach sizes.

Similar to Xu *et al.* (2018), Ma *et al.* (2022) use ARMA-GARCH to describe the severity of cyber incidents between 2012 and 2018 reported by PRC and apply spatial clustering to address the geographical dependence. Individual severity models for each state in the United States are built using the ARMA-GARCH process, which are then combined in the K-means clustering process and reestimated on a cluster-based level. The inter-arrival times of incidents in each cluster are described by ACD models. By applying the clustering method, the statistical correlation between inter-arrival time and severity of cyber incidents across clusters is successfully removed, making the pricing process less complicated than the copula approach for insurers.

Zeller and Scherer (2022) propose a comprehensive model that is based on marked point processes for idiosyncratic incidents and systemic events. In the model, the arrival of idiosyncratic incidents at each firm is assumed to follow a Poisson process with a time- and covariate-dependent rate, and the arrival of systemic incidents follows a separate non-homogeneous Poisson process. Each arrival time point is equipped with a mark space, formed by the set of event strengths and the affected subsets, such that the resulting process is a marked point process. The severities of both idiosyncratic incidents and systemic events are modeled by a mixed distribution with a generalized Pareto distribution for exceedances and a truncated lognormal for other observations. Through fictitious insurance portfolios, the authors demonstrate the necessity to consider systemic events so that potential accumulation risk can be captured.

The presence of the accumulation phenomena is also internalized in the model of Bessy-Roland *et al.* (2021), who use multivariate Hawkes processes with specific kernel choices to describe the frequency of cyber attacks in the PRC data between 2010 and 2019. The Hawkes processes are able to reproduce clustering and autocorrelation between inter-arrival times, and they capture both shocks and persistence after shocks through their self-exciting property. The Hawkes process for the number of data breaches for group  $i$  in time interval  $[0, t]$  features an intensity process  $\lambda_t^{(i)}$ , which has a baseline intensity  $\mu_t^{(i)}$  and kernels  $\phi_{i,j}(t)$  that quantify the contagion in group  $i$  caused by a data breach in group  $j$ . The intensity process is specified as

$$\lambda_t^{(i)} = \mu_t^{(i)} + \sum_{j=1}^d \int_{[0,t]} \phi_{i,j}(t-s) dN_s^{(j)}.$$

### 3.5 Epidemic models

Although the previously discussed methods can capture statistical properties and temporal trends of cyber losses, they fall short in accounting for the propagation phase of the risk. Studying the spread of cyber risk enables us to pinpoint critical factors that influence the epidemic size and the final losses, thus enhancing the design of risk assessment and mitigation strategies accordingly. The approach necessitates interdisciplinary collaboration, as it may require knowledge about system vulnerability, network engineering, social behaviors, and natural disease spreading. Epidemiology naturally lends itself to the study of cyber risk propagation, since both epidemics and cyber attacks exhibit contagiousness, interdependence of exposure, and dynamic evolution.

Inspired by techniques for analyzing associations between gene mutations and diseases in genetic epidemiology, Gil *et al.* (2014) introduce a statistical framework for studying the susceptibility of a single node. The authors treat the network services running on a host as the defining risk factor on its susceptibility to certain cyber threats, a concept that is borrowed from a dominant model of genetic penetrance, which presumes that the presence of a dominant variant of a gene is sufficient to produce an associated phenotype, regardless of the number of copies present.

Barracchini and Addessi (2014) modify the multistate models for health insurance to capture characteristics of cyber risk. They classify the possible states of a computer damage as nd (no damage), rd (repairable damage), prd (partially repairable damage), and nrd (not repairable damage), and further divide the prd state into  $n$  levels,  $\text{prd}^{(i)}$  for  $i = 1, \dots, n$ , to indicate different levels of partially repairable damage (a higher  $i$  corresponds to a higher level of damage). This classification leads to a state space of  $S = \{\text{nd}, \text{rd}, \text{prd}^{(1)}, \dots, \text{prd}^{(n)}, \text{nrd}\}$ , formulating a Markov process from which the dynamic probability that a computer moves from one state to another can be inferred.

Liu *et al.* (2016a) apply concepts in epidemiology to devise a novel compartmental mathematical model for malware propagation. The authors argue that susceptible computers in a network should be viewed as heterogeneous rather than homogeneous with respect to the level of protection. Depending on its level of protection, a computer being modeled can be either weakly

protected susceptible (a W-node), strongly -protected susceptible (an S-node), or infected (an I-node). A W-node has a higher probability of infection compared to an S-node, and these two types of nodes can communicate with each other. The infection probabilities of a W-node and an S-node in a unit time are specified as

$$P_W = 1 - (1 - \beta_W)^{I(t)}$$

and

$$P_S = 1 - (1 - \beta_S)^{I(t)},$$

respectively, where  $\beta_W$  and  $\beta_S$  are the assumed infection rates for W-nodes and S-nodes, respectively, and  $I(t)$  is the number of infected nodes at time  $t$ . The average number of secondary infections by an infectious computer over its infectious period can be calculated as

$$R_0 = \frac{\beta_W \alpha + \beta_S \epsilon}{\gamma(\alpha + \epsilon)},$$

where  $\alpha$  is the probability of an S-node converting to a W-node,  $\epsilon$  is the probability of a W-node converting to an S-node, and  $\gamma$  is the probability that an infected node recovers to an S-node. The authors propose that there exists a unique malware equilibrium if  $R_0 > 1$ . If  $R_0 \leq 1$ , then the malware-free equilibrium is globally asymptotically stable. Their results are similar to those produced by Mishra and Pandey (2014), who use a susceptible-exposed-infectious-susceptible-with-vaccination epidemic transmission model to describe worm propagation in a network.

Fahrenwaldt *et al.* (2018) model the spread of cyber infections with an interacting Markov chain and claims with a marked point process. In their framework, the spread process is of a pure jump-type with exponential waiting times between jumps, whereas transitions are described by the susceptible-infected-susceptible (SIS) network model. Dependence is modeled by an undirected network where each node represents a firm, a system of computers, or a single device, and each edge represents a possible transmission channel. The authors apply tractable mean-field approximations for the Markov process and higher order polynomial approximations for claim functions, allowing computation of expected aggregate losses and hence cyber contract prices. Through a simulation study, the authors demonstrate that network topology plays an important role in determining insurance prices and designing risk management strategies.

Xu and Hua (2019) study cyber risk using the SIS network model with nonzero exogenous infection rates. In their modeling approach, the dynamics of an epidemic spread over the internet are captured by Markov and non-Markov processes, whereas risk dependence is addressed with copulas. In the Markov model, the following Poisson processes are used to capture infection and recovery for node  $v$ , respectively:

$$I_v(t) : 0 \rightarrow 1 \text{ at a infection rate of } \beta \sum_{j=1}^n a_{vj} I_j(t) + \epsilon_v,$$

$$I_v(t) : 1 \rightarrow 0 \text{ at a recovery rate of } \delta_v,$$

where  $\beta$  and  $\epsilon_v$  are the rates of infection due to threats inside and outside the network, respectively,  $I_j(t)$  is an indicator function which equals 1 if node  $j$  is infected at time  $t$  and 0 otherwise, and  $a_{vj}$  is another indicator function which equals 1 if nodes  $v$  and  $j$  can attack each other and 0 otherwise. From the Markov model, the authors obtain a dynamic upper bound for the infection probability and a stationary probability that can be used as a proxy to estimate infection probabilities in practice. On the other hand, in the non-Markov model, the authors assume that there exist  $D_v$  infected neighbors around each node  $v$  and that attacks on a node will stop if the node becomes infected. Time to internal infection is modeled by random variables  $Y_{v_1}, \dots, Y_{v_{D_v}}$  with the same marginal distribution, whereas time to external infection is described by another random

variable  $Z_v$ . The expected time to infection for node  $v$  is the expected minimum of time to internal infection from the  $D_v$  infected neighbors and the time to external infection. The authors further employ copulas to model the joint survival function in order to account for multivariate dependence among risks. Ultimately, an upper bound for the infection probability under the non-Markov model is also obtained. Their simulation shows dependence among infection processes affects the propagation of the epidemic and the resulting loss. Additionally, through experimentation with the Enron email network, the authors demonstrate that the recovery rate has a substantial impact on insurance premiums.

The Markov-based model developed by Xu and Hua (2019) is enriched by Antonio *et al.* (2021), who introduce a clustering coefficient factor into the SIS process with nonzero exogenous infection rates. The network structure is characterized by the individual-level clustering coefficients, which influence the efficiency of epidemic spreading. This inhibitory effect is described by incorporating different epidemic inhibition functions in the transition probability of the Markov process. The dynamic equation for the infection probability with clustering structure is obtained using  $N$ -intertwined mean-field approximation, and the dynamic upper bound is solved to be used as a conservative estimate in pricing. They adopt the cost function proposed by Xu and Hua (2019) to price cyber risk based on losses caused by infection and losses caused by system downtime. Experiments on simulated regular network and real email-Enron network validate the improvement in premiums when considering the clustering structure with inhibition.

Jevtić and Lanchier (2020) model cyber risk propagation in small- and medium-sized enterprises with a bond percolation process. In their model, the aggregate loss up to any given time is recorded as a continuous-time Markov chain with contagion times following a Poisson process, contagions in the physical layer are modeled with a homogeneous bond percolation process on a random tree that depicts the network infrastructure, and losses in each breached node are described by a heterogeneous loss topology. Using simulations, the authors demonstrate the robustness of their model in the context of cyber insurance pricing.

Chen (2019) investigates the differences between discrete-time and continuous-time epidemic models for modeling malware propagation and information dissemination. The author finds that real-life malware propagation is more aligned with discrete-time epidemic models, as worms usually take some time to spread. The author further focuses on the susceptible-infectious model and identifies three key drivers of model performance: time intervals, spatial dependence among nodes, and linearization. Small time intervals often lead to an overestimation of propagation speed and need to be accompanied by spatial dependence assumption. Consequently, continuous-time epidemic models do not provide accurate forecasts as they ignore both time intervals and spatial dependence.

Hillairet and Lopez (2021) propose a framework to design accumulation scenarios, each of which represents a global failure of the portfolios in question. The authors use the Gaussian approximation theory to approximate the evolution of the number of infected policyholders through time, and a susceptible-infected-removed (SIR) model to describe the spread of an attack at a global level. Specifically, for a population of size  $N_t = s_t + i_t + r_t$  at time  $t$ , where  $s_t$  is the number of susceptible individuals in the population,  $i_t$  is the number of infected individuals, and  $r_t$  is the number of recovered ones, the SIR model is characterized by the following set of differential equations:

$$\begin{aligned}\frac{\partial s_t}{\partial t} &= -\beta s_t i_t, \\ \frac{\partial i_t}{\partial t} &= \beta s_t i_t - \gamma i_t, \\ \frac{\partial r_t}{\partial t} &= \gamma i_t,\end{aligned}$$

where  $\beta$  and  $\gamma$  represent the contagion rate and the recovery rate, respectively. The centered processes of the cumulative number of the infected, the recovered, and the infected and not yet recovered converge in distribution toward Gaussian processes with time-dependent covariance structures. The authors hence derive the asymptotic distributions of the cost functions and illustrate through a simulation study that the insurer's response strategy has a crucial impact on the insurer's cost.

In the continuity of Hillairet and Lopez (2021), Hillairet *et al.* (2022) develop a multigroup SIR process to describe the contagion dynamic by segmenting the whole population into groups. The model further integrates digital dependence using a contagion matrix, which also contains information on the network topology formed by the multigroup population. Reaction to the crisis by taking countermeasures is also modeled, taking into account reactions to both in-group and out-group warnings. The simulation of a Wannacry episode is conducted, on the basis of model parameters that are calibrated from the OECD data. The simulation results show that the speed of infection and the final epidemic size in each sector are affected by the network connectivity patterns and group reaction strategies.

A growing interest in the use of epidemic models in cyber risk modeling can be observed in recent years, but the paucity of data limits the practical application and renders the field almost entirely theoretical. It is challenging, if not impossible, to obtain the building blocks of epidemic models, such as states, state transition probabilities, the evolution of the number of individuals in each state, and the underlying network structure. Consequently, current studies rely heavily on simulations.

### 3.6 Discussion

This section presents a review of methods used to model the empirical properties of cyber risk, where actuarial contribution is substantial. Cyber study within actuarial science has undergone a shift in focus toward more elaborate techniques including copula models, EVT approaches, stochastic processes, and epidemic models, to compensate for the insufficiency of the conventional approach using standard distributions in representing and pricing cyber risk. In particular, over a third of papers published within the past five years surveyed in this section are dedicated to epidemic models (Chen, 2019; Xu & Hua, 2019; Jevtić & Lanchier, 2020; Antonio *et al.*, 2021; Hillairet & Lopez, 2021; Hillairet *et al.*, 2022). These techniques have proven useful in drawing a more complete picture of cyber risk, but their implementation demands more comprehensive incident datasets. For example, the copula-based dependence structures and the EVT models derived from fitting the number of breached records (such as the PRC data) may be significantly different when applied to actual severity data. The issue of data scarcity is most prominent in the epidemic modeling approach, as the currently available cyber data is insufficient for identifying the network structure and inferring the state transition rates in an epidemic model. Validation of models based on simulation using real data can be an interesting path to take.

Different from risk prediction methods that focus on the prediction of risk at an individual level, risk modeling mostly considers losses at a collective level. Enterprise-level modeling is paramount to collective modeling, as it provides implications for risk management with respect to budget allocation, as well as the underwriting for cyber policies. Sparsity of enterprise-level data poses a challenge for individual loss estimation using models surveyed in this section. This difficulty might be surmounted through integrating statistical modeling approaches with prediction techniques surveyed in Section 2.

Lastly, interdisciplinary collaboration can be instrumental in advancing the mathematical and statistical modeling of cyber risk. Insights from computer science and engineering experts could help identify additional key factors in technological aspects that impact the resulting losses and risk dependencies. The identification of such factors could assist risk managers to better understand what risks are inherent in the system and may not be eliminated, thereby making

more informed insurance decisions. Moreover, in regard to data scarcity at both individual and collective levels, data mining could be leveraged to support numerical investigation. Effective interpretation and processing of the raw data obtained through mining often require collaboration between actuarial researchers and information technology specialists.

#### 4. Cyber Risk Management

The ultimate objective of cyber risk prediction and modeling is to formulate an effective framework to manage cyber risk. A report by Munich RE (2020) points out that in 2019 the global average loss was US\$4m per data breach event and US\$1m per ransomware attack. The evolution of 5G technologies, artificial intelligence, and cloud services increase the efficiency of attacks in terms of speed and scope. Attacker tactics are also expected to become more sophisticated and targeted in response to these new technologies. To keep in pace with the evolution of cyber risk, further endeavors in the development of cyber risk management schemes are demanded.

McShane *et al.* (2021) conduct a survey on cyber risk management methods from both technical and economic perspectives. The authors subsume cyber risk management methods into five categories, namely, avoidance, mitigation, transfer, retention, and exploitation, and identify the importance of integrating individual steps in the risk management process. In our review of cyber risk management methods, we take a slightly different approach in the sense that we separate different steps in the risk management process while layering the process according to the sequence that a cyber incident triggers responses from different stakeholders, including the external technology suppliers, managers, employees, and insurers of a victimized entity. The series of steps outlined in this section include risk assessment, risk mitigation, and risk retention. Risk assessment entails the detection and quantification of risk within the current cyber environment, drawing on expertise from the domains of computer science and engineering. The assessed risk may be mitigated through technological solutions, which are the primary focus of computer science and engineering experts (Cavusoglu *et al.*, 2004b; Liu *et al.*, 2016b; Krutilla *et al.*, 2021), or through operational practices that researchers in RMI are more interested in (August *et al.*, 2019; Böhme *et al.*, 2019; Egan *et al.*, 2019; Eling & Schnell, 2020; Eling & Jung, 2022). Risk remaining can either be retained or transferred, depending on the organization's risk appetite. The topic of risk retention has seen growing contributions from the domain of business and economics (Rosati *et al.*, 2017, 2019; Kamiya *et al.*, 2020).

With respect to managing residual cyber risks post-implementation of controls, cyber insurance has been posited as a risk transfer tool in some of the earliest papers. Grzebiela (2002) proposes using cyber insurance as an instrument to transfer risks associated with integrity and availability violations. Gordon *et al.* (2003) outline a generic framework for using cyber insurance to manage information risk. On top of insurance, Romeike (2000) argues that the limitation of insurance coverage may be resolved by applying alternative risk financing products. Section 5 of this paper provides an exposition on the current status of the market, pricing methodologies, and challenges confronting cyber insurers.

##### 4.1 Risk assessment

Risk assessment is a crucial step in any sort of risk management. Conventionally, risk is assessed by the probability of the risk event and the magnitude of the loss event (Kaplan & Garrick, 1981). This way of risk assessment is referred to as the pure premium approach in an insurance context. However, this risk assessment approach might not be adequate for cyber risk, as the risk is not straightforward to be modeled mathematically. Therefore, alternative risk assessment methods for cyber risk are in demand.

Mateski *et al.* (2012) design a generic threat matrix which assigns a threat level between 1 (most capable) to 8 (least capable) against some attributes to different categories of cyber threats. The

attributes used to score a threat are related to the commitment level and the resources that are readily available to pursue the goal of the threat.

Schatz and Bashrouh (2017) divide economic approaches to risk assessment into nine categories: analytic hierarchy processes,<sup>4</sup> decision support systems, game theory, net present value (NPV), return on attack,<sup>5</sup> return on investment (ROI),<sup>6</sup> a mix of ROI and NPV, real options theory,<sup>7</sup> and utility maximization. Both ROI and NPV approaches emphasize the ‘benefit and cost’ element, whereas game theory approaches exhibit a high reliance on the ‘function’ element.

Akinrolabu *et al.* (2019) propose a novel risk assessment method called the cyber supply chain cloud risk assessment (CSCCRA) for cloud service providers. The CSCCRA method consists of three components: a cloud supply chain mapping that graphically represents the cloud service through the lens of the supply chain, a cloud supplier security assessment that provides a quantitative measurement of cloud suppliers security according to certain criteria, and a cloud quantitative risk analysis that estimates a probability distribution for the variable of interest.

Yamashita *et al.* (2020) construct a framework to measure the systemic risk of switching attacks in a cyber-physical system based on security technologies deployed using Petri net models. The framework uses four models to depict the coordinated attack against IP-based substations: modified firewall model, modified password model, extended password model on intelligent electronic device, and honeynet models. The detailed statuses and state transitions within each component are described by Markov processes, and the steady-state probabilities for a cyber-physical attack at any substation are derived as a means to assess security risks.

Besides cyber losses resulting from cyber events, non-cyber losses triggered by cyber activities that are not explicitly excluded from insurance clauses should also be considered. This type of risk is referred to as non-affirmative risk, or “silent” cyber risk by insurance practitioners (Lemnitzer, 2021). Recognizing the fact that insurers often fail to adequately address their silent exposure to IT-related threats when pricing non-cyber contracts, Cartagena *et al.* (2020) propose a three-phase framework that aims to provide a consistent non-affirmative cyber risk assessment prototype for organizations. In the first phase, underwriters define exposure, evaluate wordings of clauses, and review policy levels. In the second phase, the organization contextualizes the defined exposure into possible scenarios to establish a comprehensive understanding of non-affirmative risk it encounters. In the last phase, the management develops a risk appetite relating to non-affirmative risk and monitors the implementation of the framework.

In addition to technological vulnerability, human factors may also give rise to cyber risk. Georgiadou *et al.* (2022) identify insider threat types and contributing factors through a meta analysis, and link the contributing factors to a cyber security culture framework for assessing and evaluating the current security readiness of an entity’s workforce. The framework encompasses an organizational level and an individual level. The former considers dimensions including assets, continuity, access and trust, operations, defense, and security governance, whereas the latter includes dimensions such as attitude, awareness, behavior, and competency.

Ganin *et al.* (2020) present a multi-criteria decision-analysis-based approach to bridge the gap between risk assessment and risk management pertaining to cyber security. The criteria include three components: threats, vulnerabilities, and consequences. The first component describes the ease of attack and benefits of a successful attack. The second component considers vulnerabilities in the targeted system, including the hardware, software, and personnel-related levels. The third component considers consequences that pertain to violations of confidentiality, integrity, and availability. Apart from risk assessment, the authors implement a countermeasure scoring system that could assess the effectiveness of risk management.

<sup>4</sup>An analytic hierarchy process is a structured method that aggregates sub-problem solutions into a conclusion (Schatz & Bashrouh, 2017).

<sup>5</sup>A return on attack measures an attacker’s gain and loss (Schatz & Bashrouh, 2017).

<sup>6</sup>A return on investment indicates the efficiency of a security investment (Schatz & Bashrouh, 2017).

<sup>7</sup>Real options theory goes through flexibility evaluation in the decision making process (Schatz & Bashrouh, 2017).



## 4.2 Risk mitigation

Organizations may take technical and internal controls to reduce the likelihood of cyber security breaches and/or constrain the severity of losses from attacks. Böhme and Schwartz (2010) describe the probability that a node in an organization incurs a loss in an accident by a function of the network environment and the security level. Their model implies that the effectiveness of a firm's technical controls over its network environment is directly related to its exposure to cyber risk. Technical controls can be either detective or preventative. The former alarms users of security violations, whereas the latter proactively defends specific vulnerabilities from attacks (Cavusoglu *et al.*, 2004b).

An organization may not necessarily have control over technical vulnerabilities, because software is often supplied by external vendors (Böhme *et al.*, 2019). However, there are some actions that can be taken within the organization's reach, such as frequent backup of information, installation of security technology, and constant renewal of antivirus systems (Liu *et al.*, 2016a). An effective control technique is to install software that is designed to detect and prevent attacks. For example, Liu *et al.* (2016b) propose a detection scheme named ActiveTrust that can effectively avoid black hole attacks, which can result in incoming and/or outgoing traffic being silently discarded. Furthermore, an appropriate use of technology is as important as an installation of relevant systems. Bilge *et al.* (2017) find that in general, software and files signed by rarely known vendors as well as rarely updated machines tend to have a higher probability of malware infection.

The amount of investment on technical controls is also of great substance. Cavusoglu *et al.* (2004b) use a game theoretic approach to describe the strategic investment decisions of an organization. Their results suggest that given the quality parameters of each technology and the organization-specific parameters, an organization should select the security technology that maximizes the cost saving, which is defined as the difference between the cost with the technology chosen implemented and the cost without security technology. This conclusion is in line with that of Gordon and Loeb (2002), which indicates that the optimal level of security investment is where the difference between benefit and cost is maximized. Gordon and Loeb (2002) also demonstrate that for a given threat level, the optimal investment in information security does not necessarily increase with the system's vulnerability and the amount generally does not exceed 37% of expected loss.

Krutilla *et al.* (2021) provide a dynamic extension of the investment model formulated by Gordon and Loeb (2002), incorporating discounting and depreciation effects in the benefit–cost analysis of cyber security investment decisions over the long run. The economically efficient maximum level of cyber security capital in the dynamic setting is found to be  $\frac{37\%}{r+\delta}$ , where  $r$  is the discount rate and  $\delta$  is the depreciation rate. On the contrary, Maillart and Sornette (2010) come up with a view that all entities are evenly vulnerable irrespective of their level of information security. They advocate the size effect, which states that the severity of a breach is largely affected by the size of the organization in question.

Risk mitigation from the suppliers' side is considered by Böhme and Kataria (2006) and August *et al.* (2019). Böhme and Kataria (2006) suggest that the government should promote competition, because an increasing diversity of software products could effectively reduce both local and global correlation of cyber risk. August *et al.* (2019) point out that the majority of users do not update their systems in a timely manner because the deployment of system patches is not an economically optimal option for them. They argue that a potential incentive structure is for suppliers to charge for patching rights, so that consumers who elect to relinquish patching rights and have their systems automatically patched pay less than those who choose to retain the rights. Such a price differentiation may reduce the unpatched population, thereby nurturing a safer network environment.

Restraining the impact of cyber crises with technical controls should be complemented by inputs from the management and operational level. Cyber risk management in enterprises ranges

from a single centralized department to an all-level framework. An organization can take a top-down or bottom-up approach to manage cyber risk, depending on whether the organization focuses on the risk management process or risk identification and quantification (Böhme *et al.*, 2019). Stoneburner *et al.* (2002) highlight the importance of technical training, risk ownership, periodic reviews, and audits in the prevention of security breaches on the management level. Controls on the operational level, such as regular data backup and physical securities, are overseen by the management level.

Organizational control in insurance companies is investigated by Egan *et al.* (2019), who develop a framework based on the Chief Risk Office Forum cyber incident taxonomy (CRO Forum, 2016) and the National Institute of Standards and Technology framework (National Institute of Standards and Technology, 2014). The authors carry out three case studies that are representative of the current risk landscape of the industry: insider leaks at a general insurer, cyber extortion at a life insurer, and telematics device hack at a motor insurer. They determine that protective actions should be taken on information assets to prevent successful insider attacks, cyber extortion can be mitigated by installing defense and detection mechanisms, and hacking attacks could be eliminated by constant upgrading, access monitoring, and regular security testing.

Eling and Schnell (2020) take a different perspective on cyber risk mitigation by considering the capital requirement regime within the applicable regulatory framework. The authors consider three regulatory capital models: Solvency II, U.S. risk-based capital standards, and Swiss Solvency Test. They find that regulatory models in general underestimate cyber risk since they do not reflect heavy tails and interdependence of the risk appropriately. All three regulatory models exhibit insufficient capital requirements to ensure solvency, particularly for small cyber insurance portfolios and high policy limits. However, the authors argue that increasing capital requirements may hinder the growth of an immature cyber insurance market rather than maintaining solvency of insurers. They therefore suggest building sufficiently large portfolios in developing markets and increasing capital requirements in developed markets.

Eling and Jung (2022) also examine the capital implications for firms when considering cyber risk as part of their operational risk frameworks. They analyze the cyber loss data in the financial industry documented in the SAS OpRisk database. Specifically, they fit a Tweedie model to the data and subsequently perform a regression analysis through the generalized linear model to determine the key drivers of cyber loss severity. The results show individual heterogeneity, particularly concerning firm size, interconnectivity, and legal liability level, could largely impact the capital requirement of a financial institution. As a consequence, the authors suggest firm-specific risks should be reflected in insurance prices as well as capital requirement, and a loss-distribution-based approach is better suited for achieving this goal.

### 4.3 Risk retention

Risks that cannot be reduced or transferred are retained by organizations. Hence, in addition to preventive measures, responsive schemes are important in containing the aftermath of a cyber crisis. Cavusoglu *et al.* (2004a) find evidence that within two days of a security breach announcement, the breached firm faces an average decrease of 2.1% in market value and an average loss of \$1.65 billion in market capitalization. Similar views are held by Campbell *et al.* (2003), Spanos and Angelis (2016), Rosati *et al.* (2017) and Rosati *et al.* (2019). Furthermore, Kamiya *et al.* (2020) discover that a successful external attack could affect the entire industry if it reveals personal financial information. Their finding reinforces the indispensability of post-incident controls.

Knight and Nurse (2020) construct a corporate communication framework to limit the loss arising from a data breach incident. The proposed decision making process for an organization has two stages: before crisis and after crisis. Before a data breach incident, the organization should establish the goals they aim to achieve post-breach, maintain the relevant knowledge base, involve corporate partners in the process, and ensure security basics are in place. When a crisis happens,

the organization has to decide whether, what, when, and how to disclose the breach. The authors suggest that accepting responsibility rather than pointing fingers at external parties can reduce reputational damage. The organization should also cater to different demographic groups by making their announcement easy to understand. Although there exist conflicting views that suggest the market does not necessarily penalize firms victimized by cyber attacks (Hovav & D'Arcy, 2003, 2004), the framework provides practical guidelines on damage control.

A more comprehensive system considering preventive, supportive, and responsive measures is proposed by Onwubiko and Ouazzane (2020), who argue that there are three key elements in an effective ERM framework: incident governance command, incident sharing, escalation and reporting, and incident management. The first element involves a hierarchical structure that allocates responsibility to different management levels. In the second element, organizations are recommended to establish a sharing partnership with all sectors to ensure consistent communication. The third component is consonant with the second, emphasizing the fact that enterprise cyber risk management is a cooperative task rather than independent work.

#### 4.4 Discussion

In this section, works pertaining to the assessment, mitigation, and retention of cyber risk have been reviewed. The surveyed works approach the topic from different viewpoints, including those of technology suppliers, organizational users, and insurance providers. Based on the review, several gaps in cyber risk management research are identified.

First, we notice a lack of works on the assessment of the spillover effect of cyber risk. A firm may face cyber threats or endure cyber losses as a result of a security breach in another organization. The risk arises from associations with external parties, which cannot be measured internally within the entity. Therefore, it is imperative to design a risk assessment framework at the industry level to provide an indicative estimation. The process will require industry cooperation that may be coordinated and overseen by authorities.

Second, the quantification of the impact of risk mitigation measures is understudied. A fundamental contribution is made by Gordon and Loeb (2002), who develop a mathematical model to measure the loss reduction resulting from technical security investments and determine the optimal level to invest. Their work is among one of the most referenced papers about cyber risk mitigation, and it continues to be adopted and enriched in recent papers (Bentley *et al.*, 2020; Dou *et al.*, 2020; Krutilla *et al.*, 2021). Testing of these quantitative models using real data, and formulating a more comprehensive framework that incorporates a broader range of security enhancement tools and risk reduction metrics, are interesting directions to follow.

Lastly, integrating cyber risk management into the existing ERM framework is a pressing issue (McShane *et al.*, 2021). The integration is complicated by the underdevelopment of reliable quantification methods for estimating the likelihood and magnitude of the risk when compared to other operational risks. More accurate measurement of cyber risk can be achieved by using prediction and modeling techniques surveyed in Sections 2 and 3. In turn, these models can benefit from the integrated risk management framework, in a way that the framework helps identify risk determinants from the management side. Additionally, the interdependence between cyber risk and other components of the ERM framework poses another challenge. For instance, credit risk could be affected in the case of a security breach, even though it is listed as a separate element (Stine *et al.*, 2020). Such interdependence should be thoroughly considered and addressed when adapting the ERM framework to incorporate cyber risk.

## 5. Cyber insurance

The first standalone cyber insurance was introduced in 1998 by the International Computer Security Association (Marotta *et al.*, 2017). Since then, the cyber insurance market has been

developing gradually. Still, unlike conventional insurance, the cyber insurance market is far from being efficient due to a lack of volume and liquidity (Anderson & Moore, 2006). The purchase of cyber insurance has not become common because organizations are set back by wide variations of coverage and policy terms, high costs, and limited coverage (Marotta *et al.*, 2017; Xie *et al.*, 2020).

Although the global cyber insurance market is still in its embryonic stages, many authors envisage the potential for the industry. Biener *et al.* (2015) argue that cyber insurance has virtues of creating incentives for risk-appropriate behavior and raising organizations' awareness of cyber threats. Marotta *et al.* (2017) believe that cyber insurance may serve as an indicator of the quality of security protection and propel the advancement in standards regulating cyber security. They also argue that cyber insurance encourages organizations to invest more in information security, thereby improving social welfare through the positive spillover effect. A different view by Baker and Shortland (2022) posits that insurance can function as a form of regulation or governance by providing *ex post* risk reduction support, while preserving policyholders' autonomy in security decisions. From insurers' viewpoint, Cole and Fier (2020) assert that cyber insurance is a profitable line of business.

### 5.1 Status quo

Verizon (2021) reports a total of 29,207 incidents in 2020, among which 5,258 were confirmed data breaches that were mostly caused by external actors. The entertainment sector was the most targeted industry, followed by the public administration sector (Verizon, 2021). The rising number of incidents is accompanied by an expansion of the cyber insurance market. The number of cyber insurance providers increased from less than 50 in 2015 to more than 150 in 2018 (Daffron *et al.*, 2019). Munich RE (2020) envisions that the size of the global cyber insurance market will reach US\$20 billion in 2025, with strong growth expected in Asia and Europe.

However, the rise in supply does not translate to a mature market. Although the insurance industry perceives the new product category as lucrative and abundant in growth opportunities, insurers and reinsurers do not strive their utmost to promote the sector due to uncertainty about their knowledge in the risk and its drivers (Lemnitzer, 2021). According to Böhme *et al.* (2019), the underdevelopment of the cyber insurance market is chiefly ascribed to the lack of demand and claims in the 1990s. The lack of claims is an indication of cumulative risk, also known as "cyber hurricane," causing reinsurers to cease providing protection for cyber insurers in the early 2000s (Böhme & Schwartz, 2010).

To gain insights into the industry, Marotta *et al.* (2017) survey cyber insurance policies available on the market in 2017. The authors outline the approaches used by insurers and group them according to their emphases: risk/security level specification and game theoretic approaches for premium specification. They also identify the peculiarities in underwriting and claim handling processes that set cyber insurance apart from conventional insurance policy writing.

Romanosky *et al.* (2019) perform a thematic analysis of 235 cyber insurance products that are filed with state insurance commissioners in New York, Pennsylvania, and California from 2007 to 2017. The authors compare the coverage and exclusions, risk assessment processes, and premium determination methods among the products under consideration. They find that most insurance policies are consistent in terms of coverage, but different sublimits are often applied to different coverage areas. More variations are observed in the policy exclusions. In line with the findings of Gordon *et al.* (2003), most of the cyber insurance policies under consideration cover first-party<sup>8</sup> and/or third-party<sup>9</sup> losses arising from a firm's cyber-related operations. The authors find that during the process of underwriting, insurance providers are mostly interested in the amount and type of data, particularly sensitive data related to debit and credit card transactions, managed by

<sup>8</sup>First party coverage covers losses incurred directly by the insured in the incident (Romanosky *et al.*, 2019).

<sup>9</sup>Third party losses incur as a result of litigation by alleged injured third parties (Romanosky *et al.*, 2019).

an organization; little attention is given to the technical infrastructure, the business environment, and the security budget. However, the sample may not be representative for the current condition of national and global cyber insurance markets. A lack of regulatory standards and guidelines governing capital requirements is another impediment for market growth (Eling & Schnell, 2020).

## 5.2 Pricing cyber insurance

Underwriting risk is one of the most significant risks underlying cyber insurance products for insurers (Eling & Schnell, 2016). To maintain solvency and profitability, an insurer should always charge adequate premiums in the first place. On the other hand, an overly high premium may not be economically effective for the insured organization (Böhme *et al.*, 2019). Due to a lack of reinsurers and heavy tails of cyber losses, extant cyber insurance products tend to charge high premiums to compensate for the extreme risk, thereby discouraging entities from buying cyber insurance. To establish a balance between adequate premiums and business volume, reliable pricing methods for cyber insurance are demanded.

Böhme and Kataria (2006) model the insurer's cost in a single period by the aggregate amount of the expected loss, sum of all administrative costs, and the interest on the safety capital required to settle all claims in the worst-case scenario considering ruin. They show by simulation the viability of cyber insurance in managing risks with high internal and low global correlation.

Herath and Herath (2011) propose to model the net premium for the first party loss excluding profits and expenses by the discounted value of the sum insured multiplied by the probability of a cyber incident. The discount period is the time between policy issuance and security breach, modeled by a Poisson process. In the absence of historical records on frequency of claims or annual claims paid by the insurer, the authors propose an integrated copula-based simulation algorithm to price the first-party loss.

Romanosky *et al.* (2019) identify five main factors in insurance pricing: external sources, estimation, competitor behavior, experience, and reference to prices from other insurance lines. Out of the policies surveyed by the authors, insurers typically price their policies either on a flat rate basis or using the base rate approach. The flat rate premium can deviate across different groups of risks, whereas the base premium is calculated as a product of a firm's annual revenue or assets and modification factors reflecting the firm's risk level. For both approaches, the firm's asset value or revenue seems to be the most important determinant of insurance premiums, as it can be regarded as a proxy for the firm's size and hence risk level.

Utility-based approaches are often used in insurance pricing. Carfora *et al.* (2019) estimate cyber insurance premiums using nonlife insurance pricing principles and assess the insured's acceptable prices using the indifference principle in utility theory. In a similar vein, Böhme *et al.* (2019) suggest that from a utility-theory perspective, with optional market insurance, a risk averse firm will choose an optimal security investment  $s^*$  and a sum insured  $x^*$  to maximize its expected utility. An insurer would price their policies such that the premium to charge for losses up to a limit  $x$  is  $\pi x$ , where  $\pi = D(s^*)(1 + \lambda)$  and

$$(s, x)_\sigma^* = \operatorname{argmax}_{s, x} \left( D(s)U_\sigma(w - l - s + (1 - \pi)x) + (1 - D(s))U_\sigma(w - s - \pi x) \right).$$

In the formula above,  $\sigma > 0$  reflects the risk aversion of the firm,  $U_\sigma(\cdot)$  is the utility function,  $w$  is the firm's initial wealth,  $D(s)$  is the defense function mapping a security investment  $s$  to a probability of loss,  $l$  is the monetary loss, and  $\lambda$  is the loading factor. Dou *et al.* (2020) also design an insurance pricing scheme that is based on expected utility theory.

Eling *et al.* (2022) point out that an aspect of cyber risk that traditional pricing principles do not consider is the heterogeneous relationship between the cost of cyber risk events and firm-specific factors. The authors unravel such heterogeneity through a quantile regression model, which can capture non-central locations of the cost distribution that the ordinary linear regression method cannot, on the severity data provided by Cowbell Cyber Incorporation. The dataset

used contains firm-specific security information that enables identifying the effect of individual security measures on cyber costs. Compared to insurance prices computed from a Tweedie model and a two-part GLM, quantile-based pure premiums are generally larger as they embed the heterogeneous impact of firm size, industry and security level.

Lau *et al.* (2020) argue that traditional actuarial insurance principles are inadequate in dependence consideration that could expose insurers to insolvency situations. They propose a probabilistic and game-theoretic approach to calculate premium principles. A Stackelberg security game, in which the system owner as the defender moves first, and the attacker follows, is devised to obtain the optimal stochastic allocation of defense resources across target substations. The optimal defense resource allocation is used to evaluate the potential monetary loss of a cyber attack. Ultimately, cyber insurance premiums via value-at-risk or tail value-at-risk based on total loss are derived and then allocated to individual transmission companies, thus mitigating the insolvency risk.

Yang *et al.* (2020) propose to price cyber insurance for cyber-physical power systems considering ruin of the insurer. The probability distribution of the claim size, which incorporates the expected mean time to restore power to normal steady-state conditions for each substation, of hypothesized substation outage is computed based on the cyber-reliability assessment. The ruin probability of the insurance company is formulated using the distribution of claim sizes and a lump sum premium calculated from the expectation principle, which employs a feasible loading factor identified in the ruin probability calculation.

Another premium structure that could mitigate insolvency issues is designed by Lau *et al.* (2022) for protecting power systems against cyber attacks. A stochastic epidemic network model tailored to the reliability-based load loss estimation is built to describe the risk spreading in a cyber-physical system. The reliability results are used in the estimation of the Shapley mutual insurance premium principle, which utilizes the solution of a Shapley cooperative game to guarantee a lower cost with mutual insurance than the cost without. The affordability and efficiency in restraining insolvency risk are verified in case studies.

Most policies only cover first- and third-party losses. Böhme and Schwartz (2010) point out that not covering secondary losses<sup>10</sup> could be problematic, as firms may choose not to disclose breaches if the expected secondary costs exceed the sum insured for primary losses. Bandyopadhyay and Mookerjee (2019) expand the scope of cyber insurance policies to include the impact of secondary losses. They observe that a cyber event often comes with a significant damage to reputation and loss of investor confidence, but these secondary losses are often excluded from the policy coverage. The insured may choose to abstain from claim disclosure in fear of reputational loss, a behavior that is described as a hidden action problem of information asymmetry by Moore (2005). In this case, the operationalized deductible (the contracted deductible plus the secondary loss) is higher than the contracted deductible. The authors argue that incorporating this off-contract hidden behavior can reduce overpricing and hence improve the efficiency of a cyber insurance contract.

While the mainstream research community considers risk dependence as a major obstacle for cyber insurance, Khalili *et al.* (2019) propose an alternative view. The authors suggest that the best strategy for a cyber insurer is to insure both the service provider and its customers. Rather than applying the base premium approach that is widely adopted by cyber insurers, they develop their own risk-adjusted model which contains an incentive factor. On the basis of this model, a service provider is offered a premium discount (a larger incentive factor) if it agrees to invest more in its security posture. The probability of breach decreases with the incentive factor, an outcome that directly affects all of the service provider's customers. As a result, a cyber insurer may enjoy a greater reduction in risk if it insures both the service provider and its customers at the same time.

<sup>10</sup>A secondary loss is the indirect consequence of a cyber incident, such as a reputational loss and reduction in share prices (Marotta *et al.*, 2017).

The authors argue that their proposed approach not only increases the insurer's profit but also enhances social welfare as a result of a safer network overall.

### 5.3 Challenges

Multiple concerns have been raised regarding the development of the cyber insurance industry. Gordon *et al.* (2003) underline moral hazard and adverse selection faced by insurers offering cyber insurance. Böhme and Schwartz (2010) suggest insurers' lack of experience and insufficient historical data for competitive premium pricing hinder the development of the cyber insurance market. Biener *et al.* (2015) assess the insurability of cyber risk on the basis of the insurability criteria set out by Berliner (1982). They identify data scarcity, risk dependence, moral hazard, adverse selection, and insufficient insurance coverage as factors that may impair the insurability of cyber risk. Marotta *et al.* (2017) also find that randomness of loss occurrence, information asymmetry, and coverage limits are the most significant issues hampering the development of the cyber insurance market. The dual impact of cyber risk on cyber insurers through their channels of policy underwriting and IT system operations is another handicap to overcome (Eling, 2020). In what follows, we discuss three most recognized challenges: information asymmetry, underinvestment in self-protection, and risk dependency.

#### 5.3.1 Information asymmetry

The unbalanced information between insurance providers and policyholders is particularly discernible in the cyber insurance market, since firms are reluctant to share the details of their security structures and security breaches (Marotta *et al.*, 2017). Information asymmetry can give rise to moral hazard, a problem that arises due to the insured's inclination to reckless behaviors after purchasing protection.

Anderson and Moore (2006) argue that moral hazard in cyber security originates from misaligned incentives, which are at the expense of insurers. The authors suggest that the liability of managing cyber threats should be assigned to the party that is the most capable of the task, but the misaligned incentives between parties give rise to a poor allocation of risk management. Another common issue in cyber security lies in the information asymmetry between consumers and software vendors, which deters developers from thorough code checking and security testing since the unmeasurable quality at a cost of premium does not appeal to buyers. They argue that software providers should hire fewer but more competent programmers and more software testers.

Dou *et al.* (2020) argue that pure insurance contracts cannot encourage the insured to increase self-protection investment if there is no moral hazard. To address this issue, the authors investigate how entities may be incentivized to invest in self-protection through a mathematical model based on the utility theory. They further propose an optimal cyber-insurance contract scheme that maximizes the expected utility of users.

To cope with the lack of transparency, Panda *et al.* (2019) propose a Bayesian game with incomplete information to model the strategic interaction between the insured and the insurer and to identify the misrepresentation of security posture by policyholders. The model possesses audit power that addresses the discrepancy between the applicant's reported security level and the applicant's compliance with the policy terms *ex post* insurance purchase. The authors apply the indifference principle of utility to determine the optimal auditing strategy for an insurer in the event of a breach.

#### 5.3.2 Entity's security investment

Meland *et al.* (2015) claim that firms typically use a two-pronged approach to manage IT security risk: they first invest in security technologies and then purchase cyber insurance to cover the residual risk. A firm's security level affects its risk exposure and ability to withstand attacks.

Consequently, it affects the risk faced by its insurer. Firms tend to underinvest in cyber security due to a lack of understanding of the security interdependence arising from physical and logical interconnection of computers (Ögüt *et al.*, 2005). The authors find that mechanisms such as fines and information sharing can effectively encourage firms to invest at the socially optimal level.

A mantra in economics is that competition drives prices down. However, Ögüt *et al.* (2005) argue that a more competitive cyber insurance market does not necessarily lead to lower premiums or a higher coverage, because as prices are driven down, firms may opt for less costly cyber insurance in lieu of investing in self-protection. Such behavior in turn raises the riskiness of insurers' portfolios, an effect that will ultimately drive insurance prices back up.

Ögüt *et al.* (2011) find that firms underinvest in self-protection because of the difficulty of proving losses from a cyber incident. While Gordon *et al.* (2003) believe that cyber insurance is complementary to self-protection, Ögüt *et al.* (2011) reveal that self-protection and cyber insurance are only complementary if self-protection is observable to an insurer, whereas nonobservable self-protection behaves as a substitute to cyber insurance. Firms tend to invest less in both self-protection and insurance when the protection behavior is observable to insurers. On the other hand, if the behavior is unobservable to insurers, firms invest less in self-protection but purchase more insurance.

Apart from an entity's own investment strategy, public policies play an important role in facilitating a safer network. The work of Srinivas *et al.* (2019) sheds light on the importance of standardization, regulation, and government intervention in cyber security. Regulatory mechanisms such as fines may encourage the insureds to invest more in self-protection (Marotta *et al.*, 2017).

### 5.3.3 Risk dependency

While conventional insurance can provide protection for weakly correlated risks such as automobile accidents, the high correlation exhibited by cyber risk poses a significant challenge for the development of cyber insurance (Ögüt *et al.*, 2005; Biener *et al.*, 2015; Eling & Schnell, 2016; Marotta *et al.*, 2017). Ögüt *et al.* (2005) discover that interdependence of risk leads firms to underinvest in cyber security and buy less insurance. Additionally from another perspective, the positive externalities caused by interdependence of cyber security facilitates free riding, meaning that firms try to take advantage of other market participants' security postures. While self-protection features positive externalities, Hofmann and Rothschild (2019) suggest that purchasing insurance does not induce positive spillovers, making it harder to assess the efficiency of insurance contracts as well as self-protection investments.

Eling (2018) pinpoints the importance of considering the macroeconomic impact of cyber insurance. As IT systems are increasingly connected, an excessive issuance of cyber insurance can lead to systemic risk. Monoculture in computing technologies contributes to dependence of cyber risk, in a way that if the vulnerabilities of a system or software are identified by malicious attackers, all of its user nodes fall under threats and consequently the risk level for insurers is significantly increased (Bandyopadhyay & Mookerjee, 2019). Nevertheless, Khalili *et al.* (2019) hold an opposite view that risk dependency can be leveraged to enhance the overall risk reduction when the insurer provides protection to both service providers and their customers.

## 5.4 Discussion

In this subsection, works on pricing mechanisms and the cyber insurance market are reviewed. Although the subject of insurance is rooted in actuarial science, inputs from the realm of computer science are found to be substantial. There is an agreement among the referenced papers that classical premium structures based on mean and variance of losses are inapplicable to cyber risk. In the past five years, there is a growing number of publications focusing on the development



of more advanced pricing principles, including utility-based models (Böhme *et al.*, 2019; Carfora *et al.*, 2019; Dou *et al.*, 2020), game theoretic approaches (Lau *et al.*, 2020, 2022), and premiums with an incentive factor (Khalili *et al.*, 2019). Notwithstanding the considerable scholarly attention it has received, cyber insurance research is still in its formative stages. Despite the need for more accurate risk quantification methods, potential directions of future study on cyber insurance are discussed here.

Certainly, government intervention could play an important role in facilitating the growth of the cyber insurance market. Similar to legislative regulation of data breach notification, government oversight may be needed in data collection and data sharing in the industry. Implementing standardized protocols for information gathering in the insurance underwriting process and establishing a knowledge sharing platform for insurers could foster a more transparent and competitive cyber insurance market. Furthermore, regulatory efforts could be made to reduce moral hazard through mandating minimal standards of risk mitigation (Eling & Schnell, 2016). Such regulatory intervention has the potential to alleviate information asymmetry between insurers and policyholders, promoting a healthier insurance environment and more efficient market.

Additionally, the dual impact of cyber risk on insurers has not been adequately investigated (Eling, 2020). Cyber risk insurers face cyber risk arising from not only their policyholders but also their own information systems. The dual impact of cyber risk requires a careful assessment of insurers' security investment and capital requirement. In future research, it would be interesting to investigate how the dual impact of cyber risk may be mitigated through some sort of governmental support (e.g., provision of coverage for the cyber risk entailed in insurers' internal systems) or reinsurance mechanisms.

## 6. Conclusion

In this paper, we take an interdisciplinary approach to review works on cyber risk prediction, quantification, management, and insurance, with a goal to inspire future research that can be applicable to actuarial practice and insurance businesses. Overall, we find that contributions from the computer science world account for a large part of the literature, while actuarial input is on the rise. Machine learning applications, more sophisticated modeling techniques, and more advanced premium principles continue to appeal to scholars and industry practitioners from different backgrounds. Through the review, we identify a number of research gaps that are common across different themes. These research gaps are summarized in the rest of this section.

First, a major challenge in the study of cyber risk resides in the lack of reliable loss data. Many numerical studies in the existing literature are based on simulated (fictitious) data, but results derived from real data are more relevant to real-life applications. Most of the previous cyber risk studies that are conducted using real data are based on cyber incident data from the U.S., because of a lack of records in other regions. The reliance on U.S. data means that these studies may not be representative to other parts of the world. Geographical differences may be unearthed if more data becomes available globally. Furthermore, in existing data sets, some aspects of incidents are not properly recorded due to the difficulty of collection or reluctance of victims to disclose. For example, the number of employees or annual revenue of the breached entity is often used as a proxy for the size of infection (number of devices infected). The relevance of past data to future situations is also challenged by the dynamic nature of cyber risk. This problem is exacerbated by delays in disclosure. As data breach notification laws are now enforced in many jurisdictions, it is likely that additional data sets will become available in the future. In future research, it is warranted to revisit some of the cyber risk modeling and prediction problems with such data sets.

Second, while primary losses of cyber risk have been fairly extensively studied, little effort has been made on investigating secondary losses. The indirect economic damage of a cyber incident can constitute a significant proportion of the total loss, especially in the event of a zero (primary) loss breach. These losses are even harder to document and manage than primary losses,

as they generally involve a loss of reputation and subsequent reduction of stock prices. In light of the quantification complexity, insurers usually exclude secondary losses from their policy coverage, a business decision that may disincentivize organizations from purchasing insurance. We believe that mandatory data breach notifications would facilitate the collection of secondary loss data, and that an interesting direction following data collection is to explore the correlation between secondary losses across an industry and the practicability of including these losses in cyber insurance.

Third, management of catastrophic cyber events should be explored in more depth. Insurance on its own may not provide sufficient coverage for catastrophic cyber losses, such as cyber warfares and attacks on national infrastructures. Risk transfer methods for insurers, including reinsurance and cyber-linked derivatives, are nearly an unexplored field. The underdevelopment of cyber reinsurance manifests reinsurers' fear of excessive losses, which may require new reinsurance structures to limit their risk exposure. Alternative avenues of investment in the form of cyber-linked derivatives include options, vanilla options, swaps, and futures (Pandey & Sneekenes, 2016). Catastrophe bonds designed on EVT approaches are developed by Xu and Zhang (2021) and Liu *et al.* (2021) to transfer tail risks faced by cyber insurers to the capital market. Despite these attempts, the viability of using these financial instruments to manage cyber risk demands significant further research.

Finally, more interdisciplinary collaboration is clearly needed to resolve the limitations of current cyber studies. Collaborative efforts from different disciplines, not limited to computer science, business studies, risk management, and actuarial science, present an interesting research area. Risk modeling and management must adapt to the rapidly evolving landscape of cyber threats, which requires inputs from computer scientists and engineers. On the other hand, identification of technical vulnerabilities may need insights from a risk management perspective. Furthermore, the use of insurance as a tool to manage cyber risk should be explicitly accounted for when designing the risk management framework. Another potential area of interdisciplinary research is the study of systemic cyber risk arising from the connectivity of networks, possibly by drawing on knowledge from various fields, including social behavior, biological science, and information security.

## References

- Aas, K., Czado, C., Frigessi, A. & Bakken, H. (2009). Pair-copula constructions of multiple dependence. *Insurance: Mathematics and Economics*, **44**(2), 182–198.
- Advisen (2022). *Cyber loss data*. <https://www.advisenltd.com/data/cyber-loss-data/>
- Ahmed, A. A. & Zaman, N. A. K. (2017). Attack intention recognition: A review. *International Journal of Network Security*, **19**(2), 244–250.
- Akinrolabu, O., Nurse, J. R., Martin, A. & New, S. (2019). Cyber risk assessment in cloud provider environments: Current models and future needs. *Computers & Security*, **87**, 101600.
- Anderson, R. & Moore, T. (2006). The economics of information security. *Science*, **314**(5799), 610–613.
- Antonio, Y., Indratno, S. W. & Saputro, S. W. (2021). Pricing of cyber insurance premiums using a Markov-based dynamic model with clustering structure. *PLoS ONE*, **16**(10), e0258867.
- August, T., Dao, D. & Kim, K. (2019). Market segmentation and software security: Pricing patching rights. *Management Science*, **65**(10), 4575–4597.
- Awiszus, K., Knispel, T., Penner, I., Svindland, G., Voß, A. & Weber, S. (2023). Modeling and pricing cyber insurance: Idiosyncratic, systematic, and systemic risks. *European Actuarial Journal*, **13**(1), 1–53.
- Bakdash, J. Z., Hutchinson, S., Zaroukian, E. G., Marusich, L. R., Thirumuruganathan, S., Sample, C., Hoffman, B. & Das, G. (2018). Malware in the future? Forecasting of analyst detection of cyber events. *Journal of Cybersecurity*, **4**(1), ty007.
- Baker, T. & Shortland, A. (2022). Insurance and enterprise: Cyber insurance for ransomware. *The Geneva Papers on Risk and Insurance - Issues and Practice*, **48**, 275–299.
- Ballardie, T. & Crowcroft, J. (1995). Multicast-specific security threats and counter-measures. In *Proceedings of the Symposium on Network and Distributed System Security* (pp. 2–16).

- Bandyopadhyay, T. & Mookerjee, V.** (2019). A model to analyze the challenge of using cyber insurance. *Information Systems Frontiers*, **21**(2), 301–325.
- Barracchini, C. & Addressi, M. E.** (2014). Cyber risk and insurance coverage: An actuarial multistate approach. *Review of Economics & Finance*, **4**, 57–69.
- Beirlant, J. & Teugels, J. L.** (1992). Modeling large claims in non-life insurance. *Insurance: Mathematics and Economics*, **11**(1), 17–29.
- Bentley, M., Stephenson, A., Toscas, P. & Zhu, Z.** (2020). A multivariate model to quantify and mitigate cybersecurity risk. *Risks*, **8**(2), 61.
- Berliner, B.** (1982). *Limits of insurability of risks*. Prentice Hall.
- Bessy-Roland, Y., Boumezoued, A. & Hillairet, C.** (2021). Multivariate Hawkes process for cyber insurance. *Annals of Actuarial Science*, **15**(1), 14–39.
- Biener, C., Eling, M. & Wirfs, J. H.** (2015). Insurability of cyber risk: An empirical analysis. *The Geneva Papers on Risk and Insurance - Issues and Practice*, **40**(1), 131–158.
- Bilge, L., Han, Y. & Dell'Amico, M.** (2017) Riskteller: Predicting the risk of cyber incidents. In *Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security* (pp. 1299–1311).
- Böhme, R.** (2005) Cyber-insurance revisited. In *Workshop on the Economics of Information Security (WEIS)*.
- Böhme, R. & Kataria, G.** (2006) Models and measures for correlation in cyber-insurance. In *Workshop on the Economics of Information Security (WEIS)*.
- Böhme, R., Laube, S. & Riek, M.** (2019). A fundamental approach to cyber risk analysis. *Variance*, **12**(2), 161–185.
- Böhme, R. & Schwartz, G.** (2010) Modeling cyber-insurance: Towards a unifying framework. In *Workshop on the Economics of Information Security (WEIS)*.
- Campbell, K., Gordon, L. A., Loeb, M. P. & Zhou, L.** (2003). The economic cost of publicly announced information security breaches: Empirical evidence from the stock market. *Journal of Computer Security*, **11**(3), 431–448.
- Canali, D., Bilge, L. & Balzarotti, D.** (2014). On the effectiveness of risk prediction based on users browsing behavior. In *Proceedings of the 9th ACM Symposium on Information, Computer and Communications Security* (pp. 171–182).
- Cárdenas, A. A., Amin, S., Lin, Z.-S., Huang, Y.-L., Huang, C.-Y. & Sastry, S.** (2011) Attacks against process control systems: Risk assessment, detection, and response. In *Proceedings of the 6th ACM Symposium on Information, Computer and Communications Security* (pp. 355–366).
- Carfora, M. F., Martinelli, F., Mercaldo, F. & Orlando, A.** (2019). Cyber risk management: An actuarial point of view. *Journal of Operational Risk*, **14**(4), 77–103.
- Cartagena, S., Gosrani, V., Grewal, J. & Pikinska, J.** (2020). Silent cyber assessment framework. *British Actuarial Journal*, **25**(2), e2. doi: [10.1017/S1357321720000021](https://doi.org/10.1017/S1357321720000021)
- Cavusoglu, H., Mishra, B. & Raghunathan, S.** (2004a). The effect of internet security breach announcements on market value: Capital market reactions for breached firms and internet security developers. *International Journal of Electronic Commerce*, **9**(1), 70–104.
- Cavusoglu, H., Mishra, B. & Raghunathan, S.** (2004b). A model for evaluating IT security investments. *Communications of the ACM*, **47**(7), 87–92.
- Cebula, J. L. & Young, L. R.** (2010). *A taxonomy of operational cyber security risks*. Technical report, Software Engineering Institute, Carnegie Mellon University.
- Chen, Y.-Z., Huang, Z.-G., Xu, S. & Lai, Y.-C.** (2015). Spatiotemporal patterns and predictability of cyberattacks. *PLoS ONE*, **10**(5), e0124472.
- Chen, Z.** (2019). Discrete-time vs. continuous-time epidemic models in networks. *IEEE Access*, **7**, 127669–127677.
- Chockalingam, S., Pieters, W., Teixeira, A. & van Gelder, P.** (2017) Bayesian network models in cyber security: A systematic review. In *Secure IT Systems: 22nd Nordic Conference, NordSec 2017, Tartu, Estonia, November 8-10, 2017, Proceedings 22* (pp. 105–122). Springer.
- Cole, C. R. & Fier, S. G.** (2020). An empirical analysis of insurer participation in the U.S. cyber insurance market. *North American Actuarial Journal*, **25**(2), 232–254.
- Cremer, F., Sheehan, B., Fortmann, M., Kia, A. N., Mullins, M., Murphy, F. & Materne, S.** (2022). Cyber risk and cyber-security: A systematic review of data availability. *The Geneva Papers on Risk and Insurance - Issues and Practice*, **47**(3), 698–736.
- CRO Forum** (2016). Concept paper on a proposed categorisation methodology for cyber risk. Author.
- Daffron, J., Ruffle, S., Andrew, C., Copic, J. & Quanttrill, K.** (2019). *Bashe attack: Global infection by contagious malware*. Cambridge Centre for Risk Studies, Lloyds of London and Nanyang Technological University.
- Daley, D. J., & Vere-Jones, D.** (2003). *An introduction to the theory of point processes: Volume I: Elementary theory and methods*. Springer.
- Denning, D. E.** (1987). An intrusion-detection model. *IEEE Transactions on Software Engineering*, **SE-13**(2), 222–232.
- Deneux, T.** (2019). Logistic regression, neural networks and dempster-shafer theory: A new perspective. *Knowledge-Based Systems*, **176**, 54–67.
- Dou, W., Tang, W., Wu, X., Qi, L., Xu, X., Zhang, X. & Hu, C.** (2020). An insurance theory based optimal cyber-insurance contract against moral hazard. *Information Sciences*, **527**, 576–589.

- Edwards, B., Hofmeyr, S. & Forrest, S. (2016). Hype and heavy tails: A closer look at data breaches. *Journal of Cybersecurity*, 2(1), 3–14.
- Egan, R., Cartagena, S., Mohamed, R., Gosrani, V., Grewal, J., Acharyya, M., Dee, A., Bajaj, R., Jaeger, V.-J. & Katz, D. (2019). Cyber operational risk scenarios for insurance companies. *British Actuarial Journal*, 24(6), e6. doi: 10.1017/S1357321718000284
- EIOPA (2022). *Discussion paper on methodological principles of insurance stress testing*. [https://www.eiopa.europa.eu/consultations/discussion-paper-methodologies-insurance-stress-testing-cyber-component\\_en](https://www.eiopa.europa.eu/consultations/discussion-paper-methodologies-insurance-stress-testing-cyber-component_en)
- Eisenbach, T. M., Kovner, A. & Lee, M. J. (2022). Cyber risk and the U.S. financial system: A pre-mortem analysis. *Journal of Financial Economics*, 145(3), 802–826.
- Eling, M. (2018). Cyber risk and cyber risk insurance: Status quo and future research. *The Geneva Papers on Risk and Insurance - Issues and Practice*, 43(2), 175–179.
- Eling, M. (2020). Cyber risk research in business and actuarial science. *European Actuarial Journal*, 10(2), 303–333.
- Eling, M. & Jung, K. (2018). Copula approaches for modeling cross-sectional dependence of data breach losses. *Insurance: Mathematics and Economics*, 82, 167–180.
- Eling, M. & Jung, K. (2022). Heterogeneity in cyber loss severity and its impact on cyber risk measurement. *Risk Management*, 24(4), 273–297.
- Eling, M., Jung, K. & Shim, J. (2022). Unraveling heterogeneity in cyber risks using quantile regressions. *Insurance: Mathematics and Economics*, 104, 222–242.
- Eling, M. & Loperfido, N. (2017). Data breaches: Goodness of fit, pricing, and risk measurement. *Insurance: Mathematics and Economics*, 75, 126–136.
- Eling, M. & Schnell, W. (2016). What do we know about cyber risk and cyber risk insurance? *Journal of Risk Finance*, 17(5), 474–491.
- Eling, M. & Schnell, W. (2020). Capital requirements for cyber risk and cyber risk insurance: An analysis of Solvency II, the U.S. risk-based capital standards, and the Swiss Solvency Test. *North American Actuarial Journal*, 24(3), 370–392.
- Eling, M. & Wirfs, J. (2019). What are the actual costs of cyber risk events? *European Journal of Operational Research*, 272(3), 1109–1119.
- Fahrenwaldt, M. A., Weber, S. & Weske, K. (2018). Pricing of cyber insurance contracts in a network model. *ASTIN Bulletin: The Journal of the IAA*, 48(3), 1175–1218.
- Fang, X., Xu, M., Xu, S. & Zhao, P. (2019). A deep learning framework for predicting cyber attacks rates. *EURASIP Journal on Information Security*, 2019(5). doi: 10.1186/s13635-019-0090-6
- Fang, Z., Xu, M., Xu, S. & Hu, T. (2021). A framework for predicting data breach risk: Leveraging dependence to cope with sparsity. *IEEE Transactions on Information Forensics and Security*, 16, 2186–2201.
- Farkas, S., Lopez, O. & Thomas, M. (2021). Cyber claim analysis using generalized Pareto regression trees with applications to insurance. *Insurance: Mathematics and Economics*, 98, 92–105.
- Ganin, A. A., Quach, P., Panwar, M., Collier, Z. A., Keisler, J. M., Marchese, D. & Linkov, I. (2020). Multicriteria decision framework for cybersecurity risk assessment and management. *Risk Analysis*, 40(1), 183–199.
- García-Teodoro, P., Díaz-Verdejo, J., Maciá-Fernández, G. & Vázquez, E. (2009). Anomaly-based network intrusion detection: Techniques, systems and challenges. *Computers & Security*, 28(1-2), 18–28.
- Georgiadou, A., Mouzakitis, S. & Askounis, D. (2022). Detecting insider threat via a cyber-security culture framework. *Journal of Computer Information Systems*, 62(4), 706–716.
- Ghafir, I., Kyriakopoulos, K. G., Lambbotharan, S., Aparicio-Navarro, F. J., AsSadhan, B., BinSalleeh, H. & Diab, D. M. (2019). Hidden markov models and alert correlations for the prediction of advanced persistent threats. *IEEE Access*, 7, 99508–99520.
- GhasemiGol, M., Ghaemi-Bafghi, A. & Takabi, H. (2016). A comprehensive approach for network attack forecasting. *Computers & Security*, 58, 83–105.
- Gil, S., Kott, A. & Barabási, A.-L. (2014). A genetic epidemiology approach to cyber-security. *Scientific Reports*, 4(1), 5659.
- Gordon, L. A. & Loeb, M. P. (2002). The economics of information security investment. *ACM Transactions on Information and System Security (TISSEC)*, 5(4), 438–457.
- Gordon, L. A., Loeb, M. P. & Sohail, T. (2003). A framework for using insurance for cyber-risk management. *Communications of the ACM*, 46(3), 81–85.
- GOV.UK (2020). *Countries in the EU and EEA*. <https://www.gov.uk/eu-eea>
- Grzebiela, T. (2002). Insurability of electronic commerce risks. In *Proceedings of the 35th Annual Hawaii International Conference on System Sciences*.
- Hansman, S. & Hunt, R. (2005). A taxonomy of network and computer attacks. *Computers & Security*, 24(1), 31–43.
- Herath, H. & Herath, T. (2011). Copula-based actuarial model for pricing cyber-insurance policies. *Insurance Markets and Companies*, 2(1), 7–20.
- Herath, H. S. & Herath, T. C. (2007). Cyber-insurance: Copula pricing framework and implication for risk management. In *Workshop on the Economics of Information Security (WEIS)*.

- Hillairet, C. & Lopez, O. (2021). Propagation of cyber incidents in an insurance portfolio: Counting processes combined with compartmental epidemiological models. *Scandinavian Actuarial Journal*, **2021**(8), 671–694.
- Hillairet, C., Lopez, O., d'Oultremont, L. & Spoorenberg, B. (2022). Cyber-contagion model with network structure applied to insurance. *Insurance: Mathematics and Economics*, **107**, 88–101.
- Hofmann, A. & Rothschild, C. (2019). On the efficiency of self-protection with spillovers in risk. *The Geneva Risk and Insurance Review*, **44**(2), 207–221.
- Hovav, A. & D'Arcy, J. (2003). The impact of Denial-of-Service attack announcements on the market value of firms. *Risk Management and Insurance Review*, **6**(2), 97–121.
- Hovav, A. & D'Arcy, J. (2004). The impact of virus attack announcements on the market value of firms. *Information Systems Security*, **13**(3), 32–40.
- Howard, J. D. (1997). *An analysis of security incidents on the Internet 1989-1995*. Carnegie Mellon University.
- Howard, J. D. & Longstaff, T. A. (1998). *A common language for computer security incidents*. Technical report, Sandia National Labs, United States.
- Huang, K., Zhou, C., Tian, Y.-C., Yang, S. & Qin, Y. (2018). Assessing the physical impact of cyberattacks on industrial cyber-physical systems. *IEEE Transactions on Industrial Electronics*, **65**(10), 8153–8162.
- Husák, M., Komárková, J., Bou-Harb, E. & Čeleda, P. (2018). Survey of attack projection, prediction, and forecasting in cyber security. *IEEE Communications Surveys & Tutorials*, **21**(1), 640–660.
- Jacobs, J. (2014). Analyzing Ponemon cost of data breach. *Data Driven Security*, **11**, 5.
- Jang-Jaccard, J. & Nepal, S. (2014). A survey of emerging threats in cybersecurity. *Journal of Computer and System Sciences*, **80**(5), 973–993.
- Jevtić, P. & Lanchier, N. (2020). Dynamic structural percolation model of loss distribution for cyber risk of small and medium-sized enterprises for tree-based LAN topology. *Insurance: Mathematics and Economics*, **91**, 209–223.
- Jung, K. (2021). Extreme data breach losses: An alternative approach to estimating probable maximum loss for data breach risk. *North American Actuarial Journal*, **25**(4), 580–603.
- Kamiya, S., Kang, J.-K., Kim, J., Milidonis, A. & Stulz, R. M. (2020). Risk management, firm reputation, and the impact of successful cyberattacks on target firms. *Journal of Financial Economics*, **139**(3), 719–749.
- Kaplan, S. & Garrick, B. J. (1981). On the quantitative definition of risk. *Risk Analysis*, **1**(1), 11–27.
- Khalili, M. M., Liu, M. & Romanosky, S. (2019). Embracing and controlling risk dependency in cyber-insurance policy underwriting. *Journal of Cybersecurity*, **5**(1), tyz010.
- Knight, R. & Nurse, J. R. (2020). A framework for effective corporate communication after cyber security incidents. *Computers & Security*, **99**, 102036.
- Krutilla, K., Alexeev, A., Jardine, E. & Good, D. (2021). The benefits and costs of cybersecurity risk reduction: A dynamic extension of the Gordon and Loeb model. *Risk Analysis*, **41**(10), 1795–1808.
- Lau, P., Wang, L., Wei, W., Liu, Z. & Ten, C.-W. (2022). A novel mutual insurance model for hedging against cyber risks in power systems deploying smart technologies. *IEEE Transactions on Power Systems*, **38**(1), 630–642.
- Lau, P., Wei, W., Wang, L., Liu, Z. & Ten, C.-W. (2020). A cybersecurity insurance model for power system reliability considering optimal defense resource allocation. *IEEE Transactions on Smart Grid*, **11**(5), 4403–4414.
- Leau, Y.-B. & Manickam, S. (2015) Network security situation prediction: A review and discussion. In *International Conference on Soft Computing, Intelligence Systems, and Information Technology* (Vol. **516**, pp. 423–435).
- Lemnitzer, J. M. (2021). Why cybersecurity insurance should be regulated and compulsory. *Journal of Cyber Policy*, **6**(2), 118–136.
- Li, Q., Tian, Y., Wu, Q., Cao, Q., Shen, H. & Long, H. (2020). A Cloud-Fog-Edge closed-loop feedback security risk prediction method. *IEEE Access*, **8**, 29004–29020.
- Liu, J., Li, J. & Daly, K. (2022). Bayesian vine copulas for modelling dependence in data breach losses. *Annals of Actuarial Science*, **16**(2), 401–424.
- Liu, W., Liu, C., Liu, X., Cui, S. & Huang, X. (2016a). Modeling the spread of malware with the influence of heterogeneous immunization. *Applied Mathematical Modelling*, **40**(4), 3141–3152.
- Liu, Y., Dong, M., Ota, K. & Liu, A. (2016b). Activetrust: Secure and trustable routing in wireless sensor networks. *IEEE Transactions on Information Forensics and Security*, **11**(9), 2013–2027.
- Liu, Y., Sarabi, A., Zhang, J., Naghizadeh, P., Karir, M., Bailey, M. & Liu, M. (2015) Cloudy with a chance of breach: Forecasting cyber security incidents. In *Proceedings of the 24th USENIX Conference on Security Symposium* (pp. 1009–1024).
- Liu, Z., Wei, W. & Wang, L. (2021). An extreme value theory-based catastrophe bond design for cyber risk management of power systems. *IEEE Transactions on Smart Grid*, **13**(2), 1516–1528.
- Ma, B., Chu, T. & Jin, Z. (2022). Frequency and severity estimation of cyber attacks using spatial clustering analysis. *Insurance: Mathematics and Economics*, **106**, 33–45.
- Maillart, T. & Sornette, D. (2010). Heavy-tailed distribution of cyber-risks. *The European Physical Journal B*, **75**(3), 357–364.
- Malavasi, M., Peters, G. W., Shevchenko, P. V., Trück, S., Jang, J. & Sofronov, G. (2022). Cyber risk frequency, severity and insurance viability. *Insurance: Mathematics and Economics*, **106**, 90–114.

- Marotta, A., Martinelli, F., Nanni, S., Orlando, A. & Yautsiukhin, A.** (2017). Cyber-insurance survey. *Computer Science Review*, **24**, 35–61.
- Mateski, M., Trevino, C. M., Veitch, C. K., Michalski, J., Harris, J. M., Maruoka, S. & Frye, J.** (2012). *Cyber threat metrics*. Technical report, Sandia National Laboratories, United States.
- McShane, M., Eling, M. & Nguyen, T.** (2021). Cyber risk management: History and future research directions. *Risk Management and Insurance Review*, **24**(1), 93–125.
- Meland, P. H., Tondel, I. A. & Solhaug, B.** (2015). Mitigating risk with cyberinsurance. *IEEE Security & Privacy*, **13**(6), 38–43.
- Mirkovic, J. & Reiher, P.** (2004). A taxonomy of DDoS attack and DDoS defense mechanisms. *ACM SIGCOMM Computer Communication Review*, **34**(2), 39–53.
- Mishra, B. K. & Pandey, S. K.** (2014). Dynamic model of worm propagation in computer network. *Applied Mathematical Modelling*, **38**(7-8), 2173–2179.
- Moore, D., Shannon, C., Brown, D. J., Voelker, G. M. & Savage, S.** (2006). Inferring internet Denial-of-Service activity. *ACM Transactions on Computer Systems*, **24**(2), 115–139.
- Moore, T.** (2005) Countering hidden-action attacks on networked systems. In *Workshop on the Economics of Information Security (WEIS)*.
- Mukhopadhyay, A., Chatterjee, S., Saha, D., Mahanti, A. & Sadhukhan, S. K.** (2006) e-risk management with insurance: A framework using copula aided Bayesian belief networks. In *Proceedings of the 39th Annual Hawaii International Conference on System Sciences (HICSS'06)* (p. 126a).
- Mukhopadhyay, A., Chatterjee, S., Saha, D., Mahanti, A. & Sadhukhan, S. K.** (2013). Cyber-risk decision models: To insure IT or not? *Decision Support Systems*, **56**, 11–26.
- Munich RE** (2020). *Cyber insurance: Risks and trends*. Author.
- National Institute of Standards and Technology** (2014) Framework for improving critical infrastructure cybersecurity. In *ITL Bulletin*.
- National Institute of Standards and Technology** (2020). *2019 NIST/ITL Cybersecurity Program Annual Report*.
- NetDiligence** (2022). 2022 cyber claims study. Author.
- Office of the Australian Information Commissioner** (2018). *Mandatory data breach notification comes into force this thursday*. <https://www.oaic.gov.au/updates/news-and-media/mandatory-data-breach-notification-comes-into-force-this-thursday/>
- Ögüt, H., Menon, N. & Raghunathan, S.** (2005) Cyber insurance and IT security investment: Impact of interdependence risk. In *4th Workshop on the Economics of Information Security (WEIS)*.
- Ögüt, H., Raghunathan, S. & Menon, N.** (2011). Cyber security risk management: Public policy implications of correlated risk, imperfect ability to prove loss, and observability of self-protection. *Risk Analysis*, **31**(3), 497–512.
- Okutan, A., Yang, S. J. & McConky, K.** (2017) Predicting cyber attacks with Bayesian networks using unconventional signals. In *Proceedings of the 12th Annual Conference on Cyber and Information Security Research*.
- Onwubiko, C. & Ouazzane, K.** (2020). SOTER: A playbook for cybersecurity incident management. *IEEE Transactions on Engineering Management*, **69**(6), 3771–3791.
- Panda, S., Woods, D. W., Laszka, A., Fielder, A. & Panaousis, E.** (2019). Post-incident audits on cyber insurance discounts. *Computers & Security*, **87**, 101593.
- Pandey, P. & Sneekenes, E.** (2016). Using financial instruments to transfer the information security risks. *Future Internet*, **8**(2), 20.
- Park, H., Jung, S.-O. D., Lee, H. & In, H. P.** (2012) Cyber weather forecasting: Forecasting unknown internet worms using randomness analysis. In *IFIP International Information Security Conference* (pp. 376–387).
- Paté-Cornell, M.-E., Kuypers, M., Smith, M. & Keller, P.** (2018). Cyber risk management for critical infrastructure: A risk analysis model and three case studies. *Risk Analysis*, **38**(2), 226–241.
- Peng, C., Xu, M., Xu, S. & Hu, T.** (2017). Modeling and predicting extreme cyber attack rates via marked point processes. *Journal of Applied Statistics*, **44**(14), 2534–2563.
- Peng, C., Xu, M., Xu, S. & Hu, T.** (2018). Modeling multivariate cybersecurity risks. *Journal of Applied Statistics*, **45**(15), 2718–2740.
- Polatidis, N., Pimenidis, E., Pavlidis, M., Papastergiou, S. & Mouratidis, H.** (2018). From product recommendation to cyber-attack prediction: Generating attack graphs and predicting future attacks. *Evolving Systems*, **11**(3), 479–490.
- Privacy Rights Clearinghouse** (2019). *Chronology of data breaches*. <https://privacyrights.org/data-breaches>
- Qin, X. & Lee, W.** (2004) Attack plan recognition and prediction using causal networks. In *20th Annual Computer Security Applications Conference* (pp. 370–379).
- Rhode, M., Burnap, P. & Jones, K.** (2018). Early-stage malware prediction using recurrent neural networks. *Computers & Security*, **77**, 578–594.
- Riek, M. & Böhme, R.** (2018). The costs of consumer-facing cybercrime: An empirical exploration of measurement issues and estimates. *Journal of Cybersecurity*, **4**(1), ty004.
- Romanosky, S.** (2016). Examining the costs and causes of cyber incidents. *Journal of Cybersecurity*, **2**(2), 121–135.

- Romanosky, S., Ablon, L., Kuehn, A. & Jones, T. (2019). Content analysis of cyber insurance policies: How do carriers price cyber risk? *Journal of Cybersecurity*, 5(1), tyz002.
- Romeike, F. (2000). It-risken und grenzen traditioneller risikofinanzierungsprodukte. *Zeitschrift für Versicherungswesen*, 51(17), 603–610.
- Rosati, P., Cummins, M., Deeney, P., Gogolin, F., van der Werff, L. & Lynn, T. (2017). The effect of data breach announcements beyond the stock price: Empirical evidence on market activity. *International Review of Financial Analysis*, 49, 146–154.
- Rosati, P., Deeney, P., Cummins, M., Van der Werff, L. & Lynn, T. (2019). Social media and stock price reaction to data breach announcements: Evidence from us listed companies. *Research in International Business and Finance*, 47, 458–469.
- Rustad, M. L. & Koenig, T. H. (2019). Towards a global data privacy standard. *Florida Law Review*, 71, 365.
- Sarabi, A., Naghizadeh, P., Liu, Y. & Liu, M. (2016). Risky business: Fine-grained data breach prediction using business profiles. *Journal of Cybersecurity*, 2(1), 15–28.
- Sarkar, S., Almukaynizi, M., Shakarian, J. & Shakarian, P. (2019). Predicting enterprise cyber incidents using social network analysis on dark web hacker forums. In *The Cyber Defense Review* (pp. 87–102).
- Schatz, D. & Bashroush, R. (2017). Economic valuation for information security investment: A systematic literature review. *Information Systems Frontiers*, 19, 1205–1228.
- Schuster, M. & Paliwal, K. K. (1997). Bidirectional recurrent neural networks. *IEEE Transactions on Signal Processing*, 45(11), 2673–2681.
- Shalev-Shwartz, S. & Ben-David, S. (2014). *Understanding machine learning: From theory to algorithms*. Cambridge University Press.
- Shu, K., Sliva, A., Sampson, J. & Liu, H. (2018). Understanding cyber attack behaviors with sentiment information on social media. In *Social, Cultural, and Behavioural Modeling* (pp. 377–388).
- Sklar, A. (1959). Fonctions de répartition à n dimensions et leurs marges. *Publications de l'Institut de Statistique de l'Université de Paris*, 8, 229–231.
- Spanos, G. & Angelis, L. (2016). The impact of information security events to the stock market: A systematic literature review. *Computers & Security*, 58, 216–229.
- Srinivas, J., Das, A. K. & Kumar, N. (2019). Government regulations in cyber security: Framework, standards and recommendations. *Future Generation Computer Systems*, 92, 178–188.
- Stine, K., Quinn, S., Witte, G. & Gardner, R. (2020). *Integrating cybersecurity and enterprise risk management (ERM)*. Technical report, National Institute of Standards and Technology. NIST Interagency or Internal Report (NISTIR) 8286.
- Stoneburner, G., Goguen, A. & Feringa, A. (2002). *Risk management guide for information technology systems*. Special Publication. National institute of Standards and Technology.
- Subroto, A. & Apriyana, A. (2019). Cyber risk prediction through social media big data analytics and statistical machine learning. *Journal of Big Data*, 6, 50.
- Sun, D., Wu, Z., Wang, Y., Lv, Q. & Hu, B. (2019). Risk prediction for imbalanced data in cyber security: A Siamese network-based deep learning classification framework. In *2019 International Joint Conference on Neural Networks (IJCNN)* (pp. 1–8).
- Sun, H., Xu, M. & Zhao, P. (2021). Modeling malicious hacking data breach risks. *North American Actuarial Journal*, 25(4), 484–502.
- Sun, N., Zhang, J., Rimba, P., Gao, S., Zhang, L. Y. & Xiang, Y. (2018). Data-driven cybersecurity incident prediction: A survey. *IEEE Communications Surveys & Tutorials*, 21(2), 1744–1772.
- Verizon (2021). *2021 data breach investigations report*. Author.
- Voss, W. G. & Houser, K. A. (2019). Personal data and the GDPR: Providing a competitive advantage for U.S. companies. *American Business Law Journal*, 56(2), 287–344.
- Wheatley, S., Maillart, T. & Sornette, D. (2016). The extreme risk of personal data breaches and the erosion of privacy. *The European Physical Journal B*, 89, 7.
- Woods, D. W., Moore, T. & Simpson, A. C. (2021). The county fair cyber loss distribution: Drawing inferences from insurance prices. *Digital Threats: Research and Practice*, 2(2), 1–21.
- Xie, X., Lee, C. & Eling, M. (2020). Cyber insurance offering and performance: An analysis of the U.S. cyber insurance market. *The Geneva Papers on Risk and Insurance - Issues and Practice*, 45, 690–736.
- Xu, M. & Hua, L. (2019). Cybersecurity insurance: Modeling and pricing. *North American Actuarial Journal*, 23(2), 220–249.
- Xu, M., Hua, L. & Xu, S. (2017). A vine copula model for predicting the effectiveness of cyber defense early-warning. *Technometrics*, 59(4), 508–520.
- Xu, M., Schweitzer, K. M., Bateman, R. M. & Xu, S. (2018). Modeling and predicting cyber hacking breaches. *IEEE Transactions on Information Forensics and Security*, 13(11), 2856–2871.
- Xu, M. & Zhang, Y. (2021). Data breach CAT bonds: Modeling and pricing. *North American Actuarial Journal*, 25(4), 543–561.
- Yamashita, K., Ten, C.-W., Rho, Y., Wang, L., Wei, W. & Ginter, A. (2020). Measuring systemic risk of switching attacks based on cybersecurity technologies in substations. *IEEE Transactions on Power Systems*, 35(6), 4206–4219.

- Yang, Z., Liu, Y., Campbell, M., Ten, C.-W., Rho, Y., Wang, L. & Wei, W.** (2020). Premium calculation for insurance businesses based on cyber risks in IP-based power substations. *IEEE Access*, **8**, 78890–78900.
- Zängerle, D. & Schiereck, D.** (2022). Modelling and predicting enterprise-level cyber risks in the context of sparse data availability. *The Geneva Papers on Risk and Insurance - Issues and Practice*, **48**, 434–462.
- Zeller, G. & Scherer, M.** (2022). A comprehensive model for cyber risk based on marked point processes and its application to insurance. *European Actuarial Journal*, **12**, 33–85.
- Zhan, Z., Xu, M. & Xu, S.** (2013). Characterizing honeypot-captured cyber attacks: Statistical framework and case study. *IEEE Transactions on Information Forensics and Security*, **8**(11), 1775–1789.
- Zhan, Z., Xu, M. & Xu, S.** (2015). Predicting cyber attack rates with extreme values. *IEEE Transactions on Information Forensics and Security*, **10**(8), 1666–1677.
- Zhang Wu, M., Luo, J., Fang, X., Xu, M. & Zhao, P.** (2023). Modeling multivariate cyber risks: Deep learning dating extreme value theory. *Journal of Applied Statistics*, **50**(3), 610–630.