JOURNAL OF CYBERSECURITY

Journal of Cybersecurity, 2022, 1–12
https://doi.org/10.1093/cybsec/tyac016
Research Paper

Research Paper

The nature of losses from cyber-related events: risk categories and business sectors

Pavel V. Shevchenko[1], Jiwook Jang[1], Matteo Malavasi[1], Gareth W. Peters[2,*], Georgy Sofronov[3] and Stefan Trück[1]

[1]Department of Actuarial Studies and Business Analytics, Macquarie Business School, Macquarie University, Sydney NSW 2109, Australia, [2]Department of Statistics and Applied Probability, College of Letters and Science, University of California Santa Barbara, Santa Barbara, CA 93106, USA and [3]School of Mathematical and Physical Sciences, Faculty of Science and Engineering, Macquarie University, Sydney NSW 2109, Australia

*Correspondence address. Department of Actuarial Mathematics and Statistics, School of Mathematical and Computer Sciences, Heriot-Watt University, Edinburgh, Scotland, EH14 4AS, UK; E-mail: garethpeters@ucsb.edu

Received 3 February 2022; revised 8 October 2022; accepted 26 October 2022

Abstract

In this study, we examine the nature of losses from cyber-related events across different risk categories and business sectors. Using a leading industry dataset of cyber events, we evaluate the relationship between the frequency and severity of individual cyber-related events and the number of affected records. We find that the frequency of reported cyber-related events has substantially increased between 2008 and 2016. Furthermore, the frequency and severity of losses depend on the business sector and type of cyber threat: the most significant cyber loss event categories, by number of events, were related to data breaches and the unauthorized disclosure of data, while cyber extortion, phishing, spoofing, and other social engineering practices showed substantial growth rates. Interestingly, we do not find a distinct pattern between the frequency of events, the loss severity, and the number of affected records as often alluded to in the literature. We also analyse the severity distribution of cyber-related events across all risk categories and business sectors. This analysis reveals that cyber risks are heavy-tailed, i.e. cyber risk events have a higher probability to produce extreme losses than events whose severity follows an exponential distribution. Furthermore, we find that the frequency and severity of cyber-related losses exhibit a very dynamic and time-varying nature.

Key words: cyber risk, frequency and severity, risk categories, business sectors, heavy-tailed distributions

Introduction

According to a recent estimate provided in the Global Risk Report by the World Economic Forum [1], losses from cyber-related risks might reach US$ 6 trillion in 2021. Due to the digitalization of business and economic activities via the Internet of Things (IoT), cloud computing, mobile, blockchain, and other innovative technologies, cyber risk is inherent and extreme. In general, one may refer to cyber risk as any risk of financial loss, disruption to operations, or damage to the reputation of an organization due to failure of its information technology (IT) systems, as defined by the Institute of Risk Management (IRM) in [2, 3]. Financial losses from malicious cyber activities result from IT security/data/digital assets recovery costs, liability with respect to identity theft and data breaches, reputation/brand damage, legal liability, cyber extortion, regulatory defence, and penalties coverage and business interruption. In the financial sector, cyber risk is classified by the Basel Committee on Banking Supervision [4] as a category of operational risk, for instance affecting information and technology assets that can have consequences for the confidentiality, availability, and integrity of information and information systems [5].

In this study, we seek to understand from an empirical perspective, the relationships and dynamics of cyber risk loss processes. In order to achieve this, we undertake a study of both the behavior of the occurrence of events, known as frequency analysis, as well as the magnitude of events, known as severity analysis. As such, the

study undertaken provides a thorough analysis of the frequency and severity of cyber-related events across different risk categories and business sectors. This is important, as cyber risk reaches all aspects of corporate, industry and government sectors at both the institutional level, down to an individual level.

Furthermore, requirements on the need to mitigate, report, and respond to emerging cyber threats differs in the regulatory intensity and requirements across each sector of industry and society. Therefore, one would expect differing patterns in the dynamics of cyber risk loss processes when viewed from differing industry and risk category perspectives. One may also expect the emergence of different trends in various industries to occur as new attack vectors or attacker capabilities are developed. Therefore, given the increasing importance of cyber threats to businesses, government, and individuals, a statistical analysis of the nature of such events will help to develop appropriate risk management strategies and support investment into optimal mitigation of the risks. Without adequately quantifying the risks from different cyber threat categories for business sectors, it is impossible to manage the identified risks to be within acceptable levels.

Another key focus of this article is to demonstrate the main features of cyber risk loss processes, based on the conducted frequency and severity analysis. There are various statistical models for frequency and severity used in operational risk practice suitable for cyber risk; see e.g. monographs devoted to such models [6, 7]. Cohen et al. [8] also suggest that cyber losses and non-cyber losses from operational risk share a similar fundamental risk profile. This would suggest that modeling techniques that have been originally developed for operational risk modeling may also be adequate for cyber-related threats. Whilst in this work we undertake a purely empirical analysis, the results we present add insight into the nature of the statistical models (such as considered in [9]) that will be suitable to capture cyber risk loss processes adequately and the challenges faced in trying to achieve this in a non-stationary emerging dynamic threat domain.

The frequency of malicious cyber activities is rapidly increasing, with the scope and nature dependent on an organization's industry, size, and location. According to the Allianz Global Risk Barometer 2021 [10], cyber incident (including cybercrime, IT failure/outage, data breaches, fines, and penalties) is currently a top-three global business risk. It is, therefore, critical that corporations and governments focus on IT and network security enhancement. Unless public and private sector organizations have effective cyber security plans and strategies in place, and tools to manage and mitigate losses from cyber risks, cyber events have the potential to affect their business significantly, possibly damaging hard-earned reputations irreparably [11].

Due to the impact of COVID-19, business and economic activities will be also accelerated in cyber space, which could significantly increase the frequency and impact of cyber events around the globe, with alarming consequences for public and private sector organizations [12]. Concerns about higher frequency and severity of cyber catastrophes demand a re-examination of characterizing cyber risks. A significant challenge ahead is to quantify individual cyber-related events in order to better understand the current and emerging risk landscape in cyber space and to minimize potentially catastrophic losses from cyber activities.

The lack of historical data on losses from cyber risk is another challenge to model the frequency and severity of individual cyber-related events [13–16]. For example, in Australia, it only became mandatory for breached organizations to notify their data breaches details in February 2018 (see [17]). Many countries around the world are in a similar situation, such that often only very limited data on losses from cyber-related events is available. This makes the design of adequate models for the quantification of cyber risks very difficult.

McShane et al. [11] provide a comprehensive review of the literature on managing cyber risks, focusing in particular on work, i.e. related to risk identification, risk analysis, and risk treatment. For each of these steps, an appropriate quantification of potential losses from cyber-related events is paramount to consider and operationalize. The modeling and quantification of cyber losses is also a key component required to incorporate cyber risk into an overall enterprise risk management process and to facilitate the prioritization of investment decisions aimed at reducing the impact of cyber attacks.

This study provides a thorough analysis of the frequency and severity of individual cyber-related events for different cyber threats and business sectors. To do this, we use one of the most comprehensive databases on losses from cyber-related threats provided by Advisen[1], hereafter referred to as Advisen Cyber Loss Data (https://www.advisenltd.com/data/cyber-loss-data/). Romanosky [18] examined an early version of this dataset (i.e. over 12 000 cyber events from 2004 to 2015) that includes data breaches, security incidents, privacy violations, and phishing crimes with a regression analysis. In contrast, the dataset analyzed in this study comprises over 132 126 cyber events from 2008 to 2020, affecting 49 496 organizations, with >80% of the organizations represented in the dataset residing in the USA. This is both a more current analysis and we argue more complete in coverage than previous studies. This is important given the dynamic and non-stationary nature of cyber loss processes, as we will demonstrate in the analysis and studies undertaken.

Our findings provide new insights on the relationship between the frequency and severity of losses from cyber events and factors such as the number of affected records by a cyber attack, or risk types and business sectors. We, first, evaluate the relationship between the frequency and severity of individual cyber-related events and the number of affected records. We find that the frequency and severity of the events are not independent of the business sector and type of cyber threat. Second, analysing the severity distribution of cyber-related events across all risk categories and business sectors provides us that cyber risk types are heavy tailed, i.e. cyber risk events have a higher probability to produce extreme losses than events whose severity follows an exponential distribution (for an approachable book length discussion on heavy tails see, e.g. [19]). We also summarize detailed findings from the analysis of Advisen Cyber Loss Data, which are beneficial for a better understanding of the nature of cyber risk. We close this article with some key recommendations to cyber risk management decision-makers in private and public sector organizations.

The Data

In our empirical analysis of the structure of losses from cyber-related threats, we use the Advisen Cyber Loss Data, which is one of the most comprehensive datasets on cyber events. The dataset contains incidents collected from reliable and publicly verifiable sources, such as news media, governmental and regulatory sources, state data breach notification sites, and third-party vendors. For cyber loss data, Ad-

1 Advisen is a leading provider of data, media, and technology solutions for the commercial property and casualty insurance market. Advisen's proprietary datasets include "*Cyber Loss Data*," "*Casualty Dataset*," "*Private D&O Loss Data*," "*Public D&O Loss Data*" and "*Loss Insight*."

visen adopts a granular classification based on the type of cyber risk threat.

Classification of cyber risks

Cyber risk involves a wide variety of risk factors and touches on nearly every sector of the public and private domains, presenting many facets, combining technical know-how with behavioral and cultural aspects [20–23]. This multidimensional heterogeneity makes defining and classifying cyber-related events a non-unique task, to the point where a globally accepted and standardized classification of cyber risk is not yet achieved universally across all industry domains and sectors of society. Instead, there are classifications and taxonomies, which have been developed from different industry perspectives [4, 5, 13, 24–27]. Furthermore, many public and private institutions have tried to produce classifications addressing the most prominent cyber risk aspects relevant to their own stakeholders [20, 23, 28, 29].

For our empirical analysis, we decided to follow the classification that has been suggested in the Advised Cyber Loss Data.[2] This classification comprises the following 16 cyber risk categories.

- Privacy—unauthorized contact or disclosure: cases when personal information is used in an unauthorized manner to contact or publicize information regarding an individual or an organization without their explicit permission.
- Privacy—unauthorized data collection: cases where information about the users of electronic services, such as social media, cellphones, websites, and similar is captured and stored without their knowledge or consent, or where prohibited information may have been collected with or without their consent.
- Data—physically lost or stolen: situations where personal confidential information or digital assets have been stored on, or may have been stored on, computer, peripheral equipment, data storage, or printouts, which has been lost, stolen, or improperly disposed of.
- Data—malicious breach: situations where personal confidential information or digital assets either have been or may have been exposed or stolen, by unauthorized internal or external actors whose intent appears to have been the acquisition of such information.
- Data—unintentional disclosure: situations where personal confidential information or digital assets have either been exposed, or may have been exposed, to unauthorized viewers due to an unintentional or inadvertent accident or error.
- Identity—fraudulent use/account access: identity theft or the fraudulent use of confidential personal information or account access in order to steal money, establish credit, or access account information, either through electronic or other means.
- Industrial controls and operations: losses involving disruption or attempted disruption to "connected" physical assets such as factories, automobiles, power plants, electrical grids, and similar (including "the internet of things").
- Network/website disruption: unauthorized use of or access to a computer or network, or interference with the operation of same, including virus, worm, malware, digital denial of service (DDOS), system intrusions, and similar.
- Phishing, spoofing, social engineering: attempts to get individuals to voluntarily provide information, which could then be used illicitly, e.g. phishing or spoofing a legitimate website with a close replica to obtain account information, or sending fraudulent emails to initiate unauthorized activities (aka "spear phishing").
- Skimming, physical tampering: use of physical devices to illegally capture electronic information such as bank account or credit card numbers for individual transactions, or installing software on such point-of-sale devices to accomplish the same goal.
- IT—configuration/implementation errors: losses resulting from errors or mistakes, which are made in maintaining, upgrading, replacing, or operating the hardware and software IT infrastructure of an organization, typically resulting in system, network, or web outages or disruptions.
- IT—processing errors: losses resulting from internal errors in electronically processing orders, purchases, registrations, and similar, usually due to a security or authorization inadequacy, software bug, hardware malfunction, or user error.
- Cyber extortion: threats to lock access to devices or files, fraudulently transfer funds, destroy data, interfere with the operation of a system/network/site, or disclose confidential digital information such as identities of customers/employees, unless payments are made.
- Denial of service (DDOS)/system disruption: a disruption, which is too widespread to be accounted to individual organizations/entities.
- Digital breach/identity theft: widespread hacking or identity theft, which targets a large number of companies, or individuals.
- Undetermined/other.

The dataset is also classified using the following 20 business sectors proposed by the North American Industry Classification (NAIC) system [30]:

- Agriculture, forestry, fishing, and hunting
- Mining, quarrying, and oil and gas extraction
- Utilities
- Construction
- Manufacturing
- Wholesale trade
- Retail trade
- Transportation and warehousing
- Information
- Finance and insurance
- Real estate and rental and leasing
- Professional, scientific, and technical services
- Management of companies and enterprises
- Administrative and support and waste management and remediation services
- Educational services
- Health care and social assistance
- Arts, entertainment, and recreation
- Accommodation and food services

2 Note that the proposed cyber loss event and business line categories are distinct from those defined by the Basel II regulatory framework [4] for operational risk. Thus, when working with a taxonomy distinct from that specified in regulation for the banking sector, financial institutions subject to such regulatory reporting requirements will have to consider carefully the mapping exercise to move from the Advisen taxonomy to the Basel II required reporting taxonomy. Basel II business lines are largely non-representative of the taxonomy adopted by Advisen which covers a much wider selection of sectors including the financial services. As such, the integration of other cyber loss data collections such as those collected over the last 15 years in the banking industry by consortiums such as, e.g. ORX https://managingrisktogether.orx.org/ should be carefully considered.

- Other services (except public administration)
- Public administration

The great conundrum for modeling and analysing cyber-related loss data is that whilst the prevalence of such events and their impact is seemingly growing over time, the access to national, standardized public domain databases for such loss data is scarce and often prohibitively expensive to obtain [14, 16, 31, 32]. Moreover, given that a widely accepted cyber risk definition and taxonomy does not exist universally across different sectors subject to different regulatory considerations and bodies, a dataset containing uniform and systematic information on cyber event severity and frequency is hard to find and to work with. As such, we believe that the dataset considered in this study represents an industry gold standard in this regard and, therefore, should act as a meaningful representation of the current status of cyber risk loss process evolution.

Preliminary analysis

The dataset analyzed in this study contains 132 126 cyber events from 2008 to 2020, affecting 49 496 organizations, with >80% of the organizations represented in the dataset residing in the USA. It is important to also note that given the nature of cyber risk, the reporting requirements, and the methods of data collection utilized in compiling this dataset under study, it can be assumed that a relatively high share of events may not be recorded. It is well known in fact that enterprises and companies seldom and reluctantly report cyber-related events to avoid, among other things, a loss of reputation and trust from their counterparties.

Figure 1 illustrates the number of events in the Advisen database by country. The vast majority of the recorded events in this database occurred in the USA (83.09%), while only a minority of events is recorded for the entire European Union (2.65%), Asia (3.17%), or Oceania (1.04%). As mentioned earlier, given the focus of the dataset on the USA, companies presented in the dataset have been classified according to the NAIC system. In the following, we focus our analysis on non-zero losses in the dataset, i.e. 4667 cyber events. Note that the share of these losses in the entire database is only 3.53% of the total events. However, given our emphasis on the severity of cyber-related events, we have to rely on events where information on the magnitude of the loss was provided.

Table 1 provides descriptive statistics of non-zero losses for each cyber risk category. We find substantial differences for the number of non-zero loss events across the different risk categories. While we observe over 1900 non-zero losses for the category *Privacy—Unauthorized Contact or Disclosure*, only six non-zero losses are observed for the category *Industrial Controls* throughout the sample period. We also find heterogeneity in the magnitude of losses across the different categories. All risk categories exhibit a mean loss, i.e. higher than the median, indicating that the loss distribution is skewed to the right, potentially exhibiting heavy tails. In some cases, this effect is so pronounced that the mean is >100 times higher than the median. Table 1 also illustrates that losses from cyber-events typically have a very high standard deviation, positive skewness paired with high kurtosis. Overall, the descriptive statistics in Table 1 also seem to confirm earlier results on cyber-related losses typically following heavy-tailed distributions; see, e.g. [31–33].

Characterization of Cyber Risks

Affected records, frequency, and severity of events

In a first step, we evaluate the relationship between the frequency and severity of individual cyber-related events and the number of affected records. Figure 2 shows the business sector ranked by frequency and severity of cyber events. Each circle represents a business sector, and its area corresponds to the average number of records affected by a cyber event, i.e. the larger the circle the more records have been affected.

The sectors with the highest average cyber loss are: "Information," "Manufacturing," "Transportation and Warehousing," and "Wholesale Trade." In terms of records affected: "Information," "Professional, Scientific, and Technical Services," "Agriculture, Forestry, Fishing, and Hunting," and "Accommodation and Food Services." Figure 2 also depicts the fact that monetary losses and the number of records affected vary across business sectors. Business sectors in the top right corner of the graph in Fig. 2 share some common features: they exhibit high average loss and high number of events, and a high average number of records affected (the bubbles have larger sizes than the sector in the top left corner of the graph). This seems to indicate that depending on the intrinsic nature of the business sectors, for some sectors there is a connection between a high number of records stolen, which translates into high losses. However, for other sectors, a larger number of records does not necessarily translate into greater losses. For instance, records stolen in sectors such as "Mining, Quarrying, Soil and Gas Extraction," "Agriculture, Forestry, Fishing, and Hunting," and "Construction" have a lower monetary value than records stolen in "Information" and "Professional, Scientific, and Technical Services."

Figure 3 shows the Advisen cyber risk threat types ranked by frequency and average severity. Each circle represents a business sector, and the area of the circle corresponds to the average number of records affected. The cyber risk type with the highest average loss and average number of records affected is "Digital Breach/Identity Theft." Looking at Fig. 3, cyber risk types can be divided into three groups according to their average loss:

(1) average loss lower than $2 million: "Cyber Extortion," "Denial of Service(DDOS)/System Disruption," "Privacy—Unauthorized Contact or Disclosure," "Data-Unintentional Disclosure," "Identity Fraudulent Use/Account Access," and "Skimming, Physical Tampering";
(2) average loss between $10 and $100 million: "Phishing, Spoofing, Social Engineering," "IT-Configuration/Implementation Error," "Network/Website Disruption," "Data-Malicious Breach," "Privacy-Unauthorized Data Collection" and "IT-Processing Error"; and
(3) average loss greater than $100 million: "Digital Breach/Identity Theft."

Overall, there seems to be no clear-cut relationship between the frequency of events, loss severity, and the number of affected records. The relationship depends also on the business sector and type of cyber threat.

Figure 4 proposes an adapted version of the categorization matrix by the Australian Cyber Security Center (ACSC), in terms of business sectors and average cyber event severity [29]. Advisen risk categories are mapped to ACSC risk types as follows: Disruption of Services ("IT Configuration/Implementation Errors", "Network/Website Disruption", "IT - Processing Errors", "Industrial Controls and Operations"); Damage of Key Sensitive Data ("Identity - Fraudulent Use/account access", "Cyber Extortion"); Malware ("Data - Physically Lost or stolen", "Data - Malicious Breach", "Skimming and Physical Tampering"); Low Level Malicious Attack ("Privacy - Unauthorized Data Collection", "Phishing, Spoofing, social engineering", "Data- Unintentional Disclosure"); Scanning ("Privacy - Unauthorized Contact or Disclosure"). Figure 4 illustrates the heterogeneity in

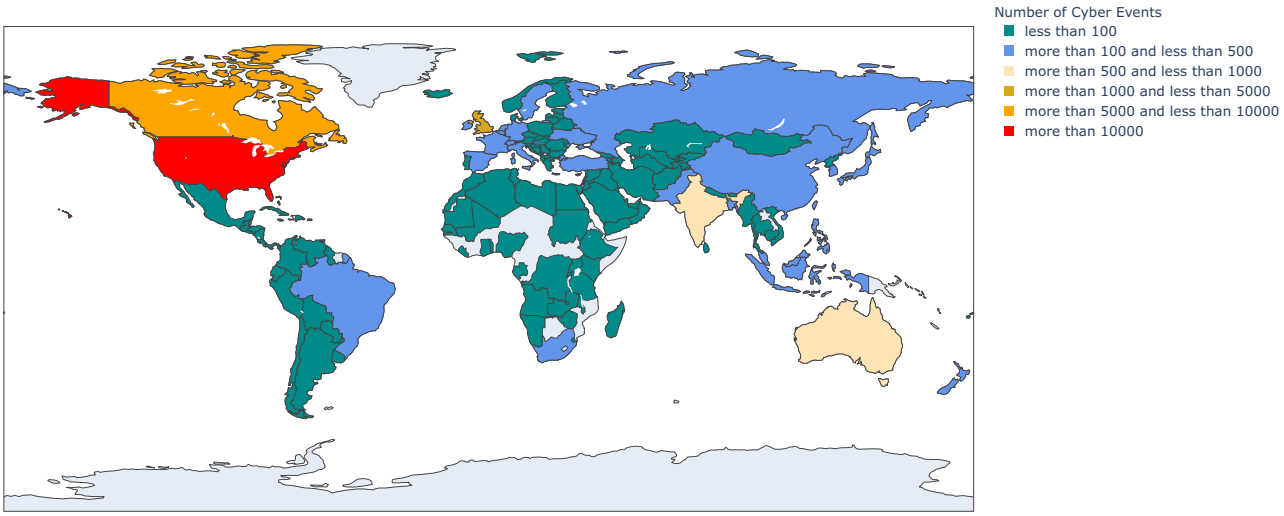

Figure 1: Number of cyber events by country during the period 2008–2020 across all loss categories.

Table 1. This table reports descriptive statistics of cyber risk-related losses aggregated by categories. All dollar values are reported in million dollars. The losses exhibit great variability in terms of median and first four moments across the considered risk types. "Digital Breach/Identity Theft," "IT—Processing Errors," and "Privacy—Unauthorized Data Collection" have the highest average loss amongst all cyber risk categories

Risk category	*N*	Mean	Median	SD	Skew	Kurt
Phishing, spoofing, social engineering	202	12.36	0.57	79.3	9.72	95.53
Privacy—unauthorized contact or disclosure	1916	3.05	0.03	23.8	31.75	1185.92
Data—unintentional disclosure	217	1.34	0.1	8.81	12.73	172.27
Privacy—unauthorized data collection	133	46.77	0.45	434.07	11.18	124.5
Data—malicious breach	858	22.13	0.5	171.64	17.33	360.13
Identity—fraudulent use/account access	689	1.2	0.03	6.55	10.28	124.79
Data—physically lost or stolen	97	23.91	0.24	202.14	9.63	91.16
Skimming and physical tampering	91	1.72	0.05	6.08	6.14	42.59
IT – Processing Errors	44	76.55	0.66	264.77	5.32	29.25
IT—configuration/implementation errors	63	17.06	0.8	43.3	3.23	10.41
Network/website disruption	181	18.77	0.16	68.85	4.76	23.32
Cyber extortion	137	0.52	0.01	2.78	6.86	48.24
Digital breach/identity theft	11	469.22	30.0	1064.11	2.62	5.24
Denial of service (DDOS)/system disruption	1	0.39	0.39	–	–	–
Undetermined/other	21	1.53	0.65	2.43	3.25	10.68
Industrial controls and operations	6	30.7	2.07	62.39	1.78	1.18

the severity of cyber events both for cyber risk type and business sector. In particular, sectors with high average loss in Fig. 2 report higher average losses for every ACSC cyber risk type, than those sectors with low average loss in Fig. 2. This empirical fact emphasizes the dependence of monetary losses on specific company features, since companies operating in different business sectors have a different business model, a different internal structure, and different levels of cyber risk resilience.

The frequency and severity of cyber events

In the following, we examine the frequency and severity of cyber events in more detail.

Figure 5 illustrates the distribution of the number of cyber attacks per company between 2008 and 2020. More than 40% of the companies suffered from cyber crimes more than once during this period, with almost 3% of the firms being affected >10 times. It is important to note that the dataset contains only information regarding cyber risk-related events, which have been publicly disclosed, and it would be safe to assume that the real number of events is much higher.

Figure 6 shows the number of cyber events for the five business sectors that were most affected. We find that the frequency of reported cyber-related events has substantially increased between 2008 and 2016 (4800 reported events in 2008 versus 16 800 reported events in 2016). There is also a significant delay in the reporting of events that needs to be taken into account when drawing conclusions on the risks. As it can be seen from the graph, the number of events appears to be decreasing after 2016 (the period corresponding to the dashed area in the graph). Given that the decline is consistent across all business sectors, this seems to suggest the presence of a reporting delay, rather than a systematic improvement in cyber threat prevention, detection, and response mechanisms common for every business sector. Such reporting delay can be attributed to numerous factors, such as the reluctance of enterprises and companies to report cyber risk-related events, and the data collection procedure

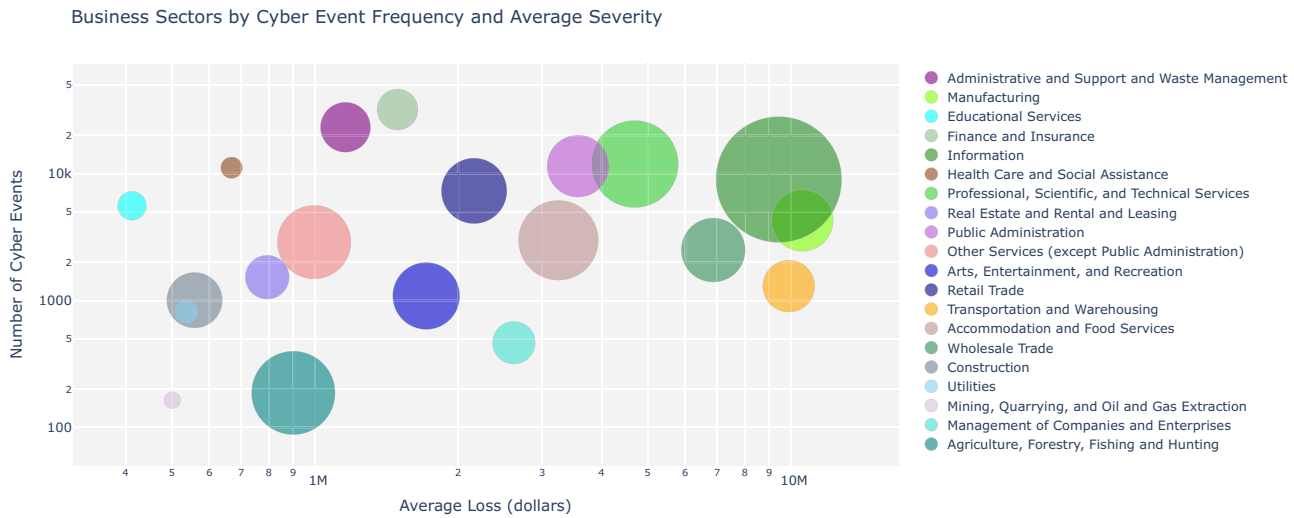

Figure 2: Frequency and severity of individual cyber-related events and the number of affected records (indicated by the size of the circle) across business sectors.

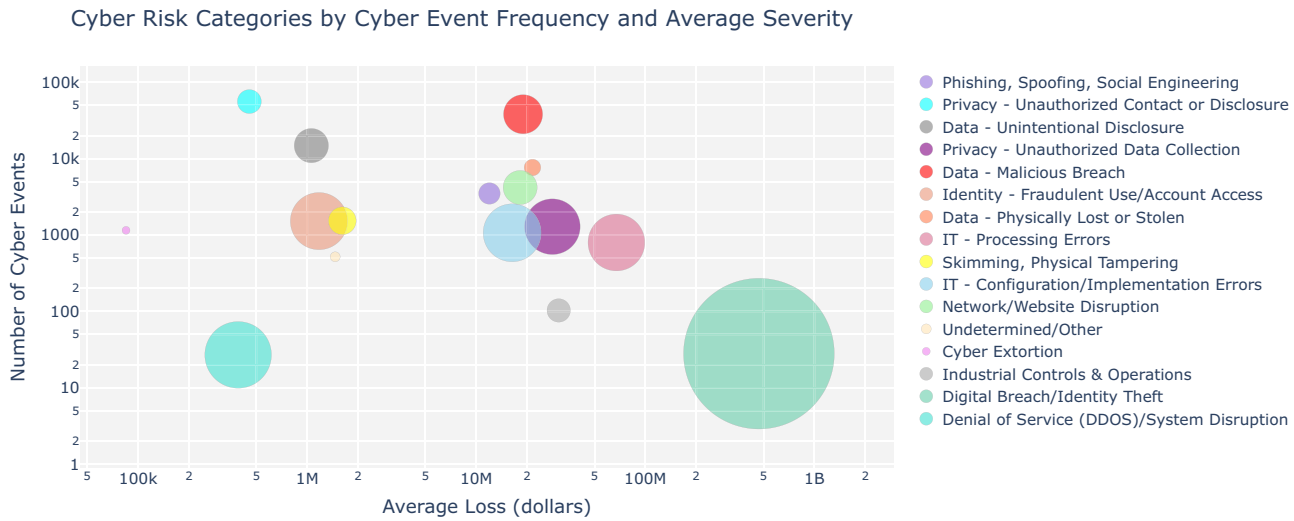

Figure 3: Frequency and severity of individual cyber-related events and the number of affected records (indicated by the size of the circle) across risk categories.

employed by Advisen that abide to the United States of America Freedom of Information Act regulation [34]. As a matter of facts, while US-domiciled companies have a 60-day window between discovering the data breach and reporting it to affected parties, non-US-domiciled entities do not have such strict requirement.

Figure 7 illustrates the percentage of events that falls into a specific cyber risk category for the period 2008–2020. The figure also shows the dynamic nature of cyber risk, with substantially changing shares for different event types. In particular we find that cyber risk categories such as "Data—Unintentional Disclosure," "Data—Malicious Breaches," "Network/Website Disruption" have become increasingly more common since 2008. Moreover, in recent years, "Cyber Extortion" and "Phishing, Spoofing, Social Engineering" are on the rise, reflecting the capability of cyber criminals to adapt and create new forms of cyber threats. At the same time, the share of events for the category "Data—Physically Lost or Stolen" that played a major role in the years 2008–2011 has dropped significantly.

Cyber attacks are time varying in nature, and so are also the root causes of losses. Figure 8 reports the share of total cyber-related losses that can be attributed to the different risk types for each year. Recall that for the frequency of different cyber risk categories we found a relatively clear structure as indicated by Fig. 7. However for the severity of events, there is much more heterogeneity in the cyber risk categories across the time period. Recall that some losses from cyber events are extremely high, leading to a situation where in some years very few events (or even an individual event) can make up a relatively high percentage of the total loss during that year. Nonetheless, we find that losses from "Data—Malicious Breaches" are typically among the highest, while this risk category can also be classified as the most severe risk type over the period 2017–2020. For other years, a high share of losses could be attributed to "Phishing, Spoofing, Social Engineering" in 2008, "Digital Breach/Identity Theft" in 2012, "Privacy – Unauthorized Data Collection" in 2013, and "Network/Website Disruption" in 2017.

Not only does the nature of cyber risks change through time, but also companies in different business sectors suffer losses due to cyber events. Figure 9 shows the share of total cyber-related losses that can

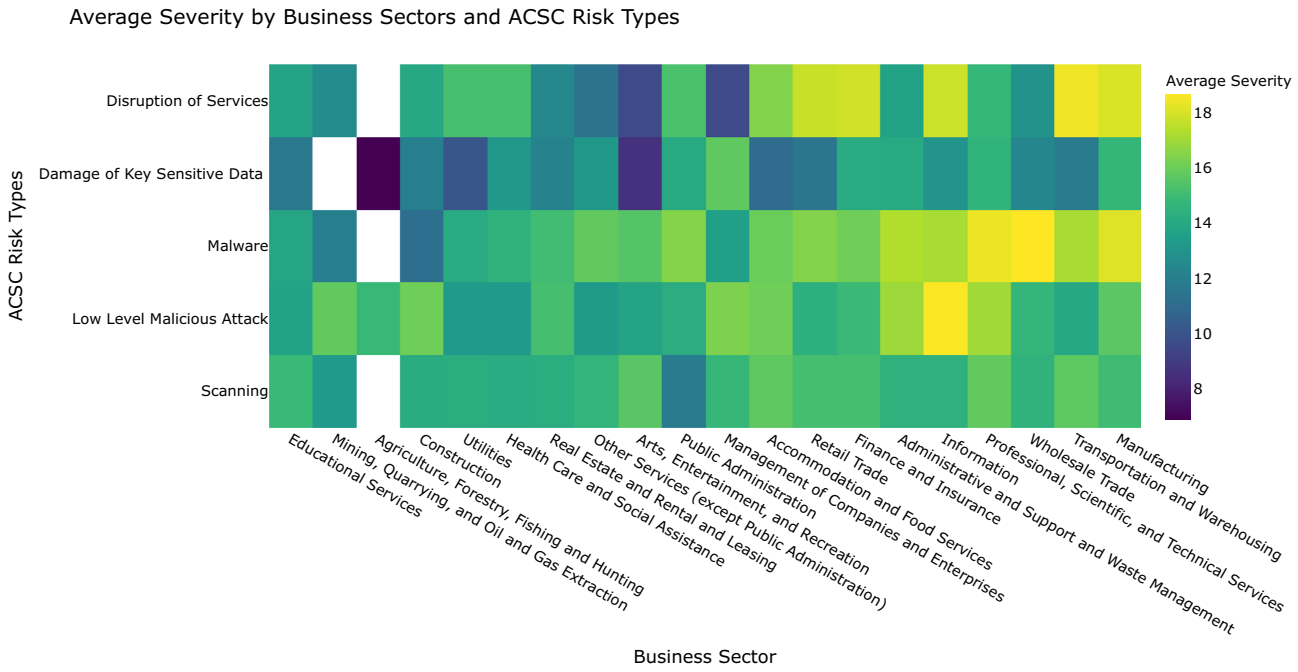

Figure 4: Average severity (on the log scale) of cyber-related events by business sectors according to the ACSC classification. Sectors with high average losses show higher average severity in all the ACSC cyber risk types, than those with low overall average losses. Risk cells with no data available are indicated by white color.

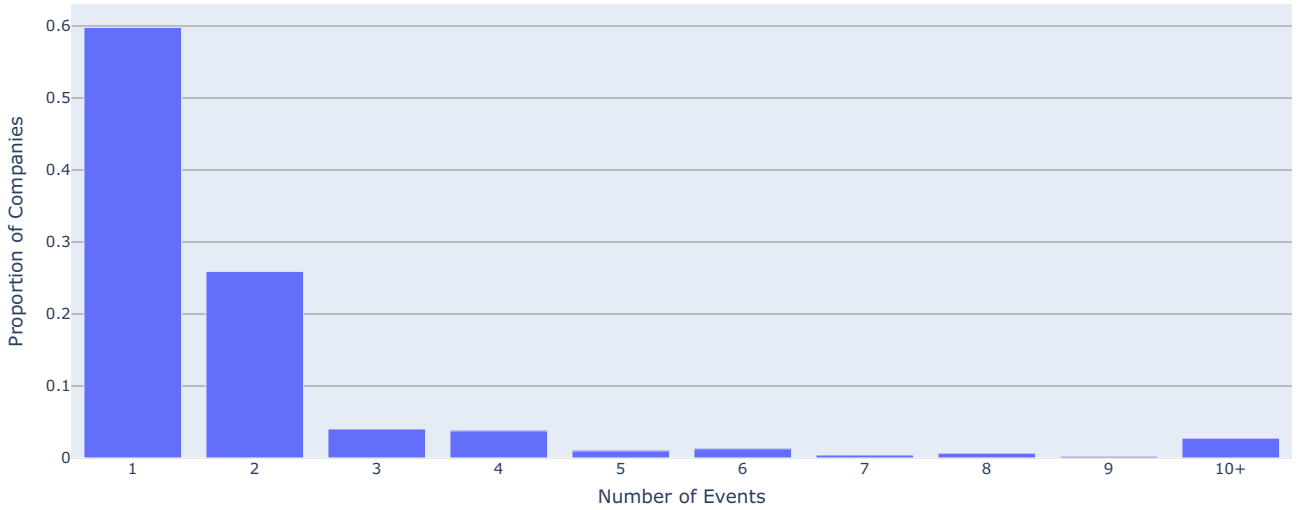

Figure 5: Distribution of the number of cyber-attacks per company.

be attributed to a specific business sector. Our results indicate that the "Information," "Professional Scientific and Technical Services," and the "Finance and Insurance" sectors typically seem to be among the most affected business sectors. However, this figure also illustrates that other sectors can be heavily affected by cyber events, e.g. "Public Administration" in 2015 and "Transportation and Warehousing" in 2020.

Overall, the frequency and severity of cyber-related losses exhibit a very dynamic and time-varying nature. While the occurrence of events seems to be dominated by certain risk categories, extreme losses occur in various cyber risk categories or business sectors. This behavior also makes it particularly difficult to predict the nature or magnitude of losses from cyber-related events.

The severity distribution

Finally, we look at the severity distribution of cyber-related events across all risk categories and business sectors. Considering our sample, the majority of losses are typically relatively small, i.e. 85% of events cause losses <$2 million. However, we also observe a number of more extreme losses in the database: 5% of losses exceed $10 million, while 1.4% of cyber-related losses exceed $100 million, and 0.17% of events cause losses that are >$1 billion. Thus, the distribution of losses is clearly asymmetric, and contains some extreme observations.

Figure 10 shows the fit of a lognormal probability density function to the severity of cyber events (blue line). The gray vertical lines correspond to, in ascending order of magnitude, the estimated me-

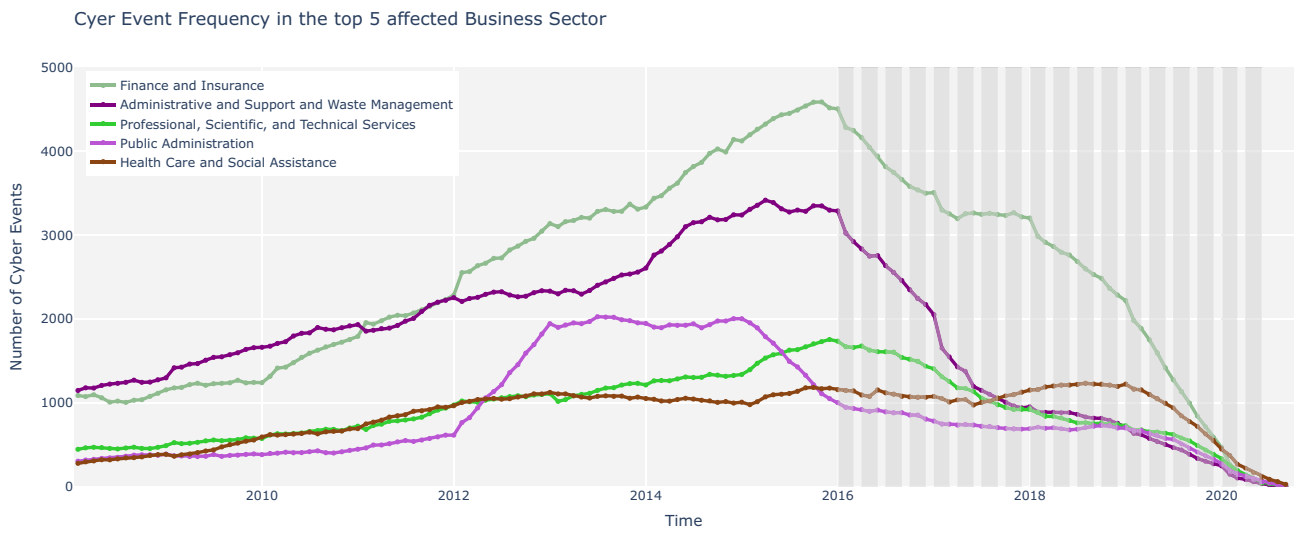

Figure 6: Number of cyber events (based on a 1-year rolling window with 1 month step) for the five business sectors that were most affected.

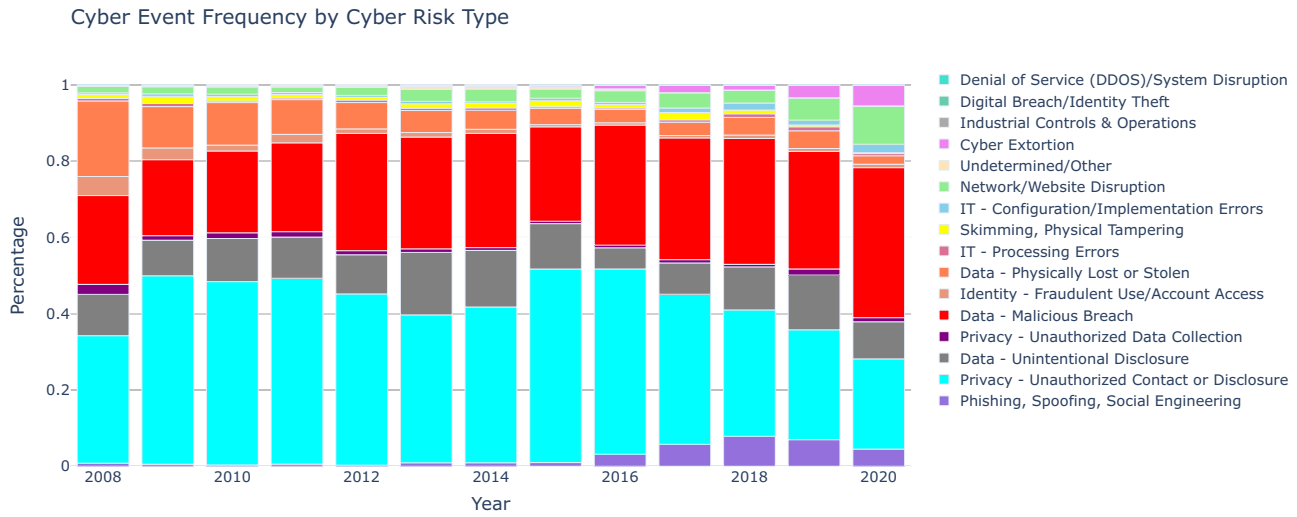

Figure 7: Share of cyber events for different cyber risk types for the period 2008–2020. "Data—Unintentional Disclosure," "Data—Malicious Breaches," "Network/Website disruption," and "Phishing, Spoofing, Social Engineering" have become increasingly more common.

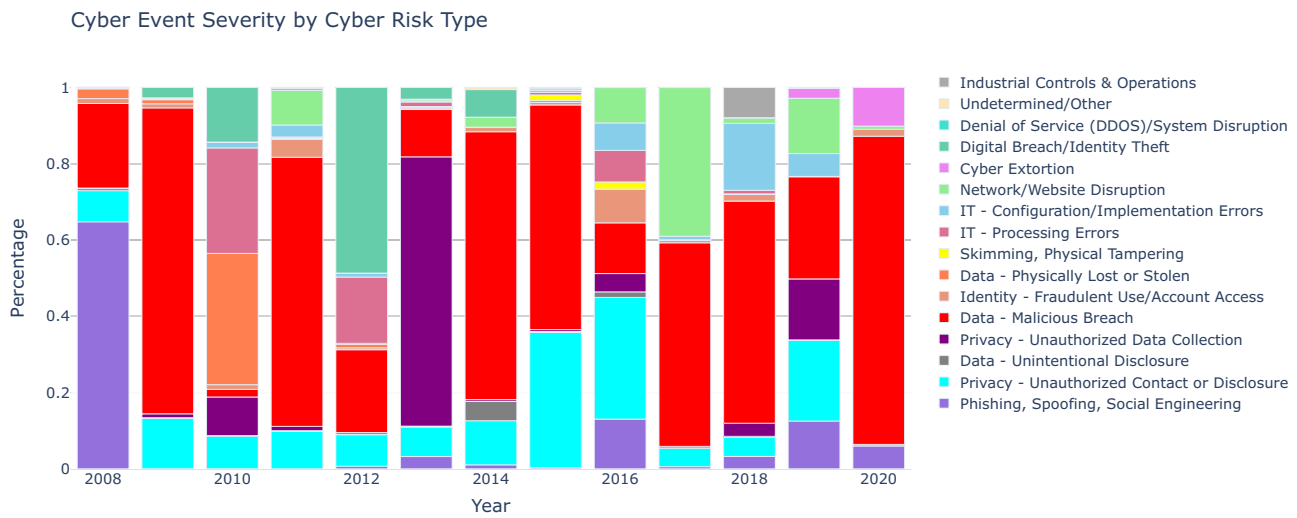

Figure 8: Share of total cyber-related losses that can be attributed to individual risk types for each year 2008–2020.

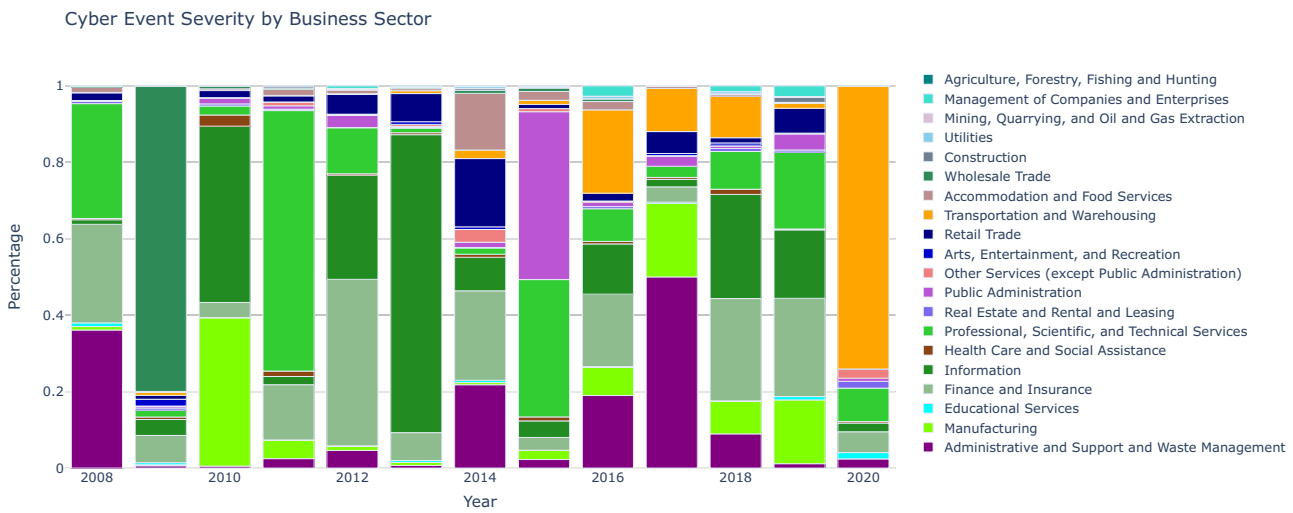

Figure 9: Share of total cyber-related losses that can be attributed to individual business sectors for each year 2008–2020. "Information," "Professional Scientific and Technical Services," and "Finance and Insurance" are the most affected business sectors.

dian of losses from cyber events ($0.11 million), the estimated 90% quantile ($6.36 million), and the estimated 95% quantile ($20.18 million). The estimated Akaike information criterion (AIC) for model comparison is 3320. Interestingly, the mean of the loss distribution ($16.84 million) is substantially higher (around 160 times) than the median. Note that the mean loss is even higher than the 90% quantile of the distribution, confirming the substantial influence of a small number of very extreme events on the loss distribution. Moreover, the cumulative top 0.5% highest losses yield approximately the same amount as the cumulative bottom 99.5%.

This empirical result is consistent with other findings in the literature, where the loss distribution associated with cyber events usually presents heavy tails (see, for instance [9, 14, 32]). In insurance, this is also reflected in the well known concept of "one loss causes ruin," where the probability of one single event, having the potential to trigger losses so extreme that the company could fail to recover from is greater than zero. This is also consistent with results of statistical regression models considered in [9], where cyber event severity was found to follow a distribution so heavy tailed that, depending on risk types and company characteristics, it may present infinite mean. As a robustness check, Fig. 10 also shows the fitted probability density function for three other commonly used distributions in modeling losses: generalized Pareto (in red), loggamma (in green), and Weibull (in purple). The estimated AIC is 3864, 23 983, and 4075, for the generalized Pareto, loggamma, and Weibull distribution, respectively. In terms of quantiles, loggamma and Weibull present comparable estimates to the lognormal case, while the 90% and 95% quantile estimates in the generalized Pareto case are much higher than those of the other distributions, with the 95% returning a value of $108.6 million. This confirms further that appropriate cyber event loss distributions present heavy tails, and that cyber event losses have the potential to be catastrophic.

Conclusions and Policy Recommendations

In this article, we used a comprehensive dataset of cyber-related events to study the nature of cyber risk losses across different risk categories and business sectors. In particular, we focused on the relationship between the frequency and severity of individual events and the number of affected records.

The studies undertaken demonstrated that over 60% of companies that recorded cyber-related losses have suffered from cyber-attacks more than once in the period 2008–2020. This suggests that governance processes relating to mitigation of such events could be enhanced and that regulation and reporting around best practices as it emerges could help mitigate repeated events of the same nature from reoccurring.

It is also clear from the analysis that even with increasing scrutiny and increasing regulatory guidance that occurred in many industry sectors over the period of study, the rate of cyber crime has not abated. In fact, the frequency of reported cyber-related events has substantially increased between 2008 and 2016 (from 4800 reported events in 2008 to 16 800 reported events in 2016). Furthermore, the reporting of such events for modeling purposes could be improved as there appears to be a significant delay in the reporting of events that needs to be taken into account when drawing conclusions on the risks.

Furthermore, we found no distinct pattern or clear-cut relationship between the frequency of events, the loss severity, and the number of affected records. Contrary to assumptions often made in practice, the reported loss databases do not demonstrate a direct proportional relationship between total loss incurred from a cyber event and attributes from the event such as the number of compromised records (data records breached or stolen), the number of employees in a corporation or the number of units of a company affected. This finding shows that all companies, no matter the volume or size of data record can be susceptible to significant incurred loss from cyber events.

The frequency and severity of the events depend on the business sector and type of cyber threat. The most significant cyber loss event categories, by number of events, were "Privacy—Unauthorized Contact or Disclosure" and "Data—Malicious Breach." Data-related breaches have become increasingly more common since 2008, while "Cyber Extortion," "Phishing, Spoofing, Social Engineering" practices also continue to increase, the pace at which malicious breach-related events have occurred has now surpassed these other prominent categories of loss event risk type in recent years. In terms of business sectors, the "Information" sector, "Professional Scientific and Technical Services," and "Finance and Insurance" have suffered most of the financial damage during the sample period 2008–2020.

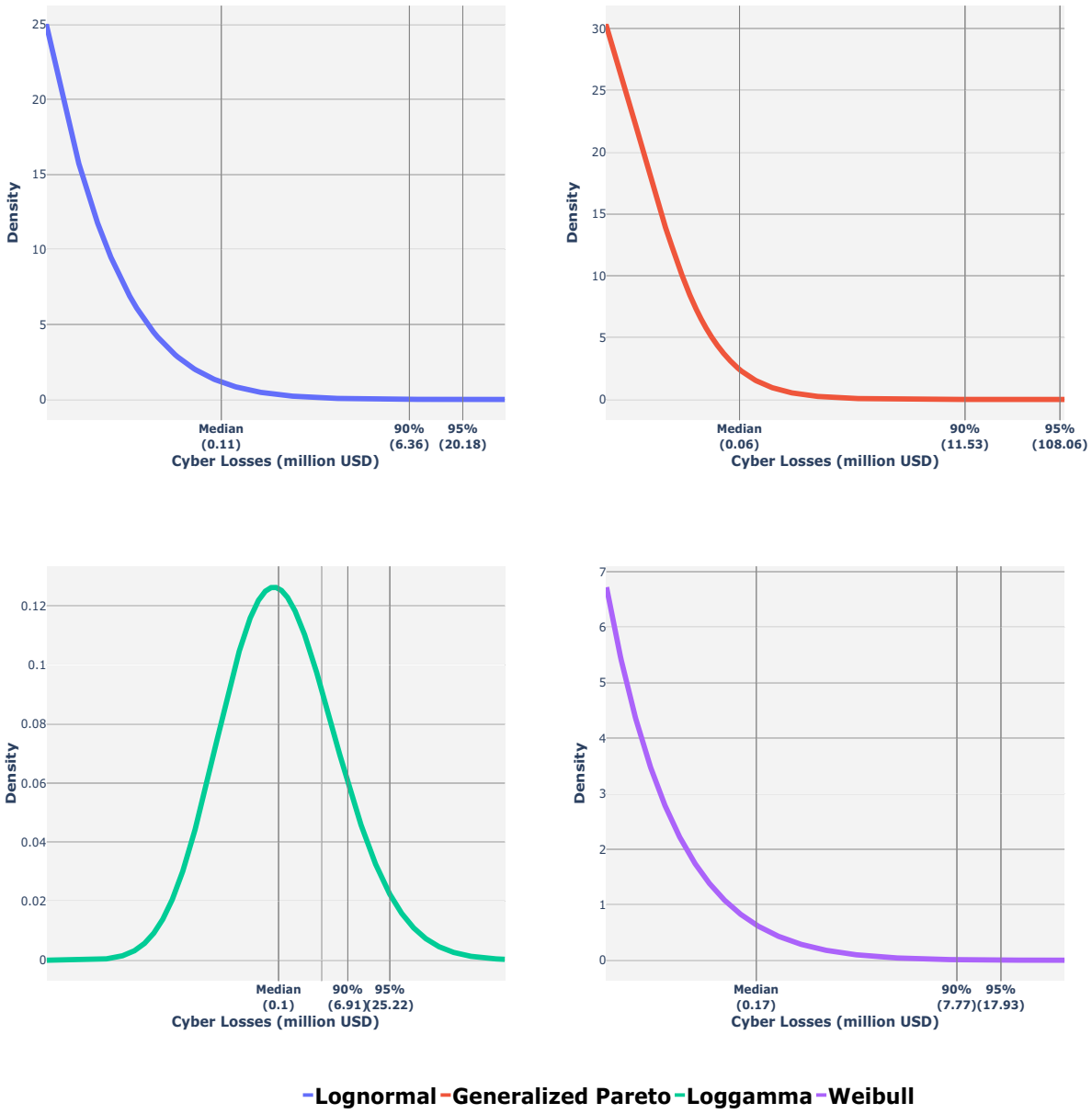

Figure 10: Fitted probability density functions for cyber event severity in the case of lognormal (blue), generalized Pareto (red), loggamma (green), and Weibull (purple) distributions. The gray lines correspond to, in ascending order of magnitude, the estimated median, the estimated 90% quantile, and the estimated 95% quantile in each case.

Furthermore, the findings of the analysis re-affirmed that losses from cyber-related events are heavy-tailed. We found that the majority of losses were relatively small, i.e. 85% of events cause losses <$2 million, while a very small number of events caused losses that were even greater than $1 billion. Furthermore, the mean of the loss distribution (~$16.8 million) was around 160 times higher than the median (~$108 000). Thus, cyber losses are well-represented by the expression "one loss causes ruin" adage. As such, in all categories of cyber loss type and in all sectors of the economy it was found that loss severity is often dominated by large individual events.

Our findings also lead to a number of policy recommendations and suggestions for future work. Currently, data collection and databases on losses from cyber events have an unbalanced recording of samples with the strongest emphasis on the US data collection. However, cyber risk is international in nature affecting both industry as well as government agencies across all sectors of the economy. Similar to other risks such as credit or operational risk a concerted effort should be made to develop an adequate data collection and classification process for cyber-related risks in the international landscape. This should include detailed information on the event types, the failure modes that led to the loss events, and the components of the loss events broken down into categories. Thus, it would be beneficial for sector specific regulators to continue to develop a working taxonomy and reporting framework specific to the risk profiles and needs of different industry sectors. These should not be unified across all sectors of the economy but tailored to particular sectors to capture the heterogeneous nature of cyber risk data event types and loss behavior profiles.

An effort should also be made in order to increase the awareness of cyber risk, even among small business and entities. Due to their

heavy-tailed nature, losses from cyber events can have catastrophic consequences against which even appropriate insurance policies might not be adequate to cover such losses. As cyber losses are currently significantly under-insured, there is a potential large exposure gap that could amount to tens of billions of dollars. This problem should be addressed in the near future: with maturity of the reporting frameworks and consistency in the loss data collection, insurers will be able to more reliably price and design insurance contracts to mitigate some of the losses incurred from cyber risk, which in tandem with improving risk governance is a key component of risk transfer for this category of risk. Currently, the provided cyber insurance products have grown as a market exponentially, however, their scope of coverage is extremely limited and bespoke in nature, making insurance premiums prohibitive to many markets. Consequently, at present, not all losses can be covered by insurance as there are extreme risks that pose a new challenge to the financial viability of insurers.

Finally, the dynamically changing nature of cyber risk dictates that enterprises, insurers, and government agencies have to constantly update their cyber hygiene practices. The attribution of cyber-related losses to different risk types and business sectors seems to change over time, making it particularly challenging to accurately predict the magnitude of losses from cyber-related events. Estimation of cyber risk based on historical losses only is backward looking and might miss this dynamic nature and changing landscape of cyber attacks. A way to overcome this problem could be the use of scenario analysis to account for a constantly changing environment. There are various techniques in operational risk practice that allow to combine historical data and scenario analysis for estimation of risk frequency and severity; see e.g. [35, 36] or for a book length treatment ([6], chapters 14, 15). Using both historical data as well as additional information based on opinions of experts in the field might help to provide a more forward-looking quantification and better management of the risks from cyber-related events.

Acknowledgment

This research has been conducted within the Optus Macquarie University Cyber Security Hub and funded by its Risk Management, Governance, and Control Program.

Authors' Contributions

P.V.S. (conceptualization, formal analysis, funding acquisition, investigation, methodology, project administration, resources, supervision, and writing – original draft, writing – review and editing), J.J. (conceptualization, formal analysis, investigation, methodology, and writing – original draft, writing – review and editing), M.M. (conceptualization, data curation, formal analysis, investigation, methodology, software, validation, visualization, and writing – original draft, writing – review and editing), G.W.P. (conceptualization, data curation, formal analysis, investigation, methodology, project administration, software, supervision, validation, visualization, and writing – original draft, writing – review and editing), G.S. (conceptualization, formal analysis, investigation, methodology, and writing – original draft, writing – review and editing), and S.T. (conceptualization, formal analysis, funding acquisition, investigation, methodology, project administration, resources, supervision, and writing – original draft, writing – review and editing).

Conflict of Interest

Authors declare no conflict of interest.

References

1. World Economic Forum. Global risk report. 2020. https://www.weforum.org/reports/the-global-risks-report-2020. Accessed 1st August 2022.
2. Allison A, Chatzilia A, Canham D. *et al.* Cyber risk resources for practitioners. Technical Report. London: Institute of Risk Management, 2014a.
3. Allison A, Chatzilia A, Canham D. *et al.* Cyber risk executive summary. Technical Report. London: Institute of Risk Management, 2014b.
4. Basel Committee on Banking Supervision. International convergence of capital measurement and capital standards: a revised framework. Technical Report. Basel: Bank for International Settlements, 2006.
5. Cebula JL, Young LR. A taxonomy of operational cyber security risks. Technical Report. Pittsburgh: Pittsburgh Software Engineering Institute, Carnegie Mellon University, 2010.
6. Cruz MG, Peters GW, Shevchenko PV. *Fundamental Aspects of Operational Risk and Insurance Analytics: A Handbook of Operational Risk*. New York: John Wiley & Sons, 2015.
7. Shevchenko PV. *Modelling Operational Risk Using Bayesian Inference*. Berlin: Springer Science & Business Media, 2011.
8. Cohen RD, Humphries J, Veau S. *et al.* An investigation of cyber loss data and its links to operational risk. *J Oper Risk* 2019;**14** (3): 1–25.
9. Malavasi M, Peters GW, Shevchenko PV. *et al.* Cyber risk frequency, severity and insurance viability. *Insur Math Econ* 2022;**106**: 90–114.
10. Allianz Risk Barometer. 2021. https://www.agcs.allianz.com/news-and-insights/reports/allianz-risk-barometer.html. Accessed 1st August 2022..
11. McShane M, Eling M, Nguyen T. Cyber risk management: history and future research directions. *Risk Manag Insur Rev* 2021;**24**: 93–125.
12. Lallie HS, Shepherd LA, Nurse JRC. *et al.* Cyber security in the age of COVID-19: a timeline and analysis of cyber-crime and cyber-attacks during the pandemic. *Comput Secur* 2021;**105**:102248.
13. Biener C, Eling M, Wirfs J. Insurability of cyber risk: an empirical analysis. *Geneva Pap Risk Insur Iss Pract* 2015;**40**:131–58.
14. Eling M, Wirfs J. What are the actual costs of cyber risk events?. *Eur J Oper Res* 2019;**272**:1109–19.
15. Gordon LA, Loeb MP, Sohail T. A framework for using insurance for cyber-risk management. *Commun ACM* 2003;**46**:81–5.
16. World Economic Forum. Cyber information sharing: building collective security. 2020. https://www.weforum.org/reports/cyber-information-sharing-building-collective-security. Accessed 1st August 2022.
17. Parliament of Australia. Privacy Amendment (notifiable data breaches) Bill 2016. 2017. https://www.aph.gov.au/Parliamentary_Business/Bills_Legislation/Bills_Search_Results/Result?bId=r5747. Accessed 1st August 2022.
18. Romanosky S. Examining the costs and causes of cyber incidents. *J Cybersecur* 2016;**2**:121–35.
19. Resnick SI. *Heavy-Tail Phenomena: Probabilistic and Statistical Modeling*. Berlin: Springer Science & Business Media, 2007.
20. JRC. European cybersecurity centres of expertise map. 2018. https://publications.jrc.ec.europa.eu/repository/handle/JRC111441. Accessed 1st August 2022.
21. Peters G, Shevchenko PV, Cohen R. Understanding cyber-risk and cyber-insurance. In: Maurice D, Fairman D, Freund J (eds), *Fintech: Growth and Deregulation*. London: Risk Books, 2018, 303–330.
22. Peters G, Shevchenko PV, Cohen R. *et al.* Statistical machine learning analysis of cyber risk data: event case studies. In: Maurice D, Fairman D, Freund J (eds), *Fintech: Growth and Deregulation*, London: Risk Books. 2018, 75–99.
23. JRC. A proposal for a European cybersecurity taxonomy. 2019. https://publications.jrc.ec.europa.eu/repository/handle/JRC118089. Accessed 1st August 2022.
24. Cebula JJ, Popeck ME, Young LR. A taxonomy of operational cyber security risks version 2. Technical Report. Pittsburgh: Pittsburgh Software Engineering Institute, Carnegie Mellon University, 2014.

25. CRO. Cyber resilience: the cyber risk challenge and the role of insurance. 2014. https://www.thecroforum.org/2014/12/19/cyber-resilience-cyber-risk-challenge-role-insurance/. Accessed 1st August 2022.
26. CRO. CRO Forum Concept Paper on a proposed categorisation methodology for cyber risk. 2016. https://www.thecroforum.org/wp-content/uploads/2016/06/ZRH-16-09033-P1_CRO_Forum_Cyber-Risk_web-2.pdf. Accessed 1st August 2022.
27. Cyentia. Information risk: insight study. 2016. https://www.cyentia.com/iris/. Accessed 1st August 2022.
28. NIST. Standards for security categorization of federal information and information systems. 2004. https://csrc.nist.gov/publications/detail/fips/199/final. Accessed 1st August 2022.
29. Australian Cyber Security Centre. Annual cyber threat report. 2020. https://www.cyber.gov.au/acsc/view-all-content/reports-and-statistics/acsc-annual-cyber-threat-report-july-2019-june-2020. Accessed 1st August 2022.
30. Executive Office of the President Office of Management and Budget. North American Industry Classification System. 2020. https://www.census.gov/naics/. Accessed 1st August 2022.
31. Edwards B, Hofmeyr S, Forrest S. Hype and heavy tails: a closer look at data breaches. *J Cybersecur* 2016;**2**:3–14.
32. Eling M, Loperfido N. Data breaches: goodness of fit, pricing, and risk measurement. *Insur Math Econ* 2017;**75**:126–36.
33. Maillart T, Sornette D. Heavy-tailed distribution of cyber-risks. *Eur Phys J B* 2010;**75**:357–64.
34. The U.S. Department of Justice. The Freedom of Information Act. 2016. https://www.foia.gov/. Accessed 1st August 2022.
35. Shevchenko PV, Wüthrich MV. The structural modeling of operational risk via Bayesian inference: combining loss data with expert opinions. *J Oper Risk* 2006;**1** (3):3–26.
36. Lambrigger DD, Shevchenko PV, Wüthrich MV. The quantification of operational risk using internal data, relevant external data and expert opinion. *J Oper Risk* 2007;**2** (3):3–27.