JOURNAL OF CYBERSECURITY

Journal of Cybersecurity, 2018, 1–15
doi: 10.1093/cybsec/tyy006
Review article

Review article

A taxonomy of cyber-harms: Defining the impacts of cyber-attacks and understanding how they propagate

Ioannis Agrafiotis[1,*], Jason R. C. Nurse[2], Michael Goldsmith[3], Sadie Creese[4], and David Upton [5]

[1]Department of Computer Science, University of Oxford, Oxford, OX1 3QD, UK, [2]School of Computing, University of Kent, Canterbury, CT2 7NF, UK, [3]Department of Computer Science, University of Oxford, Oxford, OX1 3QD, UK, [4]Department of Computer Science, University of Oxford, Oxford, OX1 3QD, UK and [5]Saïd Business School, University of Oxford, Oxford, OX1 1HP, UK

*Corresponding address: Department of Computer Science, University of Oxford, Oxford, OX1 3QD, UK.
E-mail: ioannis.agrafiotis@cs.ox.ac.uk

Received 25 August 2017; revised 23 July 2018; accepted 6 September 2018

Abstract

Technological advances have resulted in organizations digitalizing many parts of their operations. The threat landscape of cyberattacks is rapidly changing and the potential impact of such attacks is uncertain, because there is a lack of effective metrics, tools and frameworks to understand and assess the harm organizations face from cyber-attacks. In this article, we reflect on the literature on harm, and how it has been conceptualized in disciplines such as criminology and economics, and investigate how other notions such as risk and impact relate to harm. Based on an extensive literature survey and on reviewing news articles and databases reporting cyber-incidents, cybercrimes, hacks and other attacks, we identify various types of harm and create a taxonomy of cyber-harms encountered by organizations. This taxonomy comprises five broad themes: physical or digital harm; economic harm; psychological harm; reputational harm; and social and societal harm. In each of these themes, we present several cyber-harms that can result from cyber-attacks. To provide initial indications about how these different types of harm are connected and how cyber-harm in general may propagate, this article also analyses and draws insight from four real-world case studies, involving Sony (2011 and 2014), JPMorgan and Ashley Madison. We conclude by arguing for the need for analytical tools for organizational cyber-harm, which can be based on a taxonomy such as the one we propose here. These would allow organizations to identify corporate assets, link these to different types of cyber-harm, measure those harms and, finally, consider the security controls needed for the treatment of harm.

Key words: cybersecurity; risk; cyber-attack impacts; harm; organisational security; information systems

Introduction

Society depends heavily on technology for interaction, commerce and industry. While technology has led to significant advances in these areas, particularly through the use of Internet, it also has exposed organizations and individuals to a host of new risks resulting from attacks through digital interfaces. These include, for example, denial-of-service (DoS) attacks on networks, data breaches on corporate and personal devices, and viruses that can cripple computer infrastructures [1]. Theft of corporate secrets, sabotage of systems in order to compromise services and systems integrity, and the copying of customer data to sell their identities on the dark web (in order to facilitate other crimes) are all examples of the kinds of acts

that are perpetrated and can all result in harm to an enterprise which is dependent on digital technologies to conduct their business, and which are often custodians of people's data and metadata about people. We initially define cyber-harm as the damage that arises as a direct result of an attack conducted wholly or partially via digital infrastructures, and the information, devices and software applications that these infrastructures are composed of. Understanding the nature of such cyber-harm is critical to ensure that the controls and methods of mitigation we deploy are effective and proportionate to the risks. This article surveys the literature with a view to elucidate the nature of cyber-harm and to underpin further research aimed at analytical frameworks for reasoning about such harm.

Approaches to identifying risk arising from cyber-attacks

To address risks arising from cyber-attacks, many and various solutions have been proposed. These include processes and technologies designed to prevent unauthorized and potentially threatening actors from accessing the digital systems and assets. They also include novel intrusion-detection and prevention systems designed to identify emergent threats and help organizations limit any resulting harm. There is a general acceptance that digital infrastructures are socio-technical systems, and therefore the people involved must also be considered an attack surface for the purpose of preventing cyber-attacks and mitigating cyber-risk.

Threats and attacks have traditionally been at the centre of organizational security and cyber-risk discussions, as noted by the US National Institute of Standards and Technology (NIST) [2]; looking at these is an intuitive response, since to prevent cyber-harm we must know how we might be attacked in cyberspace. One approach to assess the resulting harm is to be able to anticipate such threats and their likely intent. An alternative to such a threat-driven approach is to focus security risk analysis on assets and impacts first. Here, the process involves the identification of the impacts on business assets if they are compromised, and then consideration of the threats that could lead to those impacts [2]. Such analysis identifies and prioritizes those components that are critical for the organizations' mission. One advantage of an impact-oriented approach is that the range of impacts that can be identified in an organization is not driven solely by the knowledge of threats and attacks (which is necessarily incomplete, as no one can be sure that they have complete knowledge due to dynamic threat-landscape where novel attacks are developed frequently). In an environment in which the threat landscape for organizations changes rapidly and novel attack-patterns continually emerge, understanding the potential impact of these attacks on organizational assets may alleviate the associated uncertainty, at least initially in risk-management activities.

Regardless of whether the risk analysis begins with threats to assets, or with potential impact on assets, the ultimate result is the enumeration and estimation of the greatest risks faced by an organization. Controls are then selected to address the risks deemed most significant. The primary advantage of such risk-based approaches (whichever is followed) is that the security budget and response are set to be proportionate to the risks faced. Both critically depend upon our ability to accurately prioritize such risks.

However, there exists very little data on the effectiveness of risk controls once they are deployed, and how they might actually result in lower risk exposure across all assets and functions of an organization. This means that we lack the scientific framework or foundation upon which to select and compare the relative benefits of these controls. We suggest that there is a comparable lack of knowledge of the harm, which might result from cyber-attacks. It is this lack of knowledge that may result in the deployment of controls incapable of mitigating the overall harm. Such limitations may prevent us from identifying and understanding all the potential harms that can result and the relationships that might exist between them. Essentially, we may be selecting our risk treatments and controls based on knowledge that does not fully take account of the ways in which harm can emerge, nor of the breadth of harms that can result from a single cyber-attack. If one simply takes each risk and treats it in isolation, one may not see the connection between various risks and the cascade of harms that can result.

Why we need a taxonomy of cyber-harm for organizations

In this article, we present a prototype taxonomy of organizational cyber-harm which should help researchers and practitioners alike to consider the full range of harms that might result from cyber-attacks, when developing risk treatments. This is necessary to underpin our assessment of risk, and also our ability to quantify the harm resulting from such risks. We explore the topic of cyber-harm, with the intention of developing a more holistic understanding of what constitutes organizational cyber-harm than is available in the extant literature. In what follows, we critically examine cyber-harm, including how it and related topics such as cyber-risk, criminology and cyber-economics, feature in existing research and practice as documented in the literature. Next, we focus specifically on defining a taxonomy of the various types of organizational cyber-harm. This is required to adequately model and reason about harms. We present and draw insights from four case studies in order to provide initial indications about how different types of harm in our taxonomy are connected and how cyber-harm may propagate.

Finally, we conclude our work with a brief discussion of the need for analytical tools for organizational cyber-harm. One such tool, in the form of a conceptual model based on our taxonomy and general reflection, is considered which could enable organizations to better understand, achieve and enhance their cybersecurity. We expect to reveal nuances about how these harms may be linked and how their negative impacts on organizations might be measured, and ultimately to support cybersecurity tasks such as harm reduction and the prioritization of cyber-risk for treatment.

What is cyber-harm for organizations?

Traditional definitions of harm

Harm is a concept that has been researched in-depth in various fields including philosophy, psychology, sociology and law; but significantly less in cybersecurity. In the dictionary definitions of harm, the most common relates it to hurt, injury or damage of some sort [3]. Although these definitions may be regarded as accurate representations of the meaning being conveyed, they arguably oversimplify a complex concept that has been the subject of significant thinking and research effort over the last few decades.

In law, for instance, even though the definitions of harm concentrate on injury and damage (as described above), they often extend this to consider the 'subject' of harm, i.e. an individual or the interests of a collective [4]. An example of 'ultimate harm' in the context of an individual, therefore, is death. The medical domain maintains a similar interpretation of harm and focuses on ill-treatment or impairment of an individual's health [5]. Harm is so core to the practice of medicine that many regard the primary duty of a physician as

ascribing to the principle of non-maleficence, i.e. literally doing no harm [5].

Kleinig provides one of the more critical and philosophical discussions on harm, and synthesizes traditional definitions as well as existing research from several disciplines including law, ethics, health and philosophy [6]. Based on his comprehensive reflection, he suggests that harm may be understood as the impairment of the welfare interests of a being, with welfare interests regarded as those necessary to the functioning of individuals as purposeful, self-reflective and responsible agents. This description is insightful for at least two reasons. First, it highlights the conventional use of harm to define a negative consequence (as a result of some action), and secondly, it centres on beings or individuals as the typical subject of harm. This definition accurately captures the use and understanding of harm in other areas such as psychology (e.g. self-harm), medicine and law (e.g. harm to individuals) [7].

In recent years, harm has increasingly been applied in broader contexts, such as harm to companies or industries. For instance, there has been research exploring how environmental violation events harm the reputation of an organization [8] and more topically, analysis of how cyber-attacks can result in harm to businesses and even to the economy of a nation [9, 10].

The relationship between harm, impact and risk in organizations

Narrowing our focus to the enterprise context, two concepts closely related to harm are 'impact' and 'risk'. Both of these concepts feature prominently in the literature and practice of organizational information security. Broadly speaking, impact is the effect of an action by one person or thing upon another and can be either positive or negative. This characterization of impact as a generic term is supported by others in security across academia and government [11, 12].

The European Union for Network and Information Security Agency (ENISA) defines impact as the result of an unwanted incident [13]; this is a definition it borrows from the International Organization for Standardization and the International Electrotechnical Commission (ISO/IEC) [14]. Whilst not definitive, the arguable suggestion here is that impact is adverse. For the NIST, developing an understanding of impact is a significant component of the risk management process for organizations. They describe impact as the 'harm that can be expected to result' from consequences of unauthorized actions or loss of confidentiality, integrity or availability [2]. Their appreciation of impact is clearly oriented on harm, potentially with the intent of stressing 'impact' as undesirable or an impairment of organizational interests. A significant observation that can be made based on our reflection so far is that, although impact is a non-specific term, in security, it often implies a negative outcome [11]. On occasion, this adverse meaning is made explicit through the use of words such as harm.

The term 'risk' is associated with many of the concepts presented above and its theoretical underpinnings are provided by the seminal works of Beck and Giddens [15–17]. According to Beck [17], risk is a modern concept that presupposes decision-making and is a result of the speed of modernization that has transformed our society to a risk society. The increased influence of science and the technological innovations have resulted in two major transformations that define the era of risk society. The first transformation, namely the end of nature refers to the fact that almost all aspects of the physical world are influenced by human interventions, shifting the focus of attention from 'what nature could do to us' to 'what we have done to nature' . The second transformation, namely the end of tradition, describes 'a process of individualisation' [16], where people question traditions, institutions and old societal norms.

A risk society that has experienced these two transformations experiences 'uncontrollable risk' because the risks are now 'manufactured, second-ordered and unnatural' [17]. Unanticipated advances in technology can increase the gap between actual and perceived risks, transform visible risks to invisible to virtual and render these risks borderless—a concept described by Giddens as the 'scientization of nature', 'the colonization of nature' or 'the end of nature' [16]. Therefore, the traditional concept of risk perceived as the probability of an adverse event multiplied by the magnitude of impact must be expanded. In order to expand our understanding of risk, Beck and Giddens suggest that manufactured risks can be analysed in three dimensions: spatial, temporal and social. Spatial, because these risks cross national borders and can affect the globe; temporal because manufactured risks may influence generations that have not been born yet; and social because the effects are a combination of actions of many individuals that shift individual risks to systemic risks.

A major concern with manufactured risks is that societies experience a denial of responsibility from organizations and individuals for creating these risks which results in avoidance of action in terms of risk management, a concept coined as 'organised irresponsibility' [15]. Organized irresponsibility disincentives organizations to invest in controls to mitigate harms and to provide compensation for individuals, despite the fact that they acknowledge the reality of catastrophes [15]. Whilst the risks that Beck and Giddens describe in risk societies are inspired by advances in nuclear, chemical and biomedical technologies, advances in information technology and cyberspace share the same characteristics. Therefore, all these concepts can be adopted from the risk community to help understand the nature of cyber-harm to an organization, by designing a taxonomy on cyber-harms. We believe that such a taxonomy will help enterprises to engage in security risk management tasks intended to identify, assess, prioritize and treat the various risks that they face.

Insights from criminology and white collar crimes

A stream of literature where harm has a pivotal role is criminology in general and the study of white-collar crime in particular. Criminologists, due to difficulties in defining crimes and identifying their detrimental impact, propose to depart from the notion of crime and focus on that of social harm [18–21]. Therefore, harm is key to social policy and observations of different types of harm occurring from crimes shape practical guidance [19], rendering the development of sound methods to systematically assess harm of increasing importance. Greenfield *et al.* [19], present a framework comprising a set of processes to empirically assess harm. They identify five key dimensions where harm may manifest, namely: functional integrity; material support and amenity; freedom from humiliation; privacy or autonomy; and reputation. They also define five magnitude levels of these types of harm and examine the cascading nature of harm by examining real-world crimes that have caused severe harm to society.

In a similar vein, Van Slyke *et al.* [18] construct a taxonomy of harms for white-collar crimes by focusing on the victimization element of these crimes. They examine a series of white-collar crimes and list the costs arising from these offences. They complement desktop research with victim surveys, and focus on the severe lasting effect of harms in certain individuals. Further insights are provided by suggesting that harms can be conceptualized as a pyramid, with chronic

harms at the top, 'one-off' victims who suffered severe losses in the middle and victims who are unaware of the fraud, or have incurred small costs, at the bottom. Secondary effects of harm are also considered, with the authors suggesting that these relate to victims who experience great losses or suffer psychological effects.

Furthermore, Van Slyke's study considers harms that may relate not only to individuals but also to other stakeholders, such as communities, neighbourhoods, governments and society at large. Specific focus is also given to calculating the costs of crime, with the authors arguing for three types of costs, those incurred in anticipation of a crime, those incurred as a consequence of it and those in responding to it. They suggest two approaches to calculating these costs: 'bottom-up' based on surveying crime cases and estimating individually different harms; and 'top-down', trying to estimate how much the public is willing to pay to avoid or reduce these crimes. Brenner presents the first approach to identify metrics for estimating crime that originates in cyberspace [22]. Although she acknowledges that designing metrics and scales for cybercrime is extremely difficult, due to 'apprehension', scale and evidence issues, she proposes a simple taxonomy of harms consisting of three main types, namely individual, systemic and inchoate.

The researchers in the discipline of criminology studied in this survey all concur that estimating the cost of crimes, as well as providing models for assessing harms, present significant technical and methodological challenges [18–20]. These challenges arise due to limited utility of conventional research tools such as surveys, poor statistical data obtained by law-enforcement agencies and the tendency of individuals to conceal crimes from the authorities due to embarrassment or lack of ways to report these crimes [18]. In addition, only a small percentage of cases are prosecuted and there is no consolidated source of information aggregating different crimes or incidents. The keen reader will have recognized the stark similarities with incidents in cyberspace. There are several lessons to be learnt from the discipline of criminology, but we need to emphasize that all the approaches from this context determine harm arising from specific crimes, whereas in this article, we present an asset-driven approach. There are clearly parallels between non-cyber-crime and cyber-crimes from a harm perspective (since their victims are common), which can be used to design a taxonomy of cyber-harm.

Cyber-economics

Felici *et al.* [23] emphasize the need to further explore the field of economics by focusing on cyber-incidents. They postulate that ICT stimulates new markets and is integrated into current economic sectors that foster growth. They argue that the field of cybersecurity economics is essential in assisting ICT to hold this dual role. They further suggest that challenges in this field require a multidisciplinary approach and that models created by researchers must acknowledge the new information regarding cyber-incidents, their impact and their relations to the dynamics of other cyber-actors.

Anderson *et al.* [24] are pioneers in providing a first approach to measuring costs of cyber-incidents. In their article, they highlight the difficulties in assessing impact due to the fast pace of technological developments and the large asymmetries between estimating costs and revenues and their real values. Similar to the models presented in the criminology literature, in their model Anderson *et al.* equate harm with cost and consider direct and indirect costs, defence and crime costs, as well as costs to society. They extend their work by considering concepts from economics such as the 'moral-hazard effect', the hidden-action problem and network neutrality, amongst others, to provide a holistic understanding on the economics of information security [25].

In a similar vein, Moore highlights further challenges in the field of economics of cybersecurity [26]. Drawing from concepts from the field of economics, Moore identifies challenges, inter alia, misaligned incentives such as the natural tension between efficiency and resilience in IT systems, information asymmetries and externalities. He suggests that to overcome these challenges regulatory intervention is necessary. Moore further identifies online identity theft, industrial cyber-espionage, critical infrastructure protection and botnets as the most persistent threats in cybersecurity and proposes a series of regulatory solution options.

Other efforts focus on the evolution of risk frameworks, modelling the resilience of business systems [27]. In these models, researchers try to understand how catastrophes may disrupt globally critical services by examining the interconnectedness of assets. A threat-based model is created and each threat: is attributed with different mechanisms of destruction; is related to specific vulnerabilities; and presents different challenges for the resilience of systems. The taxonomy of threats is developed through an extensive review of historical incidents extended as far back as in 1000AD. Similarly to crime taxonomies, correlations and triggering mechanisms for various types of catastrophes are sought. One of the many classes of threat examined by this article is cyber-threat.

A similar approach is proposed by Lloyds of London, where they consider fictional but realistic scenarios to understand the concept of cyber-risk aggregation [28]. The authors of the report note that cyber-risk is a growing global threat due to the increase in cyber-incidents during the last years. They utilize two fictional scenarios, namely a 'cloud service provider' hack and a 'mass vulnerability', and seek to calculate direct and indirect costs for both organizations and insurers. They conclude that the potential for a cyber-attack to sweep through many organizations and the secondary effects of the attack due to interdependencies between organizations could have disastrous consequences.

There are a few institutes, which provide aggregate data and publish annual reports of cyber-incidents. For example, the Cyber Security Breaches Survey (CSBS) from the UK Government annually captures trends in cyber-incidents and details of cybersecurity risks [29]. The report presents statistics about how organizations operate in cyberspace and identifies common types of threat. To comment briefly on key findings in the 2017 report, the survey highlights that all UK businesses are potentially exposed to cyber-threats. Government sources of guidance on cybersecurity threats remain few, but 75% of the organizations, which take advantage of this information, find it useful. They have identified that a sizeable proportion of businesses still lack security controls despite the fact that the vast majority of them have increased their cybersecurity budget. The most common types of successful attacks are related to staff receiving fraudulent emails (in 72% of cases where firms identified a successful attack or an attempt). The next most common issue is related to viruses, spyware and malware (33%), people impersonating the organization in emails or online (27%) and ransomware (17%).

Based on such reports and drawing on their previous work [24, 25], Anderson *et al.* provide a set of recommendations in order to address the lack of statistical data in the European Union (EU) and to further the field of security economics [30]. They propose to the EU the introduction of a comprehensive security-breach notification law and the publication of loss statistics. They also identify that common vulnerabilities can trigger cascading effects in cyber-attacks and propose diversity as a security measure. Finally, they

highlight the problem of moral hazard in Critical National Infrastructure (CNI) and propose the regulation of best practice approaches to cybersecurity for these stakeholders.

Focusing on the incentives for CNI and regulatory approaches, Laube *et al.* examine the economics of mandatory security-breech reporting to authorities [31]. They design a principal-agent model able to describe conflicts of interest between regulators and organizations. Their model considers security investment and firms' interdependence, mandatory security-breach reporting and security audits. They conclude that laws, which enforce mandatory security-breach reporting are essential for high-security interdependent firms with the premise that disclosure costs are low.

Kshetri attempts to define a cost-benefit calculus using a similar methodology to Laube *et al.* [32], but he focuses on the perspective of the attacker. He identifies characteristics of cyber-criminals, cyber-crime victims and law-enforcement agents and argues that these three classes of entity, when they interact, lead to a vicious circle of cyber-crime. He provides a calculus that considers the benefits and costs to an attacker and reasons about whether a cyber-crime may occur. It is worth noting that the authors suggest that psychological effects as well as criminal conviction are part of an attacker's benefits or losses.

Edwards *et al.* [33], explore a publicly available dataset of data breaches and apply a Bayesian Generalised Linear Model to unveil trends in data breaches. They conclude that the size and frequency of data breaches has been stable in recent years, but their impact is growing due to the ability of threat-actors to monetize personal information better and to the increasing number of electronic financial transactions. An interesting approach, based on the 'top-down' methodology described in the criminology field, is presented by Nguyen *et al.* [34]. The authors attempted to elicit 'premiums' that some users would be willing to pay to protect their assets from cyber-incidents. Their results show that participants in their survey were willing to pay a premium of between $9 and $11 monthly to protect their social-media accounts, while they were willing to wait between 8 and 9 additional minutes to receive their emails, provided these would be free of spam and phishing emails.

Much of the research on cyber-economics is naturally intended to be viewed through a societal or supply-chain economy lens, but it has consequences for organizations as well and places harm located at a single organization in the context of the wider societal actors who can implement levers which can help organizations mitigate such harms and mandate or incentivize behaviours necessary for success.

Monetizing cyber-incidents

The ability to quantify harm would allow an organization to make better decisions regarding the treatment of a particular risk. We have reflected on current literature to determine the extent to which techniques exist to quantify cyber-harm (or indeed, attack impact). Generally, we found that there is a lack of effective metrics, tools and frameworks for estimating the harm from cyber-attacks on organizations. The approaches that we have identified are either quantitative or qualitative in nature. Most approaches endeavour to monetize the metric output values, in terms of financial loss, in order to be able to compare harm between cyber-incidents. These approaches consider direct and indirect costs emanating from a cyber-attack for different harms [35–38].

Fluctuations in stock market prices have attracted the interest of many researchers, the idea being to compare the price of the stock before and after a cyber-attack. Telang and Wattel [36] focus on software-development companies and report that on average vendors lose 0.6% of their market value when software vulnerabilities are exposed. Regarding exposure or leakage of customer data, Acquisti *et al.* [37] provide significant statistical evidence that there is a negative short-term impact on the value of stocks, but this effect decreases rapidly over time. Further evidence of the negative effects on the market value of an organization that may arise from a cyber-breach once it is made public is presented by Cavusoglu, Mishra and Raghunathan [38]. More recently, there have been concerns regarding associating fluctuations in stock prices with cyber-incidents and, in particular, with data breaches [39].

There are types of attacks, however, that do not seem to have an impact on the value of the stock of organizations, such as DoS [40]. In a similar vein, Campbell *et al.* [41] suggest that there is no impact when the security breach concerns non-sensitive data. There is a distinguishable difference, though, when the breach concerns confidential data, causing the market value of the organization to drop briefly. Finally, Kannan, Rees and Sridhar argue that there is no significant difference in the loss of market value depending on whether the security breach affects the confidentiality, availability or integrity of data [42]. These are all interesting points, but they attest to the difficulty of characterization and quantification of cyber-harm.

Other approaches have focused on 'measuring' harm by way of qualitative severity levels (or brackets, similar to high, medium, low) based on whether certain attacks have harms within defined criteria thresholds. One article, for example, outlines six main levels of risk impact from minor (1) to business-critical (6), and attributes for impact criteria include reputation, human capital and financial [14]. For minor impact, the thresholds are as follows: the reputation threshold is zero to limited negative publicity and no impact on the institution's reputation; human capital threshold is that the attack affects less than 5% of employees and there is no impact on recruitment or retention of staff; and the financial threshold is an annual loss of less than $1 million in the current fiscal year. Each of these thresholds (and associated values) increases as the rating progresses from minor through to moderate, substantial, serious, severe and business-critical. The advantage of such a quasi-quantified approach lies in the fact that accuracy in metrics is not required and it may be possible to obtain a rough estimate of the harm quickly. These thresholds would, of course, change depending on the enterprise.

A very promising approach to quantifying harm is detailed in a report published by the World Economic Forum [43]. The aim of their approach is to understand the benefits from digitalizing functions and services of organizations, the costs that may occur when attacks may be realized, determining the threat imposed to organizations and to try to find the optimal investment in cybersecurity. They introduce the notion of the cyber-Value-at-Risk (VaR) a 'risk measure for a given portfolio and time horizon as a threshold loss value' [43]. VaR considers the probability that a loss will exceed the profits in a given time. Those authors outline the properties that the Cyber-VaR value should have, but highlight that they do not provide the means to quantify and compute these properties. A completed model would be able to provide answers such as 'given a successful cyber-attack, a company will lose not more than X amount of money over a period of time, with 95% accuracy' [43]. The core components of such a model are quantifying the assets under threat, computing the vulnerabilities and creating threat profiles of attackers. In terms of harms, they provide an example of how the assets of an oil company may be impacted and identify harms regarding future revenue loss, litigation and public relations

costs, business interruption costs and reputational damage, even bankruptcy if the attack is persistent for a certain number of days.

It is evident that models reasoning about harm are scarce and are either based on fictional scenarios or try to reason about harms based on statistical data about costs. However, the quantification of harm is still an unsolved problem for organizations. Most approaches have focused on insight from stock-market prices; however, they fall short in estimating the harm related to cyber-attacks and incidents. This is because usually drops in stock-market prices are brief [40, 41], while costs that relate to other types of harm such as physical damages or incident response costs are neglected. Cyber-VaR is promising but much more needs to be done before this becomes a viable option for organizations.

In summary, therefore, we believe that a model that is asset-driven may provide a different perspective on the notion of cyber-harm and insights from criminology and other fields can underpin such efforts. Further research is required on the topic of the quantification of harms (both direct and indirect), potentially through the linkage with assets and threats. We will return to this observation at the conclusion of this article as it provides inspiration for how to evolve from a cyber-harm taxonomy to a model capable of underpinning analytics on cyber-harm and the effectiveness of risk controls in addressing it.

Emergence of cyber-harm as a concept in organizations

The origin of cyber-harm is firmly rooted in the psychological domain and describes the harm or negative impact to individuals that might occur as a result of interactions in cyberspace (e.g. cyberbullying) [44, 45]. In recent years, that term, similarly to 'harm' itself, has been expanded and applied to more general contexts. The adaptation of cyber-harm to cybersecurity more broadly builds on this conceptualization, and aims to focus on the adverse impacts of cyber-attacks across all stakeholders, including individuals, communities, organizations and nations. For instance, there is literature exploring cyber-harm in the domain of cyber-warfare [46, 47]. Here, cyber-harm is loosely perceived as harm perpetrated via the Internet or similar electronic means, most often involving some form of cyber-incident or intentional attack (such as an outsider hacking into an enterprise or an insider inserting an infected drive into a workstation). This description encompasses other research work that suggests that cyber-harm may also be caused via other means, such as cyber-exploitation, where the goal of the attack is primarily to obtain data from the targeted system [48].

To consider cyber-harm in the context of organizations, therefore, is to consider the detrimental impacts resulting from cyber-events or incidents that could take place that would involve the organization in any fashion. Incidents could be intentional attacks such as compromising systems, or unintentional due to mistakes, user errors or broadly natural phenomena, and may derive from external parties as well as from within the organization. This distinction of harm to intentional and unintentional has traditionally been localised to cyber-assets: for instance, a computer network might be infected or a web server forced offline because of a DoS attack. But the reliance of society on technology has positioned such harm also in the physical sphere. The consequence of this is that as cyber and physical spaces overlap, attacks on enterprises using cyberspace can have a tangible, offline harm. As the US Department of Homeland Security states, such harm could also include physical damage to property or bodily harm [49]. Our understanding of cyber-harm should not be limited to the online components of a system, but rather should be extended to include the offline components as well.

There have been several attacks that have exemplified the physical reality of cyber-harm. Two of the most prominent are the recent Ukrainian blackout [50], where malware facilitated the shut down of a power plant and prevented essential systems from rebooting; and the remote hijacking of the Jeep Cherokee, where white hat hackers obtained full control of the vehicle, resulting in car manufacturer Chrysler recalling 1.4 million vehicles before any malicious attack was attempted [51]. The Chrysler attack drew the attention of the automotive industry to the risks that Internet-of-Things (IoT) may pose to all manufacturers. These add to the other better-known impacts of attacks including damaged corporate reputation, loss of customers and business partners, and (financial) compensation to affected parties; as witnessed by Sony, Target and Ashley Madison [52]. It is evident that cyber-harm is potentially more than the sum of the impacts considered in traditional risk assessments, and that a novel taxonomy focusing on understanding the full spectrum of cyber-harm is required. A similar rationale is presented in [53], where the authors reflect on the notion of cyber-harm from a national perspective.

Defining a taxonomy of organizational cyber-harm

To facilitate more effective reasoning about cyber-harm and to address the various challenges identified regarding modelling it, it is useful to describe a taxonomy for organizational harm. This should outline the range of categories of harm and structure them in a way that allows cascading harms to be considered, and in a format that organizations would be able to apply during security risk analyses. A key advantage would also be that it would force consideration of harms not usually deemed 'corporate' and thus rarely properly assessed. A good example of this is the psychological harm to individuals resulting from cyber-attacks. We present such a cyber-harm taxonomy in this section. To support this research, in addition to the literature considered above, we have conducted a comprehensive survey of known cyber-incidents found in publicly available databases [54, 55], in combination with case studies and news reports.

Taxonomy of cyber-harm

There have been several attempts to define the impacts of cyber-attacks [2, 12, 56], however, their use and adoption has been limited. For our taxonomy, we have created and analysed a dataset of news articles, literature and databases of cyber-incidents. More specifically, we have collected news articles, such as [57], published in major newspapers and security magazines, which target national and international audiences. To identify these articles, we searched for articles that contained phrases such as, inter alia, 'cyber attack', 'cyber incident' and 'hackers', commonly used when cybersecurity incidents are discussed. We reviewed literature focusing on taxonomies of harm ranging from white-collar crimes to psychology. Finally, datasets such as Hackmageddon [54] and those from the VERIS Community Database (VCDB) [55], albeit limited in the variety of the cyber attacks they contain, were utilized due to the absence of more holistic datasets. VCDB is a public effort to collect cybersecurity incident reports with a specified structure. Verizon RISK team is responsible for the maintenance of the database, which contains more than 5 000 incidents. Out of these incidents, we focused on the most contemporary reports that contained information relevant to our taxonomy, excluding incidents whose source were physical attacks. Hackmageddon is a well-known cyber-incident website that collects public reports and document on a monthly

basis. The same rationale as that applied with the VCDB regarding extracting relevant incidents was followed here, and we again focused on contemporary reports.

We then applied content analysis [58] to process the sources in our dataset. Content analysis is a qualitative data analysis technique, aiming to identify key 'themes' in documents. There are three approaches to content analysis: the first is the inductive approach that is based on 'open coding', meaning that the categories or themes are freely created by the researcher. In open coding, headings and notes are written in the transcripts while reading them and different categories are created to include similar notes that capture the same aspect of the phenomenon under study. The second approach is deductive content analysis that requires the prior existence of a theory to underpin the classification process. This approach is more structured than the inductive method and the initial coding is crafted by the key features and variables of the adopted theory [58, 59]. In the process of coding, excerpts are ascribed to categories and the findings are dictated by the theory or prior research. However, there could be novel categories that may contradict or enrich a specific theory. Therefore, if deductive approaches are followed strictly, these novel categories that offer a refined perspective may be neglected. This is the reason why we opted for the third type, which is a mixture of the deductive and inductive approaches.

We used harms identified in the literature of white-collar crimes [18, 19] and other taxonomies of harm [2, 11, 60] as core themes for our deductive approach. Themes to which we could not match excerpts from articles and cyber-incident datasets were excluded from our taxonomy. We then considered excerpts that were not allocated to any themes. This process was iterative; we created themes based on an inductive approach and in the following iterations, we merged themes which described substantially identical notions. We concluded the process when there was an iteration in which no further themes could be merged. Two researchers were involved in the process of content analysis. The first person identified the themes and the second verified the content by independently using the proposed thematic schema to replicate the results of the first researcher.

Once we obtained all the relevant themes, we divided the harm types into categories to form hierarchies of harm. Subsequently, we reflected on the resulting structure in the context of a smaller set of cyber-incidents to determine whether the harm from these incidents could be modelled, and incorporating any refinements (e.g. identify incidents that could not be described by the types of harm in the taxonomy) necessary. The hierarchies that we define in our taxonomy contribute to the novelty of the research given that existing models (e.g. [2, 12, 56]) only focus on lists of impacts and losses from cyber-attacks. We believe that the provision of structure through a harm taxonomy is useful, particularly in engaging with different types of stakeholders who may be affected in different ways by cyber-attacks. Moreover, it allows us later to consider how harms propagate across and between different high- and low-level categories in the time period after an attack has occurred. In Fig. 1, we present our taxonomy for organizational cyber-harm, where the main categories are coloured in orange and the subtypes of harm in yellow.

To structure our taxonomy, we have taken inspiration from existing research on categorizations of harm [18, 19]. The main harm types we include are:

- Physical or Digital harm (i.e. harm describing a physical or digital negative effect on someone or something)
- Economic harm (i.e. harm that relates to negative financial or economic consequences)
- Psychological harm (i.e. harm which focuses on an individual and their mental well-being and psyche)
- Reputational harm (i.e. harm pertaining to the general opinion held about an entity)
- Social and Societal harm (i.e. a capture of harms that may result in a social context or society more broadly) [2, 61, 62].

For each one of these types, we identified several sub-types that characterized that harm in further detail. In Tables 1–5 below, we present and describe the main sub-types as well as including appropriate references to articles that exemplify them. Harm types are designed to be distinctive, however, all types may be attempted to be interpreted in economic terms. Thus, economic harm may overlap with other harm types.

Briefly reflecting on a selection of the harm definitions contained above: examples of Reputational harms that an organization may suffer as a result of a cyber-incident are damaged public image of an organization (e.g. an organization may be regarded as insecure or incapable of protecting customer data) and reduced corporate goodwill (i.e. the business becomes one that others are reluctant to interact or trade with). Harms in the Social and Societal space range from negative changes of public perception (e.g. after an attack, the public may view a certain type of technology as unreliable or insecure), to the disruption of the daily lives of the public. For instance, the cyber-attack on a Ukrainian power company caused a blackout that affected 700 000 homes, numerous communities and society as a whole in the country [50]. This attack had imminent impact on society and the harm caused is analogous to the speed of detection and the effective mitigation controls in place. As nations, and subsequently organizations, vary in their cybersecurity maturity, the extent of these harms will vary as well.

Physical or Digital harm is one of the most familiar types of harm for organizations, and examples of it are: damaged or unavailable systems; corrupted data files; exfiltration or theft of sensitive or customer data; and bodily injury to employees or customers. From these examples, it can be seen that at the current description level of the taxonomy (as shown in Fig. 1) assets are not specifically named. This is intentional and enables users of the taxonomy either to maintain a separate asset listing (or asset taxonomy) and map the two as necessary, or to add a different category in this taxonomy to detail the relevant assets that may be harmed in that particular way. Our decision was informed by the fact that management of such assets is achieved by different methodologies in organizations. Abstracting the taxonomy as arranged above follows a similar approach to one of the most well-known computer incident taxonomies [75].

One of the intended advantages of our taxonomy is its clear mapping of the key types and sub-types of cyber-harm. In the face of an incident, therefore, organizations could quickly obtain some general understanding of the types of resulting harm that they may face. This is also important because it may force consideration of aspects not usually deemed 'corporate' and thus rarely properly assessed. Moreover, this broadens understanding of risk and could be incorporated during initial risk assessment phases as well. A good example of this is the psychological harm to individuals. If a business is victim to a cyber-attack, this not only impacts them but also individuals including customers and employees. In the attack on UK Internet Service Provider TalkTalk in 2015, customers not only experienced financial loss, but felt worried and upset about the attack and TalkTalk's response [70, 71]. This could be of interest to an organization because such harms could be prolonged and further impact company reputation and repeat customer business, or result in customers recommending that their friends and colleagues

Figure 1. Taxonomy of organizational cyber-harms.

completely avoid the company. Social-media platforms such as Twitter can exacerbate this harm due to the great visibility they give to customers and the public [35]. This highlights a subset of the wide span of consequential harms, captured in the taxonomy, that result from cyber-incidents.

The propagation of cyber-harm

As the literature from criminology and cyber-economics suggests [18, 19, 24, 25], harm has interesting characteristics that relate to cascading effects. In this section, we consider four case studies of real-world attacks, which provide initial insights into how our taxonomy can be used to identify propagation sequences of different types of cyber-harm, thus illustrating how cyber-harm can emerge and cascade. The four case studies were chosen based on the detailed accounts of the impact of cyber-attacks in the organizations that was publicly available, and because of the long-lasting effects of these attacks. Using the harms in our taxonomy shown in Fig. 1, we identify the assets that were targeted in the case studies, which types of harm occurred first and how these harms in turn triggered different types of harm. Our aim is to explore common sequences of harms, which may be likely to result given that an initial harm has

Table 1. Defining elements in the taxonomy for the physical or digital harm type

Cyber-harm type	Cyber-harm sub-type
Physical or digital	Damaged or unavailable – The asset has been physically or digitally affected to the point where it is not available to fulfil its intended purpose [57] Destroyed – The asset has been physically or digitally ruined [12] Theft – The asset has been physically or digitally stolen [63] Compromised – The asset has been physically or digitally affected [63] Infected – The asset has been physically or digitally contaminated [50] Exposed or leaked – The asset has been physically or digitally disclosed [64] Corrupted – The asset has been physically or digitally debased or its integrity affected [50] Reduced performance – The asset has had its ability to function lowered [57] Bodily injury – The body of the human asset has been wounded [12] Pain – The human asset has experienced agony [12] Loss of life – The human asset is no longer alive [65] Prosecution – Legal proceedings have been launched against an individual or organization [57, 66] Abuse – The asset has been physically or digitally misused [67] Mistreatment – The asset has been physically or digitally brutalized [67] Identity theft – The theft of personal identity information [67]

Table 2. Defining elements in the taxonomy for the economic harm type

Cyber-harm type	Cyber-harm sub-type
Economic	Disrupted operations – The operational assets (e.g. processes) are not functioning as expected [12] Disrupted sales or turnover – The amount of sales or turnover of the organization has been reduced [52] Reduced customers – The number of customers of the organization has dropped [52] Reduced profits – The profits of the organization have dropped [52, 68] Reduced growth – The growth of the organization has dropped [68] Reduced investments – The investments made by external parties into the organization have dropped [67] Fall in stock price – The stock price of the organization has dropped [67] Theft of finances – Finances of the organization have been stolen [69] Loss of finances or capital – Finances or capital have been diminished [67] Regulatory fines – Fines levied by regulatory bodies that the organization is liable to pay [12] Investigation costs – The fees payable by the organization for investigating an incident [67] PR response costs – The fees payable by the organization for engaging a public relations after an incident [67] Compensation payments – The costs that the organization has had to pay as compensation to those affected by the incident [70] Extortion payments – The costs that the organization has had to pay to continue its operations (e.g. after ransom-related incidents) [65] Loss of jobs – The organization has had to reduce its number of employees [12] Scam victims – The organization or its stakeholders have been conned [65]

Table 3. Defining elements in the taxonomy for the psychological harm type

Cyber-harm type	Cyber-harm sub-type
Psychological	Confusion – Disarray experienced by the organization's stakeholders [70, 71] Discomfort – Uneasiness experienced by the organization's stakeholders [35, 70, 71] Frustration – Dissatisfaction experienced by the organization's stakeholders [57] Worry or anxiety – Nervousness experienced by the organization's stakeholders [57] Feeling upset – Anger experienced by the organization's stakeholders [70, 71] Depressed – Low-spiritedness experienced by the organization's stakeholders [65] Embarrassed – Humiliation experienced by the organization's stakeholders [65] Shameful – Disgracefulness experienced by the organization's stakeholders [65] Guilty – Regret or remorsefulness experienced by the organization's stakeholders [65] Loss of self-confidence – Lack of courage or certainty experienced by the organization's stakeholders [50] Low satisfaction – Lack of contentment experienced by the organization's stakeholders [72] Negative changes in perception – An adverse change in how stakeholders regard a stakeholder [65]

occurred. We perform this analysis here in order to demonstrate that the taxonomy can adequately characterize harms arising in such scenarios. This could, however, also be used in gaining a better understanding of the broader risk facing the organization along the dimensions proposed by Beck and Giddens [15–17].

The Sony cases

In April 2011, amid unstable economic conditions, Sony announced that personal information for 77 million PlayStation Network (PSN) subscribers as well as 24.6 million Sony Online Entertainment accounts had been exposed due to an external breach [64]. The data

Table 4. Defining elements in the taxonomy for the reputational harm type

Cyber-harm type	Cyber-harm sub-type
Reputational	Damaged public perception – An adverse change in how the public regards the organization [12] Reduced corporate goodwill – A negative change in the established reputation of an organization [67] Damaged relationship with customers – An adverse change in relationship between the organization and its customers [67] Damaged relationship with suppliers – An adverse change in relationship between the organization and its suppliers [62] Reduced business opportunities – A negative change in the chances for organizational expansion and growth [67] Inability to recruit desired staff – Difficulty to attract and recruit appropriate employees for roles within the organization [73] Media scrutiny – Media outlets continuously examining the organization [12] Loss of key staff – Key employees within the organization have either been let go, reassigned, or have resigned [74] Loss or suspension of accreditation or certifications – The organization has had its accreditation or certifications removed temporarily or permanently [12] Reduced credit scores – Stakeholders associated with the organization have had or are at risk of having their credit scores negatively impacted [68]

Table 5. Defining elements in the taxonomy for the social and societal harm type

Cyber-harm type	Cyber-harm sub-type
Social and societal	Negative changes in public perception – An adverse change in how society generally regards the organization [52] Disruption in daily life activities – Daily life activities and services in a society not functioning as expected [68] Negative impact on nation – An adverse impact on how a nation (including its services, etc.) functions [50] Drop in internal organization morale – A reduction how employees within the organization perceive that organization [57, 66]

breach involved information about account logins, passwords, credit card details, purchase histories and billing addresses. Sony's facilities in Japan were also heavily impacted from the earthquake of March 2011, resulting in the suspension of several critical operations, which rendered the cyber-attack well timed to inflict maximum damage. Sony had to place its PSN services offline the day following the attack [67] to assess the extent of the incident, resulting in loss of revenue; incurred response costs regarding identifying and addressing the vulnerabilities exploited and notifying the customers; a rough estimate of the costs is $171 million. This figure, however, does not include punitive damages from lawsuits, costs from identity theft or any other misuse of stolen credit cards, nor the loss of business and market capitalization [67].

In late April 2011, Sony provided a comprehensive recovery plan and an accurate calculation of the costs inflicted from the earthquake, but they were still not yet able to calculate the full organizational harm from the cyber-attack [64]. The aggregated impact of the earthquake and the data breach resulted in a significant decrease in Sony's market evaluation as depicted in stock-exchange markets. Sony's share price dropped 19% after the earthquake, a drop equivalent to the general Japanese stock exchange market, but soon recovered 50% of this loss [64]. After the cyber-attack, however, Sony's price sustained a 12% loss (this time it was not a reflection of the rest of the Japanese economy), and the revelation of the security weaknesses once Sony had restored service prolonged the recovery phase [64].

Three years after these incidents, in November 2014, confidential data from Sony Pictures were once again leaked. The data included more than 30 000 internal documents, 170 000 emails, social-security numbers of Sony's employees, personnel reviews and medical histories, and movies which had not yet been released. The same cyber-attack paralysed all of Sony's systems, rendering the online database of stock footage unsearchable, the telephone system offline, computers and servers unusable; this was described by the FBI as an 'unprecedented digital assault that would have felled 90 per cent of companies it hit' [57].

Sony was forced to replace a large number of its systems, set up a hotline for identity fraud, provide psychological counselling for employees and organize seminars on data security. Following the attack, Sony's employees received emails threatening their families if they did not denounce Sony, their credit cards were available for sale on Dark Net markets, and some witnessed their bank accounts exceeding credit limits. A survey conducted by the Identity Theft Resource Center regarding victims of identity theft, reported that victims' experienced 'denial, frustration, rage, fear, betrayal, and powerlessness in the days, weeks, and years after the violation' [57]. Class-action lawsuits from employees were filed, either because Sony did not notify those whose data was leaked, or over fears of how personal leaked information could be potentially used. This also contributed to the fact that some key staff left the company; and furthermore, the press discovered Sony's diversity issues, which were discussed extensively in the content of the leaked emails [57, 66].

The JP Morgan case

JP Morgan Chase, one of the largest banks in the USA, reported that hackers obtained administrator access to several of their servers. Information regarding names, phone numbers, email and physical addresses of account holders was exfiltrated, affecting 76 million households and seven million small businesses. JP Morgan had announced an increase in their cybersecurity budget of $250 million per year just before the attack occurred [76]. The company was forced to replace the majority of its IT infrastructure, a process that was time-consuming and hindered the daily lives of employees. The remaining budget was spent hiring more than 1000 employees to monitor the company's systems [74]. Of significant interest are the two long-term effects, which resulted from this hack. The majority of the customers whose information was leaked were obliged to monitor their finances in fear of fraud, while they received fake emails directing them to impostor websites for financial exchanges. As a result, many became victims of financial fraud. The second

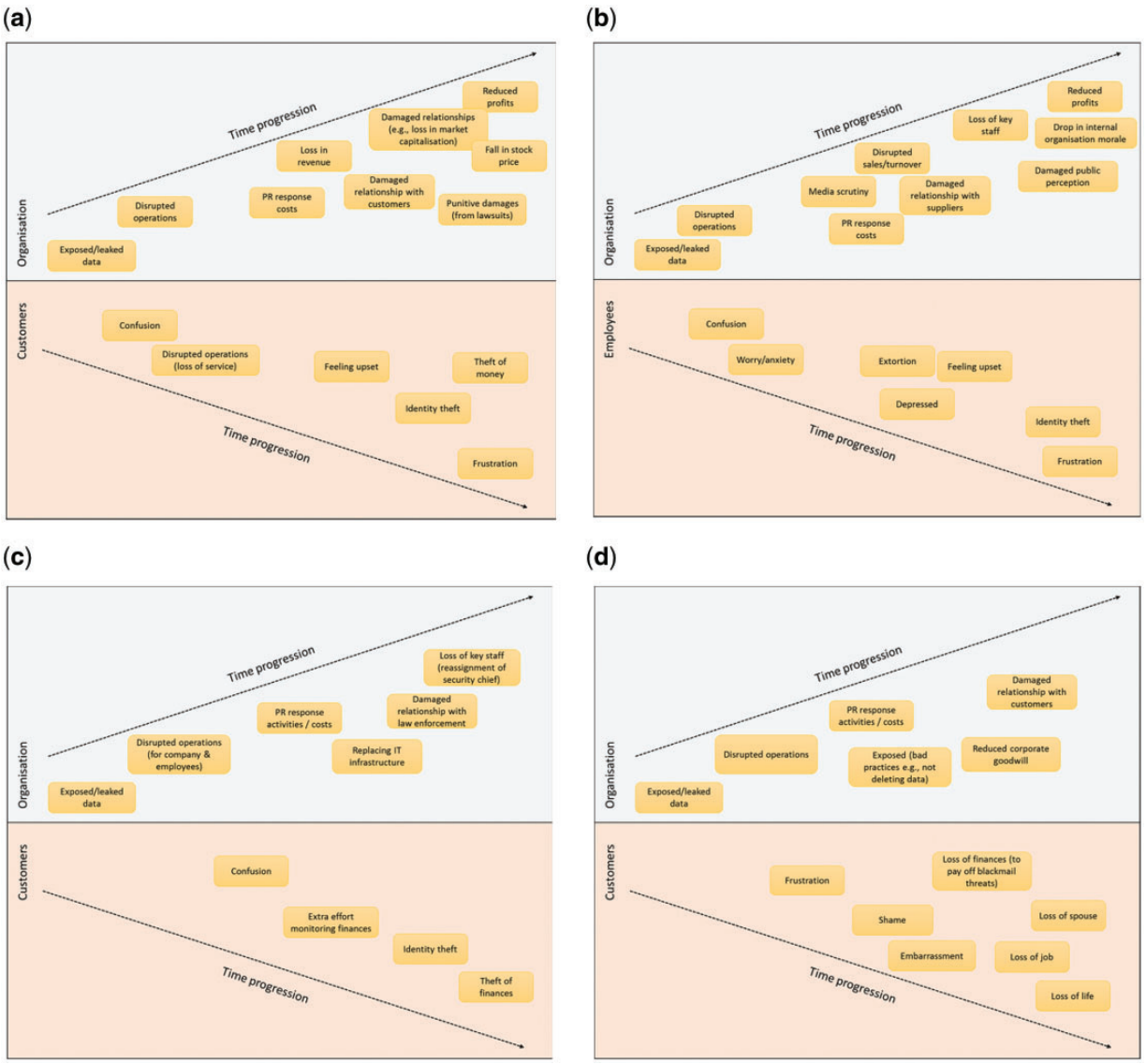

Figure 2. Propagation of harm after the cyber-attacks on Sony in 2011 (a) and 2014 (b), JPMorgan (c) and Ashley Madison (d).

effect was the replacement of their chief information security officer because of his inadequate collaboration with federal authorities in an attempt to try to control the investigation and obscure the leakage of information [74].

The Ashley Madison case

In July 2015, details of 33 million accounts and personal information about people registered on Ashley Madison, a website facilitating extramarital affairs, were leaked [63]. A core principle of Ashley Madison's business model was privacy and security, through which they would build a trust relationship with their customers. The cyber-attack, therefore, had dramatic consequences for the reputation of the company, not only because it exposed the vulnerabilities of the system, but because it proved that Ashley Madison's promise to delete data upon customers' request was not kept [77]. As a result of this practice, Ashley Madison became liable to lawsuits [77], with many organizations soliciting litigants on Twitter [72]. What are of great interest in this case, however, are the repercussions of what was coined as 'collateral damage' which are peculiar to the nature of the services the website offered.

Once the data was publicly available and easily searchable, customers became susceptible to blackmail, with professional and personal ramifications [72]. Many of the leaked email addresses contained the '.mil' domain, indicating people who serve in the US military. Adultery, however, is a crime in the US military and members of Ashley Madison were subject to a year of confinement or dishonourable discharge [77]. In a similar vein, owners of 1, 200 '.sa' email addresses were exposed to a potential death sentence, which is the punishment in Saudi Arabia for adultery. New practices of cybercrime emerged, with criminals threatening to expose people whose email addresses were found in the Ashley Madison dataset to their 'significant other', unless $225 were paid in bitcoin [65]. Public figures were coerced into 'painful personal admissions', others were divorced, while the Toronto police reported two suicides potentially linked to the cyber-attack [65].

Analysis of case studies for propagation of harm

We start our analysis with a digest of the different types of harm arising from the case studies and their impact on the organization and its employees and customers. This is presented below as a visual

in Fig. 2, and then discussed in general in the remainder of the section.

There are several salient points that can be seen in the cases assessed. Focusing on one of the most prevalent classes of cyber-attack in the literature, i.e. data breaches (e.g. details of JP Morgan customers or employees at Sony), the direct type of harm which occurs based on our taxonomy is 'exposure or leakage of digital information'. As it is evident in the case studies presented above, different entities and stakeholders were affected by the various harms that occurred (e.g. the organization under attack, its employees, customers and suppliers).

We commence our analysis for the subsequent types of harm from an organization's perspective, since they are the main targets of data breaches. The most prominent type of harm is 'reputational damage', which may lead to 'damaged relationships with employees and customers'. By the time the attack is publicly announced, 'economic harms' may be triggered due to potential regulatory fines from 'law enforcement' (as happened in the case of Sony), which may be amplified from relevant harms including 'PR response costs' (to give notice of the incident and to manage the company's response including both online and offline media), 'reduced numbers of customers, falls in stock prices and reduced growth'. A key point to note here is that this harm propagation also alludes to the temporal dimension present with risk more generally, as discussed in Section 'What is cyber-harm for organizations?'.

Departing from organizations and focusing our perspective on employees and customers, 'psychological' harm is the most common type of harm following 'leakage of digital information'. People feel 'confusion, discomfort, frustration' and 'worry' and the magnitude of these types of harm depends on the environment within which the attack was realized. For example, when a financial institution such as JP Morgan has been breached, psychological harms were more severe than in the Sony case. In the cases where individuals were blackmailed additional types of harm such as 'extortion payments' occurred. In a similar vein, 'identity theft' may be experienced and this can result in compensation payments by the banking sector. In extreme circumstances, the ultimate example being that of Ashley Madison, where psychological harms resulted in 'loss of life' because individuals felt 'shamed' and 'embarrassed'. Regarding 'social harm', it may occur in situations where not all the aforementioned types of harm are addressed appropriately and in a timely manner. An example where such a harm was manifested is the Sony case, where there was 'disruption of daily lives' and a 'drop in internal organization morale'.

It should be evident that in the cases presented above that the sequence of types of harm, which occurred when information was leaked, is similar, the main difference being the impact and the length of the chain describing the propagation of different types of harm. These attributes depend on how well, and timely, stakeholders who were responsible for addressing harmful situations respond to the events that unfold. Thus, as alluded to in Fig. 2, there is a temporal element that is critical to the propagation of harm which is also related to the quality of controls thatorganizations have in place to mitigate harms.

In a similar vein, we can observe how types of harm unfold when the assets under attack are 'destroyed'. Starting from an organization's perspective, emerging harms are 'disrupted operations, deteriorating sales' and 'loss of key staff' (in cases where they are forced to resign). It is important to note that the types of direct harm that manifest in most cases depend on the assets exploited by the attacks. The presence of subsequent waves of harm is influenced by the remediation measures, which organizations have in place. As a pattern, 'physical' harms lead to 'economic' harms, which if not addressed may lead to 'reputational' harms for organizations. When 'psychological' harms for employees occur after 'physical harms', then 'economic' and 'physical' harms may follow for employees and customers. The presence of such types of harm may amplify the 'economic', 'reputational' and more scarcely 'social' harms that organizations already experience.

To reflect more generally on the cases in terms of commonalities in harm propagation, exposed or leaked data, especially when it contains personal information, usually has a significant impact on the organization and its customers. Customers often feel confused and frustrated, and this may escalate significantly depending on the data that has been leaked (sometimes it may be identity theft, and other times loss of life). As certain personal information is held for life such as names and social security numbers, the harm associated with a cyber-attack can last for years—this is particularly why more companies are offering credit-monitoring services after data leaks. Similar broad propagation effects can be also seen in more recent hacks including that of Equifax in 2017 [78] and the Singapore government's national health database in 2018 [79].

In terms of the organizational harms, the leakage of data typically has some negative impact on operations, and tends to involve PR response costs and loss in revenue in some. A subsequent harm that is often incurred by the organization is a damaged relationship with customers, suppliers and the public. Possibly the largest difference in the cases is the exact harm that can result. With Ashley Madison, this related in loss of life due to suicide, which may be understandable given the personal nature of the data leaked. For JPMorgan, however, we saw the reassignment of the executive in charge of protecting their network. To consider our earlier reflections on risk in Section 'What is cyber-harm for organizations?', it is unclear whether this represented some instance of 'organised irresponsibility' or poor management of incident response.

To consider the situation today, there are several recent examples of similar propagations of organizational harm—for instance, the case of Facebook and Cambridge Analytica in 2018 [80]. Here, the 'end' harm of this incident was the closure of latter, and regulatory fines and severe public criticism for the former. It is the unknown nature of such attack consequences that may lead to some organizations attempting to avoid harm propagation using other —potentially questionable—means. The Uber breach is an intriguing example of this, where the company opted to secretly pay hackers rather than publicly revealing the leak of details of 57 million customers and drivers [81].

It is important to understand the propagation trends of harm, as we may be able to ascertain what the likely harm is in future attacks and put in place measures to mitigate it. This is a very different lens to that of a kill-chain [82], which seeks to explain the phases of an attack. If we were to orientate a defensive strategy solely around a kill-chain then we might find ourselves investing in the defensive measures and incident responses, which are not actually tightly coupled with limiting the harm to the organization.

Conclusions, reflection and future work

Technological advancements have forced organizations to digitalize parts of their functionality and operations. While investments in IT may result in profit and prosperity, there is always the risk lurking of cyber-attacks and incidents. The threat landscape of cyber-attacks is rapidly changing and the impact of such attacks is uncertain. There is, however, as we showed on Section 'What is

cyber-harm for organizations?', a lack of effective metrics, tools and frameworks seeking to understand and assess the harm organizations face from cyber-attacks.

According to the CUNA president and CEO Jim Nussie, organizations are not incentivized to invest in and prioritize security [83]. It is of paramount importance for board members to obtain a comprehensive cost-benefit analysis on how cutting-edge technologies and investments in implementing strong cybersecurity practices may hedge the risk of a cyber-attack and its harmful impact. The case studies presented in Section 'The propagation of cyber-harm' illustrated that organizations lack sufficient models to estimate the harm, direct and indirect, from cyber-attacks. What it is further evident from our analysis of the case studies is that organizations remain oblivious to the harms that consumers or their employees experience. Therefore, it is impossible without a holistic understanding of all possible harms for organizations to prioritize controls to mitigate these harms. Current practices which organizations adopt either myopically calculate the harm from a cyber-attack or estimate financial damages from the stock-market exchanges. In this way, they neglect the indirect harms resulting from cyber-attacks and the harms that consumers experience; these harms are not always visible and may have more longitudinal effects.

Based on a thorough literature review and on analysing a series of cyber-incidents, we have presented a taxonomy of cyber-harms aimed at providing further insight into the direct and indirect harms which organizations and individuals may experience. Our expectation is that our taxonomy should provide the essential broad knowledge of harms for organizations, enable them to consider indirect harms to consumers and other corporate and non-corporate actors, as well as shift the current tendency of organizations to remain inactive or tolerate harms which impact non-corporate actors. We hope to avoid situations and perspectives such as the following, where the former executive director of Sony Pictures was reported stating '[I]t's a valid business decision to accept the risk of a security breach. I will not invest $10 million to avoid a possible $1 million loss' in 2005 [83]. The reality is that cyber-attacks can have much more significant and long-lasting harms beyond what is initially perceived. Our taxonomy would help to elucidate these, and thereby support better decision-making in risk management and the selection of security controls.

While we believe that our taxonomy elucidates many of the key aspects of cyber-harm for organizations, we emphasize that this version is especially intended to motivate further discourse on this topic in the field. As such, there are several outstanding questions still to be addressed. For example, is this taxonomy of harm able to capture and usefully structure all the types of harm that may occur to organizations as a result of a cyber-incident? Although we sought to be comprehensive in our research, by considering real cases and relevant literature, we appreciate that discussions with business and security professionals in organizations, particularly those that have suffered a cyber-incident, may lead to an expanded set of harm categories or a refined harm structure. A key activity, therefore, is the expansion of the taxonomy in Fig. 1, and characterization of more rigorous and useful harm quantification metrics and magnitudes.

Although there has been significant research in the space of understanding the impact of cyber-incidents, as discussed in previous sections, the lack of a model which can support analytics regarding the detection, measurement, prediction and prioritization of cyber-harms is evident. The taxonomy developed and presented in this article is essential to the creation of such a model, which can then underpin analytics—such analytics include a more functional understanding of how we might go about modelling the interconnections that exist between harms, and so the possible cascading effects.

Therefore, our next steps are to extend this research by designing an asset-oriented model. Our decision is based on the fact that such an approach encourages organizations to focus on their core assets, and think beyond current threats to consider the full range of harms that might potentially result to assets. Reflecting on our taxonomy and the case studies presented in the article, we believe that such a model should comprise six different stages in defining and assessing the notion of cyber-harm. These are: identifying core assets; identifying direct harm to assets; determining the stakeholders that hold an interest in direct harm; identifying different types of cyber-harm occurring from the direct harm; measuring the overall indirect harm (i.e. propagating harm) for all the stakeholders; and understanding this variety of cyber-harm and security controls in place that might be able to treat it.

Every stakeholder may perceive or experience harm differently, and the consequences of cyber-attacks should be assessed based on their views, resulting in the existence of different 'lenses' to examine cyber-harm. We believe that such a model is crucially required if organizations are to optimally structure their cybersecurity controls for minimizing harms. This is especially relevant as technologies such as the IoT and Artificial Intelligence (AI) mature and become widely deployed, and organizations look to manage risk—be it through internal methods or investment in cyber-insurance [84]. Our review of the literature suggests that the majority of successful cyber-attacks exploit well-known vulnerabilities and the inertia of organizations to provide appropriate cybersecurity policies due to the misconception of the risks that may emerge. It is, therefore, crucial for board members to obtain an accurate estimate of direct and indirect harm from cyber-attacks before reconsidering the threat landscape their organizations face. We believe a taxonomy of harms is a decisive first step towards this direction.

References

1. ISACA. *State of Cybersecurity: Implications for 2015*. http://www.isaca.org/cyber/Documents/State-of-Cybersecurity_Res_Eng_0415.pdf (13 July 2018, date last accessed).
2. National Institute of Standards Technology. *Special Publication 800-30 Revision 1: Guide for Conducting Risk Assessments*. http://dx.doi.org/10.6028/NIST.SP.800-30r1 (13 July 2018, date last accessed).
3. Oxford English Dictionary. Definition of harm. 2013.
4. Schulhofer SJ. Harm and punishment: a critique of emphasis on the results of conduct in the criminal law. *UPa L Rev* 1974;**122**:1497–1607.
5. Sharpe V, Faden A. *Medical Harm: Historical, Conceptual, and Ethical Dimensions of Iatrogenic Illness. 1998*. New York: Cambridge University Press, 1974.
6. Kleinig J. Crime and the concept of harm. *Am Philos Q* 1978;**15**:27–36.
7. Gratz KL. Risk factors for and functions of deliberate self-harm: an empirical and conceptual review. *Clin Psychol* 2006; **10**:192–205.
8. Zou HL, Zeng RC, Zeng SX. How do environmental violation events harm corporate reputation? *BSE* 2015; **24**:836–54.
9. Andoh-Baidoo FK, Amoako-Gyampah K, Osei-Bryson K-M. How Internet security breaches harm market value. *IEEE Secur Priv* 2010; **8**: 36–42.
10. U. K. Government. *Chancellor's Speech to GCHQ on Cyber Security*. https://www.gov.uk/government/speeches/chancellors-speech-to-gchq-on-cyber-security (13 July 2018, date last accessed).
11. New Zealand Government. Risk Assessment Process: Information Security. https://www.ict.govt.nz/assets/ICT-System-Assurance/Risk-Assessment-Process-Information-Security.pdf (13 July 2018, date last accessed).

12. UVM. *Enterprise Risk Management Program: Guide to Risk Assessment & Response*. https://www.uvm.edu/sites/default/files/UVM-Risk-Management-and-Safety/Guide_to_Risk_Opportunity_Assessment_Response.pdf (13 July 2018, date last accessed).
13. ENISA. *Security Risk Management Glossary*. https://www.enisa.europa.eu/activities/risk-management/current-risk/risk-management-inventory/glossary (13 July 2018, date last accessed).
14. ISO/IEC. ISO/IEC 13335-1: 2004 Part 1: Concepts and models for information and communications technology security management, 2004.
15. Beck U. *Risk Society: Towards a New Modernity*. **Vol. 17**. London: Sage, 1992 .
16. Giddens A. *The Consequences of Modernity*. Cambridge: Polity Press, 1990.
17. Beck U. The terrorist threat: world risk society revisited. *Theory Cult Soc* 2002;**19**:39–55.
18. Van Slyke SR, Van Slyke S, Benson ML. *The Oxford Handbook of White-Collar Crime*. Oxford University Press, 2016.
19. Greenfield VA, Paoli L. A framework to assess the harms of crimes. *Br J Criminol* 2013;**53**:864–885.
20. Levi M. Social reactions to white-collar crimes and their relationship to economic crises. In: Deflem M (ed.), *Economic Crisis and Crime*. Sociology of Crime, Law and Deviance, Volume 16. Emerald Group Publishing Limited, 2011, 87–105.
21. Spalek B. White-collar crime and secondary victimization: an analysis of the effects of the closure of BCCI. *Howard J Crim Just* 2001;**40**:166–79.
22. Brenner SW. Cybercrime metrics: old wine, new bottles? *Va. JL & Tech* 2004;**9**:13–13.
23. Felici M, Wainwright N, Cavallini S, *et al*. What's new in the economics of cybersecurity? *IEEE Secur Priv* 2016;**14**:11–13.
24. Anderson R, Barton C, Böhme R, *et al*. Measuring the cost of cybercrime. *The Economics of Information Security and Privacy*, 2013; 265–300.
25. Anderson R, Moore T. The economics of information security. *Science* 2006;**314**:610–613.
26. Moore T. The economics of cybersecurity: principles and policy options. *IJCIP*2010;**3**:103–17.
27. Punter A, Coburn A, Ralph D. Evolving risk frameworks: modelling resilient business systems as interconnected networks. *Centre for Risk Studies, University of Cambridge* 2016. http://cambridgeriskframework.com/page/17 (13 July 2018, date last accessed).
28. Lloyds of London. *Counting the cost*. https://www.lloyds.com/news-and-insight/risk-insight/library/technology/countingthecost (13 July 2018, date last accessed).
29. Klahr R, Shah J, Sheriffs P, *et al*. Cyber security breaches survey 2017: main report, 2017. https://www.gov.uk/government/statistics/cyber-security-breaches-survey-2017 (13 July 2018; date last accessed).
30. Anderson R, Böhme R, Clayton R, *et al*. Security economics and European policy. In: Pohlmann N., Reimer H., Schneider W. (eds). *ISSE 2008 Securing Electronic Business Processes* 2009; 55–80.
31. Laube S, Böhme R. The economics of mandatory security breach reporting to authorities. *J Cybersecur* 2016;**2**:29–41.
32. Kshetri N. The simple economics of cybercrimes. *IEEE Secur Priv* 2006;**4**:33–39.
33. Edwards B, Hofmeyr S, Forrest S. Hype and heavy tails: a closer look at data breaches. *J Cybersecur* 2016;**2**:3–14.
34. Nguyen KD, Rosoff H, Richard SJ. Valuing information security from a phishing attack. *In: International Conference on Applied Human Factors and Ergonomics*. Cham: Springer, 2017.
35. Why it pays to complain via Twitter. *BBC News*, 2014. http://www.bbc.co.uk/news/business-27381699 (13 July 2018, date last accessed).
36. Telang R, Wattal S. An empirical analysis of the impact of software vulnerability announcements on firm stock price. *IEEE Trans Softw Eng* 2007;**33**:544–57.
37. Acquisti A, Telang R, Friedman A. Is there a cost to privacy breaches? An event study. *Proceedings of the 3rd International Conference on Intelligent Systems (ICIS)*, 2006.
38. Cavusoglu H, Mishra B, Raghunathan S. The effect of internet security breach announcements on market value: capital market reactions for breached firms and internet security developers. *IJEC* 2004; **9**:70–104.
39. Kvochko E, Pant R. Why data breaches don't hurt stock prices. *Harvard Business Review*, 2015. https://hbr.org/2015/03/why-data-breaches-dont-hurt-stock-prices (13 July 2018, date last accessed).
40. Hovav A, D'Arcy J. The impact of denial-of-service attack announcements on the market value of firms. *RMIR* 2003;**6**:97–121.
41. Campbell K, Gordon LA, Loeb MP, *et al*. The economic cost of publicly announced information security breaches: empirical evidence from the stock market. *J Comput Secur* 2003;**11**:431–448.
42. Kannan K, Rees J, Sridhar S. Market reactions to information security breach announcements: an empirical analysis. *IJEC* 2007;**12**:69–91.
43. World Economic Forum. *Partnering for cyber resilience towards the quantification of cyber threats*. http://www3.weforum.org/docs/WEFUSA_QuantificationofCyberThreats_Report2015.pdf (13/07/2018, last accessed)
44. Çetin B, Yaman E, Peker A. Cyber victim and bullying scale: a study of validity and reliability. *Comput Educ* 2011;**57**:2261–2271.
45. Harvard Mental Health Letter. Protecting children and teens from cyber-harm. *Harvard Health Pubs* 2008;**25**:4–5.
46. Gartzke E. The myth of cyberwar: bringing war in cyberspace back down to earth. *Int Secur* 2013;**38**:41–73.
47. Charles P, Pfleeger SL. *Analyzing Computer Security: A Threat/Vulnerability/Countermeasure Approach*. Upper Saddle River, NJ: Prentice Hall, 2012 .
48. Kesan JP, Hayes CM. Thinking through active defense in cyberspace. In: *Proceedings of the Workshop on Deterring Cyberattacks: Informing Strategies and Developing Options*, Washington, DC: The National Academies Press; 2010, 327–42.
49. US Department of Homeland Security. *Cyber risk management and cybersecurity insurance*. http://www.dhs.gov/cybersecurity-insurance (13 July 2018, date last accessed).
50. Titcomb J. Ukrainian blackout blamed on cyber-attack. *The Telegraph*, 2015 http://www.telegraph.co.uk/technology/news/12082758/Ukrainian-blackout-blamed-on-cyber-attack-in-world-first.html (13 July 2018, date last accessed).
51. Greenberg A. Hackers remotely killed a jeep on the highway- with me in it. *Wired*, 2015; http://www.wired.com/2015/07/hackers-remotely-kill-jeep-highway/ (13 July 2018, date last accessed).
52. Lee T. Forget the Ashley Madison or Sony Hacks – a crippling cyberattack is imminent in the US. *The Guardian*, 2015. http://www.theguardian.com/technology/2015/jul/26/cybercrime-hacking-internet-of-things-target (13 July 2018, date last accessed).
53. Agrafiotis I, Bada M, Cornish P, *et al*. Cyber harm: concepts, taxonomy and measurement. *Saïd Business School Working Paper* 2016; **23**. doi: http://dx.doi.org/10.2139/ssrn.2828646.
54. Paolo Passeri, Hackmageddon. *Cyber attacks timeline*, 2016; http://www.hackmageddon.com/category/security/cyber-attacks-timeline (13 July 2018, date last accessed).
55. Veris Community D. http://veriscommunity.net/vcdb.html (13 July 2018, date last accessed)
56. UK Government and Marsh, Ltd. *UK cyber security: the role of insurance in managing and mitigating the risk*. https://www.gov.uk/government/publications/uk-cyber-security-the-role-of-insurance (13 July 2018, date last accessed).
57. Hess A. Inside the Sony hack.*Slate*, 2015 [CrossRef][10.3998/mij.15031809.0002.203] http://www.slate.com/articles/technology/users/2015/11/sony_employees_on_the_hack_one_year_later.html (13 July 2018, date last accessed).
58. Elo S, Kyngäs H. The qualitative content analysis process. *J Adv Nurs* 2008;**62**:107–15.
59. Hsieh HF, Shannon SE. Three approaches to qualitative content analysis. *Qual Health Res* 2005;**15**:1277–88.
60. Pemberton S. Social harm future (s): exploring the potential of the social harm approach. *Crime Law Soc Change* 2007; **48**:27–41.
61. The Parliament of the Commonwealth of Australia. *Privacy Amendment (Notification of Serious Data Breaches) Bill* 2015. https://www.ag.gov.au/Consultations/Pages/serious-data-breach-notification.aspx (13 July 2018, date last accessed).
62. Gizmodo. *Last month's massive target hack was the heating guy's fault*, 2014. http://gizmodo.com/last-months-massive-target-hack-was-the-heating-guys-1516926877 (13 July 2018, date last accessed).

63. InfoSec Institute. Ashley Madison revisited: legal, business and security repercussions,. 2015, 8. http://resources.infosecinstitute.com/ashley-madison-revisited-legal-business-and-security-repercussions (13 July 2018, date last accessed).
64. Dark Reading. *Sony data breach cleanup to cost $171million*, 2011 http://www.darkreading.com/attacks-and-breaches/sony-data-breach-cleanup-to-cost-\$171-million/d/d-id/1097898 (13 July 2018, date last accessed).
65. Ashley M. *Aftermath: confessions, suicide reports and hot on the hacker's trail. National Post*, 2015 http://news.nationalpost.com/news/canada/ashley-madison-aftermath-confessions-suicide-reports-and-hot-on-the-hackers-trail (13 July 2018, date last accessed).
66. Variety. *Sony hack attack opens minefield of legal questions that has hollywood worried*, 2015,07–13. http://variety.com/2015/biz/news/sony-hack-attack-opens-minefield-of-legal-questions-that-has-hollywood-worried-1201471664 (13 July 2018, date last accessed).
67. PwC. *Limiting the impact of data breachesthe case of the Sony Play Station Network*, 2011 http://www.strategyand.pwc.com/reports/limiting-impact-data-breaches-case (13 July 2018, date last accessed).
68. The Huffington Post. *A look back at the target breach*, 2015. http://www.huffingtonpost.com/eric-dezenhall/a-look-back-at-the-target_b_7000816.html (13 July 2018, date last accessed).
69. White L. and Bergin T. Tesco says $3million stolen in cyber theft, resumes service. *Reuters*, 2016. http://www.reuters.com/article/us-tesco-bank-idUSKBN1331TX (13 July 2018, date last accessed).
70. Talktalk hackers go on £600 spending spree with stolen card details as boss says its too early to consider compensation. *The Mirror*, 2015. http://www.mirror.co.uk/news/uk-news/talktalk-hackers-go-600-spending-6694321 (13 July 2018, date last accessed).
71. McDaid L. Talktalk cyber-attack: county Londonderry man targeted. *BBC News*, 2015. http://www.bbc.co.uk/news/uk-34613921 (13 July 2018, date last accessed).
72. Top data security expert fears traumatic aftermath in Ashley Madison hack. *The Guardian*, 2015. https://www.theguardian.com/technology/2015/aug/19/ashley-madison-hack-outcome (13 July 2018, date last accessed)
73. Sony seeking more cybersecurity staff amid hack. *The Wall Street Journal*, 2014. http://blogs.wsj.com/cio/2014/12/22/sony-seeking-more-cybersecurity-staff-amid-hack-fallout/ (13 July 2018, date last accessed).
74. JP Morgan security exec reassigned after breach. *Europe TechWeek*, 2015. http://www.techweekeurope.co.uk/e-management/jobs/jp-morgan-exec-reassigned-171644 (13 July 2018, date last accessed).
75. Howard JD, Longstaff TA. *A common language for computer security incidents*. Sandia National Laboratories, 1998. https://prod.sandia.gov/techlib-noauth/access-control.cgi/1998/988667.pdf (13 July 2018, date last accessed).
76. JP Morgan Chase reveals massive data breach affecting 76m households. *The Guardian*, 2014. http://www.theguardian.com/business/2014/oct/02/jp-morgan-76m-households-affected-data-breach (13 July 2018, date last accessed).
77. The Verge. *The mind-bending messiness of the Ashley Madison data dump*,2015. http://www.theverge.com/2015/8/19/9178855/ashley-madison-data-breach-implications (13 July 2018, date last accessed).
78. Krebs on Security. Breach at Equifax May Impact 143M Americans, 2017. https://krebsonsecurity.com/2017/09/breach-at-equifax-may-impact-143m-americans/ (13 July 2018, date last accessed).
79. Singapore personal data hack hits 1. 5m, health authority says. *BBC News*, 2018. https://www.bbc.com/news/world-asia-44900507 (13 July 2018, date last accessed).
80. Revealed: 50 million Facebook profiles harvested from Cambridge Analytica in major date breach. *The Guardian*, 2018. https://www.theguardian.com/news/2018/mar/17/cambridge-analytica-facebook-influence-us-election (13 July 2018, date last accessed).
81. Uber concealed huge data breach. *BBC News*, 2017. https://www.bbc.co.uk/news/technology-42075306 (13 July 2018, date last accessed).
82. Hutchins EM, Cloppert MJ, Amin RM. Intelligence-driven computer network defense informed by analysis of adversary campaigns and intrusion kill chains. In: Ryan J. (ed.). *Leading Issues in Information Warfare & Security Research* Vol. **1**, Reading, UK: Academic Publishing International Limited; 2011.
83. Infosec Institute. *How harmful can a data breach be?*, 2016. http://resources.infosecinstitute.com/the-cost-of-a-data-breach-how-harmful-can-a-data-breach-be (13 July 2018, date last accessed).
84. Woods D, Agrafiotis I, Nurse JR, *et al*. Mapping the coverage of security controls in cyber insurance proposal forms. *JISA* 2017;**8**:8.