

Analysing cyber-insurance claims to design harm-propagation trees

Louise Axon, Arnau Erola, Ioannis Agraftotis, Michael Goldsmith, Sadie Creese
Department of Computer Science, University of Oxford, UK

`{firstname.lastname}@cs.ox.ac.uk`

Abstract—With a continuously changing threat landscape, companies must be prepared for the most unforeseen cyber events. Harm originating from cyberspace varies in magnitude and type, with potential for systemic consequences. While the adoption of security controls may partially mitigate the impact of cyber-attacks, a nuanced understanding of how events unfold during and after an incident will help organisations to better estimate the risk they face and implement advanced incident response strategies. A better estimation of risk is of particular importance to the insurance community because the costs from claims due to cyber-events vary significantly. Towards this end, we collected and analysed more than 70 claims against an insurance company, extracting different types of harm and their characteristics. We then reconstructed the claims based on these types of harm in order to obtain patterns of how cyber-harm propagates. The result is a graph indicating the most common paths that harm follows on multiple events. The findings can help policy-makers and insurance companies to understand how harm propagates, estimate more accurately the value-at-risk and adopt the necessary controls to mitigate these harms.

Index Terms—cyber-insurance, harm trees, cyber threats

I. INTRODUCTION

Companies face an increasingly varied range of threats to their cybersecurity, and measuring their exposure to cyber-risk – the likelihood that a company will be negatively affected by a cyber-incident – is an important yet complex problem. Developing reliable and accurate approaches to reasoning about cyber-risk exposure is critical. A better understanding of this problem would enable companies to improve their cybersecurity posture and incident-response strategies, in order that they may have the best chance of protecting themselves against and responding to damaging cyber incidents. Exposure to cyber-risk must also inform the provision of cyber-insurance to companies. Understanding a company’s posture – their assets, risk-controls and exposure to threats – is important to understanding the likelihood that a cyber-incident may occur. In understanding cyber-risk, it is also important to understand how events may unfold after a particular type of cyber-incident has occurred: the harms that companies may experience directly or indirectly as a result of the cyber-incident, and the way in which these harms may propagate.

By cyber-harms we refer to the damaging consequences that a company can experience as a result of a cyber-incident, which may take the form of financial losses, loss of business reputation and customers, or psychological damage to employees, for example [2]. Based on past cases of high-

profile cybersecurity attacks and breaches, there is some understanding of the types of harm that may directly result from a cybersecurity incident. It is also clear that there is propagation of cyber-harm in cyberspace. An example is the data-breach at Ashley Madison, a commercial website facilitating extramarital affairs, in which the initial harm – the breach of confidential customer information – resulted in further harms including damage to the company’s reputation and loss of customers; extortion and public humiliation of customers; and damages resulting from the ensuing lawsuits filed by customers [1].

In order that we can more accurately understand the ensuing harms that companies may experience as a result of cyber-incidents, we need to understand how and when harm propagates in cyberspace. There is also a need to understand the contexts in which certain harms may lead to further harms (whether harm manifests and propagates differently within different sectors, or following different types of attack, for example). To gain initial insight on this topic, we examined the occurrence of harm in 70 cyber-insurance real claims. We used descriptions of the security event that occurred, and of the processes that ensued, to identify the types of harm experienced in the case of each claimant and the way in which harms propagated.

By understanding the harms that are caused by cybersecurity events, in which cases, and how these harms propagate to result in further harms, we aim to produce initial insight into the space of harms that may result from specific cyber incidents, and the likelihood and severity (in financial terms) of these harms. The aim is to move towards devising a method for predicting the likely harm trees that would occur from an incident in a particular context.

The next section presents a literature review on harm in the cyberspace. Section III introduces the methodology we followed to collect the data and Section IV discusses our findings. Future work and discussion are presented in Section V.

II. BACKGROUND

The concept of harm has been studied in a number of disciplines, inter alia, philosophy, sociology, medicine, psychology, law and economics. The majority of research works perceive harm as injury and damage for a specific ‘subject’. Subjects in a broader range encompass individuals, societal

groups and organisations [3]. Kleinig offers a nuanced approach by reflecting on how harm is defined in multiple disciplines and reconciling the differences [4]. Kleinig’s definition of harm is more nuanced and posits that harm should be conceptualised as the impairment of the welfare of a subject, where welfare is deemed to be interests which are essential to the well-being of individuals and society in general [4].

Adopting a more general definition of harm enables researchers to explore the complex characteristics of harm and the different forms in which it can manifest. A stream of literature where harm has been explored in-depth is criminology and more specific white-collar crime [5]–[7]. Harm from white collar crimes has been realised in many different forms and researchers have focused on examining social harms in order to systematically assess the impact from these crimes. Studies in criminology have considered harms that are experienced not only by individuals but by a number of stakeholders, such as communities, vulnerable societal groups and governments [5]. Criminology is also one of the first disciplines where the cost of harm through crime has been monetised. These costs mainly reflect actions which subjects undertook to anticipate and prevent a crime, and actions that aimed to mitigate the harm.

When considering harm that originates from cyberspace, literature is still in its infancy. In the field of economics, researchers tried to explore cyber-incidents and assess how they can hinder ICT growth or influence market prices [8]. Felici et al., suggest that any effort to estimate cyber harm must be multidisciplinary, and should consider the dynamic nature of harm and the wide range of stakeholders it may influence [8]. Anderson et al. [9], in their seminal work, attempt to measure the costs of cyber-incidents. In their model, they perceive harm as costs that occur from direct and indirect consequences, as well as costs regarding mitigation and defence mechanisms for a number of stakeholders, including society. Agraftiotis et al. [2], examine a database of cyber-incidents and provide a taxonomy of cyber-harms. They further elaborate on the characteristics of harm and explain how these may influence its magnitude, with the most important being the cascading effect of harm.

In another stream of literature, researchers have evolved risk frameworks and resilience models to consider the impact of cyber-attacks [10]. Such models examine how events may result in systemic risks and disrupt services and critical infrastructure globally. Fictional but plausible scenarios are designed to highlight the interconnectedness of assets and understand how harm may propagate to cause systemic risks [11].

Finally, in the insurance sector there are models such as AIR and FICO which collect publicly available data from cyber-incidents and try to correlate this cost with information regarding the organisations, such as revenue, number of trained employees in cybersecurity and adherence to security standards [12], [13]. In doing so, these models are able to predict the value-at-risk (VaR) for organisations. Such models however, rely on the ability of third parties to accurately estimate the harms that organisations experience.

Based on the literature, there is a gap in frameworks that can guide the holistic assessment of cyber-harm. There are a number of quantitative or qualitative models that consider direct losses or focus on a scenario to calculate possible contextualised indirect losses. In this paper we analyse how real events, as captured in claim forms filed for cyber incidents, unfolded. Based on this analysis, we identify patterns and create a number of harm trees. This is our first attempt towards creating a library of harm trees to capture more systematically how harm propagates. Harm-trees could then be used to estimate more accurately the risk which organisations face from cyber-attacks.

III. METHODOLOGY

The data used in this study was collected from real cyber-insurance claims from an insurance company. In order to comply with General Data Protection Regulation (GDPR) requirements and with a Non-Disclosure Agreement (NDA), we transcribed only the description of the event excluding any personal information and anonymising events where necessary. No digital records were created. We obtained access to different sources depending on availability for each case: claim forms, forensic reports, and correspondence between insurers, underwriters and legal counsel. We targeted claims from a variety of sectors including healthcare, technology and finance.

We examined more than 70 claim forms related to incidents with malicious intent, as well as incidents with benign motive (caused by IT errors, or human mistakes for example). A small sample of the examined claims were rejected due to policy violations or because the total financial loss did not surpass the excess amount. For the latter, although a description of the incident was provided, the information regarding detailed financial costs was limited. This is due to the fact that insured companies are obliged to submit a claim form once an incident has been detected, including estimates of costs for reserve protection, but they do not usually provide an updated report if the final costs are lower than the excess or if the incident is not covered by the policy.

Since claims report monetary-loss information, the types of harm we were able to identify were mainly financial. We analysed each case to identify the unique harms that were realised. Individual harms that were similar based on the description provided in the reports were merged. We then reconstructed all the cases using the unique harms. By mapping out the occurrence of harms, the order in which harms occurred, and the harms that occurred as a result of other harms, we established patterns in how the incidents unfolded and created harm-trees. Moreover, whenever information was available, we collected: monetary impact and length of time the cyber event lasted, the number of employees, annual revenue, high level description of technical controls, and North American Industry Classification System (NAICS) code (a code describing the type of industry) to allow us to examine correlations between premiums, harms and characteristics of organisations.

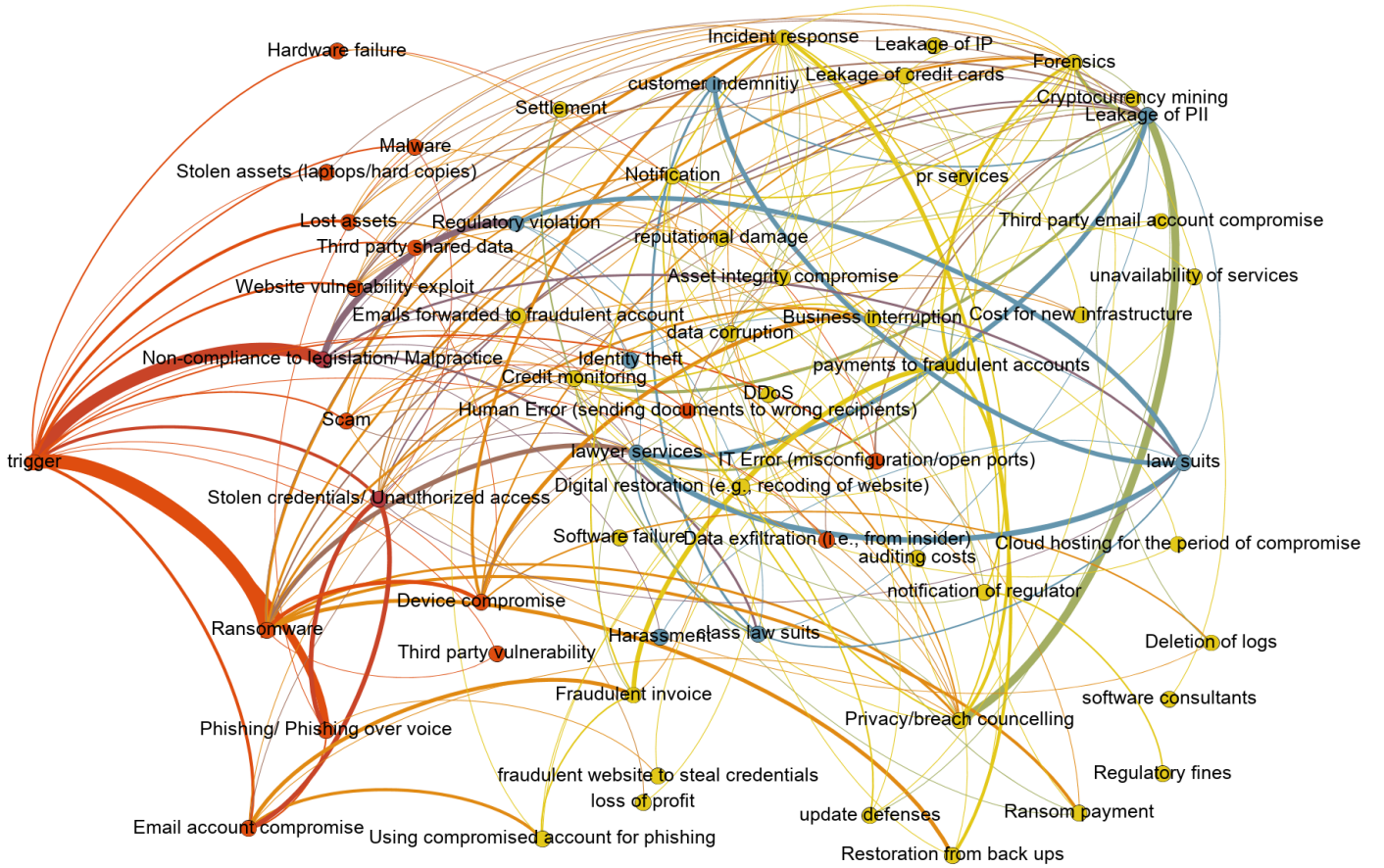


Fig. 1. Graph representation of the most common harm trees observed in the collected data.

IV. RESULTS

We analysed the data collected, identifying the types of harm recorded by all researchers, to produce a classification of all types of harm present in the claims cases. All harms collected were iteratively revisited and those representing similar harms clustered. In total we identified 63 distinct harms present in the claims. By identifying the harms that occurred and which harms had occurred as a result of previous harms, we specified an individual chain of events for each of the 70 claim cases. We then synthesised this information to obtain patterns; we classified claim forms that had common initial step and considered this the root of a tree; then for each class we added as the next level of the tree the events that were second in the chain of the claim forms; we continued this process iteratively for all the classes and we obtained the propagation of harms. Thus, the root of each tree is the attack that triggers the harms, and all other nodes are harms that occur in succession or in parallel from this attack.

We obtained a set of harm-trees occurring as a result of cyber-incidents in the claim cases analysed. We combined these harm trees to produce the graph in Figure 1 displaying the harms and frequency of their occurrences, and also the frequency of relationships between harms (i.e., the frequency with which a particular type of harm led to another). In

1	Ransomware
2	Non-compliance legislation / Malpractice
3	Breach of PII
4	Phishing / Phishing over voice
5	Stolen credentials / Unauthorised access

TABLE I
MOST COMMON TRIGGERS IN THE DATASET.

Figure 1, all nodes connected to the *trigger* node (left in the figure) are attacks, and the rest of the nodes are harms. *Gephi* tool was used to plot the graph [14]. Edges are directed, the reader should interpret the direction of the curved edges clockwise. The thickness of the edges represents the frequency of that link, i.e. the thicker the edge the more frequent the occurrence of this harm. Colours are only used to better differentiate the edges.

Table I and Table II show the most common triggers (first step in an attack tree) and most common harms identified in our dataset respectively. Note that frequency is determined by the addition of weights of all edges running into a node.

1	Lawyer services
2	Privacy / Breach counsel
3	Regulatory violation fine
4	Fraudulent invoice
5	Law suits

TABLE II
MOST COMMON LOSSES IN THE DATASET.

V. LIMITATIONS AND FUTURE WORK

This paper has presented our initial research on harm-propagation trees. While at this stage the number of observations is not large enough to generalise the results, we contribute an initial insight into the development of harm trees to understand the types of harm experienced as a result of cyber incidents, and the ways in which these harms commonly propagate.

Since our data was collected from claim forms, the types of harm identified are limited to the harms that the insurance policy covered, as well as to the coverage that the insured requested. We intend to build on this initial work to extend the list of harms and develop a theory for the occurrence and propagation of cyber-harms experienced by companies as well as the factors that affect harm: we anticipate that this may include the type of attack, the sector of the company, or the size of the company, for example. The next step would be to enrich our dataset with more observations from different insurance companies and other sources, to avoid limiting our data to policies coverages, and create a harm graph that can provide useful insights into mitigating these harms.

ACKNOWLEDGEMENT

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

REFERENCES

- [1] S. Mansfield-Devine, "The Ashley Madison Affair", Network Security. Elsevier, vol. 9, pp. 8–16, 2015.
- [2] I. Agraftiotis, J. R. C. Nurse, M. Goldsmith, S. Creese, and D. Upton, "A taxonomy of cyber-harms: Defining the impacts of cyber-attacks and understanding how they propagate", Journal of Cybersecurity. Oxford University Press, vol. 4, no. 1, pp. tty006, 2018.
- [3] S.J. Schulhofer "Harm and punishment: a critique of emphasis on the results of conduct in the criminal law". UPa L Rev; 122:14971607, 1974.
- [4] J. Kleinig, "Crime and the concept of harm". Am Philos Q;15:2736, 1978.
- [5] SR. Van Slyke, S. Van Slyke, ML. Benson, "The Oxford Handbook of White-Collar Crime", Oxford University Press, 2016.
- [6] VA. Greenfield, L. Paoli, "A framework to assess the harms of crimes". Br J Criminol; 53:864885., 2011.
- [7] M. Levi, "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, 87105, 2011.
- [8] M. Felici, N. Wainwright, S. Cavallini, "Whats new in the economics of cybersecurity?" IEEE Secur Priv;14:1113, 2016.
- [9] R. Anderson, C. Barton, R. Bohme, "Measuring the cost of cybercrime", The Economics of Information Security and Privacy; 265300, 2013.

- [10] A. Punter, A. Coburn, D. Ralph, "Evolving risk frameworks: modelling resilient business systems as interconnected networks", Centre for Risk Studies, University of Cambridge, 2016. (<http://cambridgeriskframework.com/page/17>, last accessed: 25/02/2019)
- [11] Lloyds of London, Counting the cost. (<https://www.lloyds.com/news-andinsight/risk-insight/library/technology/countingthecost>, last accessed: 25/02/2019)
- [12] AIR, "Verisk cyber exposure data standard and AIR Preparers Guide", 2016 (<https://www.air-worldwide.com/Models/Cyber/>, last accessed: 25/02/2019).
- [13] FICO, Cyber Risk Score (<https://www.fico.com/en/products/cyber-risk-score>, last accessed: 25/02/2019).
- [14] Bastian M., Heymann S., Jacomy M., "Gephi: an open source software for exploring and manipulating networks", International AAAI Conference on Weblogs and Social Media, 2009.