Data Analytics for Cyber Risk Analysis Utilizing Cyber Incident Datasets*

Melissa Portalatin
University at Albany
Albany, NY
mportalatin@albany.edu

Omer Keskin
University at Albany
Albany, NY
okeskin@albany.edu

Sneha Malneedi
University at Albany
Albany, NY
smalneedi@albany.edu

Owais Raza
University at Albany
Albany, NY
oraza@albany.edu

Unal Tatar
University at Albany
Albany, NY
utatar@albany.edu

Abstract— The imperative factors of cybersecurity within institutions have become prevalent due to the rise of cyber-attacks. Cybercriminals strategically choose their targets and develop several different techniques and tactics that are used to exploit vulnerabilities throughout an entire institution. With the thorough analysis practices being used in recent policy and regulation of cyber incident reports, it has been claimed that data breaches have increased at alarming rates rapidly. Thus, capturing the trends of cyber-attacks strategies, exploited vulnerabilities, and reoccurring patterns as insight to better cybersecurity. This paper seeks to discover the possible threats that influence the relationship between the human component and cybersecurity posture. Along with this, we use the Vocabulary for Event Recording and Incident Sharing (VERIS) database to analyze previous cyber incidents to advance risk management that will benefit the institutional level of cybersecurity. We elaborate on the rising concerns of external versus internal factors that potentially put institutions at risk for exploiting vulnerabilities and conducting an exploratory data analysis that articulates the understanding of detrimental monetary and data loss in recent cyber incidents. The human component of this research attributes to the perceptive of the most common cause within cyber incidents, human error. With these concerns on the rise, we found contributing factors with the use of a risk-based approach and thorough analysis of databases, which will be used to improve the practical consensus of cybersecurity. Our findings can be of use to all institutions in search of useful insight to better their risk-management planning skills and failing elements of their cybersecurity.

Keywords— Cyber Security, data analytics, VERIS database, etc.

I. Introduction

Cyber risk has become one of the most emerging and prevalent concerns that has directly affected several corporations. Cybercriminals are becoming more sophisticated and practicing new tactics to intrude on networks and disrupt the flow of businesses. The information that is being targeted by cybercriminals is of most importance to the corporations, and the incidents result in loss of data and reputation. The malware used in attacks and other tactics are becoming more intricate and are posing major threats to the cybersecurity posture of the targets. This year alone, 72% of security breaches involved large businesses, and 45% of all security breaches featured hacking [1]. The risk posture of businesses and institutions is growing annually. Organizations need to make an adequate cyber risk management plan. With detection and prevention of malware becoming more strenuous, these businesses and institutions must incorporate and practice an improved style of data-based approaches within their cyber risk management plan.

Proper cyber risk management requires identifying potential risks, analyzing the brunt of those risks, and organization of an effective plan that could be facilitated if necessary [2]. When developing a cyber risk management plan, an efficient data-based approach must be incorporated to ensure consistency [3]. With the current use of conflicted data-based approaches, it is of best interest when developing a cyber risk management plan that all steps are prioritized accordingly. Unlike the maturity-based approach, the data-based approach targets specific areas and capabilities of risk with testing and monitoring [3]. This assists with developing a cyber risk management plan that secures specific potential high risks within institutional systems. In this paper, we used the exploratory data analysis technique within the VERIS dataset to audit the benefits of the data-centric approach within cyber risk management.

II. Literature Review

Within the cybersecurity field, there are several risks and tactics that corporations must be aware of. When developing an efficient plan of defense, these subjects are crucial to be aware of because they inform the focus on specific areas of high risk. Due to the recent stringent overview of policy and regulations on reporting incidents, the number of publicly known data breaches has increased [1]. Cybercriminals choose their targets and techniques strategically, which focuses on the organization's vulnerabilities. As cybercriminals become more sophisticated with their plans, institutions must develop a higher level of security. Instead of solely attributing to defense, it is imperative to create a plan that screens all potential threats of cybersecurity to promote proper mitigation and recovery. The objective of this type of risk management plan is to evaluate the cyber-attack strategies, vulnerabilities, and occurring consequences that distinctively affect the institution as a whole.

This material is based upon work supported by the National Science Foundation under Grant No. 1948261.

Risk management has increasingly become a focal point of research within the context of cybersecurity due to the increasing sophistication of cyber threats [4]. Recent research in this area has centered on developing the most effective risk management methodologies by identifying susceptibility to cyber threats through estimations and patterns of characteristics relating to the targets of these breaches. Liu et al. [5] focused their study on characterizing cybersecurity breaches with the end goal of determining to what extent cybersecurity incidents can be predicted based on externally observable properties of an organization's network. Similarly, Moody et al. [6] focused their study on characterizing candidates susceptible to phishing. The study found candidate constructs, mainly a few personality and other individual factors, are important predictors of phishing susceptibility. Allodi and Massacci [7] performed a study to highlight the limitations of risk assessment procedures. Ultimately this study concludes that quantitative risk estimates are more effective than the currently used qualitative methods because these current methods result in estimates that are widely recognized as unrealistic. Laube and Bohme [8] focused on cyber risk information sharing and how this technique impacts defenders' strategies. The study concluded that there is a need to make existing mechanisms of information sharing more effective. The current cyber risk information-sharing mechanisms neglect information sharing, considering it as a technical problem rather than an economic one [9], [10]. Taking the building blocks of previous research further, Tonn et al. [11] explored prevention of and recovery from cyber attacks. This study found that cyber mitigation and insurance options have been implemented but are generally insufficient as organizations do not have the means to rigorously assess and manage cyber risk. To address the lack of data, Tatar et al. utilized a Monte-carlo reliability analysis technique to calculate the economic impact of cyber incidents in power generation systems [12].

In order to collectively gather information on how to perform proper risk management, researchers have analyzed human components that affect cybersecurity posture. In doing so, *The Vocabulary for Event Recording and Incident Sharing* (VERIS) has also been used to gather cyber incident data and analyze tactics and patterns of incidents that occurred in the previous years. According to Aziz, Lee, and Akkuzu [13], the data within the VERIS database is community-based, which means that it covers many types of cyber-attacks, from attacks on individual people to large organizations like companies and governments. However, this also means that the data is incomplete and possibly inaccurate in many areas due to human error and withheld information. The most incomplete statistics include the victims of the attack and its impact [13].

A risk-based approach is used when gathering information because this focuses on high-risk elements that require attention immediately. Risk management can lead to a reduction in expected loss from security failure incidents [14]. When using a risk-based approach, it is imperative that the focus relies on observing risk and damage, IT solutions, and calculating the expected loss from cyber incidents. Two correlating concepts that are analyzed when using the risk-based approach are incident type and bypass rate, which allows IT professionals to efficiently calculate the return on investment within an institution's security solutions in order to advance their cybersecurity accordingly [14]. Ultimately, institutions can save time and money when using a risk-based approach when developing an efficient risk management plan. Acquiring the use of this new framework is an unconventional concept that is based on avoided risk rather than increased productivity. This approach will also influence the cybersecurity of an institution to revamp security countermeasure investments and reduce spending without forfeiting protection [14].

Researchers have addressed similar issues while analyzing similar reports in order to conclude the problematic issues that have occurred within cyber incidents. Beazley [15] has developed research on the exploratory data analysis of a unified host and network dataset to promote the observation of detrimental tactics and the flow of cyber incidents. Romanosky [16] has contributed to the literature by examining the cost and causes of cyber incidents, which attributes to the understanding of the monetary investments lost during cyber incidents. Murukannaiah [17] also developed research developing a machine learning model for a privacy incident database. Researchers Wu, Kang, and Li [18] developed crucial information on the analysis of risk assessment methods for the cybersecurity of cyber-physical systems. Bapat et al. [19] devised a system that would allow systems to easily detect botnet malware by finding anomalies that often go undetected. Using these studies and reports will assist with developing a conclusion that will elaborate on specific areas of threats and vulnerabilities within institutions and their cybersecurity.

Human-related activities that take place during cyber incidents can deter the effectiveness and outcome of a cyber-attack. Hadlington [20] concluded that the top three types of cyber attack tactics were phishing, malware, and spoofing. Phishing and spoofing can be identified as the contributing factor of human influence. As a whole, the information security community has stated that the weakest link in cybersecurity is humans [20]. In relevance of human factor, it was found that additional aspects can conspire with the lack of understanding from employees about the importance of data, software, and systems within the institution's system, negligent knowledge of the level of risk adhere to direct responsibilities, and unaware of employee behaviors that can attribute to influence cyber risk [20]. Another important concept related to the human factor is the "insider threat," which is also known as the growing concern for vulnerabilities within the institution [21].

The information that was recorded in the IBM Cost of a Data Breach Report [22] and Verizon Data Breach Investigation Report [1] exemplifies the losses caused by data breaches. Along with Advisen Cyber Loss Dataset [23], which contributes to the findings of cyber risks, it can be used to conclude the major effect of these repercussions of a cyber incident. Using similar studies that can overview specific issues within cybersecurity, we revealed explanation and visuals that show human components and the effect it has on cybersecurity posture of institutions. The interesting factors that come with this study are not only of significance to human aspects but also valuable for other detrimental causes of cyber incidents because it is the measurement of high-risk areas overall.

III. Methodology

The aim of this research is to find patterns and behaviors of cyber incidents in areas of weaknesses and to distribute mitigation, defense, and response system against cyber-attacks in correlation with previous experiences. The data analysis method that we have in this study is the risk-based approach which promotes the prioritizing of investments based on the cyber program's effectiveness. After discovering specific areas of weaknesses, institutions must monetarily invest in the areas of high risk that need the most attention. In order to develop a risk-based approach that is effective for an institution, these steps of analyses should be taken to create the prioritized steps needed to cover problematic areas of lacking cybersecurity. In addition, we evaluated human factors in cybersecurity that could enhance risk management plans when not properly practiced that would significantly increase the risks [20].

A. *Dataset*

The VERIS dataset is a publicized metric system designed to provide a common language for elaborating on security incidents that have perpetuated institutions [24]. When first launched in 2010, it allowed institutions to collect, compare, and analyze the shared information given to acclimate a more structured and repeatable practice in their cybersecurity. This dataset was developed and owned by Verizon in order to create a communal database that is publicized and used in practice when evaluating risks. Interpreting this report along with human aspects will assist with the observation of tactics and patterns that were incorporated during these attacks. This information collected from the Verizon report enables interactive capabilities that contribute to the findings of overall patterns of cyber incidents in high-risk areas. Other aspects, such as the US Department of Health and Human Services and Attorneys Generals that assist with breach notifications, were also included in this dataset to capture beneficial information that can possibly help other institutions improve their cybersecurity. This dataset includes the continuous patterns of assets, threats, impacts, and controlling landscapes of previous cyber incidents. The ultimate goal of VERIS is to show visuals of the foundation that was developed previously while using that as leverage to learn from experiences to measure and manage risk formally [20].

The Verizon Data Breach Investigations Report [1] correlates with the VERIS dataset because it includes several incidents that were discovered using VERIS. After a thorough analysis of the VERIS dataset, it will devote to the findings of how organizations can potentially have more leverage to predict cyber incidents, along with the most efficient way to calculate cyber risk. Exploratory data analysis was also conducted during this research to capture the latest trends among cyber incidents. The overall goal of exploratory data analysis of the VERIS dataset and similar studies is to create awareness of emerging threats and vulnerabilities to protect the cybersecurity of organizations. Therefore, it focuses on weaknesses within the institutions' cybersecurity rather than concentrating on the institution's overall cybersecurity.

B. *Exploratory Data Analysis of Cyber Incidents*

The demographics of this research can contribute to the outcome that reflects the nature of risk. These contributing factors would be based on targeted industries, such as healthcare, finance, and information sectors that have been involved in a cyber incident [25]. This information gave us leverage in order to find the effectiveness of a risk-based approach when developing a risk management plan. We found that the root cause of this issue is the inconsistent framework that is implemented when developing a risk-management plan that does not target specific areas of weakness but focuses on the overall cybersecurity of the institution. This directly conflicts with the understanding of efficient cybersecurity and how to implement the idea of a risk-based approach [26]. This study showed that institutions should create risk management plans that focus on high-risk threats and likelihood [26]. In VERIS, it shows the examination of evidence and post-incident analysis that can be used to develop a useful risk management plan. In the IBM Cost of a Data Breach Report [22], it is stated that the average cost of a data breach in 2020 was $3.35 million, which was a 9.8% increase from the previous year. During the same timeframe, it was found that the cost of each lost or stolen record was $163, contributing to the factor with a 3.8% increase from the last year. 80% of these breaches included the exposure of customers' personally identifiable information and concluded with an average of 211 days to identify and contain the breach within the institution [22].

Figure 1 shows the number of cases over time. Figure 2 presents the typical size of the victim organizations. Figure 3 shows the sectors that have been targeted. These graphs were developed to create a visual comprehension of the factors that contribute to these aspects of cyber incidents.

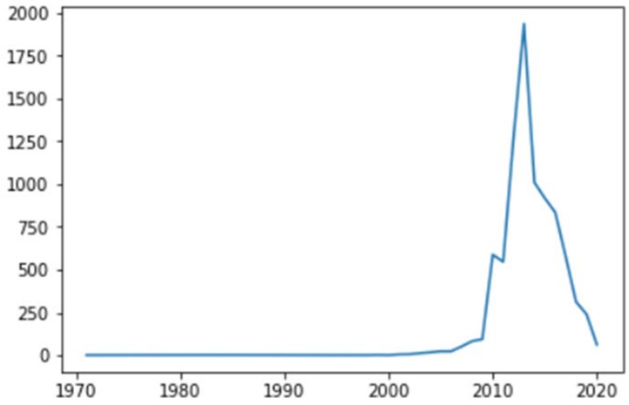

Fig. 1. Number of cases over time

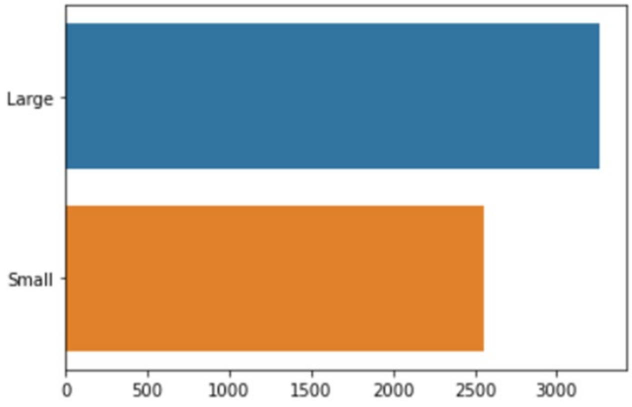

Fig. 2. Organization Size

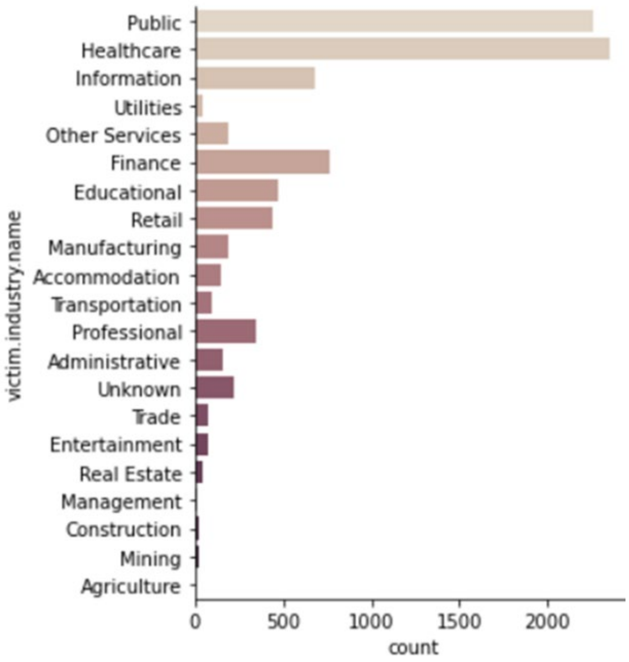

Fig. 3. Targeted Sectors

Incidents commonly disclose private and healthcare information, which is a major loss that institutions need to consider when practicing an effective risk management plan. In the Advisen Cyber Loss Dataset [23], we found that the financial losses that attributed to response cost, litigation, economic losses, fines, and penalties were a total of $20 billion. In the Verizon Data Breach Investigations Report [1], it was concluded that there were 157,525 cyber incidents in 2020. Of those incidents, 3,950 were confirmed to be data breaches, and 86% of those were financially motivated. Collectively, the Verizon Threat Research Advisory Center concluded that because of thorough monitoring and proper research, the trends of cyber-attacks are becoming more familiar. This can assist with bringing awareness to institutions about the commonality between patterns and targets of cyber incidents, resulting in proper use of the risk-based management approach.

IV. Findings

In this paper, it is clear that institutions are sacrificing proper cybersecurity for monetary assets. We found that 97% of breaches included the loss of confidentiality regarding Confidentiality, Integrity, and Availability (CIA Triad). Confidentiality ensures only the authorized users can access the data; integrity ensures that only the authorized users can modify the data, and availability is about the systems being operable whenever the users need them [27]. Common causes of these cyber incidents are human error (30%) and machinery misuse (21%), by which the majority are human vulnerabilities that are primarily exploited by phishing attacks via email.

In Figure 2, it is shown how cyber incidents have been attributed to the CIA Triad and the attribution to what area is of most risk during cyber incidents.

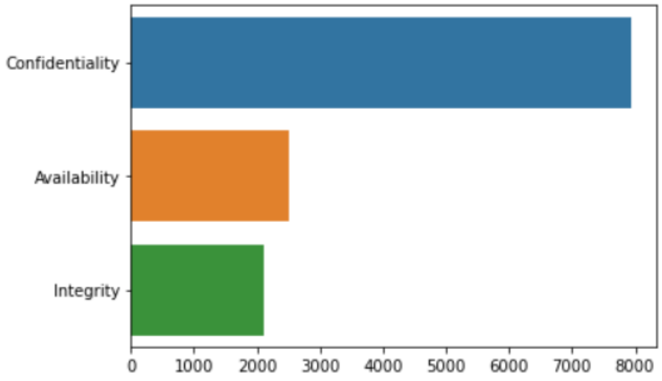

Fig. 4. Distribution of the breaches based on attribute

The categories in which these action types fall are error, hacking, physical, malware, social, environmental, and unknown (Figure 3). This is a visual graph that we have created to emphasize the tactics that are prevalently used when exploiting vulnerabilities

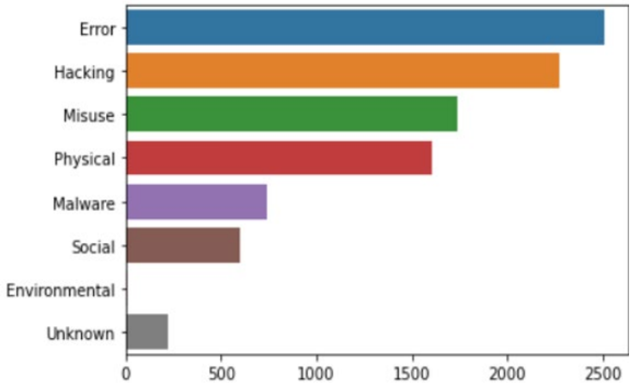

Fig. 5. Distribution of the breaches based on action types

Figures 4 and 5 exemplify the relationships between action by attribute and attribute by action, in a correlation with the CIA triad and action type previously listed.

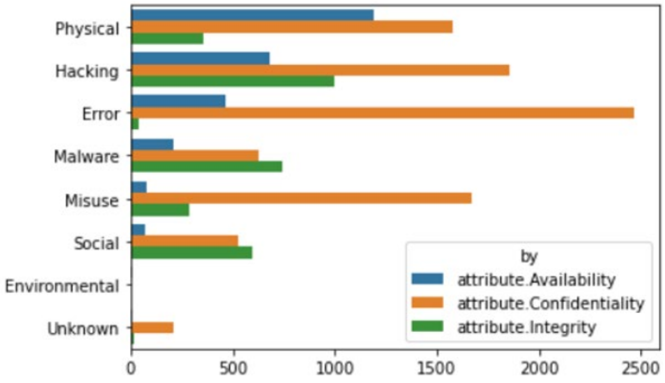

Fig. 6. Action by attribute

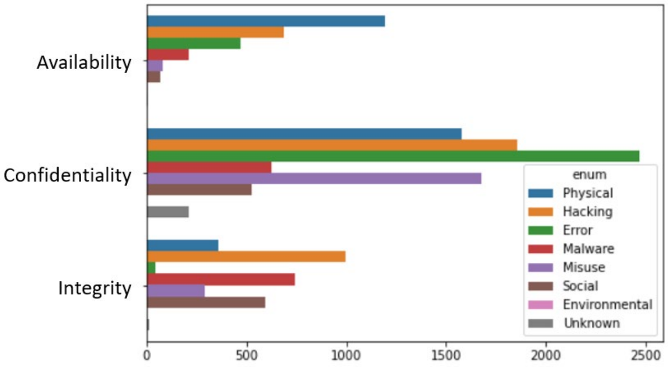

Fig. 7. Attribute by action

In addition, we evaluated the relationship between external actors versus internal actors that contribute to an institution's vulnerabilities. External actors have been twice as more vulnerable to error than internal actors. External actors have been twice as more vulnerable to hacking attacks than internal actors. External actors have also caused more incidents than internal actors in the subject of computer misuse.

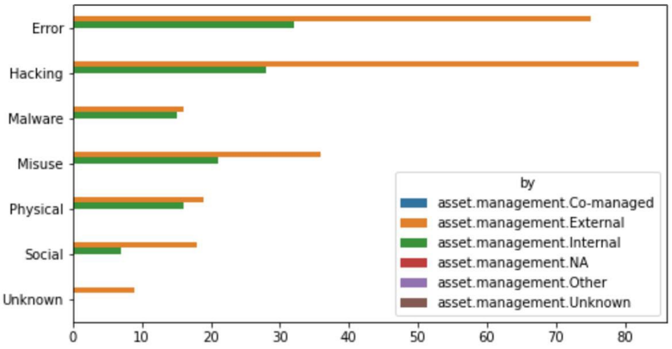

Fig. 8. Distribution over external and internal actors

Figure 6 represents a visual of the relationship between external versus internal actors that attribute with the different levels of asset management.

V. Conclusion

Cyber risk assessment plans have become more challenging over time due to the progression of sophisticated cyber threats and trends. We found that the process of analyzing cyber incident databases thoroughly can benefit an institution's cybersecurity because it familiarizes IT professionals with the prevalent strategies and trends that have been used to exploit vulnerabilities. Risk managers can then analyze the data of previously exploited vulnerabilities to calculate the likelihood and consequences of a potential future cyber incident. Although this framework could be claimed as unconventional, institutions should take advantage and use this as leverage to design adequate mitigation actions for future attacks.

In accordance with the human factor and other components, we found that it is essential that all employees, regardless of position, are aware of cyber risk and their devoting responsibilities that can potentially taint the institution's cybersecurity and mitigation plan, as well as being fully knowledgeable and well trained on the framework incorporated in the institution's risk management plan. As previously stated, the human factor is one of the "weakest links" in cybersecurity, and in order to prioritize this issue, we found it is best to consistently be transparent with the entire institution on the subject of trending malicious cyber tactics. If humans are not taught to analyze and react to cyber threats accordingly, then this cycle will remain unbroken. Cybercrime is increasing at elevating rates, and institutions must assimilate to this framework in order to skillfully protect and precisely prioritize mitigation.

References

[1] Verizon, "2020 Data Breach Investigations Report," *Verizon Enterprise*, 2020. https://enterprise.verizon.com/resources/reports/dbir/ (accessed Jan. 05, 2021).

[2] Advisen, "Cyber Risk Data Methodology for Insurance & Risk Analysis." [Online]. Available: https://in.advisenltd.com/cyber-risk-data-methodology/.

[3] B. Fischer, "Risk-Based Approach to Cyber and Information Security - SCA - Call Now," Apr. 16, 2019. https://www.scasecurity.com/risk-based-approach/ (accessed Jan. 12, 2021).

[4] Ü. Tatar and B. Karabacak, "An hierarchical asset valuation method for information security risk analysis," in *International Conference on Information Society (i-Society 2012)*, 2012, pp. 286–291.

[5] Y. Liu *et al.*, "Cloudy with a Chance of Breach: Forecasting Cyber Security Incidents," in *24th {USENIX} Security Symposium ({USENIX} Security 15)*, 2015, pp. 1009–1024.

[6] G. D. Moody, D. F. Galletta, and B. K. Dunn, "Which phish get caught? An exploratory study of individuals′ susceptibility to phishing," *European Journal of Information Systems*, vol. 26, no. 6, pp. 564–584, Nov. 2017, doi: 10.1057/s41303-017-0058-x.

[7] L. Allodi and F. Massacci, "Security Events and Vulnerability Data for Cybersecurity Risk Estimation," *Risk Analysis*, vol. 37, no. 8, pp. 1606–1627, 2017, doi: https://doi.org/10.1111/risa.12864.

[8] S. Laube and R. Böhme, "Strategic Aspects of Cyber Risk Information Sharing," *ACM Comput. Surv.*, vol. 50, no. 5, pp. 1–36, Nov. 2017, doi: 10.1145/3124398.

[9] U. Tatar, B. Karabacak, and A. Gheorghe, "An Assessment Model to Improve National Cyber Security Governance," in *11th International Conference on Cyber Warfare and Security: ICCWS2016*, 2016, p. 312.

[10] U. Tatar, B. Karabacak, P. F. Katina, and A. Igonor, "A complex structure representation of the US critical infrastructure protection program based on the Zachman framework," *International Journal of System of Systems Engineering*, vol. 9, no. 3, pp. 221–234, Jan. 2019, doi: 10.1504/IJSSE.2019.102869.

[11] G. Tonn, J. P. Kesan, L. Zhang, and J. Czajkowski, "Cyber risk and insurance for transportation infrastructure," *Transport Policy*, vol. 79, pp. 103–114, Jul. 2019, doi: 10.1016/j.tranpol.2019.04.019.

[12] U. Tatar, H. Bahsi, and A. Gheorghe, "Impact assessment of cyber attacks: A quantification study on power generation systems," in *2016 11th System of Systems Engineering Conference (SoSE)*, Jun. 2016, pp. 1–6, doi: 10.1109/SYSOSE.2016.7542959.

[13] B. Aziz, J. A. Lee, and G. Akkuzu, "Evaluating the Quantity of Incident-Related Information in an Open Cyber Security Dataset," in *International Conference on Business Information System*, Cham, 2019, vol. 373, pp. 531–542, doi: 10.1007/978-3-030-36691-9_45.

[14] A. Arora, D. Hall, C. A. Piato, D. Ramsey, and R. Telang, "Measuring the risk-based value of IT security solutions," *IT Professional*, vol. 6, no. 6, pp. 35–42, Nov. 2004, doi: 10.1109/MITP.2004.89.

[15] C. Beazley *et al.*, "Exploratory Data Analysis of a Unified Host and Network Dataset," in *2019 Systems and Information Engineering Design Symposium (SIEDS)*, Charlottesville, VA, USA, Apr. 2019, pp. 1–5, doi: 10.1109/SIEDS.2019.8735640.

[16] S. Romanosky, "Examining the costs and causes of cyber incidents," *Journal of Cybersecurity*, vol. 2, no. 2, pp. 121–135, Dec. 2016, doi: 10.1093/cybsec/tyw001.

[17] P. Murukannaiah, C. Dabral, K. Sheshadri, E. Sharma, and J. Staddon, *Learning a Privacy Incidents Database*. 2017.

[18] W. Wu, R. Kang, and Z. Li, "Risk assessment method for cyber security of cyber physical systems," in *2015 First International Conference on Reliability Systems Engineering (ICRSE)*, Oct. 2015, pp. 1–5, doi: 10.1109/ICRSE.2015.7366430.

[19] R. Bapat *et al.*, "Identifying malicious botnet traffic using logistic regression," in *2018 Systems and Information Engineering Design Symposium (SIEDS)*, Charlottesville, VA, Apr. 2018, pp. 266–271, doi: 10.1109/SIEDS.2018.8374749.

[20] L. Hadlington, "The 'Human Factor' in Cybersecurity: Exploring the Accidental Insider," *Psychological and Behavioral Examinations in Cyber Security*, 2018. www.igi-global.com/chapter/the-human-factor-in-cybersecurity/199881 (accessed Jan. 18, 2021).

[21] A. Sasse, S. Brostoff, and D. Weirich, "Transforming the 'Weakest Link' — a Human/Computer Interaction Approach to Usable and Effective Security," *BT Technology Journal*, vol. 19, Aug. 2001, doi: 10.1023/A:1011902718709.

[22] IBM, "Cost of a Data Breach Report 2020," 2020. https://www.ibm.com/security/digital-assets/cost-data-breach-report/ (accessed Jan. 05, 2021).

[23] Advisen, "Cyber Loss Data," *Advisen Ltd.* https://www.advisenltd.com/data/cyber-loss-data/ (accessed Apr. 09, 2021).

[24] "The VERIS Framework." http://veriscommunity.net/index.html (accessed Apr. 09, 2021).

[25] B. Karabacak and U. Tatar, "Strategies to Counter Cyber Attacks: Cyber Threats and Critical Infrastructure Protection," *Critical Infrastructure Protection*, vol. 116, p. 63, 2014.

[26] M. U. Aksu *et al.*, "A quantitative CVSS-based cyber security risk assessment methodology for IT systems," in *2017 International Carnahan Conference on Security Technology (ICCST)*, Oct. 2017, pp. 1–8, doi: 10.1109/CCST.2017.8167819.

[27] National Institute of Standards and Technology, "Glossary | CSRC." https://csrc.nist.gov/glossary (accessed Apr. 13, 2021).

Author Information

Melissa Portalatin is a senior at the College of Emergency Preparedness, Homeland Security, and Cybersecurity at the University at Albany. Melissa is the recipient of the Presidential Award for Undergraduate Research in 2021.

Omer Keskin is a Ph.D. candidate in Engineering Management at Old Dominion University and an M.S. student in cybersecurity at the University at Albany. His area of research includes organizational cybersecurity risk management, economics of cybersecurity, cyber insurance, cybersecurity education, supply-chain security, and blockchain technology.

Sneha Malneedi is a high school senior looking to pursue a degree in computer science starting in the fall. Her current interests within the field include cybersecurity, artificial intelligence, and game development.

Arthur Owais Raza is a graduate student studying Cyber Security Threat Analysis with an expected graduation date of Fall 2021.

Dr. Unal Tatar is an assistant professor of cybersecurity in the College of Emergency Preparedness, Homeland Security, and Cybersecurity at the University at Albany. Dr. Tatar worked as a principal cybersecurity researcher in government, industry, and academia for over 15 years. He is the former coordinator of the National Computer Emergency Response Team of Turkey. Dr. Tatar's research is funded by the National Science Foundation, National Security Agency, Department of Defense, NATO, and Society of Actuaries. His main topics of interest are cybersecurity risk management, economics of cybersecurity, cyber insurance, privacy, cybersecurity education, supply-chain security, and blockchain. Dr. Tatar holds a BS in Computer Engineering, an MS in Cryptography, and a Ph.D. in Engineering Management and Systems Engineering.