Reaction to Data Breaches: A Practitioner-Public View of Organizational Lapses in Security and Ransomware Attacks in 2020

Early stage paper

Paras Bhatt
The University of Texas at San Antonio
paras.bhatt@utsa.edu

Rohit Valecha
The University of Texas at San Antonio
rohit.valecha@utsa.edu

H. Raghav Rao
The University of Texas at San Antonio
hr.rao@utsa.edu

ABSTRACT

There have been numerous data breach incidents and ransomware attacks during the last few years, which have eroded trust in organizations and caused anguish and concern. Using a data driven approach we study the reaction to data breaches by practitioners and the public by analyzing two datasets composed of Verizon's Data Breach Investigation Report (DBIR) 2021 and social media discourse from Twitter. In the DBIR, the ransomware and data breach incidents are discussed by practitioners with detailed summaries about the incidents. In contrast social media discourse from Twitter is by the public. In this paper we study reactions to these incidents focused primarily on organizational lapses in security and on ransomware attacks. Since data breach incidents and ransomware attacks can affect any organizations and individuals irrespective of their cyber defenses it is important to understand how practitioners and the social media users discuss these incidents. Based on an LDA topic modeling approach we observe that topical differences in opinions with regard to practitioners and public discourse exist in issues such as loss, laws, information compromise, and cost of cyber threats. Our findings indicate that (a) public reactions on social media discuss personal aspects of data breaches such as their private information or credentials leaking online, and the security threats & targets of ransomware attacks; and (b) practitioners' reports discuss the information compromised in data breaches and how ransomware attacks are increasingly being deployed to disrupt organizations' ability to use data. These similarities and differences regarding public and practitioner viewpoints can help in creating actionable cyber threat intelligence.

Keywords

Data breach, ransomware, public reaction, social media, Twitter.

INTRODUCTION

During recent years, there have been numerous data breaches as is evidenced by the breach incidents (Bassett et al, 2021). Several data breaches have led to the exposure of healthcare data (Alkinoon et al, 2021) which violated several Fair Information Practice Principles (FIPPs)[1] (such as security, transparency, quality and integrity) and delayed the adoption of contact tracing and surveillance applications. The latest Verizon DBIR (Bassett et al, 2021) lists 347 data breach incidents. Of them many were perpetrated by adversaries through ransomware and others occurred as a result of organizational lapses in security.

Prior research has looked at the adverse consequences of data breaches (Nikkah and Grover, forthcoming) from the point of view of organizations; whether their reputation is affected or not, whether their sales are affected or not (Janakiraman et al, 2018; Syed, 2019), etc., but there are limited studies that seek to understand how people respond to and discuss data breaches (Bachura et al, 2022). Some of the literature is focused on studying data breaches as a singular incident with

[1] https://iapp.org/resources/article/fair-information-practices/

a single point of failure – the organization – assuming that the organization was not able to protect users' personal data and as a result of their lack of oversight such a data breach occurred (Bentley et al, 2018; Hammouchi et al, 2019). However, in the recent past, the public social media discourse has changed to include the narrative focused on the victim as well as the narrative focused on the aggressor. This is possibly because the aggressors have become more prominent and reveal themselves partially, implying that they are known to an extent. In the previous case of data breaches, the aggressors (virus, worms, black hat hackers, etc.) were anonymous, now in the case of ransomware attacks they are pseudo anonymous because the hackers need to monetize the ransom transactions.

In this paper we suggest that the narrative shift is captured in both cyber security threat reports such as DBIR and public reactions on social media platforms like Twitter that discuss data breach and ransomware incidents. Therefore, a study of the practitioner and public discourse between ransomware attacks and data breaches is a gap that needs to be understood. We believe that an analysis of data breaches and ransomware incidents is needed to enable a comparison between the practitioners' response and the public response to cyber threats, to create actionable cyber threat intelligence and prevent harms from such incidents.

In order to develop defenses against data breaches, it is necessary to understand how such attacks affect citizens and it is important to study people's responses to such breaches of personal data. In this regard, social media platforms have become a reliable avenue for collecting and analyzing data regarding data breaches (Bachura et al, 2022). Social media datasets, for example from Twitter (Bhatt et al, 2022), provide an opportunity to better understand social media users' conversations regarding potential problems arising from the processing of personally identifiable information (PII) (e.g., unanticipated revelation of personal data, lapse in data security, loss of self-determination and trust) from data breaches.

We collect one dataset of data breach incident reports from the Verizon's DBIR of 2021 that describes the major data breach incidents of the year 2020, and we collect a second data set of social media conversations from Twitter using keywords of ransomware and data breaches. In this regard, we aim to answer the following research question:

- RQ 1 – How do major topics of discussion differ between practitioner reports and the public reaction related to data breaches and ransomware attacks?
- RQ 2 – What are the similarities and differences in major topics of discussion that arise during discourse regarding data breaches that occur as result of organizational lapses in security as compared to discourse regarding ransomware attacks?

We use a data driven approach to identify major topics associated with data breaches and ransomware attacks. Using datasets collected from Twitter streaming API and data breach incidents from Verizon's DBIR we build n-gram models that uncover aggregate social media conversations regarding the major topics of discussion within each. We use advanced aspect-based topic modeling built on transformers of Google's BERT engine to understand the key discussions about data breaches and ransomware incidents. We compare and contrast the results from two types of cyber threats: ransomware vs. organizational lapses in security to analyze how social media conversations depict and discuss such incidents. We draw upon natural language processing for this investigation to extract major topics of discussion regarding data breaches and ransomware. We also link these topics to the major FIPPs such as security, minimization, accountability, transparency and quality.

Data breach is an ever-increasing problem in the cybersecurity ecosystem (Cranor et al, 2015) and through this paper we provide several contributions as well as theoretical and practical implications for studying such breaches. First, our work in understanding the differences between practitioner and public-oriented discourse will show how data breaches and ransomware attacks are perceived and reported. This will help to spur research in how security issues such as ransomware attacks and massive breaches of personal data should be communicated effectively to users. This is helpful for both organizations and policymakers for determining the best approach to communicate data breach events without causing mass panic and hysteria or loss of confidence and trust in online services and portals, which is so often seen in incidents that involve personal data breaches (Bachura et al, 2022). Second, this paper recognizes that data breach events can be seen from multiple perspectives, for example, in a victim-oriented perspective data breaches are seen as being within organizational control and preventable, but organizations fail to do so, and in an aggressor-oriented perspective, data breaches are outside the organizational control wherein massive ransomware attacks dilapidate the entire security infrastructure of organizations. The distinction between multiple perspectives can help to develop and test automatic methods of ensuring data security and preemptively fixing issues that may result in such data breach incidents occurring in the first place. Specifically, extracting and understanding the major topics of discussion in data breaches using longitudinal data will enable an in-depth analysis of social media users' and practitioners conversations about protection of their personal data and help in developing better definitions and usage of security controls available to users. Also, such understanding can be used to respond to personal data breaches in future and ensure robust data security.

The rest of the paper is structured as follows. In the next section, we present the literature on ransomware, organizational errors and data breach discussions on social media platforms followed by the research model. Then the following section presents research methods and techniques as well as our data collection and analysis, which is followed by the results section in which we discuss our major findings. We then discuss the implications of our work and research contributions in the discussion section followed by the conclusion section.

LITERATURE REVIEW

Causes of Data Breaches

Ransomware and Adversarial Attacks on Personal Data

Ransomware incidents have become a global incidence and have risen since the last decade (Popoola et al, 2017). The last two years of the recent Covid pandemic have seen a significant jump in the number of ransomware attacks particularly targeted at the healthcare industry (Spence et al, 2018; Branch et al, 2019). Ransomware mostly affects personal data of users and organizations and researchers have called for ransomware to be considered as a data security breach issue (Brewczyńska et al, 2019). Prior literature suggests that the next frontier of ransomware attacks will increasingly be on users' personal data stored on mobile devices (Faghihi and Zulkernine, 2021) which contain sensitive and personal data as compared to organizational data. Kozlowska (2018) notes that such attacks will remain frequent until there is a fundamental change in organization's information security policies and how they protect users' personal data.

In order to study the wide scale effects of ransomware attacks there does not exist a comprehensive framework that takes into account social and technical perspectives of the attack. Some studies focus on understanding the technical metrics that allow malicious hackers to mount these attacks (Kharraz1 et al, 2018). They focus on mainly on encryption and communication techniques and suggest that the modus operandi of ransomware attacks have largely remained the same over the years: locking all the files on a network and asking for a ransom to unlock them. Another issue is that there is a lack of reporting requirements or the availability of a standardized format that describes the critical aspects of an attack that must be reported either to regulatory agencies or the wider community as a whole (Branch et al, 2019). In this regard, it becomes important to understand what metrics of ransomware are discussed in practitioner reports and what implications they can provide for information security research. To this end, in this paper we focus on classifying threat reports using techniques such topic modeling and hierarchical clustering. We use natural language processing tools to automatically extract the data beach related aspects in these reports as well as from social media discussions regarding such breaches in general.

Organizational Lapses in Security and Loss of Personal Data

Lack of training, outdated technology resources, unwillingness to buy advanced protection software applications and lax security policies are all examples of errors that can be directly attributed to the organization. As a result of these organizational errors there have been several cases of data breaches. Researchers have noted that users are the weakest link in the security chain (Moody et al, 2020) but often organizations themselves are to blame for being targeted by cyber-attacks (Popoola et al, 2017). From OPM hacks to several other wide scale attacks that resulted in the unwanted disclosure of private data (Bachura et al, 2022), people have criticized the organizations' role in responding to the attack or mitigating the harmful effects of unwanted data disclosure.

In this paper, we use content analysis techniques (Azeez and Van der Vyver, 2019) to investigate the type of information put forward by data breach reports as well as in social media discourses to determine the key issues in the context of organizational lapses in security and loss of personal data.

Data Breaches Discussions on Social Media

Data breach incidents are heavily discussed on social media platforms and give rise to emotional conversations among users in which they express a variety of emotions from anger and sadness and to disgust and fear (Bachura et al, 2022). They also result in the loss of trust and an erosion of privacy. Often, data breaches are communicated by organizations as trivial incidents and sometimes by expressing feelings of remorse. Researchers have noted that there is a strong connection between what companies say (express via written communique regarding a data breach they experienced) and what they do after detecting a breach (actions taken to prevent further data leakages) (Ayaburi and Treku, 2020). Thus engaging and communicating on social media networks with users affected by the personal data breach is an important communication mechanism for organizations.

RESEARCH MODEL

The research model for this paper is discussed in Figure 1. We use a data driven model and analyze two datasets. The first dataset is derived from Verizon's Data Breach Investigation Report for the 2020-21. It includes a detailed description of the cyber-attacks that occurred during this period and has comprehensive information relating to each cyber threat incident such as the incident summary, identity of the victim/company affected by the breach, the industry it normally serves, whether it is a government entity and its location. The second dataset is composed of users' conversations collected from Twitter for the period under study. The data includes various attributes such as a unique identifier for the tweet, a timestamp of when it was posted, the username and the text it contains.

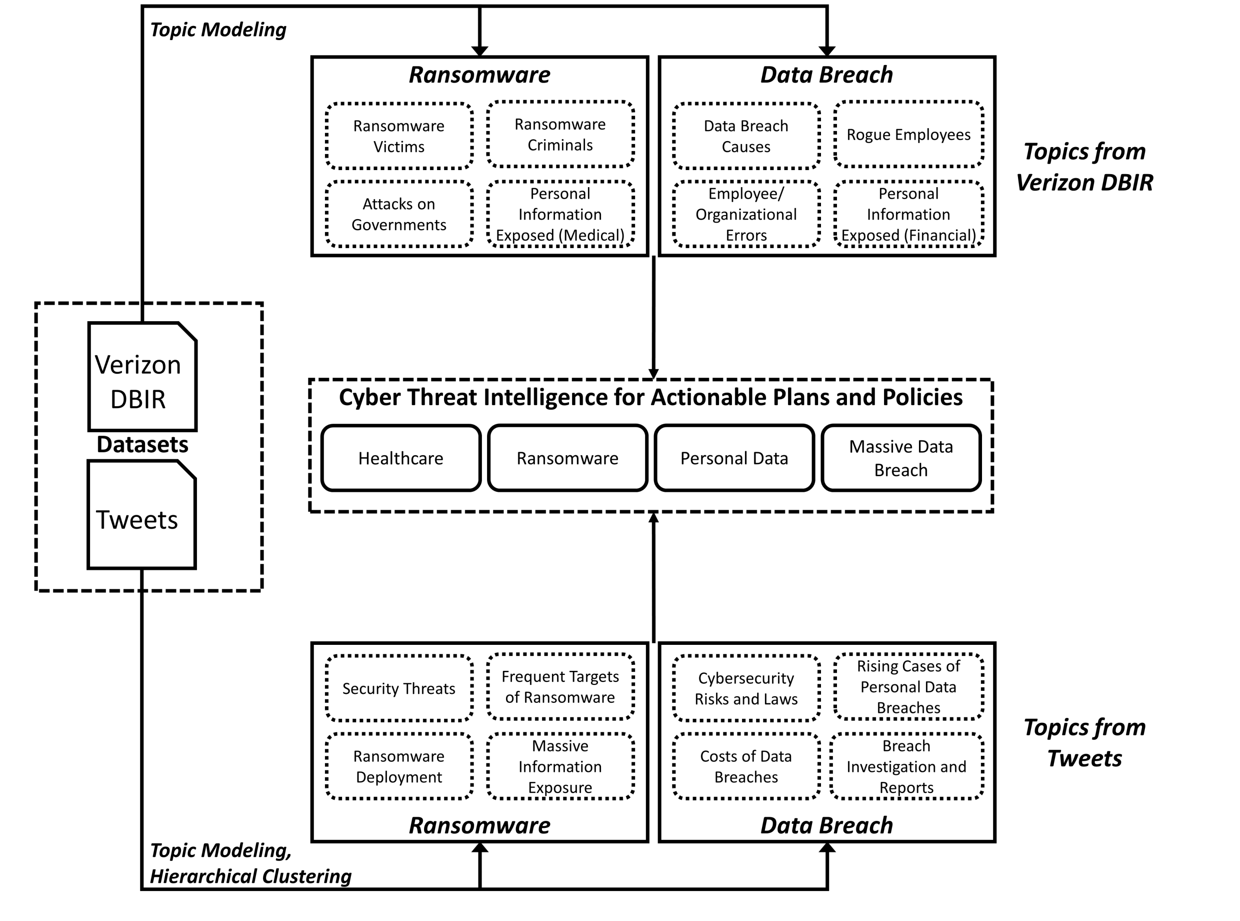
Topic Modeling
Ransomware
Ransomware Victims
Ransomware Criminals
Attacks on Governments
Personal Information Exposed (Medical)
Data Breach
Data Breach Causes
Rogue Employees
Employee/ Organizational Errors
Personal Information Exposed (Financial)
Topics from Verizon DBIR
Verizon DBIR
Datasets
Tweets
Cyber Threat Intelligence for Actionable Plans and Policies
Healthcare
Ransomware
Personal Data
Massive Data Breach
Security Threats
Frequent Targets of Ransomware
Ransomware Deployment
Massive Information Exposure
Ransomware
Cybersecurity Risks and Laws
Rising Cases of Personal Data Breaches
Costs of Data Breaches
Breach Investigation and Reports
Data Breach
Topics from Tweets
Topic Modeling, Hierarchical Clustering

Figure 1. Data Driven Model

We use various natural language processing (NLP) tools for cleaning and preprocessing both the datasets. Specifically in Python we use regex (regular expression) to remove text characters and handles that are not recognized (#,@,$,/,;)) while running the topic modeling and hierarchical clustering tasks.

After cleaning the data, we proceed with our in-depth data analysis. We use a Latent Dirichlet Allocation (LDA) topic modeling approach to extract key topics discussed in reports of data breach and ransomware incidents from the DBIR and social media conversations about such incidents.

Upon analyzing our results, we provide several propositions regarding the practitioners and public view on data breaches and ransomware attacks and map our extracted topics to share actionable cyber threat intelligence (Figure 1).

METHODOLOGY

Data

We base our data collection from the Verizon Data Breach Investigation Report (Basset et al, 2022) for collating information about threats and attacks over the period of one year. We also collected another dataset focused on social media conversations related to data breaches and ransomware attacks in particular. We used Twitter streaming API and developer accounts to collect the tweets using hashtags (for example #privacyMatters) or using event identifiers (for example search strings like 'data breach'). We identified a total of N = 341 data breach incidents in the DBIR report. Of this we have N = 241 as ransomware and adversarial attacks and N = 141 incidents as attacks focusing on personal data breaches. In some of the incidents, there exists an overlap between ransomware incidents that affect personal data as well.

The number of topics extracted in the social media dataset is proportionate to the total number of tweets that discussed data breaches and ransomware incidents. The topic modeling resulted in the extraction of 10 major topics for DBIR which were later combined into 4 major topic clusters for ransomware and 4 for data breaches. Similarly, we extracted 9 topics for the social media dataset which were also combined into 4 major clusters for ransomware and 4 for data breaches (in data breach 2 similar topics were grouped as one) which discuss key issues related to ransomware attacks and personal data breach, as perceived by the users. This framework can be utilized for discussing the negative effects of data breaches and how to mitigate or plan for them.

Dataset	No.	No. of topics	Keywords Searched
DBIR	347 (incidents)	10	Ransomware, Data Breach, Information, Security, Breach
Social media	735,661 (tweets)	9	Ransomware, Data Breach, Personal Data, Information, Security, Breach

Table 1. Data Metrics

Techniques

Topic Modeling

We used LDA Topic modeling (Blei, 2012) for both the datasets to extract the major topics of discussion within each. Such techniques have increasingly been used to study discussions on social media platforms especially on Twitter (Bhatt et al, 2022) as it allows limited characters which makes it easier to efficiently summarize users' conversations. In essence, topic models provide sets of keywords that are closely related to individual tweets and can be used as proxy to establish the point of discussion within the said tweet. For the DBIR dataset we ran a Python lda_mallet model over 100 iterations (iter=100) that went through the entire dataset 10 times (passes=10) with hyperparameters set to extract 15 topics (n=15) and based on the coherence value we selected the model with 10 topics. We replicated this model with the Twitter dataset but had to combine two similar topics and then proceeded with 9 topics of analysis.

Across the topics extracted from the two datasets, the authors systematically analyzed the keywords by their weightage and classified the tweets into various topic clusters related to data breaches and ransomware that were related to the data breach discussions on Twitter and the threat investigation DBIR report. We further proceeded to conduct hierarchical clustering analysis of the extracted topics to provide a much more granular view of the topics of discussion related to such incidents.

Hierarchical Clustering

In this paper, we have further augmented our n-gram topic modeling approach for text classification by using the hierarchical clustering technique (Singh et al, 2018) to derive the central topics within social media conversations that relate to the discussions around the data breaches and threat reports. Using the model, a general distribution can be obtained that showcases what key central themes social media users discuss when they talk about data breaches in particular that occur as a result of ransomware attacks versus those which occur as a result of organizational errors. The use of hierarchical clustering technique is apt to capture the user discussions on Twitter as has been evidenced by previous studies on social media considering the growth seen in the magnitude users posts and discussions during critical events (Albanese and Feurstein, 2021). For each of the topic models across our datasets, we have central clusters that have been generated so that patterns in user discussions about data breach themes can be observed.

RESULTS

Our results from the topic modeling for data breaches reports and user conversations on Twitter show that there are 10 major topics that are discussed in the data breach reports and 9 major topics in the Twitter dataset. These topics depict the nature of the discussions with regard to the personal data exposed or the consequences of the data breach and ransomware for an organization. From our results we observed several differences between ransomware and data breaches (See Table 2 and Table 3) in terms of the topics extracted. These differences can be attributed to the nature of the discussions wherein users' perception about data breaches and ransomware attacks diverge based on who is affected. From an organizational point of view, when ransomware attacks happen users tend to talk about the severity of the attack (economic loss). In contrast, when data breaches happen, users lament the effect it has on their personal data (privacy loss).

Topic	**Description**
Ransomware	
Security Threats	These discussions were about the growing number of ransomware attacks and data breaches that have frequently been occurring.
Massive Information Disclosures	Focused on how massive personal data repositories are being attacked and personal information is being disclosed online.
Frequent Targets of Ransomware	Focused on the major companies that have been struck by an ransomware attacks or had its data exposed.
Ransomware Deployment	The discussions were about the manner of ransomware deployment wherein organizational systems were hijacked or data centers were held hostage.
Data Breach	
Cybersecurity Risks and Laws	These tweets focused on discussing the various increasing cybersecurity risks associated with the sharing of personal data and whether there are enough laws and are they strong enough to prevent data breaches.
Costs of Data Breaches	These tweets focused on how data breaches are costing so much money for individuals.
Breach Investigation and Report	These tweets actively discussed whether the breach investigations have revealed any information about the compromised personal data.
Credentials Exposed Online	Focused on how data breach incidents are increasingly resulting in the unwanted exposure of user credentials online.
Personal Information Exposed	These tweets discuss how sensitive personal information circulates online after data breach incidents.

Table 2. Twitter Major Topics

Topic	Description
Ransomware	
Attacks on Governments and Public Institutions	These reports focused on several government run services and departments that had been targeted by such cyber-attacks.
Ransomware Victims	Focused on who the victims are and their current state of affairs.
Ransomware Criminals	These reports discussed who or which group of hackers were responsible for the ransomware attacks mentioned in the report.
Personal Information Exposed (Medical and Grades)	These presented information regarding the personal information that was compromised as a result of the data breaches such as medical healthcare information and student grades.
Data Breach	
Data Breach Causes (Stolen APIs/ Misconfiguration)	These reports discussed the major reasons for data breach such as stolen application programming interfaces and misconfiguration of security settings or applications.
Personal Information Exposed (Financial Information Exposed)	These reports presented information regarding the personal information that was compromised as a result of the data breaches such as financial information like loans and credits.
Employee Errors/ Organizational Errors	Focused on attacks and breaches that happened as a result of employee negligence or fault and organizational oversight or lapse in security.
Rogue Employees	Focused on the rogue employees that perpetrated/assisted in data breaches.

Table 3. DBIR Major Topics

DISCUSSION AND CONTRIBUTIONS

Data breaches and ransomware attacks primarily violate many of the FIPPs and cause several potential problems for organizations that severely affect users' personal information. In this regard, the FIPPs of security, transparency, accountability, minimization, quality and integrity are directly relevant for protecting users' privacy and security of data. Data breaches in particular call into question the security measures put in place by organizations to prevent unwanted access and disclosure of personal information of users. Also, if such a breach of personal data occurs, the organizations is accountable for such lapses in security. In this work, we have coded the topics extracted from both the DBIR reports and the aggregate user conversations on Twitter that discuss topics that are linked to several FIPPs (see Table 4). Based on the number of topics extracted from the two datasets, the top 3 major topics the present the topical differences between the two cases and the FIPPs associated with it are shown below.

Dataset	**Top 3 Major Topics Extracted**	
	Ransomware	**Data Breach**
DBIR Dataset	**i)** Attacks on Government (*City of Las Vegas)* and Public Institutions (*Hospitals*) – *Accountability*; **ii)** Ransomware Victims (*Patient Details Exposed*) – *Minimization*; **iii)** Ransomware Criminals (*Exposed List of Organ Donors*) – *Security*.	**i)** Data Breach Causes (*Stolen APIs, Misconfiguration) – Quality*; **ii)** Organizational Errors (*Employee Errors, Rogue Employees*) – *Integrity*; **iii)** Personal Information Exposed (*Medical History, Student Grades, Financial Loans*) – *Security*.
Social media Dataset	**i)** Security Threats (*Linux Systems*) – *Security*; **ii)** Frequent Targets of Ransomware (*Hospitals and Medical Institutions*) – *Minimization*; **iii)** Ransomware Deployment (*Virus, Malware, Email, Files*) – *Security*.	**i)** Cybersecurity Risks (*Patient Data*) – *Security*; **ii)** Personal Details (Stealing Personal Information,: *User accounts and credentials compromised) – Transparency;* **iii)** Rising Cases of Personal Data Breaches (Healthcare data: *quarantine exemption data released*) – *Minimization*.

Table 4. Ransomware and Data Breaches Topics in DBIR and Twitterverse

Through the analysis of our topic modeling results, we can formulate a set of propositions which are based on the results from our topical cluster analysis.

For the first proposition, we focus on the loss from data breaches and ransomware incidents. While practitioners' reports on these incidents describe the different types of information lost during data breaches and ransomware attacks, the public response is concerned with the loss of their personal information, such as user credentials, and its disclosure. A sample description of the public and practitioner response in this regard is presented below following which we present our first proposition.

Sample Public Response (Tweet): "*Bigbasket faces potential data breach; details of 2 crore users likely to have been leaked, put for sale on dark web.*"

Sample Practitioner Response (Report): *Over eight million patients in India had their personal and medical details exposed after security researchers discovered multiple vulnerabilities in a government-run COVID-19 surveillance system...... The research team found two main problems: an unsecured git repository containing code for the platform as well as plain text admin credentials and a separate index of CSV files containing daily COVID-19 patient reports, which was accessible without a password. Personal data exposed included full names, addresses, phone numbers, diagnoses, symptoms and medical records. Even worse, the passwords in the git repository were listed twice, once in easy-to-crack, unsalted MD5 hashes. Most were simply four-digit numbers, often linked to the same code as that of the platform's administrators, the report noted.*

Loss - Proposition P1A: The public response is focused on the loss of personal information that is circulating online (Topics: Personal Information Exposed – User Credentials, Massive Information Disclosures);

P1B: the practitioners' response is focused on the type of information lost (Topics: Personal Information Exposed - Medical, Grades, Financial).

The second set of propositions focus on the legal and criminal aspects of data breaches and ransomware incidents. While the practitioners reports focus on who is responsible for such attacks either ransomware criminals or rogue employees, the public response centers around the existence of laws governing the security of personal information that could deter these cyber-attack incidents. A sample description of the public and practitioner response in this regard is presented below following which we present our second proposition.

Sample Public Response (Tweet): "*#6 add whistleblower protection for violations of California data broker, data breach notification and privacy (#CPRA) laws"*

Sample Practitioner Response (Report): *VT San Antonio Aerospace Inc., which provides maintenance, repair and overhaul services to aircraft, was hit with a ransomware attack affecting its U.S. commercial operations. A criminal group known as Maze gained unauthorized access to our network and deployed a ransomware attack, according to a Friday statement by Ed Onwe, vice president and general manager of the company, which is a subsidiary of the North American headquarters of Singapore's ST Engineering Ltd.*

Laws - P2A: The public response is focused on the laws governing the security of personal information (Topics: Laws and Cybersecurity Risks);

P2B: the practitioners' response is focused on the criminals responsible for the cyber-attacks (Topics: Ransomware Criminals and Rogue Employees).

The third set of propositions are related to the information that is compromised during incidents of data breach and ransomware attacks. Herein, the practitioners' reports generally describe how the data and information is compromised, for example either through stolen APIs or misconfigurations. The public response to such incidents is focused on which type of information is compromised either user credentials or personally identifiable information. A sample description of the public and practitioner response in this regard is given below following which we present our third proposition.

Sample Public Response (Tweet): *“😳😳😳 MASSIVE DUMP! 2659 passwords with emails were just leaked in a public paste: 😳 ‘‰ https://t.co/5h1LSTnlxd #infosec #cybersecurity #gdpr #databreach #security #leak #breach https://t.co/caQEX5bSgL”*

Sample Practitioner Response (Report): *US-based virtual learning platform Playground Sessions’s data leak exposed nearly 4,100 user records through s3 bucket misconfiguration*.

Information Compromise - P3A: The public response is focused on the type of personal information that is compromised (Topics: Personal Information Exposed – User Credentials, Massive Information Disclosures);

P3B: The practitioners’ response is focused on how the information is compromised (Topics: Data Breach Causes - Stolen APIs, Misconfiguration).

The final set of propositions relate to the costs associated with data breach and ransomware incidents. While the practitioners’ reports focus on how organizational or employee errors cost a lot for organizations and public institutions, the public response is focused on the rising costs of data breaches for individuals. A sample description of the public and practitioner response in this regard is given below following which we present our fourth proposition.

Sample Public Response (Tweet): "*A large-scale ransomware attack or data breach could cripple an organization and saddle them millions of dollars in costs. If this were to happen, it wouldn't just be employee bonuses that were threatened, but the jobs of everyone at the organization.*"

Sample Practitioner Response (Report): *About 17GB of data has been exfiltrated from Anglicare Sydney, a Christian not-for-profit that supports people across the greater Sydney and Illawarra regions, to a remote location during a ransomware attack, according to a statement put out by the organisation….. Anglicare Sydney said only its systems had been affected and not those of the government……. Contacted for comment, a NSW Government spokesperson said: "Department of Communities and Justice cyber security staff were quick to act on potential threats posed by the Anglicare cyber-attack. DCJ took immediate protective action to ensure the cyber breach did not impact their systems. "At this point Cyber Security NSW is not aware of any impacts on NSW Government systems or services from the Anglicare cyber-attack. "Cyber Security NSW, together with DCJ, is working closely with Anglicare to assist with their investigation and response to the incident, including engaging with NSW Police."*

Cost - P4A: The public response is focused on the costs of personal information disclosure (Topics: Costs of Data Breaches);

P4B: The practitioners' response is focused on the organizations whose information is compromised (Topics: Attacks on Governments and Public Institutions, Organizational/Employee Errors).

The results from this paper showcase the importance of studying the difference in public reaction to ransomware and data breaches. There are several topics that show a difference in public reaction and opinions about ransomware and personal data breach incidents and thus are perceived differently by people on social media. It is essential to understand how the data breaches and ransomware attacks are discussed in the social media conversations and in practitioner reports because they provide valuable insight for fortifying the defenses of organizations and individuals alike. For example, the topical analysis on the DBIR reports highlights several causes of data breaches such as stolen APIs and misconfigurations. An analysis of the DBIR reports and the tweet discourses can provide cyber threat intelligence that can be used to create actionable plans and policies for securing healthcare and personal data, and protecting against massive data breaches and ransomware attacks.

Further, by analyzing the public reaction between aggressor-oriented discourse and victim-oriented discourse we can segregate how data breaches in general affect both organizations and people alike. This segregation is useful for crisis managers to pacify the public by crafting a response to a data breach crisis that focuses on efforts to protect personal information from further leakage. In GDPR and other privacy regulations it is necessary to disclose events that compromise users' data and privacy (Schmitz-Berndt eta l., 2021) such as these cyber-attacks and our work can guide responders to draft such disclosure in the most effective way.

CONCLUSION

We have identified several information assurance issues being discussed in the social media domain as well as in official cybersecurity reports such as the DBIR. In the ever-increasing landscape of ransomware attacks it is essential to understand the different public responses to such breach incidents. The responsibility of protecting data rests with organizations therefore attempts to fortify their defenses are paramount. In this respect, our work has implications for policymakers and based on the insights from this paper, they can establish better response and recovery efforts in cases of data breach incidents as well as ransomware attacks. With respect to our findings, we provide several propositions that highlight the key differences in the practitioners view and the public response to data breaches and ransomware. For example, the laws and costs associated with such incidents are points of distinction between the practitioners' view and the public response.

Our work also furthers the understanding of cyber-attacks such as ransomware and data breaches. We see that there are considerable differences between the two where people discuss diverging viewpoints based on who is to blame for data breaches and ransomware attacks, the organizations or cyber attackers. Understanding such differences can help spur research in automatic methods of data protections using AI based solutions.

In our future work we will add a dynamic range of emotion analysis for evaluating the differences between sentiments across the two cases. We will also analyze the several data breach and ransomware incidents over a longer time period to understand the temporal differences between perceptions of ransomware attacks and data breaches. Also, we will focus on validating emotion-based analysis with actual data collected for a specific data security breach incident such as the Solarwinds vulnerability found last year.

ACKNOWLEDGMENTS

The authors thank the review team for their comments that have helped to greatly improve the paper.

REFERENCES

Ayaburi, E. W., & Treku, D. N. (2020). Effect of penitence on social media trust and privacy concerns: The case of Facebook. International Journal of Information Management, 50, 171-181.

Azeez, N. A., & Van der Vyver, C. (2019). Security and privacy issues in e-health cloud-based system: A comprehensive content analysis. Egyptian Informatics Journal, 20(2), 97-108.

Bachura E., Valecha, R., Chen R., and Rao, H. R. (2022). The OPM Data Breach: An Investigation of Shared Emotional Reactions on Twitter. MIS Quarterly Forthcoming.

Bassett, G., Hylender, C. D., Langlois, P., Pinto, A., & Widup, S. (2021). Data breach investigations report. Verizon DBIR Team, Tech. Rep.

Bentley, J. M., Oostman, K. R., & Shah, S. F. A. (2018). We're sorry but it's not our fault: Organizational apologies in ambiguous crisis situations. Journal of Contingencies and Crisis Management, 26(1), 138-149.

Bhatt, P., Vemprala, N., Valecha, R., Hariharan, G., & Rao, H. R. (2022). User Privacy, Surveillance and Public Health during COVID-19–An Examination of Twitterverse. Information Systems Frontiers, 1-16.

Branch, L., Eller, W., Bias, T., McCawley, M., Myers, D., Gerber, B., & Bassler, J. (2019). Trends in malware attacks against United States healthcare organizations, 2016-2017. Global Biosecurity, 1(1).

Brewczyńska, M., Dunn, S., & Elijahu, A. (2019). Data privacy laws response to ransomware attacks: A multi-jurisdictional analysis. In Regulating New Technologies in Uncertain Times (pp. 281-305). TMC Asser Press, The Hague.

Brooks, S., Brooks, S., Garcia, M., Lefkovitz, N., Lightman, S., & Nadeau, E. (2017). An introduction to privacy engineering and risk management in federal systems (pp. 1-49). MD, USA: US Department of Commerce, National Institute of Standards and Technology.

De Simone, D. M. (2019). Data Breaches Are Not Just Information Technology Worries!. Pediatric Nursing, 45(2).

Faghihi, F., & Zulkernine, M. (2021). RansomCare: Data-centric detection and mitigation against smartphone crypto-ransomware. Computer Networks, 191, 108011.

Gkikas, D. C., Tzafilkou, K., Theodoridis, P. K., Garmpis, A., & Gkikas, M. C. (2022). How do text characteristics impact user engagement in social media posts: Modeling content readability, length, and hashtags number in Facebook. International Journal of Information Management Data Insights, 2(1), 100067.

Hammouchi, H., Cherqi, O., Mezzour, G., Ghogho, M., & El Koutbi, M. (2019). Digging deeper into data breaches: An exploratory data analysis of hacking breaches over time. Procedia Computer Science, 151, 1004-1009.

Janakiraman, R., Lim, J. H., & Rishika, R. (2018). The effect of a data breach announcement on customer behavior: Evidence from a multichannel retailer. Journal of Marketing, 82(2), 85-105.

Kharraz, A., Robertson, W., Balzarotti, D., Bilge, L., & Kirda, E. (2015, July). Cutting the gordian knot: A look under the hood of ransomware attacks. In International conference on detection of intrusions and malware, and vulnerability assessment (pp. 3-24). Springer, Cham.

Kozlowska, I. (2018). Facebook and data privacy in the age of Cambridge Analytica. The Henry M. Jackson School of International Studies, Seattle, 30.

Mamonov, S., & Benbunan-Fich, R. (2018). The impact of information security threat awareness on privacy-protective behaviors. Computers in Human Behavior, 83, 32-44.

Moody, G. D., Siponen, M., & Pahnila, S. (2018). Toward a unified model of information security policy compliance. MIS quarterly, 42(1).

Nadir, I., & Bakhshi, T. (2018, March). Contemporary cybercrime: A taxonomy of ransomware threats & mitigation techniques. In 2018 International Conference on Computing, Mathematics and Engineering Technologies (iCoMET) (pp. 1-7). IEEE.

Nikkhah, H. R., and Grover, V. "An Empirical Investigation of Company Response to Data Breaches," MIS Quarterly, forthcoming.

Popoola, S. I., Iyekekpolo, U. B., Ojewande, S. O., Sweetwilliams, F. O., John, S. N., & Atayero, A. A. (2017). Ransomware: Current trend, challenges, and research directions. In Proceedings of the World Congress on Engineering and Computer Science (Vol. 1, pp. 169-174).

Spence, N., Niharika Bhardwaj, M. B. B. S., & Paul III, D. P. (2018). Ransomware in healthcare facilities: a harbinger of the future?. Perspectives in Health Information Management, 1-22.

Syed, R. (2019). Enterprise reputation threats on social media: A case of data breach framing. The Journal of Strategic Information Systems, 28(3), 257-274.