information

MDPI

Article

Analysis of Insider Threats in the Healthcare Industry: A Text Mining Approach

In Lee

School of Computer Sciences, Western Illinois University, Macomb, IL 61455, USA; i-lee@wiu.edu

Abstract: To address rapidly growing data breach incidents effectively, healthcare providers need to identify various insider and outsider threats, analyze the vulnerabilities of their internal security systems, and develop more appropriate data security measures against the threats. While there have been studies on trends of data breach incidents, there is a lack of research on the analysis of descriptive contents posted on the data breach reporting website of the U.S. Department of Health and Human Services (HHS) Office for Civil Rights (OCR). Hence, this study develops a novel approach to the analysis of descriptive data breach information with the use of text mining and visualization. Insider threats, vulnerabilities, breach incidents, impacts, and responses to the breaches are analyzed for three data breach types.

Keywords: insider threats; PHI; protected health information; data breach; healthcare provider; text mining; visualization; vulnerabilities; cybersecurity

1. Introduction

As the healthcare industry grows, the protection of private patient data has become a significant challenge [1]. Under HIPAA, any patient-identifying and personal medical information, such as demographic data, medical histories, and insurance information, is considered protected health information (PHI) [2]. As digital technologies advance, the healthcare service industry has been converting paper-based protected health information to computer-based electronic protected health information (ePHI). A host of healthcare information systems are currently using ePHI to store patients' data, treat patients, file insurance claims, and conduct many other operational functions. However, along with the use of the Internet for healthcare information systems, there has been an increase in high-profile data breach incidents, and the development of evidence-based data security practices became a research priority [3].

Data breaches emanate from insider and outsider threats. Although many breach incidents are caused by outsiders, the most damaging incidents are often caused by insiders [4]. A recent report suggests that over half of healthcare breaches come from insiders of the organization [5]. Data breaches by insiders are often difficult to detect. However, negligent and malicious employees are of great risk as witnessed in the case of the Capital One—Amazon cloud data breach [6]. According to a study conducted by Shred-it [7], more than 85% of senior executives and 515 small business owners admit employee negligence is one of their most serious information security threats.

Previous studies on the data breaches in the healthcare industry mainly focused on analyzing well-structured breach data such as hospital types, breach types, and breach locations [8–12]. Recently, text mining approaches have been used in the healthcare domain to automate the process of gleaning insights from unstructured textual data [13]. The Latent Dirichlet Allocation (LDA) has been used for topic modeling in the healthcare domain [14,15]. Long Short-Term Memory (LSTM) and various classification techniques such as random forest, support vector machine, and logistic regression have been used to analyze the sentiment using a Twitter dataset associated with coronavirus [15–17]. While

check for updates

Citation: Lee, I. Analysis of Insider Threats in the Healthcare Industry: A Text Mining Approach. *Information* **2022**, *13*, 404. https://doi.org/10.3390/info13090404

Academic Editor: Willy Susilo

Received: 21 July 2022
Accepted: 24 August 2022
Published: 27 August 2022

the multi-year data breach report publicly available from the U.S. Department of Health and Human Services (HHS) Office for Civil Rights (OCR) includes summary descriptions of the data breach incidents and responses to the incidents, there has been no attempt to systematically analyze the descriptions and the evolution of the insider threats in the healthcare industry. Analyzing thousands of data breach descriptions would be time-consuming and costly, but may provide healthcare providers with useful information for the enhancement of data security.

To overcome the challenges of analyzing the descriptions and the insider threats manually, this study adopts a text mining approach. This study demonstrates that the text mining approach facilitates analysis of the textual description of the data breaches and the results shed valuable insights into the insider threats, data breaches, and organizational responses. Specifically, this study attempts to address the following research questions:

RQ1: What are the characteristics of insider threats in data breach incidents?

RQ2: How have insider threats, vulnerabilities, data breach incidents, impacts/losses, and responses to the data breaches evolved?

To answer these questions, this study utilizes the data breach incidents reported at the U.S. Department of Health and Human Services (HHS) Office for Civil Rights (OCR) website (https://ocrportal.hhs.gov/ocr/breach/breach_report.jsf, accessed on 5 January 2022)

From a methodological standpoint, this study is the first of its kind to analyze the unstructured textual descriptions contained in the breach incident report in the healthcare industry. In Section 2, past studies are reviewed and research gaps are identified. In Section 3, the research methods of this study are discussed. In Section 4, insider threats in data breach incidents of three data breach types are analyzed. In Section 5, a summary of the findings, practical implications, and future research directions are discussed.

2. Related Works

According to the CERT Guide to Insider Threats [18], insider threats is defined as "A malicious insider threat is a current or former employee, contractor, or business partner who has or had authorized access to an organization's network, system, or data and intentionally exceeded or misused that access in a manner that negatively affected the confidentiality, integrity, or availability of the organization's information or information systems." In addition to the malicious insider threats, Unintentional Insider Threats: A Foundational Study [18] defines unintentional insider threat as "An unintentional insider threat is a current or former employee, contractor, or business partner who has or had authorized access to an organization's network, system, or data and who, through action or inaction without malicious intent, causes harm or substantially increases the probability of future serious harm to the confidentiality, integrity, or availability of the organization's information or information systems."

The origin of the data breach threats can be classified into an outsider [19–21], an insider [22–25], and a mix of an insider and an outsider [26–28]. The mix of insiders and outsiders takes place when the threats come from the outside, but the data breaches are realized by the actions of insiders (e.g., phishing email). Insider threats are further classified into accidental errors, ignorance, unintentional non-malicious (negligent) threats, and malicious threats [29]. The threats from outsiders and insiders have the potential to adversely impact operations, organizational assets, or individuals via unauthorized access, destruction, disclosure, modification of information, and/or denial of service (https://csrc.nist.gov/glossary/term/threat, accessed on 29 April 2022). However, insider threats have been easily overlooked or ignored due to various reasons such as ease of implementation, high chance of success, inaccurate solutions, and less chance to detect and prevent [30]. The insiders may have legitimate access to their organization data, but can misuse their credentials to conduct malicious attacks such as IT sabotages and theft of confidential information [30].

Detecting insider threats is challenging. Three categories of data sources are considered to detect insider attacks [31]: (1) host, (2) network, and (3) contextual. Host-based data sources are individual hosts (e.g., computers), and can reflect how a host behaves and the human user's interactive behavior with the host [31]. Network analytics are playing an increasingly important role in addressing cybersecurity threats. For example, [32] presents Beehive, which addresses the problem of automatically mining and extracting knowledge from the dirty log data produced by a wide variety of security products in a large enterprise [32]. Finally, contextual data sources provide contextual information such as human resources (HR) and psychological data for various intent analyses [31].

To minimize the threats from insiders and outsiders, organizations need to address vulnerabilities to data security [33]. Vulnerabilities are weaknesses in an information system, system security procedures, internal controls, or implementation that could be exploited or triggered by a threat source (https://csrc.nist.gov/glossary/term/vulnerability, accessed on 5 May 2022). Vulnerabilities can be manifested in the form of human, organizational, physical, and technical vulnerabilities [34,35].

In many organizations, the human aspect of cybersecurity is one of the weakest links [36]. An empirical study shows that one of the reasons that malicious attacks continue to occur at an alarming rate in IoT systems is the poor compliance with information security policies that are mainly caused by behavior issues and the severe lack of security awareness [37]. Another empirical study shows there is a low level of IoT cybersecurity awareness by working adults on the IoT and proposes a cybersecurity e-brochure playbook as a possible solution [38]. It is also necessary to develop people-centric workplaces where desirable security behaviors are disseminated among the employees [39].

Human factors contribute to the majority of security violations in healthcare [40]. Based on responses to an online questionnaire survey of 212 healthcare staff, [40] assess various factors that affect security and privacy behavior among healthcare workers. Their study revealed that work emergency and conscientiousness have a positive correlation with IS conscious care behavior risk. But agreeableness is negatively correlated with information security knowledge risk and information security attitude risk. The authors of [41] explore how human factors affect data locations via linear regression and rank data location vulnerability using collaborative filtering. They find that human factors play a major role in data location breaches.

The organization aspect of data security focuses on senior management support [42], security readiness [43], security policy [44], information security governance [45], and security culture [46,47]. These organizational factors are known to positively affect the attitudes and behaviors of internal users and employees as well as external users and customers. Organizational readiness to handle cyberattacks has become an integral part of enterprise risk management [48]. Sustained support from senior management is crucial to ensuring that action plans are in place to mitigate the risk of cyberattacks [36].

Technical vulnerabilities refer to security weaknesses in software, hardware, networks, or a system. Technical vulnerabilities arise due to flaws or the incorrect design of software and the limitations of the device or application [49]. Due to the evolving nature of the technical vulnerabilities, it is hard to provide an exhaustive list of vulnerabilities [50]. A vulnerability scanning is a process and technical measure implemented for the timely detection of vulnerabilities within the organization, infrastructure network, and system components [51]. Noting that disseminating medical data beyond the protected cloud of institutions poses severe risks to patients' privacy, [52] propose a blockchain-based data sharing framework that addresses the access control challenges and permits users to request sensitive patients' data stored in the cloud after their identities and cryptographic keys are verified.

3. Research Methodology

This study introduces a five-step methodology for data breach analysis. The uniqueness of our methodology is that the text mining is conducted in two steps using two widely used and validated text mining tools: VOSviewer (Developer: Centre for Science and Technology Studies (CWTS), Leiden University, EZ Leiden, The Netherlands) and NVivo

(Developer: QSR International, Doncaster, Australia). First, objective keyword identification is conducted with VOSviewer to increase the accuracy of the keyword selection and the selected keywords and clusters of the keywords are visualized to understand the keyword relationships. Second, the selected keywords are searched in the web descriptions using NVivo to analyze the context surrounding the keywords. By combining VOSviewer and NVivo, this methodology could achieve a complexity reduction in keyword identification and contextual analysis without losing the accuracy of the analysis. The research methodology consists of the following five main steps, starting from collecting the data to analyzing the descriptions of the data breach incidents.

3.1. Step1: Data Collection

Data collection is the first step of the methodology. To analyze the evolution of the breach incidents, this paper collected the breach incidents from the HHS website (Source: https://www.hhs.gov/hipaa/for-professionals/breach-notification/breach-reporting/index.html, accessed on 5 January 2022). The public report is in CSV file format and can be simply downloaded without web scraping. For this study, the area of inquiry is a web description of the data breaches in healthcare. The data from 21 October 2009 to 15 October 2010 were dropped from the analysis to take into account the lag time to the full effect of the 2009 HIPAA Breach Notification Rule [53]. The HHS website shares the data breaches collected according to the HIPAA Breach Notification Rule, since October 2009. HIPAA Breach Notification Rule requires that HIPAA-covered entities and their business associates in the US that store data on human health report any data breach that compromises the confidentiality of 500 or more patients/human subjects no later than 60 days following a data breach. Each data breach incident consists of the name of the covered entity, state, covered entity type, individuals affected, breach submission date, type of breach, location of breached information, business associate present, and web description. The data breach report consists of 2003 data breaches from the healthcare provider sector.

3.2. Step 2: Data Pre-Processing and Cleaning

The next phase of our methodology includes data cleaning and filtering out breach incidents that do not contain the breach description. The data breach report in the original CSV file format was converted into a Microsoft Excel file to sort the records based on the contents of the web description and remove records without description. This phase identified a total of 1136 data breach incidents that contain data breach descriptions.

3.3. Step 3: Segmentation of Data

The third phase of the research methodology involves the segmentation of data for detailed analysis. The data breach types include Hacking/IT Incident, Improper Disposal, Loss, Other, Theft, and Unauthorized Access/Disclosure. Among the six data breach types, this study focuses on Hacking/IT Incident, Theft, and Unauthorized Access/Disclosure. The other three types, Improper Disposal, Loss, and Other, were not included due to the limited number of data breach incidents. To further segment the data, temporal aggregations of the data breaches into three years were made. The three years were chosen to see the trends of the insider threats and the description of data breaches over time and the periodic comparison of data breaches. Period 1 is between 16 October 2010 and 15 October 2013, and contains 231 data breach incidents with descriptions. Period 2 is between 16 October 2013 and 15 October 2016, and contains 457 data breach incidents with descriptions. Period 3 is between 16 October 2016 and 15 October 2019, and contains 458 data breach incidents with descriptions.

3.4. Step 4: Measurement of Keyword Co-Occurrences and Clustering of the Keywords

This step starts with a visualization of the web descriptions related to data breach incidents. VOSviewer is used to provide a visualized overview of the keywords identified in the web descriptions. VOSviewer measures keyword co-occurrences and clustering of the keywords [54]. VOSviewer is used to measure the co-occurrence of keywords among the descriptions of the data breach incidents. The operationalization of VOSviewer for this study is as follows. The first step of VOSviewer is to choose the option "Create map based on text data". Then, "Read data from VOSviewer files" is chosen and the CSV file is loaded. Next, binary counting is chosen as the counting method to count each keyword only once in each data breach incident. The minimum threshold was set at three occurrences so that a large number of keywords would be considered for selection. Keywords with a high relevance score tend to represent specific topics covered by the data breach incidents. To focus on more relevant keywords, 60% of the keyword are included based on their relevance score. The minimum number of co-occurrences of keywords is set to obtain at least 20 keywords for analysis. The generated keywords are verified keyword by keyword for relevancy and accuracy. Finally, the network visualization is created along with clusters of keywords. Clustering is performed to group the keywords based on the keyword co-occurrences. The cluster size is set at a minimum of five keywords per cluster. Clustered groups are color-coded for clear visualization. Using the zooming and scrolling functions of VOSviewer, the clusters and links of keywords are examined to understand the keyword relationships. The keyword co-occurrences and clustering of the keywords provide the basis for the contextual analysis of the breach incidents in Step 5.

3.5. Step 5: Analysis of Descriptions of Data Breaches

This step utilizes NVivo for contextual analysis of the descriptions of the data breaches with the list of keywords and clusters created in Step 4. The web descriptions were imported into NVivo to conduct keyword searches of data breach incidents in the web descriptions. The list of keywords identified in VOSviewer is used as an initial coding template. The contextual analysis is conducted for the context in which the keywords occur, and new keywords are added for analysis. For example, assume that in the previous phase, "PHI" and "employee" were clustered as co-occurring keywords. Since these keywords alone do not provide insightful information about the insider threats, vulnerabilities, data breaches, and impacts of data breaches, and the healthcare providers' responses, the descriptions of the data breach incidents related to the co-occurring two keywords are identified and analyzed further. NVivo is used to search through all data breach incidents and identify data breach incidents whose descriptions match "PHI" and "employee". The descriptions of the identified data breach incidents are analyzed in terms of the insider threats, vulnerabilities, data breaches, impacts of data breach incidents, and the healthcare providers' responses.

A table for the insider threats, vulnerabilities, data breaches, impacts of data breach incidents, and the healthcare providers' responses was created. The new keywords of the data breaches are coded in terms of the insider threats, vulnerabilities, data breaches, impacts of data breaches, and responses to the incidents. The duplicates (or similar words) of keywords identified in this step are removed from the table to create a non-redundant list of the keywords.

4. Analysis of Data Breach Types

This section measures keyword co-occurrences and clustering of the keywords and analyzes the description of data breach types in terms of the insider threats, vulnerabilities, data breaches, impacts of data breaches, and responses to the incidents. VOSviewer was used to analyze co-occurring keywords and clusters of the keywords relevant to three data breach types—hacking/IT incident, theft, and unauthorized access/disclosure. Then, NVivo was used to identify data breach incidents and the insider threats related to the co-occurring keywords in each cluster.

4.1. Analysis of Hacking/IT Incident

Table 1 shows that 25 co-occurring keywords are grouped into three clusters in Period 1, 25 co-occurring keywords into three clusters in Period 2, and 27 co-occurring keywords into three clusters in Period 3. The grouping of the co-occurring keywords into specific clusters does not mean certain keywords co-occur only in a certain cluster but indicates that the keywords co-occur more frequently within that cluster than in other clusters. It is noted that insider threats are identified in seven clusters among nine clusters. For example, staff was identified as an insider threat of Cluster 1 of Period 1, and the employee was identified as an insider threat of Cluster 2 of Period 1. Business associates in Cluster 3 of Period 2 and Cluster 2 of Period 3 are considered an insider since they are contractors who have authorized access to the hospital's network, system, or data. In addition to identifying the insider threats in each cluster, other related characteristics of data breach incidents were identified in each cluster. For example, 'risk analysis' became an important cybersecurity activity in Period 2 and Period 3. In Period 1 and Period 2, 'malware' became a serious data breach and triggered an investigation by OCR, and in Period 3 'ransomware' and 'ransomware attack' became a serious data breach and co-occurred with 'computer server' and 'server'. 'Malware' was not a cybersecurity issue for 'computer server' and 'server', but a data breach issue for personal computing devices such as laptops and desktops. These keywords co-occur with specific insider threats, providing ample ground for further analysis.

Table 1. Clusters of Co-occurring Keywords in Hacking/IT Incident in Three Periods.

Period 1	**Period 2**	**Period 3**
25 items (3 clusters)	25 items (3 clusters)	27 items (3 clusters)
Cluster 1 (11 items) access breach notification corrective action plan ephi firewall hacker health information office policy server staff	**Cluster 1 (10 items)** address assurance birth breach notification corrective action employee name phi protected health information social security number	**Cluster 1 (13 items)** access assurance clinical information computer server corrective action ephi health information policy ransomware ransomware attack risk analysis safeguard server
Cluster 2 (7 items) address assurance birth corrective action employee name social security number	**Cluster 2 (9 items)** clinical information ephi health information malware ocrs investigation policy risk analysis staff technical assistance	**Cluster 2 (8 items)** breach notification business associate employee ocrs investigation phi protected health information technical assistance technical safeguard
Cluster 3 (7 items) malware medication ocrs investigation phi protected health information system treatment information	**Cluster 3 (6 items)** business associate computer server diagnosis server treatment information unauthorized access	**Cluster 3 (6 items)** address birth name social security number unauthorized user workforce member

Figure 1 shows the visualization of the keyword co-occurrence and clustering of the keywords in Hacking/IT Incident in Period 1. Each cluster is in a different color. The occurrences of a keyword refer to the number of data breach incidents in which the keyword occurs [49]. In Figure 1, keywords with higher occurrences are shown more prominently with larger labels and circles than keywords with lower occurrences. The stronger the link between two keywords, the thicker the line is in the visualization map. For example, In Figure 1, it is easy to see that 'name' occurs most frequently and its label is the largest in Period 1. Figure 2 shows that 'employee' of Cluster 2 in Period 1 has various link strengths with other keywords in its cluster and other clusters. It is noted that most of the links are with the keywords in its cluster and only a few links are with keywords in other clusters and the links are stronger in its cluster than in other clusters.

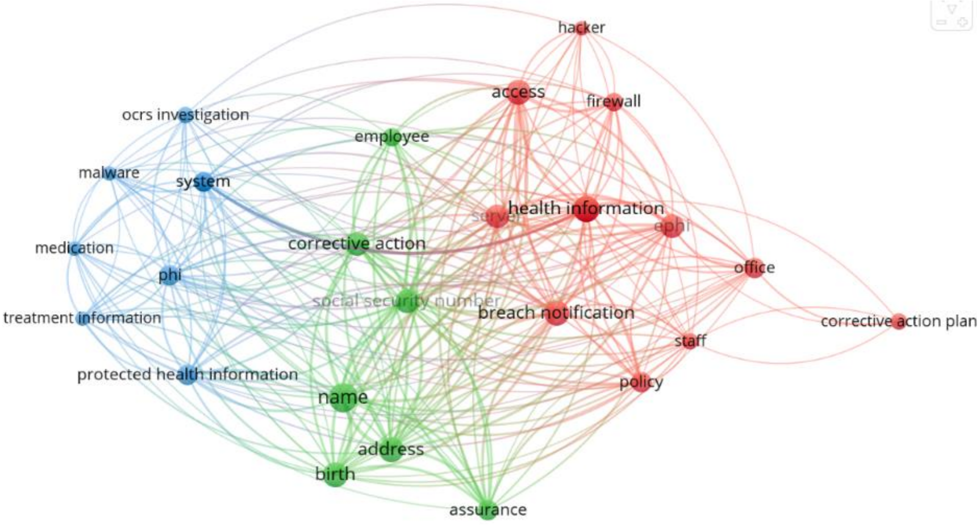

Figure 1. Keyword co-occurrences of Hacking/IT Incident in Period 1.

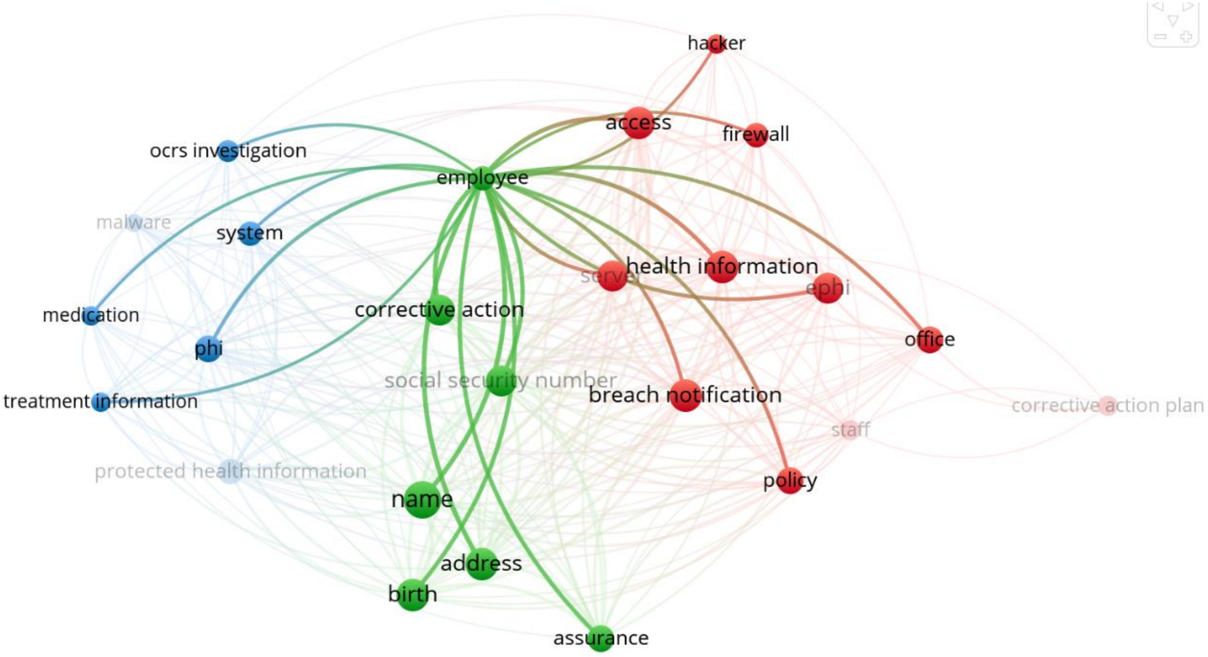

Figure 2. Various Link Strengths of 'Employee'.

To analyze the insider threats and characteristics of data breaches in detail, NVivo was used to identify descriptions related to keywords in each cluster and analyze the descriptions in terms of insider/outsider threats, vulnerabilities, data breaches, impacts of the breaches, and responses to the breaches.

In Table 2, the third column shows insider/outsider threats. Insider threats include staff, employees, workforce members, and business associates. Outsider threats include hackers, unknown outsiders, and ransomware attackers. Human vulnerabilities include a lack of social engineering awareness and a lack of security awareness of employees. Technical vulnerabilities include a lack of firewall, unencrypted software, weak security control, lack of technical security measures, weak safeguards for a system, and open firewall ports. Organizational vulnerabilities include a lack of a security management process and malicious partners. Specific data breaches include email hacking, phishing, malware, unauthorized access to the system, and ransomware. The sixth column shows the impacts of the data breaches. The impacts of the data breaches include the loss and stealing of health information, treatment information, medication information, and demographic information.

Table 2. Insider/Outsider Threats, Vulnerabilities, Breach Incidents, Impacts, and Responses in Hacking/IT Incident.

Hacking/IT Incident	Cluster	Insider/Outsider Threats	Vulnerabilities	Breach Incidents	Impacts	Responses
Period 1	Cluster 1	hacker; staff	lack of firewall; weak security control	hacking	ePHI; health information	a settlement with the OCR; corrective action plan

	Cluster 2	employee	lack of a security management process	email phishing	date of birth; name; social security number	risk analysis;
						corrective action;
						a settlement with OCR

	Cluster 3	unknown outsider	unencrypted software; lack of technical security measures	malware	medication; PHI; treatment information	OCR's investigation; security software to detect, prevent, and mitigate malware on computers
Period 2	Cluster 1	employee	lack of social engineering awareness	email phishing	address; date of birth; name; PHI; social security number	assurance of corrective action; software to scan Internet addresses in employees' emails
	Cluster 2	staff	weak safeguards for a system	malware	ePHI; health information	technical assistance by OCR; risk analysis; OCR's investigation; upgraded antivirus software
	Cluster 3	business associate	weak security control	unauthorized access to the system	diagnosis; treatment information	identity recovery services; stricter password policies; the installation of an active traffic-monitoring solution for its network

Period 3	Cluster 1	ransomware attacker	weak firewall system;
			open firewall port
				ransomware	clinical information;
					ePHI
					health information
						risk analysis;
						replacement of firewall;
						anti-malware software;
						assurance of corrective action;
						revised policy and procedures

	Cluster 2	business associate; employee	malicious partners (e.g., vendors); lack of security awareness of employee	email phishing, impermissible access to PHI	PHI	OCR's investigation; technical assistance by OCR; technical safeguard to PHI
	Cluster 3	workforce member	lack of social engineering awareness	email phishing	address; date of birth; name; social security number	improved safeguards; updated policies and procedures; training of its workforce members on better practices to safeguard PHI

The last column shows the responses to the data breach incidents. Responses to insider threats are also categorized into the human, technical, and organizational responses. Human responses include training its workforce members on better practices to safeguard PHI. Technical responses include deployment of security software to detect, prevent, and mitigate malware on computers, software to scan Internet addresses in employees' emails, technical assistance by OCR, upgrade of antivirus software, the installation of a network traffic monitoring solution, upgrade of firewall, anti-malware software, and technical safeguard to PHI. Organizational responses include stricter password policies, a corrective action plan, a settlement with the OCR, risk analysis, identity recovery services, and revised policy and procedures.

In summary, it is noted email phishing has been a persistent and prominent problem for the insiders throughout the three periods and attacks occur through organizational and human vulnerabilities of the insiders such as a lack of a security management process and malicious vendors, a lack of social engineering awareness, and a lack of security awareness of employees. Email phishing leads to the loss or stealing of demographic information such as an address, date of birth, name, and social security number as well as other protected health information. Malware and unauthorized access to the system by the business associates and staff occur through technical vulnerabilities such as weak safeguards for a system and weak security control. It is also noted that while technical vulnerabilities require proper technical responses, human vulnerabilities require more comprehensive responses including technical, organizational, and human responses to effectively manage data security.

It is noted that the responses in the last column in Table 2 are supported by the previous validated results such as training to heighten awareness and reduce human error, the usability of software and tools to reduce human error, management practices to reduce the likelihood of human error, email safeguards (anti-phishing, anti-malware), firewalls, and antivirus/anti-malware protection equipment [55].

4.2. Analysis of Theft

Table 3 shows that 26 co-occurring keywords are grouped into three clusters in Period 1, 25 keywords into two clusters in Period 2, and 24 keywords into three clusters in Period 3. It is noted that insiders ('employee,' and 'workforce member'), 'computer,' 'laptop,' and 'laptop computer' are frequently co-occurring keywords in all three periods. In comparison to the keywords of Hacking/IT Incident, 'computer,' 'laptop,' 'laptop computer,' 'encryption,' 'physical security,' and 'unencrypted laptop computer' are keywords uniquely occurring in Theft.

Table 3. Clusters of Co-occurring Keywords in Theft in Three Periods.

Period 1	**Period 2**	**Period 3**
26 items (3 clusters)	25 items (2 clusters)	24 items (3 clusters)
Cluster 1 (12 items) computer encryption ephi health information laptop laptop computer physical security police report policy risk analysis safeguard workforce member	**Cluster 1 (13 items)** address assurance birth breach notification corrective action employee laptop computer name phi protected health information social security number staff substitute notice	**Cluster 1 (12 items)** assurance breach notification clinical information computer corrective action employee ocrs investigation phi physical security policy protected health information substitute notice
Cluster 2 (8 items) assurance breach notification corrective action employee phi privacy protected health information staff	**Cluster 2 (12 items)** computer diagnosis ephi health information laptop ocrs investigation password policy safeguard technical assistance unencrypted laptop workforce member	**Cluster 2 (9 items)** address birth ephi health information laptop laptop computer name social security number workforce member
Cluster 3 (6 items) address birth diagnosis name ocrs investigation social security number		**Cluster 3 (3 items)** diagnosis medical record number staff

Table 4 shows the main insider/outsider threats, vulnerabilities, breach incidents, impacts of data breaches, and responses to the theft incidents. For theft, insider threats include staff, employees, and workforce members. Insider threats to Theft occur in all the clusters in the three periods. Human vulnerabilities include negligent employees and malicious former employees. The impacts of the data breaches include loss and stealth of demographical information and PHI. Technical vulnerabilities include an unencrypted laptop computer, unsecured disposal of PHI, unencrypted ePHI, unencrypted portable computer drive, unencrypted external hard drive, and unencrypted desktop computer. Organizational vulnerabilities include deficiencies in the HIPAA compliance program and weak physical security.

Table 4. Insider/Outsider Threats, Vulnerabilities, Breach Incidents, Impacts, and Responses in Theft.

Theft	Cluster	Insider/Outsider Threats	Vulnerabilities	Breach Incidents	Impacts	Responses
Period 1	Cluster 1	workforce member	deficiencies in HIPAA compliance program; unencrypted laptop computer; weak physical security; unencrypted ePHI; negligence of employees	theft of laptop and computer	health information; ePHI	The settlement with OCR; encryption of laptop computers; ongoing security awareness training for all staff; enhanced physical security; risk analysis; comprehensive compliance program
	Cluster 2	employees; staff	negligence of employees	theft of laptop and computer	PHI	retraining of all staff on privacy and security policies and procedures; remote access policy; electronic data backup policy
	Cluster 3	employees	unencrypted portable computer drive; unencrypted desktop computer; malicious employees; unsecured disposal of PHI	theft of medical files	address; date of birth; name; social security number; diagnosis	encryption-capable USB drives; securely locked storage facilities for mobile devices; policies preventing the removal of devices from the office
Period 2	Cluster 1	employee; staff	unencrypted laptop; unencrypted ePHI	theft of laptop	address; date of birth; name; social security number; PHI	assurance of corrective action; encryption of all unencrypted electronic devices; update of the policy on safeguarding ePHI
	Cluster 2	workforce member	unencrypted laptop; unencrypted desktop computer; backup computer hard drive	theft of laptop and desktop	diagnosis; health information	technical assistance by OCR; procedures for safeguarding mobile devices; retraining of the employee on the physical security of laptops; retraining of relevant IT personnel on standard encryption configuration processes; update of password policy; revision of HIPAA policies and procedures
Period 3	Cluster 1	employee	unencrypted desktop computer; external computer hard drives	theft of desktop and hard drive	clinical information; PHI	encrypted workstations and computers; enhanced network security; encrypting data at rest on computers; physical safeguards such as surveillance cameras and locks to deter and prevent unauthorized access; complimentary credit monitoring and identity theft protection services
	Cluster 2	workforce member	unencrypted hard drive of a laptop; unencrypted external hard drive; negligence of workforce member	theft of hard drive	address; date of birth; name; social security number; ePHI; health information	sanction of its workforce member; update of security rule policy; encryption software on all laptops and media storage devices; retraining of workforce members; encryption of a laptop; a cloud-based electronic health record system; risk analysis
	Cluster 3	staff	negligence of employee; unencrypted laptop	theft of laptops	diagnosis; medical record number	notifying local law enforcement of the breach; retraining staff; blocking the laptop from accessing the internal computer network

Human responses include ongoing security awareness training for all staff, retraining of all staff on privacy and security policies and procedures, and sanction of its workforce member. Technical responses include encryption of laptop computers, use of encryption-capable USB drives, encryption of electronic devices, enhanced network security, encrypting data on computers, installation of encryption software on all laptops and media storage devices, a cloud-based electronic health record system, and blocking laptops from accessing the internal computer network. Organizational responses include enhanced physical security, risk analysis, comprehensive compliance program, securely locked storage facilities for mobile devices, policies preventing the removal of devices from the office, assurance of corrective action, update of the policy on safeguarding ePHI, revision of HIPAA policies and procedures, physical safeguards such as surveillance cameras and locks to deter and prevent unauthorized access, and notifying local law enforcement of the breach.

In summary, it is noted that the insiders such as staff, employees, and workforce members are the predominant threats to theft. The majority of the insider threat comes from the negligence of the employees, followed by malicious employees. Thefts of medical files, laptops, computers, and devices occur through technical, organizational, and human vulnerabilities such as unencrypted PHI files, unencrypted laptop computers and devices, weak physical security, and negligence of employees. The various thefts lead to the loss or stealth of demographic information and other protected health information. While the frequency of thefts declined from Period 1 to Period 3, the theft of laptops and computers is a major concern in the three periods. Theft arising from the negligence of employees requires organizational and human responses such as retraining of all staff on privacy and security policies and procedures and development of remote access policy; rather than technical responses. For the Theft arising from malicious employees, organizational responses include securely locked storage facilities for mobile devices and the development of policies preventing the removal of devices from the office. While business associates are insider threats in Hacking/IT Incident, they are not a major part of insider threats in Theft.

4.3. Analysis of Unauthorized Access/Disclosure

Unauthorized Access/Disclosure is the third most frequently occurring data breach type.

Table 5 shows that 22 co-occurring keywords are grouped into three clusters in Period 1, 24 keywords into three clusters in Period 2, and 24 keywords into three clusters in Period 3. It is noted that 'staff', 'employee', and 'workforce member' are major co-occurring data breach incidents in all three periods. 'Business associate' started to co-occur in Period 2 and Period 3. 'Email' and 'Email address' started to co-occur in Period 2 and Period 3 and are a major source of unauthorized access/disclosure.

Table 6 shows the main insider/outsider threats, main vulnerabilities, breach incidents, impacts of data breaches, and responses to Unauthorized Access/Disclosure. Insider threats occur in All clusters in the three periods. No outsider threats are identified. Major insider threats include staff, employees, and workforce members. Minor insider threats are business associates. Human vulnerabilities include malicious employees, malicious former employees, errors of the staff, negligent employees, malicious business associates, disgruntled former business associates, and errors of business associates. Technical vulnerabilities include insecure websites and security holes in applications. Organizational vulnerabilities include weak access termination protocol and insecure records room. The impacts of the data breaches include unauthorized or suspicious access to ePHI, and accidental disclosure of medical files and email addresses.

Human responses include termination of the offending employee, retraining of the workforce, employee sanctions, and criminal charges. Technical responses include improvement of operation software, a program to track anomalies to detect inappropriate use or access, a new workflow in mailing processes to reduce the number of manual steps, password protection for electronic files, a secure online portal, two-factor email authentication, and encryption and tools to monitor Internet traffic and compliance. Organizational responses include assurance of corrective action, updated access termination protocol, improvement of HIPAA training materials, risk analysis procedure, revision of email policies, restricted workforce access to the patient folder, increased restrictions to access to PHI, and measures to improve internal security and limit employee access to records rooms.

Table 5. Clusters of Keyword co-occurrences of Unauthorized Access/Disclosure in Three Periods.

Period 1	**Period 2**	**Period 3**
22 items (3 clusters)	24 items (3 clusters)	24 items (3 clusters)
Cluster 1 (8 items)	**Cluster 1 (9 items)**	**Cluster 1 (11 items)**
corrective action plan	birth	assurance
ephi	clinical information	breach notification
health information	diagnosis	corrective action
ocrs investigation	employee	email
policy	ephi	email address
risk	health information	ocrs investigation
staff	social security number	policy
technical assistance	technical assistance	staff
	workforce member	substitute notice
		technical assistance
		workforce member
Cluster 2 (7 items)	**Cluster 2 (8 items)**	**Cluster 2 (8 items)**
access	access	address
assurance	assurance	birth
breach notification	breach notification	business associate
corrective action	business associate	diagnosis
employee	corrective action	ephi
phi	name	health information
protected health information	phi	name
	protected health information	social security number
Cluster 3 (7 items)	**Cluster 3 (7 items)**	**Cluster 3 (5 items)**
address	address	access
birth	email	clinical information
credit monitoring	ocrs investigation	employee
diagnosis	policy	phi
name	safeguard	protected health information
social security number	staff	
workforce member	substitute notice	

Table 6. Insider/Outsider Threats, Vulnerabilities, Breach Incidents, Impacts, and Responses in Unauthorized Access/Disclosure.

Unauthorized Access/Disclosure	Cluster	Insider/Outsider Threats	Vulnerabilities	Breach Incidents	Impacts	Responses
Period 1	Cluster 1	staff	insecure website; malicious staff	unauthorized access to PHI via the public website	ePHI	a settlement to OCR; a corrective action plan
	Cluster 2	employee	weak access termination protocol; malicious employee	unauthorized access to an appointment reminder system after employment ended; identity theft	PHI	updated access termination protocol; termination of the offending employee; retraining of the workforce on HIPAA policies; improvement of HIPAA training materials, risk analysis procedure, operation software, and auditing methods
	Cluster 3	workforce member	malicious employee	unauthorized access to patient medical records	PHI	free credit monitoring services for a year; a program to track anomalies to detect inappropriate use or access; termination of the offending employee and criminal charges against him

Table 6. *Cont.*

Unauthorized Access/Disclosure	Cluster	Insider/Outsider Threats	Vulnerabilities	Breach Incidents	Impacts	Responses
Period 2	Cluster 1	employee; workforce member	malicious employee; negligent physician	suspicious access to ePHI; accidental disclosure of medical files via email	ePHI; medical records; patients' names and clinical information	assurance of corrective action; termination of the responsible individuals' employment; employee sanctions according to its policy and procedure
	Cluster 2	business associate	a malicious business associate (BA); errors of BA	unauthorized access to PHI; erroneous disclosure of another patient's name in letters to patients	PHI	employee sanctions; HIPAA refresher training; a new workflow in mailing processes to reduce the number of manual steps
	Cluster 3	staff	errors of a staff member; negligent employee	erroneous emailing; unsecured email file transfer	PHI; email addresses	retraining staff on its encryption policy; a revised policy regarding electronic transmission of patient information; password protection for electronic files; sanction of the staff member, retraining the entire department; revision of email policies
Period 3	Cluster 1	staff	errors of employees	disclosure of email address; phishing email	email addresses of patients; PHI;	a secure online portal; training staff; two-factor email authentication; technical assistance by OCR; restricted workforce access to the patient folder; employee sanctions
	Cluster 2	business associate	errors of BA; malicious BA; disgruntled former BA; security hole of applications	illegal access to ePHI; mailing error; hacking	ePHI	identity theft protection services to affected individuals; encryption and tools to monitor Internet traffic and compliance
	Cluster 3	workforce member	malicious employee; negligent employee; malicious former employees; insecure records room	impermissibly accessed ePHI as well as paper PHI; access to the records room	demographic and clinical information; ePHI; paper PHI	employee sanctions; revised policy to detect inappropriate access; increased restrictions to access to PHI based on workforce member role and work location; training to all its employees regarding role-based access; measures to improve internal security and limit employee access to records rooms.

In summary, it is noted that insiders such as employees, staff, and workforce members are the major threats in Unauthorized Access/Disclosure. The unauthorized accesses/disclosures lead to the disclosures of email addresses, paper PHI, ePHI, and other demographic and clinical information. While the frequency of unauthorized accesses/disclosures low in Period 1 and Period 2, it was increased rapidly in Period 3. While negligence of employees is a major insider threat in Theft, malicious employees and malicious business associates are major insider threats in Unauthorized Access/Disclosure. It is noted that due to the prevalence of malicious employees and malicious in Unauthorized Access/Disclosure, human and organizational responses to insider threats are stricter than those of Hacking an IT Incident and Theft, including termination of the offending employee, retraining of the workforce, employee sanctions, criminal charges, and updated access termination protocol.

Our findings in Table 6 are supported by the previously validated results such as access control vulnerabilities (e.g., coworkers' computers unattended while logged in and insufficient disabling of electronic and physical access at termination) [56]. The responses in Table 6 are also supported by the validated report such as effective security practices (e.g., two-way authentication for access), training continuously to maintain a proper level of knowledge, implementation of security best practices throughout the organization, use of anti-malware software, and training and awareness on risk perception and cognitive biases that affect a decision [55].

5. Conclusions

The number of data breaches in the healthcare industry is increasing. A total of 712 healthcare data breaches were reported in 2021, exceeding the previous year's total by 10.9%, and 45,706,882 healthcare records were exposed or impermissibly disclosed [57]. Among industries that faced huge operational changes during the pandemic, the healthcare industry experienced the most substantial increase in data breach costs with $9.23 million per incident—a $2 million increase over the previous year [58].

In light of the rising data breaches in the healthcare industry, this study presented a novel approach to the analysis of the descriptive information posted on the website for the HHS OCR data breach reports. Most previous studies have focused on the aggregated statistics of data breach types and locations in the healthcare industry. While those studies were valuable in understanding the data breaches and establishing high-level cybersecurity measures, they could not provide an in-depth analysis of the insider/outsider threats, vulnerabilities, incidents, impacts, and responses to the data breaches. However, despite the insightful information on breach incidents on the HHS OCR data breach reports, it was difficult to analyze due to the unstructured textual contents. Given the current research gap, this study utilizes text mining and data visualization to analyze unstructured text to address two research questions: RQ1: What are the characteristics of insider threats in data breach incidents? RQ2: How have insider threats, vulnerabilities, data breach incidents, impacts/losses, and responses to the data breaches evolved? VOSviewer was used to analyze co-occurrences of keywords in all breach incidents and visualize the clusters of keywords, and NVivo was used to identify the incidents with keywords and analyze the descriptive contents and the insider threats.

The analysis of data breaches in Hacking/IT Incident shows that the major threats are from insiders such as staff, employees, and workforce members. Email phishing has been a persistent and prominent insider threat throughout the three periods. Human vulnerabilities of insider threats include the lack of social engineering awareness, malicious partners (e.g., vendors), and a lack of security awareness of employees. Malware and unauthorized access to the system by the business associates and staff occur through technical vulnerabilities such as weak safeguards for a system and weak security control. Organizational vulnerabilities of insider threats include a lack of a security management process and weak security control.

The analyses of data breaches in Theft also show that the insiders such as staff, employees, and workforce members are the predominant threats to theft. The majority of the insider threat comes from the negligence of the employees, followed by malicious employees. Thefts of medical files, laptops, computers, and devices occur through technical, organizational, and human vulnerabilities such as unencrypted PHI files, unencrypted laptop computers and devices, weak physical security, and negligence of employees. Theft arising from the negligence of employees requires organizational and human responses such as retraining of all staff on privacy and security policies and procedures and development of remote access policy; rather than technical responses. For Theft arising from malicious employees, organizational responses include securely locked storage facilities for mobile devices and the development of policies preventing the removal of devices from the office. While business associates are insider threats in Hacking/IT Incident, they are not a major part of insider threats in Theft.

Insiders such as employees, staff, and workforce members are the major threats in Unauthorized Access/Disclosure. While negligence of employees is a major insider threat in Theft, malicious employees and malicious business associates are major insider threats in Unauthorized Access/Disclosure. The various unauthorized accesses/disclosures lead to the disclosures of email addresses, paper PHI, ePHI, and other demographic and clinical information. Due to the prevalence of malicious employees and malicious in Unauthorized

Access/Disclosure, human and organizational responses to insider threats are stricter than those of Hacking an IT Incident and Theft, including termination of the offending employee, retraining of the workforce, employee sanctions, criminal charges, and updated access termination protocol.

Previous studies have shown that insider threats are a major issue in data security in several fields, and periodic reports have also shown similar results [59]. However, the previous studies did not address how the inside threats lead to specific mitigation efforts. This paper explicitly addressed how the healthcare providers' specific insider threats are related to their vulnerabilities, data breaches, impacts of data breach incidents, and responses. This study showed that different types of insider threats call for different mitigation efforts. With the proposed approach, an individual healthcare provider can develop a comprehensive data security plan to minimize insider threats and prioritize investment budgets for high-impact data security projects. For example, if a healthcare organization has identified particular vulnerabilities to insider threats, it would be better prepared to minimize the vulnerabilities and protect its patient data from insider threats. While our study is limited to three data breach types, future research may analyze other data breach types. Investigation of the relationships between data breach types and data breach sources may shed further insights on the insider threats, vulnerabilities, data breaches, impacts, and responses to the insider threats.

While our study is limited to three data breach types, an analysis can be extended to other data breach types such as emails, servers, or mobile devices and may reveal how insider threats affect mitigation efforts in these data breaches. An analysis of insider threats can be extended to other sectors such as retail, government, or banking for the generalization of the findings. Furthermore, future studies may also compare the usefulness of different text mining tools and their techniques for a richer analysis of insider threats.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Conflicts of Interest: The author declares no conflict of interest.

References

1. Vora, J.; Italiya, P.; Tanwar, S.; Tyagi, S.; Kumar, N.; Obaidat, M.S.; Hsiao, K.F. Ensuring Privacy and Security in E-Health Records. In Proceedings of the 2018 International Conference on Computer, Information and Telecommunication Systems (CITS), Colmar, France, 11–13 July 2018; pp. 1–5. [CrossRef]
2. Bowman, M.A.; Maxwell, R.A. A beginner's guide to avoiding Protected Health Information (PHI) issues in clinical research—With how-to's in REDCap Data Management Software. *J. Biomed. Inform.* **2018**, *85*, 49–55. [CrossRef] [PubMed]
3. Bai, G.; Jiang, J.X.; Flasher, R. Hospital Risk of Data Breaches. *JAMA Intern. Med.* **2017**, *177*, 878–880. [CrossRef] [PubMed]
4. Choi, S.; Martins, J.T.; Bernik, I. Information security: Listening to the perspective of organisational insiders. *J. Inf. Sci.* **2018**, *44*, 752–767. [CrossRef]
5. In Healthcare, Breach Dangers Come from inside the House. *Modern Healthcare*, 9 April 2018. Available online: https://www.modernhealthcare.com/article/20180410/NEWS/180419999/in-healthcare-breach-dangers-come-from-inside-the-house (accessed on 24 January 2022).
6. Capital One Says Breach Hit 100 Million Individuals in U.S. *Bloomberg*, 29 July 2019. Available online: https://www.bloomberg.com/news/articles/2019-07-29/capital-one-data-systems-breached-by-seattle-woman-u-s-says (accessed on 15 February 2022).
7. Security Tracker 2018. Available online: https://www.shredit.com/en-us/resource-center/original-research/security-tracker-2018 (accessed on 24 January 2022).
8. Yaraghi, N.; Gopal, R.D. The Role of HIPAA Omnibus Rules in Reducing the Frequency of Medical Data Breaches: Insights from an Empirical Study. *Milbank Q.* **2018**, *96*, 144–166. [CrossRef]
9. McCoy, T.H.; Perlis, R.H. Temporal Trends and Characteristics of Reportable Health Data Breaches, 2010–2017. *JAMA* **2018**, *320*, 1282–1284. [CrossRef]
10. Gabriel, M.H.; Noblin, A.; Rutherford, A.; Walden, A.; Cortelyou-Ward, K. Data breach locations, types, and associated characteristics among US hospitals. *Am. J. Manag. Care* **2018**, *24*, 78–84.

11. Ayyagari, R. An Exploratory Analysis of Data Breaches from 2005–2011: Trends and Insights. *J. Inf. Priv. Secur.* **2012**, *8*, 33–56. [CrossRef]
12. Wikina, S.B. What Caused the Breach? An Examination of Use of Information Technology and Health Data Breaches. *Perspect. Health Inf. Manag.* **2014**, *11*, PMC4272442.
13. Elbattah, M.; Arnaud, É.; Gignon, M.; Dequen, G. The Role of Text Analytics in Healthcare: A Review of Recent Developments and Applications. August 2022, pp. 825–832. Available online: https://www.scitepress.org/Link.aspx?doi=10.5220/0010414508250832 (accessed on 13 July 2022).
14. Xue, J.; Chen, J.; Chen, C.; Zheng, C.; Li, S.; Zhu, T. Public discourse and sentiment during the COVID-19 pandemic: Using Latent Dirichlet Allocation for topic modeling on Twitter. *PLoS ONE* **2020**, *15*, e0239441. [CrossRef]
15. Gourisaria, M.K.; Chandra, S.; Das, H.; Patra, S.S.; Sahni, M.; Leon-Castro, E.; Singh, V.; Kumar, S. Semantic Analysis and Topic Modelling of Web-Scrapped COVID-19 Tweet Corpora through Data Mining Methodologies. *Healthcare* **2022**, *10*, 881. [CrossRef]
16. Chakraborty, K.; Bhatia, S.; Bhattacharyya, S.; Platos, J.; Bag, R.; Hassanien, A.E. Sentiment Analysis of COVID-19 tweets by Deep Learning Classifiers—A study to show how popularity is affecting accuracy in social media. *Appl. Soft Comput.* **2020**, *97*, 106754. [CrossRef] [PubMed]
17. Imran, A.S.; Daudpota, S.M.; Kastrati, Z.; Batra, R. Cross-Cultural Polarity and Emotion Detection Using Sentiment Analysis and Deep Learning on COVID-19 Related Tweets. *IEEE Access* **2020**, *8*, 181074–181090. [CrossRef]
18. CERT Definition of 'Insider Threat'—Updated. *SEI Blog*. Available online: https://insights.sei.cmu.edu/blog/cert-definition-of-insider-threat-updated/ (accessed on 19 July 2022).
19. Dang, Q.-V. Intrusion Detection in Software-Defined Networks. In *Future Data and Security Engineering*; Springer: Cham, Switzerland, 2021; pp. 356–371. [CrossRef]
20. Alkadi, O.; Moustafa, N.; Turnbull, B. A Review of Intrusion Detection and Blockchain Applications in the Cloud: Approaches, Challenges and Solutions. *IEEE Access* **2020**, *8*, 104893–104917. [CrossRef]
21. Ellerby, Z.; McCulloch, J.; Wilson, M.; Wagner, C. Exploring How Component Factors and Their Uncertainty Affect Judgements of Risk in Cyber-Security. In *Critical Information Infrastructures Security*; Springer: Cham, Switzerland, 2020; pp. 31–42. [CrossRef]
22. Al-Mhiqani, M.N.; Ahmad, R.; Zainal Abidin, Z.; Yassin, W.; Hassan, A.; Abdulkareem, K.H.; Ali, N.S.; Yunos, Z. A Review of Insider Threat Detection: Classification, Machine Learning Techniques, Datasets, Open Challenges, and Recommendations. *Appl. Sci.* **2020**, *10*, 5208. [CrossRef]
23. Homoliak, I.; Toffalini, F.; Guarnizo, J.; Elovici, Y.; Ochoa, M. Insight into Insiders and IT: A Survey of Insider Threat Taxonomies, Analysis, Modeling, and Countermeasures. *ACM Comput. Surv.* **2019**, *52*, 30:1–30:40. [CrossRef]
24. Soh, C.; Yu, S.; Narayanan, A.; Duraisamy, S.; Chen, L. Employee profiling via aspect-based sentiment and network for insider threats detection. *Expert Syst. Appl.* **2019**, *135*, 351–361. [CrossRef]
25. Saxena, N.; Hayes, E.; Bertino, E.; Ojo, P.; Choo, K.-K.R.; Burnap, P. Impact and Key Challenges of Insider Threats on Organizations and Critical Businesses. *Electronics* **2020**, *9*, 1460. [CrossRef]
26. Ajayi, O.; Abouali, M.; Saadawi, T. Secured Inter-Healthcare Patient Health Records Exchange Architecture. In Proceedings of the 2020 IEEE International Conference on Blockchain (Blockchain), Rhodes, Greece, 2–6 November 2020; pp. 456–461. [CrossRef]
27. Nathiya, T.; Suseendran, G. An Effective Hybrid Intrusion Detection System for Use in Security Monitoring in the Virtual Network Layer of Cloud Computing Technology. In *Data Management, Analytics and Innovation*; Springer: Singapore, 2019; pp. 483–497. [CrossRef]
28. Deep, G.; Mohana, R.; Nayyar, A.; Sanjeevikumar, P.; Hossain, E. Authentication Protocol for Cloud Databases Using Blockchain Mechanism. *Sensors* **2019**, *19*, 4444. [CrossRef]
29. Prabhu, S.; Thompson, N. A primer on insider threats in cybersecurity. *Inf. Secur. J. A Glob. Perspect.* **2021**, *2021*, 1971802. [CrossRef]
30. Gunasekhar, T.; Rao, K.T.; Basu, M.T. Understanding insider attack problem and scope in cloud. In Proceedings of the 2015 International Conference on Circuits, Power and Computing Technologies [ICCPCT-2015], Nagercoil, India, 19–20 March 2015; pp. 1–6. [CrossRef]
31. Liu, L.; de Vel, O.; Han, Q.-L.; Zhang, J.; Xiang, Y. Detecting and Preventing Cyber Insider Threats: A Survey. *IEEE Commun. Surv. Tutor.* **2018**, *20*, 1397–1417. [CrossRef]
32. Yen, T.F.; Oprea, A.; Onarlioglu, K.; Leetham, T.; Robertson, W.; Juels, A.; Kirda, E. Beehive: Large-scale log analysis for detecting suspicious activity in enterprise networks. In Proceedings of the 29th Annual Computer Security Applications Conference, New Orleans, LA, USA, 9–13 December 2013; pp. 199–208. [CrossRef]
33. Allodi, L.; Massacci, F. Security Events and Vulnerability Data for Cybersecurity Risk Estimation. *Risk Anal.* **2017**, *37*, 1606–1627. [CrossRef] [PubMed]
34. Malatji, M.; Marnewick, A.; von Solms, S. Validation of a socio-technical management process for optimising cybersecurity practices. *Comput. Secur.* **2020**, *95*, 101846. [CrossRef]
35. Švábenský, V.; Čeleda, P.; Vykopal, J.; Brišáková, S. Cybersecurity knowledge and skills taught in capture the flag challenges. *Comput. Secur.* **2021**, *102*, 102154. [CrossRef]
36. Esteves, J.; Ramalho, E.; de Haro, G. To Improve Cybersecurity, Think Like a Hacker. *MIT Sloan Manag. Rev.* **2017**, *58*, 71–77.
37. Jeremiah, P.; Samy, G.N.; Shanmugam, B.; Ponkoodalingam, K.; Perumal, S. Potential Measures to Enhance Information Security Compliance in the Healthcare Internet of Things. In *Recent Trends in Data Science and Soft Computing*; Springer: Cham, Switzerland, 2019; pp. 726–735. [CrossRef]

38. Dorasamy, M.; Joanis, G.C.; Jiun, L.W.; Jambulingam, M.; Samsudin, R.; Cheng, N.J. Cybersecurity Issues Among Working Youths in an IoT Environment: A Design Thinking Process for Solution. In Proceedings of the 2019 6th International Conference on Research and Innovation in Information Systems (ICRIIS), Johor Bahru, Malaysia, 2–3 December 2019; pp. 1–6. [CrossRef]
39. Dang-Pham, D.; Pittayachawan, S.; Bruno, V. Impacts of security climate on employees' sharing of security advice and troubleshooting: Empirical networks. *Bus. Horiz.* **2016**, *59*, 571–584. [CrossRef]
40. Yeng, P.K.; Fauzi, M.A.; Yang, B. A Comprehensive Assessment of Human Factors in Cyber Security Compliance toward Enhancing the Security Practice of Healthcare Staff in Paperless Hospitals. *Information* **2022**, *13*, 335. [CrossRef]
41. Hughes-Lartey, K.; Qin, Z.; Botchey, F.E.; Dsane-Nsor, S. An Assessment of Data Location Vulnerability for Human Factors Using Linear Regression and Collaborative Filtering. *Information* **2020**, *11*, 449. [CrossRef]
42. Rothrock, R.A.; Kaplan, J.; van der Oord, F. The Board's Role in Managing Cybersecurity Risks. *MIT Sloan Manag. Rev.* **2018**, *59*, 12–15.
43. Hasan, S.; Ali, M.; Kurnia, S.; Thurasamy, R. Evaluating the cyber security readiness of organizations and its influence on performance. *J. Inf. Secur. Appl.* **2021**, *58*, 102726. [CrossRef]
44. Sharma, S.; Warkentin, M. Do I really belong? Impact of employment status on information security policy compliance. *Comput. Secur.* **2019**, *87*, 101397. [CrossRef]
45. AlGhamdi, S.; Win, K.T.; Vlahu-Gjorgievska, E. Information security governance challenges and critical success factors: Systematic review. *Comput. Secur.* **2020**, *99*, 102030. [CrossRef]
46. Nasir, A.; Arshah, R.A.; Hamid, M.R.A.; Fahmy, S. An analysis on the dimensions of information security culture concept: A review. *J. Inf. Secur. Appl.* **2019**, *44*, 12–22. [CrossRef]
47. Georgiadou, A.; Mouzakitis, S.; Bounas, K.; Askounis, D. A Cyber-Security Culture Framework for Assessing Organization Readiness. *J. Comput. Inf. Syst.* **2020**, *2020*, 1845583. [CrossRef]
48. Bodeau, D.J.; Graubart, R.D. Cyber Prep 2.0: Motivating Organizational Cyber Strategies in Terms of Threat Preparedness. February 2017. Available online: https://www.mitre.org/publications/technical-papers/cyber-prep-20-motivating-organizational-cyber-strategies-in-terms-of (accessed on 24 January 2022).
49. Villegas-Ch, W.; Ortiz-Garces, I.; Sánchez-Viteri, S. Proposal for an Implementation Guide for a Computer Security Incident Response Team on a University Campus. *Computers* **2021**, *10*, 102. [CrossRef]
50. Ani, U.P.D.; He, H.M.; Tiwari, A. Review of cybersecurity issues in industrial critical infrastructure: Manufacturing in perspective. *J. Cyber Secur. Technol.* **2017**, *1*, 32–74. [CrossRef]
51. Shojaeshafiei, M.; Etzkorn, L.; Anderson, M. Cybersecurity Framework Requirements to Quantify Vulnerabilities Based on GQM. In *National Cyber Summit (NCS) Research Track*; Springer: Cham, Switzerland, 2020; pp. 264–277. [CrossRef]
52. Xia, Q.; Sifah, E.B.; Smahi, A.; Amofa, S.; Zhang, X. BBDS: Blockchain-Based Data Sharing for Electronic Medical Records in Cloud Environments. *Information* **2017**, *8*, 44. [CrossRef]
53. O. for C. Rights (OCR). Breach Notification Rule. *HHS*, 14 September 2009. Available online: https://www.hhs.gov/hipaa/for-professionals/breach-notification/index.html (accessed on 2 May 2022).
54. van Eck, N.J.; Waltman, L. VOSviewer Manual. p. 51. Available online: https://www.vosviewer.com/documentation/Manual_VOSviewer_1.6.8.pdf (accessed on 5 May 2022).
55. Unintentional Insider Threats: A Foundational Study. Available online: https://resources.sei.cmu.edu/library/asset-view.cfm?assetid=58744 (accessed on 13 July 2022).
56. Moore, A.P.; Cappelli, D.M.; Trzeciak, R.F. The 'Big Picture' of Insider IT Sabotage Across U.S. Critical Infrastructures. In *Insider Attack and Cyber Security*; Springer: Boston, MA, USA, 2008; p. 46.
57. December 2021 Healthcare Data Breach Report. *HIPAA Journal*, 18 January 2022. Available online: https://www.hipaajournal.com/december-2021-healthcare-data-breach-report/ (accessed on 15 February 2022).
58. IBM Report: Cost of a Data Breach Hits Record High During Pandemic. *IBM Newsroom*. Available online: https://newsroom.ibm.com/2021-07-28-IBM-Report-Cost-of-a-Data-Breach-Hits-Record-High-During-Pandemic (accessed on 15 February 2022).
59. 2021 Data Breach Investigations Report. *Verizon Business*. Available online: https://www.verizon.com/business/resources/reports/dbir/ (accessed on 15 February 2022).