Heliyon 7 (2021) e06522

Contents lists available at ScienceDirect

Heliyon

journal homepage: www.cell.com/heliyon

Research article

Human factor, a critical weak point in the information security of an organization's Internet of things

Kwesi Hughes-Lartey [a,d], Meng Li [a,b], Francis E. Botchey [a,d], Zhen Qin [a,b,c,*]

[a] School of Information and Software Engineering, University of Electronic Science and Technology of China, Chengdu, China
[b] Institute of Electronic and Information Engineering UESTC in Guangdong, China
[c] Network and Data Security Key Laboratory of Sichuan Province, China
[d] Computer Science Department, Koforidua Technical University, Koforidua, Ghana

ARTICLE INFO

Dataset link:
https://www.kaggle.com/archangell/hipaa-breaches-from-20092017

Keywords:
Data breach
Human behavior
Human factors
Information security
Internet of things

ABSTRACT

Internet of Things (IoT) presents opportunities for designing new technologies for organizations. Many organizations are beginning to accept these technologies for their daily work, where employees can be connected, both on the organization's premises and the "outside", for business continuity. However, organizations continue to experience data breach incidents. Even though there is a plethora of researches in Information Security, there "seems" to be little or lack of interest from the research community, when it comes to human factors and its relationship to data breach incidents. The focus is usually on the technological component of Information Technology systems. Regardless of any technological solutions introduced, human factors continue to be an area that lacks the required attention. Making the assumption that people will follow expected secure behavioral patterns and therefore system security expectations will be satisfied, may not necessarily be true. Security is not something that can simply be purchased; human factors will always prove to be an important space to explore. Hence, human factors are without a doubt a critical point in Information Security. In this study, we propose an Organizational Information Security Framework For Human Factors applicable to the Internet of Things, which includes countermeasures that can help prevent or reduce data breach incidents as a result of human factors. Using linear regression on data breach incidents reported in the United States of America from 2009 to 2017, the study validates human factors as a weak-point in information security that can be extended to Internet of Things by predicting the relationship between human factors and data breach incidents, and the strength of these relationships. Our results show that five breach incidents out of the seven typified human factors to statistically and significantly predict data breach incidents. Furthermore, the results also show a positive correlation between human factors and these data breach incidents.

1. Introduction

Internet of Things (IoT) has been gaining grounds rapidly over the years concerning the Internet or Computer Networks [1] and according to [2] IoT mainly refers to the augmentation of physical objects and devices, where these objects and devices have sensing, computing and communicating abilities and are connected in a network to utilize them collectively. Over the years, there has been a lot of research on IoT and they have mainly been from the perspective of thing-oriented. These researches have covered areas such as identification of objects, tracking of objects, privacy control, sensing data visualization and object networking [1]. However, the interaction between humans and IoT is an arena that has a lot to be explored [2], not to even mention or talk about information security vulnerabilities on IoT as a result of human factors. The actual concept of IoT was conceptualized by the Auto-ID Center at Massachusetts Institute of Technology (MIT), which began to design and propagate across a company radio frequency Identification Infrastructure. The concept was to make all objects in the world network connected and to represent a vision where the Internet makes an extension into the real world concerning everyday objects [3]. The whole idea of 'things' in a network structure refers to either real or virtual actors, such as real-world objects, virtual data, intelligent software and human beings being participants. This is to create an environment that provides access to basic information from one object to the other, to facilitate information sharing with others in the real-world effectively [4]. Nicolescu et al. [5] provide a better and current working definition of

* Corresponding author at: School of Information and Software Engineering, University of Electronic Science and Technology of China, Chengdu, China.
E-mail address: qinzhen@uestc.edu.cn (Z. Qin).

https://doi.org/10.1016/j.heliyon.2021.e06522
Received 26 May 2020; Received in revised form 9 August 2020; Accepted 11 March 2021

IoT. The term IoT can mean different things to different actors and the values associated with IoT do not merely vary with the more obvious technological, economic, and political factors, but also with behavioral patterns and cultural practices across individuals, communities, and demographics.

It must be noted that IoT does not have a unique definition. Nevertheless, a broad understanding of IoT is one that provides any service over the traditional Internet which enables or provisions human-to-thing, thing-to-thing or thing-to-things communications. Due to the connections and communications among these actors, there are many potential threats and attacks against the security and privacy of information or things, which has become a source of concern. Hence, these concerns need to be investigated or studied and addressed appropriately. These studies should make simple the design and development of IoT objects that will enable a plethora of services for human beings in different sectors such as the health sector in which different end-user devices (smartphones, laptops, tablets, etc.) and their periphery.

Security breach of sensitive data such as personal data has become a phenomenon that occurs regularly over the last few years. These acts are perpetrated by malicious cyber actors who attack organizations' information systems through different means for information ex-filtration [6]. It affects almost every sector, especially those with 'valuable' data such as the health sector. Today cyber threats keep increasing and the vulnerability of industries such as healthcare is apparent. In 2015, Anthem Inc. was hacked, and this led to an exposure of millions of data, where individuals lost their Protected Health Information (PHI), potentially putting them at risk for identity theft and fraud [7]. Modern research shows how the focus of information security is mostly geared towards providing solutions through the technology, such as deep learning network architecture proposed for human activity recognition based on mobile sensor data [8], collaborative privacy-preserved deep neural network architecture (dubbed MSCryptoNet) based on a fully homomorphic cryptosystem [9], deep learning framework to identify smartphone users based on the original smartphone sensor data, acquired when users shake their smartphones or perform some daily actions [10], predicting demographic information by leveraging the perspectives of smartphone application usage [11], lightweight device authentication protocol; speaker-to-microphone (S2M) by leveraging the frequency response of a speaker and a microphone from two wireless IoT devices as the acoustic hardware fingerprint [12], partially hidden policy to protect private information in an access policy [13], attribute hiding predicate encryption with equality test is formulated to provide the privacy preservation of user attributes and flexible search capability on ciphertexts simultaneously [14], semi-supervised generative adversarial network (GAN) for channel state information (CSI)-based activity recognition (CsiGAN) based on the general semi-supervised GANs [15] and a hybrid framework, Super-Recognition of Pedestrian Re-Identification (SRPRID), to strengthen pedestrian re-identification based on multi–resolution images captured by disparate cameras [16].

People, just like information technology, are part and parcel of information security and even though there have been many technological advances in information technology with such sophistication that makes it difficult for data breach incidents to occur on a technological level, the same cannot be said about people. Data breach offenders are becoming more and more aware of the fact that human factors, whether error or behavior, may be a weak point of an information security structure for their success [17]. This is a clear indication that the strength of any good information security system is in the hands of those who use it and not just the technology. Even though the popularity of mobile devices and other sensors can be used to collect biometric information to ensure security [18], human behavior is still critical. The ideology that good technological systems can solve an organization's security problem is an indication of a lack of understanding of the problem and also a lack of understanding of technology [19]. [20] claims that the major type of information breach that organizations face is to some extent related to the exploitation of human resources in terms of the error they commit or their user behavior. The staff of an organization is seen as the Achilles hill for information security breaches [20, 21, 22, 23]. Research also shows that most of the time, organizations have the 'habit' of consistently overlooking human factors as a key cause of security breaches and would rather prioritize their resources on technological controls and solutions [24]. The inescapable tie between information security and humans cannot just be overlooked. And when it comes to information security system evaluations, organizations will mostly evaluate the technological aspect of the system and will perform very little or no evaluation of human factors, which can greatly impact the vulnerability of the security system [25]. The protections of data or information is dependent on a good information security plan, which must be one that does not overlook human factors and the various controls on user behavior and habits apart from the usual practice of technological controls [26]. This work makes use of data made available by [27] on Kaggle's official website which contains the official dataset from the Department of Health and Human Services (DHHS) on all reported Protected Health Information (PHI) data breaches from medical centres, including dental centres in the United States of America. The data provides an archive of reported Health Insurance Portability and Accountability Act of 1996 (HIPAA). HIPAA is a US Federal Law that sets standards of how healthcare plans, healthcare clearing houses and healthcare providers protect the privacy of their patients' health information. HIPAA privacy rule allows the recognition that it is not practicable to remove all risks of incidental disclosure. Also when there are policies with reasonable safeguards and appropriate limits of how PHI is used and disclosed, then incident disclosure does not violate the rule [28].

According to [29] the National Conference of State Legislatures passed by all 50 states, Puerto Rico, District of Columbia, and the U.S. Virgin Islands, required government and private entities to notify individuals who have been impacted by information security breaches that may compromise their personally identifiable information. Typically the laws define what is classified as personally identifiable information in each state, what entities are required to comply with, what specifically constitutes a breach, the timing and method of notice required to individuals and regulatory agencies, and consumer credit reporting agencies, and any exemptions that apply, such as exemptions for encrypted data. Also business entities or organizations report or notify an entity designated by the US Homeland Security information about information security incidents, threats, and vulnerabilities. These entities shall promptly notify and provide that same information to the United States Secret Service, the Federal Bureau of Investigation, and the Commission for civil law enforcement purposes, and shall make it available as appropriate to other federal agencies for law enforcement, national security, or computer security purposes [29].

The motivation of this study is to make available to organizations or business entities information about the potential connections between data breach incident types and human factors. The study provides an analysis covering nine years of data breach incidents. This is done by predicting the relationship between human factors and five data breach incident types. This paper also examines the relationship between human factors and 'other' and 'unknown' breach types. These breach types are explained in section 4. The dataset used, to the best of our knowledge, was last updated in October 2017, before and during its analysis. The classification of the dataset is discussed in section 4. The three main contributions of this paper are summarized as follows:

- To propose a holistic information security framework for human factors and technology based on literature. The proposed framework may be useful in reducing the number of data breach incidents due to the increase in human factors.
- To predict the relationship between data breach incidents and human factors. The study takes into consideration different types of data breach incidents that happened as a result of human action and model the relationship between incidents and human factors.

The predicted relationship will help us understand the extent to which human factors affect data breach incidents and information security at large. By doing this, the paper explains from the dataset, the variation in data breach incidents that can be attributed to human factors or to determine whether human factors have no relationship with data breach incidents.

- To evaluate the strength of the relationship between data breach incidents and human factors. By establishing the dependence or the association of data breach incidents and human factors, the analysis will not determine a causation but rather the strength of the relationship.

The rest of this paper is arranged as follows: a review of related work in section 2. In section 3, the study proposes a framework for human factors and technology. In section 4, a classification of the dataset used is explained and also an evaluation of the linear regression is provided. In Section 5, a validation of human factors as a weak link in information security by providing an analysis of different data breach incidents with human factors underpinning them and predicting the relationship between human factors and data breach incidents. A discussion on the regression models in section 6 and the conclusion in section 7.

2. Related work

In the world of IoT, information communications and collaborations among hospital staff is a problem in the healthcare industry. One of the factors that contributes to this, is the lack of computer and information security knowledge on how to use health information and also the lack of controls over the transmission and receipt of information during high loading work periods and most often leads to a lot of difficulties, including the use of electronic health information, hence compromising the security of the system. The study further indicated that nurses, pharmacists and public health workers used health information more than physicians [30]. [31] elaborates on the importance of preserving privacy in IoT, and human behavior in the use of IoT can not be downplayed.

2.1. Developments in cyber risk assessment for Internet of things

According to [5] IoT technology is associated with an entire spectrum of values that is yet to be assessed, and should be assessed on three main domains: economic, technical and social. The value of IoT can not be minimized to one or two of these domains, even though such practices go on in today's world. It is important to note that social and cultural customs can norm and limit not only economic aspects of IoT but also the technological part and this becomes critical for good information security. This implies that research into the security of IoT must be broadened to social aspects, a domain where human factors can be addressed. They adapted IoT of both Cyber Value at Risk model, a well-established model for measuring the maximum possible loss over a given time, and the MicroMort model, a widely used model for predicting uncertainty through units of mortality risk. The resulting new IoT MicroMort for calculating IoT risk is tested and validated with real data from the BullGuard's IoT Scanner of over 310,000 scans, and the Garner report on IoT connected devices. With these two calculations were developed, the current state of IoT cyber risk and the future forecasts of IoT cyber risk. Therefore, their work focused on the advances in the efforts of integrating cyber risk impact assessments and offer a better understanding of economic impact assessment for IoT cyber risk. [32] proposed a decomposed cyber security risk assessment standard in which there is a combination of concepts of building a model for building standardization of impact of assessments. The proposed model is identified to have two problems: new design principles for assessing cyber risk, and the identification of different risk vectors. Their paper focused on the analysis of the best approach for quantifying the impact of cyber risk in the IoT space. The model and the documented process represents a new design for mapping IoT risk vectors and optimizing IoT risk impact assessment. Radanliev et al. [33] argue that designing a holistic model for IoT risk assessment and risk management remain a challenge. The design of any assessment model must focus on, IoT economic impact, IoT machine ethics, IoT sensor networks, IoT safety, IoT cyber security and IoT equipment. They discussed how interdisciplinary research could prove to be very beneficial to help more people to understand and consider the many issues around the risk in IoT systems and ultimately, make a contribution to the design of a holistic approach to IoT risk assessment.
These trends and advances in managing risks associated with internet security provide no or little parameters or vectors for human factors as a major component of the assessments.

2.2. Behavior

According to [26], the notion that suggests informal behavior is a central theme in describing those characteristics of people, organizations, and acts of communication which affect information. This means that the management of information security connotes the management of the integrity of communication. The argument therefore then continues that behavior and communication should be considered as opposite sides of the same coin and that, any kind of discordance in behavioral patterns could potentially lead to security breaches. There is hence, an understanding that a cause and effect relationship between unwarranted behavior and breakdown in communication may lead to a security breach. Furthermore, consideration for information systems and communication must be made, understanding that the consideration of both to be the same thing is not a new idea. Organizations must recognize that information system facilitates communication and must be interlaced from threads of communication. It will therefore not be a far-fetched idea that any problem with the system of communication will most certainly affect information security systems that facilitate it directly and vice-versa. Organizations and businesses that need to protect themselves against attacks, often do so because of the wealth of resources at their disposal to protect their information technology. However, these resources barely have any link what so ever to their investment in making their staff immune to data breach incidents. By this, attackers know that people are likely to be the weakest linkage in the chain of information security and invest many hours to track down and exploit their vulnerability and carelessness even without any guarantee of success. This is because when dealing with people, they can be confident of discovering any number of vulnerabilities or careless behaviors, while being just a little creative [34].

2.3. Human factors

In academic literature, several theories postulated after investigations by researchers and reported human factors to have had an impact on user behavior, both negative and positive. According to [35] as cited by [36]), one of these theories is security culture (cultural factor). This is a human factor that is associated positively with an employee's willingness to follow laid down security procedures. In every organization, corporate culture can exist whether employees are aware of it or not. In other words, they may not necessarily be aware consciously of such a culture but may be operating in it [37]. Organizational culture is not the only parameter to be considered when dealing with the cultural aspect of human factors, but also factors such as national culture, regional location and religion. National culture has a direct effect on the usefulness of the level of information protection and behavior. Studies conducted mostly on Western culture and Asian culture indicate that Western organizational cultures are more individualist while Asian organizational cultures are more of a collective one [38]. Another human factor to consider is personality. [36] claims that five traits are often used to describe people according to their psychology; these being openness, agreeableness, extraversion, conscientiousness and neuroticism. They further go

on to explain the relationships of these personality traits and information security compliance behavior based on a study conducted by [39]. This research sampled 120 users with a research model based on the five major personality traits aforementioned. Their results revealed that conscientiousness and agreeableness have a significant impact on user compliance with information security policy. Alotaibi et al. also explained in another study by [40], in which the study was designed to understand personality traits that underpin behavior and the extent to which it affects users' intention to comply with information security policy. They did this by implementing and empirically validating a comprehensive theoretical model that aimed to assess the impact of the personality factors. The results of their research on 481 participants showed that more open, conscientious and agreeable participants were likely to comply with information security policy. Conversely, the participants who were more extroverted and neurotic often tend to violate information security policy [40]. Alotaibi et al. considers perception as yet another human factor which was investigated by [41]. The study considered perception to be a key component of human behavior and a major part of intelligence. Proctor argues that human interpretation or recognition of sensory information has a substantial impact on user behavior. So the perception of employees for information technology has a great impact on their behavior and decisions [36]. This concept is complemented by [42], where the investigation showed that when organizations are dealing with users' perception of information security, their perception is determined by several factors, such as awareness, knowledge, controllability, severity and possibility, which will, in turn, become an influencing factor to their behavior and decisions. It is very important to understand that when users have a complete picture and full awareness of what is happening in an information security policy space, it will positively impact their ability to recognize potential threats. Therefore, perception can be considered as knowledge about a particular domain and as such employees should keep up to date with the latest threat patterns and the consequential security requirements. An important human factor that also influences user behavior cited in academic literature in the form of reported incidents is gender. [43] Hanley et al. in their study found that, 94 percent of insider incidents were associated with males while a technical report by [44] also found that the majority of insider incidents were male initiated. [43]. However, [45] makes a counter-argument that both genders pose an equal threat to information security. Their study found that 50 percent of insider threats were associated with females and 50 percent with males. An examination of habits theory proposes that humans perform many actions without making conscious decisions and then get familiar with executing these actions. The argument explains that information technology usage is directly related to habits. The actual behavior of users is highly influenced by their technology usage habits. And so some researchers think that habitual behavior explains information security policy non-compliance [36]. Pahnila et al. studied factors that impact users' compliance using a theoretical model and one of the factors was users' habits. The study was an empirical one, provided by a model of over 245 participants from a Finnish company. The results exposed that users' habits have a significant impact on intention to comply with information security policy [46]. Looking at employee satisfaction, which is a component of human factors as defined by [35]. It is the employee's overall feeling of well-being while at work. It is widely accepted that an employee who is satisfied with his or her employer is most likely to conform to the organization's information security policy. Users who report positive feelings about their organization are expected to have a good sense of their responsibilities, especially in terms of conforming to information security policy. [35] further argues that some studies have investigated the relationship between job satisfaction and employee conformity. They provided empirical support for the claim that job satisfaction has a positive impact on compliance with security policy. Their examples examined the influence of job satisfaction on user's information security policy compliance decisions and in their theoretical research model, they hypothesized that satisfaction is positively associated with security compliance intention. The research model was tested on 223 survey participants, and the results suggested that job satisfaction contributes to security policy compliance. The result further found a strong relationship between users' intention to conform to information security and job satisfaction.

The last human factor discussed, is technology democracy. [47] explains that systems and applications that are used at work and home have converged and have become intertwined over the years. Applications that are used in home environments are now used in business systems as well, which potentially creates a challenge to the status quo of the use of technology in many organizations. Users today demand more freedom to use a wider variety of applications and devices to do their work more effectively. This is classified as 'technology democracy'. Again, when there is a mixture of work and home environments, employees will more likely demonstrate unintelligent behavior towards security [47].

3. Organizational information security framework for human factors in an Internet of things

In this section, an information security framework for human factors is proposed for an IoT as shown in Fig. 1. The framework mainly focuses on the non-technological aspects because information technology is much more protected than the users who use it [34]. Therefore, in this framework, the discussion is centered on countermeasures to the breaches mentioned in section 4. The IoT part mainly represents the technological aspect and is divided into four parts, with all needing or requiring an appropriate security:

- Technology: Technology represents the type of processor chips, sensors, Radio-Frequency Identification (RFID), Near Field Communication (NFC), and cyber-physical systems [48].
- 'Things': These are objects such as wearables, televisions, laptops, tablets, smartphones, cars, e.t.c. [48]
- Infrastructure: IoT infrastructure consists of access technologies, data storage and processing, data analytics, and security. These are the pillars that enable growth for future IoT solutions [49].
- Software: A complete IoT system requires software. It addresses the domains of networking and action through platforms, embedded systems, middleware, and partner systems. The individual and master applications are accountable for data acquisition, device integration, real-time analytics, and process extension in an IoT network [49].
- Security: Security for IoT found in all the domains mentioned above. It is designed to ensure the steady working of all the functionalities in an operational system, so that devices that are connected can give a business a real boost. Any thing or device connected to the Internet can be exposed to cyber-attacks. IoT security is the technology area that provides 'safety' for the connected devices and networks in the Internet of Things [48, 49].

3.1. Hacking

3.1.1. User awareness

Organizations need to have a consistent policy that focuses on having their employees trained or educated, and updated on the best practices in protecting themselves from hackers which will in turn be a protection for the organization. To ensure that users don't become weak links in an IoT, they must be equipped with the relevant knowledge. This must be done to shield and reduce user susceptibility to hacking activities. [26, 36, 50] all highlight the impact of security awareness on employee behavior and its significance in influencing their intention to comply with the best practices. Employees are likely to violate security policies when there is a lack of awareness and knowledge.

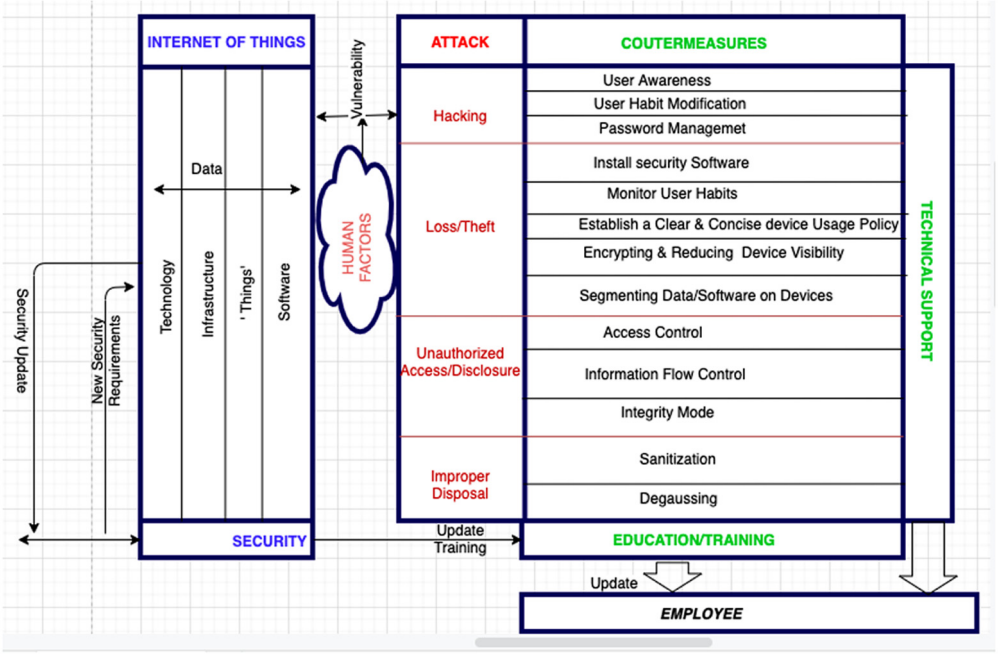

Fig. 1. Organizational Information Security Framework For Human Factors in an IoT.

3.1.2. User habits

User habits in an IoT is crucial to its security. Once again education is essential. Educating and training users on how to behave online will help in the modification of their online habits. This will particularly help when it comes to a hacker tracking users' habit online. For example, when users are contacted to verify an account, they need to be well informed not to comply but to contact the appropriate entity in the organization from which the apparent verification came from to have its legitimacy verified. User habits, such as clicking on hyperlinks will have to be discouraged, since hackers may use these links deceptively in obtaining their user credentials. Rather they need to be well informed as much as possible to always type the organization's correct web address directly in a web browser's address bar, in order not to be vulnerable to phishing [51].

3.1.3. Password management

[34] makes the argument of how many people have insufficient understanding of the inner workings of a computer and due to that, it is quite difficult for many to appreciate computer security principles. Often, user understanding is fuzzy in rather basic things, such as password management. Again education remains the fundamental theme. User training should be aimed at helping end-users understand some of the best practices in 'simple but complex' password credentials. Password management must include the following:

- Not writing down passwords anywhere
- Changing password at least every 3 months
- Not using words or phrases associated with the user
- Not using the same password for every account
- Every password must be at least 8 characters
- Passwords must be a mix of letters (uppercase and lowercase), numbers and special characters
- Not using default passwords

Users must be made aware that they are the weakest link in the chain of security and that hackers are willing to spend many hours tracking, monitoring, and exploiting new vulnerabilities without guaranteed success. They are confident of discovering a number of passwords since people are involved.

3.2. Loss and theft

The loss and theft of mobile or portal devices are threats that could result in data loss. When it comes to data loss, one must ascertain whether it bothers on data-at-rest and data-in-motion to ensure the confidentiality and integrity of the data. However, portable devices are more sophisticated than that. This must involve protecting data on the device, data in the applications, and data over the network [52]. Organizations can set out the following to help mitigate device loss or theft.

3.2.1. Installing security software on portable or mobile devices

This is an important countermeasure, where IT experts of an organization should pay equal attention to just like other hardware pieces like servers on the corporate network.

3.2.2. Monitoring user behavior

Employees most of the time are oblivious when their devices are compromised, hence, putting themselves at risk. For this, a consistent monitoring of user behavior can show anomalies that can be indicative that an attack is on the way. Furthermore, automated monitoring may also prove crucial when making sure your organizations' IoT security policies are not infringed upon.

3.2.3. Establishing a clear and concise portable or mobile device usage policy

Organizational security policies must include a mobile or portable device usage policy. This should sufficiently cover the acceptable use of anti-loss and anti-theft procedures and guidelines and a mandatory security sitting. The guidelines should also implement a compliance monitoring and remediation of deficiencies.

3.2.4. Encrypting and reducing visibility into devices that have access to the organizational network

It is best if a malicious user cannot easily access data on the device, in cases where a device gets lost or stolen. The taking over of a lost or stolen device should also not allow the malicious user to have a 'walk in the network' of the organization. To achieve this, user and device identities must be placed in a comprehensive identity and access management (IAM) system.

3.2.5. Segmenting data and software in user devices that participate on the organizational network

To minimize the exposed attack surface area when a device is lost or stolen, data segmentation, by placing users with mobile or portable devices into role-based groups with different levels of access privileges, can be employed. When device software is segmented, it prevents users from installing unwanted software that might cause interference into the corporate network.

3.3. Unauthorized access or disclosure

To safeguard the system from unauthorized access or disclosure, there must be an implementation of logical access control to an organization's critical and confidential information to reduce the impact when there is a security breach. This will be a control over who and what is accessed to a specific IoT resource as well as controlling the type of permitted access. To do this, the control must be embedded into the software such as applications, database management systems, operating systems, or implemented in network devices like routers [53]. For logical access to be effective the following must be done:

3.3.1. Access control model

According to [53], the access control model is one that must define the rules and guidelines of how objects are accessed by subjects. It must provide confidentiality and integrity while ensuring that there is accountability in three main ways: discretionary, mandatory, and role-based.

3.3.2. Information flow model

This is a model that must ensure information flow direction and security levels to ensure the confidentiality of information. By doing this, it will prevent the flow of higher security level information down to a lower level where a read permission allows a subject at a higher level to read an object at an equal or lower level, while a write permission allows a subject at a lower level to write up an object at an equal or lower level and the only subject that can make changes to the resource's security label will be the object, hence, ensuring confidentiality and not integrity.

3.3.3. Integrity model

Unlike the information flow model that ensures confidentiality, the integrity model does not ensure the same, but data integrity. In a read permission, integrity is achieved by allowing the subject to read when its integrity is equal or higher than the object and for write permission, the subject is allowed to write objects that it has an equal or lower integrity.

3.4. Improper disposal

Proper device disposal is critical for every organization. An improper disposal could potentially lead to data confidentiality issues with both legal and ethical implications. Therefore, having a policy that provides a proper cleaning or destruction of devices with sensitive and confidential data and licensed software on them is important and policies can be developed around the following areas:

3.4.1. Sanitization

Organizations must ensure that when devices are to be disposed of, the data on them must be removed using different methods such as overwriting and erasing data by utilizing methods prescribed by the National Institute of Standards and Technology (NIST) special publication 800-88 [54].

3.4.2. Degaussing

Proper methods must be used when storage media is subjected to a powerful magnetic field to remove the data on the media by rearranging the magnetic field on electronic media to completely erase its content. For example, computer hard drives and other electronic storage devices such as computer tapes store data within magnetic fields containing layers of magnetic materials [54].

4. Classification of breach incident types of dataset

The dataset used in this work consists of over 1600 recorded cases of data breaches, specifying the name of the covered entity (CE), the state the entity is located in, the number of individuals affected, date of submission of the breach, type of breach, location of breach, business associate present and the description of the breach from October 2009 to November 2017. To stay within the objective of predicting how human factors influence data breaches in organizations, only a selected number of parameters are considered; date of submission of the breach, the type of breach, and the description. The descriptive parameter narrates what led to the breach. A few of the records had missing values in all the columns except for the year (date of submission of breach). Such records were removed and not considered in this study. To clean data in a way that will be supported by quantitative analysis, the descriptive column, which is a string format was examined, record by record, case by case and where it was indicative of human factors such that if the underlying cause of the breach was directly due to human error or behavior, a score of 1 was assigned, otherwise 0. The data was then extracted according to the type of breach, the year the breach happened, and the number of human factors associated with it for that particular year. An assumption that even though undetected and unreported data breach incidences may be significant to the findings of this study, the reported data breach provides a confidence that typify data breach incidences in general.

An analysis of variance (ANOVA) for linear regression is used for the analysis of this study and the study uses Pearson's r which measures a linear relationship between two continuous variables. The regression line used is, $DATA = FIT + RESIDUAL$, that is:

$$(y_i - \overline{y}) = (\hat{y} - \overline{y}) + (y_i - \hat{y}_i) \tag{1}$$

Where the first term is the total variation in the dependent variable(s) y from the dataset, the second term is the variation in the mean observation, while the third term is the residual value, then square each of the given terms in equation (1) and add them over all the observations n, which gives the equation

$$\sum (y_i - \overline{y})^2 = \sum (\hat{y}_i - \overline{y})^2 + \sum (y_i - \hat{y}_i)^2 \tag{2}$$

Equation (2) can be rewritten as $SST = SSE + SSM$, where SST is the notation for the total sums of square, SSE error sums of square and SSM is the model sums of squares. The sum of the samples is equal to the ratio of the model's sums of square, $r^2 = SSM/SST$. With this, there is a formalization that the interpretation r^2 which explains the fraction of the variability in the data, that is explained by the regression model. The variance s_y^2 is given by:

$$\frac{\sum (y_i - \overline{y})^2}{n-1} = \frac{SST}{DFT} \tag{3}$$

Where DFT is the total degree of freedom.

$$MSM = \sum \frac{(\hat{y} - \overline{y})^2}{1} = \frac{SSM}{DFM} \tag{4}$$

Where DFM is a model degree of freedom. In equation (4) the mean square model (MSM) applies because the regression model has one explanatory variable x. The corresponding mean square error (MSE) is the estimate of the variance of the population of the regression line (σ^2)

$$\sum \frac{(y_i - \hat{y}_i)^2}{n-2} = \frac{SSE}{DFE} = MSE \tag{5}$$

The ANOVA calculations for the regression are shown in Table 1.

Table 1. ANOVA for Regression of Human Factors and Types of Breach.

Dependent Variable		Sum of Squares	df	Mean Square	F	Sig
HITi	Regression	2235.679	1	235.679	13.259	0.008[b]
	Residual	1180.321	7	168.617		
	Total	3416.000	8			
ImD	Regression	74.096	1	74.096	8.173	0.024[b]
	Residual	63.460	7	9.066		
	Total	137.556	8			
Loss	Regression	442.096	1	442.096	12.406	0.010[b]
	Residual	249.459	7	35.637		
	Total	691.556	8			
UAD	Regression	8720.635	1	8720.635	21.530	0.002[b]
	Residual	2835.365	7	405.052		
	Total	11556.000	8			
Theft	Regression	7804.797	1	7804.797	3.788	0.093[b]
	Residual	14423.426	7	2060.489		
	Total	22228.222	8			
Other	Regression	169.549	1	169.549	1.159	0.317[b]
	Residual	1024.007	7	146.287		
	Total	1193.556	8			
Unkwown	Regression	53.715	1	53.715	0.988	0.353[b]
	Residual	380.507	7	54.358		
	Total	434.222	8			

[b] Predictors: (Constant), HF.

$$r_{jk} = \frac{s_{jk}}{s_j s_k} = \frac{\sum_{i=1}^{n}(x_{ij} - \overline{x}_j)(x_{ik} - \overline{x}_k)}{\sqrt{\sum_{i=1}^{n}(x_{ij} - \overline{x}_j)^2}\sqrt{\sum_{i=1}^{n}(x_{ik} - \overline{x}_k)^2}} \quad (6)$$

Equation (6) is used to compute the correlation matrix of all the dependent variables. It is a Pearson correlation matrix between the variables x_j and x_k.

4.1. Characterization of breached incident types

The study characterizes the different types of breaches according to the breach type and its description as reported in the dataset:

4.1.1. Theft

These are breaches that occurred as a result of an electronic device being physically stolen and subsequently leading to the breach of information. Some of the devices that were stolen were desktop computers from front desk areas, backup tapes, stolen records from an entities office, laptops from offices, employee vehicles, USB drives, and external hard drives containing the PHI of several individuals.

4.1.2. Loss

A breach classified as loss is one which involved the misplacement of data that may have led to data being compromised. It is important to note, that the dataset does not explicitly refute the possibility of it being stolen, which would mean classifying it as theft. Neither does it imply loss itself a causation of other types of attacks. As a result, this work classifies the loss as a type of breach based on the reported cases in HIPAA. So where it is not known as to how data got missing and later being compromised is thereby classified as loss.

4.1.3. Unauthorized access or disclosure

Breaches that happened as a result of former workforce members, while still employed, downloading the names and certain personal information of its clients are classified as Unauthorized Access or Disclosure (UAD). UAD also includes employees or CE sharing PHI with authorized people. In a case study, some software vendors and business associates (BA) for the CE failed to disable a software switch, which allowed Google to index files on the CE's hosted website containing the electronic Protected Health Information (ePHI) of thousands of individuals. The ePHI included individual names, addresses, zip codes, Medicaid numbers, and primary care physician's names and addresses. Other cases of unauthorized access included employees sending Medicaid reports to their email, leading to a breach that affected over 270,00 individuals and the types of protected health information (PHI) involved in the breach included names, addresses, phone numbers, social security numbers, and their Medicaid identification numbers.

4.1.4. Improper disposal

A breach that happened as a result of the CE mailing envelopes containing PHI that arrived at the contracted provider's address damaged, with the contents missing is classified as Improper Disposal (ImD). Envelopes that were damaged at the postal facility where they were processed and contained member claim information of individuals, including members' names, identification numbers, claim numbers, dates of service, procedure codes, charges, and provider information is also ImD. Breaches that occurred as a result of employee erroneously distributing emails containing ePHI of thousands of individuals to the wrong recipients are classified as same. The last category of ImD are cases where electronic devices that were classified as "spoilt" were trashed with data still accessible on them, which led to a breach and after an investigation by the CE, it was found that the way the devices were disposed of was the cause of the breach.

4.1.5. Hacking or IT incident

Breaches that are classified as hacking or IT incidents (HITi) includes events such as a foreign Internet Protocol (IP) address accessing a CE's website, which contained a database containing PHI of clients or an unknown assailant associated with a foreign IP address that attempted to bypass the security mechanisms of a computer server of a former third party administrator and BA. A lot of individuals were affected by such breaches. The servers contained PHI regarding some of the CE participants such as names, addresses, social security numbers, and clinical information, including information regarding healthcare providers and types of service. When file servers at the entities' offices are compromised and impermissibly accessed and there is a compromise that potentially exposes the prescription records of thousands of individuals to an unauthorized source via electronic transmission, classified as HITi. In such cases the PHI involved in the breach included names, addresses diagnostic codes, name of medication prescribed, medication costs, and some social security numbers. Cases that included computer malware that was detected on the CE unencrypted billing software program are also classified as HITi. In these incidents the CE did not know when the malware entered its system. And thousands of individuals were potentially affected by this malware. The types of PHI involved included demographic, financial (claims information), and clinical information (diagnoses/conditions, medications, lab results, and other treatment in-

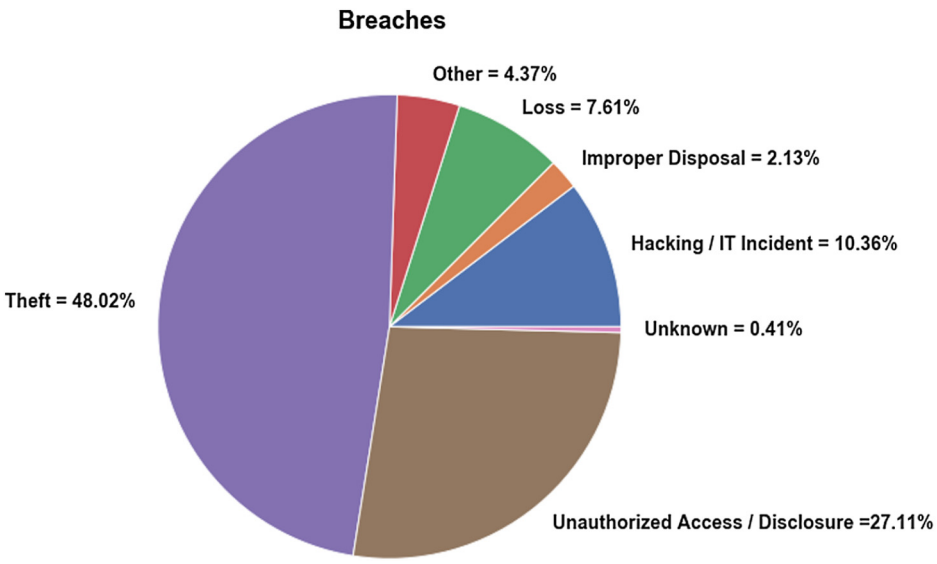

Fig. 2. Distribution of Breaches applied to Human Factors.

formation). Finally, instances where database web servers containing the ePHI of many clients were breached by an unknown external person(s) for use as game servers. The ePHI on the database web servers included names, dates of birth, types of x-rays, and dates of x-rays.

4.1.6. Other and unknown

Other and Unknown are also types of breaches in which the former are breaches that are not classified or close to two or more of the aforementioned types, and the later are breaches that even though detected that a breach had taken place, there was no way of knowing the actual breach that took place nor its proximity to any of the previous classifications. In other words, breaches that were reported with most of the parameters given but a missing type of breach were classified in this study as unknown.

5. Validation of human factors as a weakness in information security

An extraction of nine (9) observations was made from the 1722 records that were reported from 2009 to 2017 comprising of nine (9) years. The number of reported cases for each breach incidence computed for each year in separate columns and the overall number of human factors that were associated with that year's breach incidents were also computed in one column based on the descriptive column of the dataset. Human factors (HF) are used as an independent variable, while the different types of breaches used as dependent variables in a linear regression computation.

5.1. Results

5.1.1. Analysis of data breach incidents

The distribution of data breaches shown in Fig. 2 typifies the weaknesses that human factors pose in an information security set-up. In this case, human factors that led to a breach, attributed 48.02% to a breach of theft and 27.11% to UAD, giving them a combined share of 75.13%. Thus, theft and UAD are the two most common breaches to occur when a data breach is a result of human factors. HITi attacks formed 10.36% of breach cases when human factors are at the center of a breach. It may not be as large as theft and UAD, but still reasonably high. And clearly showing how human factors can easily make a 'secure' information security vulnerable to such attacks. The study also revealed that 7.61% of breaches were attributed to loss, 2.13% to ImD and others making up 4.37% of breaches caused by human factors.

The distribution illustrated in Fig. 2 only indicates the overall percentiles from 2009 to 2017. Fig. 3 shows the yearly distribution of human factors applied to the different types of breaches for each year as they were reported from 2009 to 2017.

- Theft
  The study revealed that human factors underpinning a breach of theft were closely distributed, with the lowest being 2016, 2009, and 2017 year periods accounting for 6.77%, 2.54%, and 0.21% respectively. And 2010, 2013, 2014 and 2015 recording 18.18%, 14.59% ,12.47% and 12.47% in that order while 2012 and 2011 had 13.11% each.
- Loss
  The breach of loss caused by human factors was very high in the year 2015 accounting for 28%, followed by 2014 with 18.67% and 14.67% attributed to 2013. The rest of the years' results were 13.33% for 2016, 12% for 2012, while 5.33%, 4%, 2.67% and 1.33% were the recorded for 2010, 2011, 2017 and 2009 respectively.
- Hacking or IT Incident
  In 2016 there was a huge rise in human factors concerning information security. The study revealed that 42.16% of human elements that led to HITi happened in 2016 while 2013, 2015, and 2014 had 14.71%, 13.73%, and 12.75% respectively. The results also showed 2012 with 7.84%, and 2011 and 2010 having 3.92% each, while 2017 accounting for 0.98%
- Improper Disposal
  In the years 2015 and 2013, the largest percentages for ImD breaches of 28.57% and 23.81% respectively were recorded, and in 2016, 19.05% of ImD breach as a result of human factors. The year 2014 assumed 14.29%. The rest of the years, 2012 and 2010 were 9.52% and 4.76% respectively.
- Unauthorized Access or Disclosure:
  Except 2009, the study revealed that from 2009 to 2017, a breach of UAD with the descriptive parameter of the dataset alluding to human factors as the problem had the year 2015 accounting for 27.34%, 2016 had 23.60% and 2014 with 17.98%. 2013 assumed 16.48% while 2012, 2011, 2017 and 2010 comprised of 6.74%, 3.75%, 2.62% and 1.50% respectively.
- Other
  Our study further showed that from the period under consideration, 2009 to 2017, five (5) of the years had other breaches that were caused by underlying human factors as reported in the de-

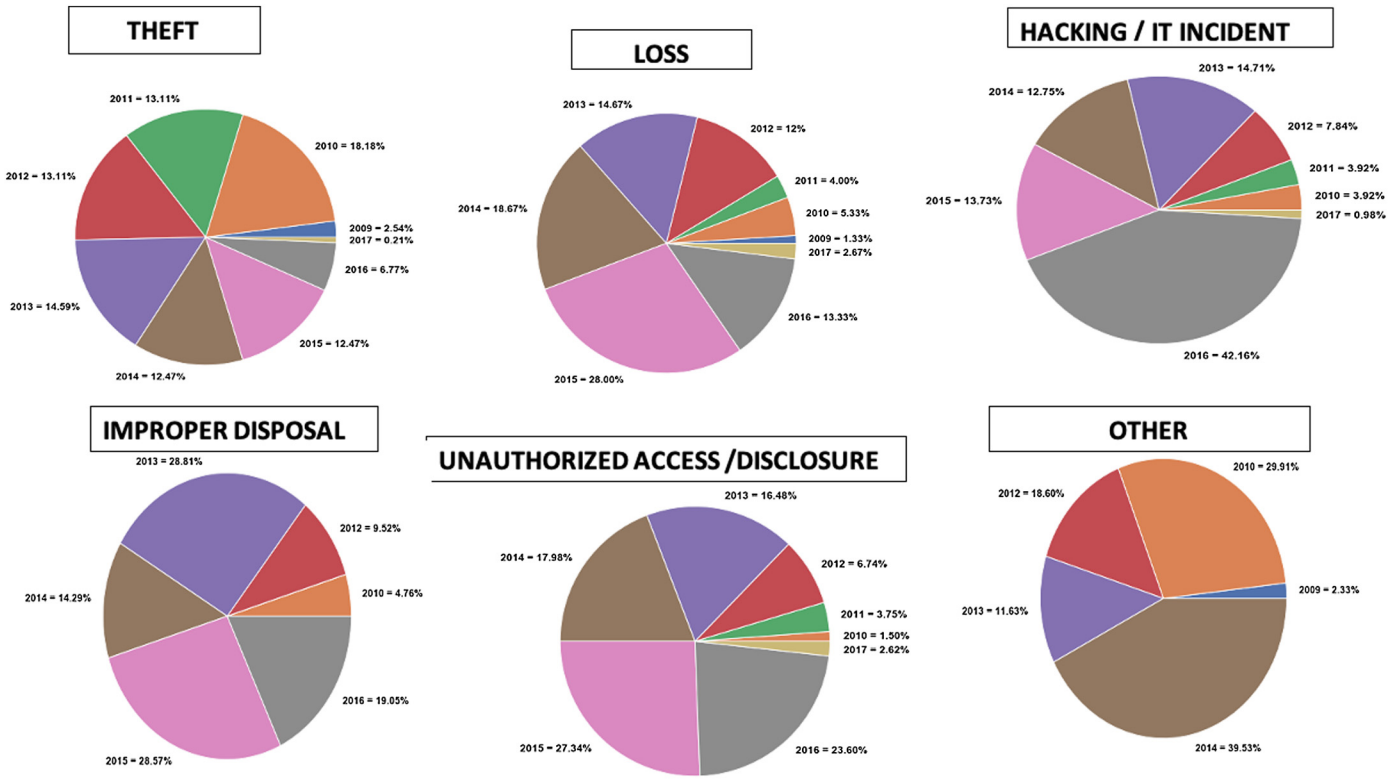

Fig. 3. Distribution of Yearly Human Factors Applied to Breaches 2009 to 2017.

Table 2. Model Summary of Human Factors and Types of Breach.

Dependent Variable	R	R Square	Adjusted R Square	Std. Error of the Estimate	Change Statistics		
					R Square Change	F Change	df
HITi	0.809[a]	0.654	0.605	12.985	0.654	13.259	1
ImD	0.734[a]	0.539	0.473	3.011	0.539	8.173	1
Loss	0.800[a]	0.639	0.588	5.970	0.639	12.406	1
UAD	0.869[a]	0.755	0.720	20.126	0.755	21.530	1
Theft	0.593[a]	0.351	0.258	45.393	0.351	3.788	1
Other	0.377[a]	0.142	0.019	12.095	0.142	1.159	1
Unknown	0.352[a]	0.124	-0.001	7.373	0.124	0.988	1

[a] Predictors: (Constant), HF.

scriptive parameters of the dataset, where 2014 had 39.53%, 2010 with 29.91%, 2012 with 18.60%, 2013 having 11.63% and 2.33% for 2009.

5.2. Relationship between human factors and data breach incidents

5.2.1. Human factors statistically predict breach types

From the ANOVA for the linear regression in Table 1, the following observations can be deduced between HF and the dependent variables. Firstly, with HITi, it can be established that HF could statistically and significantly predict HITi, the F statistics are equal to 13.259, with a distribution of [1, 7) and the probability of observing the value is greater or equal to 13.259 being less than 0.01. Secondly, ImD computation proved that HF could statistically and significantly predict ImD, giving an F statistics of 8.173 and a distribution of [1, 7), which gives a probability of observing the value that is greater or equal to 8.173 less than 0.05. Next it can be seen that the independent variable HF with the breach type loss, a dependent variable. The analysis shows that HF could statistically and significantly predict loss, with an F statistics being equal to 12.406, a distribution of [1, 7), and a probability of observing the value which is greater or equal to 12.406, to be less than 0.05. The analysis then shows that HF could statistically and significantly predict UAD, giving an F statistic of 21.530, and again a distribution of [1, 7). The probability of observing this value being greater or equal to 21.530 is less than 0.005. The next dependent variable measured with HF is theft. Unlike the previous analysis, HF could not statistically and significantly predict theft with an F statistic of 3.788 and a distribution of [1, 7). The probability of observing the value greater or equal to 3.788 is greater than 0.05. Now with an F statistic of 1.159 and a distribution of [1, 7), HF could not statistically and significantly predict others. The probability of observing the value that is greater or equal to 1.159 is greater than 0.05. Finally, the ANOVA for the linear regression between HF and unknown as indicated that HF could not statistically and significantly predict unknown, with $F[1, 7) = 0.988$, and the probability of observing the value being greater or equal to 0.988 is greater than 0.05.

5.2.2. Variation explained by human factors

Table 2 has the measurement of the proportion of the variations in the dependent variables that are explained by the independent variable, which in this case is HF. HF accounted for 60.5%, 47.3%, 58.8%, 72.0%, 25.8%, 1.9% and −0.01% of the explained variability in HITi, ImD, loss, UAD, theft, other and unknown respectively. In other words, non-human factors account for 39.5%, 52.7%, 41.2%, 28%, 74.2%, and 98.1% of the unexplained variability of HITi, ImD, loss, UAD, theft and

Table 3. Coefficients of Human Factors and Types of Breach.

		Unstandardized Coefficients		Standardized Coefficients		
Dependant Variable	independent variable	B	Std. Error	Beta	t	Sig.
HITi	(Constant)	−4.716	8.996		−0.524	0.616
	HF	0.270	0.074	0.809	3.641	0.008
ImD	(Constant)	1.995	2.086		0.956	0.371
	HF	0.049	0.017	0.734	2.859	0.024
Loss	(Constant)	4.453	4.136		1.077	0.317
	HF	0.120	0.034	0.800	3.522	0.010
UAD	(Constant)	13.943		−0.529	0.613	
	HF	0.534	0.115	0.869	4.640	0.002
Theft	(Constant)	32.902	31.447		1.046	0.330
	HF	0.505	0.260	0.593	1.946	0.093
Other	(Constant)	2.870	8.379		0.343	0.742
	HF	0.074	0.069	0.377	1.077	0.317
Unknown	(Constant)	2.104	5.108		0.412	0.693
	HF	0.042	0.042	0.352	0.994	0.353

Table 4. Correlations Matrix.

	Theft	Loss	HITi	ImD	UAD	Other	Unknown	HF
Theft	1							
Loss	0.850**	1						
HITi	0.087	0.440	1					
ImD	0.871**	0.884**	0.263	1				
UAD	0.283	0.664	0.901**	0.513	1			
Other	0.732*	0.679*	−0.133	0.786*	0.152	1		
Unknown	0.719*	0.746*	0.017	0.712*	0.364	0.746*	1	
HF	0.593	0.800**	0.809**	0.734*	0.869**	0.377	0.352	1

** Correlation is significant at the 0.01 level (2-tailed).
* Correlation is significant at the 0.05*level* (2-tailed).

others respectively. These non-human factors will require an empirical study to ascertain the degree to which they affect the aforementioned data breaches or attacks. This result clearly establishes that the success of attacks like hacking, unauthorized access are hugely influenced by human factors and so organizations must adopt an information security framework that comprehensively covers human factors.

5.2.3. Regression of human factors and the breach types

In the analysis on the data shown in Table 3, the regression for each of the dependent variables and HF, such that the equation predicted $HITi = -4.716 + 0.270x(x = HF)$. And so for each change or increase of human factors, the average change in the mean of hacking or IT incident is about 0.270. The regression equation for ImD, predicted $ImD = 1.995 + 0.049x(x = HF)$, indicating the average change in the mean of improper disposal to be about 0.049 for every increase in human factors. The dependent variable Loss saw an average change in its mean of about 0.120 for each change in human factors, given by its equation which predicted $Loss = 4.453 + 0.120x(x = HF)$. The regression equations also predicted that $UAD = -7.380 + 0.534x(x = HF)$, $theft = 32.902 + 0.505x(x = HF)$, $other = 2.870 + 0.074x(x = HF)$ and $unknown = 2.104 + 0.042x(x = HF)$.

5.2.4. Evaluation of the strength of the predictions

A Pearson correlation coefficient computed as shown in Table 4 to evaluate the strength of the relationship between HF the dependent variables that is when a breach occurred. The result indicated that there was a positive correlation between the HF and HITi, where $r = 0.809$, p being significant at 0.01. HF and ImD, the result indicated that there is a positive correlation between the two variables, where $r = 0.734$, and p is significant at 0.05. The correlation also showed that the strength of the relationship between HF and Loss, when a breach occurred, is positive, with a correlation $r = 0.8$ and p being significant at 0.01. The strength of the relationship between HF and UAD when a breach occurred is a strong positive correlation where $r = 0.869$ and p significant at 0.005. The correlation coefficient computed on HF and theft when a breach occurred indicates a positive correlation between the two variables, $r = 0.593$. However, p is not significant. Also the Pearson correlation coefficient computed on HF and other, when a breach occurred, indicated a minimal positive correlation between the two variables, $r = 0.377$ with p not being significant. Finally, HF and unknown breach types showed a minimal positive correlation between the two variables, $r = 0.352$, and yet again p is not significant. Therefore, an increase in four variables HITi, ImD, Loss, and UAD is positively correlated with an increase in HF, and increases in the remaining three variables, theft, other, and unknown did not correlate with increases in HF.

Table 5. Significance Level of Breach Types.

0.01	0.05	0.1	0.5
HITi=0.008	Loss=0.010	Theft=0.093	Other=0.317
	UAD=0.02		Unknown=0.353
	ImD=0.024		

5.2.5. T-test

Table 6 shows the results of a T-test. HiTi was a reported breach with underlining human factors (M = 24, SD = 20.664) in all the breach incidents reported as a whole, t(8) = 3.484, p = 0.008.ImD was a reported breach with underlining human factors (M = 7.222, SD = 4.147) in all the breach incidents reported as a whole, t(8) = 5.225, p = 0.02. Loss was also a reported breach with underlining human factors (M = 17.22, SD = 9.298) in all the breach incidents reported as a whole, t(8) = 5.557, p = 0.01. UAD was a reported breach with underlining human factors (M = 49.33, SD = 38) in all the breach incidents reported as a whole, t(8) = 3.894, p = 0.05. Again from the all the reported breach incidents, Theft was a reported breach with underlining human factors (M = 86.556, SD = 52.712), t(8) = 4.926, p = 0.01. Other reported breaches with underlining human factors (M = 10.78, SD = 12.215) in all the breach incidents reported as a whole computed t(8) = 2.647, p = 0.029, and Unknown reported breaches with underlining human factors (M = 6.56, SD = 7.367) in all the breach incidents reported as a whole computed t(8) = 2.669, p = 0.028.

Table 6. T-Test of Human Factors in Types of Breach.

	t	df	Sig(2 tail)	Mean Difference	Std. Div	95% Confident interval Lower	of the Difference Upper
HITi	3.484	8	0.008	24.000	20.664	8.12	39.88
ImD	5.225	8	0.01	7.222	4.147	4.03	10.41
Loss	5.557	8	0.01	17.222	9.298	10.08	24.37
UAD	3.894	8	0.05	49.333	38.007	20.12	78.55
Theft	4.926	8	0.01	86.556	52.712	46.04	127.07
Other	2.647	8	0.029	10.778	12.215	1.39	20.17
Unknown	2.669	8	0.028	6.556	7.367	0.89	12.22

6. Discussion

6.1. Socio-technical systems

According to Shin [55], the technical parts are usually the focus, when it comes to investigating a system and its applications. The normal approach is to rather highlight the technological interactions, while ignoring the people who use it. It is important to note that the working conditions affect the whole system. The type of system and environmental factors could include laws and regulations, market competition, or human factors. A holistic approach to system analysis is fundamental, since both technology and people define the overall performance of a system [56].

Sommerville and Dewsbury [57] also argue that there needs to be a cross-disciplinary framework that represents all the aspects of technological systems. This should include the technical equipment, the market, the people, and the society for which the system was created or adopted. Hence, failure is inevitable if all aspects of the system are not adequately examined.

The growth of IoT has major socio-technical implications not only for individuals, but also for organizations and society. IoTs have developed in a way that enables new ways of working, to increase safety and to facilitate coordination. This may however lead to interference with established work practices, undermine security, productivity, and individual satisfaction, creating an unforeseen impact on relations of behavior, power, and control. This is a question of socio-technical perspective. These perspectives, however, are rarely addressed in the development and research for IoT [55]. This study has clarified a practical point of view of the conceptualization of the IoT as a human-centered system, by clarifying a series of data breach incidents and how they affect computing, including IoT.

6.2. Vulnerability

The findings of this study provide evidence that human factors present a great threat to the information security system of organizations and in this case, it is most significant at 0.05 as shown in Table 5. It creates an avenue by which the security of an organization becomes vulnerable and ultimately making it easy for information to be compromised. There is an ever-increasing threat of data ex-filtration through loss, improper disposal, unauthorized user access or disclosure, and hacking or other information technology incidents. However, there is no evidence that human factors are a major player in data breach of theft, others, or unknown as depicted in Table 5. When the perceived value of data on the black market is very high, the probability that the threat to organizational data will decrease shortly is very low. Organizations or companies may use 'modern' techniques to frustrate breaches on a network or their information system. However, there will always be dedicated attacks on valuable data, due to their worth on the black market [6]. Thus, human factors become a critical point in the prevention of data breach and data ex-filtration.

Also sensitive data is shared among various participants and actors. Data or information sharing and external collaboration with other entities, which have become more and more common in today's businesses make data ex-filtration issues worse. Furthermore, as human resources are also becoming more mobile, where employees are allowed to work from outside the organization's premises, it increases the potential for data to be breached [58].

According to [50], it is a natural thing to want to make people behave in a way that results in more security. But the truth is that when there is an increase in user awareness, it does not often lead to sufficient secure behavior and the reality of behavioral transformation is a complex phenomenon altogether. It is difficult for people to quit habits that are detrimental to their health, despite the abundance of information on the risks associated with such habits or behaviors due to their short term gratification. This is also true and applicable in the world of information security. It does not mean changing habits or behaviors is not possible, but usually requires that one chooses a veracious intervention for the job at hand. It should be an intervention that changes a behavior which directly targets the actor or one that can indirectly affect the actor's behavior through technological or organizational solutions, fitting both purpose and use.

6.3. Increasing user security awareness

User security awareness is critical to the overall security of any organization. Information security awareness should be a preventive measure that must be used by organizations to firmly establish correct security procedures and security principles in the minds of all employees. It is essential because any security technique can be misused or misconstrued, thereby not benefiting from its real value. Increased awareness minimizes user-related security threats and maximizes the efficiency of security techniques from the human point of view. [26]
To increase user security awareness, McLean provides a proposal which 'sells' information security to people via campaigns. These campaigns can prove very useful in terms of security education, and providing a positive impetus to information security. Thereby maintaining the importance of security in the eyes of all employees. Campaigns are good measures for improving attitudes in organizations [59]. Campaigns can also be based on the Hammer theory, which aims to make information security an 'in' topic in an organization. The theory is such that when a new concept is properly introduced in an organization, everybody is interested to use it [60]. Campaigns and 'in' topics can be used synchronically within awareness programs, and they are critical in providing incentives for end-users and in invigorating the importance of these factors in people's minds.

7. Limitations

A notable limitation in this paper is the sole focus on only reported breaches of HIPAA. The study does not also attempt to identify behavioral elements that may play a critical role in behavior, which may lead to a breach, and so these factors have not been discussed in the current study. Furthermore, sociological forces that may shape an individual's perceptions of organizational abuse and discipline have not been considered.

8. Conclusion

Even though there are very good technologies that organizations can employ to protect sensitive data from breaches on their network, it

only solves one part of the problem. As long as human beings are a part of IoT, good information security solutions must incorporate human factors in them. This paper has, through the analysis, predicted the relationship between data breach incidents and human factors that provides an understanding of how human factors affect information security. The study has also shown the strength of the relationship between human factors and the different types of data breach incidents. The paper also proposed a framework that integrates technology and human factors that may be useful in reducing the number of data breach incidents due to an increase in human factors.

Author contribution statement

K. Hughes-Lartey: Conceived and designed the experiments; Performed the experiments; Analyzed and interpreted the data; Contributed reagents, materials, analysis tools or data; Wrote the paper.

M. Li, F. E. Botchey: Conceived and designed the experiments; Analyzed and interpreted the data.

Z. Qin: Conceived and designed the experiments; Analyzed and interpreted the data; Contributed reagents, materials, analysis tools or data.

Funding statement

This work was supported in part by the National Natural Science Foundation of China (No. 61672135), the Frontier Science and Technology Innovation Projects of National Key RD Program (No. 2019QY1405), the Sichuan Science and Technology Innovation Platform and Talent Plan (No. 20JCQN0256), and the Fundamental Research Funds for the Central Universities (No. 2672018ZYGX2018J057).

Data availability statement

Data associated with this study has been deposited at https://www.kaggle.com/archangell/hipaa-breaches-from-20092017.

Declaration of interests statement

The authors declare no conflict of interest.

Additional information

No additional information is available for this paper.

Acknowledgements

This work was supported in part by the National Natural Science Foundation of China (No. 61672135), the Frontier Science and Technology Innovation Projects of National Key R&D Program (No. 2019QY1405), the Sichuan Science and Technology Innovation Platform and Talent Plan (No. 20JCQN0256), and the Fundamental Research Funds for the Central Universities (No. 2672018ZYGX2018J057).

References

[1] L. Atzori, A. Iera, G. Morabito, The Internet of things: a survey, Comput. Netw. 54 (15) (2010) 2787–2805.
[2] B. Guo, D. Zhang, Z. Wang, Z. Yu, X. Zhou, Opportunistic iot: exploring the harmonious interaction between human and the internet of things, J. Netw. Comput. Appl. 36 (6) (2013) 1531–1539.
[3] L. Yang, S.-H. Yang, L. Plotnick, How the internet of things technology enhances emergency response operations, Technol. Forecast. Soc. Change 80 (9) (2013) 1854–1867.
[4] J. Wang, D. Rosca, W. Tepfenhart, A. Milewski, M. Stoute, Dynamic workflow modeling and analysis in incident command systems, IEEE Trans. Syst. Man Cybern., Part A, Syst. Hum. 38 (5) (2008) 1041–1055.
[5] R. Nicolescu, M. Huth, P. Radanliev, D. De Roure, Mapping the values of iot, J. Inf. Technol. 33 (4) (2018) 345–360.
[6] T. Floyd, M. Grieco, E.F. Reid, Mining hospital data breach records: cyber threats to us hospitals, in: 2016 IEEE Conference on Intelligence and Security Informatics (ISI), IEEE, 2016, pp. 43–48.
[7] E. Nakashima, Security firm finds link between china and anthem hack, Washington Post (2015).
[8] Z. Qin, Y. Zhang, S. Meng, Z. Qin, K.-K.R. Choo, Imaging and fusing time series for wearable sensor-based human activity recognition, Inf. Fusion 53 (2020) 80–87.
[9] O.-A. Kwabena, Z. Qin, T. Zhuang, Z. Qin, Mscryptonet: multi-scheme privacy-preserving deep learning in cloud computing, IEEE Access 7 (2019) 29344–29354.
[10] Z. Qin, L. Hu, N. Zhang, D. Chen, K. Zhang, Z. Qin, K.-K.R. Choo, Learning-aided user identification using smartphone sensors for smart homes, IEEE Int. Things J. 6 (5) (2019) 7760–7772.
[11] Z. Qin, Y. Wang, H. Cheng, Y. Zhou, Z. Sheng, V.C. Leung, Demographic information prediction: a portrait of smartphone application users, IEEE Trans. Emerging Topics Comput. 6 (3) (2016) 432–444.
[12] D. Chen, N. Zhang, Z. Qin, X. Mao, Z. Qin, X. Shen, X.-Y. Li, S2m: a lightweight acoustic fingerprints-based wireless device authentication protocol, IEEE Int. Things J. 4 (1) (2016) 88–100.
[13] H. Xiong, Y. Zhao, L. Peng, H. Zhang, K.-H. Yeh, Partially policy-hidden attribute-based broadcast encryption with secure delegation in edge computing, Future Gener. Comput. Syst. 97 (2019) 453–461.
[14] J. Sun, Y. Bao, X. Nie, H. Xiong, Attribute-hiding predicate encryption with equality test in cloud computing, IEEE Access 6 (2018) 31621–31629.
[15] C. Xiao, D. Han, Y. Ma, Z. Qin, Csigan: robust channel state information-based activity recognition with gans, IEEE Int. Things J. 6 (6) (2019) 10191–10204.
[16] Z. Qin, W. He, F. Deng, M. Li, Y. Liu Srprid, Pedestrian re-identification based on super-resolution images, IEEE Access 7 (2019) 152891–152899.
[17] J.J. Gonzalez, A. Sawicka, A framework for human factors in information security, in: Wseas International Conference on Information Security, Rio de Janeiro, 2002, pp. 448–487.
[18] Z. Qin, G. Huang, H. Xiong, Z. Qin, K.-K.R. Choo, A fuzzy authentication system based on neural network learning and extreme value statistics, IEEE Trans. Fuzzy Syst. (2019).
[19] B. Schneier, Schneier on Security, John Wiley & Sons, 2009.
[20] R. Klahr, J. Shah, P. Sheriffs, et al., Cyber Security Breaches Survey 2017: Main Report, 2017.
[21] N.S. Safa, R. Von Solms, L. Futcher, Human aspects of information security in organisations, Comput. Fraud Secur. 2016 (2) (2016) 15–18.
[22] C.I. Canfield, B. Fischhoff, A. Davis, Quantifying phishing susceptibility for detection and behavior decisions, Hum. Factors 58 (8) (2016) 1158–1172.
[23] M. Evans, Y. He, L. Maglaras, H. Janicke, Heart-is: a novel technique for evaluating human error-related information security incidents, Comput. Secur. 80 (2019) 74–89.
[24] D. Liginlal, I. Sim, L. Khansa, How significant is human error as a cause of privacy breaches? An empirical study and a framework for error management, Comput. Secur. 28 (3–4) (2009) 215–228.
[25] A.E. Speed, B.L. Woo, C.G. Kouhestani, J.J. Stubbs, G.C. Birch, Human factors in security, in: 2018 International Carnahan Conference on Security Technology (ICCST), IEEE, 2018, pp. 1–5.
[26] H.A. Kruger, L. Drevin, S. Flowerday, T. Steyn, An assessment of the role of cultural factors in information security awareness, in: 2011 Information Security for South Africa, IEEE, 2011, pp. 1–7.
[27] Archangell, Hipaa breaches from 2009-2017-complete archives of reported hipaa breaches, https://www.kaggle.com/archangell/hipaa-breaches-from-2009201, 2017.
[28] M.L. Network, Fact sheet medical privacy of protected health information, https://www.cms.gov/outreach-and-education/medicare-learning-network-mln/mlnproducts/downloads/hipaaprivacyandsecurity.pdf, 2018.
[29] D. Guardian, Definitive guide to u.s. state data breach laws, https://web.archive.org/web/20180905170224/https://info.digitalguardian.com/rs/768-OQW-145/images/the-definitive-guide-to-us-state-data-breach-laws.pdf, 2018 [internet archive].
[30] R. Hamdan, Human factors for iot services utilization for health information exchange, J. Theor. Appl. Inf. Technol. 96 (8) (2018).
[31] Z.-K. Zhang, M.C.Y. Cho, C.-W. Wang, C.-W. Hsu, C.-K. Chen, S. Shieh, Iot security: ongoing challenges and research opportunities, in: 2014 IEEE 7th International Conference on Service-Oriented Computing and Applications, IEEE, 2014, pp. 230–234.
[32] P. Radanliev, D.C. De Roure, J.R. Nurse, R.M. Montalvo, S. Cannady, O. Santos, P. Burnap, C. Maple, et al., Future developments in standardisation of cyber risk in the internet of things (iot), SN Appl. Sci. 2 (2) (2020) 169.
[33] P. Radanliev, D.C. De Roure, C. Maple, J.R. Nurse, R. Nicolescu, U. Ani, Cyber Risk in iot Systems, 2019.
[34] L. Neumann, Cyber Security. Simply. Make It Happen, 2017.
[35] J. D'Arcy, A. Hovav, D. Galletta, User awareness of security countermeasures and its impact on information systems misuse: a deterrence approach, Inf. Syst. Res. 20 (1) (2009) 79–98.
[36] M. Alotaibi, S. Furnell, N. Clarke, Information security policies: a review of challenges and influencing factors, in: 2016 11th International Conference for Internet Technology and Secured Transactions (ICITST), IEEE, 2016, pp. 352–358.
[37] S. Furnell, K.-L. Thomson, From culture to disobedience: recognising the varying user acceptance of it security, Comput. Fraud Secur. 2009 (2) (2009) 5–10.

[38] R.E. Crossler, A.C. Johnston, P.B. Lowry, Q. Hu, M. Warkentin, R. Baskerville, Future directions for behavioral information security research, Comput. Secur. 32 (2013) 90–101.
[39] J. Shropshire, M. Warkentin, A. Johnston, M. Schmidt, Personality and it security: an application of the five-factor model, in: AMCIS 2006 Proceedings, 2006, p. 415.
[40] M. McBride, L. Carter, M. Warkentin, Exploring the role of individual employee characteristics and personality on employee compliance with cybersecurity policies, RTI Int.-Inst. Homeland Security Solut. 5 (1) (2012) 1.
[41] R.W. Proctor, J.D. Proctor, Sensation and perception, in: Handbook of Human Factors and Ergonomics, 2006, pp. 51–88.
[42] Q. Hu, T. Dinev, P. Hart, D. Cooke, Managing employee compliance with information security policies: the critical role of top management and organizational culture, Decis. Sci. 43 (4) (2012) 615–660.
[43] M. Hanley, T. Dean, W. Schroeder, M. Houy, R.F. Trzeciak, J. Montelibano, An Analysis of Technical Observations in Insider Theft of Intellectual Property Cases, 2011.
[44] D. Cappelli, A. Moore, R. Trzeciak, T.J. Shimeall, Common Sense Guide to Prevention and Detection of Insider Threats, CERT, Jan 2009.
[45] E. Kowalski, D. Cappelli, A. Moore, Insider threat study: Illicit cyber activity in the information technology and telecommunications sector, Tech. Rep., Carnegie-Mellon Univ Pittsburgh PA Software Engineering Inst., 2008.
[46] S. Pahnila, M. Siponen, A. Mahmood, F. Oulun, et al., Employees Behavior Towar ds Is Security Policy Compliance University of Oulu Department of Information Processing, 2007.
[47] C. Colwill, Human factors in information security: the insider threat–who can you trust these days?, Inf. Secur. Tech. Rep. 14 (4) (2009) 186–196.
[48] M. Ceniceros, Business innovation the internet of things ecosystem: the value is greater than the sum of its “things”, https://www.business2community.com/business-innovation/internet-things-ecosystem-value-greater-sum-things-0829370, 2014. (Accessed 30 July 2020).
[49] Internet of things infrastructure backbone, https://jungleworks.com/internet-of-things-infrastructure-backbone/. (Accessed 30 July 2020).
[50] K. Young-mclear, G. Wyman, Advances in Human Factors in Cybersecurity, 2016.
[51] M. Bhardwaj, G. Singh, Types of hacking attack and their countermeasure, Int. J. Educ. Plann. Admin. 1 (1) (2011) 43–53.
[52] M.T. Raggo, Mobile Data Loss: Threats and Countermeasures, Syngress, 2015.
[53] R.F. Rights, Global information assurance certification paper, 2003.
[54] R. Kissel, M.A. Scholl, S. Skolochenko, X. Li, Sp 800-88 rev. 1. Guidelines for Media Sanitization, 2006.
[55] D. Shin, A socio-technical framework for internet-of-things design: A human-centered design for the internet of things, Telemat. Inform. 31 (4) (2014) 519–531.
[56] H.R. Schindler, J. Cave, N. Robinson, V. Horvath, P. Hackett, S. Gunashekar, M. Botterman, S. Forge, H. Graux, Europe’s policy options for a dynamic and trustworthy development of the internet of things, Smart 2012 (0053) (2013).
[57] I. Sommerville, G. Dewsbury, Dependable domestic systems design: a socio-technical approach, Interact. Comput. 19 (4) (2007) 438–456.
[58] L. Cheng, F. Liu, D.D. Yao, Enterprise data breach: causes, challenges, prevention, and future directions, Wiley Interdiscip. Rev. Data Min. Knowl. Discov. 7 (5) (2017).
[59] M.T. Siponen, J. Kajava, Ontology of organizational it security awareness-from theoretical foundations to practical framework, in: Proceedings Seventh IEEE International Workshop on Enabling Technologies: Infrastucture for Collaborative Enterprises (WET ICE’98) (Cat. No. 98TB100253), IEEE, 1998, pp. 327–331.
[60] J. Kajava, M.T. Siponen, Effectively implemented is security awareness-an example from university environment, in: Proceedings of IFIP-TC, Vol. 11, 1997, pp. 105–114.