Khazar University

Department of Computer Science

Master of Computer Science

Amir Masoud Rahmani

Adepoju Alamu Luke

“The impact of security and privacy issues on data management in fog Computing"

In partial fulfillment of the requirements for the degree of Master of Science in Engineering in Computer Science

May, 2022

Abstract

With the increased growth of the application domains of IoT and the associated volumes of data generation, IoT systems are complicated and have small storage and recycling capacity. The cloud, a primary IoT storage medium with countless benefits, is not ideal for processing real time IoT data without delays. Capacity of data generated by IoTs keep increasing rampantly with associated security risks and privacy-preserving problems. Therefore, privacy maintenance, confidentiality and integrity of user's data, improved latency and bandwidth restrictions are some of the major respective challenges of cloud computing.

Fog computing is therefore a novel paradigm and an extension of the cloud. Which aims to improve cloud efficiency by enabling IoTs to locally process data before cloud transmission. However, some of the issues present in cloud such as the establishment of connection between edge devices often raise security and privacy concerns are also inherent in fog.

The goal of this study, however, is to look at the state of data management security and privacy in a fog computing environment by reviewing existing security frameworks and data privacy procedures. This study lays bare the security vulnerabilities that exist inside the fog environment, creating hazards to user data privacy and security, and in lieu of that, this study incorporates features of data in addition to the acquired facts and statistics. Privacy-preservation is key to the continued use of services within the context of internet usage, as a result respondents indicated that they were experienced internet users who have been using the internet and its associated resources for various purposes, however respondents neither agreed nor disagreed with the possibility of the tracking or monitoring of their usage of the internet. The perception of respondents influenced the usage of the internet and various computing devices.

Abstrakt

IoT-nin tətbiq domenlərinin artması və əlaqəli həcmdə məlumat istehsalı ilə IoT sistemləri mürəkkəbdir və kiçik saxlama və emal qabiliyyətinə malikdir. Saysız-hesabsız faydaları olan əsas IoT saxlama vasitələri olan bulud, gecikmədən real vaxt IoT məlumatlarının işlənməsi üçün əlverişli deyil. IOT-lar tərəfindən istehsal olunan məlumatların qabiliyyəti əlaqəli təhlükəsizlik riskləri və şəxsi məlumatların qorunması problemləri ilə geniş şəkildə artmaqda davam edir. Buna görə də, şəxsi məlumatların qorunması, istifadəçinin məlumatlarının məxfiliyi və nöqsansızlığı, təkmilləşdirilmiş yubiley və bant genişliyi məhdudiyyətləri bulud hesablamasının əsas müvafiq

çətinliklərindən bəziləridir. Duman hesablaması buna görə də roman paradiqması və buludun genişləndirilməsidir. Hansı ki, bulud ötürmədən əvvəl IOT-lərin məlumatları yerli-yerində emal etməsinə imkan verməklə bulud effektivliyini artırmaq məqsədi daşıyır. Lakin buludda olan bəzi məsələlər, məsələn, kənar cihazlar arasında əlaqənin qurulması çox vaxt təhlükəsizliyi artırır və şəxsi məlumatlarla bağlı narahatlıqlar da dumana səbəb olur.

Bu araşdırmanın məqsədi isə mövcud təhlükəsizlik çərçivələrini və məlumatların gizliliyi prosedurlarını nəzərdən keçirərək, dumanlı hesablama mühitində məlumatların idarə edilməsi təhlükəsizliyi və məxfilik vəziyyətinə baxmaqdır. Bu araşdırma dumanlı mühit daxilində mövcud olan təhlükəsizlik zəifliklərini çılpaq qoyur, istifadəçi məlumatlarının gizliliyi və təhlükəsizliyi üçün təhlükələr yaradır. Bunun qarşısında isə bu araşdırma əldə edilmiş faktlar və statistika ilə yanaşı, məlumatların da xüsusiyyətlərini özündə cəmləşdirib. Məxfiliyin qorunması internet istifadəsi kontekstində xidmətlərdən davamlı istifadə üçün əsasdır. Belə ki, nəticədə respondentlər müxtəlif məqsədlər üçün internetdən və onunla əlaqəli resurslardan istifadə edən təcrübəli internet istifadəçiləri olduqlarını bildiriblər. Lakin respondentlər onların internetdən istifadəsinin izlənilməsi və ya monitorinqinin mümkünlüyü ilə nə razılaşıblar, nə də razılaşmayıblar. Respondentlər haqqında təsəvvür internetin və müxtəlif hesablama cihazlarının istifadəsinin təsirinə səbəb olmuşdu.

Table of Contents

List of Tables

List of Figures

Chapter One

1.0 Background of study

The world is estranged in a paradigm of the existence and generation of excess data, but the means to make good use of it has not been developed and has not achieved full utilization if developed. The evolution of data over the past few decades has further made it exceedingly difficult to quantify the exact volume of data generated, which is beneficial to making accurate estimations and predictions. This abundance can be attributed to the growing plethora of pervasive computing devices capable of operating autonomously or under guidance.

Internet of Things (IoT) is a network of homogeneous nodes with sensory intelligence and can exchange data over the internet (Pagallo et al., 2017). Defined by Clark (2016), IoT is a giant network of devices and people interacting in real time to generate data through embedded sensors. To this end, the IoT can be referred to as any device or devices that have sensor intelligence and computational capabilities in addition to a capability to communicate with others over the internet. Sensory nodes that operate autonomously or with guidance have outgrown the traditional connection of two or more computers to mobile phones, smart monitoring devices, tablets, smart TV's, smart refrigerators and other smart gadgets. IoT devices are primarily expected to collect information about their environment, be able to process and analyze the information gathered, and be able to transmit messages over the Internet (Pagallo et al., 2017).

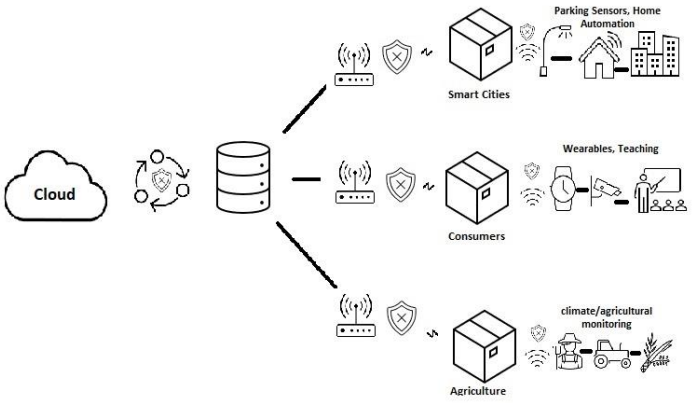

Figure 1.1. Structure of an IoT ecosystem

According to statista.com the globally estimated number of actively connected IoTs between 2010 and 2015 was 12 billion devices, shortly afterwards the number increased to 40.4 billion devices. Contrasting these figures to non-connected IoTs which has experienced a stunt growth over the

years, devices are more connected now more than ever. Regardless, the number of actively connected IoTs is further expected to plumate to about 105.3 billion devices whilst the number of non-connected devices increases by 1% statista.com (2022). However, previous estimates by Georgiev (2021) were 20.4 billion by 2020 and 75 billion by 2025 respectively. The nearness of the estimates by Georgiev have been surpassed in multiple folds as estimated by statista.com Georgiev (2021).

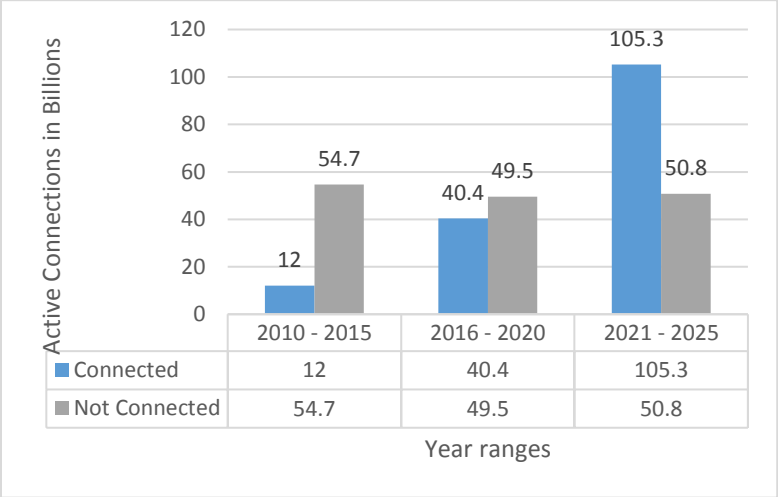

Figure 1.2. Trend of IoT growth (statista.com)

The proliferation of IoTs creates a new dimension of consideration, big data which is regarded as a major paradigm that counters the deployment and application domains of IoTs. The interaction of these pervasive computing devices is expected to generate vast amounts of data, since they are computationally and storage intensive. Miri & Pazzi, (2021) estimate this quantity as being ten times higher than projected. Big data comes from the interaction of many IoTs in an IoT ecosystem, and to make accurate and reliable decisions, businesses and corporations need this big data.

Processing and storing data becomes the next ideal paradigm after generation from the interaction of the IoTs which is primarily in the cloud. The cloud described by NIST, the National Institute of Standards and Technology is the national standards institute as the provision of on-demand services to customers through an actively established internet connection (Mell and Grance, 2011). Storage and processing are among the functionalities provided by cloud service providers (Akhtar

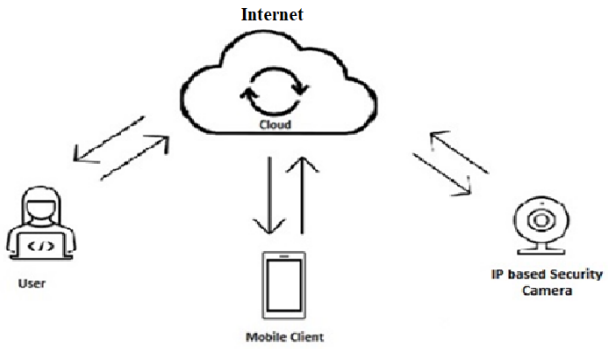

Figure 1.3: Structure of a public cloud

(Valkonen, 2013)

However, cloud computing as an ideal medium for IoT and the division of cloud into Public, private, and hybrid cloud computing models exist, as well as the availability of a variety of platforms that suit users' needs, such as infrastructure-as-a-service, platform-as-a-service, and software-as-a-service. Regardless of the benefits associated with cloud usage and the clouds ability to attempt to cope with the instantaneous and real-time processing, storage and analyzing needs of data from IoTs, it is not ideal for time sensitive data (Miri & Pazzi, 2021). Privacy and security associated with cloud computing emerges as a serious challenge for cloud computing as the management of data presents a unique problem because of the enormity of the volume of data generated by IoTs, the unpredictable nature and complexities associated with cloud transmissions, processing constraints due to bandwidth and latency setbacks. Moreover, data privacy, security, availability, location, and data transmission are the main concerns in cloud computing (Akhtar et al, 2021).

The latency, bandwidth and other lapses identified in cloud computing has necessitated the development and propounding of a novel paradigm of computing. A paradigm with the ability to cater for the network requirements of IoTs and also support instantaneous processing of data from IoTs without disrupting the existing cloud infrastrature. Fog computing intends to improve the efficiency of cloud computing by placing servers at the edge of the network. These servers act as intermediaries between connected devices and the cloud thereby removing the bottleneck of overloading cloud servers increasing efficiency (Mukherjee, et al., 2017).

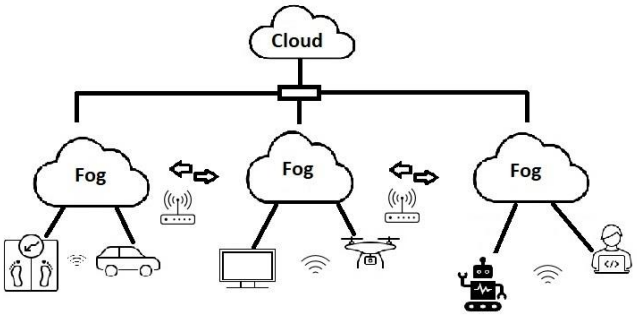

Figure 1.4. Structure of Fog

Fog computing is distinguished from its parent environment, the cloud in its placement of servers and transit channels in that it is more decentralized than cloud computing. In fog computing, heterogeneous and portable systems can locally process data within multiple hardware (Khan et al, 2017). The placement and ability of fog systems to process, storage and transmit data from IoT sources makes it an ideal choice for IoTs. However, fog computing although an extension of cloud, possesses its own security threats at all the bridge levels and among the connected devices. An instance of security and privacy attacks in the cloud, intensifies within fog systems regarding fog data, network, and malware (Khan et al, 2017).

1.1 Problem statement

With the increasing growth of IoT applications and the associated volumes of data generation, IoT systems are complicated and have small storage and recycling capacity (Abdulqadir, et al., 2021). The cloud serves as a primary storage medium for IoTs and although there are some benefits associated with cloud computing such as the convenience of relatively accessing configurable computing resources (Akhtar et al, 2021), massive scalability (Carlin & Curran, 2011), easy data recovery as well as multiple access to data schemes (Mukherji & Shashwat, 2015), the cloud is not ideal for processing real time data without delays for time sensitive applications (Miri and Pazzi, 2021). These sensitive applications require low latencies for real time data exchanges and as such necessitate the provision and placement of computational services and resources next to them (Nadeem and Saeed, 2016). This need has necessitated the proposing of several approaches to mitigate the problems experienced by IoTs during data processing and transmissions to the cloud such as the introduction of fog computing.

Fog computing intends to improve the efficiency of cloud computing by providing IoTs the ability to locally process data before cloud transmission. (Abdulqadir, et al., 2021). Fog removes the

bottleneck of overloading cloud servers increasing efficiency (Mukherjee, et al., 2017). However, some of the issues present in cloud are also inherited by fog computing (Khanum et al., 2021). The establishment of connection between edge devices often raise security and privacy concerns such as the susceptibility of the medium of data exchange to to malicious attacks intrusions such Denial of Service attacks, eavesdropping, data hijacking, and masquerade attacks (Yassein, et al., 2020; Mukherjee, et al., 2017; Stojmenovic, We, Huang, & Luan, 2015). These attack vectors lay the groundwork for more privacy and security intrusions, with additional concerns arising from fog networks' limited visibility, fog servers' ineffective threat detection, the lack of mechanisms allowing users to engage in selective data gathering, virtualization issues, and fog node masquerading because these exposed faults or vulnerabilities are exploited to become entry points for more attacks (Chanal and Kakkasageri, 2015).

Meanwhile, the inclusion of data changes the dimension of the mentioned attack vectors although there are presently limited studies on security and privacy in fog computing (Yi, et al., 2015). Security attacks often damages the integrity and quality of data in fog networks (Kapil, Agrawal, & Khan, 2018). The data generation capacity of IoTs keep increasing rampantly with associated security risks and privacy-preserving problems. Maintaining the privacy, confidentiality and integrity of user's data is one of the major challenges in big data (Fang et al., 2017). Furthermore, the heightening of security and privacy problems by the 4Vs (volume, variety, velocity, and veracity) of big data (Zhang, 2018) renders traditional protection systems difficult (Yin, Zhang, Xi, & Wang, 2017) and obsolete in some cases. The effects are often the leakage of user's data leading to Internet harassment, and more serious is the loss of property lead to all kinds of nuisance in life, (Li, 2021). However, the privacy and security of data should focus on all aspects of security related to big data i.e., security of the infrastructure, network security, data security, privacy, log management, etc (Bhatia & Sood, 2018).

1.2 Research Objectives

The aim of this study is to explore the impact of security and privacy issues on data management in fog computing. The objectives of the study are to:

1. Examine the existing security frameworks for data management in fog computing
2. Outline the security and privacy threats to the management of data in fog computing
3. Recommend appropriate solutions for data management in fog computing

1.3 Research Questions

The following questions were asked to guide the study:

1. What are the available security frameworks for data management in fog computing?
2. What are the security and privacy threats to the management of data in fog computing?
3. What are the most appropriate privacy and data security solutions for the data in fog computing?

1.4 Purpose of the Study

The purpose of this study is to investigate into the state of security and privacy of data management in fog computing environment through the review of existing security frameworks and privacy operations on data to provide insights about the safety and security measures adopted by service providers through the fog computing environment to maintain confidentiality and integrity of user's data.

1.5 Organization of the Study

The study is divided into five chapters. Chapter 1 presents a background of the study, a problem statement, a research question, a significance statement, a limitation, and a plan of organization. The relevant related literature is reviewed and presented in Chapter 2. The research design and methodology are presented in Chapter 3. Results and conclusions are presented in chapter 4. The conclusion, recommendations, and suggestions for further study are presented in Chapter 5.

Chapter Two

2.1 Review of Related Literature

The focus of this chapter is to review literature related to the study. The review of existing literature is categorized mainly according to the themes of the specified objectives: examining the existing security frameworks for data management in fog computing, outlining the security and privacy threats to the management of data in fog computing and finally outlining of appropriate solutions for data management in fog computing. The 21st century is earmarked with data abundance of originating sources such as sensors, mobile devices, video/audio, networks, log files, transactional applications, the web, and social media (IBM, n.d). The enormity of the data generated is so huge the traditional processing software and approaches find it extremely difficult to cope with. According to a publication by Lori Lewis on AllAccess, every minute on the internet is

characterized with the generation of data in millions, with over 21 million texts messages sent and received, 69 million messages exchanged via WhatsApp and Facebook Messenger as well as about 198 million Emails exchanged, 3.4 million snaps from Snapchat, 3 million views on Imgur (Lewis, 2021).

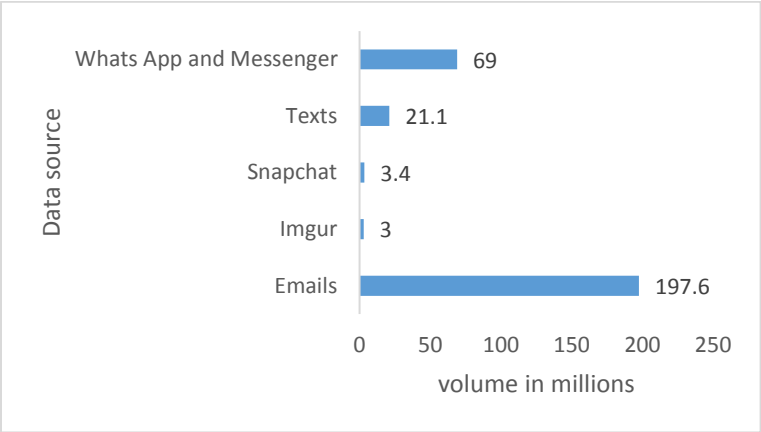

Figure 2.1. Data statistics on the Internet (Lewis, 2021)

According to Desjardins (2019), the active and passive interaction of pervasive computing devices generates a voluminous amount of accumulated data in the digital universe estimated at 4.4 zettabytes of data in 2013. This quantity of data has increased exponentially over the past years, with the total approximation of data around 2.5 quintillion bytes each day in 2018. Still, in 2020, the total volume of data generated was about 44 zettabytes. This approximation figure is expected to grow to 175 zettabytes of data in 2025 further (Desjardins, 2019). See, (2021) also postulates that the quantity estimation of global data generation between 2010 and 2015 was about 50.5 zettabytes and this number has further skyrocketed to about 183 zettabytes between 2016 and 2020 and is further projected to plumate to about 624 zettabytes by 2025 (See, 2021).

This data enormity in such unanticipated volumes over the years has resulted to the coining of the term “big data” to describe the quantity of data generated and in circulation. Big data results from the extensive and voluminous generation of data by distributed computing devices through an increase in its usage and interaction. The term big data is a complex word blend of 'big' and 'data' described from the quantity of data from generation sources, storage, and processing capacities of the sources. Big data is further described by Vuleta, (2021) as “the systematic quantification, extraction, and analysis of data from large and intricate sources per unit of time the traditional data processing software and approaches cannot easily process” (Vuleta, 2021). Therefore, an undisputed exponential growth in data is fueled primarily by the interaction of IoTs and other forms

of distributed computing devices (Desjardins, 2019) to manage and process data with low latency (IBM, n.d).

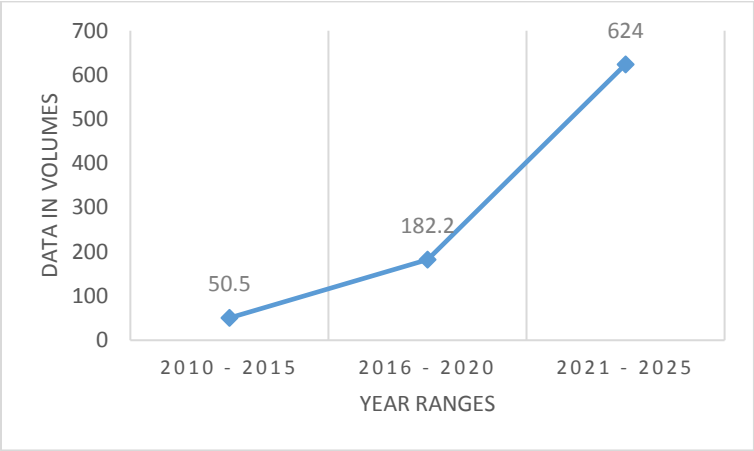

Figure 2.2. Growth Statistics of Data over the years

The impact of big data is extensive in diverse fields like retail marketing, entertainment, health care industry, etc. In retail marketing, convenience is key and a major drive for big data study. Supermarkets have big customer bases, and customer data provide purchase insights and facilitate the study of demand and supply dynamics for efficient resource allocations. This is realized from the real-time study of transactional database records of purchases. In entertainment, however, data is harnessed and monitored by content creators and suppliers in a bid to understand viewing habits. Personalization analytics, messaging, content delivery, and device analytics present content providers with the ability to predict and make recommendations based on the viewing preferences of customers (Marr, 2016).

Although big data is grouped into structured, semi-structured and unstructured with each type differentiated by similar characteristics such as high velocity: rate of data generation and transmission over a medium, volume: quantity of data produced and circulated per unit, usually in minutes of time, value: data characteristics of interest, veracity: validity and reliability of data.), and variety however is about the availability of different formats (Ishwarappa and Januradha, 2015; Marr, 2014; Thomas, 2018). The complexity of data sources as well as the nature of data renders traditional data processing approaches ineffective because more data is produced at unanticipated volumes and rates because of the intervention or artificial intelligence (AI), more mobile devices in circulation, the proliferation of social media and the internet of things (ibm, n.d). A stark example of data originators will be from sensory devices, audio/video, network, log files,

transactional applications, the world wide web and from social media apps and websites as well but most of the generated data is volumnious in nature and also require instantaneous processing (IBM, n.d). With this in mind, it is essential to bring to bare the destination of big data for its transmission, processing, and storage needs. The cloud is regarded as the defacto storage and processing medium for big data, reinforcing the need to maintain a safe and secure cloud. The term cloud is defined by the Cambridge English dictionary as a mass of water vapor that floats above the ground, the difference with fog however the distance from earth's surface is. The adoption of the name cloud however is used to describe the availability of servers for various computing functionalities. Armbrust et al., (2010) defines cloud computing to encompass the delivered applications and services over the internet as well as the hardware and software systems designed to provide cloud services (Armbrust, et al., 2010). The cloud is a combination of technology and data to create endless opportunities for internet users (Sudeep, 2020). It is also described by the National Institute of Standards and Technology (NIST) as the provision of on-demand services to customers through an actively established internet connection (Mell and Grance, 2011).

Cloud computing operates through virtualization technologies by providing virtual machines (VM) equipped with its own computing resources to users through a physically "non-existent" computer. Virtual machines are hosted on computers and sandboxed from one another to limit the interaction between host and virtual machines. Files and applications are not made visible to each other on each platform although they operate from the same physical host machine (CloudFlare, n.d). The distinction between the hardware and software technologies in cloud computing is referenced to tangible components of the infrastructure as data centers and logical infrastructure as the cloud. The data centers and cloud services are the responsibility of cloud service providers although cloud services are made available to public users or customers through pay to use services and hence the name public cloud. Private clouds however are the internal data centers purposefully designed for organizations and corporations and is not made available to the public; they are internalized for company utilization (Armbrust, et al., 2010; Sudeep, 2020; CloudFlare, n.d). Storage and processing are among the functionalities provided by cloud service providers (Akhtar et al, 2021). The division of cloud are: public, private and hybrid clouds, with each type dedicated to providing cloud services for a target group.

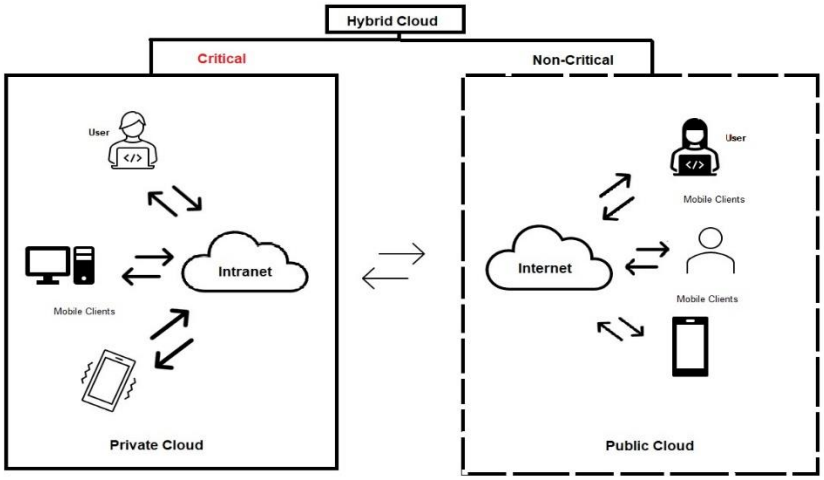

Figure 2.3. Types of cloud (Valkonen, 2013)

Therefore, there is the need to maintain high security and trust mechanisms to attain these heights. Dabhi et al. (2017) assert that heterogeneous nodes with varied computing capabilities characterize a fog computing ecosystem present a challenge of maintaining homogenous security and trust-based systems for each device. Although the public key infrastructure and the creation of trusted environments for the safe execution of tasks are potential solutions, the challenge of failure of sensors, networks, service platforms, and general use applications is eminent in operations. Privacy focuses on effectively managing users' data in remote environments.

However, the attempt to collect, store, and process huge volumes of data with a corresponding exponential growth rate concerning time reveals the inherent security problems in cloud computing. (Hamlen et al. 2010) revealed that the security vulnerabilities present in many technologies are also inherent in cloud computing, network security and vulnerability issues, security and privacy of operating systems, virtualization problems, optimum scheduling of available resources, load balancing, management of memory systems, and concurrency controls. As established by many authors, cloud computing is a computing paradigm that brings all computing resources to users' doorsteps. Big data is mainly attributed to the proliferation of IoTs with data transmission in huge volumes. The unpredictable nature of big data is a central problem that fog computing was introduced to remedy.

2.3 Fog Data Management

According to the Technical white paper by the Poznan Supercomputing and Networking Center data management is one of the core services of fog computing, although initially designed to resolve the technicalities and complexities of cloud computing (Alwakeel, 2021). Fog computing

being regarded as an intermediate layer between IoTs and Cloud, although provides semi-permanent storage for IoTs (Ema, et al., 2019). The identification and taxonomical approach to studying fog data management topics as suggested by Sadri et. al (2021), itemizes data processing, data security and data storage as the broad areas encapsulating other themes and subtopics in an attempt to outline the various approaches to data management in Fogging (Sadri, et al, 2021).

Nguyen et al., (2018) summarizes the data management process in four (4) distinctive stages with each stage in the data transit process in a fog computing environment playing distinctive roles such as the bottom stage acting as a source and acquisition medium for data, the data generated from connected devices are collectively or spontaneously acquired and ready for transmission to the fog servers, the next stage however performs data aggregation by way of accepting the transmitted data from multiple sources and grouping, however data is scrutinized partially at this stage by through preprocessing, further management of the preprocessed data is performed still with the middle part of the proposed architecture. Accumulated data undergoes active scrutiny as well as reorganizing and structuring through management processes and procedures before actually sent to the cloud for interpretation (Nguyen, Salcic, & Zhang, 2018).

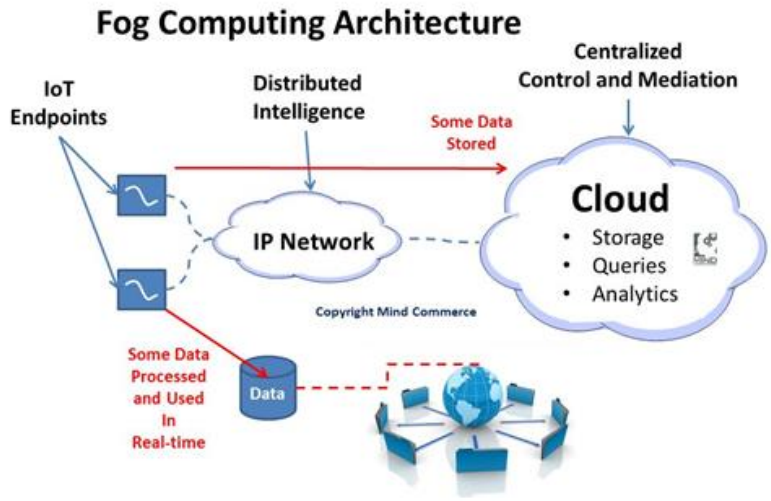

Figure 2.4. The architecture of the fog data management platform (Christensen, nd.)

2.4 Data Concepts

Other fog data management topics include data quality, data cleansing, data reliability, data sharing, and data locality

2.4.1 Data Quality

Providing facts is an essential premise for decision-making, enhanced by data characteristics such as openness, reliability, trustworthiness, and validity. The subsequent process, as well as the

resulting consequences, are comparable to the previously discussed qualities as a result of data assimilation. Individuals, corporations, and organizations must collect accurate, transparent, and open data in order to make informed decisions about productivity, profit, and growth opportunities (informatica, nd.; Wang & Strong, 1996). However, data redundancy, incompleteness, inaccuracy, irrelevant (unwanted characteristics) commonly leads to opposite results, resulting in higher costs and time waste associated with ensuring compliance with quality standards (Simplilearn, 2021). Data quality is a data management concept which focuses on the validity, reliability, and trustworthiness of data (Foote, 2021). Wang and Strong, (1996) define data quality as an attribute of being fit for use and as such quality data is often considered an essential asset to possess (Cichy and Rass, 2019).

2.4.2 Data Cleansing

When data is unfit for use, it is often the result of the removal of impurities and other extraneous materials which holds a high possibility of affecting the expected outcomes and results (talend, nd.). There is therefore the need to identify and rid off inherent impurities in data before being considered fit for use, data cleansing is a process of cleaning, removing, correcting improperly formatted, incomplete and duplicated data before it is processed for decision-making (Foote, 2021; talend, nd.).

2.4.3 Data Acquisition

With data acquisition mainly concerned with the collection of information, a series of conventionally ascribed steps have been adopted to produce and ensure smooth conveyance of information from source to destination (Fiandrino, et al., 2017), Fog computing platform is a widely dispersed platform with many heterogeneously connected devices participating in data exchanges both amongst devices and the cloud as well (Neware, 2019).

The transit of data between source and destination commences with the detection of physical and environmental changes by IoTs, the conversion from mere environmental or physical signals to corresponding binary for further processing by a computer is generally described as data acquisition, the systems that facilitates the work and process of data acquisition is known as Data Acquisition System and commonly abbreviated as DAS or DAQ to mean the same and refers to a series of stages that are regarded as most fundamental and crucial in the description of the influx of data from IoT sources participating in data exchanges and Transfers.

The data acquisition process is regarded as the initial steps towards the transition of data from IoTs to cloud for processing, Smith defines DAS as a process involving the sampling of signals for the measurement of a real-world physical phenomenon, conversion of observed data to a corresponding digital signal for processing or manipulation by a computer (Smith, 2020). Omega.com defines DAS as the process of gathering information from an environment which can be done with a sensor, a computer, and a measurement device (omega, 2019).

A data acquisition system consists of sensors designed to gather information from a physical environment, a device that measures the electrical properties of a given object, and a computer that processes the acquired signals (omega, 2019). Data acquisition stage reveals the heterogeneity of connected devices and multiple data generation sources, the case of car parking occupancy, data can be obtained from different kinds of carparks available in a city (Nguyen et al., 2018). These functions are imbibed into fog nodes at the edge of the network, IoTs although are low powered and possess low computational functionalities, gather information from the environment through sensors for onward conversion of signals by a transducer for manipulation by a computer with the aid of an application software.

The DAS plays an important role in the data transit process because it is primarily concerned with the collection and storage of data from diverse IoTs, manages as well as prepare the data from the sensors for onward transmission to the fog and ultimately to the cloud (Mohammadreza, et al., 2020). Data acquisition in fog computing facilitates fog node's ability to support real-time, post-recording visualization and analysis of data (Smith, 2020). In the field of medical imaging as reported by Shi et al., (2021) the data acquisition systems help in the early detection of symptoms of infections through automated scanning of patients and in the case of COVID-19, it protects medical practitioners and front-liners through minimal in-person contact (Shi, et al., 2021).

Within the fog computing paradigm, data collection and monitoring systems such as Recirculating Aquaculture System (RAS) are often employed to solve the problems by enabling heterogeneous communication of connected devices on IoT platforms, addressing data collection gaps whiles acting as a storage unit to handle the initially and semi-processed data for further transmission to cloud for real-time processing (Wu et al., 2020). Al-Hussaini et al, (2018) asserts data acquisition tools can be easily constructed utilizing a low-cost system like Raspberry Pi because of the benefits of size, cost, portability, high efficiency, and low power consumption (Al-Hussaini et al., 2018).

Wu et al., (2020) proposes an architecture for large and complex machines within the fog environment for data acquisition, a differential system from the conventional data acquisition setup to comprise of three layers for IoT data; cloud, fog and edge layers respectively, with the cloud being the ultimate processing medium. Under the proposed architecture by Wu et al., (2018), fog nodes shall collect data from dumb equipment's with the assistance of practical mathematical models, purposely for the conversion of electrical signals to binary for computer use, juxtaposed to multithreading and flexibilities aimed to improve performance. (Wu et al., 2020). This architecture focuses on total system performance particularly measured on metrics such as; response times, disk space use, processing time, data upload time of the proposed system to the conventional system and accordingly the results of the experiment confirmed a correlational increase in overall productivity (Wu et al., 2020). This experimental construct by Wu et al., 2020 did not however focus on data filtering, security and privacy of the system although it was indicated security and privacy will be much more explored in further studies.

2.4.4 Data Processing

The next logical step in the data transit process after data acquisition is processing. Data processing is described as the series of steps aimed at transforming acquired data (raw data) to meaningful data with insightful and endless opportunities. The cloud although extensively regarded inappropriate for massive IoT data demanding real-time exchanges, the adoption of fog computing however seeks to bridge the gap between source and destination by the provision and placement of cloud servers closer to the end devices (Pfandzelter and Bermbach, 2019) because of bandwidth limitations and latency requirements of the cloud although the privacy and security of user's data on the cloud remains questionable (Pfandzelter and Bermbach, 2019).

In an attempt to address the problems of latency and bandwidth restrictions, Desikan et al., (2017) proposed a dynamic Distributed Latency-Aware Data Processing (DLA-DP) for fog enabled gateways with the use empirical mathematical models. The DLA-DP model premised on latency-aware data processing algorithm uses an algorithm which updates the status of data at every gateway transit process of fog data. Also, the DLA-DP performs data forwarding by employing scheduling metrics such as arrival time, response time and hop count. For viability of the proposed model for IoT data after acquisition, Desikan et al., (2017) contrasted the DLA-DP model against two major existing models; First Hop Count (FHC) and Pure Cloud Processing (PCP) models respectively with the stipulated comparison metrics.

Comparison Metric	PCP	FHC	DLA-DP
Arrival Time	Cloud computing processes all data	As sensors collect data, gateways process the data until storage capacity is reached and forward the rest to the cloud	Uses probabilistic data forwarding between gateways only when the limit of the first gateway is reached
Average System Response	Same system response time with FHC but with resource saturation	Least and negligible. When the limit of the first gateway is reached, the excess is forwarded to the next router and then to the cloud.	Significant response time achieved by system
System Effectiveness	Single gateway resource saturation with an established 200, 400 or 1600 data per second arrival.	Same data arrival with PCP	Increased system effectiveness after multiple gateway saturation hence optimal performance of 4100 data per second is attained

Table 1.1. Contrasting performance metrics of DLA-DP with FHC and PCP (Desikan et al., 2017)

Inheriting the shared features of Desikan et al., (2017), Lan, et al., (2021) also proposed a hybrid data processing framework which consolidates the processing of both fog and cloud called the Process Engine Data Flow (PEDF). In this research, Lan, et al., (2021) indicated that the PEDF is flexible and can be used for multiple application scenarios because of its adoptation of the Direct Acyclic Graph method of routing data packets. Owing to the fog-cloud based architecture, the model structure of PEDF takes into account the heterogeneity of available hardware to ease the complexities of data processing applications by equipping each processing unit with its own PEDF (Lan, et al., 2021). One of the benefit of PEDF is the exploitation of cloud and fog semantics to create an on-demand multi-

platform process engine data flow with various resource constraints. The shortfall of this proposed model is the lack of security and privacy mechanisms to curtail the data in its possession.

The impact of data processing is experienced across multiple disciplines and the likes of big data companies are also constantly processing data in huge quantities without regard for the safety and privacy of originators, the case of Facebook and Zuckerberg having faced privacy complaints over the decades such as acquisition of user's data without permission. A $90 million decade-long class action lawsuit was won against the Tech giant and an agreement to delete all wrongfully collected data (Duffy, 2022; Jewers, 2022) the fundamental interpretation of this incidence is the realization of the importance of data and how much more is needed for various purposes.

Pfandzelter and Bermbach (2019) proposes the determination of answers to two important questions in data processing; the identification of location for data processing in an attempt to guide the filtration and aggregation of data at the edge of the network before transformation at fog and ultimately cloud aggregation. The determination of the tools and services become accessories to the initial stage as well as the utilization of data processing tools and services through the entire processing sequence (Pfandzelter and Bermbach, 2019). This suggestion is based on the premise of the distributed and heterogeneous nature of fog nodes in the entire data ecosystem. Fog nodes or IoTs enjoy the privilege of localizing the processing of its data at source and based on the heterogeneity of the data generated, there is a common pattern, structure and unbounded in most cases (Tonjes, et al., 2014). The conventional data processing structure has three (3) stages with storage being an auxiliary stage.

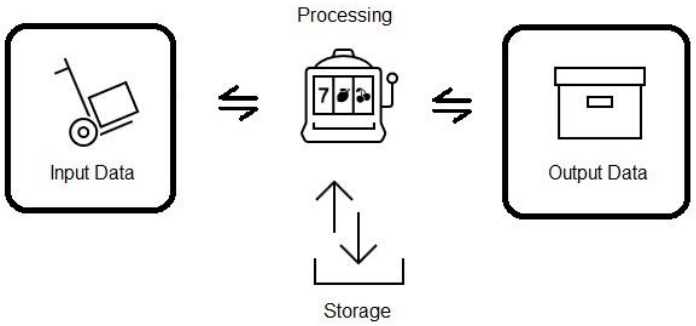

Figure 2.5. Stages of Data Processing (Duggal, 2021)

The exploitation of data to produce meaningful results is the primary function of data processing however, the main benefits of data processing in fog computing is to address the problem of network congestion cannot be overly emphasized in lieu of the associated growth of IoT nodes, however data security, data tampering, data privacy leakages (Zhang, et al., 2020).

2.5 Proposed Solutions for data concepts under fog data management

Functional Encryption with Public Keys, proposed by Sahai and Hakan (2010), is worry-free encryption. According to the authors, present public key encryption relies on the Key Generating Authority's capacity to retain a chain of trust. This is because a compromise on the authority's part decrypts the contents of the message, jeopardizing the security and privacy of messages. However, empirical probabilistic solutions eliminate the necessity for public keys for decryption and eliminate pure reliance on the central trust. Because it is secure against chosen-cipher text attacks (CCA), the worry-free encryption works for arbitrary polynomial-time functions. The worry-free encryption allows the sender to encrypt a message using the hidden access policy, which allows the recipient to decode the message after successful verification using the IND-polynomial CPA's time computable policy. The worry-free encryption strategy uses the random oracle model to allow the sender to encrypt a message using the hidden access policy, which allows the recipient to decrypt after successful verification using polynomial time computable policy under the IND-CPA public key encryption scheme. The authors' proposed encryption approach has the advantage that if a certificate authority is compromised, the message is not easily deciphered by an unintended receiver. By constructing a wearable Telehealth equipped with an Intel Edison processor, a low-power embedded computer (Shen, Su, & Zhang, 2018), Dubey et al. (2015) suggested a service-oriented architecture for fog. The body sensor network (BSN) for data acquisition, a fog gateway computer for real-time processing, and a cloud server for storage and analysis make up this service-oriented architecture. The fog's services include connecting patients with physicians for diagnosis, evaluation, and determination using data filters to save bandwidth by transmitting only important bits of information to the cloud. This is accomplished by employing Dynamic temporal warping (DTW), a pattern-mining technique for time-series data, Clinical speech processing chain (CLIP) for data filtering, and fog data compression prior to cloud transmission for decompression and processing. The suggested architecture collects data from patients in the form of time-series with the help of a low-power computer, and then performs a pattern finding analysis on the data before

sending it. This proposed architecture focuses on reducing (filtering) data for storage and transmission while consuming less power without sacrificing efficiency (Dubey, et al., 2015).

Kafhali et al. (2019) also presented an architecture for effective resource management by combining software defined networking (SDN) and network function virtualization (NFV). The authors built their architecture in a blockchain, fog, and cloud computing enclave, resulting in a more superior and efficient platform for managing the massive influx of IoT data. Considering the lack of a unified security solution for IoT data, the author's claims that this architecture enables a verifiable, secure, and permanent method of storing data processing records through smart contract enforcement. The benefits of integrating SDN and NFV include that it allows IoTs to communicate with their surroundings and process their own data without requiring human involvement, as well as making the fog environment more efficient, highly adaptable, low-cost, on-demand, and safe in the distributed cloud. According to the authors, the integration solves the problem of network congestion and security in the core network. Furthermore, security, privacy, and validity automatically adjust to the threat type, obviating the need for human evaluations and configurations at the network's edge aimed at improved latency.

Owing to the diversified nature of the fog environment as well as the latency challenges experienced within fog, Mahmud et al. (2018) proposed a policy called, latency-aware application module management policy aimed at improving communication latency between nodes without compromising deadline-based Quality of Service (QoS) and optimization of resource within fog. This policy is modelled on the structure of an IoT with much emphasis on the application layer and developed within the framework of the iFogSim environment as well as the ability of a fog node to independently execute tasks. This policy is implemented using the Module forwarding algorithm which uses a forwarding strategy to relocate modules in order to maximize the number of computationally active fog nodes. Two algorithms have been developed in support of the proposed application management policy. The first is about Application Module placement and the second one simplifies a constrained based optimization problem in forwarding modules towards the inactive resources of idle modules.

Li et al. (2018) proposed that resource optimization within the framework of machine learning models and algorithms may be used to achieve data collecting and management. Pattern/feature recognition and selection approaches with integration with swarm intelligence and decision table classifier, also known as Swarm Decision Table, were used to produce the proposed solution

(SDT). According to the authors, SDT was created with the goal of assisting in the selection of a suitable data mining model in a fog environment. To examine the performance of several swarm feature selection methods (BestFirst, Elephant and Harmony) with the decision table model, a simulation was run using the Waikato Environment for Knowledge Analysis (Weka 3) platform. The accuracy, Kappa statistic, time cost, recovery ability, fluctuation degree, and successive assessment measures were applied. Each SDT model demonstrated a unique ability in terms of recovery ability, fluctuate degree and successive time to the dataset used for the study.

2.6 Fog Data Security

Owing to the extensive growth in computing devices such as the IoT, there has been a corresponding increase in the quantum of data generated. And with the cloud being the default storage and processing medium for IoT data, there is the need to reinforce the need to maintain a safe and secure cloud. Sadly, the cloud is not an ideal storage and processing medium for time, latency sensitive applications. The Fog computing paradigm however was developed by Cisco Systems not as a replacement but as an extension of the cloud in its placement of servers and other valuable computing resources at the edge of the network (Khan et al., 2017). However, the underlying tradeoff between the cloud and fog are; latency, bandwidth restrictions, homogenous devices with diverse security requirements, which further leads to the postulation that the challenges inherent in cloud are also present in fog. Dabhi et al. (2017) asserts that heterogonous nodes with varied computing capabilities within a fog computing ecosystem present a challenge of maintaining uniform security and trust-based systems for each device.

2.6.1 **Traditional Approaches to Securing Fog Data**

Security within the fog environment has remained a major concern for researchers and security professionals. This is due to the growing number of attacks on the internet with low powered and inferior security equipped nodes being the easy targets or entry points for malicious attackers. According to a live map depiction of the growing global number of cyber-attacks as of 12th March 2022 by Kaspersky, Russia was ranked first with Brazil, the United States of America, Germany, and China followed respectively in order of magnitude. Attacks on these countries targeted financial services, consulting, Telecom industry, manufacturing and insurance companies and as reported on the Fortinet threat map, over 552,337+ records from each of the specified categories have been compromised (Fortinet, 2022; Kaspersky, 2022).

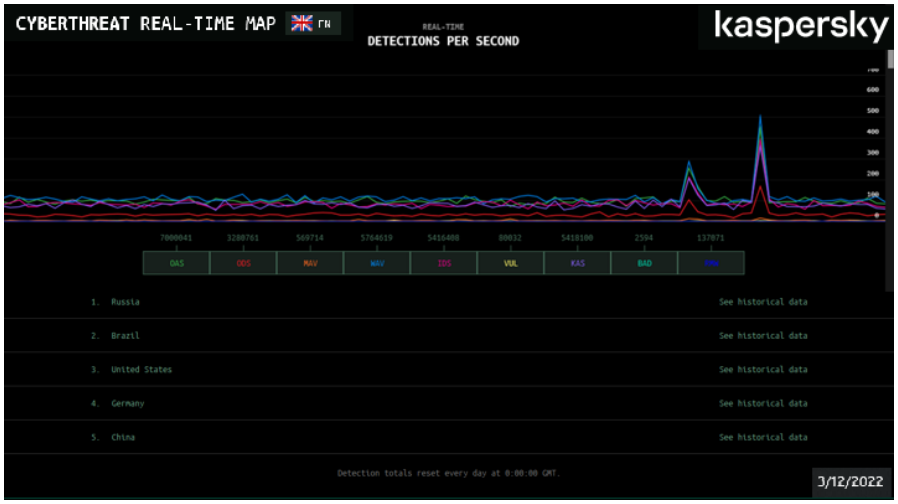

Figure 2.6. Cyber Threat Live Map (Kaspersky, 2022)

Owing to the growing number of attacks and its impact, clearly the traditional security measures can be described as both outmoded and ineffective to curb modern attacks. Traditional security approaches to securing fog data is largely premised on the adoption of public key infrastructure (PKI) and low-level cryptographic encryption techniques.

The protection of data in earlier systems was primarily to unplug a device to curb the impact of a security and data breach. But with advancements in modern computing, this is considered redundant and such data breaches need not much time to execute and cause the desired destruction, for this alternative protection systems were developed employing cryptographic techniques like encryption.

2.6.2 Encryption

Cryptography is a science for maintaining the confidentiality of information, and the process is called encryption (Malik, et al., 2020). Encryption is a cryptographic technique of converting information or data into a code, especially to prevent unauthorized access (Malik, et al., 2020). Cloudflare describes encryption as a set of techniques used to convert human readable text to a sequence of incomprehensible code (cloudflare, n.d). From the above, it might be concluded that encryption refers to the techniques used to alter data so that when intercepted by an unintended recipient the data is rendered inaccessible and does not convey the true meaning.

As a traditional method of securing data, encryption techniques and algorithms ensures that malicious and unintended parties are prevented from accessing sensitive data (CyberEdu, n.d) therefore enforcing; privacy, security, as well as protecting the integrity of data in transit, and with authentication measures the intended recipient deciphers the encoded data for comprehension.

Kaspersky security labs outlines encryption algorithms to include but not limited to: the Advanced Encryption Standard (AES), Rivest, Shamir, and Adleman (RSA), (3DES) Triple Data Encryption Standard (kaspersky, 2022). Encryption in its entirety is divided into two sub themes: public and private encryption. Wikipedia differentiates the two encryption techniques by the type of key pairs shared for decryption at destination (Wikipedia, nd.).

In Cybersecurity, the Confidentiality, Integrity and Availability of information (CIA triad) is regarded as crucial for the protection of data, this is enforced with the use of encryption techniques (Stickney, 2021). However, with growing systems and modern advancements in computing have rendered legacy security encryption methods are ineffective. Considering the number of bits of DES being 56 bits which results in 2^{56} key combinations, which is due to recent improvements takes less time to decipher. The RSA however is a comparatively better algorithm because the number of combinations is 2^{2048} combinations, which is next to impossible for normal computers to decipher, however the advances in quantum computing have proved that with time, the security offered by the RSA will soon be rendered obsolete.

In a wireless network medium however, the implementation of RC4 cipher in wireless networks also brings to question the security of the network. The implementation of RC4 in WEP renders it easy to crack, subsequent improvement in WPA-TKIP strengthens the security of the network by making the data more difficult to crack, with a high possibility of successful outcomes obtained within a short period of time. WAP2 uses the AES cipher and strengthens a network much more than WEP and WPA-TKIP however the upgrade to WPA3, prevents attackers from decrypting packets even when the shared key is known.

2.6.3 Public Key Infrastructure (PKI)

Imbibed into the PKI is the generation and management of a key used for encoding and decoding data. The Pcmag encyclopedia defines public key infrastructure as a security framework which protects data exchanges by employing public cryptographic encryption techniques (Pcmag, nd.). Techtarget defines PKI as the use of public encryption techniques to securely manage data and ensure data confidentiality during exchanges and transfers (Techtarget, 2021). The Public Key Infrastructure is essentially made up of policies, hardware, software, procedures and entities effectively coordinated to ensure secured data transfers through verifications and revoking of certificates. Data in its entirety is not encrypted, instead a certificate is issued by a certificate issuing authority (CA) attached to the data for transfers. A digital key is used to refer to the locking

and unlocking of data (encryption and decryption), however as described by Pcmag, the PKI uses public encryption by issuing a public and a private key are mathematically related and used to verify the authenticity of data between a source and its destination (Microsoft docs, 2021; tutorialspoint, n.d). The outline of the use of PKIs are explicitly explained in the Request for Comments by the Internet Task force which is indicated in RFC 2527, a public key certificate and therein a digital certificate is used to verify the identity of the originator with the associated private key (Chokhani and Ford, 1999). The problem with PKI is that the entire process and components operate on a chain of trust based on identity verification, the compromise of a single node results in the distrust of all nodes. Moreover, certificate authorities are supposed to be trust-worthy, and a breach of security breaks the chain of trust. A stark example is the 2011 and 2017 detection of fake certificates by a Dutch CA, digiNotar and Symantec's CA respectively. Certificates in use with an origin of the mentioned are blocked on all devices initiating transfers with from this CA are discarded (Techtarget, 2021).

2.6.4 Bypassing Traditional Perimeter Defenses

Accessibility is integral to security. Cloud environments are thoroughly connected, which facilitates bypassing traditional perimeter security models, while traditional environments are controlled through perimeter security models. There are several threats to system and data, including malicious insiders, account hijacks, inaccurate identity management, and unsafe APIs.

2.7 Security Frameworks for data management in Fog Computing

The national institute for standards and technology defines a security framework to comprise all pre-determined procedures and guidelines designed to assist in the management and reduction of security risks (The National Institute of Standards and Technology, nd.).

A security framework is distinguished from a standard in that it's a pre-established set of processes and procedures that defines the policies and procedures surrounding the implementation and management of security controls (Techtarget, nd.). The characteristics of a security framework is that they are voluntary, based on existing standards, guidelines and best practices aimed at the management and reduction of risks (Dawson, 2019). Moreover, due to the underlying challenge of a lack of uniform standard to cater for the security needs of heterogeneous fog nodes, the adoption of security frameworks bridges this gap with the aim of managing and reducing the security risks experienced in a fog ecosystem. There are numerous security frameworks aimed at preserving

privacy of individuals as well as their data from interferences and is further described as the blueprint for risk management as well as policies used for vulnerability reduction (Dawson, 2019).

2.7.1 NIST Cybersecurity Framework

The NIST Cybersecurity Framework was developed in February 2014 by the National Institute of Standards and Technology for use by all organizations and individuals keen on boosting their security defenses to mitigate threats (Dawson, 2019). An executive order signed by former President Trump directed all federal agencies to adopt and use this framework in the management of risks, the private sector also adopts it as a heterogeneous guide for privacy preservation. (Swenson, 2018). The NIST cybersecurity framework facilitates the open communication and collaboration between executives, experts and industry associations (Swiss Cyber Institute, 2021). The NIST cybersecurity framework has 3 primary components (Core, Profiles, and Implementation Tiers) and a five (identify, protect, detect, respond, and recover) step process for mitigating security risks and maintaining a secure system is notably adopted in many industries because it is cost-effective and flexible (Dawson, 2019; The National Institute of Standards and Technology, nd.).

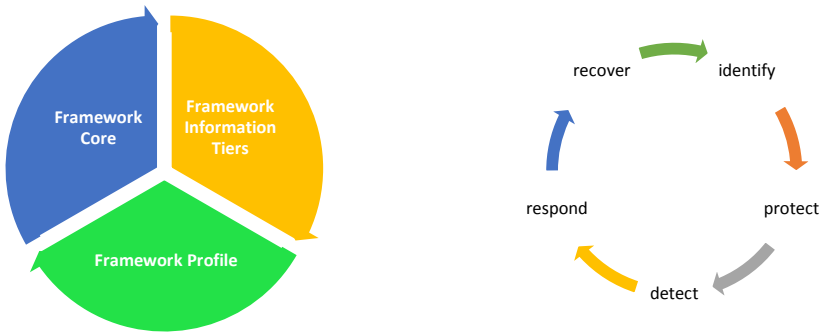

Figure 2.7. Structure of the NIST Cybersecurity Framework (The National Institute of Standards and Technology, n.d)

The core of the NIST cybersecurity framework adopts a plain and easy to understand language to assist industries and individuals to tackle cyber risks and complement the existing cybersecurity and management processes (The National Institute of Standards and Technology, nd.). The implementation tier component allows for the evaluation of the current cybersecurity processes and provides reasons for the inclusion and adoption of best practices whilst the framework profile component enables organizations to develop a blueprint for minimizing the cyber risks that are aligned with organizational goals (Swiss Cyber Institute, 2021). The adoption of this framework in the fog environment is premised on the established fact that, IoTs operate within a trust-based

system, where a single break in trust could ruin the entire system. Although the NIST cybersecurity framework is a plan-driven approach, it helps in tackling data risks within a fog environment (Swiss Cyber Institute, 2021). Cisco's Industrial Threat Defense (CITD) is modeled on the provision and deployment of maximum network resources to meet the flexibility need of existing networks (Cisco, nd.). The impact of CITD rests on the ability of managers to assess associated risks as well as outline preventive steps to curb the possibility of spreading through leveraging of existing tools and practices within any adopted environment (Cisco, nd.).

2.7.2 International Organization of Standardization (ISO) 27000 Series

Another key framework for the protection of data within the context of managing the safety and privacy of systems is the ISO 27000 series, developed by the International Organization of Standardization and the International Electrotechnical Commission (Dawson, 2019). The ISO 27000 series frameworks are flexible and have been adopted by a lot of industries because the focus of the centralized nature of this framework suite (Kirvan and Joseph, 2021). The Information Security Management System (ISMS) within the ISO enclave ensures proper audit and compliance to established procedures for implementation and is described as the primary standards comprised mainly of ISO 27001 AND ISO 27002 (Kirvan & Joseph, 2021). An externally approved third party audits and certifies the process to ensure compliance with ISO 27000 series standards. Approximately 60 standards are included in the ISO 27000 series. One of them is ISO 27018, a standard for cloud computing. A disaster recovery standard based on ISO 27031 discusses IT disaster recovery, Digital evidence is addressed in ISO 27037, and storage security in ISO 27040 and Healthcare is covered by ISO 27799.

2.8 Recommend appropriate solutions for data management in fog computing

The introduction of each new technology has an accompanied potential for data security breach. The security considerations of mobile phones are different from those of computers and even servers or other smart devices. The convenience however afforded with the use of wireless networks (Wi-Fi) has more avenues for security attacks than the traditional wired media. With most security measures heavily targeting external threats and attacks, malevolent downloads, internally emanating threats costs twice as much. An internal attack could be in the form of deletion of sensitive data and release of confidential information. Password sharing or granting access could lead to escalation of privileges leading to unintended access.

To prevent such instances of intended and unintended problems, developing data securities policy is more important than ever. To provide an overview of the acceptable norms regarding the use of mobile technology as well as password security, wireless access policies for protecting confidential data, the primary role of this policy should be to provide an overview of the acceptable norms pertaining to the use of this technology. In contrast to the encryption and algorithms used, Sahai and Hakan, (2010) proposed the Worry-Free Encryption to provide the needed assurance of an uncompromised data exchanges. The Worry-Free Encryption uses public encryption scheme through the authorization of credentials with a handshake setup, Pre, Auth, CheckAuth, Enc, Dec functionalities (Sahai and Hakan, 2010).

2.7.1 Blockchain Integration

The IP addresses of the parties can be traced despite the presence of firewalls and Network Address Translators, which is a privacy breach to a large extent. These concerns can be further explained in the lack of international regulatory standards to ensure conformity and uniformity of the devices and the complexities of operating the systems. These growing unemployment concerns are directly proportional to the ever-increasing number of innovations emanating from the field daily. Privacy and security present itself as a major challenge as it directly impacts users' safety (Quek, 2017). Attempts to bridge the security gaps in IoT have led to the introduction and adoption of blockchain technologies (Atzori, 2016). Blockchain is defined by (blockchainresearchinstitute, n.d) as Distributed networks that act both as a digital ledger and a mechanism for all assets to be transferred securely without the involvement or interruption of an unauthorized third party. A blockchain network strictly enforces transparency and openness by providing a single forensic record of all transactions in real-time. A typical blockchain network performs the following operations – generate tokens, storage, and exchange of tokens. However, the token represents an encryption string that is very difficult to break, to enforce the storage and exchange of records on the network. The adoption, however, has been perceived as a panacea to the security problems in IoT, as a replica of its functionalities is enforced in exchanging data in the interaction of IoT devices. The enhanced security-based functioning of IoT has made it the preferred choice in almost all aspects resulting in the transmission of data-intensive events in real-time. The overall goal of its implementation is to improve efficiency through resource optimization and improvement in the quality of services. Notable features of IoT are its ability to function as a decentralized system while ensuring diversity in its applications and the diversity of its ecosystem. These characteristics

often result in numerous challenges such as network-based complexities, resource constraints, privacy, and security vulnerabilities. Blockchain in its integration in IoT has been perceived as a solution to IoT networks' security and privacy-preserved challenges.

2.8 Integrating Machine Learning in IoTs

Machine learning as an artificial learning branch is concerned with the development of models, the ability to train and test models as well as trusting the model to make accurate predictions with a certain amount of confidence. The underlying principle of machine learning models is systems are aided with the ability to learn from the available datasets to make decisions such as identify patterns with minimal or no human involvement or other underlying connotations.

However, the most notable limitation of ML models is the heavy dependence on the availability of datasets for learning before final deployment into the real-world issues. The data for training a model may be hazardous in some instances as they pose a few security threats and on which attackers could leverage on to carry out attacks (Hussain et al., 2020) which begets the question of data reliability (Roberto et al., 2018). The introduction of deep learning is however to address the limitations of machine learning models because deep neural network models imitate the workings of a functional human brain in its ability to identify patterns and make decisions (Hussain et al., 2020). Machine Learning and deep learning techniques can provide embedded intelligence to IoT devices using networks.

Chapter Three

3.0 Introduction

Based on the theoretical underpinnings and design strategies presented in Chapter one (1), the research design, sample size and sampling procedures, population used in this study are outlined. A discussion follows on the sources of data, tools used for data collection, and concluding with a description of how to collect and analyze data.

3.1 Research Design

MacMillan and Schumacher (2001:166) define a research design as a series of steps which involves the determination of participants, sites and data collection procedures aimed at answering the questions posed in research. They go on to say that the goal of a good research design is to produce credible results. The goal of this study, however, is to look at the state of data management

security and privacy in a fog computing environment by reviewing existing security frameworks and data privacy procedures. This study lays bare the security vulnerabilities that exist inside the fog environment, creating hazards to user data privacy and security, and in lieu of that, this study incorporates features of data in addition to the acquired facts and statistics. Based on the foregoing, a hybrid research design that incorporates both quantitative and qualitative characteristics is the most appropriate research strategy for this study. The researcher's choice of research design is since statistical data was employed to conduct the study within the context of the data and internet growth trends and the prevalence of IoTs.

3.2 Research Methodology

According to Schwardt (2007:195) a research methodology is a theory of procedural inquiry comprising the analysis of assumptions, principles, and procedures in a particular approach to inquiry. The research design type for this study is a descriptive and explanatory analyzed through mixed (qualitative and Quantitative) methods. Bouchrika (2020) defines explanatory research as the attempt of a research to explain findings and ideas aimed at expanding on an existing theory by exploring the limits of a subject to present answers central to the research's topic. Questionnaires were used to evaluate participant's perception of online privacy, data security and to determine their levels of satisfaction in the management of their data within the fog environment.

3.3 Population

(Banerjee & Chaudhury, 2010) defined a research population to comprise all potential groups from which information is to be ascertained. The population of a study can also be described as all groups with either homogenous or heterogenous characteristics beneficial to the study and from which the researcher leverages on to acquire such needed information. This study aims to look at the state of data management security and privacy in a fog computing environment by reviewing existing security frameworks and data privacy procedures, therefore the population for this study was selected from the global population of internet users estimated at about 5,258,006,498. As found on the website of (WebsiteSetup, 2021), Asia is the only continent with the largest percentage of internet users counting for about 51.2%, whiles 14.8% are from Europe, with 12.8% from Africa and 9.5% in Latin America and the Caribbean, 6.8% in North America, 3.7% in the middle east and 0.6% from Oceania and Australia. (WebsiteSetup, 2021).

Owing to the geographically dispersed nature of the target population, cost limitations, and the cumbersome nature of the target population were also reasons for the adoption of this sampling

technique for the creation of continental clusters to group respondents. With discretion, a random number of 1000 was chosen from the population of internet users. This was scaled down to 500 due to cost and some other anticipated constraints. From the population of 500 potential respondents, clusters of 5 comprising 100 potential respondents were estimated and a simple random method was used to select a cluster from the 5 clusters to ultimately become the population to be studied.

3.4 Sampling Technique and Sample Size

An analysis or research study can be conducted by choosing a sample from a population. As a general rule, the population samples should be selected in a way that allows conclusions or inferences drawn from the study to be generalized. Samples are chosen from a large population, so that they represent the entire sample, as defined by Leady (1993).

Researchers use sampling techniques to gather information from a smaller group that can fairly represent the entire group. Cluster sampling was used in this study. In cluster sampling, researchers divide a population into groups and then obtain a representative sample from each group. The cluster is a subgroup that represents the diversity of the whole population, while the set of clusters are similar (Frost, nd.). The choice of sampling technique was based on the dispersed nature and geographical distribution of the population. The simple random method was further used to assist in the selection of potential participants from each cluster to be included in the study, aimed at facilitating the variation of, between and within clusters as well afford all members under consideration the equal chance of being selected. The researcher adopted Yamane Formula in the estimation of sample size from which inferences could be made. The formula that would be used to calculate the sample of the study was developed by Yamane (1973). The formula is produced below.

$$n = \frac{N}{1+N(e)^2} \qquad (3.1)$$

n = sample size
N = the population size
e = allowable error (which in this study is estimated to be 0.05).
The sample size is estimated as.

$$n = \frac{100}{(1+100(0.05)^2} = \frac{100}{1+100(0.0025)} = \frac{100}{1+0.25} = \frac{100}{1.25} = 80.$$

A proportional figure of 50% was calculated again against the original student sample size of eighty.

$$\frac{50}{100} \times 80 = 40$$

3.5 Sources of data and Instrument for data collection

The main source of data was mainly from secondary and primary sources respectively. The main primary source was from the administration of questionnaires from which responses were collected through an online administration of questionnaires where questions were mostly close ended to prevent respondents from giving vague answers. The secondary data sources were existing literatures obtained from journal and article publications, online (Internet) searches and other literatures obtained from books and book sections respectively.

3.5.1 Questionnaire

The questionnaire was found to be appropriate for this study because the study employed an explanatory design (Fraenkel & Wallen). The questionnaire was a close-ended type which was in the form of a Likert scale which was built on the key themes raised in the research questions; familiarity with fog computing, perceptions about fog data, providing personal information over the internet and perceptions about internet privacy.

3.6 Procedure for Data Collection

Data was collected by administering an online questionnaire to respondents. A google form was created with the questions and a link to the questionnaire was sent out to respondents through the various social media platforms like What's app, Telegram and Facebook. A consent was sought from each respondent with the purpose of the study explained before answering commenced. The researcher explained the purpose of the study and assured the participants of the necessary confidentiality on the information to be gathered.

3.7 Analysis of acquired data

Data was coded, analyzed, and reported using the Statistical Package for Social Science (SPSS) Statistics software and Microsoft Excel. The SPSS software was chosen for the data analysis because it is reasonably user friendly and does most of the data analysis one needs as far as quantitative analysis is concerned (Dadzie-Bonney, 2015). Furthermore, SPSS is among the most widely used statistical data analysis tools in educational research (Muijs, 2004). The data entries were done by the researcher in order the check the accuracy of the data. The study adopted the descriptive statistical tools such (frequencies, mean, mean of means, standard deviation and

percentages) to analyze the data collected, and also to answer the research questions raised. To facilitate the discussion, the answers given to "Strongly Agree" (SA) and "Agree" (A) on the Likert-scale were combined, and the responses given to "Strongly Disagree" (SA) were combined, too.

Chapter Four

4.0 Results, Analysis and Discussions

In this chapter the researcher presents results, analysis and discussions. The discussions would be based on the results from the data collected. A total of forty (40) respondents were chosen to participate in the online survey with the realization of a 100% response rate.

4.1 Analyzing the reliability of data collection instrument

Data collection was primarily conducted through the use of a questionnaire. Having a reliable instrument was of the utmost importance for obtaining accurate results and predictions since this was the main instrument for collecting data. Oden (nd.) describes the reliability test of a questionnaire as determining whether or not it is able to produce the same or similar results under the same or similar conditions (Oden, nd.). In light of this, Cronbach's alpha was adopted as well as employed to evaluate the reliability of the items contained in the questionnaire. An evaluation of Cronbach's alpha's internal reliability is usually based on the scale of test items and measured between 0 and 1 (Goforth, 2015; Tavakol and Dennick, 2011). Additionally, the Cronbach's alpha checks to ensure that the instrument is accurate and reliable (Pires, Colussi, & Calvo, 2014). Cronbach's alpha is estimated using the formula

$$\alpha = \left(\frac{k}{k-1}\right)\left(1 - \frac{\sum_{i=1}^{k} \sigma_{yi}^2}{\sigma_x^2}\right) \qquad (4.1)$$

Where:

$k = number\ of\ scale\ items$
$\sigma_{yi}^2 = Variance\ associated\ with\ i\ item$
$\sigma_x^2 = variance\ associated\ with\ observed\ total\ scores$

The use of this formula was facilitated by the Statistical Package for Social Science (SPSS) Statistics which computed the values for alpha and the inter-item correlation matrices respectively in Tables 4.16, 4.17, 4.34 and 4.35 respectively.

4.2 Demographic Characteristics

The relevant demographic characteristic considered important by the researcher was continent of respondent's origin and age group. Table 4.10 is the distribution of continental origin of respondents.

Table 4.1. Respondent's continent of origin

	Frequency	Percent
Africa	33	83
Asia	3	7
Between Europe and Asia	1	3
Europe	3	7
Total	**40**	**100**

Source: Field Data, 2022

From Table 4.1 out of a total of 40 respondents, 33 respondents of the total respondent population were from Africa, 3 respondents were from Asia and Europe respectively and one (1) respondent between Europe and Asia. From the Table 4.10 it can be said that the African participants were marginally more than the other represented participants from the other mentioned continents. As part of the demographical data, respondents were asked of the age group they belonged to.

This question was used to determine the age representation of respondents the research participants. This is summarized in the table 4.2;

Table 4.2. Age distribution of Respondents

	Frequency	Percent
18 to 24	4	10
25 to 34	29	72.5
35 to 44	7	17.5
Total	**40**	**100**

Source: Field Data, 2022

From Table 4.2, majority of the respondents 29 (72.5%) were between the ages of 25 and 34, whiles age range from 35 to 44 were 7 (17.5%) of the entire respondents sample, and 4 (10%) respondents were between the ages of 18 and 24 respectively. The age distribution presents a conclusion factor that no minors but adults who were fully aware of the ramifications of online dangers participated

in the study. To ascertain the internet usage habits, respondents were asked of their familiarity with the term and usage of the internet. The responses were summarized in Table 4.3.

4.3 General Perceptions of Internet Privacy

Perception of respondents of internet privacy were sought to provide insights, the responses were tabulated in the respective tables.

Table 4.3. Are you familiar with the internet?

	Frequency	Percent
Yes	40	100
Total	**40**	**100**

Source: Field Data, 2022

From Table 4.3, all 40 (100%) respondents indicated that they were familiar with the term and usage of the internet. Respondents were also asked to indicate the years of internet usage, responses are summarized in Table 4.4.

Table 4.4. If Yes to the question above, how long have you been using Internet?

	Frequency	Percent
3 to 4 years	1	2.5
5+ years	39	97.5
Total	**40**	**100**

Source: Field Data, 2022

From Table 4.4, 39 (97.5) respondents agreed that their years of experience and familiarity with the internet usage was more than 5 years, whiles 1(2.5%) also indicated that their years of experience with internet usage ranged between 3 to 4 years. Although respondents indicated they were familiar and experienced in the usage of internet, to better understand the internet usage habits, respondents were further asked how often they were connected to the internet. The results are summarized in Table 4.5.

Table 4.5. Frequency of connection to the internet

	Frequency	Percent
Daily	20	50
Hourly	19	47.5
Yearly	1	2.5
Total	**40**	**100**

Source: Field Data, 2022

From Table 4.5, 50% of the respondent population indicated that they connected to the internet on a daily basis, whiles 47.5% indicated that they connect hourly and only 2.5% indicated that they connected to the internet on a yearly basis. The results from Table 4.5 revealed that respondents were experienced and connected more often to the internet on a daily basis. The amount of time people spend online on a daily basis was also queried, and the answers are presented in Table 4.6.

Table 4.6. Duration of connection to the Internet

	Frequency	Percent
All day long	17	42.5
All week long	2	5.0
Specific portions of the day	21	52.5
Total	**40**	**100**

Source: Field Data, 2022

According to Table 4.6, 52.5 percent of respondents said they only use the internet during specified times of the day, 42.5 percent said they use it for the entire day, and only 5% said they were online all week. Respondents were asked why they stay online and for reasons they remain connected as indicated and summarized in table 4.7. According to results summarized in the table, the most ranked reason was using the internet for educational purposes which represented 85% of the total respondent population, with using the internet for work-related research and communication purposes were ranked equally (70% each) according to the responses received, 60% of respondents also indicated that their use of the internet was for entertainment purposes whiles 52.5% responded that they also used the internet to access the happenings of current affairs, product information gathering was 50% and online shopping was 34(85%) the respectively and the least reasons for the use of the internet was for making financial Transactions 15(37.5%) and traveling reservations 9(22.5%). The results from the analysis of Table 4.7 was based on the stipulation that respondents were free to choose multiple responses provided, based on which ranking of the preferred use of the internet was determined.

Table 4.7. Ranking of reasons for the use of the Internet

Reason	Frequency	Percent
Education	34	85
Work-related research	28	70
Communication and staying in touch (social media, emails)	28	70
Entertainment	24	60

Accessing current affairs (News, sports, weather)	21	52.5
Product information gathering	20	50
Online shopping	34	85
Personal finance (Banking and business-related financial management	15	37.5
Travel reservations	9	22.5

Source: Field Data, 2022

To further understand the reasons for rankings of the internet usage as summarized in Table 4.7, respondents were then asked if the internet was confusing to use, responses were summarized in table 4.8.

Table 4.8. Complexity (confusion) about the perception of the Internet

	Frequency	**Percent**
Strongly Agree	1	2.5
Neutral	4	10
Disagree	12	30
Strongly Disagree	23	57.5
Total	**40**	**100**

Source: Field Data, 2022

Respondents were asked about their familiarity with fog computing and their responses were summarized in table 4.9. This question was to determine the familiarity of respondents within the fog computing environment. According to the responses gathered, 65 percent of respondent's total figure indicated that they were familiar with fog computing whiles 35 percent also indicated they were not familiar with both the term and its usage. The researcher took this into consideration and as such the heading of this section included a definition of both cloud and fog computing. To explain the concepts for respondents was an attempt to describe the concept for a more vivid understanding. The results however revealed that although fog computing was well known, respondents were not familiar with the term.

Table 4.9. Familiarity with fog computing

	Frequency	**Percent**
No	14	35
Yes	26	65
Total	**40**	**100**

Source: Field Data, 2022

4.3.1 Major Findings

The general aim of the study was to assess the impact of security and privacy on the management of data within the fog computing environment. The demographic results from the administered questionnaire revealed that all respondents were experienced internet users who have been using the internet and its associated resources for various purposes. Age was a main factor which affected the usage of internet, most respondents fell within the age category between 18 and 35, and hence they were the youth. The most ranked use of the internet was the use of the internet for educational purposes, followed by work related research and communication on social media, whereas the least use of the internet was for travel reservations. According to the Technology Acceptance Model (TAM; Davis, 1989), the perceived ease of use and perceived usefulness are the factors that affect a person's decision to adopt a technology with this in mind, the awareness of privacy-preservation within the context of internet usage was envisaged to assess the perceived complexities associated with the use of the internet (Schaie and Willis, 2016). The results revealed respondents strongly disagreed with the assertion of the difficulty and complexity with the use of the internet and fog computing. Elements of the technology acceptance model were further employed to assist the researcher in the determination of respondents about data privacy. Respondents were asked about the perceived ease of use of the management of data on the internet with 35 percent of the respondent population disagreeing. Privacy-preservation is key to the continued use of services within the context of internet usage, as a result respondents indicated that they neither agree nor disagree with the possibility of the tracking or monitoring of their usage of the internet. Moreover, the availability of information on various topics relative to the needs of respondents facilitated the continued reliance and use of the internet and fog computing infrastructures respectively.

4.4 Perceptions about data privacy

Data privacy is an important aspect in the determination of the safety and privacy of data within the context of fog computing, therefore respondents were asked of the complexity in the management of data generally on the internet. The results were summarized in table 4.10.

Table 4.10. It is difficult to manage data on the Internet

	Frequency	**Percent**
Strongly Agree	5	12.5
Agree	7	17.5
Neutral	10	25

Disagree	4	10
Strongly Disagree	14	35
Total	**40**	**100**

Source: Field Data, 2022

When respondents were asked if it is difficult to manage data on the internet, 14 (35%) strongly disagreed with the statement, whiles 4 (10%) also disagreed, however some respondents remained neutral towards the management of data on the internet whereas, 7 (17.5%) and 5 (12.5%) agreed and strongly agreed respectively. The results obtained indicated that, respondents were aware and could control the quantum of data shared on the internet. This further reiterates and informs the researcher that data privacy was a concept known by respondents.

Table 4.11. I am confident no one monitors what information transmit on the Internet

	Frequency	**Percent**
Strongly Agree	3	7.5
Agree	5	12.5
Neutral	11	27.5
Disagree	12	30
Strongly Disagree	9	22.5
Total	**40**	**100**

Source: Field Data, 2022

The responses from study participants on the question about confidence in the monitoring of information transmitted on the internet were recorded in table 4.11. From the table, majority of the respondents 12 which represented 30% disagreed with the statement, indicating that most of what was transmitted on the internet was monitored by their Internet service providers and the governments. However, 11 respondents which represented 27.5% remained neutral on the subject whiles 5(12.5%) and 3(7.5%) agreed that transmitted information was not monitored and they were anonymous on the internet.

Table 4.12. The Internet is expensive

	Frequency	**Percent**
Strongly Agree	22	55
Agree	8	20
Neutral	7	17.5
Disagree	3	7.5
Total	**40**	**100**

Source: Field Data, 2022

Patronizing cloud and fog services has an associated cost (McNally, 2022) reports that an average internet user in America spends about $61 on internet bills depending on location and the type of

connection used. Table 4.21 is a summary of the responses of participants on the perceived cost of accessing the internet in their respective countries. In lieu of the results from Table 4.21 as well as to understand the reasons why users spend the indicated amount of time on the internet, 22(55%) and 8(20%) respondents strongly agreed and agreed that internet is expensive, 7(17.5%) respondents remained neutral about this whiles 3(7.5%) disagreed on the cost of internet usage.

The Cornell Law School (nd.) defines pornography or "porn" as the display and distribution of sexually entailing materials purposely for stimulating sexual desires. (Cornell Law School, nd.) The internet is described as one of the easiest places for the distribution of sexually explicit materials and respondents were asked if it was easy to find pornography, the results of respondents were summarized in table 4.13.

Table 4.13. It is easy to find porn on the Internet

	Frequency	Percent
Strongly Agree	26	65
Agree	8	20
Neutral	4	10
Disagree	1	2.5
Strongly Disagree	1	2.5
Total	**40**	**100**

Source: Field Data, 2022

From the Table 4.13, 26(65%) and 8(20%) respondents strongly agreed that pornography is easy to find on the internet, 4 (10%) remained neutral on this matter whiles 1(2.5%) both disagreed and strongly disagreed about this statement. The results from the table indicated that, pornography was easily accessed and spread on the internet which further begs the question of safety on the internet, most respondents to the study were 18+, however individuals under the adult prescribed age could also easily access such contents online.

Table 4.14. There is too much information on the Internet

	Frequency	Percent
Strongly Agree	28	70
Agree	7	17.5
Neutral	3	7.5
Strongly Disagree	2	5
Total	**40**	**100**

Source: Field Data, 2022

According to the results of the purposes for which respondents used the internet, education and research based were ranked top, for this reason respondents were further asked if their choice was based on the availability of too much information on the internet. Availability and accessibility of data were crucial to the management of data within the fog computing environment. From the Table 4.14, 28(70%) and 7(17.5%) of the respondent population indicated that they strongly agree and agree with this statement. 3(7.5%) remained neutral whiles 2(5%) strongly disagreed respectively. Internet spam is described as the receiving of unsolicited emails, based on the responses on the usage pattern of respondents on the internet, they were further asked if they receive spam or unwanted messages from unknown internet users, the results are summarized in Table 4.15;

Table 4.15. I always receive unwanted messages from unknown Internet users

	Frequency	**Percent**
Strongly Agree	12	30
Agree	11	27.5
Neutral	9	22.5
Disagree	6	15
Strongly Disagree	2	5
Total	**40**	**100**

Source: Field Data, 2022

From Table 4.15, pertaining to the receipt of unwanted messages from unknown users on the internet 12 (30%) and 11(27.5%) both strongly agreed and agreed respectively to the assertion whiles 9(22.5%) remained neutral and however, 6 (15%) and 2 (5%) strongly disagreed and disagreed respectively.

4.4.1 Major Finding - Perceptions about data privacy

Data privacy is an important aspect in the determination of the safety and privacy of data within the context of fog computing and interconnection of multiple devices. Respondents indicated they were often asked to provide personal information over the internet to websites and unsuspecting individuals. Personal information also referred to as personal identifiable information (PII) is defined by any type of information that can be traced and or tracked to a specific individual, information such as name, contact, age, height, skin color, social security number, email etc. are personal identifiable to a specific individual by which agencies or third parties intend to use such acquired information from individuals in conjunction with other data elements for prescribed purposes (US Department of Labor, nd.). The study revealed that cloud, fog and generally internet

security is not an abstract concept respondent were oblivious to, hence the results about the perception of users about the complexity of data privacy-preservation in Table 4.10 revealed that a cumulative sum of 18 which represented 45 percent of the respondent population strongly disagreed whiles a cumulative sum of 12(30%) strongly agreed. This result is interpreted in relation to the indicated use of the internet from Table 4.7 about the ranking of the use of the internet.

Data Privacy is defined by emotiv (nd.) as the set of practices aimed at ensuring that the data shared by users is used for its intended purposes and not misappropriated (emotiv, nd.). In lieu of this definition of data privacy, respondents were asked if they were confident in both the management and as well as with the ability of their online activities to be monitored, a cumulative sum of 21 (52.5%) strongly objected that their online activities were not monitored whiles a sum of 8(20%) agreed that their online activities were monitored. The noted disparity between the responses and the respective rates reveals the division of uncertainty of knowledge of the concept of monitoring and tracking of online behaviors and actions by companies, websites and unsuspecting parties.

35(87.5%) of respondents further indicated that their patronage of the internet and its associated services is influenced by the availability of information while 2(5%) indicated that despite the availability and ease of use and access of information on the internet, their choice remained intact.

Other factors that impact the provision of personal information over the internet were:

1. Receiving unsolicited messages and emails from unknown parties
2. The ease of access as well as the readily accessible availability of information as indicated by some respondents was of high concern.
3. The cost associated with the usage of the internet
4. Easy access to pornographic materials and media

4.4.1.1 Test of Reliability - Perceptions about data privacy

Table 4.16. Test of reliability - Reliability Statistics

Cronbach's Alpha	Cronbach's Alpha Based on Standardized Items	N of Items
.687	**.584**	**8**

Source: Field Data, 2022.

From Table 4.16, the computed value for Cronbach's alpha is$\alpha = 0.687$, the scale of interpretation for the obtained value of alpha ranges between 0 and 1, however an obtained alpha value of more than 0.6 is acceptable and preferable (Tavakol and Dennick, 2011). In lieu of the obtained alpha value, the questions contained in this section were acceptable. The question items were further correlated in using an inter-item matrice which as further an estimation of the consistency of questions when compared to other question items. The comparism is illustrated in table 4.17. Due to width contraints, the question item headers were modified using:

Question 1: I find the Internet too confusing

Question 2: It is difficult to manage data on the Internet

Question 3: I am confident no one monitors what information i transmit on the internet

Question 4: The Internet is expensive

Question 5: It is easy to find porn on the internet

Question 6: There is too much information on the internet

Question 7: I always receive unwanted messages from unknown internet users

Question 8: My personal information can be easily stolen on the internet

Table 4.17 contains the inter-item correlation of question items pertaining to the privacy of data as well as perception of respondents within the context of internet, data processing.

Table 4.17. Inter-Item Correlation Matrix

	Question 1	Question 2	Question 3	Question 4	Question 5	Question 6	Question 7	Question 8
Question 1	1.000	.734	-.845	-.548	.801	.852	.844	.816
Question 2	.734	1.000	-.861	-.863	.932	.943	.903	.929
Question 3	-.845	-.861	1.000	.777	-.940	-.920	-.908	-.949
Question 4	-.548	-.863	.777	1.000	-.845	-.777	-.740	-.833
Question 5	.801	.932	-.940	-.845	1.000	.939	.909	.978
Question 6	.852	.943	-.920	-.777	.939	1.000	.961	.952
Question 7	.844	.903	-.908	-.740	.909	.961	1.000	.941
Question 8	.816	.929	-.949	-.833	.978	.952	.941	1.000

4.5 Providing personal information over the Internet

Respondents were further asked of the possibility and how easy it is to acquire personal information over the internet. The responses were summarized in table 4.18.

Table 4.18. My personal information can be easily stolen on the Internet

	Frequency	Percent
Strongly Agree	17	42.5
Agree	13	32.5
Neutral	7	17.5
Disagree	1	2.5
Strongly Disagree	2	5
Total	**40**	**100**

Source: Field Data, 2022

From the table 4.18, respondents both strongly agreed and agreed by indicating in the responses 17(42.5%) and 13(32.5%) respectively strongly agree and agree to easy access to personal information on the internet, 7(17.5%) remained neutral whiles 1 (2.5%) and 2(5%) strongly disagreed and disagreed respectively. The results from the table indicates respondents awareness of the volatile nature through easy access schemes to personal information shared over the internet. Regarding misappropriation of acquired data, responses were summarized in table 4.25, Confidentiality, Integrity and Availability are considered an essential components of the information security model, hence leveraging on the importance of this model, respondents were asked if someone might misuse their personal information shared on the internet, the responses were summarized as 18(45%) and 11(27.5%) strongly agreed and agreed respectively whiles 8(20%) remained neutral, however 1(2.5%) and 2(5%) strongly disagreed and disagreed respectively as shown in table 4.19.

Table 4.19. Misappropriate (misuse) personal information

	Frequency	Percent
Strongly Agree	18	45
Agree	11	27.5
Neutral	8	20
Disagree	1	2.5
Strongly Disagree	2	5
Total	**40**	**100**

Source: Field Data, 2022

Respondents were further asked if they websites frequently asked for the provision of personal information. The responses were tabulated in Table 4.20.

Table 4.20. Providing personal information to websites

	Frequency	Percent
No	10	25
Yes	30	75
Total	**40**	**100**

Source: Field Data, 2022.

From Table 4.20 when respondents were asked if they were frequently asked by websites to provide personal information, 30(75%) responded with Yes whiles 10(25%) indicated they were not asked. The summary of the responses reveals that websites were actively used as avenues for data acquisition and entry points for most websites. Respondents were further asked a follow up question about leaving their personal information on websites, they were asked to rate the level of comfortability by indicating if they were comfortable or not. Responses were summarized in table 4.21. From Table 4.21, almost all respondents 37(92.5%) indicated that they were not comfortable to share or leave their personal information on websites whiles 3(7.5%) responded they were not comfortable sharing their personal information on websites. They further indicated that there was an inherent tradeoff between respondents and the websites they visit. The services of most websites needed personal information to be able to better provide the needed resource for record keeping and for repudiated purposes as well.

Table 4.21. Do you feel comfortable sharing personal information on websites?

	Frequency	Percent
Comfortable	37	92.5
Not Comfortable	3	7.5
Total	**40**	**100**

Source: Field Data, 2022.

Respondents were asked if they indicated No from the question in Table 4.21, how long it took to produce incorrect identities. Table 4.22 presents a summary of the responses from participants.

Table 4.22. If No, how long does it take to provide a false identity to a website?

	Frequency	Percent
1 Hour	2	5
Less than an hour	38	95

Total	**40**	**100**

Source: Field Data, 2022.

From Table 4.22, 38(95%) respondents indicated that it took less than an hour to provide incorrect personal information to websites whiles 2(5%) also agreed that it takes almost an hour to provide incorrect personal details to websites. In lieu of this, it is realized that although providing personal information to some websites is important and necessary for the provision of services, respondents also had the tendency to provide incorrect information to bypass the security requirements. Respondents were asked to rank the listed conditions or reasons below which they agree to disclose personal information to websites, the results are summarized in the table 4.23. When respondents were asked to rank the listed reasons in table 4.23, 34(85%) was ranked first from the list of reasons from the respondent population strongly opposing the idea of giving out accurate personal information to websites or companies they are not trustworthy. 33(82.5%) indicated that they do not disclose their information if they are unsure what it will be used for, hence companies or websites need to explicitly define the reasons for data acquisition and processing. Moreover, 32(80%) also indicated that when the company of individual requesting for personal information is unknown, they (respondents) decline to provide personal information or will provide incorrect details. Some respondents further indicated that they were worried about eavesdropping or inappropriate acquisition of personal information by ranking this reason 4th with respondent value of 30(75%), also 29(72.5%) always declined to provide sensitive personal information such as bank account details, social security numbers amongst others to websites with 28(70%) prefer to remain anonymous online. The least ranked reasons were if the services provided does not merit the information requested 28(70%) and 26(65%) finding the time taken to complete the provision of personal information both time consuming and exhausting.

Table 4.23.Conditions for the refusal to disclose personal information to websites.

	Frequency	**Percent**
I do not trust the company or individuals	34	85
If they do not disclose what they use my information for	33	82.5
The company or individual is unknown	32	80
I am worried my information might be stolen through eavesdropping	30	75

Some information asked are sensitive	29	72.5
I prefer to remain anonymous online	28	70
The services rendered to me does not worth the information I give	28	70
Completing the requirements is time consuming and exhausting	26	65

Source: Field Data, 2022.

Respondents were asked of the importance of companies requesting for their personal information, the responses were summarized in table 4.24.

Table 4.24. Importance of Company requesting for personal information.

Labels	Frequency	Percent
No	8	20
Yes	32	80
Total	**40**	**100**

Source: Field Data, 2022

When asked about the importance of companies requesting for personal information, 32(80%) of the respondent population said yes whiles 2(20%) said they did not see the relevance of companies requesting for their personal information before rendering services to them.

4.5.1 Major Findings - Provision of personal information over the internet as a threat to security and privacy to the management of data.

The responses received from participants from this section revealed that respondents were aware of the susceptibility of personal information being stolen, compromised or misappropriated. 29(52%) of respondents strongly agreed that the possibility of their personal information being misappropriated was high but regardless, they were willing to provide their personal information to websites, companies and third parties in exchange for services rendered to them, a cummulative sum of 37(92.5%) of responses agreed that they will provide personal information to websites, whiles 3(7.5%) disagreed respectively. Although some respondents disagreed indicating that they were not comfortable sharing their personal information with websites, they further indicated that it was possible to provide fake personal information because they do not trust the individuals, websites or companies requesting for such information and when prompted, the time taken to produce such false information was less.

From the table, the most ranked condition for the refusal to disclose personal information was the absence of trust between the user and the individual, group, company or website requesting for the information. As part of the privacy-preservation practices within the context of data privacy as outlined by the US Department of Labor, (nd.) individual user(s) shares the responsibility of safeguarding and/or protecting their personal information and more so, contractors have a legal and moral obligation of respecting the privacy of users by using the acquired data for its intended purposes. The establishment of trust between transacting parties is the bed rock of data processing and management within the context of fog computing. Fog nodes operate on a trust based system for verfication, processing, storage and management of data. Any acquired personal information are processed locally by the fog nodes before onward transmission to the cloud, the lack of accurate data for fog nodes often results in the processing of wrong data with wrong insights, predictions and implementation. The least ranked condition which demanded the provision of incorrect personal information to complete the requirements which is sometimes time consuming and exhausting.

4.6 Familiarity with data processing

Regardless, to establish a foundation after considering all the risk factors if respondents would still provide personal information to websites and companies. The results were summarized in table 4.25.

Table 4.25. Willingness to provide personal information to websites and companies

	Frequency	**Percent**
No	13	32.5
Yes	27	67.5
Total	**40**	**100**

Source: Field Data, 2022.

From Table 4.25, 27(67.5%) of the respondent population agreed that they will provide personal information to websites whiles 13(32.5%) of respondents disagreed that they would not provide personal details to websites. The results from this table informs the researcher of the impact of the clarity and considerations from reflections.

Table 4.25 is a summary of the choice of respondents to provide personal information for a fee. The tabulated responses from participants in the table 4.26 indicates that 26(65%) of respondents

responded with a negative answer (No) whiles 14(35%) responded with a yes. The indication from this reveals that although the majority of respondents will not provide personal information for a fee, most would consider this option as an alternative means to an end.

Table 4.26. Provision of personal information to websites for a fee?

	Frequency	Percent
No	26	65
Yes	14	35
Total	**40**	**100**

Source: Field Data, 2022.

Out of the total number of respondents who participated in the study, 29(72.5%) responded positive to making online purchases whiles 11(27.5%) responded with No, indicating that they do not make online purchases. This is summarized in Table 4.27.

Table 4.27. Making online purchases

	Frequency	Percent
No	11	27.5
Yes	29	72.5
Total	**40**	**100**

Source: Field Data, 2022.

To determine the frequency of online purchases, respondents were asked to make a choice from the listed options in Table 4.28 as well, the responses were also summarized in table 4.28.

Table 4.28. Frequency of purchases over the Internet

	Frequency	Percent
All the time	5	12.5
Occasionally	26	65
Rarely	9	22.5
Total	**40**	**100**

Source: Field Data, 2022.

Out of the total respondents who participated in the study, majority of the respondents 26(65%) responded that they make purchases online occasionally with further 9(22.5%) indicating that they rarely make online purchases whiles 5(12.5%) also responded that they make online purchases all the time.

Using the 3-Point Likert Scale of (1 = Very Likely, 2 = Likely, 3 = Not likely) and premised on the previous question, respondents were asked if they consider making online purchases in the next six months from the survey. The responses were summarized in the Table 4.29.

Table 4.29. Likelihood of making online purchases in the next six months

	Frequency	Percent
Very likely	15	37.5
Likely	20	50
Not Likely	5	12.5
Total	**23**	**100**

Source: Field Data, 2022.

From the table 4.27, majority of respondents 20(50%) indicated that they are likely to make online purchases in the next six months, 15(37.5%) indicated they were very likely to make online purchases whiles 5(12.5%) indicated they would not likely make online purchases.

On the question of What is the significance of your consent with the under-listed conditions, Table 4.36 gives the distribution of their responses using the 5-point Likert scale of (Very Important = 1, Somewhat Important = 2, Neutral = 3, Unimportant = 4, Somewhat Unimportant = 5).

Table 4.30. Significance of Consent in giving personal information (F = Frequency, P = Percent)

	1		2		3		4		5	
	F	**P**	**F**	**P**	**F**	**P**	**F**	**P**	**F**	**P**
Sites sell/share your personal information with others	27	67.5	7	17.5	2	5	2	5	-	-
Sites track your movement around their site	24	60	6	15	3	7.5	3	7.5	2	5
Sites track your movement around the Internet	26	65	7	17.5	4	10	2	5	1	2.5
Sites track your online purchases	26	65	7	17.5	4	10	1	2.5	-	-
Sites gather in-depth personal profiles about you from other outside databases	28	70	6	15	4	10	1	2.5	1	2.5
Sites customize your online experience to your personal preferences	25	62.5	10	25	2	5	2	5	1	2.5

Source: Field Data, 2022.

When respondents were asked of the importance of consent when websites or companies decide to sell/share the personal information acquired from respondents, majority of the respondent population which represented 27(67.5%) strongly agreed that it is very important that their consent be sought before companies engage in the sale or sharing of their personal information, 7(17.5%) respondents also agreed, 2(5%) respondents each respectively remained neutral and disagreed.

When respondents were asked of the importance of consent before websites or companies track the movement of respondents around their site, majority of the respondents 24(60%) strongly agreed and 6(15%) agreed, 3(8%) remained neutral however, 3(7.5%) and 2(5%) strongly disagreed and disagreed respectively. Besides tracking individual activity on a website, it is possible to track a user's activity on other websites through browsing cookies, respondents were therefore asked of the importance of consent before the initiation of

movement tracking on the internet, 26(65%) and 7(17.5%) strongly agreed and agreed respectively, 4(10%) remained neutral, 2(5%) and 1(2.5%) strongly disagreed and disagreed respectively.

Respondents were further asked of their general consent if websites tracked through data retention of their online purchases, 26(65%) and 7(17.5%) strongly agreed and agreed respectively, 4(10%) remained neutral whiles 1(2.5%) disagreed. On the issue of websites gathering in-depth personal profile information about respondents with consent, 28(70%) and 6(15%) strongly agreed and agreed respectively, 4(10%) remained neutral whiles 1(2.5%) for each strongly disagreed and disagreed that consent was not necessary for acquiring such information. Finally on the importance of consent when sites customize respondents online experience to meet the personal preferences, 25(62.5%) and 10(25%) respectively strongly agreed and agreed, 2(5%) remained neutral to this reason whiles 2(5%) and 1(2.5%) disagreed and strongly disagreed respectively.

Table 4.31. Recording of online activities WITH Consent.

	Frequency	**Percent**
No	20	50
Yes	20	50
Total	**40**	**100**

Source: Field Data, 2022.

Respondents were asked on what constituted an invasion of privacy, either tracking respondent's online activities WITH and WITHOUT consent. Table 4.31 summarizes the responses about recording online activities WITH consent, if it constitutes privacy invasion, equal number of respondents agreed and disagreed by selecting Yes 20(50%) and No 20(50%), however table 4.32 which asked about the recording of online activities WITHOUT consent, 8(20%) responded No whiles 32(80%) responded Yes.

Table 4.32. Recording of online activities WITHOUT Consent.

	Frequency	**Percent**
No	8	20
Yes	32	80
Total	**40**	**100**

Source: Field Data, 2022.

Communication and data sharing over the internet has precedents in telephone calls, mail by mail, faxes and in-person meetings. Alternatively, respondents were asked to indicate which of the listed mediums they were concerned about the privacy of sharing personal information, using a 5-Point Likert Scale of (Much more concerned = 1, more concerned = 2, Neutral = 3, less concerned = 4, somewhat less concerned = 5) to indicate their level of concern. Table 4.31 shows a summary of the responses.

Table 4.33. Distribution of Privacy concerns over other communication media

	1		**2**		**3**		**4**		**5**		**Total**
	F	**P**	**F**	**P**	**F**	**P**	**F**	**P**	**F**	**P**	
Telephone	7	17.5	11	27.5	2	5	9	22.5	11	27.5	**40**
Mail (By Post)	3	7.5	11	27.5	3	7.5	10	25	13	32.5	**40**
Fax	2	5	9	22.5	4	10	11	27.5	14	35	**40**
In-person	7	17.5	8	20	5	12.5	10	25	10	25	**40**

Source: Field Data, 2022

The results from Table 4.33 revealed that the comparison between internet and each of the listed mediums, respondents were much less concerned with privacy on mediums either than the internet, in the case of privacy of exchanges with telephones, 11(27.5%) and 9(22.5%) responded they were somewhat less concerned and less concerned respectively, 2(5%) respondents remained neutral whiles 7(17.5%) and 11(27.5%) were much more concerned. Communication preference comparison with mail(by post), 13(32.5%) and 10(25%) responded they were somewhat less concerned and less concerned respectively 3(7.5%) respondents remained neutral whiles 11(27.5%) and 3(7.5%) were concerned and much more concerned respectively.

More so, respondents also indicated that comparing the privacy of Fax to internet, 14(35%) and 11(27.5%) were both somewhat unconcerned and unconcerned, 4(10%) remained neutral whiles 9(22.5%) and 2(5%) were very much concerned and concerned

respectively. For in-person communication, a cumulative sum of 10(25%) somewhat were unconcerned whiles 10(25%) were concerned about the privacy of their data, however 5(12.5%) remained neutral and 7(17.5%) as well as 8(20%) were strongly concerned and concerned respectively.

4.6.1 Test of reliability - Fog Data

Table 4.34. Test of reliability from Table 4.31

Cronbach's Alpha	Cronbach's Alpha Based on Standardized Items	N of Items
.985	**.986**	**4**

Source: Field Data, 2022.

The results from Table 4.34 reveals an alpha score of 0.985. The variation in alpha scores from an administered questionnaire is interpreted to mean a high level of consistency between measured items, hence dictates that the questionnaire used as well as the prediction accuracy were reliable. An inter-item correlation matrix was further estimated from the acquired data responses and summarized in table 4.33

Table 4.35. Inter-Item Correlation Matrix from Table 4.31

	Telephone	Mail (By Post)	Fax	In-person
Telephone	1.000	.939	.909	.978
Mail (By Post)	.939	1.000	.961	.952
Fax	.909	.961	1.000	.941
In-person	.978	.952	.941	1.000

Source: Field Data, 2022.

The idea of inter-item correlation is to obtain a near perfect correlation with similar test items with the same scale, the diagonal values of $r = 1.000$ depicted a perfect correlation when compared with itself. The questions within the section measured the same concept and hence the coefficients obtained were in the range of $(0.5 \geq r \leq 1)$. Telephone when compared with Mail (By Post) revealed an r value of 0.939, Fax (0.909) and in-person (0.978) respectively. Mail (By post) compared with Telephone was 0.939, Fax (0.962) and In-person (0.952) respectively. More so, Fax had an r value of 0.909 when compared with telephone, 0.961 with Mail (By Post) and 0.941 with In-person. Furthermore, In-person had r matrix comparison values of 0.978, 0.952, 0.941 for each of the respective items of Telephone, Mail (By Post), and Fax.

Interpretation of Results of Cronbach's Alpha

The reliability analysis was carried out using the Cronbach's Alpha on the comparison of privacy concerns over other communication media which comprised 8 items. The estimated alpha score of $\boldsymbol{\alpha = 0.985}$ revealed that the questions were acceptable and reliable as most question items included were worthy of being included since a low alpha score would be interpreted as a need to be excluded or deleted. The results from the Inter-Item matrix correlation further supported the estimated value since the question items recorded acceptable high r values which ranged $\boldsymbol{(0.5 \geq r \leq 1)}$.

Table 4.36. Listed reasons for choice of alternative media for transmissions

	Concerned		Neutral		Not Concerned	
	Frequency	**Percent**	**Frequency**	**Percent**	**Frequency**	**Percent**
My information could be stolen, it's not safe	31	77.5	5	12.5	4	10
It is not clear to me how my information will be used	35	87.5	5	12.5	-	-
It is unclear who I am dealing with	31	77.5	6	15	3	7.5
I don't trust the website with my personal information	31	77.5	6	15	3	7.5
Mainly because I'm unfamiliar with this modern technology	21	52.5	6	15	13	32.5
My privacy has been abused on the Internet	24	60	5	12.5	11	27.5
Someone I know had their privacy violated online	25	62.5	6	15	9	22.5

Source: Field Data, 2022

To support the choices made in Table 4.34, respondents were asked to choose from a list of suggested reasons. Table 4.34 indicates that 31 respondents were influenced by concerns about data theft occurring within the enclave of the internet and were concerned about their privacy when communicating over the selected medium, five (5) remained neutral and four (4) respondents indicated they were not concerned about privacy or security of their data. The other 35 respondents further indicated that they were concerned because there

were no clear indications as to how their information would be used, while five (5) respondents remained neutral. Three-quarters (37) of respondents indicated the actual individual, company or group was concealed, which influenced their choice of transmission media. Thirty-one (31) respondents indicated that trusting the individual or company further affected their choice. Additionally, 21 (52.5%) and 24 (62.5%) respondents indicated that being unfamiliar with modern technologies, privacy abuse on the internet, and knowledge of a close relative being abused online were also reasons.

Table 4.37. Data management by fog servers and websites.

	Concerned		**Neutral**		**Not Concerned**	
	Frequency	**Percent**	**Frequency**	**Percent**	**Frequency**	**Percent**
Education	17	42.5	7	17.5	8	20
Entertainment	26	65	7	17.5	7	17.5
Work-related Research	17	42.5	7	17.5	7	17.5
Personal Finance	18	45	5	12.5	5	12.5
Accessing Current Affairs(News, Weather, Sports)	10	25	10	25	11	27.5
Travel	12	30	9	22.5	7	17.5
Product Information Gathering	13	32.5	6	15	10	25

Source: Field Data, 2022

The results of the summary from respondents on how confident they were in the management of your personal data by fog servers with a list of suggested reasons. For education, the 42.5% were concerned, 17.5% were neutral and 20% were not concerned. For entertainment, 65% were concerned, 17.5% remained neutral and unconcerned respectively. For work-related research, 42.5% indicated they were concerned whiles 17.5% were both neutral and unconcerned respectively. Personal finance had 45% of the respondent population concerned about the privacy of data transmissions, 12.5% for both neutral respondents and not concerned respondents

respectively. Moreover, Accessing Current Affairs (News, Weather, Sports) 25% of respondents were both concerned and neutral, whiles 27.5% were not concerned. Travel further had 30% of respondents concerned, 22.5% neutral and 17.5% not concerned whiles product gathering as the final reason listed had 32.5% concerned, 15% neutral and 25% unconcerned. The results from table 4.41 indicates that reinforced notion that respondents although prefer the internet usage as a paramount means of transmitting information, they were aware of the inherent security and privacy challenges in the other media and hence depending on the use case scenario, switched between media for respective purposes.

Chapter Five

Summary, Conclusion and Recommendation

5.0 Introduction

In this chapter the researcher presents summary of the results from the data collected. Conclusions are drawn based on the results from the survey and finally make recommendations for appropriate audience.

5.1 Summary of Findings

The general aim of the study was to assess the impact of security and privacy on the management of data within the fog computing environment. The demographic results from the administered questionnaire revealed that African participants were marginally more than the other represented participants from the other mentioned continents, all respondents were experienced internet users who have been using the internet and its associated resources for various purposes. Owing to the results of the age distribution, majority of participants were between the ages 24 and 34 which indicates that no minors participated in the study, adults who were fully aware of the ramifications of online dangers participated in the study.

Privacy-preservation is key to the continued use of services within the context of internet usage, as a result respondents indicated that they were experienced internet users who have been using the internet and its associated resources for various purposes, however respondents neither agreed nor disagreed with the possibility of the tracking or monitoring of their usage of the internet. The perception of respondents influenced the usage of the internet and various computing devices.

Data privacy is an important aspect in the determination of the safety and privacy of data within the context of fog computing and interconnection of multiple devices. Respondents

indicated they were often asked to provide personal information over the internet to websites and unsuspecting individuals. The study revealed that cloud, fog and generally internet security is not an abstract concept respondent were oblivious to, hence the results about the perception of users about the complexity of data privacy-preservation. Other factors that impact the provision of personal information over the internet were:

1. Receiving unsolicited messages and emails from unknown parties
2. The ease of access as well as the readily accessible availability of information as indicated by some respondents was of high concern.
3. The cost associated with the usage of the internet
4. Easy access to pornographic materials and media

Participants were aware of the susceptibility of personal information being stolen, compromised or misappropriated. 29(52%) of respondents strongly agreed that the possibility of their personal information being misappropriated was high but regardless, they were willing to provide their personal information to websites, companies and third parties in exchange for services rendered to them.

Conclusions

Throughout this study, security and privacy were examined in relationship to the management of information within a fog computing environment. SPSS and Microsoft Excel were used to analyze data collected by an online questionnaire. Following a thorough analysis of the information gathered and analyzed, data privacy has been determined to be an important and integral aspect of fog computing and the creation of data privacy within the context of interconnected devices. In light of the fact that cloud, fog, and internet security are not abstract concepts nor is it something that should be oblivious to, the results of the study on perceptions of users regarding data privacy preservation revealed that respondents were aware of the dangers associated with the use, however they formed behavioral patterns using the medium, even though it was complemented with other mediums.

Future Works

The results in this thesis also provide a strong foundation for future work in awareness with the security and privacy lapses within the fog computing environment. `

References

Abdulqadir, H. R., Zeebaree, S., Shukur, H., Sadeeq, M., Salim, B., Salih, A., & Kak, S. (2021). A Study of Moving from Cloud Computing to Fog Computing. *Qubahan Academic Journal*, 60 – 70. doi:10.48161/qaj.v1n2a49

Admin. (2019, September 6). *Developing IoT Applications – Best Technologies and Tools for IoT Developers*. Retrieved from Vizah Engineering Solutions: https://vizah.ch/en/developing-iot-applications-best-technologies-and-tools-for-iot-developers/

Agrawal, D., Sudipto, D., & Amr , E. (2011). Big Data and Cloud Computing: Current State and Future Opportunities. *Proceedings of the 14th International Conference on Extending Database Technology - EDBT/ICDT '11 - (2011.03.21-2011.03.24)]* (pp. 22 - 24). Uppsala: ACM Press the 14th International Conference. doi:10.1145/1951365.1951432

Akhtar, N., Kerim, B., Perwej, Y., Tiwari, A., & Praveen, S. (2021). A Comprehensive Overview of Privacy and Data Security for Cloud Storage. *International Journal of Scientific Research in Science, Engineering and Technology, Volume 8*(Issue 5), 113 - 152. doi:10.32628/IJSRSET21852

Al-Hussaini, K., Zainol, S., Ahmed, R., & Daud, S. (2018). IoT Monitoring and Automation DataAcquisition for RecirculatingAquaculture System Using Fog Computing. *Journal of Computer Hardware Engineering, volume 1*. doi:10.63019/jche.v1i2.610

Alwakeel, A. M. (2021). An Overview of Fog Computing and Edge Computing Security and Privacy Issues. *Multidisciplinary Digital Publishing Institute*, 21 - 24. doi:10.3390/s21248226

Armbrust, M., Fox, A., Griffith, R., Joseph, A., Katz, R., Konwinski, A., . . . Zaharia, M. (2010, April). Clearing the clouds away from the true potential and obstacles posed by this computing capability. *Communications of the ACM*, pp. pp. 50 - 58. doi:10.1145/1721654.1721672

Arvind, N., & Vitaly , S. (2008). Robust De-Anonymization of Large Sparse Datasets. *The Proceedings of 29th IEEE Symposium on Security and Privacy.* Oakland. Retrieved from http://www.cs.utexas.edu/~shmat/shmat_oak08netflix.pdf/.

Atzori, M. (2016, October). *Blockchain-Based Architectures for the Internet of Things: A Survey*. doi:10.2139/ssrn.2846810

Banerjee, A., & Chaudhury, S. (2010). Statistics without tears: Populations and samples. *Industrial Psychiatry Journey*, pp. 60 – 65. doi:10.4103/0972-6748.77642

Bhatia, R., & Sood, M. (2018). Security of Big Data: A Review. *2018 Fifth International Conference on Parallel, Distributed and Grid Computing (PDGC)*, pp. 182 - 186. doi:10.1109/PDGC.2018.8745874

Bhojaraju, G., & Koganurmath, M. (2003). Research Gate. *Proceedings of Knowledge management in special libraries in digital environment : XXIV All India Conference of IASLIC,* (pp. pp. 385 - 398). Indian Association of Special Libraries & Information Centres. Retrieved December 23, 2021, from https://www.researchgate.net/publication/257298522_Database_Management_Concepts_and_Design

blockchainresearchinstitute. (nd.). *What is a Blockchain?* Retrieved from https://www.blockchainresearchinstitute.org/an-intro-to-blockchain-and-nfts/

Bouchrika, I. (2020, September 25). *Types of Research Design: Perspective and Methodological Approaches*. Retrieved April 2, 2022, from Research.com: https://research.com/research/types-of-research-design

Cambridge Dictionary. (n.d). *Meaning of fog in English*. Retrieved from Cambridge Dictionary: https://dictionary.cambridge.org/dictionary/english/fog

Carlin, S., & Curran, K. (2011). Cloud Computing Security. *International Journal of Ambient Computing and Intelligence*, pp 14 - 19. doi:http://dx.doi.org/10.4018/jaci.2011010102

Chan, S. (2017, November 09). *Fog brings the cloud closer to the ground: Cisco innovates in fog computing*. Retrieved January 29, 2022, from The Newsroom: Cisco's Technology News Site: https://newsroom.cisco.com/feature-content?type=webcontent&articleId=1894659

Chen, Z., Wang, G., Hu, S., & Wei, H. (2015). Independence and controllability of big data security. *Chinese Science Bulletin, volume 60*, pp. 427- 432. doi:10.1360/N972014-00812

Chokhani, S., & Ford, W. (1999). "Internet X.509 Public Key Infrastructure Certificate Policy and Certification Practices Framework", RFC 2527,. *Network Working Group: Request for Comments: 2527*. doi:10.17487/RFC2527

Christensen, G. (nd.). *Distributed Computing for IoT: Data Management in a Fog Computing Environment*. Retrieved April 1, 2022, from eogogics: https://eogogics.com/distributed-computing-iot-data-management-fog-computing-environment/

Cichy, C., & Rass, S. (2019). An Overview of Data Quality Frameworks. *IEEE Access*, pp. 1 – 1. doi:10.1109/ACCESS.2019.2899751

Cimpan, A. (2020, May 18). *How to conduct user interviews*. Retrieved from andra-cimpan: https://andra-cimpan.medium.com/how-to-conduct-and-analyze-user-interviews-a013a44a98d5

Cisco. (nd.). *Cisco Industrial Security; Protect industrial operations against cyberthreats: Gain visibility and control over your OT and ICS with Cisco Industrial Threat Defense*. Retrieved from Cisco.com: https://www.cisco.com/c/en/us/solutions/internet-of-things/iot-security.html

Clark, B. J. (2016, November 17). *What is the Internet of Things (IoT)?* Retrieved December 24, 2021, from IBM: https://www.ibm.com/blogs/internet-of-things/what-is-the-iot/

cloudflare. (n.d). *What is encryption? | Types of encryption*. Retrieved March 15, 2022, from Cloudflare: https://www.cloudflare.com/learning/ssl/what-is-ssl/

CloudFlare. (n.d). *What is the cloud? | Cloud definition*. Retrieved February 21, 2022, from CloudFlare: https://www.cloudflare.com/learning/cloud/what-is-the-cloud/

Cronbach, L. J. (1951). The Coefficient Alpha and the Internal Structure of Tests. *Psychometrika, Volume 16. No. 3*, pp. 297 - 334. Retrieved April 30, 2022, from https://link.springer.com/article/10.1007/bf02310555

CyberEdu. (n.d). *What is Data Encryption? Data Encryption Defined, Explained, and Explored*. Retrieved from Forcepoint: https://www.forcepoint.com/cyber-edu/data-encryption

Dawson, H. (2019, June 27). *The Most Influential Security Frameworks of All Time*. Retrieved March 26, 2022, from Infosecurity Group: https://www.infosecurity-magazine.com/opinions/most-influential-frameworks-1-1-1/

Desikan, K. E., Srinivasan, M., & Murthy, C. (2017). A Novel Distributed Latency-Aware Data Processing in Fog Computing-Enabled IoT Networks. *Proceedings of the ACM Workshop on Distributed Information Processing in Wireless Networks.* doi:10.1145/3083181.3083183

Desikan, K. E., Srinivasan, M., & Murthy, C. (2017). A Novel Distributed Latency-Aware Data Processing in Fog Computing-Enabled IoT Networks. *Proceedings of the ACM Workshop on Distributed Information Processing in Wireless Networks - DIPWN'17* - (pp. pp. 1 - 6). Chennai, India: [ACM Press the ACM Workshop - (2017.07.10-2017.07.14)]. doi:10.1145/3083181.3083183

Desjardins, J. (2019, April 17). *How much data is generated each day?* Retrieved December 7, 2021, from World Economic Forum: https://www.weforum.org/agenda/2019/04/how-much-data-is-generated-each-day-cf4bddf29f/

Desjardins, J. (2019, 16 July). *Why Big Data Keeps Getting Bigger*. Retrieved February 11, 2022, from Technology: https://www.visualcapitalist.com/big-data-keeps-getting-bigger/

D'Souz, F. (2022, January 15). *The IoT ecosystem in 2022: components, industry alliances and legal environments*. (Thales Group) Retrieved February 5, 2022, from THALES: https://www.thalesgroup.com/en/markets/digital-identity-and-security/iot/inspired/iot-building-blocks

Dsouza, C., Ahn, G.-J., & Taguinod, M. (2014). Policy-Driven Security Management for Fog Computing: Preliminary Framework and A Case Study. *Proceedings of the 2014 IEEE 15th International Conference on Information Reuse and Integration (IEEE IRI 2014)*, (pp. pp. 16 - 23). doi:10.1109/IRI.2014.7051866

Dubey, H., Yang, J., Constan, N., Amiri, A., Yang, Q., & Makodiya, K. (2015). Fog Data: Enhancing Telehealth Big Data Through Fog Computing. *ASE BD&SI '15: Proceedings of the ASE BigData & SocialInformatics 2015*, (pp. pp. 1 - 6). doi:10.1145/2818869.2818889

Duffy, C. (2022, February 15). *Meta agrees to pay $90 million to settle lawsuit over Facebook tracking users' online activity*. Retrieved February 2, 2022, from CNN.com: https://edition.cnn.com/2022/02/15/tech/facebook-internet-tracking-settlement/index.html

Duggal, N. (2021, December 15). *What Is Data Processing: Cycle, Types, Methods, Steps and Examples*. Retrieved March 3, 2022, from simplilearn: https://www.simplilearn.com/what-is-data-processing-article

Ema, R. R., Islam , T., & Ahmed , H. (2019). Suitability of Using Fog Computing Alongside Cloud Computing. *IEEE Conference on Electrical,Computer* , 6 - 8. doi:10.1109/ICCCNT45670.2019.8944906

EMC Education Services. (2015). *Data Science & Big Data Analytics : Discovering, Analyzing, Visualizing and Presenting Data.* Indianapolis: John Wiley & Sons, Inc.

emotiv. (nd.). *Data Privacy*. Retrieved April 28, 2022, from emotiv.com: https://www.emotiv.com/glossary/data-privacy/

Fang, W., Wen, X., Zheng, Y., & Zhou, M. (2016). A Survey of Big Data Security and Privacy Preserving. *IETE Technical Review ( )*, pp. 1 – 17. doi:10.1080/02564602.2016.1215269

Fang, W.-L., Wen, X., Zheng, Y., & Zhou, M. (2017). A Survey of Big Data Security and Privacy Preserving. *IETE Technical Review, volume 34*, 544 - 560. doi:10.1080/02564602.2016.1215269

Fiandrino, C., Anjomshoa, F., Kantarci, B., Kliazovich, D., Bouvry, P., & Matthews, J. (2017). Sociability-Driven Framework for Data Acquisition in Mobile Crowdsensing over Fog Computing Platforms for Smart Cities. *IEEE TRANSACTIONS ON SUSTAINABLE COMPUTING*, 2377-3782 . doi:10.1109/TSUSC.2017.2702060

Fiandrino, C., Anjomshoa, F., Kantarci, B., Kliazovich, D., Bouvry, P., & Matthews, J. (2017). Sociability-Driven Framework for Data Acquisition in Mobile Crowdsensing Over Fog

Computing Platforms for Smart Cities. *IEEE Transactions on Sustainable Computing*, pp. 345-358. doi:10.1109/TSUSC.2017.2702060

Foote, K. D. (2021, December 21). *The Importance of Data Quality Tools*. Retrieved May 03, 2022, from Dataversity: Data Topics: https://www.dataversity.net/the-importance-of-data-quality-tools/

Fortinet. (2022, March 12). *Live Threat Map*. Retrieved from Fortinet: https://threatmap.fortiguard.com/

Frost, J. (nd.). *Cluster Sampling: Definition, Advantages & Examples*. Retrieved April 5, 2022, from Statistics By Jim; Making statistics intuitive: https://statisticsbyjim.com/basics/cluster-sampling/

Fürber, C. (2015). In *"3. Data Quality". Data Quality Management with Semantic Technologies.* Springer.

Gandomi, A., & Haider, M. (2015). Beyond the hype: Big data concepts, methods, and analytics. *International Journal of Information Management, 35*, 137–144. doi:10.1016/J.IJINFOMGT.2014.10.007

Goforth, C. (2015, November 16). *Using and Interpreting Cronbach's Alpha.* Retrieved April 30, 2022, from University of Virginia Library - Research Data Services + Sciences: https://data.library.virginia.edu/using-and-interpreting-cronbachs-alpha/

Herzog, T., Scheuren, F., & Winkler, W. (2007). "Chapter 2: What is data quality and why should we care?". Data Quality and Record Linkage Techniques. . Springer Science & Business Media.

Hussain, F., Rasheed , H., Hassan , S., & Hossain, E. (2020). Machine Learning in IoT Security: Current Solutions and Future Challenges. *IEEE Communications Surveys & Tutorials, 22(3),*, pp. 1686 - 1721. doi:10.1109/COMST.2020.2986444

ibm. (n.d). *Big data analytics*. Retrieved from ibm: https://www.ibm.com/analytics/big-data-analytics

ibm.com. (n.d). *Leverage effective big data analytics to analyze the growing volume, velocity and variety of data for the greatest insights* . Retrieved from IBM: https://www.ibm.com/analytics/big-data-analytics

informatica. (nd.). *The Basic Components of Data Quality*. Retrieved May 03, 2022, from informatica.com: https://www.informatica.com/resources/articles/what-is-data-quality.html

Ishwarappa , K., & Januradha, A. (2015). A brief Introduction on Big Data 5vs characteristics and Hadoop Technology. *International Conference on Intelligent Computing, Communication & Convergence (ICCC - 2015) - Conference Organized by Interscience Institutue of Management and Technology* (pp. 319 - 324). Bhubaneswar, Odisha India: Elsevier B.V. doi:10.1016/j.procs.2015.04.188

Jewers, C. (2022, February 17). *Meta agrees to pay out $90million to settle lawsuit claiming it violated privacy by tracking users even AFTER they had logged off from Facebook.* Retrieved February 19, 2022, from Daily Mail UK: https://www.dailymail.co.uk/news/article-10522773/Meta-agrees-pay-90million-settle-lawsuit-claiming-violated-privacy.html?ito=social-facebook

Jie Cui, Antonio , L., Ke , G., & Lu , L. (2022, January 1). *Data Security and Privacy for Fog/Edge Computing-Based IoT.* Retrieved from Security and Communication Networks - Hindawi: https://www.hindawi.com/journals/scn/si/769058/

Kafhali, S. E., Chahir, C., Hanini, M., & Salah, K. (2019). Architecture to manage Internet of Things Data using Blockchain and Fog Computing. *Proceedings of the 4th International Conference on Big Data and Internet of Things.* (pp. pp. 3 – 24). Rabat, Morocco: Association for Computing Machinery. doi:10.1145/3372938.3372970

Kapil, G., Agrawal, A., & Khan, R. (2018). Big Data Security challenges: Hadoop Perspective.

Kaspersky. (2022, March 12). *CYBERTHREAT REAL-TIME MAP*. Retrieved from Kaspersky: https://cybermap.kaspersky.com/stats#country=213&type=OAS&period=w

kaspersky. (2022). *What is Data Encryption?* Retrieved March 15, 2022, from Kaspersky: https://www.kaspersky.com/resource-center/definitions/encryption

Khan, S., & Simon , P. (2017). Fog computing security: a review of current applications and security solutions. *Journal of Cloud Computing*, pp. 6 - 19. doi:10.1186/s13677-017-0090-3

Khan, S., Parkinson, S., & Qin, Y. (2017). Fog computing security: a review of current applications and security solutions. *Journal of Cloud Computing: Advances, Systems and Applications*, pp. 6 - 19. doi: 10.1186/s13677-017-0090-3

Khanum, N. K., Lathar, P., & Siddesh, G. (2021). Confidentiality and Safekeeping Problems and Techniques in Fog Computing. *Research Anthology on Architectures, Frameworks, and Integration Strategies for Distributed and Cloud Computing*.

Kim, J., & Kim, Y. (2015). Benefits of cloud computing adoption for smart grid security from security perspective. *The Journal of Supercomputing*, *72*, 3522 - 3534. doi:10.1007/s11227-015-1547-0

Kirvan, P., & Joseph, G. (2021, December). *Top 10 IT security frameworks and standards explained*. Retrieved March 27, 2022, from TechTarget: https://www.techtarget.com/searchsecurity/tip/IT-security-frameworks-and-standards-Choosing-the-right-one

Lan, D., Liu, Y., Taherkordi, A., Eliassen, F., Delbruel, S., & Liu, L. (2021). A federated fog-cloud framework for data processing and orchestration. *Proceedings of the 36th Annual ACM Symposium on Applied Computing.* (pp. pp. 729 - 736). New York, NY: ACM ISBN 978-1-4503-8104. doi:10.1145/3412841.3444962

Lee, W. W., Wolfgang, Z., & Henry , C. (2016). An Ethical Approach to Data Privacy Protection. *ISACA Journal | Information Technology & Systems Resources*, pp. 1 - 9. Retrieved from https://www.isaca.org/resources/isaca-journal/issues/2016/volume-6/an-ethical-approach-to-data-privacy-protection

Lewis, L. (2021, April 13). *Infographic: What Happens In An Internet Minute 2021*. Retrieved February 18, 2022, from Allaccess: https://www.allaccess.com/merge/archive/32972/infographic-what-happens-in-an-internet-minute

Li, C., Xue, Y., Jing Wang, Zhang, W.-g., & Li, T. (2018). Edge-Oriented Computing Paradigms. *ACM Computing Surveys (CSUR), Volume 38*, pp. 1 - 34. doi:10.1145/3154815

Li, J., Fong, S., Li, T., & Song, W. (2018). Data Stream Mining with Swarm Decision Table in Fog Computing Environment. *Proceedings of the 2018 2nd International Conference on Big Data and Internet of Things - BDIOT 2018* (pp. pp. 37–42.). Beijing, China: [ACM Press the 2018 2nd International Conference - Beijing, China (2018.10.24-2018.10.26)]. doi:10.1145/3289430.3289459

Li, X. (2021). The Impact of Big Data on People and Data Security Issues. *Proceedings of the 2021 5th International Seminar on Education, Management and Social Sciences (ISEMSS 2021)*. doi:10.2991/assehr.k.210806.093

Lin, H., Weng, B., Pan , J., Lin, C.-p., & Yang, Q. (2021). Application of Wireless Sensor Networks in the Sensitive Data Security of Intelligent Data Center under the Big Data Environment. *Journal of Physics: Conference Series, 1982*. doi:10.1088/1742-6596%2F1982%2F1%2F012017

Mahanti, R. (2019). In *"Chapter 1: Data, Data Quality, and Cost of Poor Data Quality". Data Quality: Dimensions, Measurement, Strategy, Management, and Governance.*

Mahmud, R., Ramamohanarao, K., & Buyya, R. (2018). Latency-Aware Application Module Management for Fog Computing Environments. *ACM Trans. Internet Technol. 19, 1, Article 9*, pp.1 - 21. doi:10.1145/3186592

Malik, M. W., Diyanatul , H., Inggrid , N., Purnama, I., Afif, N., & Anak , A. (2020). Development of Medical Image Encryption System Using Byte-Level Base-64 Encoding and AES Encryption Method. *ICCIP '20: The 6th International Conference on Communication and Information Processing* (pp. 153 - 158). New York: Association for Computing Machinery. doi:10.1145/3442555.3442580

Malik, M. W., Diyanatul , H., Purnama, I., Nurtanio, I., Afif, N., & Ratna, A. (2020). Development of Medical Image Encryption System Using Byte-Level Base-64 Encoding and AES Encryption Method. *2020 the 6th International Conference on Communication and Information Processing.* doi:10.1145/3442555.3442580

Marr, B. (2014, March 14). *Big Data: The 5 Vs Everyone Must Know*. Retrieved February 19, 2022, from LinkedIn: https://www.linkedin.com/pulse/20140306073407-64875646-big-data-the-5-vs-everyone-must-know

McNally, C. (2022, April 12). *How Much Do Internet and Wi-Fi Cost?* Retrieved April 13, 2022, from reviews.org: https://www.reviews.org/internet-service/how-much-do-internet-and-wi-fi-cost/

Mell, P., & Grance, T. (2011). *Computer Security : The NIST Definition of Cloud Commputing.* Gaithersburg: National Institute of Standards and Technology. Retrieved December 2, 2021, from https://nvlpubs.nist.gov/nistpubs/Legacy/SP/nistspecialpublication800-145.pdf

Microsoft. (2021, January 7). *Public Key Infrastructure*. Retrieved from docs.microsoft.com: https://docs.microsoft.com/en-us/windows/win32/seccertenroll/public-key-infrastructure

Miri, F., & Pazzi, R. (2021). A Comprehensive Survey on the Convergence of Vehicular Social Networks and Fog Computing. *arXiv:2112.00143v1*, 2 - 6.

Mohammadreza, A., Kumar, N., Eskandari, A., Ahmed, H., Vidal de Oliveira, A., & Chopra, S. (2020). Photovoltaic Solar Energy Conversion || Solar PV systems design and monitoring. In M. Aghaei, S. Gorjian, & Ashish Shukla (Eds.), *Chapter 5 - Solar PV systems design and monitoring*, (pp. pp. 117 – 145.). Academic Press. doi:10.1016/B978-0-12-819610-6.00005-3

Moses, B. (2020, July 17). *What is Data Reliability?; And how to use it to start trusting your data.* Retrieved from Towards Data Science: https://towardsdatascience.com/what-is-data-reliability-66ec88578950

Moura, J. A., & Carlos , S. (2015). Handbook of Research on Trends and Future Directions in Big Data and Web Intelligence. In *Security and Privacy Issue of Big Data* (pp. pp. 20 - 58). Hershey PA, USA: IGI Global. doi:10.4018/978-1-4666-8505-5.ch002

Mukherjee, M., Matam, R., Shu, L., Maglaras, L., Ferrag, M., Choudhury, N., & Kumar, V. (2017). Security and Privacy in Fog Computing: Challenges. *IEEE Access, volume 5*, 19293 - 19304. doi:10.1109/ACCESS.2017.2749422

Mukherji, S., & Shashwat, S. (2015). Pros and Cons of Cloud Computing Technology. *International Journal of Science and Research (IJSR)*(ISSN (Online): 2319-7064), pp 848 - 851. Retrieved from https://www.ijsr.net/archive/v5i7/ART2016314.pdf

Nadeem, M. A., & Saeed, M. A. (2016). Fog computing: An emerging paradigm. *Conference: 2016 Sixth International Conference on Innovative Computing Technology (INTECH)*, (pp. 83 - 86). doi:10.1109/INTECH.2016.7845043

Nandury, S. V., & Begum, B. (2017). Big Data for Smart Grid Operation in Smart Cities. *IEEE 2017 International Conference on Wireless Communications, Signal Processing and Networking (WiSPNET)* (pp. 1507–1511.). Chennai - India: International Conference on Wireless Communications, Signal Processing and Networking (WiSPNET). doi:10.1109/WiSPNET.2017.8300013

National Institute of Standards and Technology. (2014, February 12). Framework for Improving Critical Infrastructure Cybersecurity. *Cybersecurity Framework*. Retrieved from https://www.nist.gov/system/files/documents/cyberframework/cybersecurity-framework-021214.pdf

National Security Agency. (2020). *National Security Agency | CyberSecurity Information.* Retrieved March 5, 2022, from defense.gov: https://media.defense.gov/2020/Jan/22/2002237484/-1/-1/0/CSI-MITIGATING-CLOUD-VULNERABILITIES_20200121.PDF

Neware, R. (2019). Fog Computing Architecture, Applications and Security Issues: A Survey. *International Journal of Fog Computing (IJFC) 2020*. doi:10.4018/IJFC.2020010105

Nguyen, S., Salcic, Z., & Zhang, X. (2018). Big Data Processing in Fog - Smart Parking Case Study. *IEEE Intl Conf on Parallel & Distributed Processing with Applications, Ubiquitous Computing & Communications, Big Data & Cloud Computing, Social Computing & Networking, Sustainable Computing & Communications*, pp. 127 - 133. doi:10.1109/BDCloud.2018.00031

Nikoui, T. S., Rahmani, A., & Tabarsaied, H. (2019). Data Management in Fog Computing. In R. Buyya, & S. Srirama, *Fog and Edge Computing: Principles and Paradigms*, *First Edition.* John Wiley & Sons, Inc.

NSA. (2020). *National Security Agency | CyberSecurity Information.* Retrieved March 5, 2022, from defense.gov: https://media.defense.gov/2020/Jan/22/2002237484/-1/-1/0/CSI-MITIGATING-CLOUD-VULNERABILITIES_20200121.PDF

Oden, C. (nd.). *Validity and Reliability of Questionnaires: How to Check*. Retrieved April 30, 2022, from projecttopics: https://www.projecttopics.org/validity-and-reliability-of-questionnaires-how-to-check.html

Oke, A. E., Kineber, A., Al-Bukhari, I., Famakin, I., & Kingsley, C. (2021). Exploring the benefits of cloud computing for sustainable construction in Nigeria. *Journal of Engineering, Design and Technology*, n. pag.

OMEGA. (2019). *A Complete Guide to Data Acquisition (DAQ) Systems*. Retrieved February 27, 2022, from OMEGA Engineering: https://www.omega.com/en-us/resources/daq-systems

O'Toole, E., Laura , F., Kenya , H., & Rohit , N. (2018, August). Data security procedures for researchers. Retrieved from https://www.povertyactionlab.org/resource/data-security-procedures-researchers

Pagallo, U., Durante, M., & Monteleone, S. (2017). What Is New with the Internet of Things in Privacy and Data Protection? Four Legal Challenges on Sharing and Control in IoT. *[Law, Governance and Technology Series] Data Protection and Privacy: (In)visibilities and Infrastructures, Volume 36*, 59 – 78. doi:10.1007/978-3-319-50796-5_3

Pallas, F., Raschke, P., & Bermbach, D. (2020). Fog Computing as Privacy Enabler. *IEEE Internet Computing*, 1 – 1. doi:10.1109/MIC.2020.2979161

Pcmag. (nd.). *PKI*. Retrieved from PCmag: https://www.pcmag.com/encyclopedia/term/pki

Pfandzelter, T., & Bermbach, D. (2019). IoT Data Processing in the Fog: Functions, Streams, or Batch Processing? *IEEE 2019 IEEE International Conference on Fog Computing (ICFC)*, (pp. pp. 201 - 206). [Prague - Czech Republic (2019.6.24-2019.6.26)]. doi:10.1109/ICFC.2019.00033

Pires, D. A., Colussi, C., & Calvo, M. (2014). Assessment of municipal management of oral health in primary care: data collection instrument accuracy. *Ciencia & saude coletiva, Volume 19 11*, 4525-34.

planningtank.com. (2020, August 11). *Data Processing Cycle | Definition, Stages, Use & Examples*. Retrieved March 3, 2022, from Planning Tank: https://planningtank.com/computer-applications/data-processing-cycle

Quek, T. (2017, February 7). *The advantages and disadvantages of Internet Of Things (IoT)*. Retrieved March 15, 2022, from LinkedIn: https://www.linkedin.com/pulse/advantages-disadvantages-internet-things-iot-tommy-quek

RapidScale. (2015, January 30). *Cloud Computing Stats - Security and Recovery*. Retrieved March 5, 2022, from slideshare.net: https://www.slideshare.net/rapidscale/cloud-computing-stats-security-and-recovery

Razouk, W., Daniele , S., & Kouichi , S. (2017). A new security middleware architecture based on fog computing and cloud to support IoT constrained devices. , (),. *Proceedings of the 1st International Conference on Internet of Things and Machine Learning - IML '17* - (pp. 1–8). Liverpool: [ACM Press the 1st International Conference - Liverpool, United Kingdom (2017.10.17-2017.10.18)]. doi:10.1145/3109761.3158413

Roberto, C.-V., Fernando , d., Javier , P., & Juan , M. (2018). Proceedings of the 1st Workshop on Blockchain-enabled Networked Sensor Systems - BlockSys'18 - Blockchain framework for IoT data quality via edge computing. , (),. (pp. 19–24.). [ACM Press the 1st Workshop - Shenzhen, China (2018.11.04-2018.11.04)]. doi:10.1145/3282278.3282282

Sadri, A. A., Rahmani, A., Sadri, A., Saberikamarposhti, M., & Hosseinzadeh, M. (2021). Fog data management: A vision, challenges, and future directions. *Journal of Network and Computer Applications*, pp. 34 - 35.

Sahai, A., & Hakan, A. (2010). Worry-Free Encryption: Functional Encryption with Public Keys. *Proceedings of the 17th ACM conference on Computer and communications security - CCS '10 - Worry-free encryption.* (pp. 463 - 472). [ACM Press the 17th ACM conference - Chicago, Illinois, USA (2010.10.04-2010.10.08)] . doi:10.1145/1866307.1866359

Sandler, V. (n.d). *What is Cloud Misconfiguration, and How Can You Avoid it?* Retrieved from Lightspin: https://blog.lightspin.io/cloud-misconfiguration

Schaie, K. W., & Willis, S. (2016). Handbook of the Psychology of Aging (Eighth Edition) ISBN 9780124114692. Academic Press. doi:10.1016/B978-0-12-411469-2.00036-4.

Seal, A., & Mukherjee, A. (2018). On the Emerging Coexistence of Edge, Fog and Cloud computing paradigms in Real-Time Internets-of-EveryThings which operate in the Big-Squared Data space. 1- 9.

See, A. v. (2021, June 7). *Volume of data/information created, captured, copied, and consumed worldwide from 2010 to 2025*. Retrieved February 17, 2022, from statista: https://www.statista.com/statistics/871513/worldwide-data-created/

Shen, G., Su, Y., & Zhang, M. (2018). Secure and Membership-Based Data Sharing Scheme in V2G Network. *IEEE Access*, pp. 58450 - 58460. doi:10.1109/ACCESS.2018.2874622

Shi, F., Wang , J., Shi , J., Wu , Z.-x., Wang , Q., Tang , Z., . . . Shen, D. (2021). Review of Artificial Intelligence Techniques in Imaging Data Acquisition, Segmentation, and Diagnosis for COVID-19},. *IEEE Reviews in Biomedical Engineering, 14*, pp. 4-15. doi:10.1109/RBME.2020.2987975

Shojafar, M., & Mehdi , S. (2019). Internet of everything, networks, applications, and computing systems (IoENACS). *International Journal of Computers and Applications*(1206-212X (Print) 1925-7074 (Online)), pp. 1 - 3. doi:10.1080/1206212X.2019.1575621

Simplilearn. (2021, October 19). *What is Data Quality - Definition, Dimensions, Characteristics, and How to Improve It*. Retrieved May 03, 2022, from simplilearn.com: https://www.simplilearn.com/data-quality-article

Smith, G. M. (2020, March 1). *What is Data Acquisition (DAQ or DAS)? The Ultimate Guide*. Retrieved February 27, 2022, from dewesoft.com: https://dewesoft.com/daq/what-is-data-acquisition#introduction

Stickney, J. (2021, July 18). *Hashing & Integrity — The "I" in the CIA Triad*. Retrieved March 15, 2022, from jacob-e-stickney.medium.com: https://jacob-e-stickney.medium.com/hashing-integrity-the-i-in-the-cia-triad-98b722b6fe39?source=user_profile---------3----------

Stojmenovic, I., We, S., Huang, X., & Luan, H. (2015). An overview of Fog computing and its security issues. *Wiley Online Library*, 2991–3005. doi: 10.1002/cpe.3485

Sudeep, T. (2020). *[Studies in Big Data] Fog Data Analytics for IoT Applications (Next Generation Process Model with State of the Art Technologies) ||10.1007/978-981-15-6044-6()*, - (Vol. 76). (J. Kacprzyk, Ed.) Ahmedabad, Gujarat: Springer Nature Singapore Pte Ltd. doi:10.1007/978-981-15-6044-6

Swenson, J. (2018, May 6). *Key Updates to the NIST Cyber Security Framework*. Retrieved March 27, 2022, from Abstract Forward Consulting: Cyber Security, Process Improvement, & Management Consulting: https://jeremy-swenson.com/2018/05/06/key-updates-to-the-nist-cyber-security-framework/

Swiss Cyber Institute. (2021, September 23). *What is an information security framework?* Retrieved March 27, 2022, from Swiss Cyber Institute: https://swisscyberinstitute.com/blog/nist-cybersecurity-framework-components/

talend. (nd.). *What is Data Cleansing? Guide to Data Cleansing Tools, Services and Strategy*. Retrieved May 03, 2022, from talend.com: https://www.talend.com/resources/what-is-data-cleansing/

Tavakol, M., & Dennick, R. (2011). Making sense of Cronbach's alpha. International journal of medical education. *International Journal of Medical Information, Volume 2*, pp. 53–55. doi:10.5116%2Fijme.4dfb.8dfd

Techtarget. (2021, April). *Identity and access management topic 5: PKI (public key infrastructure)*. Retrieved March 13, 2022, from TechTarget: https://www.techtarget.com/searchsecurity/definition/PKI

Techtarget. (nd.). *Top cloud security standards and frameworks to consider*. Retrieved March 26, 2022, from TechTarget: https://www.techtarget.com/searchsecurity/tip/Top-cloud-security-standards-and-frameworks-to-consider

Thaploo, V. (2020, March 5). *Difference Between Cloud Security and Traditional Security: What You Need to Know*. Retrieved March 12, 2022, from cloudlytics: https://cloudlytics.com/difference-between-cloud-security-and-traditional-security-what-you-need-to-know/

The National Institute of Standards and Technology. (nd.). *Cyber Security Framework; Getting Started*. Retrieved March 26, 2022, from NIST: https://www.nist.gov/cyberframework/getting-started

Thomas, B. (2018). Les 5V du big data. *Regards croisés sur l'économie*, 27-31. doi:10.3917/rce.023.0027

Tonjes, R., Barnaghi, P., Ali, M., Mileo, A., Hauswirth, M., Ganz, F., . . . Nechifor, S. (2014). Real-time iot stream processing and large-scale data analytics for smart city applications in poster session. *European Conference on Networks and Communications.*

tutorialspoint. (nd.). *Public Key Infrastructure*. Retrieved from tutorialspoint: https://www.tutorialspoint.com/cryptography/public_key_infrastructure.htm

US Department of Labor. (nd.). *Guidance on the Protection of Personal Identifiable Information*. Retrieved April 28, 2022, from US Department of Labor: https://www.dol.gov/general/ppii

Valkonen, A. (2013). Cloud computing ecosystem: Insights from an exploratory study in SaaS and PaaS value networks.

Vodyaho, A. I., Zhukova , N., Kulikov , I., & Abbas, S. (2021). Using the Context-Sensitive Policy Mechanism for Building Data Acquisition Systems in Large Scale Distributed Cyber-Physical Systems Built on Fog Computing Platforms},. *Comput., 10*, pp. 101. doi:10.3390/computers10080101

Vuleta, B. (2021, January 28). *How Much Data Is Created Every Day? [27 Staggering Stats]*. Retrieved December 7, 2021, from SeedScientific: https://seedscientific.com/how-much-data-is-created-every-day/

Wang, R. Y., & Strong, D. (1996). Beyond Accuracy: What Data Quality Means to Data Consumers. *Journal of Management Information Systems 12(4)*, pp. 5 – 33. doi:10.1080/07421222.1996.11518099

WebsiteSetup. (2021). *Internet Stats & Facts (2021) List of Internet, eCommerce, Hosting, Mobile & Social Media Statistics for 2021*. Retrieved from WebsiteSetup: https://websitesetup.org/news/internet-facts-stats/

wikipedia. (n.d). *Fog computing*. Retrieved January 29, 2022, from wikipedia.org: https://en.wikipedia.org/wiki/Fog_computing

Wikipedia. (nd.). *Encryption*. Retrieved from Wikipedia: https://en.wikipedia.org/wiki/Encryption

Wu, H., Yan, Y., Chen, B., Hou, F., & Sun, D. (2020). FADA: A Cloud-fog-edge Architecture and Ontology for Data Acquisition. *Journal of Latex Class Files*, 2168 - 7161. doi:10.1109/TCC.2020.3014110

Xue, C. T., & Xin, F. (2016). Benefits and Challenges of the Adoption of Cloud Computing In Business. *International Journal on Cloud Computing: Services and Architecture (IJCCSA), volume 6*. doi:10.5121/IJCCSA.2016.6601

Yi, S., Zhengrui , Q., & Qun, L. (2015). *"Security and Privacy Issues of Fog Computing: A Survey."*. Retrieved from WASA: https://www.cs.wm.edu/~liqun/paper/wasa15-fog.pdf

Yi, S., Li, C., & Li, Q. (2015). A Survey of Fog Computing: Concepts, Applications and Issues. *ACM 978-1-4503-3524-9/15/06*. doi:10.1145/2757384.2757397.

Yin, C., Zhang, S., Xi, J., & Wang, J. (2017). An improved anonymity model for big data security based on clustering algorithm. *Concurrency and Computation: Practice and Experience, volume 29*. doi:10.1002/cpe.3902

Zhang, D. (2018). Big Data Security and Privacy Protection. doi:10.2991/ICMCS-18.2018.56

Zhang, Y., Wang, P., Huang, H., Zhu, Y., Xiao, D., & Xiang, Y. (2020). Privacy-Assured FogCS: Chaotic Compressive Sensing for Secure Industrial Big Image Data Processing in Fog Computing. *IEEE Transactions on Industrial Informatics*, pp. 1551-3203. doi:10.1109/TII.2020.3008914

Zhu, X., & Badr, Y. (2018). Fog Computing Security Architecture for the Internet of Things using Blockchain-based Social Networks. *IEEE Confs on Internet of Things, Green Computing and Communications, Cyber, Physical and Social Computing, Smart Data, Blockchain, Computer and Information Technology, Congress on Cybermatics* (pp. pp. 1361 - 1366). Lyon: IEEE. doi:10.1109/Cybermatics_2018.2018.00234

Appendix

QUESTIONNAIRE FOR AIR RESPONDENTS

Hello, my name is Adepoju Alamu Luke, an Ms. Computer Science student of Khazar University. Please, I am conducting a study aimed at building users' trust and confidence on the Internet. And I would like to solicit your responses on the privacy issues experienced on the internet. Please the survey is expected to take 10-15 minutes to complete. Your responses will remain anonymous, please be as honest as possible. Thank you for your participation.

1. Which continent are you located?

Africa	[ ]	Asia	[ ]	Europe	[ ]
North America	[ ]	South America	[ ]	Australia	[ ]

2. Which age group do you belong to?

18 – 24 [ ]	25 – 34 [ ]	35 – 44 [ ]	45 – 54 [ ]
55 – 64 [ ]	65+ [ ]		

3. Are you familiar with the internet?
   Yes [ ] No [ ]
4. If Yes to the question above, how long have you been using the internet?
   Less than a year [ ] 1 - 2 yrs. [ ]
   3 - 4 years [ ] 5+ yrs. [ ]
5. How often do you connect to the internet
   Hourly [ ] Daily [ ]
   Weekly [ ] Yearly [ ]
6. When connected to the internet, how long do you spend staying connected?
   Specific portions of the day [ ]
   All day long [ ]
   All week long [ ]
7. Which of the following purposes do you use the internet for? Click all those that apply
   Entertainment [ ]
   Education [ ]
   Work-related research [ ]
   Personal finance (Banking and business-related financial management) [ ]
   Accessing current affairs (News, sports, weather) [ ]
   Travel reservations [ ]
   Product information gathering [ ]
   Online shopping [ ]
   Communication and staying in touch (social media, emails) [ ]
8. Are you familiar with fog computing? Yes [ ] No [ ]

Perceptions About Fog Data

9. Using the Likert scale below, kindly indicate your preference with the following statements with a scale of **1 = disagree 2 = Disagree 3 = neutral 4 = Agree; 5 = Strongly Agree**

	1	2	3	4	5
I find the Internet too confusing					
It is difficult to manage data on the Internet					
I am confident no one monitors what information I transmit on the internet					
The Internet is expensive					
It is easy to find porn on the internet					
There is too much information on the internet					
I always receive unwanted messages from unknown internet users					
My personal information can be easily stolen on the internet					

Providing Personal Information over the Internet

A Personal information is any information that can be associated with an identifiable living individual (name, signature, address, phone number or date of birth, photographs, voice print and facial recognition biometrics, details of fianances).

10. Do you often have to provide personal information to websites?
    Yes [ ] No [ ]
11. If Yes, how much of your real self-do you feel comfortable sharing or leave on a web site?
    25% - 50% [ ] 51% - 74% [ ] 75% - 100% [ ]
12. If No, how long does it take to provide a false identity to a website?
    Less than an hour [ ] 1 Hour [ ] Week [ ]
13. I always refuse to disclose personal information under the following conditions

Kindly use the 5-Point Likert Scale of (Very Important = 1, Somewhat Important = 2 Neutral = 3, Somewhat Unimportant = 4, Unimportant = 5).

	1	**2**	**3**	**4**	**5**
The company or individual is unknown to me.					
I am not familiar with the individuals or the company					
I do not trust the company or individuals					
If they do not disclose what they use my information for					
The services rendered to me does not worth the information I give					
I prefer to remain anonymous online					
Some information asked are sensitive					
I am worried my information might be stolen through eavesdropping					
Completing the requirements is time consuming and exhausting					

14. Does the reputation of the company requesting for your personal information over the internet matter?
    Yes [ ] No [ ]
15. Are you willing to accurately provide your personal information to websites to provide advertisements that suit your tastes and interests?
    Yes [ ] No [ ]
16. Would you be comfortable to provide your personal information to websites for a fee?
    Yes [ ] No [ ]
17. Have you ever made an online purchase? Yes [ ] No [ ]
18. If yes, how often do you make purchases over the internet?
    All the time [ ] occasionally [ ] rarely [ ]
19. If No, how likely are you to purchase items online in the next six months (1 = Very likely, 2 = Likely, 3 = Not likely)
    Very Likely [ ] Likely [ ] Not Likely [ ]

Perceptions about Internet Privacy

Throughout this survey, references will be made to privacy of personal information. Privacy refers to the amount of data gathered from you. Whiles personal information is any information that can be personally attributed to you (name, address email address, names of family members, social security number, credit card number, medical history)

Very Important = 1, Somewhat Important = 2, Neutral = 3, Somewhat Unimportant = 4, Unimportant = 5

What is the significance of your consent when:	**1**	**2**	**3**	**4**	**5**
Sites sell/share your personal information with others					
Sites track your movement around their site					
Sites track your movement around the Internet					
Sites track your online purchases					
Sites gather in-depth personal profiles about you from other outside databases					
Sites customize your online experience to your personal preferences					
Sites sell/share your personal information with others					

Does recording your online activities with your knowledge constitute an invasion of privacy?
Yes [ ] No [ ]
Does recording your online activities by websites without constitute an invasion of privacy?
Yes [ ] No [ ]

Regarding the privacy of your personal information, how concerned are you about sharing your personal information on the other communication mediums.

Kindly use the 5 point Likert Scale where less concerned = 1, somewhat less concerned = 2, Neutral = 3, more concerned = 4, much more concerned = 5

	1	2	3	4	5
Telephone					
Mail (By Post)					
Fax					
In-person					

Please rate the following reasons for your decision to transmit information over the internet rather than a traditional medium.

Kindly use the 5 point Likert Scale where Strongly Agree = 1, Agree = 2, Neutral = 3, Disagree = 4, strongly Disagree = 5

	1	2	3	4	5
My information could be stolen, it's not safe					
It is not clear to me how my information will be used					
It is unclear who I am dealing with					
I don't trust the website with my personal information					
Mainly because I'm unfamiliar with this modern technology					
My privacy has been abused on the Internet					

Someone I know had their privacy violated online					

Data management by fog servers and websites for the following listed services

	Concerned	**Neutral**	**Less Concerned**
Education			
Entertainment			
Work-related Research			
Personal Finance			
Accessing Current Affairs(News, Weather, Sports)			
Travel			
Product Information Gathering			