Prevention of Personally Identifiable Information Leakage in e-commerce using Offline Data Minimization and Online Pseudonymisation

Mukuka Kangwa

This Research Report is submitted in fulfilment of the academic requirements for the degree of

Doctor of Philosophy in ICT Security Engineering

In the School of Engineering

University of Zambia

2022

DECLARATION

I declare that this research report is my own work. Where collaboration with other people has taken place or material generated by other researchers is included, the parties and/ or materials are stated with references as appropriate.

This work is being submitted for the Doctor of Philosophy in Information and Communication Technology Security Engineering at the University of Zambia. It has not been submitted to any other university for any other degree or examination.

Mukuka Kangwa	_____	**03rd May 2023**
Name	**Signature**	**Date**

Dedication

To my wife Nchimunya and my daughters Mykayla and Michal. I thank you for your support and understanding as I had to sometimes spend huge amounts of time away from you just to study and work on this project. Above all my immense gratitude to God Almighty for giving me life, grace, and the opportunity to work on this undertaking.

Supervisor Approval

As the Candidate's Supervisor, I have approved this Thesis for Publication

Name: Dr. Charles .S. Lubobya

Sign:……..

Date: …….

As the Candidate's Supervisor, I have approved this Thesis for Publication

Name: Dr. Jackson Phiri

Sign:…….

Date: ……

Abstract

The rapid adoption of electronic channels for the delivery of services by various service providers compels the consumers of these services to adapt. For one to be granted access to e-services, one must surrender part of their Personally Identifiable Information (PII) hence making their personal data susceptible to leakage. Despite several solutions being already in use to protect PII, data leakage persists. To enhance PII protection and user privacy, the research proposes employing Offline Data minimization and Pseudonymisation using physical and logical partitions implemented through a combination of hardware and software. The implementation includes the use of unique random pseudo-ID algorithm derived from the modification of the Request for Comment (RFC) time-based One Time Password (TOTP) standard RFC 6238. The random pseudo-ID can be used to transact online while preventing online profiling that is possible when using static pseudo-IDs. The Random ID generator algorithm can be used to trace the user of a given Random pseudo–ID. Data is most vulnerable to leakage when it is accessible via the Internet. The solution developed addressed the problem of PII data leakage by making sensitive data 'offline' to the internet. The methodology employed the Trusted Third Party (TTP) approach. This meant having a third party collecting PII from e-commerce users and confirming the KYC of users who would like to be granted access to e-commerce platforms hence preventing the spreading of aggregated PII across the cyber space. At the TTP, this was achieved by data Minimization of sensitive personal information and pseudonymization of information to be made available for online transactions. To keep the PII 'Offline', a multi-layered hardware approach was used; two microcontrollers were configured to create a buffer that ensured one-way traffic towards the online-sub system that held minimized pseudonymized data. To further restrict the amount of data that could flow from the offline system to the online system, the bandwidth between the microcontrollers was restricted to 9,600bps. Experiment results showed that the 'offline' system hosting the PII could not be accessed. Further, Random IDs were successfully generated to ensure privacy is maintained for users.

Key words

Personally Identifiable Information, Data Privacy, One Time Password, Data Protection, Time-based One Time Password, Firmware, and TOR

Acknowledgements

My sincere appreciation goes to my supervisors Dr. Charles Lubobya and Dr. Jackson Phiri for their guidance, support, and encouragement during my study at the University of Zambia. Without you, this journey wouldn't have reached this far.

Thanks to my family for their prayers and encouragement.

I also appreciate the critique and advice from the members of my class, particularly, Mr. Lusungu Ndovi. Thank you very much for reviewing my work. Further appreciation goes to Mr. Daniel Mwale and Mr. Brian Kanduli for the critical roles they played and their unwavering support during the planning, setup, and conduct of the research experiments.

I would also like to express special thanks to all the Lecturers from the School of Engineering who contributed to my work during presentations in various forums. These include but are not limited to, Dr. Simon Tembo, Dr. Ackim Zulu, Dr. Evaristo Musonda, and Dr. Banda.

Table of Contents

LIST OF FIGURES

LIST OF TABLES

LIST OF EQUATIONS

LIST OF ACRONYMS

ICT	Information and Communication Technology
CPU	Central Processing Unit
KPA	KYC Privacy Agency
KYC	Know Your Customer
NUI	National Unique Identifier
ODMS	Offline Data Minimisation System
PS	Pseudonymisation System
RMS	Restricted Memory System
eID	Electronic ID
eIDr	Random Electronic ID
USB	Universal Serial Bus
GDPR	General Data Protection Regulation
DLP	Data Leakage Prevention
TTP	Trusted Third Party
IoT	Internet of Things
OTP	One Time Password
OTPaaS	One Time Password as a service
Covid-19	Coronavirus disease discovered in 2019
PII	Personally Identifiable Information

Chapter One

Concept Introduction

1.0 Introduction

The rapid advancements being made in Information Communication Technology in this Information Age have witnessed an unprecedented adoption of electronic channels in the delivery of services to consumers [1][2][3][4]. One such adoption is the use of Electronic Commerce or e-commerce in short. E-commerce is the use of the internet, the World Wide Web (web), web browsers, and mobile applications hosted by devices to conduct business. Stated differently, e-commerce is digitally enabling commercial transactions among organizations and individuals [5]. E-business, often confused with e-commerce, is the enabling of transactions and processes undertaken by a firm using information systems; the information systems are under the control of that firm. In e-business, the transactions do not involve commercial interaction with other organizations or individuals [5]. Figure 1.1 below gives a high-level overview of the difference between e-commerce and e-business discussed. Technology Infrastructure is common to both E-business and e-commerce.

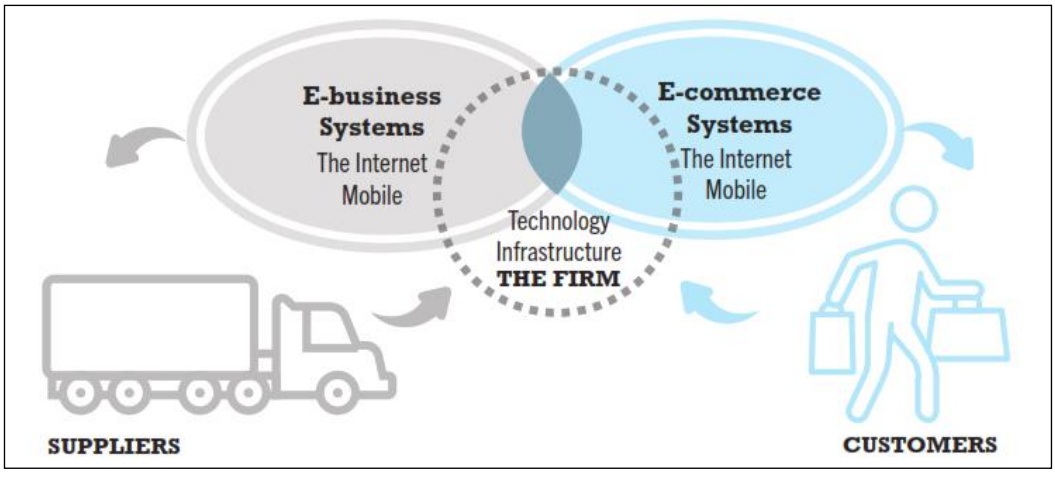


Figure 1. 1: The difference between e-commerce and e-business [5]

Simply stated e-commerce is the purchasing of goods and services digitally as shown in Figure 1.1. The transaction can be executed between individuals, organizations, and even countries. E-commerce users shop for the goods and services they would like to purchase using online platforms and pay for them digitally [6][7][8]. E-commerce

reduces costs associated with trade and provides access to a wider range of markets across the globe [9][10].

In addition to e-commerce, the Information Age has also brought about other technologically dependent activities such as social networking, financial transactions, as well as storage of medical records. Modern technological innovation is data-dependent [11]. The increased data gathering and sharing online has brought about privacy concerns. The more data is collected the higher the chances of one's privacy being violated [12][13]. Privacy is one being in a state where they can decide what is shared to who about themselves. They are in control of this and not someone else [14].

E-commerce has come with its risks such as the leakage of Personally Identifiable Information (PII) of its users. PII is user-identifying data such as names, identity numbers, and phone numbers. It is data that uniquely differentiates one from another. It is personal [15]. E-commerce platform providers usually request for PII and hold this information before granting a user access to their services [16][17][18][19]. Even chatbots require a user to submit some elements of personal data before the chat can proceed. Chatbots are now becoming common as a way companies can help reduce the cost of human resources. This is possible as some tasks are repetitive. Staff can be dedicated to tasks requiring more human reasoning that has not yet been automated [20][21].

The collection of personal data by several service providers has resulted in huge volumes of aggregated PII being vulnerable to leakage. As it stands at the moment, for any person to have access to various e-commerce platforms, they would need to provide similar PII to all the platforms they would like to access hence creating several isolated but aggregated PII for individuals [22] [23]. Several attempts are being made by hackers to leak or illegally access data stored in the cloud [24] and private storage locations. Some of the data, once accessed, might be used by the service provider as they deem fit without even consulting the owner of the data which results in the violation of the privacy of the data owner [25]. Data is leaked in several ways and due to several reasons; man-in-the-middle attack can be used to intercept sensitive data if it is not encrypted. A man-in-the-middle attack involves an attacker eavesdropping on a communication channel between two points exchanging information. If the channel being used, or the data being transmitted is not encrypted, the attacker can have access

to the data being transmitted hence leading to information leakage. Further, malware can also be used to transmit data illegitimately. Malware can be installed on user devices via illegal software, phishing emails and many other forms. Once on the user's device, it can start transmitting personal information without the knowledge of the user leading to data leakage.

Another method that can be used to steal data is SQL injection. This type of attack involves the sending of SQL instructions using data-capturing fields on a given system. The field is used to send SQL commands to the system. In addition to SQL injections, data can also be physically leaked via, for example, flash disks, compact disks, Tape, and other storage media. A good example of data being leaked is one incident involving Facebook and Cambridge Analytica where data was found to have been in the possession of a Third party. The information owners had little control over what could happen to their data the moment they rendered it to the Facebook platform [26]. This issue, together with the GDPR that came into effect in May 2018, raised international awareness of the importance of data and user privacy. [27]. More people realized that the data they were submitting to online platforms was susceptible to leakage or even abuse by data handlers into whose hands they had entrusted their PII. GDPR reaffirms the fundamental human right to privacy. Users' privacy must be respected (including the data that describes them) as a human right [28][29]. The Universal Declaration of Human rights also affirms privacy and data protection as human rights [30]. It is, therefore, serious business to ensure the privacy of PII owners is guaranteed. They should be given an opportunity to submit their PII while knowing that their privacy is guaranteed.

The GDPR and various other regulations have seen an increase in companies needing to secure the data they are storing and processing [31]. Violations of these regulations come with substantial penalties to encourage compliance. Further, users' concern with the privacy of their PII and the possibility of it being leaked can determine whether they adopt e-commerce or not [32][33][8]. Users who treasure their privacy highly are very unlikely to sign up for online platforms that involve the submission of their PII before they can be granted access. By 2019 only about 29% of e-commerce users were converting their online search into a purchase due to their concerns over the user-friendliness as well as security of the e-commerce platforms they were using. There is

a very close correlation between usability and security [7]. A user can explore an e-commerce platform and even discover some good or service they would like to purchase but abandon the mission the moment they reach a point where they must submit their PII to proceed.

In 2020 Europe experienced a 10% growth in e-commerce attributed mainly to the Covid crisis and the consequential government restrictions which forced many to start conducting transactions using online platforms [34][35][4]. To try and increase the conversion rates, some platforms try to collect more data on the behaviour of the potential customer so that they can personalize the user experience. It is also hoped that personalization would help increase repeat buyers [36][37]. One of the reasons GDPR was introduced was to win back user trust in electronic channels and enhance the digital economy [38].

Several solutions have been proposed to address the challenge of data leakage and privacy violation yet the problem persists [39]. Several incidents have occurred, and keep happening, where huge volumes of data have been leaked and user privacy breached [40]. Statistics from security companies indicate that the number of data breaches has been raising steeply in the past few years [41]. Data that is exposed to the internet, whether on the edge equipment like phones and tablets or in the cloud is at risk hence the need to provide more effective data protection methods [42].

To further define privacy; privacy by definition is any information associated with a specific subject or person [43]. This can be the date of birth, National Identity, physical address, contact details like email and phone numbers, and so on. This is information that identifies someone uniquely. It is like the PII defined earlier. PII and privacy are inseparable as they both refer to a subject in a unique manner. Therefore, the protection of data and the privacy of an individual ought to be addressed together. Loss of PII is likely to lead to the violation of privacy for the owners of the leaked PII.

Some service providers promise to delete data once the owners request that they do so. However, the deletion of data from online storage is very difficult. Just a service provider advising that the data has been deleted isn't enough. Their storage hardware might be accessed and the data they contain get recovered [44]. Ensuring data is secured requires deleting data for recoverability. That is, data should not be recoverable after

deletion. Unfortunately, not everyone knows that the so-called normal data deletion does not necessarily delete the data but simply tells the system that a given space on the storage is now available for use. The data on that space can still be recovered with special tools or software [45]. Therefore, the best way to protect data is to make that data unavailable online.

Several similar works have been produced to address the PII leakage challenge and user privacy violations. These include but are not limited to, the use of Static Pseudo-IDs while a user is online, the use of Blockchain to store user data, encryption of data while being stored in the cloud as well as the obfuscation of data while keeping it online. The drawback with some of the solutions such as the use of static Pseudo-IDs is that a user can be profiled and hence have their privacy compromised. The use of blockchain technology at this moment is not cost-effective as the technology is currently expensive and is not very scalable. The storage of PII in the cloud leaves the PII data vulnerable to leakage. Furthermore, even if data is encrypted, if it is stored online, the chances of the data being in plain text while being processed is high hence making it susceptible to leakage.

This research will focus on the design of an Offline Data Minimization System (ODMS) to work in conjunction with an Online Pseudonymisation system to offer an adequate protection of PII for e-commerce users.

In the next subsections, we will discuss data Leakage and how it affects the privacy of users in e-commerce. The subsections that will follow will outline the objectives, key research questions, motivation, problem statement, background, and application of the study. The methodology employed, the proposed approach experimented and the associated results as well as the discussion of the results will be presented in different chapters of this report. The conclusion will be given in the last chapter.

1.1 Related Work

Internet technologies are developing at a very high pace. E-commerce has been one of the great beneficiaries of this development. However, the advancement in technology has come with security, trust, and privacy challenges [46]. Many people would like to adopt e-commerce due to the benefits that accrue to its users, but the safety of the users' PII remains a huge problem as it poses a huge challenge to their privacy. The protection

of privacy for e-commerce users has been a concern since the inception of e-commerce. Almost every e-commerce transaction involves the exchange of personal data [47]. Several cases have been reported over time concerning the leaking of PII thereby compromising the privacy of the victims involved. The General Data Protection Regulation (GDPR) seeks to protect people's data from abuse by its handlers by prescribing how data must be managed and what permissions must be obtained before processing the data [48]. GDPR proposes such methods as pseudonymisation and obfuscation of PII. Pseudonymisation refers to the processing of PII in a manner that the data can no longer be traced back to the actual owner without additional information being sought to connect the dots [49]. The primary aim of pseudonymisation is to protect PII for users. The GDPR has been put in place because of the prevalent cases of data being abused by its handlers as well as users not being protected adequately by law. The regulation transfers ownership of data from handlers back to the owner [50]. The Indonesian parliament was debating their data privacy bill to achieve something similar, granting control to data owners. Although there was a plan to allow the government to have unrestrained access to personal data when the need arise as part of security [19].

E-commerce platforms and channels through which data is transmitted have some existing solutions to provide privacy to their users. The solutions in place include authentication mechanisms, server and endpoint antivirus, encryption of data while at rest and in transit as well as encryption of the channels used for data transmission. These solutions have aimed at providing confidentiality and integrity of data transmitted [51][52]. Despite the various interventions in place, PII is still being leaked hence the need for more effective solutions.

Another common solution to e-commerce security challenges that have been in use for a long time now is Firewall. Firewalls are used to separate internal networks from external networks of an organisation. In addition, traffic gets restricted to only allow traffic perceived to be harmless to the systems and the data they hold. Furthermore, in addition to firewalls, anonymous connections have also been deployed to help preserve privacy. In this scenario, mechanisms are put in place to prevent the source of the connection from being known [53].

There has been a push for the use of external parties or trusted third parties (TTP) to keep sensitive information on behalf of the user. This is premised on the assumption that the user trusts the TPP. However, trust has been broken before when data is leaked by the TTP [54]. This can happen if the TTP does not have adequate security to protect the PII it is holding. Hence, there is still a need to enhance the existing security measure implemented to secure PII.

Location privacy is yet another key technology that has been deployed to help protect one's privacy. Various technologies are being used to track the location of people. These include personal devices like smartphones, tablets, and laptops [55]. The main objective is to profile the users and understand their preferences to facilitate targeted marketing as well as offer personalized services. The approach compromises a user's privacy because if that data fell into the wrong hands, the users can be put in danger. Technologies have been developed to distort the location data that is collected on the user. This could involve widening their location grid so that the user is not put in immediate danger. The main objective is to hide the exact location of the user to maintain their privacy [56]. Nevertheless, the approach does not address the problem of PII data leakage held by various service providers. The user can hide their location but if their PII is leaked or stolen, then their privacy gets promised.

It must be noted early in this research that there is normally a trade-off between access to certain electronic services and one's privacy. For example, one might be looking for a restaurant nearest to them. This would require the service they are using to track their precise location otherwise there would be no way one would tell which is the nearest restaurant they should go to [56]. A user would have to weigh between maintaining their privacy and accessing certain personalized services. However, it is generally expected that a balance between user privacy and innovation must be achieved. Privacy regulations must not stifle innovation while at the same time, innovation must not violate user privacy [57]. The companies that are providing various technological innovations/services that require user data should be helped to be in a place where they do not keep wondering whether they are violating data protection laws or not which can hamper innovation [19]. This research provides a solution to those users and organisations who consider their privacy as a priority over any other requirements.

However, certain portions of the solution can be forgone to maintain some personalized services such as those dependent on knowing user location and so on.

This research will focus on protecting PII using Offline Data minimisation and pseudonymisation to ensure the privacy of users is guaranteed while accessing e-commerce services. The solution will also help improve consumer trust in e-services platforms and thereby encourage the use of the platforms.

1.2 Problem Statement

There has been an increase in the use of e-commerce. Very often, for one to have access to some e-commerce services, one has to provide some of their PII and this has led to e-commerce users' PII being exposed to leakage [58]. The current form of storage, analysis, and processing of data contributes greatly to the possibility of the data being leaked [59]. Most e-commerce platforms store their data online in such a way that it can be easily accessed if their system security was breached. Several cases have been and are being reported where user data has been leaked online either accidentally or deliberately [60]. The number of cases of data breaches keeps rising and an increased number each year is being reported [61]. The leaking of PII presents various risks to the users whose data is leaked; their privacy is compromised [62]. Their data can be used to defraud them financially or even cause physical harm to them by tracking their location. Furthermore, with PII in the wrong hands, the victims could have their Bank accounts emptied without their consent or knowledge.

In addition, sensitive attributes of the leaked data can be used to identify the actual owner of the data and map it with other published data to get more knowledge of the victim [63]. That is, the attacker can use leaked data to make inferences and know more information about the user that can then help them defraud their victim. These attackers can also sell the information they gather and put together for marketing purposes [64]. Users will feel their privacy is being compromised by receiving unsolicited marketing communication they never subscribed to.

Further, laws and regulations have been put in place to control how data is processed to protect the privacy of the data owners. Any violation has its consequences including tough penalties that might lead some companies into bankruptcy [62]. To avoid penalties, some technology companies might be very cautious in their innovations that

involve the processing of user data hence discouraging disruptive innovations. There is an urgent need to provide an effective solution and strike a balance between innovation and regulation.

1.3 Aim of the study

The ultimate objective of this research is to address the prevalent problem of PII leakages in e-commerce via the use of Offline Data Minimization and Online Pseudonymisation. There is a need to enhance security solutions to address the problem [65].

1.4 Research Objectives

The following are the specific objectives the study will address:

(i) Design a comprehensive and implementable model that will use Offline data minimization and online Pseudonymisation to protect PII for e-commerce users.

(ii) Develop and implement a working prototype solution of the model formulated in objective one.

(iii) Investigate the safety of the PII protected by the prototype developed in objective 2 by subjecting the system to various system assessments.

1.5 Research Questions

(i) What model can be formulated to enhance the prevention of PII data leakage for e-commerce users?

(ii) How can the proposed model be implemented and validated for its effectiveness in preventing PII leakage?

(iii) How effective is the formulated and implemented model in protecting PII and the privacy of its users?

1.6 Significance of the study

The outcome of the research will be a solution to PII leakages that can be applied not only to e-commerce but also to other electronic services such as online currency trading that require the collection of PII from their prospective clients. The primary objective of this solution is to protect the privacy of electronic service consumers by keeping their

PII away from online databases whose security can be compromised. Most organizations are collecting PII without even the knowledge of some of the users. In addition, the data is then used for commercial purposes without the consent of the owners. This puts the PII owners at risk [64].

The solution can also be applied in financial institutions like banks which normally must conduct some form of Know–Your–Customer (KYC) checks on a client before they can be granted a given service. To open an account with them, one must submit documentation such as proof of identity, residential address, workplace, contact details, and so on. Most Banks keep this information on systems that can be reached via the internet either with authorization or through back door methods by hackers. The institution holding the PII can do its best to protect the data but if the data is online, there is a probability of the data being leaked or stolen. Moreover, the chances of several unauthorized people having access are higher for online data, as the Internet has reduced the world into a global village, compared to keeping the data offline, as one would need to have physical access. It is easier to protect data from fewer points of access. Furthermore, unscrupulous individuals would be more interested in a file or database that has PII for several users than just data for one individual. This is where online aggregated data becomes of interest to hackers [47].

Using a third party to conduct KYC on the prospective users of the e-services, this solution can be used by any institution offering e-services and require KYC confirmation to be done before granting a client access. The institutions will confirm with the Third-party if they know the prospective client.

1.7 Motivation of the Research

E-commerce is here to stay and has many benefits that accrue to both the users and the service providers. The benefits include reduced cost of doing business and improved turnaround time for the delivery of electronic services like the purchase of movies and so on. However, if PII for users keeps being leaked hence putting them at risk, users are likely to get discouraged from using the services. Users want to be assured that their PII will be protected as they use e-commerce services [66]. Research has shown that the adoption of e-commerce services by users is determined by the level of consumer trust mostly built by the security and privacy provided to the users [67]. Hence,

providing solutions that can help protect users' data is key to convincing more people to adopt e-services.

This research will help address the PII leakage problem adequately and therefore provide enough comfort for users to access e-commerce services and take advantage of the benefits that come with it while minimizing the risks. PII in the wrong hands can lead to financial and physical loss to the victims depending on how determined to cause harm the perpetrators are. The best is to put the PII offline yet still offer the required e-commerce services desired by the users. The research intends to achieve this objective.

1.8 Scope of the Project

This research is limited to the prevention of PII leakage via the use of Offline data minimization and online Pseudonymisation. There are several proposed solutions to prevent the leakage of PII. Some of the solutions have been implemented with data leakage still prevalent while others remain as proposals. This study aims to enhance the privacy of an e-commerce user by using Offline data minimization in conjunction with other existing solutions. To be specific, the approach being proposed will use random code generation as well as online pseudonymisation to protect the identity of the user from the online data while employing offline data minimisation to guarantee users the safety of their PII from hackers and accidental leakages.

1.9 Theoretical and Conceptual framework

There are two approaches to data obfuscation to maintain privacy: Pseudonymization and Anonymization. The research was mainly focused on the use of pseudonymisation as the basic principle on which the solution to protect e-commerce users' PII was built. Pseudonymisation makes it difficult to trace leaked data to the original owner if implemented properly [49]. The research sought to design a model that would effectively protect users' PII while allowing them to enjoy the services of e-commerce.

With the widespread PII leakages happening due to aggregated data, privacy preservation has become paramount. Anonymization can be used to preserve user privacy [68][69][70]. Pseudonymization is another technique that can be employed to achieve personal privacy preservation. The research chose to employ Pseudonymisation and not Anonymization to develop the solution. Pseudonymisation provides a

possibility of tracing the owner by information handlers in case of need while Anonymization demands that the original owner be untraceable using the data shared [49][71][72][73]. Anonymization involves the complete and irreversible removal of or masking of identifying information from individual records [74][75][24]. This process is usually used in cases where statistical data is to be shared for research purposes without compromising the privacy of the individuals involved [76][77]. Once the data is rendered anonymous, the data protection requirements no longer apply as the data would no longer be personal [73]. It is important to ensure that a user of e-commerce can be traced via relevant authorities when the need arise [78]. Some users can commit fraud while transacting incognito hence the need for them to be traceable. Traceability must be part of the design than be an afterthought when the need arise [79].

To address the issue of the third party being untrustworthy, the study proposed the use of a KYC Privacy Agency (KPA) appointed by the national authority in each country to use the ODMS. This approach uses what is called the Trusted Model. The users must have trust in the appointed third party[54]. Figure 1.2 below depicts the concept of using a KPA.

The user requests for a service from an e-commerce provider. The e-commerce service provider requests the user to provide a random One Time Password (OTP). Then the suer generates the OTP using the platform provided by the TTP and submits it to the e-commerce service provider. The service provider contacts the TTP (the process is automated) to confirm the validity of the credentials supplied by the user. If the TTP's response is positive, the service provider grants the user access to their platform without requesting PII otherwise access is denied.

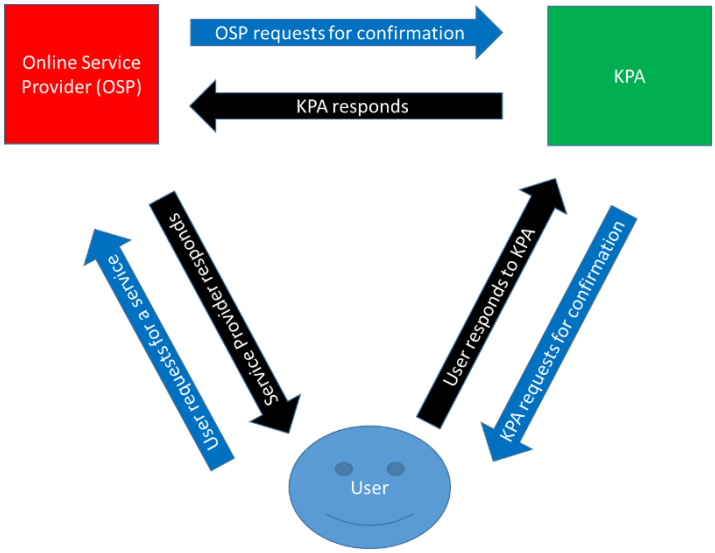


Figure 1. 2: Use of KPA as a Trusted Third Party

1.10 Operational definitions

Data Minimization: The process of reducing the amount of data that can be made available to a very minimum. That is, only providing data that is required for a particular purpose and nothing more.

Offline Data Minimization: Keeping minimized data away from online systems. That is making it inaccessible via the network be it local or via the internet. The system that does this is called Offline Data Minimization System (ODMS).

Pseudonymisation: Associating PII with a Pseudo name to prevent leaked data from being connected to the actual owner. In this process, data handlers can trace the Pseudo name back to the actual owner. Simply stated, pseudonymisation is the process of representing data in such a way that it can no longer be attributed to its original owner or subject without additional information being provided to connect the loose ends.

Anonymization: This is the process of removing Personally Identifiable Information from a data set so that the person who is associated with that data is no longer identifiable. This process is non-reversible even with additional information provided.

KYC Privacy Agency (KPA): A third party appointed or authorized to physically verify the Personally Identifiable Information of a user who would like to have their data protected while they access e-commerce services.

Restricted Memory System (RMS): This a type of memory system that can only accommodate a very minute amount of data to prevent leakage of data between offline and online systems but facilitate real-time updates of the online database for e-commerce users registering with KPA. Further, it allows the minute data to flow only in one direction. That is, from the offline system to the online system.

1.11 Ethical considerations

The research was purely laboratory-based. All the results obtained and presented in this research were obtained from experiments conducted during the research process. The process did not involve conducting surveys with individuals or organizations.

1.12 Plan of Development

The remainder of the research work is organised as outlined below:

Chapter Two- Provides detailed literature reviews on the solutions developed so far to prevent the leakage of PII and other sensitive data. In addition, this section also provides reviews on data minimisation and pseudonymisation

Chapter Three Discusses the overall methodology that was employed to design, develop, test, and implement the proposed system. Methods used to design and test the model before the actual build are described in detail.

Chapter Four Describes the detailed design of the proposed and implemented Offline data minimisation and online pseudonymisation system model. It also describes algorithms used to detach online-pseudonymized data from the offline data containing the actual PII for e-commerce users. It explains the results obtained from the validation of the model before the actual system could be developed end-to-end.

Chapter Five This chapter describes in detail how the proposed model designed in chapter four was built. It gives details of what was used and how various technologies were employed to interact and produce the Data Protection model.

Chapter Six This chapter gives detailed results from the validation of the Model that was built. The tests were done per module and then end-to-end to understand and appreciate user experience.

Chapter Seven Describes the results that were obtained through the tests conducted using the prototype system that was designed and built. It describes how the results demonstrate the effectiveness of the solution to protect PII as well as maintain user privacy.

Chapter Eight This chapter gives a conclusion of what the results entail as far as this research was concerned. It gives recommendations on how the contribution of this research can be taken advantage of to help address the predominant challenge of data leakage. It also talks about some related areas that were identified during the research conduct for consideration in future works.

1.13 Chapter Summary

The leakage of PII is a worldwide problem affecting even giants of online platforms. Rarely a day passes without a data breach being experienced or discovered by online service providers. This is despite the several different solutions that have been developed and implemented by various service providers. The leakage of PII by service providers discourages some users from adopting online services for fear of having their personal information leaked into the wrong hands.

There is a need to enhance data protection techniques to ensure that users can access online platforms without compromising their privacy. The study seeks to achieve that via the use of data minimization and pseudonymization implemented using a TTP model and Multi-layered hardware security.

Chapter Two

Literature Review

Chapter Introduction

This section explores various literature on prior works related to different data protection concepts and solutions. The material reviewed ranges from works on current data protection regulations and solutions to proposed concepts on the enhancement of personal data and privacy protection. Concepts like the use of TTP and Pseudo electronic IDs have been examined in detail. The review has also discussed works on hardware security, the use of Artificial intelligence in ICT security as well as the basic principles of ICT security. Another critical element reviewed is the storage of sensitive data; whether it should be online or offline.

2.0 Literature Review

The protection of privacy for e-commerce users has been a concern since the inception of e-commerce. Almost every e-commerce transaction involves the exchange of personal data [47]. Even communication during the conduct of business via the use of IoT devices involves the exchange of personal data. IoT devices are being used to create smart cities where data is being generated and exchanged over various communication channels [80]. There remains a serious concern regarding the privacy of users and the safety of their data on the internet [81] [80]. Several cases have been reported over time concerning the leaking of PII which has resulted in compromising the privacy of the victims involved. Studies have been conducted on how best to protect one's PII while transacting using e-commerce platforms and various approaches have been proposed to achieve the desired user privacy protection. Inadequate security in e-commerce can be a huge deterrent to its adoption[8]. The sections that follow examine the literature on various approaches used to protect information.

2.1 Trusted Third Party Models

Several scholars have undertaken research in third-party models. Frank and Michael patented a solution to help protect personal data. They proposed having a Trusted third Party that provides static Identities (ID) to users. In addition, Block chain technology was to be used to protect the data. The diagram shown in Fig 2.1 below shows a

summary of how the solution was designed to work; the user obtains an ID from the digital ID provider and submits it to the service provider as proof of identification. Then the service provider verifies with the ID provider if the user can be trusted and the response the ID provider returns determines whether a service will be offered to the user. Furthermore, an offline escrow was to be used for keeping the PII to be accessed via legally approved means. Pseudonymization (and not anonymization) is to be used to make it possible to trace a user when there is a need [22] [74].

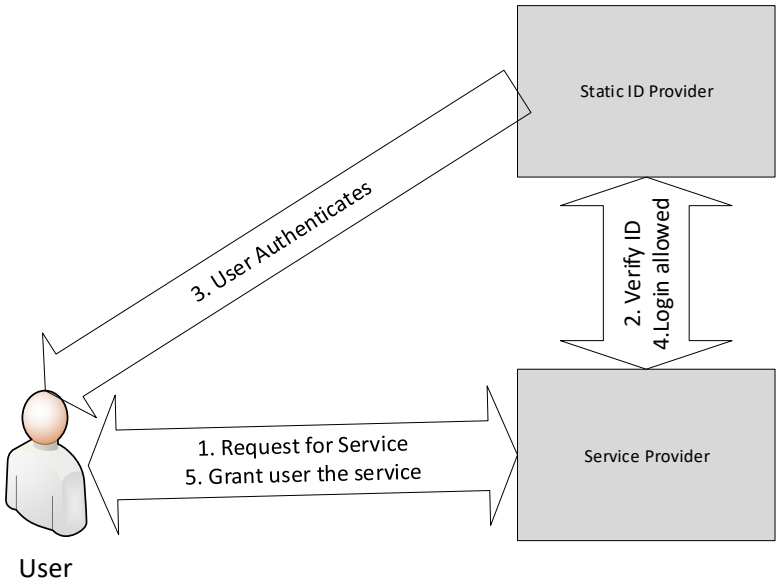


Figure 2. 1: Static ID Provider Concept [22]

The use of static electronic IDs is not adequate for providing privacy to the users as a static ID can be profiled thereby compromising the privacy of the user [82]. Furthermore, the use of Block chain technology might not be very feasible as the technology is currently resource-intensive [83]. A global solution based on this proposal would consume a huge number of resources for the proof of work to be used to protect data from being leaked or modified or even deleted. Block chain technology requires some modifications to make it feasible otherwise in its current form it would not be very useful for this application.

Further, Block chain technology has an inherent scalability challenge due to its design; it experiences time lag; when one node generates a transaction, several other nodes, about a minimum of three, need to confirm and reach a consensus before a transaction can be considered as complete. This results in transactions delaying to complete [84][50]. The solution proposed by Frank and Michael seeks to solve a global problem

by providing global access hence scalability is very key to accommodate everyone who desires to employ the solution. Block chain also faces some privacy and security issues which need addressing. Deploying this solution to address the challenge of privacy might partially resolve one problem and yet introduce more issues [83]. In addition, the complexity of Block chain makes the cost of building and maintaining it prohibitive. Cheaper ways of developing the technology need to be sought if it is to be widely adopted. It can, maybe, be built as a service where costs can be shared [85].

Most e-commerce platforms keep client information online so that they can easily authenticate the users before granting them access to any service. A lot of user data is being held in the cloud by various service providers hence making that data susceptible to leakage [86]. Attacks have been orchestrated against e-commerce platforms such as eBay which suffered a Distributed Denial of Service (DDoS) attack where their databases were scanned and client data was exposed [87]. The attack was possible as the databases were accessible via the internet. Exposure to the internet widens the attack surface as the perpetrators do not need to be in the same country as the target data. All they need is access to the internet and their reconnaissance and vulnerability exploitation tools.

Fanghan et'al [86] proposed a model depicted in Figure 2.2 that gives data owners the power to decide who to grant access to their data. They proposed the use of a cloud server to store some of the user data. The user encrypts their data before sharing and decides who can be given access by sharing their decryption keys with only the authorized users. The approach users searchable encryption that employs cryptographic primitives that allow a search of key words on encrypted data [86]. Figure 2.2 below shows how the model was designed to operate.

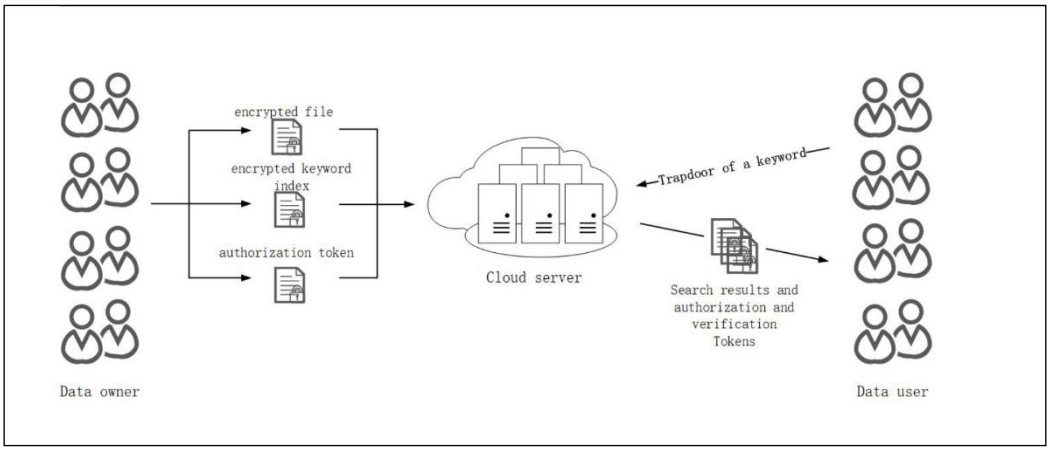


Figure 2. 2: Data Owner Controlled Access [86]

Even though the data is encrypted, if the key shared finds itself in the wrong hands, then the encrypted data, if accessed can be deciphered. In addition, it means several users can be granted access to the data thereby increasing points of possible data leak. In fact, despite being encrypted, data is most likely to be in plain text when being processed, for example in response to a request for data, hence making it vulnerable to leakage [88]. Moreover, the cloud is merely a computer system under the control of someone else. The use of the cloud in itself already partially surrenders the control of the data to the owners of the cloud [89][90][91][92]. The storage owners might promise the data owners of stringent measures to protect and not abuse their data, but the data owner has no control over the actual behaviour of the storage owners. Further, there are serious security concerns with the storage of data on the cloud. These include security issues to do with how data is stored, how data is transmitted and retrieved, and how users access this data as well as how to preserve the privacy of data owners [93][94][92][95]. The cloud being interconnected computer resources accessible via the internet, it is very possible that during data transmission, data retrieval, and data in storage can be accessed unapproved by disgruntled internet users. These concerns discourage some from adopting cloud services [96].

The cloud is a virtual pool of resources like storage and computing power based on the internet which users with computing needs share via the internet [97][98]. It mainly comprises networked applications and storage. The advantage of the cloud is that users do not need to be in the same place where the servers and datacentres they need to access are. Figure 2.3 below shows a typical cloud and how users access the various services in the cloud.

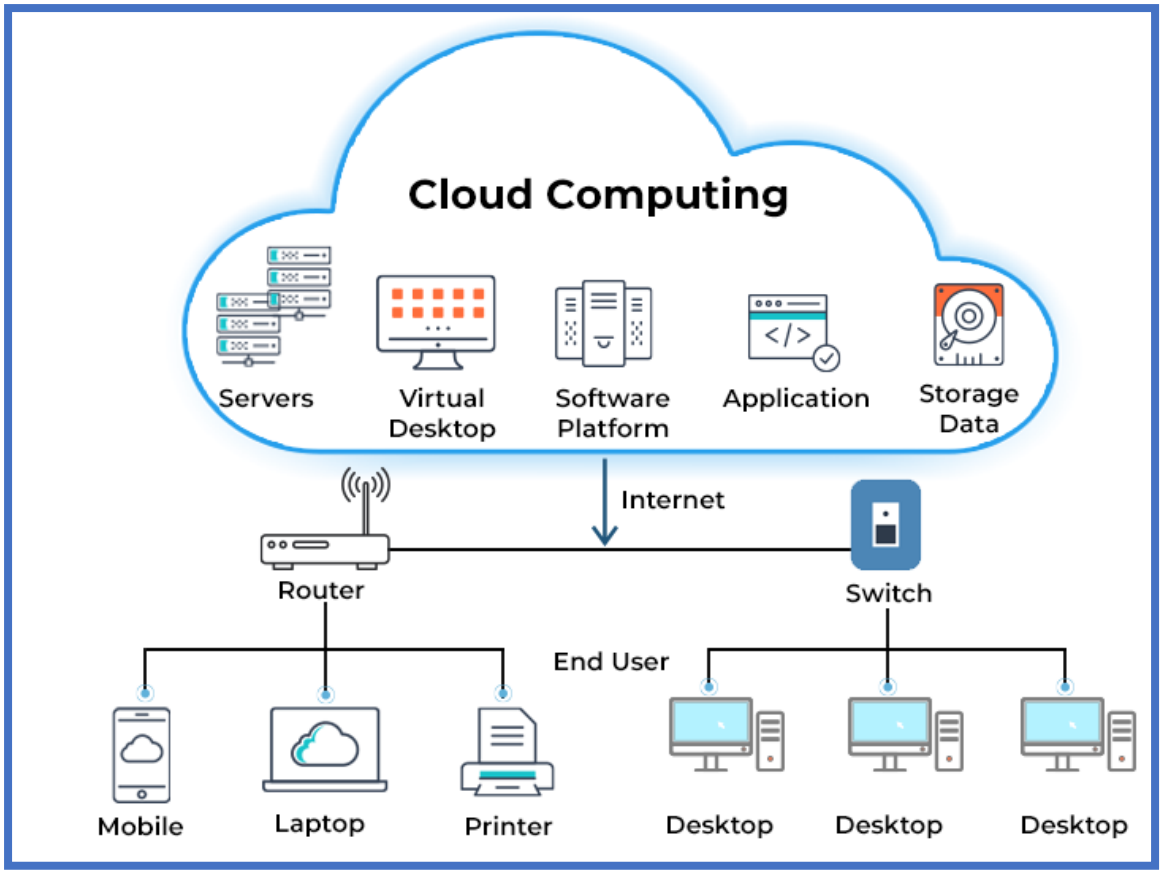


Figure 2. 3: Depiction of a Typical Cloud [99]

Users do not even need to have some of the services they need on their computers as shown in Figure 2.3. Services such as storage, software, and many others are hosted in the cloud. All they need is a computer that has access to the internet [91][96]. They can access storage, applications, virtual computer platforms, and many other services.

There are four broad categories of clouds: Public, Private, Community, and Hybrid clouds. The public cloud is accessible by users via the internet through a service provider while the private cloud is owned by an organization and is not available to public users. The community cloud is owned by a few interested parties with a common goal, and they share data. The Hybrid cloud is simply a combination of the private and public clouds [97].

The cloud has come with challenges such as data privacy, anonymity, and reliability issues; security is the issue of greatest concern. This has resulted in less trust for cloud services by some stakeholders. Most solutions used to address cloud security concerns are based on encryption to enable secure data sharing [90][100][25]. Data encryption is the conversion of plaintext (human-readable text) to cipher text (not readable by

humans). The cipher text is converted back into plaintext when it reaches its intended target. The conversion back to plaintext is called decryption [101][91]. Figure 2.4 below shows how encryption and decryption are done to securely send data over some electronic communication media.

Figure 2. 4: Secure Data Transmission [94]

Both encryption and decryption use a key which could be a shared secret or one private and another public depending on the infrastructure and setup being used. This is to protect data from being accessed and read by unauthorised individuals [101]. In the figure above, C is the cipher text resulting from the encryption algorithm working on the Plaintext P using an encryption key k1. P is the plaintext resulting from the decryption algorithm working on the cipher text C using a decryption key k2 [94]. There are two types of encryptions: Symmetric and Asymmetric. Symmetric uses one common key for both encryption and decryption while Asymmetric uses a pair of two keys; one private key for decryption and one public key for encryption [94][95][102]. Encryption is necessary before transmission as communication channels are susceptible to intrusion and unauthorised access [103]. Encryption helps protect the transmitted data from unauthorised access.

Another Trusted Third Party Model was proposed by Locher et'al [104]. They proposed the use of a distributed ledger to protect data as shown in Figure 2.5. Distributed ledger refers to having several records of the same data held by multiple independent ledgers participating in the configuration [104]. The figure below describes what a distributed ledger is.

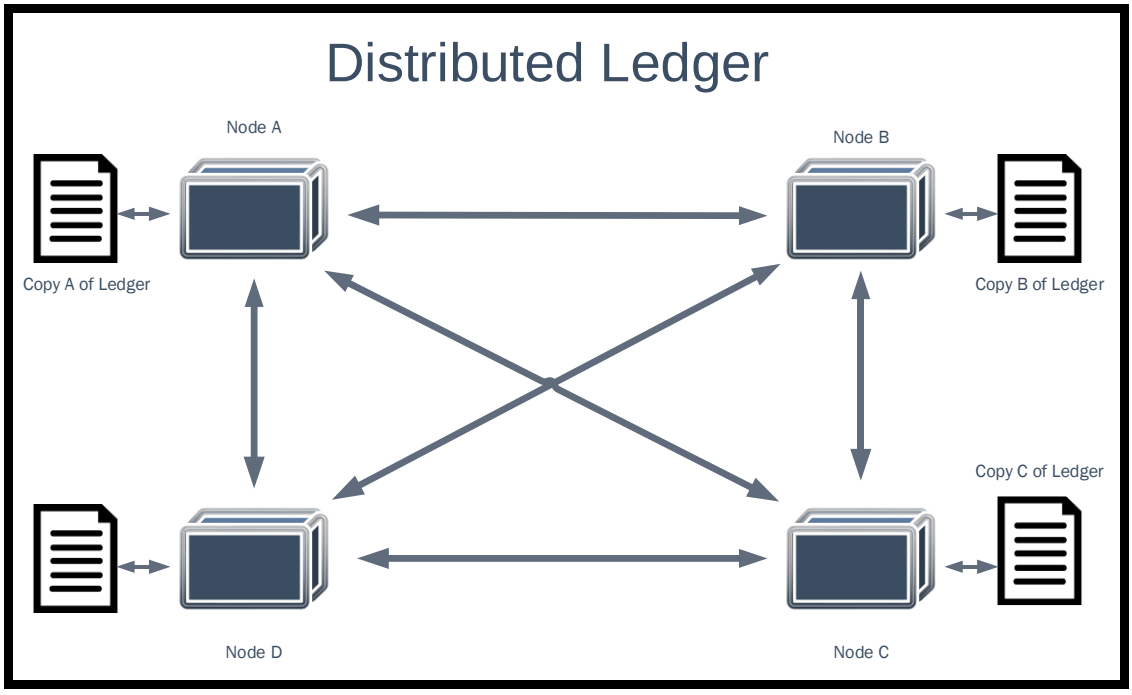


Figure 2. 5: Distributed Ledger [104]

The distributed ledger owes its security to the requirement of consensus being reached before any transaction is approved. That is before any transaction to alter any record can be completed and replicated by all participating ledgers, more than one ledger or participating node must confirm the transaction to be true. As shown in Figure 2.5, each Node keeps an exact copy of the ledger. This approach, nonetheless, results in delayed transaction confirmation as well as huge resources being required to make the technology operational [83]. Locher et'al acknowledged that despite the distributed approach of using block chain technology, users still needed to trust each other [104]. The aspect of trust is what the Trusted Third-Party model aims to address. It must also be mentioned that holding the same data in several places only makes that data more vulnerable to unauthorised access as the security measures in place might not be the same for the different participating nodes.

Aarthy et'al [105] proposed a Trusted Third Party Model that was aimed at resolving the issues of trust by users of cloud services. This model is represented in Figure 2.6. Users had huge concerns over the security and trust of cloud service providers and hence were hesitant with surrendering their data to the service providers. Their proposed model Sought to address this challenge by providing a Trusted Third Party that would monitor and assess the cloud service providers and provide assurance to users on the quality of service they would receive from their potential service providers

[105]. Figure 2.6 below gives a summary of how their proposed model was expected to work.

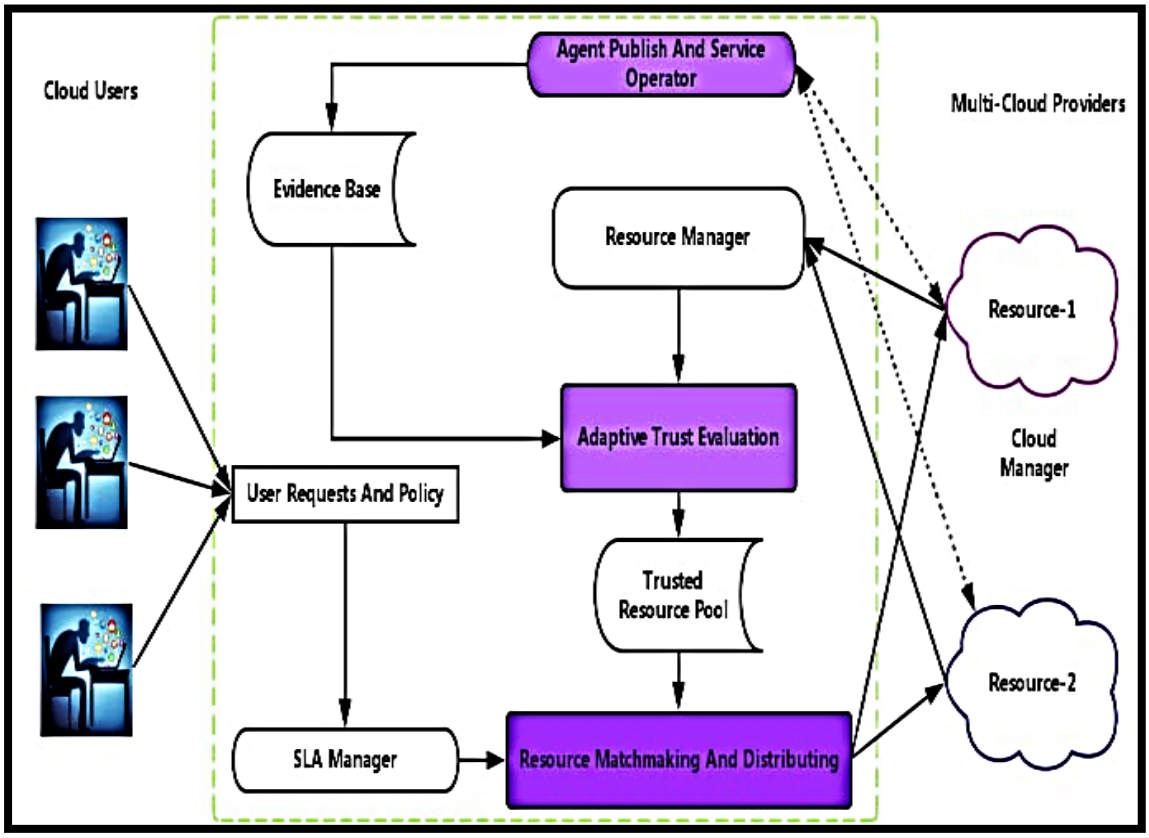


Figure 2. 6 : Trusted Third Party Model for Cloud Services [105]

The idea was to have a party that both the service providers and potential users of the services could trust. The data about the quality of service the cloud service providers are offering is aggregated together and held by a central and neutral party. The Trusted Third Party provides that service. Further, the model proposes the TTP to maintain a Service Level Agreement (SLA) with the cloud service providers against which the quality of service they provide would be evaluated. The approach was expected to improve trust and eventually the adoption of cloud services. This model can be applied in other areas such as the provision of Know Your Customer (KYC) services though in this case, the objective is to protect the PII of potential electronic service users kept in a central place. Protecting data in one place is much easier than protecting data distributed across various online platforms.

The Trusted Third Party Model was also tabled by Jamshiya et'al [106], to provide trust amongst Internet of Things (IoT) devices where security and trust establishment is a

challenge. Their model is shown in Figure 2.7. There are privacy and data security concerns around the use of IoT devices [107][108][109][110]. For these devices to connect and start sharing data, trust needs to first be established. To achieve trust, a third party that can be trusted by both parties needs to be in place. The TTP then generates a key that is distributed to all parties that need connecting. Each IoT device is first connected to the TTP and assigned an id to be identified. Then the TTP can now be used as a Trustee to tell other devices that desire to connect whom they can trust and connect with. This, of course, is done via the encryption of exchanged keys. Elliptic Curve Cryptography (ECC) is employed because it provides strong cryptography with a smaller key length [106]. Figure 2.7 below shows their proposed model:

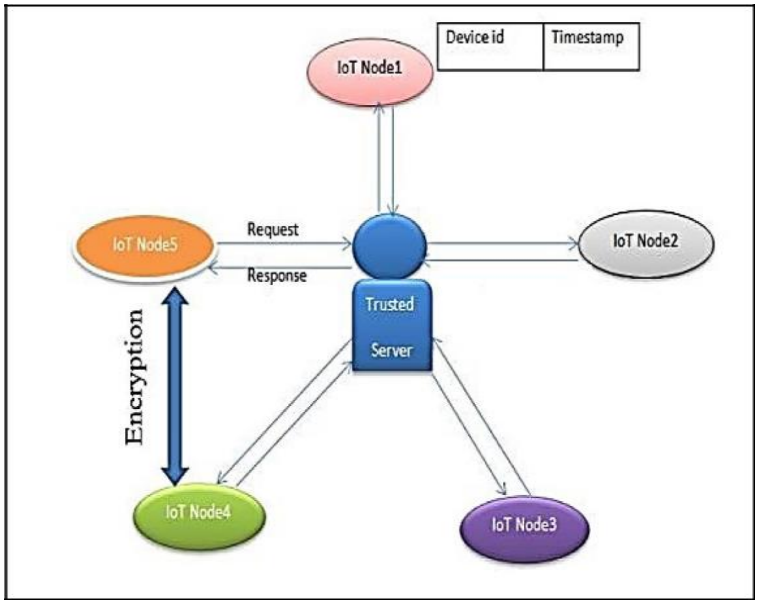


Figure 2. 7: Trusted Third Party for IoT Devices [106]

The major element in the model was the inclusion of a party that establishes trust in advance with different parties (in this case IoT devices) that might potentially connect in the future.

Block chain has also been proposed and used for the storage of data and tokenization for IoT applications to communicate with each other. However, block chain technology might not have reached a level where it can be used at the industry level for this purpose [111].

2.2 The General Data Protection Regulation (GDPR)

Peter et'al recognize the General Data Protection Regulation (GDPR) as one way of addressing the prevalent consumer privacy challenges. The European Union proposed the GDPR as a way of protecting the privacy of individuals by promoting the pseudonymization of PII in conjunction with other existing data security techniques [112]. The additional measure indicates that the existing techniques are no longer adequate hence the data leakages and privacy violations that are experienced very often. Peter et'al defined pseudonymization as morphing data in such a way that the resulting data cannot be associated with the original owner without the presence or requirement of additional information. In other words, the resulting data would not be sufficient to identify correctly who the owner of that set of data is. The authors proposed that Pseudonymization techniques be applied by various data processors such as mobile operators to protect user privacy. Techniques such as scrambling or obfuscation, blurring, masking, tokenization, and encryption were proposed [112].

Obfuscation is the transformation of data from a form that is easily readable and interpretable into a form that is unintelligible [113]. That way the resulting data becomes useless as far as identifying the owner is concerned hence preserving the privacy of the data owner. The same technique is also used by malware writers to evade antivirus detection. The malware is written is such a way that traditional anti-virus that uses signature for identification, is not able to identify as the uniquely identifying information would have been obfuscated [114]. On the other hand, data masking is simply hiding sensitive data from being accessible to the public [115]. It can involve replacing real data with similar but unreal data. For example, data with names of people and residential addresses can be masked by either putting asterisks on sensitive data elements such as names and residential addresses or replacing real names and addresses with fake ones.

Advancement in technology has resulted in the need for more effective techniques and solutions to provide security and privacy to personal and other sensitive data. The current solutions might not be sufficient to meet the required levels of privacy and security demanded by regulations such as the European GDPR [116]. The use of pseudonymization and anonymization is recommended depending on the purpose of the application. For example; Pseudonymization is preferred to anonymization when

providing a solution for the protection of Health data for people [117]. Pseudonymization provides a possibility of identifying the actual individual by using additional information when the need arises. Anonymization, on the other hand, alters data in such a way that it can no longer be traced back to the actual owner. This can be applied in a situation where data is being stored in public just to indicate what is obtained statistically without compromising the privacy of the actual individuals represented by the data [18][118]. For example, statistics showing the age groups and their preferred mode of transport or models of phones and so on. The combination of pseudonymization with other security techniques such as hashing of pseudo IDs and encryption of pseudonymized data is recommended to help enhance the protection being provided [116]. There is a need to ensure that only pseudonymized data is made online while raw identifying data is kept offline. Furthermore, necessary internal controls must also be put in place to ensure data is not leaked by internal parties.

2.3 Electronic ID (eID)

To have a presence online, especially to access most online platforms such as those used for e-commerce, one needs to have some form of electronic identity by which the accessed platforms can identify them. S. Nimalaprakasan et al [78] discussed the concept of an electronic ID (eID) which would mimic the physical National Identification Cards that are issued to citizens in each country for identity. The eID would work like the conventional ID except that it would be electronic and get used to acquiring electronic services such as having access to e-commerce and e-government services [78]. The Figure below shows the conversion from a Physical ID to an eID.

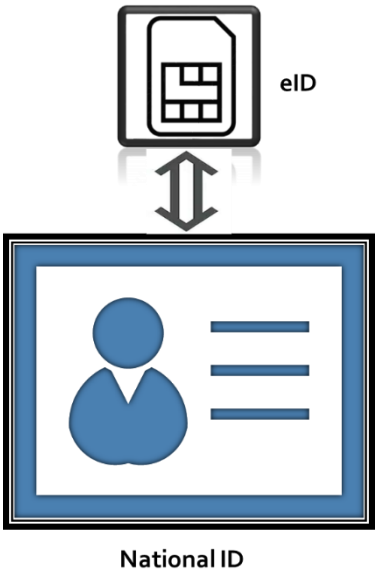


Figure 2. 8: Electronic National ID [78]

The eID would be static just like the conventional National ID. The challenge with the use of a static eID solution is that users can be tracked and profiled using their eIDs and hence have their privacy violated. Tracking and profiling are done by monitoring and recording activities associated with specific eIDs. The profile will contain the behaviour, preferences, and unique tendencies of a given ID and that would be enough to compromise the privacy of the customer by, for example, some marketing companies sending the users unsolicited adverts. The abusers of personal data have no motivation to warn the affected individuals [119].

2.4 Pseudo ID System

S. Nimalaprakasan and team [78] proposed a Pseudo System were the eID was not the real identity of the user being represented because whenever a user gives out their PII, their privacy is no longer guaranteed depending on the service they are trying to access. The Pseudo name would be different from the real identity of the user but can be associated with the actual user details by the handlers of the data [78]. Figure 2.9 below shows how a pseudo-ID operates. The actual identity of the suer is collected and converted into a Pseudo name that has a very loose connection with the actual identity. Then the user starts using the Pseudo ID for online transactions.

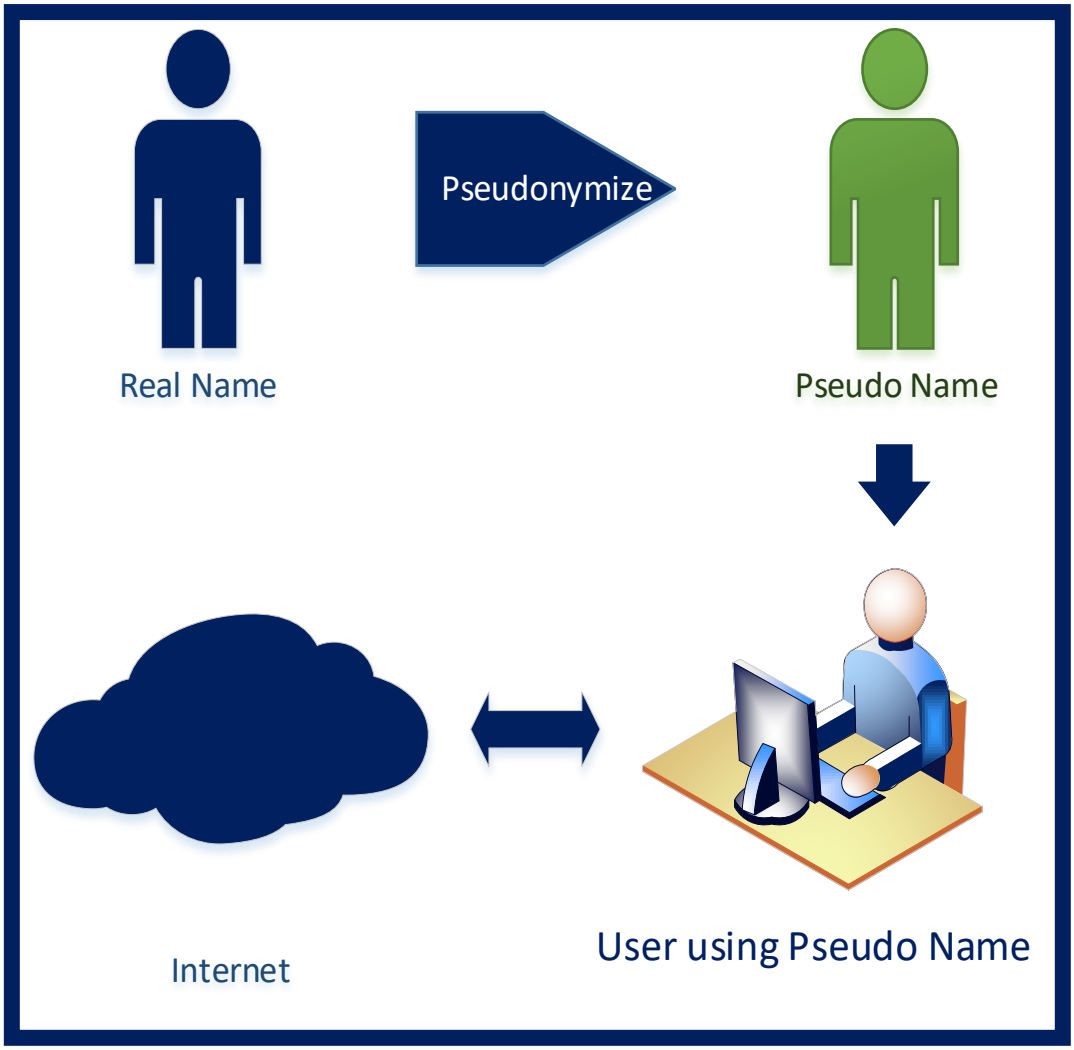


Figure 2. 9: Use of Pseudo ID while Online [78]

However, if the ID used, even though pseudo, is constant, it is possible to profile the owner of the pseudo name and hence have their privacy compromised. The solution required is to have a user access e-commerce service but still maintain their privacy even about their lifestyle, likes, and preferences without someone stalking them electronically.

2.5 Random eID per Transaction

An improved version of the eID is the use of random pseudo codes. That is, for any transaction conducted by the user, a unique pseudo name or code is generated only for that transaction. This approach addresses the challenge of the profiling that comes with static Pseudo names. User profiling is the collection of user-specific data to understand user behaviours and traits and hence enable companies to provide services suited to the subject [120]. Profiling via the use of IDs becomes close to impossible as each

transaction uses a unique code to identify the user. This approach will make the identities of the users untraceable [121]. The diagram in Figure 2.10 depicts the process of generating a Random ID.

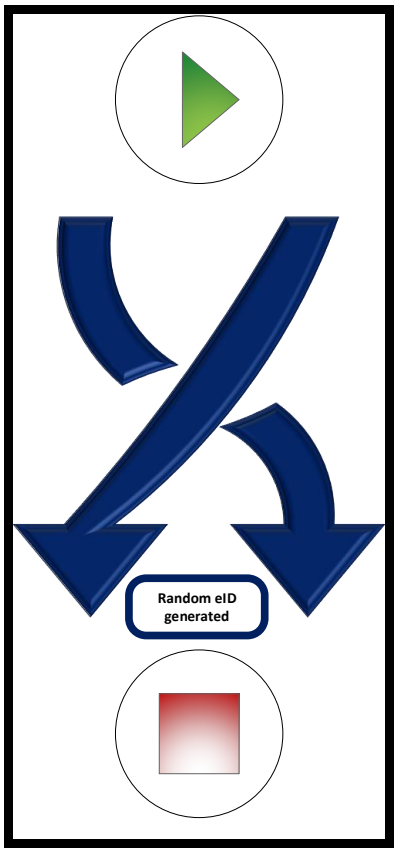


Figure 2. 10: Random eID Generation [120]

Figure 2.5.1 depicts the random generation of the eID. Whenever the user wants to transact online, they obtain a randomly generated ID which is always different from the previous ID. However, this does not address the challenge of PII leakage. The user can protect their Identity by using the Pseudo name but the handler of their PII associated with their Pseudo name can be a point of leakage by putting the PII online rendering it susceptible to hacking online perpetrators [54].

2.6 Data Leakage Prevention (DLP) Systems

Data leakage continues to be a challenge for both companies and individuals. Individuals can lose PII while companies can lose Intellectual Property (IP) which other competitor companies can take advantage of. Public awareness of data leakage can damage the reputation of a company [122]. Traditional information security protection tools such as firewalls, intrusion detection, and Virtual Private Networks (VPN) lack

proactiveness in the prevention of data leakage [60][123]. Another proactive and more effective approach to addressing leakage of data has been the development of Data Leakage Prevention Systems commonly known as DLPs. The operation of a DLP system can be based on keywords, statistical methods, or the context of the words used [124]. The context-based would be preferred as it takes advantage of the strengths of both keywords and statistical methods to derive more meaningful conclusions. Figure 2.11 below gives a description of what a context-based DLP entails and the processes it follows to prevent data leakage.

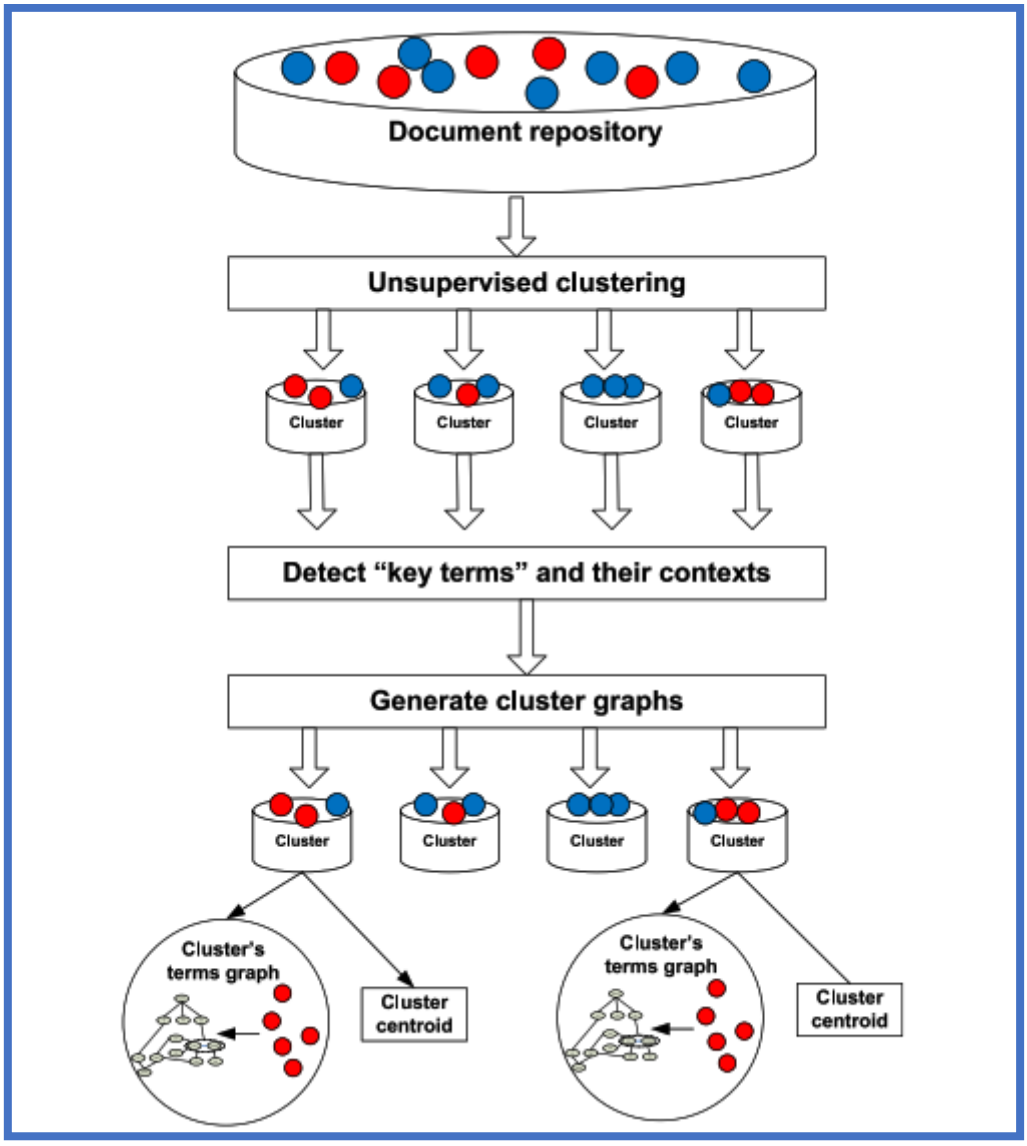


Figure 2. 11: Context-Based Content Filtering [124]

The system, as shown in Figure 2.11, has a document repository and creates clusters of key terms to detect. While the document is being inspected, cluster graphs are generated

showing the levels of presence of the key terms in the document being inspected. If the term graph shows levels above a given threshold, an alarm is raised that data leakage has been detected.

This class of technologies uses various algorithms to prevent the leakage of sensitive data. Data to be protected can be at rest, in transit, or use [125]. For example, credit card data can be blocked from being sent or an alert is sent to appropriate personnel if the defined data to be protected is being leaked via various channels such as email and moveable media like USB (Universal Serial Bus) disk drives. The figure below shows how a DLP works.

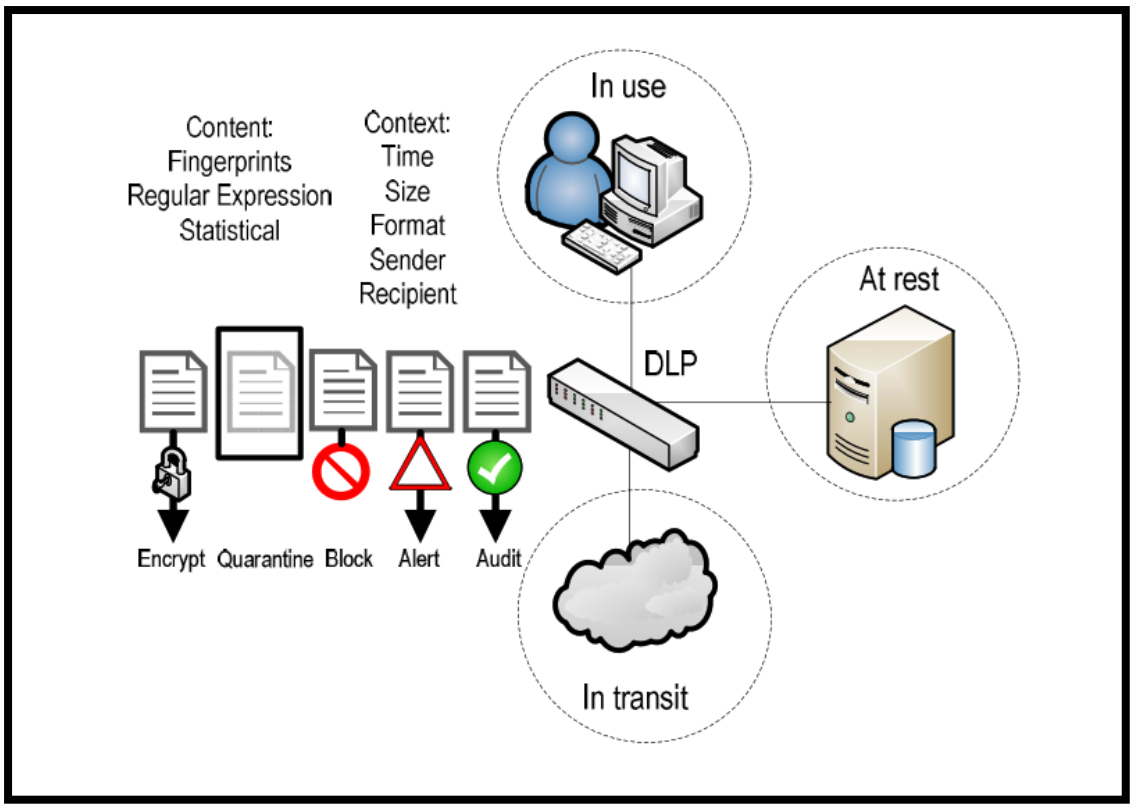


Figure 2. 12: Data Leakage Prevention System Operation [125]

Figure 2.12 shows that data can be in transit, at rest, or in use. Whatever the form, it is inspected by the DLP. The DLP conducts an audit of the data against defined terms. If a violation is detected, an alert is triggered, and appropriate action is taken. The action can be to block the sending of the data, it can be to quarantine the data, or even encrypt it.

The challenge with this approach is that sometimes the systems fail to detect sensitive data when it is encrypted or hidden in images. In addition, depending on the amount of data a DLP is dealing with, the DLP can slow down the processing speed of a system as it will have to scan, scrutinize and analyse every data element traversing through it [60]. Moreover, some DLPs can send alerts that data has been leaked but it might be too late as the data would have already been leaked anyway. Furthermore, DLPs are usually inadequate to prevent leakages of sensitive data via Peer to peer (P2P) networks as these normally use various ports including ports used by the famous internet protocol HTTP. It would, therefore, be difficult to block various ports as that would affect several services [126]. There is a need to protect PII in whatever state it is in; whether at rest in storage or memory while in use or when in transit moving from one point to another.

2.7 Privacy Enhancing Technologies (PET)

Another class of solutions developed or proposed to address the challenge of privacy violations known as Privacy Enhancing Technologies (PET) has been on the rise [127]. PETs are technologies used to protect the privacy of individuals while they are online. They help prevent the exchange of personally identifying information including life style, preferences, and so on without the consent of the data owner [128]. One solution implemented on an individual level is the Ant-tracking technology where the user prevents tracking technologies such as cookies from collecting data on their activities such as sites they are visiting and their PII fed into their web browsers. Another technology is where the users do not send identity data directly to a website but goes through an entity called anonymizer. The anonymizer in turn sends the data to the intended destination and receives the response on behalf of the users and communicates the feedback to the user. Data perturbation is another PET recommended for maintaining privacy. This is where user information is mixed with false data or the user provides completely false data when communicating on the internet to hide their real identity [54]. The challenge with all these technologies is that they do not address the weakest link of user privacy in e-commerce; the leakage of PII.

This research focused on protecting PII using Offline Data minimisation and pseudonymisation to ensure the privacy of users is guaranteed. The solution will help

improve consumer trust in e-services platforms and thereby encourage the use of the platforms.

2.8 One Time Password Security

Most e-commerce sites currently employ usernames and static passwords as authentication credentials otherwise known as single-factor authentication. Some use Unified Access management (UAM) to try and ease access management for users on multiple platforms and devices. The approach has its weaknesses such as man-in-the-middle attacks, Phishing attacks, DoS attacks, and so on [129]. Due to prevalent online identity theft, these modes of authentication are not very secure. Once the static credentials have been harvested, it becomes easy for the attacker to gain access to the victims' online accounts hosted by the affected platforms [130][131][132]. In addition to using a username and static password, one's security can be enhanced by adding another authentication factor like the One Time password (OTP) which many online service providers have adopted for use to authenticate users on to their platforms [133][134][135]. Although this approach is acknowledged to be more secure than the static password authentication method, it is considered not to be very user-friendly because of the additional steps users have to take to authenticate [136] [132]. For the user to logon to their e-commerce platform account, they would need to provide their username and associated password then an OTP would be delivered either to their phone or email box. Hence for someone to access the user account on the e-commerce platform, one would need to know the username and password and must also have access to the phone or email box [137]. The OTP is valid for only one login session after which another OTP, usually different from the previous ones, would be needed to access the account; an OTP is dynamic. That's, the value of the OTP keeps changing [130][138]. The dynamic nature of the OTP makes it more secure than the traditional static password. One would need to understand and crack the algorithm used to generate the OTP to predict the next valid OTP. Further, the dynamic nature of the OTP helps mitigate the challenge of keystroke capturing. A software or hardware key logger can capture passwords by recording the keys a user presses when logging onto their accounts. A dynamic password (OTP) becomes difficult to crack as it will change for every session [139]. These are mostly implemented in financial transactions and e-commerce by Banks to help mitigate fraud [140]. Multifactor authentication, therefore,

enhances security by making it more difficult for a hacker to get access to more than one login credential delivered via different media [141].

There are broadly two categories of OTPs; Time dependent and Time independent OTPs. Time Dependent OTPs are known as Time Based OTPS (TOTP). Time based OTPs change with time. They can be implemented to change within steps of thirty (30) seconds or any other desired period. The server and user sides do not need to be constantly communicating if they use the same time reference point. There is a need to synchronize periodically to ensure the time reference remains the same [142]. The RFC standard 6238 uses the Unix epoch time as the varying input into computing the TOTP [143]. Apart from the TOTP standard, another standard is the HMAC (hash-based message authentication code) based OTP (HOTP) which uses an incrementing value and a secret key as input to generate a dynamic OTP [144]. This one does not need periodic synchronization with the server as the incrementing value is not time dependent. However, it sometimes faces challenges when the incremental value goes out of synch creating a mismatch between the authenticating device and the platform authenticating the requesting device or person. The problem is usually addressed by allowing a given range of values to be accepted as valid OTPs when the mismatch value is within an acceptable range.

OTPs can be implemented in various ways. The most common and popular method has been the use of SMS for the delivery of OTPs to the user. However, this mode of delivery is insecure; several user SMS OTPs have been breached before. Large-scale password leakages have occurred before. Another common challenge with OTPs delivered via SMS is a delay. Since some OTPs are time sensitive, some end up getting delivered after they have already expired [141]. The delay is normally from the Mobile Network Operators' side. To worsen the challenge, users usually set very weak passwords, and this makes them vulnerable to brute force attacks. One method that can be used to deliver TOTP securely includes Authentication apps such as Google App used by Google. It must, however, be mentioned that these methods might not be 100% full-proof [144]. The App or physical device generates a TOTP concerning the UNIX epoch time and the server equally computes the TOTP based on the same inputs. Then the two values are compared when authenticating to determine if one is legally

authorised to access the given account. The inputs used are unique to that particular user account [143].

Another delivery method for OTPs is the use of QR codes. A QR (Quick Response) code is displayed on the screen then a user scans the QR code to confirm that they have access to the device where the code has been "delivered" (displayed) [141]. A match results in granting access to the user.

OTPs do not protect the transmission channel through which data is transported. It does not even prevent eavesdropping on the channel by unscrupulous individuals. It only protects the authentication section of the system for the user to be granted access. For an OTP to be effective in providing protection, the properties of the OTP such as the paraphrase used to generate it, should never be stored on any system; not even protected systems. Further, the secret key and/or the paraphrase should never cross the channels used for communication to avoid eavesdropping and eventually repeated attacks [138].

Despite the implementation of OTP, fraud still takes place. This is mainly through social engineering where users are manipulated by fraudsters/hackers into surrendering their passwords [138]. Figure 2.13 shows an example of sequences of how attacks can be orchestrated for banking services. The approach is similar to other types of services. The key element is manipulating a user to willingly surrender their password.

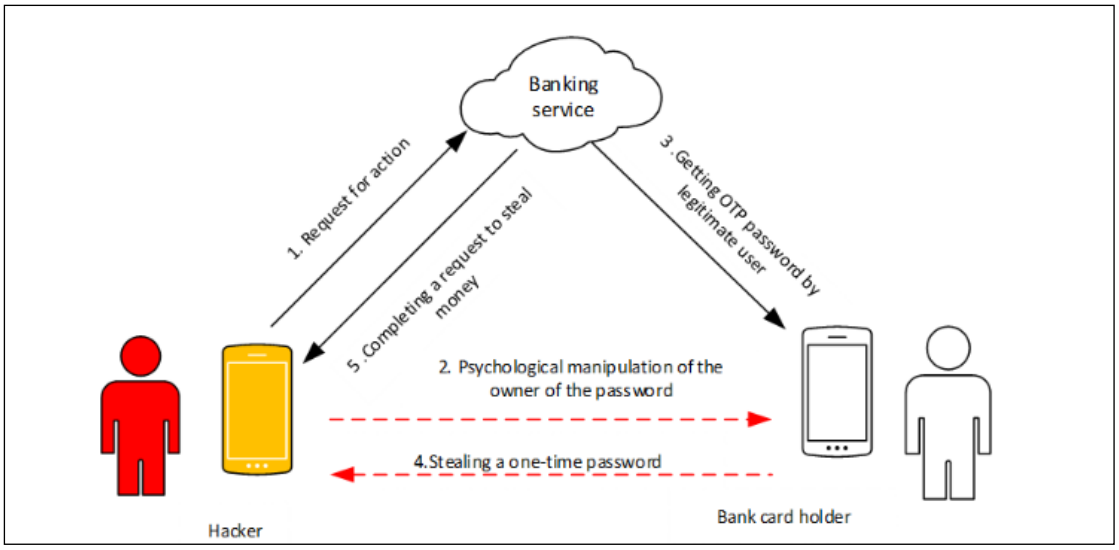


Figure 2. 13: Social Engineering Attack Sequence [138]

Social engineering attacks can be mitigated through user awareness campaigns by service providers. Users must be warned about this risk and shown how it is normally done and the symptoms it exhibits.

Other attacks on OTPs include man-in-the-middle attacks, replay attacks, impersonation attacks, theft attacks, server impersonation, Denial of service attacks, verifier theft attacks, and server modification attacks [133]. These attacks can render some OTP implementations too weak to offer reasonable protection.

To help small enterprises offer OTP services to their clients, some service providers have started offering OTP as a Service (OTPaaS). The service is cheaper to setup as the institution acquiring the service will not need to acquire and set up the hardware needed for OTP. The services are cloud-based to make it easier for different organizations to subscribe. In addition, various service providers can use the same platform hence the users would not need to manage several OTP accounts [135].

With the security issues associated with passwords, an alternate and perhaps more secure option is the use of AI voice recognition [145]. Several organizations have started using voice recognition for authentication though with some challenges such as false negatives or false positives depending on the level of sensitivity the system has been set to.

2.9 Hardware security

Hardware security refers to all actions required to unmask all hardware vulnerabilities and being able to assess their impact if exploited as well as formulating appropriate mitigations to diffuse their effects if present in a system. The best defense would be to avoid vulnerabilities through the design of the hardware. On the other hand, Hardware-based security refers to any security solution that uses hardware as the basis on which protection of information is hinged while hardware trust implies minimizing the risks posed by the hardware used to assure other systems using the hardware [13]. Figure 2.14 shows the role hardware plays in the protection of information.

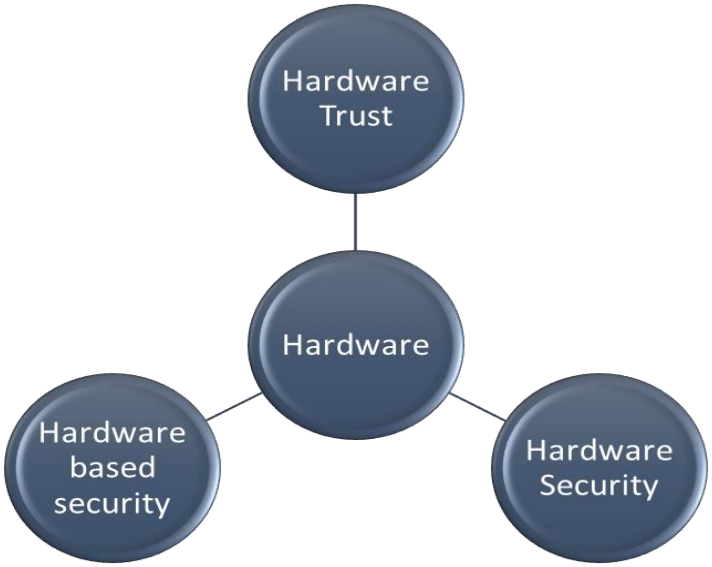


Figure 2. 14: Role of Hardware in Information Security [13]

The hardware flaws can be intentional or unintentional. Further, the flaws can be logical or physical. Logical in terms of how the system behaves given certain inputs and physical in terms of how the system is physically designed and built. The goal of these vulnerabilities is to steal, corrupt, or inhibit information or property. The attacks can be invasive (offensive) or non-invasive (passive) [13][146]. Unintentional flaws can be the result of poor design and testing that is not thorough enough to pick up and unmask bugs. Any bug that has been identified must be patched quickly to prevent its exploitation.

Hardware security has become very paramount as more Internet of Things (IoT) devices are being developed and deployed on the internet [147][148][149]. IoT is a concept where everyday life devices are interconnected via the internet exchanging data [150]. Figure 2.15 depicts the concept of IoT.

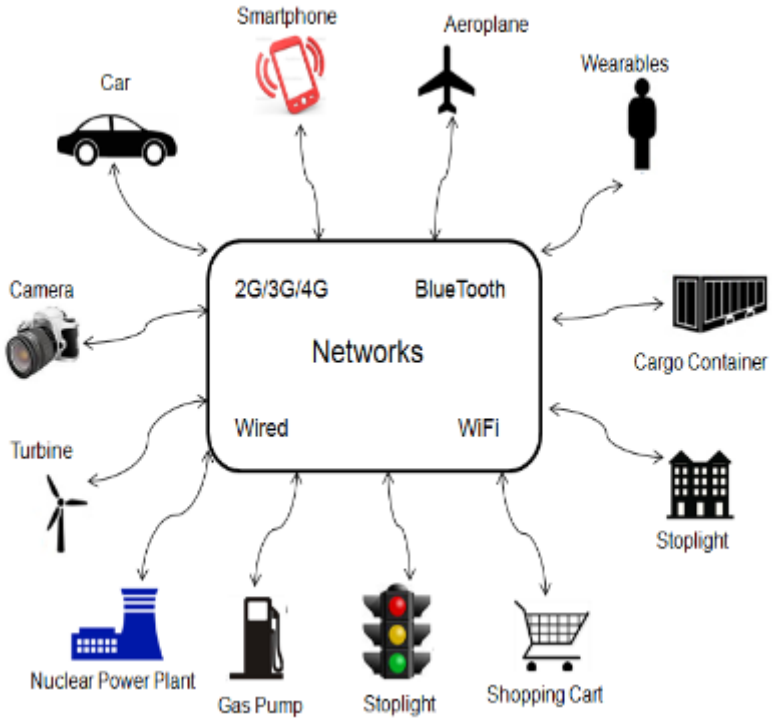


Figure 2. 15: Internet of Things Concept [150]

Software security alone is not adequate to protect users from security concerns around hardware such as IoT devices. The security of the hardware being used is equally vital especially since some Hardware Trojans (HT) can be embedded at the manufacture or design stage [151][152][153]. It, therefore, becomes very important to understand how HTs are deployed and how they operate to determine the appropriate counter measures. More hardware security challenges are popping up faster than hardware solutions are being created [154]. As more IoT devices get deployed to share data, the security of that data has become very critical[155][156]. Huge amounts of data are generated by IoT devices - being edge devices. Some of the data generated is personal hence the need to protect that data adequately to maintain privacy for the owner [100]. IoT devices are mostly controlled by Microcontrollers. A microcontroller is a small programmable processor specifically designed for embedded systems because of its small size, low power consumption, and low cost. These characteristics make them affordable and easier to handle [157][158][158] [159]. The small size of most IoT devices makes the implementation of security very difficult because they have limited memory, and limited processor power, and would make it difficult to store and process long encryption keys in addition to other security operations [160]. Lightweight security schemes are being formulated to help address this challenge. Non-cryptographic low-

level solutions are being considered to reduce the need for power, memory, and other resources [161].

An IoT device to perform its main function comprises three layers, the perception, network, and application layers. The perception layer (also known as Sensors and actuators) is responsible for perceiving things in the system and contains items such as RFIDs, cameras, various sensors, and so on. The network layer is the heart of the IoT device and transmits data gathered by the perception layer. It also manages the entire system. Network security has become a problem people must face and this layer contributes to dealing with the challenge [162][163]. To share data in whatever form, networking is at the centre. The application layer is responsible for bridging the gap between systems and human beings. It helps make the IoT device useful for human needs. [164]. The application layer can be further split into two layers namely the application layer and the data processing layer to make four layers as shown in Figure 2.16:

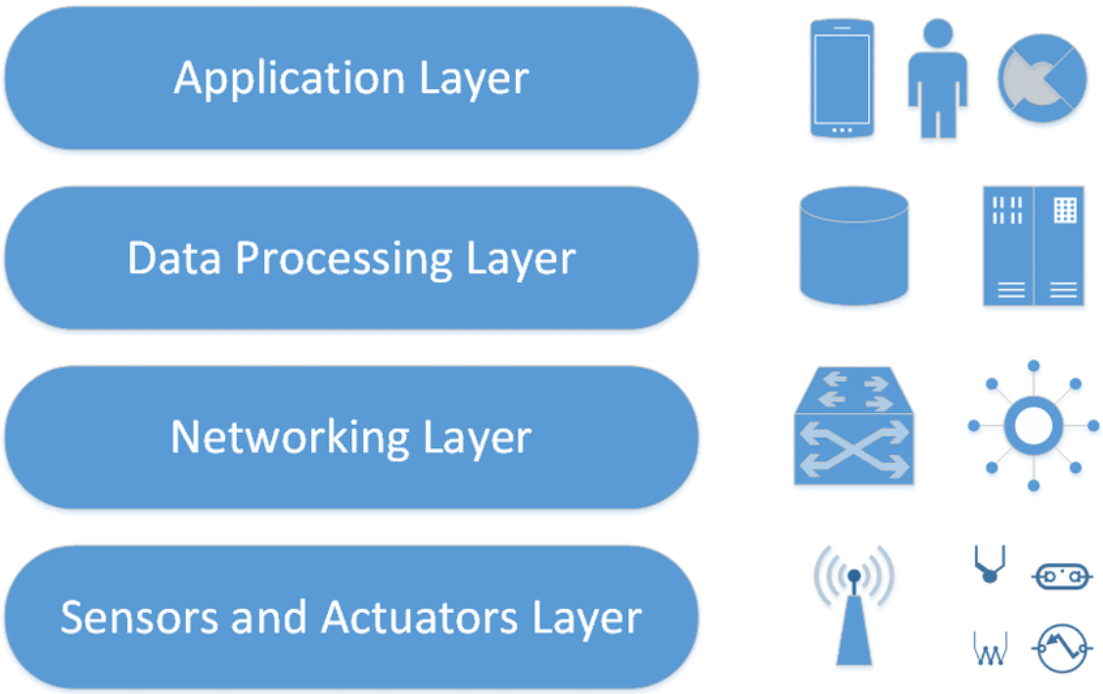


Figure 2. 16: Layers of an IoT Device [165]

To be able to provide effective security to IoT devices, one needs to understand each of the three layers and how they operate.

Microcontrollers are also known as embedded systems as they are embedded into IoT devices and other day-to-day life devices such as pressing iron, washing machine, trains, cars, and so on [166][160][146]. They are controlled by firmware which is stored inside their memory and determines how the microcontroller operates. The security of the firmware has huge ramifications for the security of people and many other institutions like governments and nations [167]. Hence there is a need to be able to audit the security of firmware to ensure that the embedded systems being used in sensitive installations are not vulnerable. Firmware is software that runs on microcontrollers and is usually written in low-level languages like C. One approach that can be used is symbolic execution. Symbolic execution is a formal way of verifying software to determine what each input gives as output. In this process, symbols are assigned for testing where each symbol represents several scenarios to determine the behaviour of the system for each given scenario. Symbolic execution is suitable for security testing of systems [168].

Secure microcontrollers are designed to protect confidential internal information. However, recent attacks have revealed that these devices are vulnerable to attacks [169][149][102]. Hidden channels can be created to be used for leaking sensitive data by sending it to the attacker's secure location [170][171]. IoT devices driven by microcontrollers are being deployed to monitor and report about the world around them. They are interconnected via the internet. However, their security has become a great concern as they are getting attacked very often. They are being targeted to be part of botnets to launch even bigger attacks [158][172][173]. These devices can be attacked in various ways such as eavesdropping, IP spoofing, and denial of service [109]. Further, even micro probing can be used to attack microcontrollers despite data in memory being encrypted. This is where equipment used for failure analysis is used to execute an attack [174][169][102]. Some have even gone to the extent of extracting the whole memory content via the use of micro probing despite the memory content being encrypted [169].

There are several other hardware (and architecture) threats that can be used to compromise the security of hardware. They include Secure Boot, Code Reuse, Speculative Execution, Firmware, and Cache attacks as well as Dynamic Random Access Memory (DRAM). Secure Boot attacks can occur when sensitive memory is

not correctly secured, and data gets erased incorrectly. The secure Boot can, therefore boot with wrong data and compromise the hardware [175][146]. Code Reuse attacks can materialize using existing bits and pieces of Code by the hacker deciding what to execute. Speculative execution can be launched when, for example, there is a race condition then an attacker speculates an attack by initiating a process that attempts to access sensitive data that it is not allowed to. Access might be granted if controls crumble due to a race condition. Firmware attacks can come about through incorrect configuration of various memories in the chip. Incorrect configurations can open loop holes that attackers can take advantage of. A cache attack occurs when a cache is being shared. Some processes can leave sensitive data such as memory addresses and traces of data they accessed. An attacker can time the victim process and observe when the desired information is in the cache then launch the exfiltration of the desired data. The attack is time driven. The DRAM attack can be executed via a cold boot then data gets extracted using another machine. It can also be staged by cooling the DRAM to reduce the leakage of current that enables the RAM to hold data [176].

The vulnerabilities being discovered can compromise system security [158]. Vulnerabilities, once discovered, can be patched [177]. However, if they have not yet been discovered, there will be no effort to address them. Even so, despite being discovered and patches being developed to address the security vulnerabilities, there remain a few users who do not try to patch their devices. It could be due to a lack of knowledge about the vulnerabilities or not being concerned with the risks that come with the vulnerabilities. There is a need to provide a solution to enhance hardware security. Patching might not be sufficient as not everyone patches their devices. Moreover, even if patches are installed, not all vulnerabilities might have been discovered and hence patched. A more effective solution is needed.

Some devices are deliberately infected with Hardware Trojans (HT) when being manufactured. HT is malicious hardware additions aimed at leaking confidential information and degrading the performance of the hardware. They can also be used to launch denial-of-service attacks [178][179][180]. The main goal of embedding HTs into hardware at manufacture is to later control and activate the HTs so that critical data can be leaked [181]. This is emanating from the new approach where hardware production is outsourced and hence hardware Trojans can be embedded from the

factory [178][179][154]. Figure 2.17 shows the various stages hardware undergoes during its life cycle. The hardware can be compromised at any stage of its life cycle.

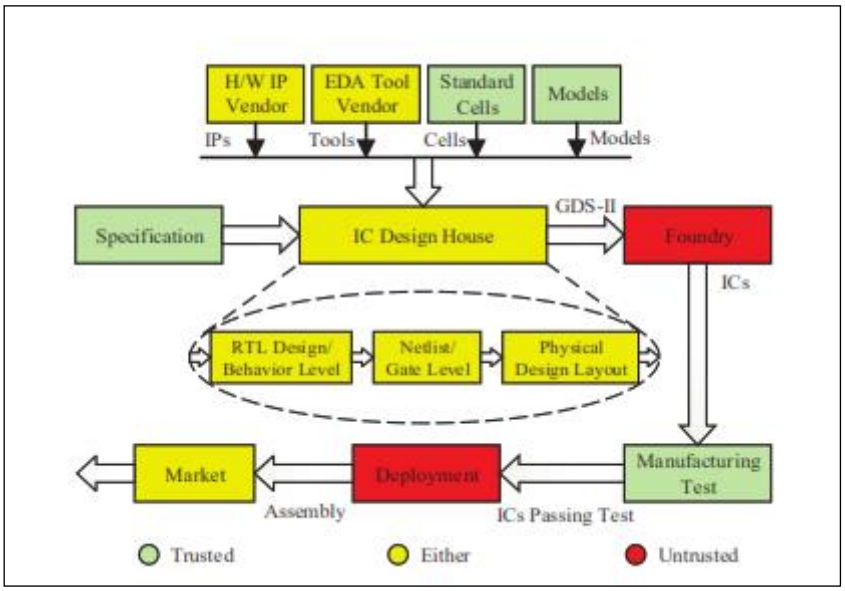


Figure 2. 17: Life cycle of Hardware [182]

The demand for more smart electric devices such as IoT home devices has led to outsourcing. There is generally a huge raise in demand globally and to meet the risen demand, most companies decided to outsource the manufacture of the hardware to smaller companies and start-ups [183][182]. Unfortunately, the approach did not analyse and assess the entire production chain to determine any weaknesses that could be introduced and provide in advance necessary mitigations. The main focus was meeting the demand and reducing production costs [13]. The approach has opened a door for HT to be inserted into the hardware at manufacture. Various attacks like the hijacking of control flow and exfiltration of secret information can also be launched. The hardware is being used as a launch pad [176].

The data leak due to insecure firmware in devices can harm user privacy as well as discourage the adoption of internet-based services [177]. Some malicious code can be installed in the firmware and run when the hardware is booted. It would be very difficult to detect such malicious code [175]. There is a need to provide a more effective way of detecting and, more importantly, preventing damage by such Trojans. One way to patch security vulnerabilities is through firmware updates. However, firmware updates can also be used to distribute malware into IoT devices. Cases abound where firmware updates have been exploited to breach the security of microcontrollers. One way is

usually to highjack the remote firmware update process and issue commands that update the firmware with malware [177][184]. Therefore, firmware updates must be carefully executed. Ensure the firmware is coming from trusted parties and has been fully tested and that there is a rollback plan in place in case the update process goes wrong. The best and most effective solution is designing a system with security in mind. There is a need for a deliberate security architecture [185][186].

It might be very difficult for a lay person, or even an expert, to notice, in good time, that there is an HT in the firmware of their devices that is leaking sensitive data. Most of these devices are deployed as single entities. This means that once one has access to this device and has managed to breach its perimeter security, then they can access all the data this device accesses. There is a need to provide a solution that would help protect the hardware devices.

One proposed solution to the embedded system security is the measuring of low-level executions that occur during the running of the firmware. The process is called ConFirm. It detects malicious code from the number of counts detected every time the firmware is run [160]. If the count is higher than expected, then there must be some malicious code executing, hence increasing the count.

Our research is proposing the use of multi-layered hardware to protect confidential data. The multi-layered approach will help prevent unauthorized access by ensuring that data can only flow in one direction: from the authorized enclave to the internet. Multi-layered Hardware security means that even if one accessed one of the hardware devices, they would still need to access the other devices to be able to gain access to the most sensitive data.

2.10 Artificial Intelligence in Cyber Security

Artificial intelligence (AI) generally refers to machines being able to perform tasks that were originally considered as needing human skills to accomplish. That is, being able to analyse and comprehend data, being able to use various languages, and finding solutions to intricate dilemmas without the need to follow pre-defined instructions. AI is often used interchangeably with Machine Learning (ML) [187].

Cyber-attacks are growing at an incredibly high rate of about 34% per annum. To compound the problem, more complex cyber-attacks are becoming prominent. What exacerbates the problem is the high shortage of skilled manpower to identify and resolve complex attacks. This shortage is not expected to get resolved soon. This is where AI comes in. AI can help address the cyber-attack challenge in several ways when combined with human ingenuity together with the capability of machines to process zettabytes of data in a short time [188]. AI can enable machines to recognise attacks in huge volumes of data, come up with a course of action and automatically execute the response plan as humans would but at a faster pace. There are already tools such as Dark Trace that are using AI to detect and respond to cyber-attacks [187].

Captcha password is another example of AI used in online security. Captcha distinguishes humans from computers by posing a puzzle difficult for computers to comprehend but very easy for humans hence making graphical passwords more secure [189][190][191]. For example, in addition to providing a username and static password, the systems also request the user to click on all trees in a picture displayed [192]. The result is binary. That is, the computer system will conclude if the user is a robot or not. If it is a robot, even if the static password and username match, access is denied. This addresses the challenge of online dictionary password attacks [193]. Online password attacks can cause DoS attacks as the system is forced to keep responding to login attempts.

Another contribution towards security by AI is voice recognition authentication. This is where voice is used as a biometric authentication measure to protect data and user privacy. Several AI voice solutions have been developed and deployed in real life. They include Alexa, Google Home, Amazon Echo, and Siri. Keypads and text-based passwords are vulnerable to hacks [145]. Voice recognition will soon become one of the preferred modes of authentication.

However, the capability of AI can also be used for harmful activities. Attackers have already started using AI-based cyber-attacks in reconnaissance, infiltration, and privilege escalation [187]. They use AI-based tools to study and understand human behaviour which they later exploit. There is a need to understand in depth how AI is being used by invaders to launch attacks to devise appropriate defense strategies. The

understanding can bring about effective detection of cyber-attacks such as the detection of suspicious URL and password flaws [188][194].

2.11 Serial Communication

Serial communication is commonly used for the transmission of digital data. Serial transmission implies that data is sent in bits one after another on the transmission line. Microprocessors process data in parallel mode hence the resulting data has to be converted from parallel to serial before transmission by the transmitter and from serial to parallel when received by the receiver as shown in Figure 2.18 [195].

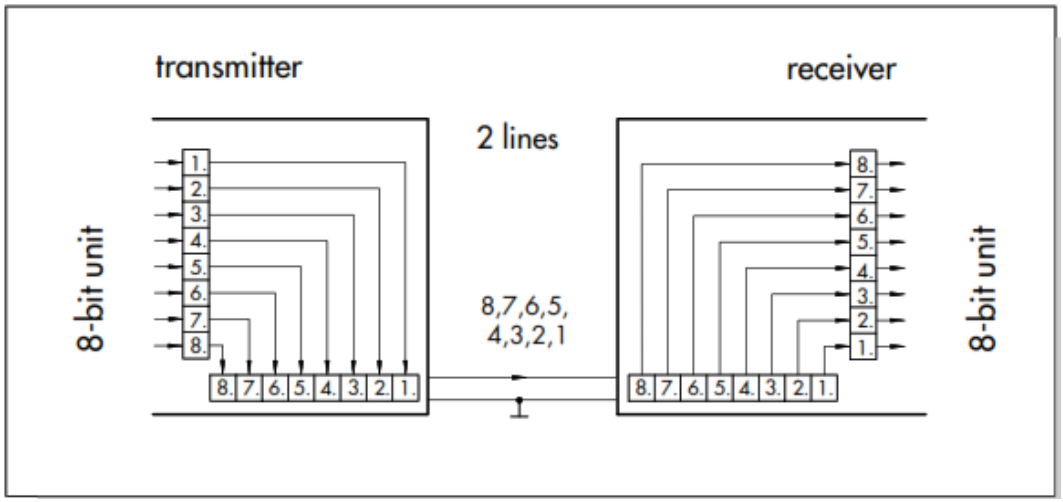


Figure 2. 18: Serial Data Transmission [195]

Serial data transmission can be used to share data between two targets or amongst several targets. There are three types of serial communication namely, Simplex. Half-Duplex and Full-Duplex [195].

In Simplex communication, data transmission is one way only while in Half-duplex communication data transmission is two ways, but each participant takes turns. The Transmission is not simultaneous. In Full-Duplex communication, data transmission is two-way and can be done simultaneously [195]. Figure 2.19 describes the three modes of serial communication.

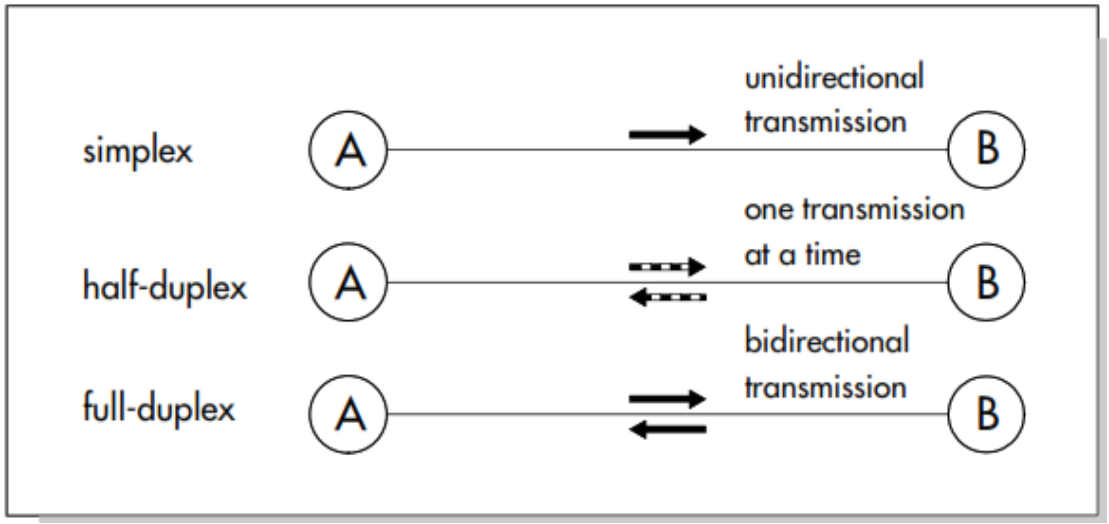


Figure 2. 19: Serial Modes of Communication [195]

In a point-to-point serial communication where data transmission is two-way, the transmitter of one device connects to the receiver of another device. This can be a Half-Duplex or Full duplex as shown in Figure 2.20.

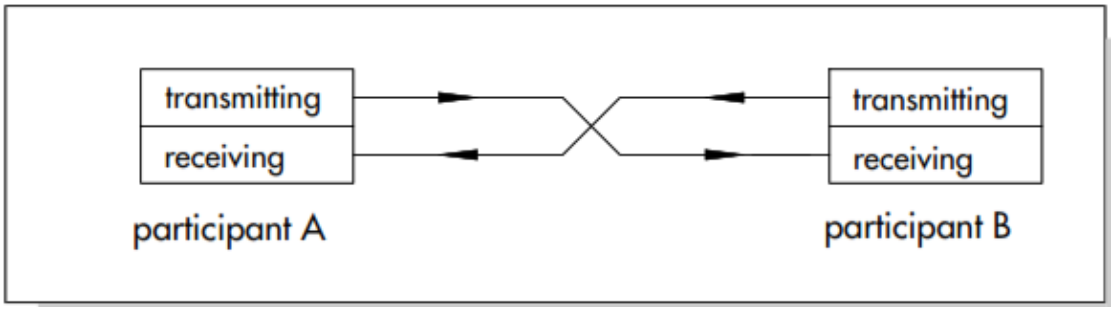


Figure 2. 20: Serial Modes of Communication [195]

Another critical aspect of serial communication is the transmission speed. Data is transmitted in bits per second denoted as “bps”. Modern Serial communication has high speeds such as 9600bps [195].

Further, for a two-wire communication control, the system can either be asynchronous or synchronous. Asynchronous control involves the receiver synchronizing with the sender through the start and stop bits while in Synchronous control, the two communication entities synchronize using a clock [195]. In asynchronous communication, speed must be agreed upon in advance while in synchronous

communication speed is controlled by the device controlling the clock for the communication [196]. Figure 2.21 shows a synchronous communication setup.

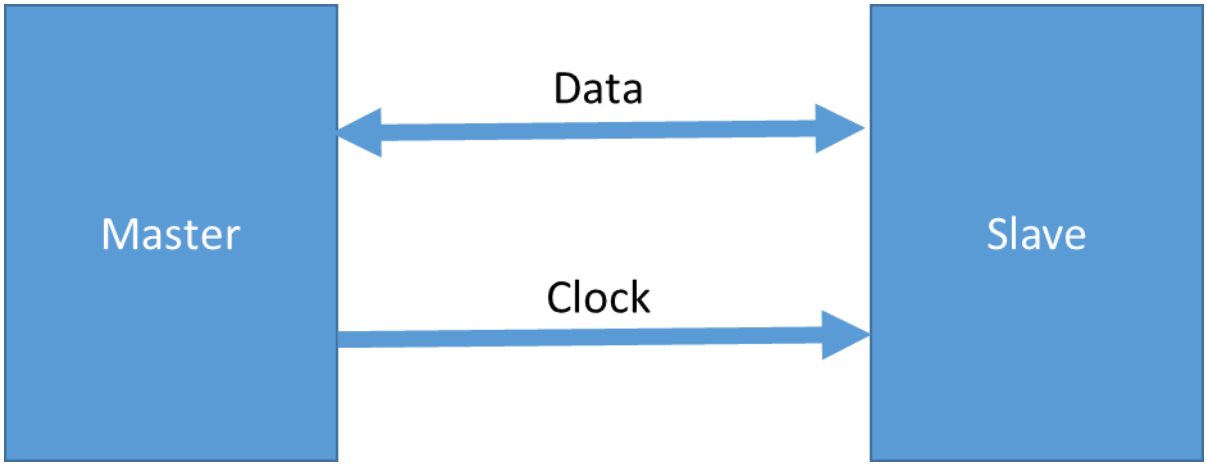


Figure 2. 21: Synchronous Serial Communication [196]

RS 232 (Recommended Standard – 232) is the approved standard interface by the Electronic industries Association (EIA) for connecting serial devices. It is normally used by devices that do not require high speed to exchange data. It has been in use for many years. The standard physical interface is the DB9 (d-type connector with 9 pins). It also uses an RJ45 [197].

There are other serial communication interfaces and modes that can transmit at higher speeds than RS 232. These include RS 422 and RS 485. RS 422 is recommended for longer distances and higher baud rates while RS 485 is recommended for multi-point communication [197].

Table 2.1 [197] gives a summary of the characteristics of various serial interfaces and modes of communication.

Table 2. 1: Properties of Various Serial Communication Types

Specifications	RS232	RS423	RS422	RS485
Mode of Operation	Single-Ended	Single-Ended	Differential	Differential
Allowed no. of Tx and Rx	1 Tx, 1 Rx	1 Tx, 10 Rx	1 Tx, 10 Rx	32 Tx, 32 Rx
Maximum cable length	50 Feet	4000 Feet	4000 Feet	4000 Feet
Maximum data rate	20 kbps	100 kbps / 10 mbps	100 kbps / 10 mbps	100 kbps / 10 mbps
Minimum driver output range	±5V to ±15V	±3.6V	±2V	±1.5V
Maximum driver output range	±25V	±6V	±6V	±6V
Tx load impedance (Ohms)	3k to 7k	>=450	100	54
Rx input sensitivity	±3V	±200mV	±200mV	±200mV
Rx input voltage range	±15V	±12V	±7V	-7V to +12V
Maximum Rx input resistance (Ohms)	3k to 7k	4k min	4k min	>=12k

The table shows that RS232, for example, is single ended meaning it only has one device at its end and hence only has 1 transmit and 1 receive terminals . Whereas RS485 has 32 transmit and 32 receive terminals and therefore can accommodate more than one device along its bus. RS232 only covers a distance of 50 feet (equivalent to about 15meters) while RS485 can cover upto 4000feet (equivalent to about 1.2km).

The solution built and tested used serial communication as the fundamental design approach to ensure one-way-data flow. This was to ensure that the offline system remained unreachable from the internet side of the online sub-system.

2.12 Offline Data Minimization

Data can be stored in various ways. The main two ways in which data can be stored from the internet access perspective are Offline and Data online Storage. Offline data storage entails storing data in such a manner that it cannot be accessed from the internet.

This can also be extended to local area networks where data is not available on the network. On the other hand, Online data storage entails making data accessible from the internet. This data can be in the cloud, on-premise, or indeed any other form of storage. Data minimization addresses three key aspects; collection of personal data, sharing of personal data, and the period the personal data is kept. Data minimization means the amount of personal data to be collected should be minimised, the sharing of personal data should also be minimized as well as the period for holding personal data for other persons must be minimized [230]. This implies that most e-commerce sites that hold personal data to grant users access are mostly violating some of the aspects of data minimizing. For example, they must collect as much data about the person they are granting access to their platform as possible, and they must hold the data for as long as they can for the user to continue accessing their platforms.

Data minimization, from the perspective of sharing as little information as possible, helps to protect user privacy by breaking linkages between minimized data and the actual data owner [230]. In other words, data minimization preserves privacy by suppressing user identity [231]. Data minimization can be combined with pseudonymization. This is where data is transformed into information that does not directly represent the original owner. In addition to providing privacy, data minimization can also help reduce the amount of data that needs to be transmitted via limited bandwidth. Reduce data elements in huge data sets can result in reduced aggregated data sizes [232].

2.13 Chapter Summary

Reports of data leakage, which usually lead to user privacy violations if PII is involved, keep emerging almost daily. Various solutions and proposals have been made to address the challenge of data leakage and user privacy violations, yet these problems persist. This chapter reviewed existing literature on various solutions that have been proposed to address the challenge and discussed some of the gaps that remain. This section will give a summary of some of the existing solutions discussed.

One proposed solution to address privacy challenges is the use of the Trusted Third-Party model where a TTP issues static IDs to users. Service providers would then get in touch with the ID issuer to verify whether the ID is valid without the user submitting their PII to the service providers. The PII is stored using block chain technology. Block chain technology faces some security and efficiency challenges. For one transaction to be confirmed, several nodes, usually a minimum of three, need to confirm hence making the solution very solution and resource intensive which would be difficult to implement world-wide. Further, static IDs can be profiled and thereby have user privacy compromised.

A Pseudo online Static ID was proposed to keep the user unknown. This addresses the challenge of the user's identity being known. However, profiling the user would reveal to the service provider lifestyle, preferences, behaviours, and other personal attributes of the user and might end up having their privacy violated. There is a need to protect the user even from online profiling by service providers and other attackers.

Random electronic IDs were also proposed as a solution to privacy violations via user profiling. This is where a different ID is generated for every transaction a user conducts online. The approach protects the privacy of the user but might be a challenge in terms of keeping logs of each random ID generated in case the user misbehaves online and they need to be prosecuted or reprimanded for their fraudulent online activity. There is a need to find a way of using random IDs without the need for keeping voluminous logs for each user and their several random IDs.

Privacy Enhancing Technologies (PET) are becoming popular for protecting the privacy of the user. PETs hide identifying information such as the identity, location, and IP address of the user and their device. They also address the issue of profiling.

For one to be given access to an online platform, they usually submit their PII and have an account created on the online platform. Most platforms require a user to define their username and password to protect their account. This approach is vulnerable hence multifactor authentication was introduced. The approach maintains a static password and username but adds another factor where one can use a One Time Password to complement the traditional credentials. Using more than one factor enhances security. Most OTPs are delivered via SMSs. This media is vulnerable. To help address this challenge, QR codes and Apps such as Google authenticator are being used to securely generate codes for authentication.

To prevent data loss, Data Leakage Prevention tools have been developed. DLPs protect sensitive data by monitoring information that is being sent from the network and taking the necessary actions that have been defined. They can send alerts and even stop data from being externalized if the action is deemed a violation of the rules set to protect sensitive data. DLPS can be resource-intensive and might end up slowing the systems on which they have been installed. Further, some data can be hidden in images which might be difficult to inspect in real time and decipher hidden information.

With the proliferation of Internet technologies, most devices are being connected to the internet. The devices are commonly known as IoT devices. Most of these devices are controlled by microcontrollers. Microcontrollers have their security challenges which must be addressed because almost every –day-to-day device is now controlled by them. They are preferred due to their miniature size, low power consumption, and therefore easy-to-handle features. Due to the high demand which most manufactures started failing to meet, most manufacturers started outsourcing production. The approach has introduced certain risks such as the introduction of Hardware Trojans at the foundry and poor testing after manufacture. Sensitive data can be externalised by the device once installed and operation. This compromises the security of devices these microcontrollers are embedded in.

Artificial intelligence has brought about a lot of advancements in the use of technology for improving lives. However, some perpetrators have taken advantage of AI to increase the success of their attacks. There is a need to understand how AI is being used for cyber-attacks to use it to address the challenge. AI is already being used by certain cyber solutions such as Dark trace. Captcha is another example of AI.

Chapter Three

Research Methodology

3.0. Chapter Introduction

The ultimate objective of this research is to address the prevalent problem of PII leakages in e-commerce via the use of Offline data Minimization and Online Pseudonymisation. The methodology employed sought to enhance the protection of PII and preserve the privacy of individuals by applying various principles and approaches. This research emphasises the need for system security to be by design. The fundamental principles applied to design the system anchor on Pseudonymization and data minimization. That is, minimize, as much as possible, data exposed to the internet so that if a breach was to happen, no damage would be made to the privacy of the individuals associated with minimized data that would have been exposed.

Further, a Trusted Third Party (TTP) model is employed to ensure less aggregated individual data is spread across the cyber space by various online service providers.

The sections that will follow will outline the methodology employed to achieve enhanced PII protection while allowing users to continue accessing online services.

3.1. System Security by Design

Data protection and user privacy must be by design and must be a prerequisite for any system that will be processing data [198][199][200]. This principle applies to many other aspects of system security. Systems must be designed to be secure at the architectural level. This approach will enable system developers drastically reduce security vulnerabilities that are common with modern software. It is vulnerabilities that invite exploitation by hackers [201].

For this to be possible, engineers equipped with appropriate tools and methods must be involved. Privacy by design as a principle means that privacy and security must be incorporated into the design, development, testing, and implementation of various products, systems, and services. It should never be an afterthought [127]. Some of the laws that have been enacted in certain jurisdictions require that privacy is by the design of the systems [202]. The GDPR demands that various techniques are used to ensure

systems are built to provide privacy to users and protect data. This by implication means these requirements must be considered when designing a system if it is to be compliant with the law when processing personal data. Therefore, any system that processes personal data must be deliberately designed to protect that data and provide user privacy as provided by law [29].

Our research focused on the use of data stupefaction by employing pseudonymization. Several techniques can be employed to modify data to ensure user privacy. These include suppression, generalization, masking, replacement, swapping, and distortion or perturbation. Generalization preserves by replacing specific data items with general data. Masking involves hiding some data elements such as replacing them with asterisks, swapping is achieved by swapping real data elements to disassociate the real elements from the right owners. Distortion or perturbation is achieved via the addition of "noise" into the data; that is, adding data that is not real [15]. Which of these techniques is suitable depends on the application. The focus of this research was the use of the solution in e-commerce hence the decision to use distortion which allows reverting to the original if the need arises [58]. The equation [15] that follows explains how data is distorted and how the original data can be recovered.

$$Dd = Do + Dn$$

Equation 3. 1

Where **Dd** is the resulting distorted data

Do is the original data and

Dn is the noise data that has been added for distortion

To derive the original data, Do, the equation below applies

$$\boldsymbol{Do = Dd - Dn}$$

Equation 3. 2

Information and Communication Technology security must be considered when designing systems to protect data. It is the technical aspect of data privacy. It addresses the critical aspects of confidentiality, integrity, and availability [203][204][13]. Availability and protection of data must be balanced to offer user-friendliness as well

as ensure there is no loss of sensitive data [205][205]. Confidentiality is the keeping of secret data undisclosed and inaccessible by unauthorised individuals or systems while integrity is maintaining the accuracy and correctness of data. It must never be overwritten or modified by unauthorised agents. Availability is ensuring that data or systems are accessible when needed [176].

Any information system has different layers and any weakness in any of the layers can be exploited [13]. Figure 3.1 shows the various layers a computer system comprises.

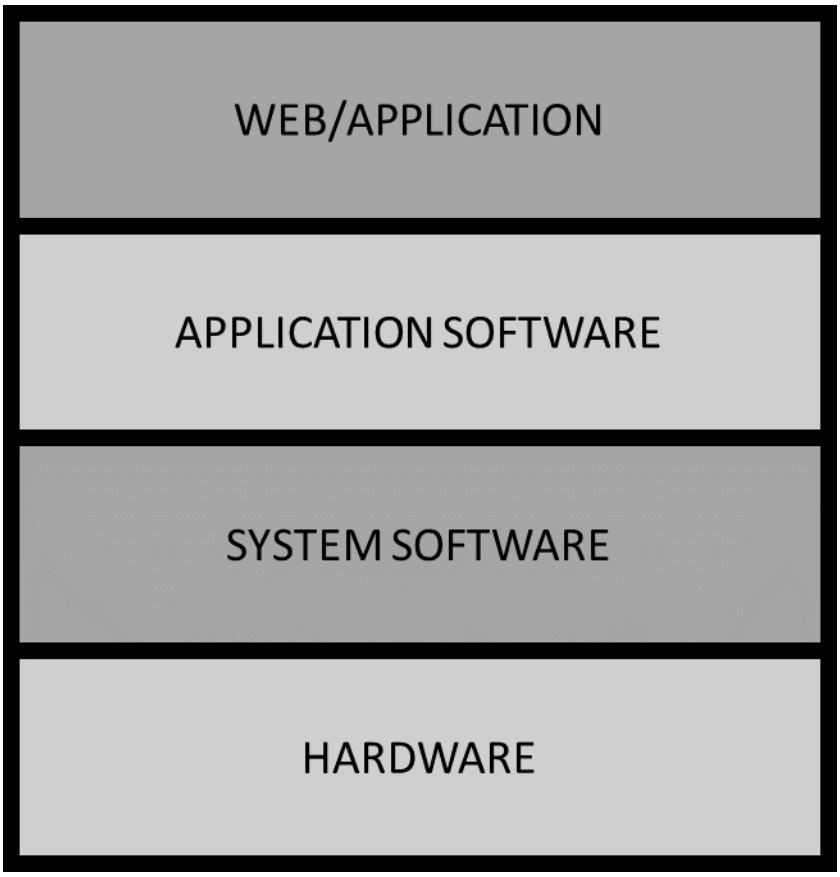


Figure 3. 1: Computer System Layers [13]

To keep the computer system secure, all layers must be secured. A security lapse in one layer can lead to the entire computer system being compromised. Over the years several security solutions have been developed and implemented in the web application, application software, and system software layers. However, hardware security has been the least attended to. Very often, software developers have put their trust in hardware manufacturers and architects to do their work as expected. However, recent attacks have shown that more needs to be done to secure the hardware and ensure that they produce

the correct results they are expected to render [13]. Hence the need for enhanced hardware security.

Privacy by Design entails that privacy must be at the centre of the system from requirements gathering, interpretation of the privacy laws, testing, and the implementation of the systems that will process data[206]. In short, privacy must be part of every stage in the development lifecycle of the systems to process personal data. Taking into account the existing as well as planned laws and regulations will ensure the implementers do not face a challenge with the law when deploying the solution into real life [207][102].

Privacy and Data Protection by Design ensures that systems are privacy friendly and therefore enable users to find them easy to use while protecting their privacy. Privacy should never be an add-on[206][208]; it must always be taken into account from the concept stage till implementation to achieve user friendliness without compromising privacy. One of the benefits of the Privacy by Design approach is that the need for labor-intensive and time-consuming patches will be reduced significantly [209].

To be effective, Privacy by Design requires interdisciplinary expertise; there is a need for legal expertise to ensure applicable laws are adhered to, software engineers are required to provide expertise on necessary techniques to comply with the legal provisions as well as user requirements, developers, system operators, and many other skills depending on the project under consideration [210].

Our design followed critical principles that govern the protection of data; Data classification, identification of exit points for the data, communication channels to be used, and devices to be used to carry or process the data as well as the application systems. This holistic approach helps ensure all vulnerable exit/entry points are sealed by design from the very start of the development process [211]. Our solution was designed to only have one entry/exit point into the storage location where sensitive data is kept and the data protector developed was placed at the exit point to maximize its effectiveness.

Another critical principle of Privacy by Design is ensuring that there is lawful processing of data. The system must be designed to make it easier to comply with applicable laws such as the GDPR. In addition, there must be a specific purpose for

which the data is being collected and limitations must be applied to ensure only appropriate data is accurately collected. Further, the collected data must be retained only for the period required for its use and transparency must be seen to be applied during the processing of that data. Data controllers and processors must be held accountable for what happens to the data they collect and process [212][213].

Further, it has been observed that most perpetrators conduct their attacks remotely and then exfiltration data from the corporate servers into their servers where they analyse the data for ransom or other activities like selling the data on the dark web [187]. In addition, it was also observed that hackers use multiple channels to transmit huge chunks of data to their secured servers. To address the highlighted risks, the solution to protect data must be designed to prevent huge volumes of data from flowing from the central data storage to external parties. It must be further designed to physically only have one channel through which data can flow from internal to external hosts.

Moreover, the isolation of systems interacting with different worlds is a very critical design principle. According to this principle, a sensitive environment must not easily interact with a normal environment such as the internet [176]. This principle further explains that hardware at two different security levels must not communicate with each other directly. Sensitive data must, therefore, be stored in a location isolated by a security-sensitive buffer from a vulnerable environment like the internet where all sorts of disingenuous crooks are found roaming.

It must be borne in mind that there is no such thing as absolute security. If a security intervention delays a hacker from having unauthorised access to an extent that by the time they gain access, their access would be of no value then security has carried the day [214]. If, for example, it takes 100 years for a hacker to illegally download confidential data, then by the time the exercise is complete, the hacker is most likely not to be alive to exploit the downloaded data. The systems would have been upgraded to suite the trends and hence the download would have been interrupted.

3.2. Trusted Party Model Approach

The research chose to employ the TTP approach model for the design and implementation of the PII protection system. Literature review revealed that there are several data breaches taking place every single day that passes and most of the data

leaked is for individuals, hence compromising their privacy. The current approach in Ecommerce is where users must submit their PII to each platform they would like to access before they could be granted access to the platform. This results in several aggregated data for individuals thereby creating a high possibility of data leakage due to too many points of data leakage being made available to the internet via the various platforms users would subscribe to. World-renowned companies have suffered data breaches hence it is not only about having the most expensive state-of-the-art security solutions, but it is also ensuring the right and most effective approach is used. This is where the TTP model comes in.

The TTP model entails having one Trusted Third Party carry out KYC activities on behalf of all users and companies and then providing confirmation to the parties that need to know about another party before a transaction such as on boarding a user onto the online platform for crypto trading can be executed. The TTP conducts the necessary KYC due diligence and registers the users on their KYC system. Each registered user is provided with minimal credentials that would enable them to access online platforms without revealing their PII. Then other providers who need KYC confirmation before granting users access can confirm with the TTP whether the user requesting access is genuine or not based on the random ID the user provides. This approach reduces several aggregated isolated PII being littered across cyberspace thereby reducing the possibility of data leakage.

The TTP model is further enhanced by utilising data minimization via pseudonymization. User data is minimized in such a way that only very minimal data is kept on the online system (API) that other service provider systems talk to for user KYC confirmation. The system does not contain identifying data for users hence even if the online system was compromised and experienced a data breach, the users' PII would still be secured on the 'offline 'system.

The TTP model has been applied in various applications; IoT pairing being one of them. This is where IoT devices that need pairing and trusting each other before data can be exchanged get pre-screened in advance by another already Trusted System. When IoT devices need to pair up, they ask the Trusted System whether the device requesting to pair with it can be trusted. The response of the trusted system determines whether or not the pairing will be successful [106].

The TTP model has also been used before to provide an online KYC system where the TTP was providing users with static electronic IDs that service providers would use to confirm with the TPP whether or not they issued the ID before granting the users access [22]. The TTP model has been used for guaranteeing the quality of service for outsourced cloud services. A Third-party system collects data on various service providers and gives potential services data on the quality of service the various providers have been offering so far. This was to help improve the adoption of outsourced cloud services by users or companies [105].

The TTP must also be trustworthy. There have been some TTP who have not been trustworthy. To address the issue of the third party being untrustworthy, the study proposed the use of a KYC Privacy Agency (KPA) appointed by the national authority in each country to use the ODMS. In a Trusted Model approach, the users must have trust in the appointed third party, or else it won't work[54]. The Figure below depicts the concept of using a KPA.

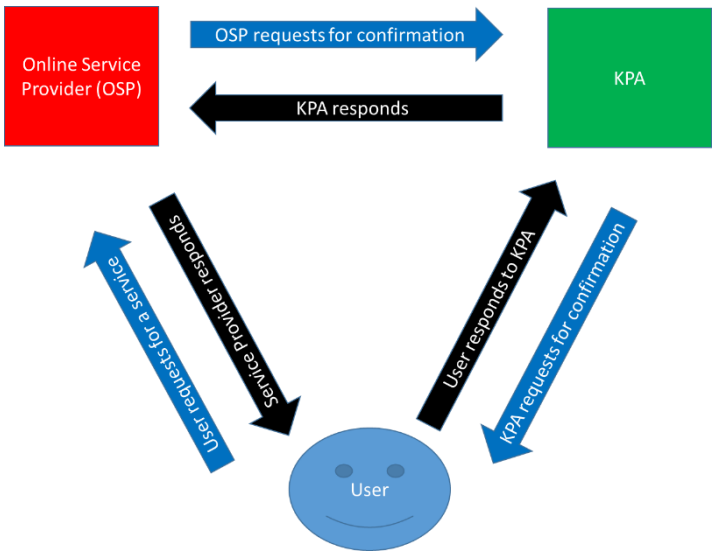


Figure 3. 2: Use of KPA as a Trusted Third Party

Our research adopted the TTP model with modifications to the various deployments that have been implemented before to enhance the protection of PII as well as ensure the preservation of privacy for the system users. The model employed ensured critical PII is kept 'offline' to eliminate the possibility of being leaked or breached by online hackers. It uses random IDs for users to preserve their privacy while online. Random IDs help prevent online profiling of static IDs.

To ensure data at rest stored on the databases is protected from those who would access systems from the backend, data is encrypted. This prevents back-office staff like ICT personnel from accessing meaningful data without audit trails.

To keep the critical PII 'offline', a buffer is created between the offline and the online system that only allows data to flow in one direction. That is, data flows from the offline system towards the online system. Data cannot flow from the online system towards the offline system to prevent Trojan viruses and hackers from being able to access it and compromise PII.

To preserve the privacy of the user, the ID given is random. To arrive at the random ID, a pseudo code for the user is used to modify the standard TOTP to arrive at the final random Id to be used for transacting. This approach enables the retracing of the anonymous user if the need arose while at the same time enabling the user to avoid being profile online if they were to use a static user ID. It further gets rid of the need to store huge volumes of audit trail data for Random IDs to be able to trace back to the user if they misbehaved while online. The used Random ID can be used to decipher the Pseudo ID used which can in turn be used to decipher the actual identity of the user by the TTP.

3.3. Mathematical Model to be employed

The Study adopted to use the Vector Space model (VSM) as it is commonly used for Information-retrieval modelling [215]. The model was used to demonstrate that mathematically that the proposed approach helps to prevent data leakage. In this model documents, and by implication, data are assumed to be part of an *n* dimensional vector space where n is the number of indexed terms (That is, the sensitive data like PII). Adrian came up with a Data Leakage system using the VSM. This research used the same system model to demonstrate that the approach used by this project was able to protect PII more effectively. The model gets web documents suspected to have been leaked and compares them with user-defined documents by the user defining the keywords. Figure 3.3 shows how the system works.

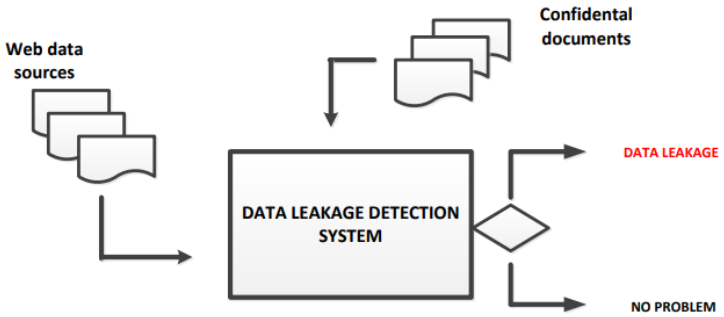


Figure 3. 3: Data Leakage Detection System

The Model determines the chances of data leakage by computing the similarity index of the defined terms and the terms found in the web document. A Threshold can be set. If the returned index is equal to or above the set threshold, then it is concluded that indeed the document was leaked. The equation [215] representing the process is shown below:

$$S_H(W_i, W_j) = \begin{cases} 1, \ if \ W_i = W_j \\ 0, \ otherwise \end{cases}$$

Equation 3. 3

Where S_H is the similarity, W_i and W_j are two keywords one user-defined and the other extracted from the web document.

The index terms will usually be several and can be called set T of index terms where $T = \{t_1 \ldots\ldots\ldots t_i \ldots\ldots\ldots t_n\}$ which is user-defined. The Web document W_j is given a vector v_j of finite real numbers as shown below:

$$v_j=(w_{ij})_{i=1,,n} = (w_{1j},\ldots\ldots\ldots w_{ij}\ldots\ldots\ldots,w_{nj})$$

Equation 3. 4

w_{ij} is the degree to which the index term embodies the web document.

Confidential User documents are represented as follows:

$$v_k=(w_{ik})i=1,\ldots,n = (w_{1k},\ldots\ldots\ldots w_{ik}\ldots\ldots\ldots,w_{nk})$$

Equation 3. 5

Therefore, the similarity check (leakage detection) between the two documents and their associated keywords will be

$$S_{jk} = s\ (v_j, v_k) > K$$

Equation 3. 6

Where K is the set similarity threshold to trigger an alert.

Using the Cayley-Klein Hyperbolic Geometry, the similarity measure is derived from the hyperbolic distance computed as follows.

$$S_{j,k} = \sigma_{j,k} = \left(ln\left(e \cdot \frac{r+\sqrt{\sum_{i=1}^{n}(w_{ij}-w_{ik})^2}}{r-\sqrt{\sum_{i=1}^{n}(w_{ij}-w_{ik})^2}} \right) \right)^{-1}$$

Equation 3. 7

Where

$$r > max_{vj} d_e(v_j v_k)$$

Equation 3. 8

$$\boldsymbol{d_e(v_j v_k)} = \sqrt{\sum_{\boldsymbol{i=1}}^{\boldsymbol{n}} (\boldsymbol{W_{ij} - W_{ik}})^{\ \boldsymbol{2}}}$$

Equation 3. 9

The assumption in this model is that documents and sensitive data are kept online in the cloud. The system detects which of the documents have been leaked out onto the web space. From the equation above the more the index of terms appears on the web the higher the similarity indicating that the document got leaked. In our proposed model, where critical data is kept offline and online Pseudonymized data is kept online, using the above model equation entails:

$$v_j = (w_{ij})_{i=1,\ldots,n} = (w_{1j}, \ldots\ldots w_{ij} \ldots\ldots\ldots, w_{nj}) \approx 0$$

will be almost zero as the PII is kept offline and hence will not find itself online.

It must be noted that the model is covering data leaked online and not via other physical means such as staff copying data from the offline system and depositing it online in cyber space.

3.4. Chapter Summary

The use of the TTP concept helps to ensure user PII is not spread across the internet which increases the possibility of the data being leaked. Different service providers have varying security measures in place to prevent data leakage. Having to provide data to every service provider to access their online platform entails having several aggregated data for individuals across cyberspace.

In addition to the concept of TTP, the solution used the concept of data minimization and pseudonymization to further enhance the protection of privacy and data leakage prevention. Data minimization reduces the amount of data to be kept online. A combination with pseudonymization further limits the damage to the privacy of victims if related data was leaked online. For example, when pseudonymized and minimized, data that can be online would comprise non-identifying elements such as Pseudo IDs that would not give away the actual IDs of data owners.

The whole solution design was premised on Security by Design. That is, the whole solution was based on ensuring PII and privacy is protected from the onset and not as an afterthought after the system is implemented. The security layer was part of the concept and not an extra layer added to the system to make it more secure.

Chapter Four

Detailed Design: Personal Data Protection Model

4.0 Chapter Introduction

This chapter gives a detailed description of the design of the solution to resolve the leakage of PII for online users. It explains why the proposed solution is effective and how it operates to protect PII. The model was designed based on the gaps identified from the real world and confirmed via literature review. The solution can be extended and adapted to other applications that process PII. It uses the enhanced TTP model to make the solution more effective and convenient for users. It helps protect user privacy while in cyber space by using Random IDs.

4.1. Research Methodology

The entire design is premised on keeping as much sensitive data as possible offline. This is achieved via a process called Data minimization. Only necessary Pseudonymized data is kept online for real-time transaction processing. Figure 4.1 below gives an overview of the process.

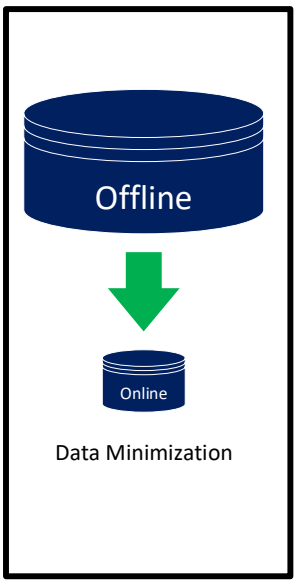


Figure 4. 1: Data Minimization Process

The Design follows the standard software development cycle using the agile approach where each component is designed, built, and tested separately before the entire integrated solution is tested for requirement conformity. Figure 4.2 below gives a summary of the stages involved in the software development cycle.

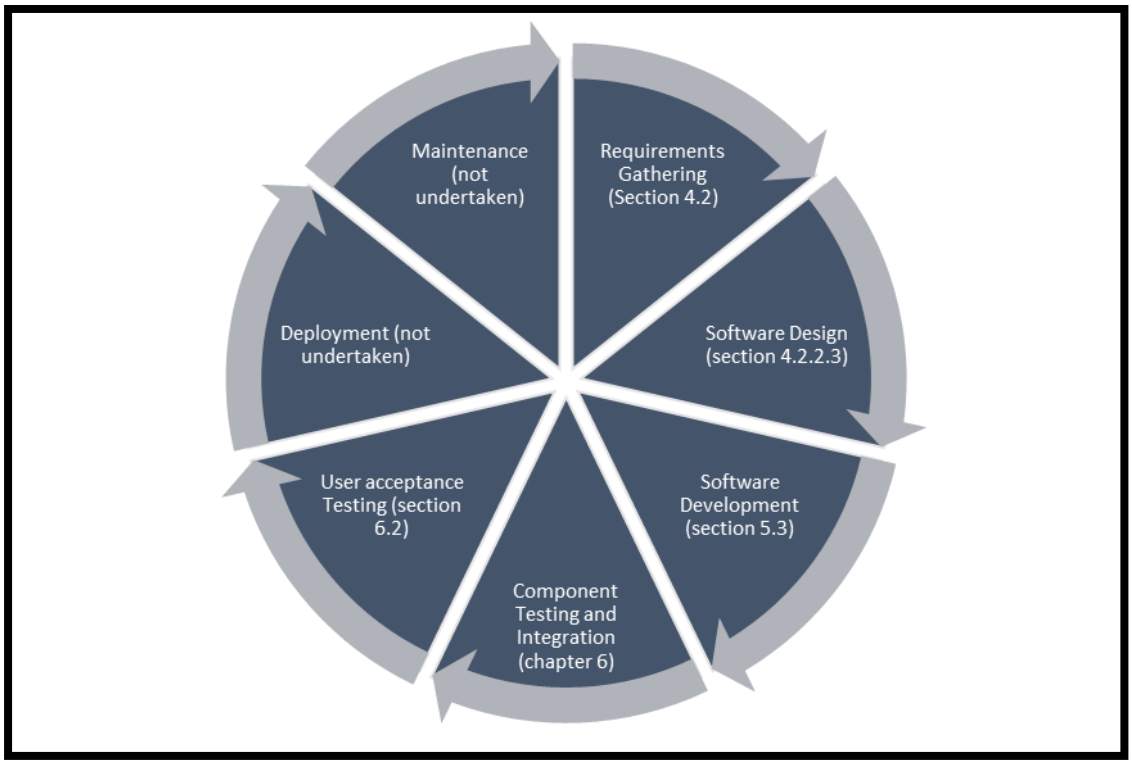


Figure 4. 2: Software Development Cycle for the KYC Prototype

The first step is gathering the business requirements of what is expected of the system to address. This includes coming up with all the necessary business processes and user stories. Once the requirements are understood, the process moves onto the design of the system aimed at meeting the requirements outlined. It was during this stage that requirements for the prototype were drafted including the user stories and process flows. The output of this stage is detailed in sections 4.1 to 4.2.2.

Once the design has been confirmed and signed-off by the system owners/users, the development of the desired software commences. At this stage, the actual solution is being built to address the outlined challenge. Various approaches are used under this stage one of them being prototyping. This is where a scaled-down system is built to help implement the user requirements gathered and allow users to provide feedback before a full-blown solution is developed. In this research, the approach was to first build a simulated prototype that only focused on the Data Protector Module. The

module was tested for technical capabilities using the simulation software called Proteus. Thereafter, the actual prototype was developed as outlined in chapter five.

Depending on how the system is being developed, the system can be divided into different components at the design stage. Development will involve the building of different components and these components will need to be tested individually before they can be integrated with other components for end-to-end testing of the entire system. This is normally done by the development team before releasing the completed system to users for testing. Users conduct User Acceptance Testing (UAT) where they determine whether the system meets their requirements. It is at this stage that the users give feedback on what needs to be corrected, improved, and so on. The successful completion of this stage leads to the deployment stage where the solution is deployed into production for use by everyone who needs the solution. The user testing component was conducted, and results were compared against the requirements earlier collated. The results are presented in chapter six. Successful deployment into production leads to system maintenance and support to ensure the system remains operational and continues to provide the quality of services desired. For this research, the process ended with user acceptance testing. Deployment would only happen on a life scale which requires further collaboration for the adoption of the solution country wide.

The research chose to divide the development of the key modules into components. That is, there was Data Protector Module, Offline Module, Online Module, and a Random ID generator module. The components were built separately and tested as distinct components before being integrated into one complete system. The Data protector module integrated with the Offline and online systems to create a complete KYC system.

The approach for this design was premised on the Security-by-Design principle. That is, the system was designed to secure PII from the onset. Security-by-Design entails the security of the key input (PII data) was of primary concern when coming up with requirements and the approach to use to build the solution. It was not an after-thought by use of Third-Party security solutions.

The methodology followed to design the system was closely looking at existing infrastructure and systems to ensure that the resulting product is a huge improvement

on the current solutions yet without the need to overhaul the existing infrastructure. For example, Random Electronic ID was designed based on the existing standard RFC 6238 with some modifications to convert the OTP into a random electronic ID that can be used to trace the user when they are anonymous. This way, the resulting system does not require a huge change to the existing systems to enhance security. It rather builds on the existing solutions to achieve more effective security of PII.

To secure PII, the design was to ensure that the PII is not easily accessible from the internet. That is keep sensitive data 'offline'. This meant that the system would comprise three key components; the offline system to hold PII, the online system to offer KYC services to online service providers who need the service, and middleware to separate the offline system and online system as well as to determine the flow of data and how much data can be transmitted at any given time.

To ensure there is a fundamental physical separation between the offline and online systems, serial communication was used to design the solution. Serial communication has various ways of communication as discussed already under the literature review section. The method chosen ensures that data can only flow in one direction. Microcontrollers were used to help achieve the one-way-dataflow desired as well as determine the bandwidth that could be used for serial communication between the offline and online systems.

The design of the solution followed the software development cycle that commenced with requirements. The sections that follow will outline how the design was put together.

4.1 High Level System Requirements

If information is available online, the possibility of someone accessing that information without authorization remains. The best way to protect information is to make it unavailable to online hackers hence the proposed design in Figure 4.3 below:

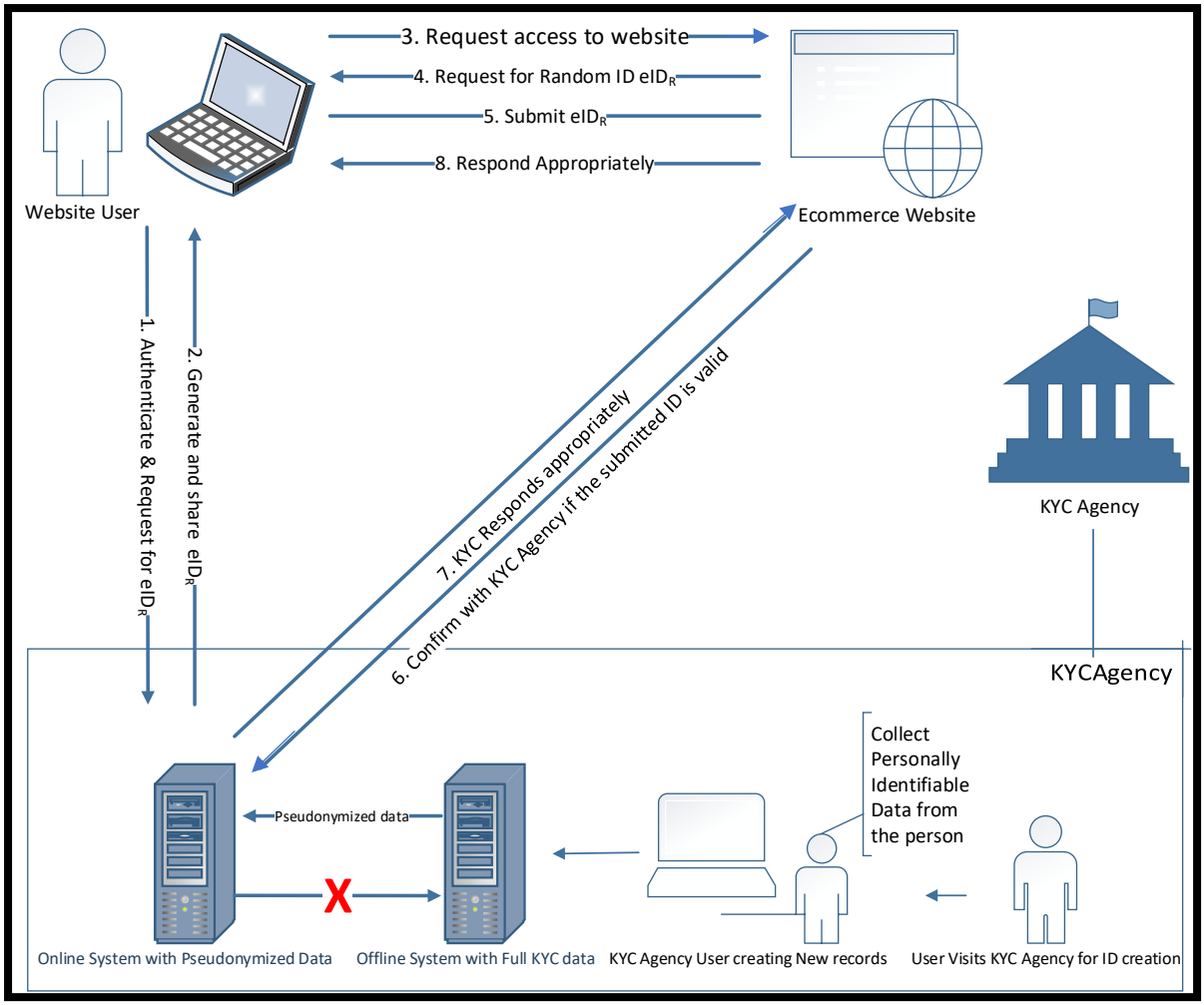


Figure 4. 3: Know Your Customer Agency Operation

The approach proposes the use of a Trusted Third Party, herein called KYC Agency, which would be a government-appointed entity tasked with the registration of its citizens to issue National IDs. In Zambia, there is a department under the Ministry of Home Affairs that is charged with the issuance of National Identity cards and passports. This would be the appropriate department to host and operate the system so that electronic IDs are generated when one is being issued with a national ID. These two must be linked. Persons under the age required to be eligible to obtain National IDs can still be registered to be issued with electronic IDs for online transactions but can be issued later with the National ID card when they are of age. This is to ensure that the protection of PII is initiated from a very young age to avoid having PII littered across cyberspace hosted by various online platforms.

The model demands that other service providers requiring KYC verification before granting a service to a user confirm with the KYC Agency if the requesting party is

genuine and the KYC agency provides assurance without sharing the PII of the requesting party. This enables the requesting party to have access to services without risking their PII and privacy by sharing it with several online platforms. The more spread the aggregated data is, the higher the chances of it getting leaked as different platforms apply different data security measures.

The TTP is proposed rather than each e-commerce organization using pseudonymisation on offline data to minimize the number of times the users will have to submit their PII. In addition, despite the e-commerce providers claiming to be using secure systems, chances are that some of them might compromise their security hence risking exposing the PII of the users. Hence, the higher the number of places the PII is held, the higher the chances of compromise. A Government agency with the support of the government is expected to provide maximum data security as proposed in this research hence minimising the likelihood of the PII being leaked.

The approach will operate as follows:

4.1.1 User Registration

The user will first register with the KYC Agency in their country of residence. They will submit Personally Identifiable Information (PII) such as their National Identification documents, Residential address, and contact details such as phone numbers and email addresses and have their image captured.

It is recommended that the KYC Agency be the same institution that issues citizens with their National Identification Documents such as passports. This will ensure that whenever a citizen is issued with an ID even for the first time, they are issued with one that can also identify them electronically.

Once the user has satisfied the requirements for registration with the KYC Agency, the KYC Agency creates a record with full Identifying Information of the user and appends a universally unique ID on the record. The data is kept on the “Offline” system that is not accessible from the internet.

Figure 4.4 gives a detailed flow of the registration process.

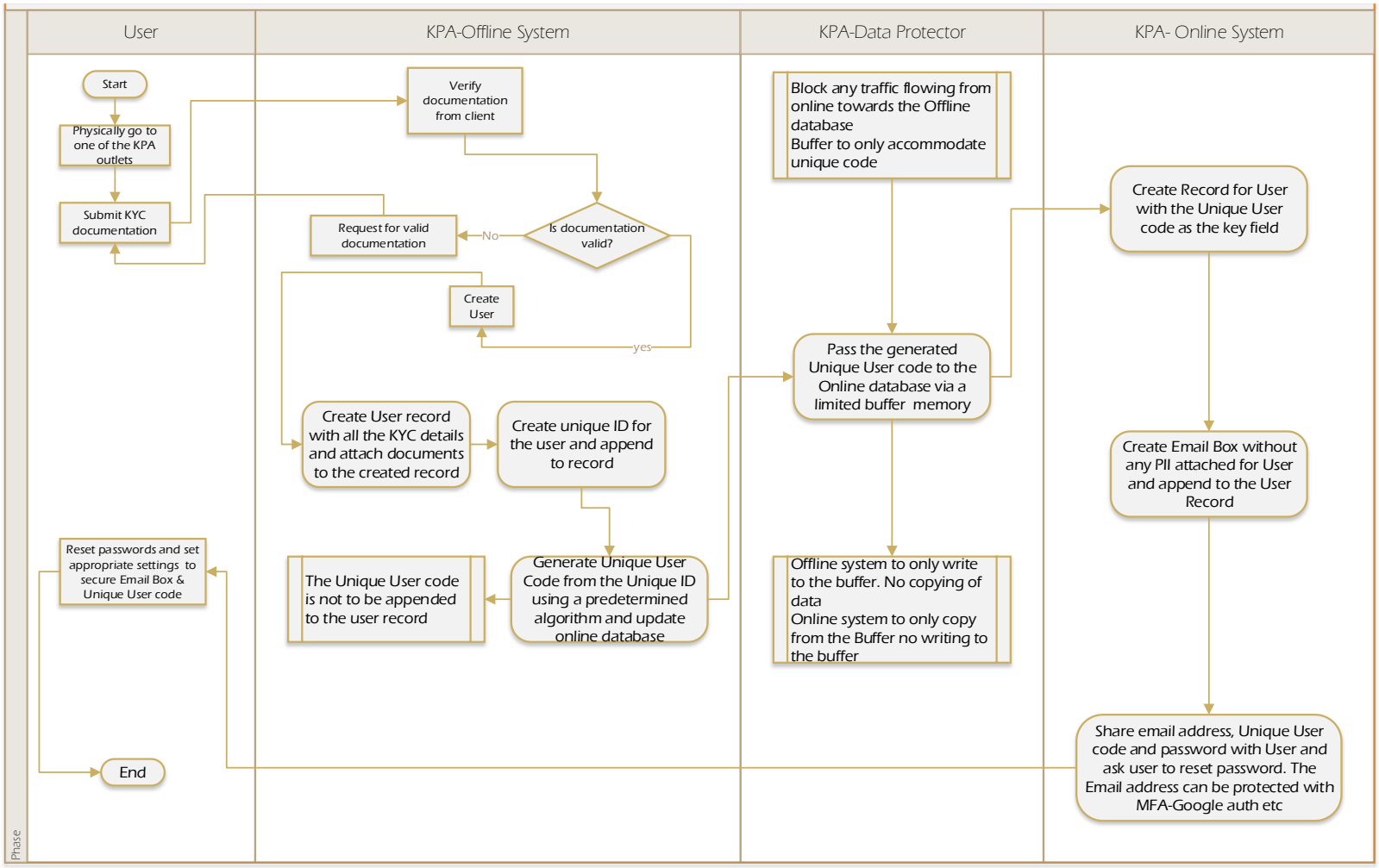


Figure 4. 4: User Registration with KPA

After the eID has been appended to the new user record, the eID is pseudonymized (eIDs) using a predetermined algorithm and sent to the online system for the creation of an online record for the user. The pseudo version of the unique ID, eIDs, is not appended to the record sitting on the offline system. This is to minimise the possibility of associating offline data to online pseudo-IDs if they are leaked for some reason by insiders.

It must be noted that the only communication between the Offline system and the only system is the automatic transmission of the Pseudo ID, eIDs, to the online system. The transmission is determined by the firmware sitting on the microcontrollers as will be explained under the operations of the Data protector system.

When the online system receives the pseudo-ID, eIDs, it automatically creates a record with the eIDs as the primary key. It then creates an anonymous email box for the user. The mailbox is to be used for delivering Random eIDr to the user. In addition, the emails can be automatically destroyed after a predetermined period to prevent the formulation of the Key/algorithm being used for the generation of random IDs in case the online system with mailboxes is compromised. A chain of emails with various random IDs might be used to crack the key and algorithm used to generate the random IDs.

4.1.2 Proposed Universally Unique electronic ID (eID)

The following format of the universally unique ID is being proposed:

The eID comprises 10 digits representing the unique ID for the person being created on the system and 3 digits representing the country the person is a citizen of as shown below:

X X X . X X X X X X X X X X

Figure 4. 5: International ID Format

The 10 digits for the universally unique portion of the ID is to accommodate for the growth of population for counties like China and India. The 3-digit prefix for the country is to accommodate the number of existing countries in the world and the new countries that might image. The first Citizen to be registered for example would have the eID shown in Figure 4.6 below:

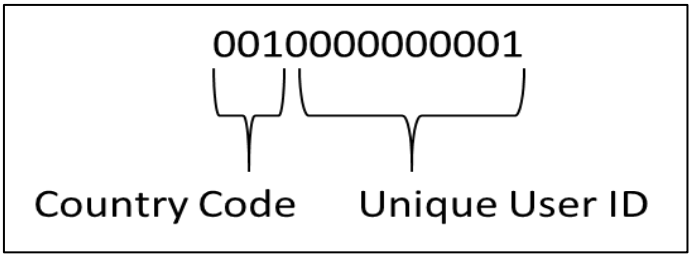


Figure 4. 6: International ID Example

4.1.3 Data Protector Operations (RMS)

The system consists of three subsystems connected logically. That is, the Data Minimisation System (DMS), RMS, and the Pseudonymisation System (PS) as shown in Figure 4.7. The Data protector is the key component of the solution. It is built using hardware to make it possible to physically block the flow of data from the online system towards the offline system hence preventing hacking of PII by internet users.

The Data protector that safeguards Personally Identifiable Information connects the offline system to the online system and operates as follows:

Data exchange between the two systems only flows in one direction; that is, it only flows from the offline system towards the online system as depicted in Figure 4.1.5. This restriction aims to ensure that no one can access the PII from the Internet. This is to reduce the possibility of a hacker accessing the PII without needing physical access to the server hosting the sensitive data.

Figure 4.7 shows the summary of how the Data protector (Restricted Memory System) will be operating:

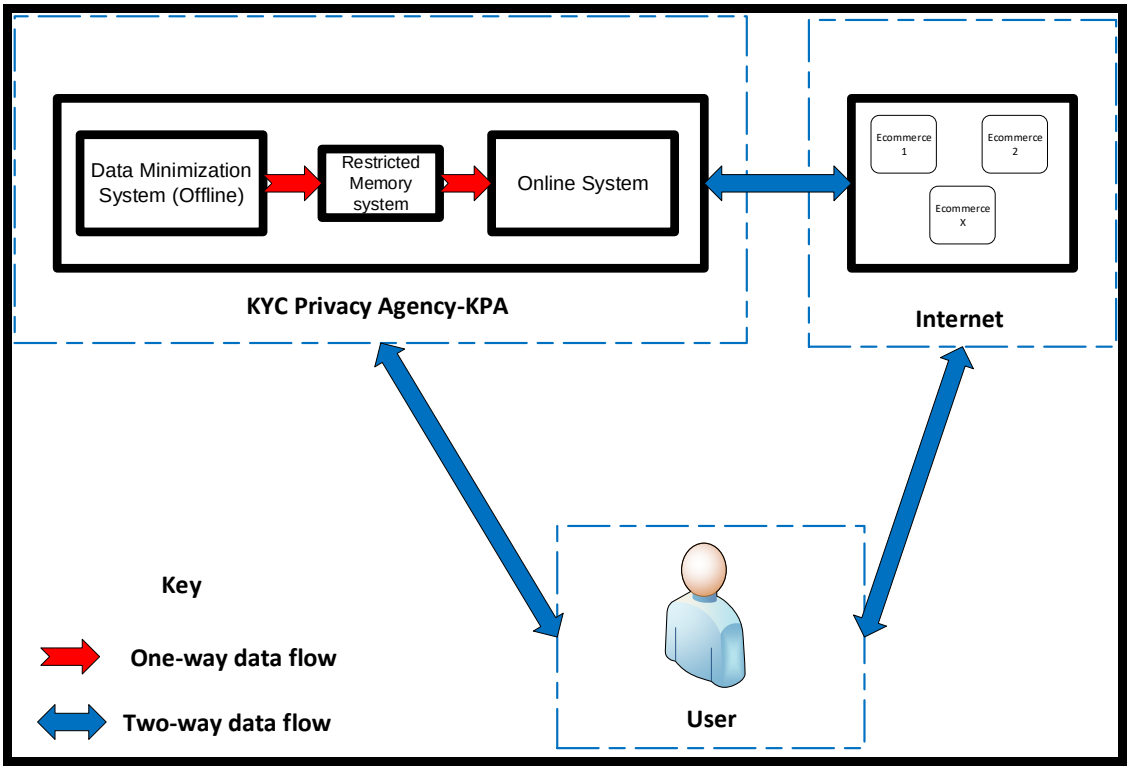


Figure 4. 7: Data Protector Operations

Furthermore, despite data being able to flow towards the online system from the offline system, to prevent huge amounts of data from being sent by disgruntled elements inside the KYC Agency using the offline system, there is a bandwidth restriction imposed between the two systems. We propose using the lowest possible serial data speed available. For example, if we wanted to send 10 gigabytes of data from the offline system to the system via a serial connection of 9600bps, it would take more than 100 days to complete the transfer of data. Slower speeds would take even longer. However, transferring pseudo-IDs would take a few milliseconds as the strings would only constitute a few kilobytes of data per unique record created at a given time. The slow rate of data transfer would discourage a hacker or disgruntled element from attempting to do so if they somehow managed to attempt, they would be discouraged by the estimated time of data transfer and hence abandon the theft.

Moreover, the system would periodically reset the connection between the two systems hence disrupting any exploitive data transfer in session as legal sessions will be expected to only last a few milliseconds.

Furthermore, data sitting on the offline databases is encrypted or hashed to prevent someone physically accessing the databases from being able to read the data and -make sense of it.

4.1.4 User Transaction with E-commerce Sites

The user will either access the KYC Agency to generate a universally unique random ID, eIDr, or first access an e-commerce site to request to transact. The site will request the user to submit their random ID issued by the KYC Agency. The sites will not be allowed to collect PII from users to prevent data leakage prevalent with online services.

The user will need to Logon into the KYC Agency system via a website or app and request a unique random ID. The KYC will authenticate the users via existing identification methods such as Google authenticator or any other multifactor authentication method. The user remains anonymous using the records kept by the online system. Once authenticated, the user generates Random ID eIDr. The KYC system sends the ID, eIDr, to an Anonymous email or is displayed on an App. Then the user enters the eIDr on the website. The website verifies with KYC Agency if they issued the eIDr supplied by the user. Depending on the feedback of the Agency, the website either grants or denies the user access to their services. Figure 4.8 gives a pictorial view of how the transaction flows from the beginning to the end.

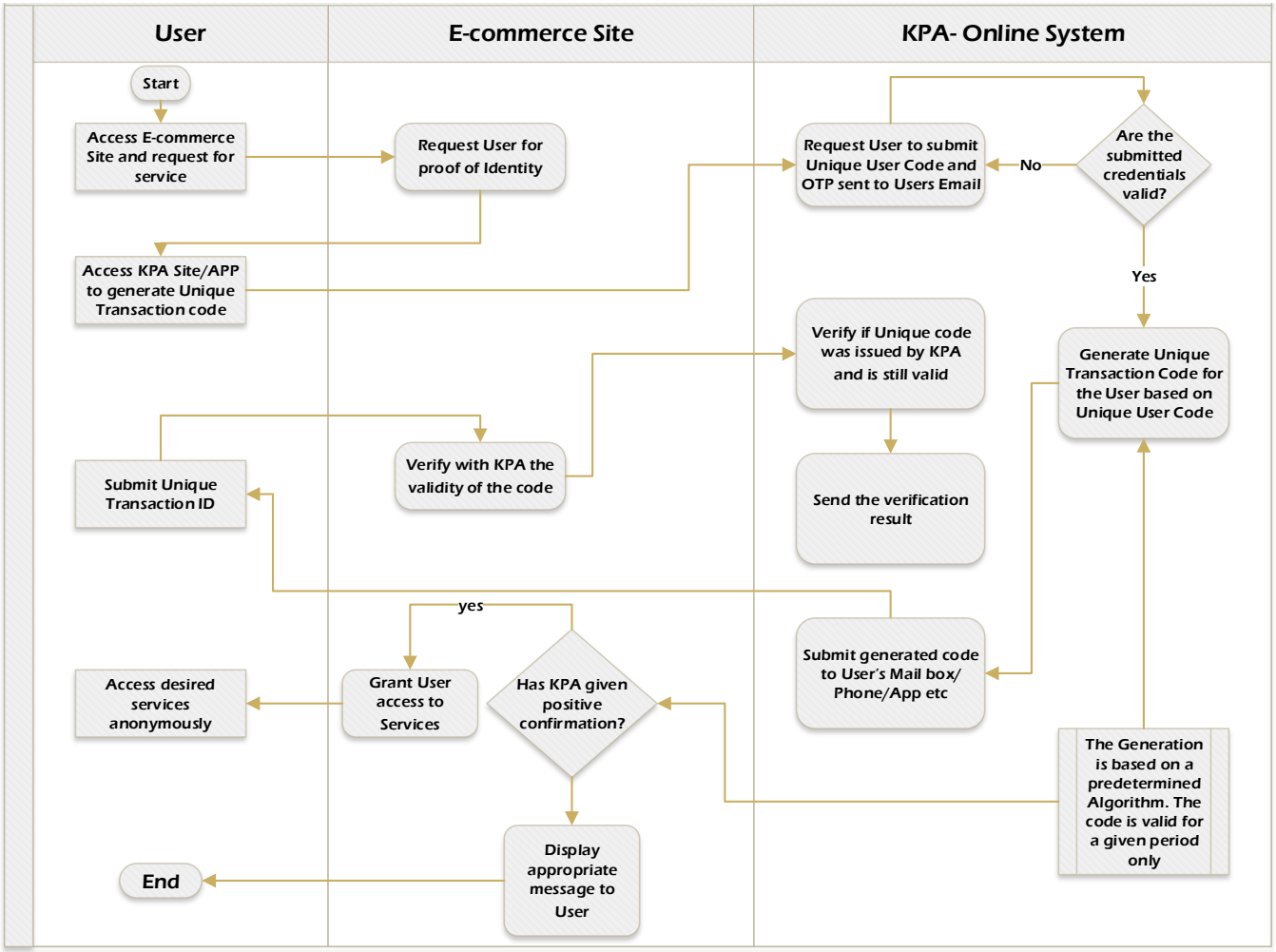


Figure 4. 8: Anonymous End-to-end E-commerce Transaction

The KYC Agency system will host the mailboxes for the users and will destroy emails containing random IDs after the predetermined validity period elapses.

There is a need to ensure that only pseudonymized data is made online while raw identifying data is kept unreachable from the internet. Furthermore, necessary internal controls must also be put in place to ensure data is not leaked by internal parties. Controls such as ensuring that the hardware system hosting PII is not fitted with external media devices such as writeable DVD drives, USB drives, Bluetooth, wireless, and so on. External tape can be connected for mass backup. The contents on the tape must be encrypted. The data on the databases must be encrypted or hashed to avoid ICT staff accessing the database directly and deciphering the actual data sitting on the servers.

Pseudo ID Generation

The Pseudo ID is generated via the concatenation of the unique Fingerprint code generated from the captured Fingerprint for the user and the unique universal ID for that user. It is critical to use the fingerprint code as it is universally unique for everyone. This will help enhance security and privacy for the user as well as ensure there is no duplication of unique Pseudo IDs for whatever reason. Figure 4.9 below outlines how a Pseudo ID is generated.

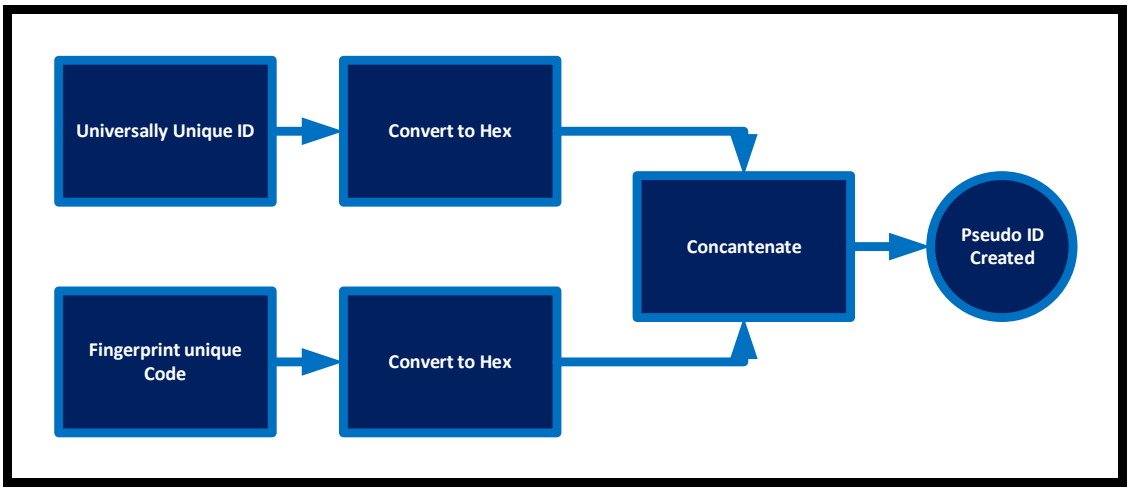


Figure 4. 9: Generation of Pseudo ID

Random eID Generation Process

Despite the user employing a pseudo-eID, if it remains static, the user can be profiled. Profiling can extract information that has monetary value. This includes demographics, shopping preferences, daily schedules, and behaviour. Having access to this data can violate the privacy of the data owner [17]. Hence there is a need to provide random eIDs to prevent the profiling of the user. Random IDs will address the challenge of profiling static IDs as the IDs will keep changing hence making it difficult to track a specific user and associate them with certain behaviours.

The generation of Random IDs is done via the use of a modified TOTP as opposed to the standard RFC6238 proposed by the Internet Engineering Task Force (IETF). The approach used is the modification of the method proposed by the standard. The standard is used to generate One Time Passwords (OTP) that are time sensitive using the Unix time as one of the variables in the generation of the passwords [143]. Our objective was to make the random OTP act as an identifier when the need arise without compromising the privacy of the user involved. Some users can commit fraud while transacting incognito hence the need for them to be traceable. Traceability must be part of the design than an afterthought when the need arises [79].

The approach is to generate the password as proposed by the standard and then modify the resulting password by passing it through a function that modifies it with an ID for that user as shown in Figure 4.10 below.

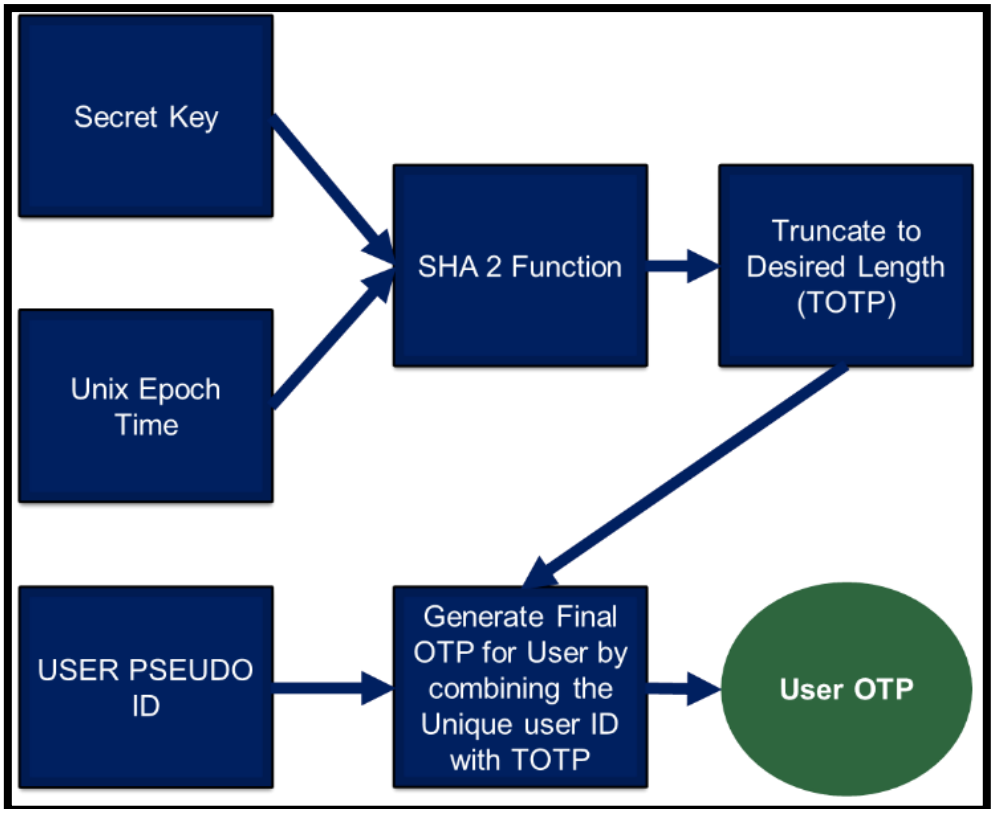


Figure 4. 10: Generation of Random IDs (User OTP)

To the user the ID is random, but the system can perform a reverse operation to generate an ID for the user of the random ID if the need arises.

4.1.4.1 Example of Random ID Generated

Using the standard TOTP generated using RFC 6238 was: **054465**

Convert to decimal (054465) = **1010100010001100101**

The Unique offline ID for the user was 001(country ID). 0010000001 (Unique User ID)

The ID is Decimal. We convert it to Hexadecimal we get **254A47A81** (Pseudo ID for the user to be stored online)

The Pseudo ID, eIDs, for the User, is 254A47A81 (0010010000001)

Appending the Pseudo ID to the standard TOTP 054465 (1010100010001100101 Base 10)

Results in: **00100100000011010100010001100101**

Converting to Base 64

Results in: **BQ3DHgeMF4D2Jo+1F**

Random User ID, eIDr, is: **BQ3DHgeMF4D2Jo+1F**

The Random ID will keep changing whenever the user generates one as one of the inputs is Unix epoch time which is continuously changing. However, the Unique user ID part will remain static hence when the need arises, it will be possible to trace back to the original user ID that generated and used that Random ID if fraud is committed. This will need to be done with the help of the KYC Agency.

Another benefit of this approach is that it will be possible to trace who generated which Random ID at any given time without keeping voluminous records of logs indicating the Random ID generated by who at what time and so on. By just examining the Random ID, the unique user ID can be deciphered by the KYC agency which will have a secret Algorithm. The Algorithm used here is just for proof of concept only.

4.1.5 Onion Routing

Despite the use of random IDs, the IP address and other unique properties of the computer being used can help trace and profile a user such as determining where they are connecting from, what sites they are accessing, and so on. To help address this challenge the use of Onion Routing (TOR) is recommended. TOR will mask the source of the requests to access services hence preventing online profiling thereby allowing a user access to services but at the same time providing anonymity [216][217][218]. The main use of the TOR network is to protect a user's privacy and it has gained popularity worldwide [219][220][221].

Figure 4.11 below shows how TOR works.

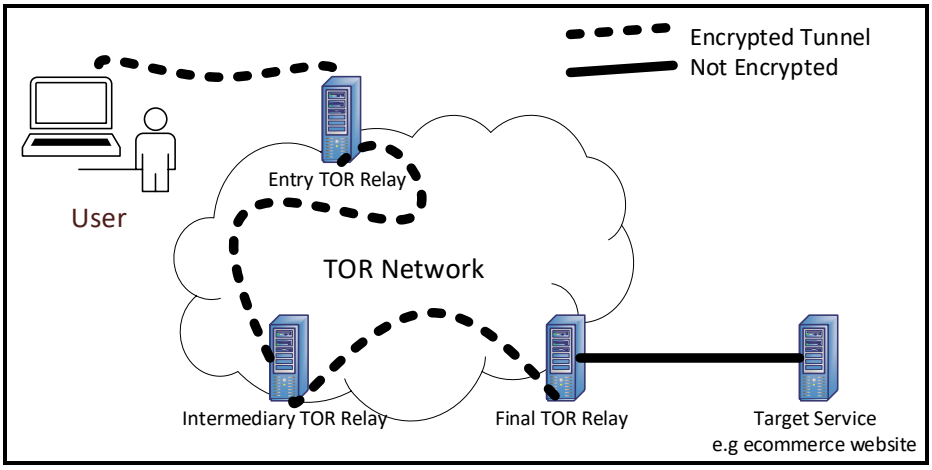


Figure 4. 11: How a TOR Network Works

TOR network employs a multi-layered architecture to provide anonymity to users. The approach not only makes the network complicated but also expensive. By masking the actual IPs of the requests by passing the information via several TOR relays, the user is shielded from being identified online not only in terms of username but also other profiling elements such as sites visited, services accessed, from where, and so on. In this manner, the TOR network anonymizes a user on the internet [222][223][224][225]. This helps further enhance user privacy.

However, the TOR network is vulnerable to end-to-end scheduling attacks by a perpetrator observing traffic from a client to the first TOR router and tracking that traffic through to the last TOR router until it reaches the target client [226]. Such an attack, if successful can unmask the identity of the clients involved and hence compromise their privacy. Similar attacks are orchestrated in Android phones via apps. The apps are de-anonymised [227].In addition, a TOR network can also be facing a Sybil attack. This is where an attacker takes control of a large number of virtual identities to obtain undue control of a network [228]. A TOR network seeks to hide one's identity while a Sybil attack seeks to de-anonymise the user via the control of virtual identities.

Furthermore, some legitimate service providers have less trust in TOR traffic as some users can abuse the anonymity feature it provides. Some users can carry out fraudulent activities online knowing that they will not be easily identified. Some users have used TOR to host sites selling drugs as well as hosting command centres for botnets. There

is a need to establish means and ways of ranking the trustworthiness of the TOR traffic source [229][221][216].

The above-proposed design can be modified by including Onion Routing (OR) to enable a user to avoid being traced via the internet addresses of the devices they are using to connect to cyber space. Figure 4.12 below considers the OR that would help keep the user devices used anonymously. The aim of OR in this configuration would mainly be to mask device identity and not necessarily user identity.

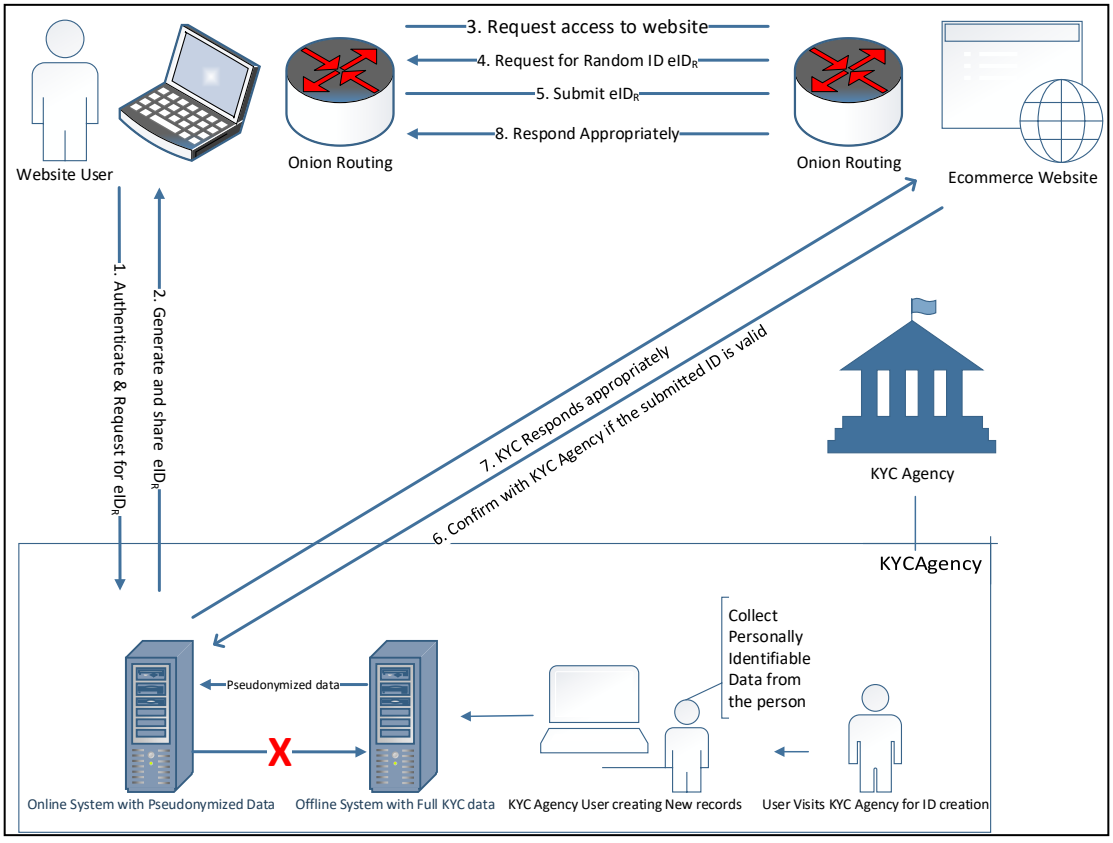


Figure 4. 12: KPA Operation with OR

4.2 Detailed Requirements

This section gives detailed requirements of the Data Protection Model designed. The Requirements contain detailed user stories, detailed descriptions, and sequence diagrams for the end –to-end process. The requirements in this section are adding more detail to the design that was outlined in the previous sections and outline the following:

(i) Detailed overall requirements for all the key system components

(ii) Sequence diagram for user registration

(iii) Sequence diagram for user interaction with the system

(iv) End-to-end user story for the system

4.2.1 Overall Detailed Requirements: All System Components

Detailed System Requirements

	Requirement-Registration System	User Story	Detailed Description	Comment
1	Login page	I want to be able to logon to the system using a combination of static password and AI based Captcha	1. Land on a page that will request for a static password and username. 2. Ask user to prove they are human by presenting a Captcha challenge 3. Grant access if provided respond to Captcha and password-username combination is correct	
2	Registration Portal	I would like to be able to register a new user onto the KYC system by capturing all the required details	1. Have a portal with all required details to be captured: a. First Name b. Surname c. Date of Birth d. Village and chief e. Physical/Home Address f. Capture live Photo of applicant g. Capture reference letters and details of referee h. Capture contact details: Phone numbers, email addresses and state preferred mode of communication i. Capture any other critical KYC data such as biometric data (fingerprint, face etc)	Refer to sequence diagrams and process flow chart
3	Generation of Pseudo ID	The system should generate a universally unique electronic ID for the successful applicant	1. If all details are correct, register member and generate universally unique electronic ID and append to their record on the offline database 2. Generate a Pseudo ID based on the electronic ID generated and send to the online System 3. DO NOT append Pseudo ID to record on the offline database system	
4	Online System Record creation	Once a new record is received from the Offline system create an online minimized account for the user with Pseudonymized data	4. Create a new account with the Pseudo ID as the key for the record 5. Append scrambled Mobile Phone number for the user 6. The scrambled Pone number must be stored under s filed called serial or anything other than phone number. The scrambling must replace numbers with letters and special characters not with numbers 7. Create Pseudo email box for the user and communicate to user on site 8. Issue user with key they will use to setup App for generating Random electronic IDs	The online record will only comprise 1. Pseudo ID 2. Pseudo Email Box 3. Scrambled Phone number

Overall Detailed Requirements: All System Components Cont'd

	Requirement-RMS System	User Story	Detailed Description	Comment
1	Communication Mode	The system must setup in such a way that data will only flow in one direction at a restricted speed	1. Data should only flow from the offline system to the online system 2. Only the Pseudo IDs and the scrambled Phone numbers must be sent to the online system 3. Microcontrollers are to be used for this purpose. 4. Bandwidth must be set to 9600bps 5. The Microcontrollers must be set in serial mode half duplex	
2	Multi-layered Hardware	The system must be setup with multiple hardware layers to enhance hardware security	1. The RMS must be built from two microcontrollers 2. They must be connected via serial. 3. The bandwidth between them must be 9600bps or less 4. They should be configured to communicate in half duplex mode 5. The microcontroller connecting directly to the offline system must be configured as the Master while the one connecting to the online system must be configured as the slave. 6. They communication WIRE from the receiving offline side of the system to the sending side of the online system MUST NOT be connected	
3	Periodic connection resets	The connection between the Master microcontroller and slave must periodically reset	1. The connection between the Master and slave microcontrollers must reset every 8 hours 2. If data transmission persists for more than 20 seconds, it must reset	
	Requirement-Random ID generating System	**User Story**	**Detailed Description**	**Comment**
1	Provide an App to be used for Authentication	I would like to be able to generate Random IDs using an authentication app like Google Authenticator	1. The App must be based on TOTP using SHA 2 2. The App must be able to generate an OTP as a Random ID for the user using the combination of the standard RFC 6238 TOTP based on UNIX epoch time and the generated Pseudo ID sitting on the online system 3. The Resulting modified TOTP (Random ID) can be delivered via the Mobile App, Phone, or pseudo email box 4. The App must allow user to setup the account on the app using the key they were issued by the KYC agency. 5. The App should not require communication with server to generate Random ID.	
2	Random ID generation	The system must generate the Random ID based on the RFC 6238 standard	1. Generate the TOTP based on the RFC 6238 standard 2. Combine the TOTP with the user's Pseudo ID 3. The resulting TOTP is the Random user ID to be delivered to the user's preferred channel. Recommended is the Mobile App 4. Create Algorithm that can decipher the Pseudo ID from the resulting TOTP	Refer to sequence diagrams and process flow chart
3	Algorithms to be created	I would like to be able to decipher the Pseudo ID from the random ID when need arise in case of fraud	1. Create Algorithm to only obtain the Pseudo ID from a random ID issued by a user to an online platform 2. Create an Algorithm that will be able to generate the actual ID from the Pseudo ID.	

4.2.2 Sequence Diagrams

The main objective of the system include:

(i) User registration after verification of appropriate KYC

(ii) Protection of User personally identifiable information

(iii) Preservation of User privacy

(iv) User interaction and verification in e-commerce transactions

4.2.2.1 User Registration and Protection

The following steps must be completed for successful user registration at a KYC Agency

(i) The user physically visits KYC Agency

(ii) The user provides valid Personal Identifying Information

(iii) KYC Agency creates a unique User record

(iv) User is given their details and information on how to use the system

(v) KYC Agency generates Pseudo ID

(vi) KYC Agency sends Pseudo ID to the Online system

(vii) The online system creates User records with minimised data

The process above is depicted in the sequence diagram in Figure 4.13 below:

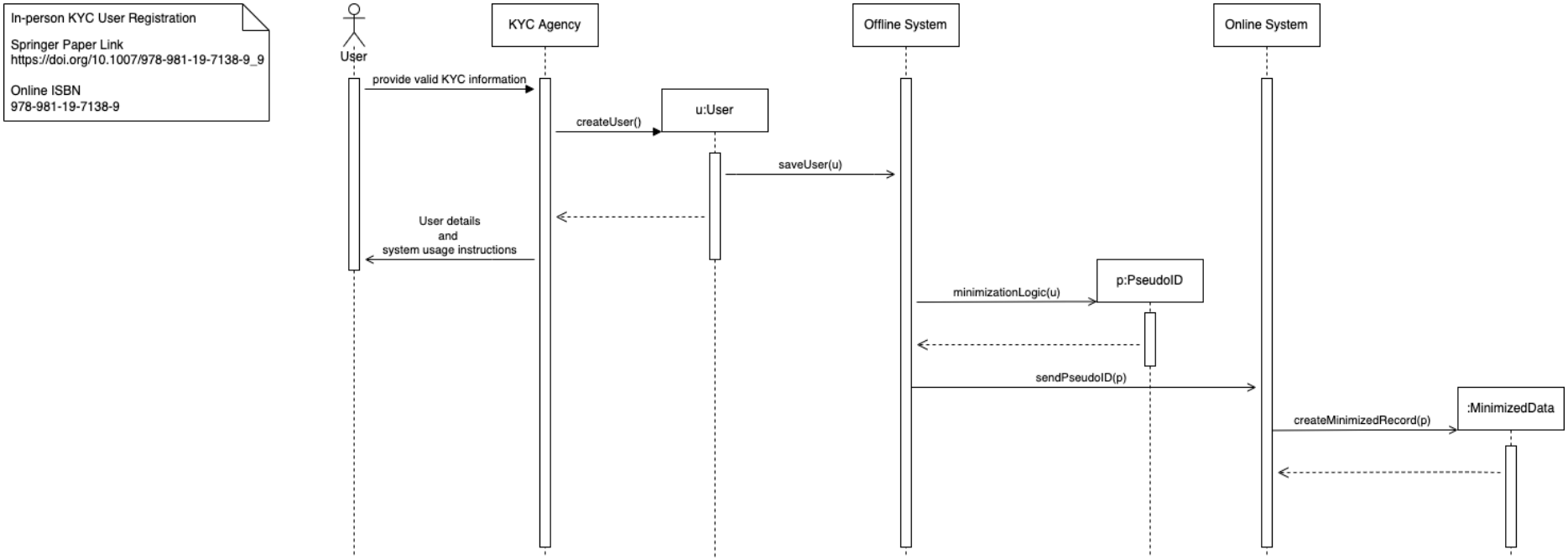


Figure 4. 13: Sequence Diagram-User Registration

4.2.2.2 User Registration and Protection

This is the core component of the system. It protects user data using a multi-layered hardware architecture and preserves user privacy using the Random ID derived from a modified RFC 6238 OTP architecture.

Once a user is created and would like to interact with the online platform, the following are the steps to be followed:

(i) User requests random ID from agency
(ii) The agency generates a random ID
(iii) The user requests e-platform access
(iv) E-platform requests authenticity confirmation from Agency
(v) The agency verifies random ID (User verification)
(vi) The agency responds positively to e-platform
(vii) E-platform provides service to the user

Figure 4.14 below demonstrates the sequence of events for the user to gain access to an online platform.

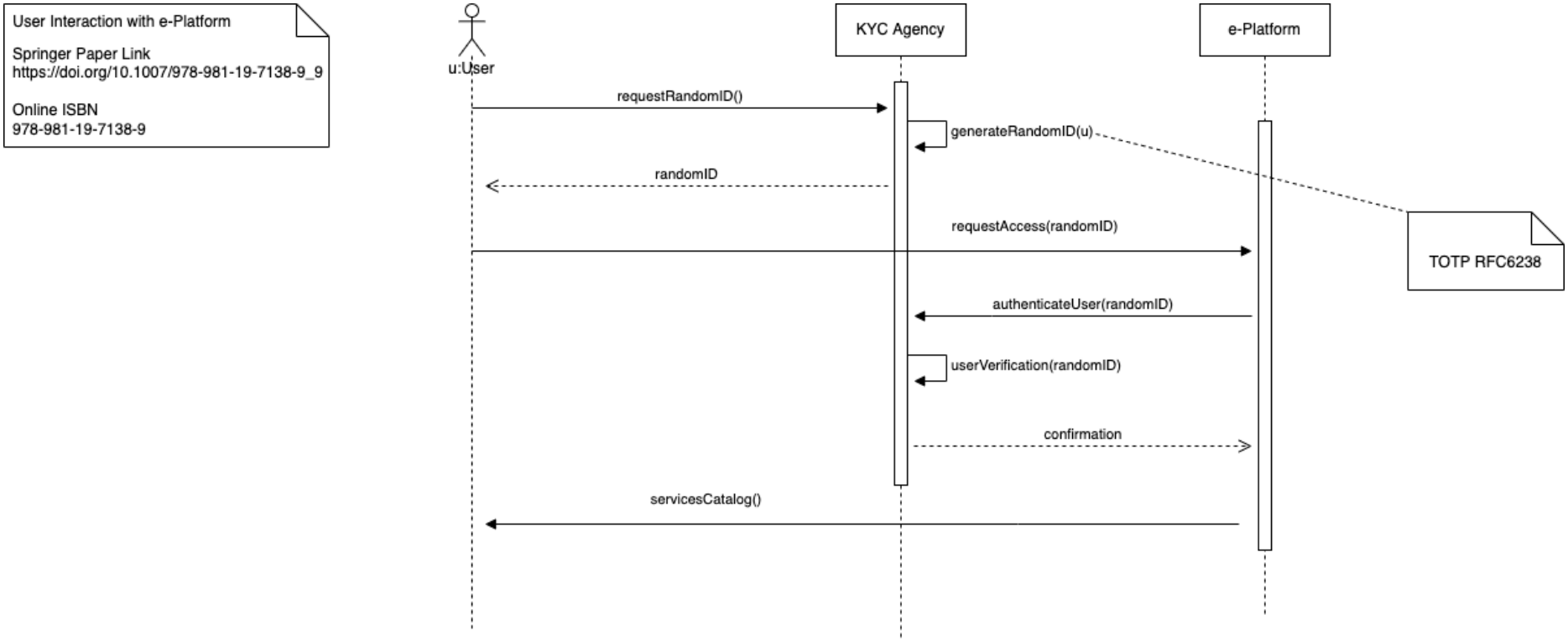


Figure 4. 14: Sequence Diagram- User Interaction and Verification

4.2.2.3 End-to-end Process Flow Chart

The end-to-end flow chart combines all the steps one would take to onboard and be able to use the system to gain access to online platforms without the need for sharing KYC data with those platforms. The flowchart in Figure 4.15 depicts the end-to-end process. That is the customer journey.

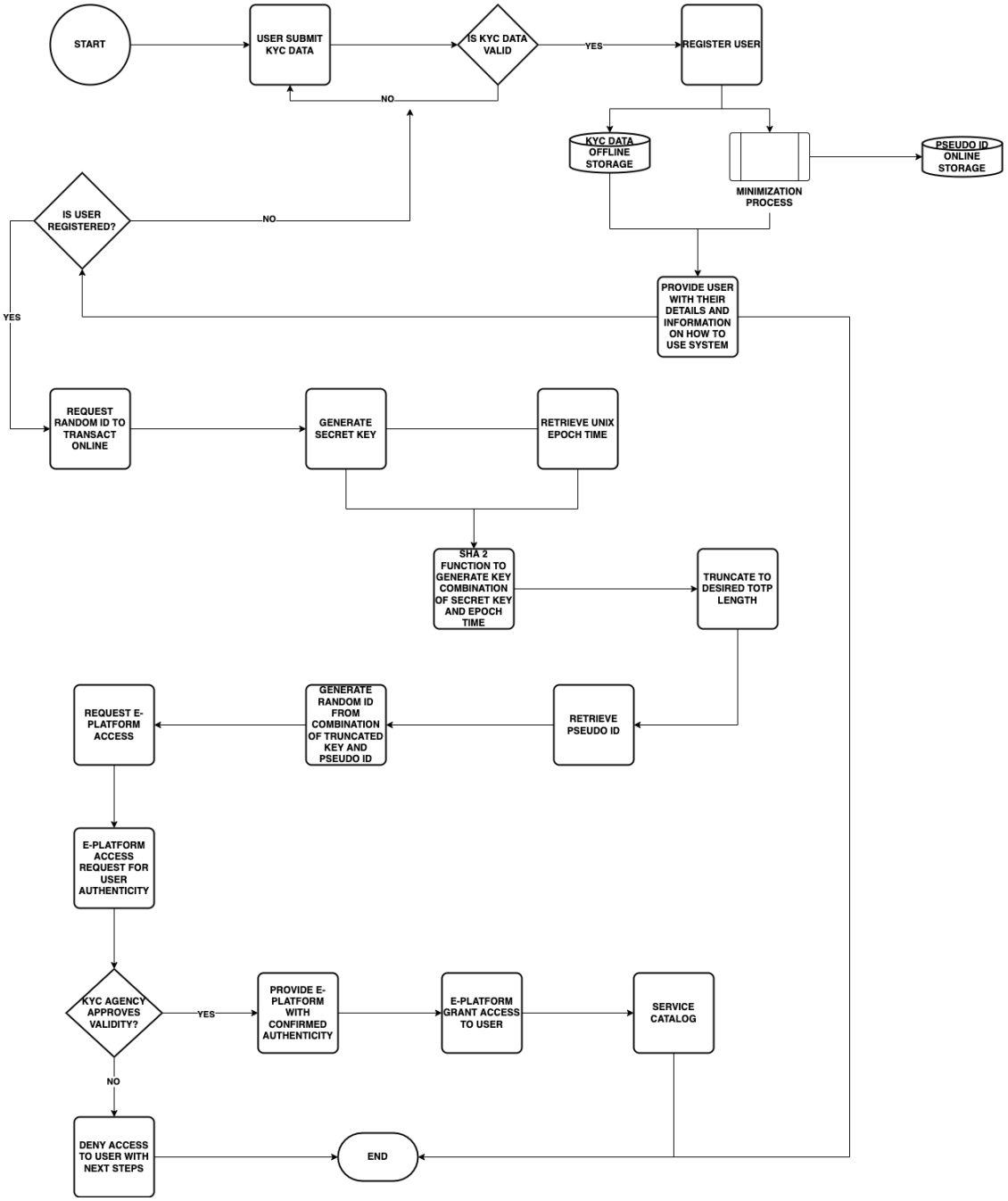


Figure 4. 15: End-to-end Process Flow Chart

4.2.3 Testing and Validation of the Design

4.2.3.1 Approach and Materials Used

Before the system could be developed and implemented, there was a need to validate and test the design to avoid wasting time building a system that might not function as desired. To validate the design of the system, the following tools were used:

(i) Proteus Simulation Software version 8 Professional
(ii) Virtual Serial Ports
(iii) Two (2) Virtual Microcontrollers (Arduino UNO)
(iv) A Laptop
(v) Arduino IDE programming platform (a language like C++)
(vi) Virtual serial Terminals

At this stage of the Software development cycle, only the key component (middle ware –Data Protector) was built on the simulation software to determine if the design would be able to work. Other Components like the Offline, online system, and code generators were built under the development and implementation of the prototype of the model designed in this chapter. Tests and details of these systems will be outlined in the later chapter dealing with the full implementation after the successful designing and simulation of the proposed model. These components were very similar to already existing solutions and would mainly be needed for the end-to-end testing and validation of the whole solution.

4.2.3.2 Proteus Setup

The system was setup as shown below in Figure 4.16 in the proteus simulation software: Two microcontrollers were configured as Master and slave in the simulation software. The Pin 11 and 10 on both the Master and slave Arduino UNOs were configured as Transmit and receive Pins respectively.

Virtual serial terminals were connected at three points of the Setup; on the entry point of Master Arduino, in the middle of the two Arduinos, and at the exit of the Slave Arduino. The entry point was being used to send some random data from Master Arduino (Offline system) to the Slave Arduino (Online system). The one on the middle

point was being used to monitor whether data was exiting Master Arduino and being sent toward the slave Arduino as were as monitor if data was exiting the slave and being sent towards the Master. The one at the Slave Arduino exit was being used to try and send data from the Slave (online system) towards the Master (offline system).

Virtual serial software was installed on the Laptop hosting the Proteus simulator. The virtual serial ports were configured on Proteus and serial port activities from Proteus could be monitored suing serial port software.

Figures 4.17 and Figure 4.18 show snippets of the code created in proteus for the Master and Slave Arduinos respectively.

The system simulation was tested for the following:

Data was sent from the serial terminal connected to Master Arduino toward the Slave Arduino. The data was successfully sent and reached the exit of the Slave Arduino and was displayed at the Terminal of the virtual serial monitor. This is depicted in Figure 4.19.

Data is sent from the Slave Arduino towards the master Arduino using the virtual serial terminal connected to the exit terminals (Pin 10 and 11) of the Slave Arduino. No data was monitored or observed at the middle and terminal connected to Master Arduino. Data was not sent successfully in the opposite direction. This is shown in Figure 4.20.

Further, the bandwidth was set at 9600bps to limit the amount of data that can be sent to the online system from the offline system. It would take more than 100 days to send 10GB of data at that rate, hence discouraging some villains.

Table 4.1 gives a summary of all the tests described for the various modules developed and tested in this research.

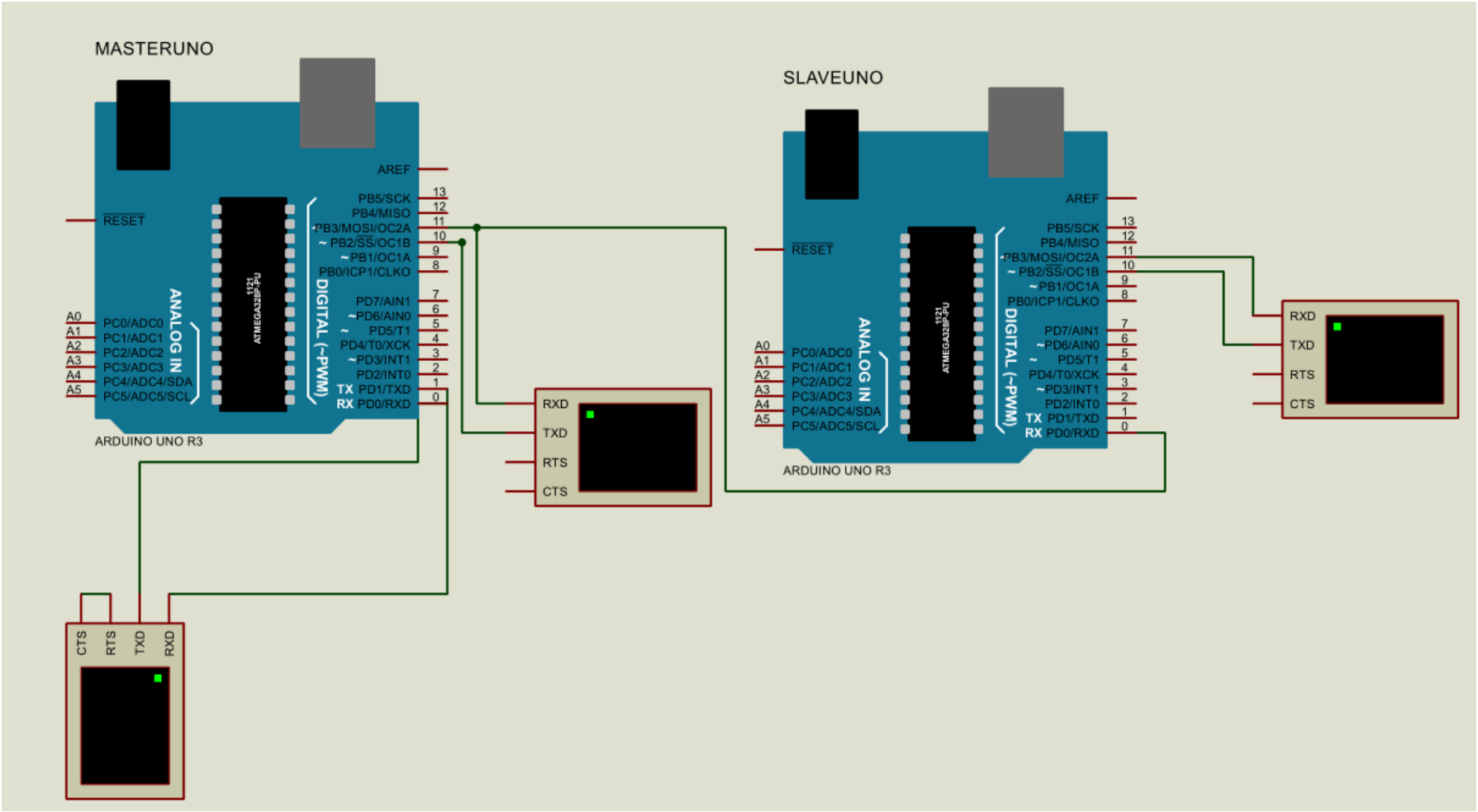


Figure 4. 16: Proteus Configuration Setup of the Middleware System

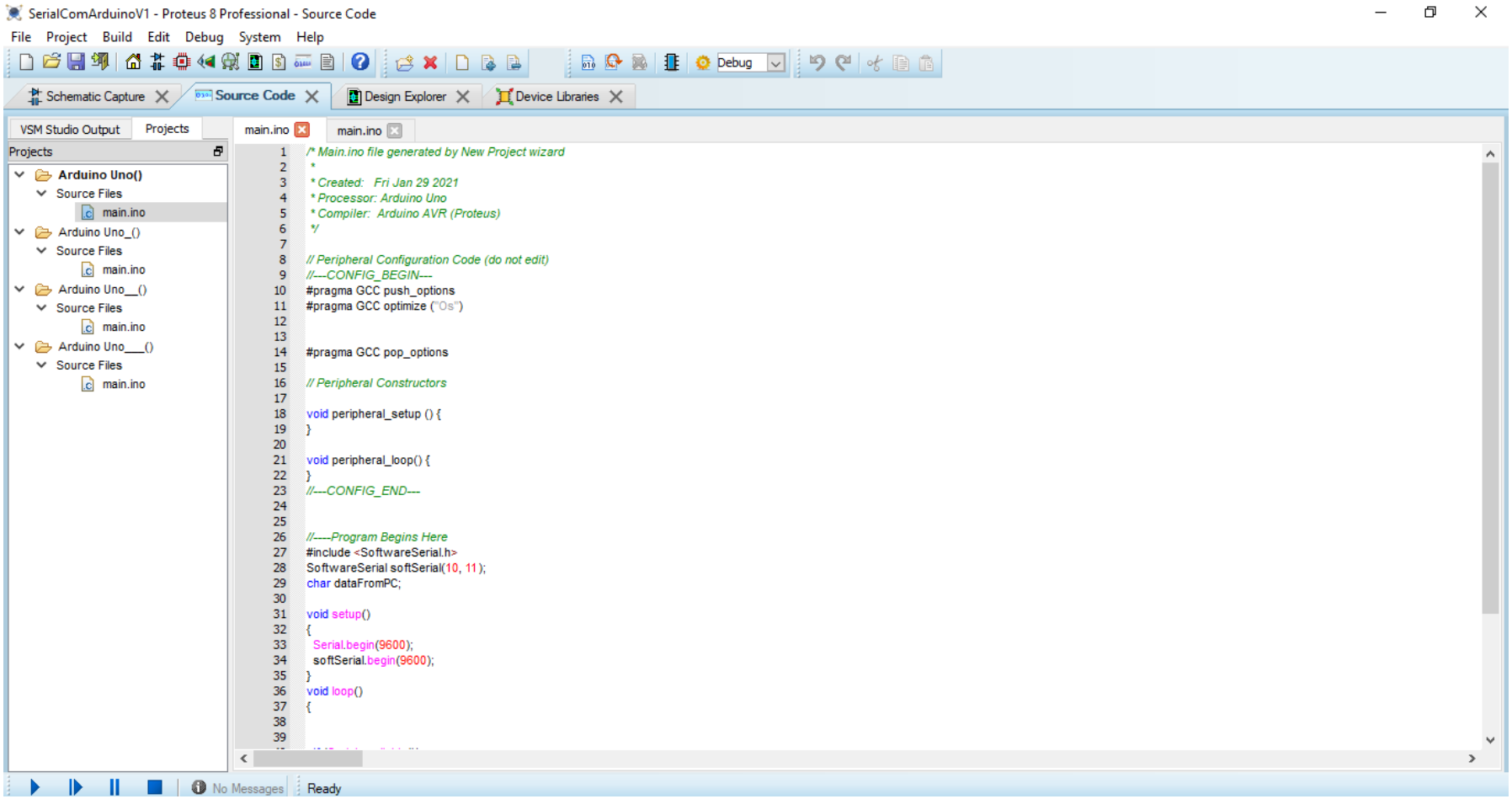


Figure 4. 17: Snippet of Master Arduino Code

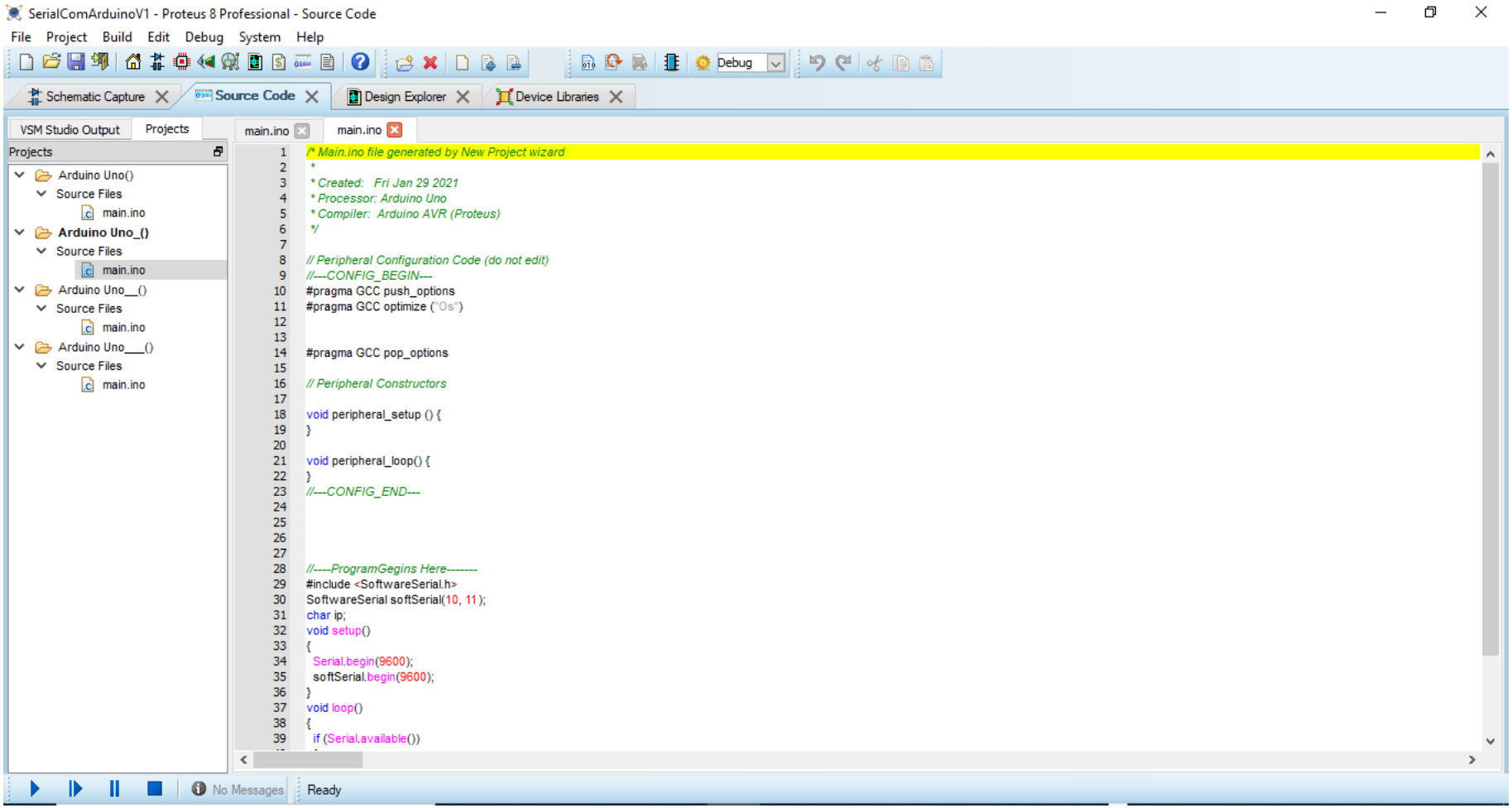


Figure 4. 18: Snippet of Code for Slave Arduino

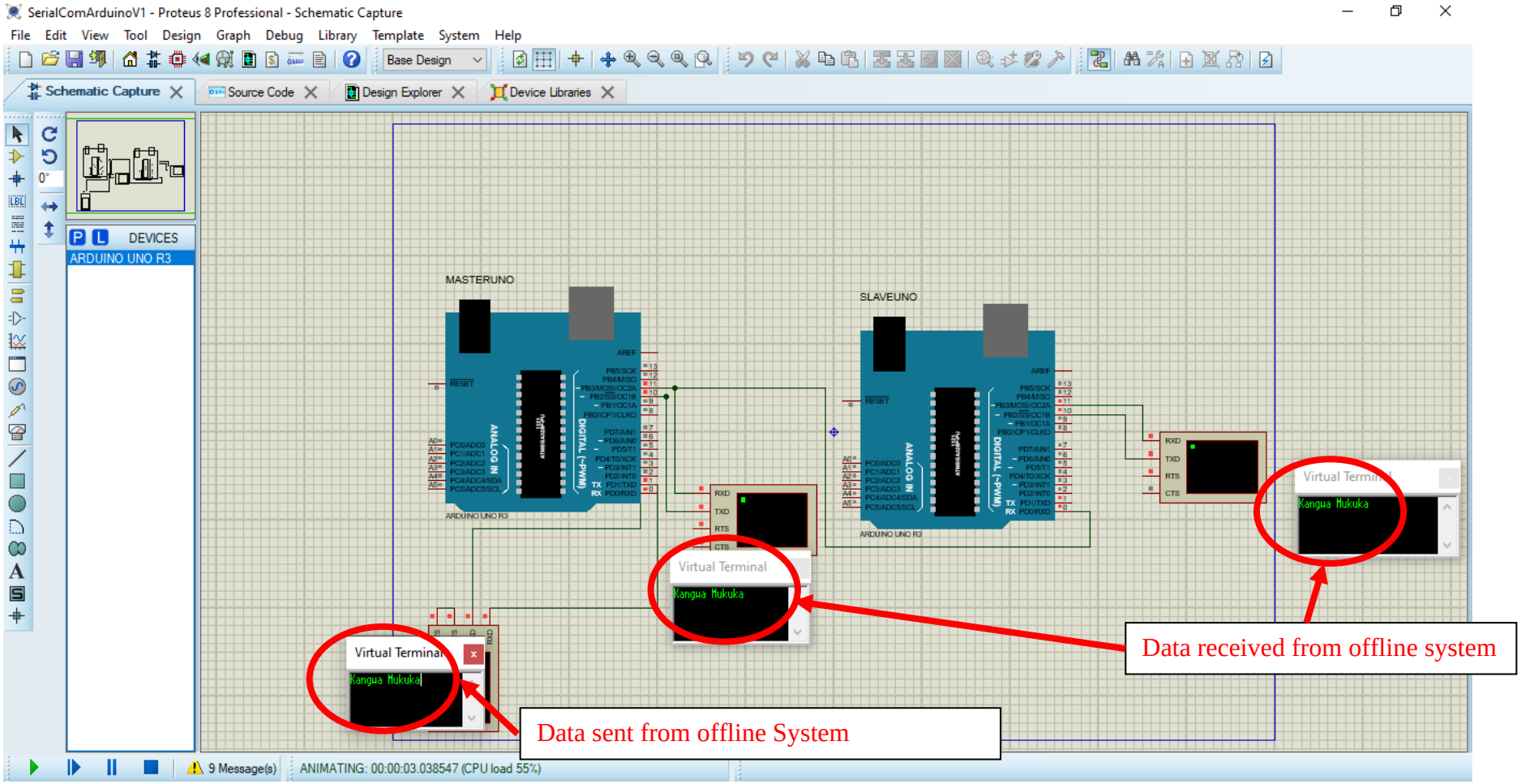


Figure 4. 19: Data Successfully Sent from Master (Online System) to Slave (Offline system)

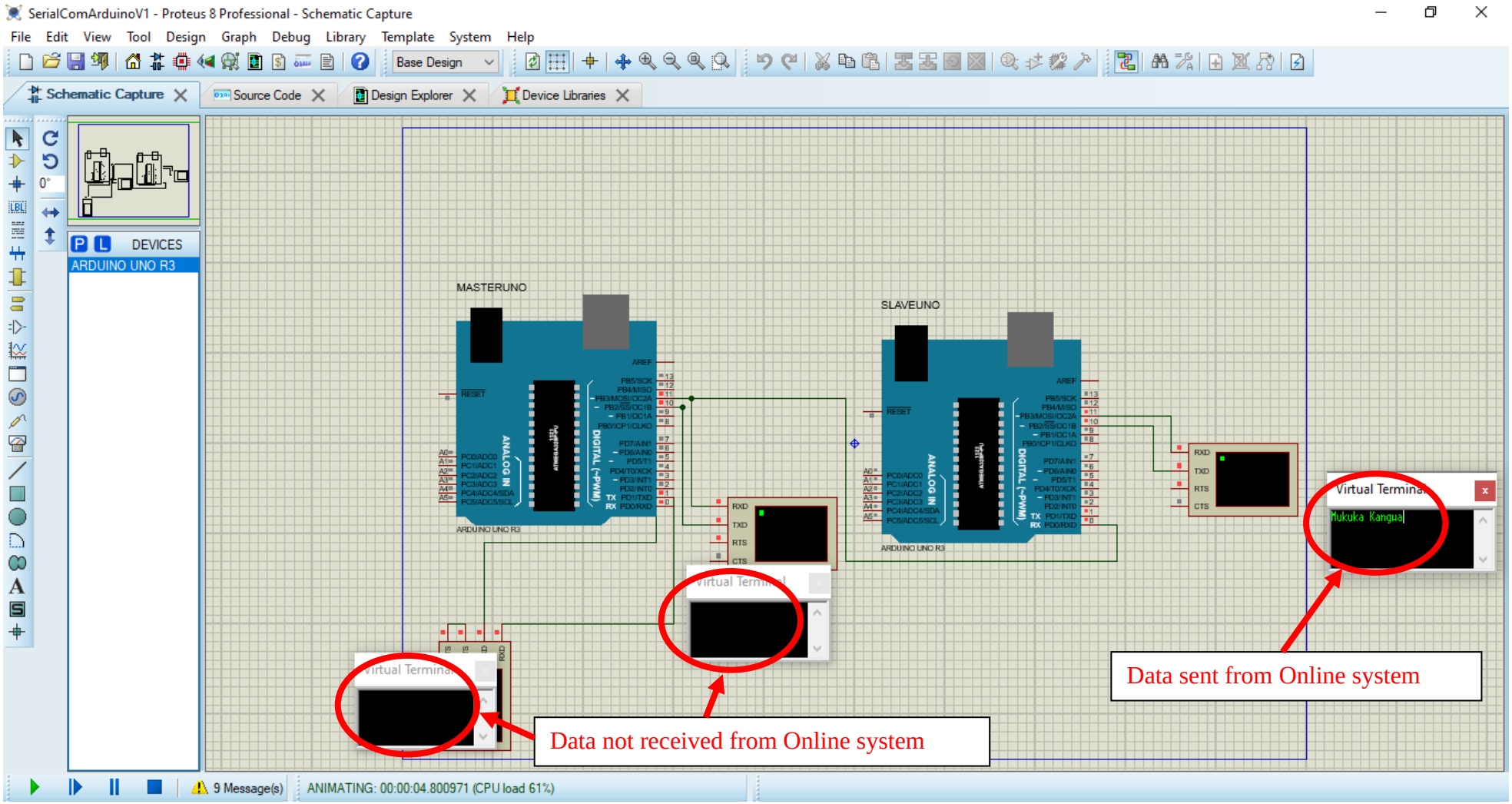


Figure 4. 20: Data could not be sent from Slave (Online System) to the Master (Offline System)

4.2.3.3 Summary of Simulation Test Results

Table 4. 1 Summary of Simulation Results

	Requirement-RMS System	User Story	Expected Results	Actual Result
1	Communication Mode	The system must setup in such a way that data will only flow in one direction at a restricted speed	1. Data should only flow from the offline system to the online system. 2. Microcontrollers are to be used for this purpose. 3. Bandwidth must be set to 9600bps. 4. The Microcontrollers must be set in serial mode half duplex	PASS PASS PASS PASS
2	Multi-layered Hardware	The system must be setup with multiple hardware layers to enhance hardware security	1. The RMS must be built from two microcontrollers. 2. They must be connected via serial. 3. The bandwidth between them must be 9600bps or less. 4. They should be configured to communicate in half duplex mode. 5. The microcontroller connecting directly to the offline system must be configured as the Master while the one connecting to the online system must be configured as the slave. 6. They communication WIRE from the receiving offline side of the system to the sending side of the online system MUST NOT be connected	PASS PASS PASS PASS PASS PASS
3	Periodic connection resets	The connection between the Master microcontroller and slave must periodically reset	3. The connection between the Master and slave microcontrollers must reset every 8 hours. 4. If data transmission persists for more than 20 seconds, it must reset	Not configured at this stage of the project

4.3 Chapter Summary

From the results obtained using simulation software, the design was plausible, and the project could proceed to build the prototype to test the solution end-to-end. The simulation results show that it is possible to protect data by isolating sensitive data from the internet using the Data protector simulated in this chapter. The solution used the multilayered hardware approach to achieve the objective of allowing data to only flow in one direction. It further protected data from leakage by ensuring that the bandwidth between the offline system and the online system was very minimal to deter would-be hackers from attempting as the exercise would be futile since a very long period would be required to transfer huge volumes of data.

The research proceeded onto building the prototype.

Chapter Five

Prototype Build

5.0 Chapter Introduction

To ascertain the effectiveness of the proposed solution, tests were conducted using various methods and tools. The major modules of the solution to be built were the Data protector, the offline system for storing data to be kept 'offline' and the online system to interact with e-commerce sites and other service consumers for user KYC confirmation. The code generator was also developed as it was a modification of the original RFC 6238-based OTP generator. Further, the formulation of unique algorithms for pseudonymization and the creation of traceable random IDs added novelty to the solution.

5.1 Materials and Method used

After the design and successful simulation of the Data protector system, it became critical to build the system end-to-end and test its effectiveness in protecting data. The following were the key components used to build the actual system:

(i) Two Laptops were used; one was used to host the offline system while another was used to host the online system: (CPU - Intel(R) Core (TM) i5-2520M CPU @ 2.50GHz 2.50 GHz, RAM- 8GB, Operating System- Windows 10 Pro, 64-bit operating system)

(ii) Two Arduino Microcontrollers (Arduino Uno): two were used to provide multi-layered hardware security to protect data from Hardware Trojans. The Arduino UNOs were used as they have a function to define software serial ports in addition to the existing serial PINs (Transmit and Receive). The solution needed more than one pair of serial PINs to achieve the desired goal.

(iii) Circuit board kit to make it easier to connect the various components.

(iv) Wire connectors were used to connect the two microcontrollers.

(v) A fingerprint reader was used to capture fingerprints.

(vi) Arduino IDE was used to program the Arduino microcontrollers.

(vii) GO programming language was used for the Offline backend system.

(viii) Html and Java script were used for the front-end part of the offline system.

(ix) Python language was used to build the online system.

(x) React web development tool was used to build the code generator.

The system was then tested against each of the key requirements outlined in the design chapter to determine if the system would meet key user requirements to protect PII.

5.2 Build Approach

The design used serial communication to build the RMS in preference to parallel communication. The aim was to ensure that data could only flow in one direction at a limited amount of bandwidth and serial communication makes this possible [195].

Two-way serial communication requires two cables physically connected between the two devices communicating with each other as shown in Figure 5.1.

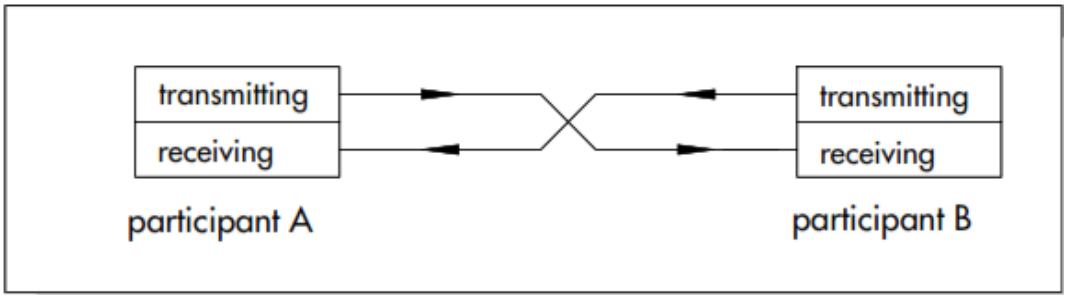


Figure 5. 1: Simplified Circuit Diagram [195]

There are three (3) basic ways of communicating serially, simplex, half-duplex, and Full duplex. In Simplex communication, data transmission is one-way only between participating devices while in Half-Duplex data the participants take turns in transmitting data. In Full-Duplex, participating devices can both transmit and receive data at the same time.

To ensure one-way communication, Simplex communication was employed by having one cable physically removed as shown in the Figure below.

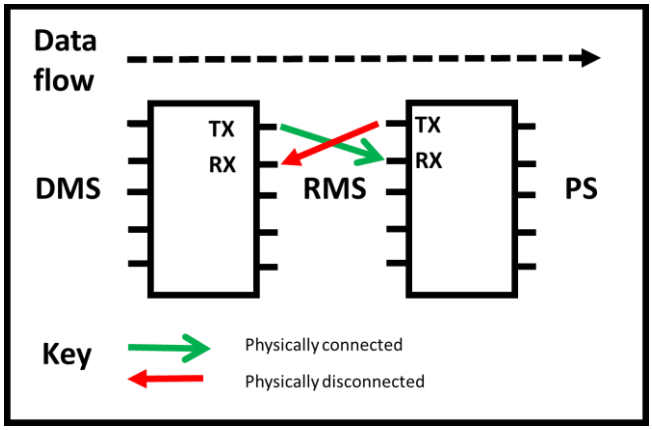


Figure 5. 2: Simplified Circuit Diagram

This would ensure that even if the online component wanted to communicate with the offline component, the communication would not be successful as there would be no means of reaching the other side. The achievement of this requirement was very critical as the offline component holds very sensitive PII. A breach of the system would result in PII being compromised via leakage or unauthorised access.

The design uses two microcontrollers instead of one to control and enhance the security of the data protection system. Hardware is susceptible to hardware Trojans. These viruses can be placed into the hardware at the manufacturing stage as the manufacturer can modify the design to put in place Trojans that can be used to deliberately leak information. For the Trojans to be activated, a hacker or disgruntled manufacturer would need access to the hardware either physically or remotely [181]. A system needs to remain resilient even when part of the setup has been compromised [102]. That is what this multi-layered hardware architecture achieves.

One Microcontroller connecting the offline side was configured as a Master while the one facing the online system was configured as a slave as shown in the circuit in Figure 5.2.3 below:

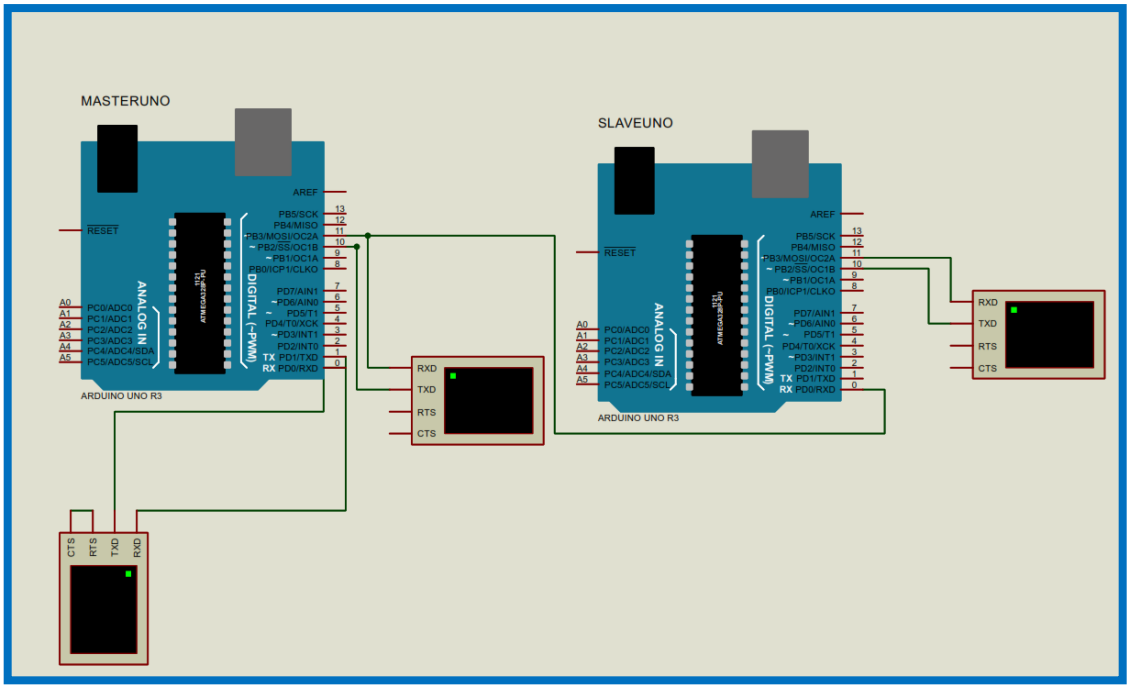


Figure 5. 3: Physical Configuration of RMS

Even if the Microcontroller facing the online system was compromised, the hacker would not be able to breach the entire connection as they would need access to the Master Microcontroller to change configurations and enable two-way communication thus making it impossible to achieve without having physical access. The multi-layered hardware approach makes the solution simple, yet very effective.

The Figure below shows the Arduino IDE source code for the Master controller.

Master Processor | Arduino IDE 2.0.3

File Edit Sketch Tools Help

Arduino Uno

Master Processor.ino | Master Processor.ino

```
#include <SoftwareSerial.h>
SoftwareSerial softSerial(10, 11);
String dataFromPC;

void setup()
{
  Serial.begin(9600);
  softSerial.begin(9600);
}
void loop()
{


  if (Serial.available())
  {
    dataFromPC=Serial.readStringUntil('\r');
    softSerial.println(dataFromPC);
  }

}
```

Figure 5. 4: Snippet of Source Code on the Master Microcontroller

PINs 10 and 11 were configured as Receive and transmit PINs respectively. This was to ensure that serial communication is employed for data transmission. A library called “software serial” had to be used to enable the creation of another serial port pair. This happens when one needs more than a single pair of serial ports. The setup required serial communication between the computers and the microcontrollers as well as serial communication between the two microcontrollers.

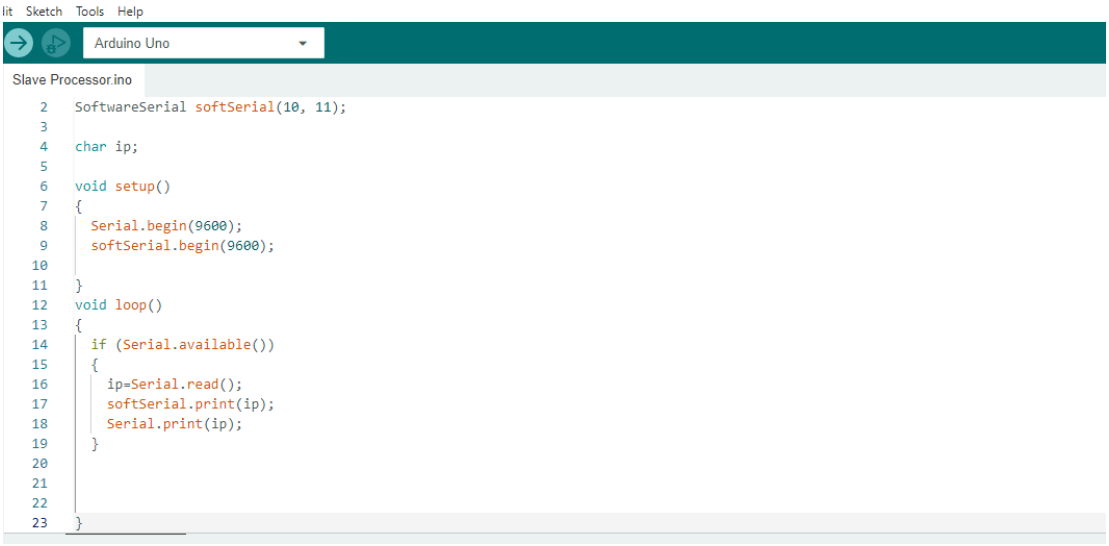


Figure 5. 5: Snippet of code for the Slave

The Figure below is the physical wiring of the actual setup:

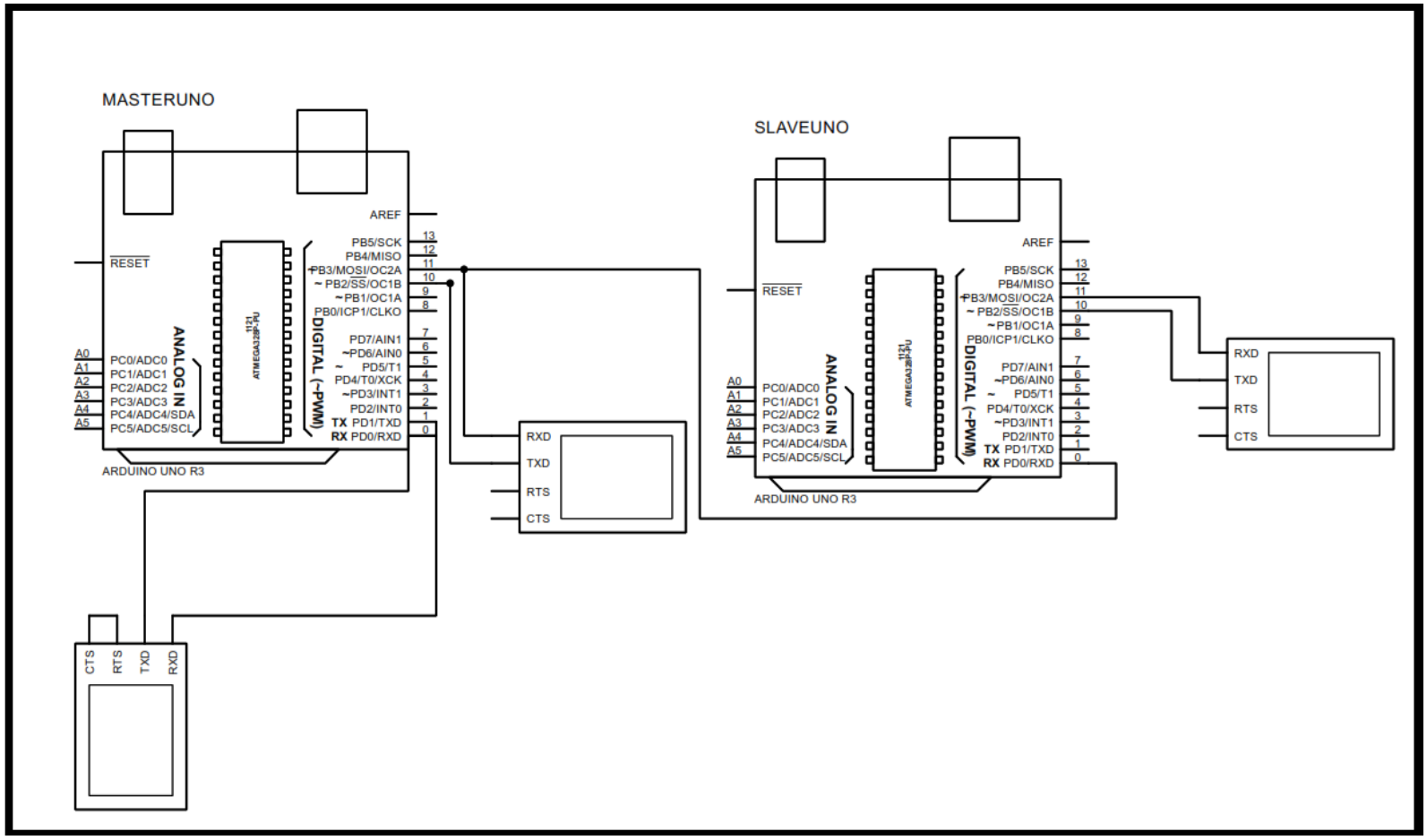


Figure 5. 6: Diagram of Physical Representation of Circuit Setup

The picture below represents the actual setup of the Data protector.

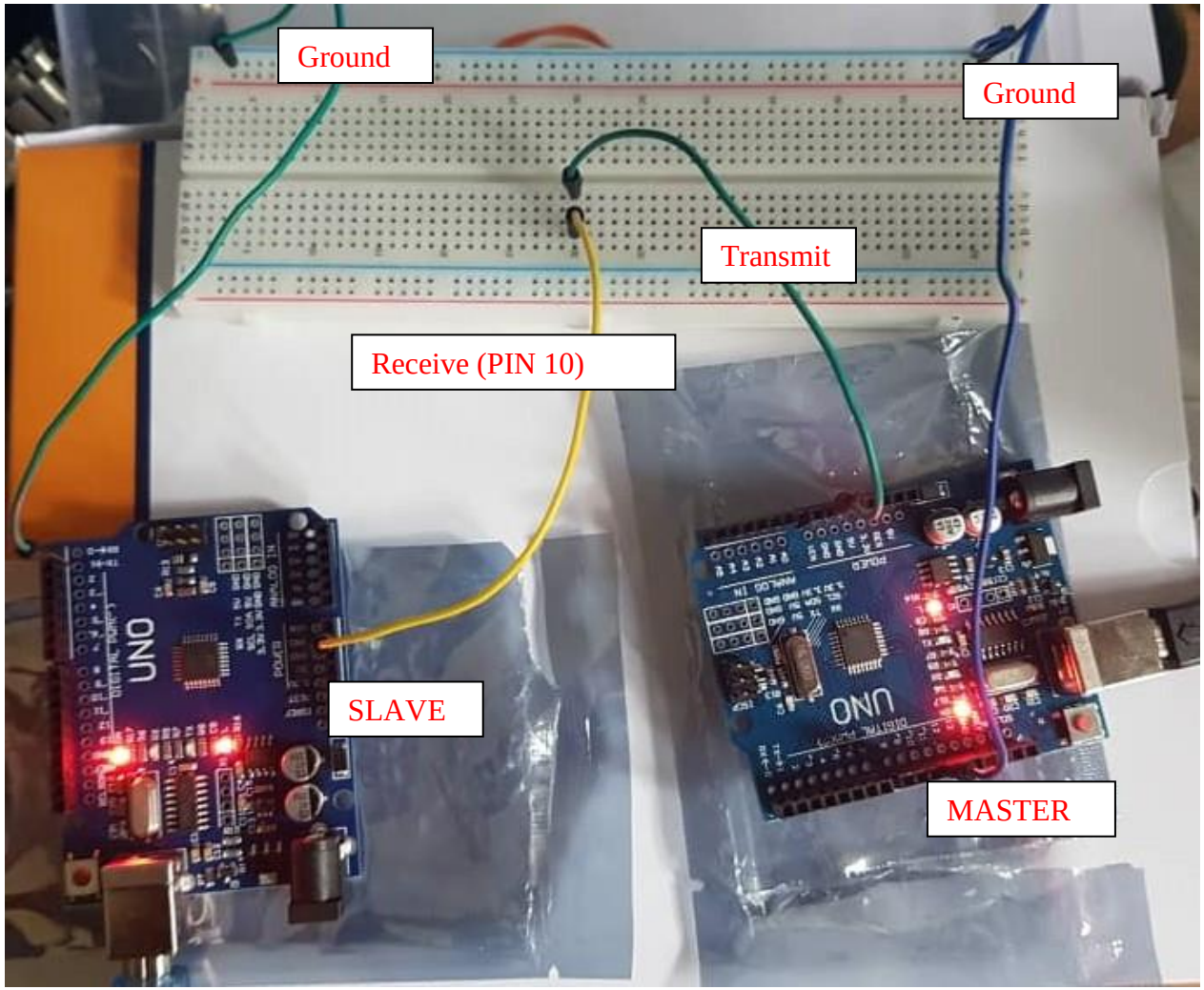


Figure 5. 7: Actual Setup focusing on the Data Protector

The picture below shows the end-to-end setup of the system. That is, both online and offline laptops were connected via the data protector shown above.

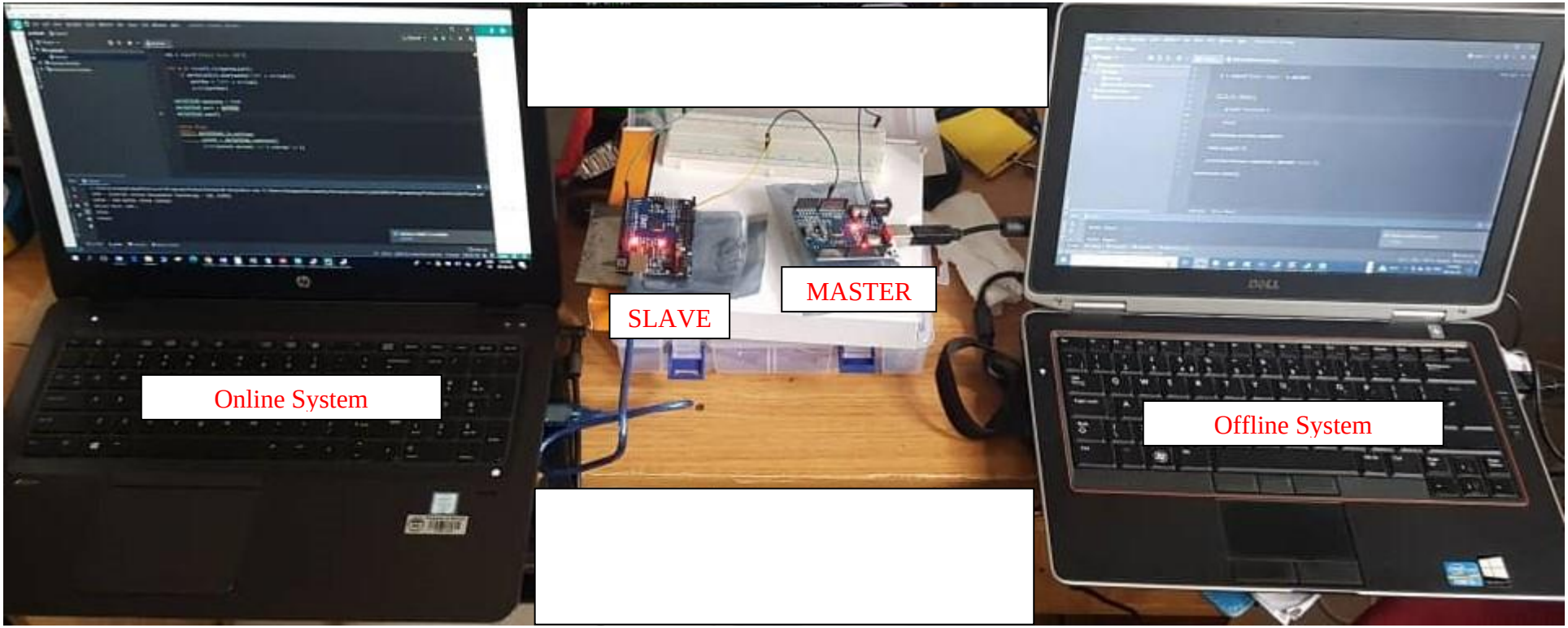


Figure 5. 8: End-to-End setup of Configuration

The end –to-end setup comprised, the data protector, and serial cables connecting the data protector to the two laptops as well as the online and offline laptops. The laptops were both connected to the data protector using serial USB cables. The same cables were supplying the required 5 volts to the two microcontrollers.

The transmitting PIN (TRX) of Master Arduino was connected to the PIN 10 (configured as Receive PIN) of the Slave Arduino. It must be noted that the grounds for both controllers were connected to the same line to ensure that the voltage between the two microcontrollers was not floating. A Floating voltage negatively affects serial communication.

The Offline system was configured with a Fingerprint scanner to capture fingerprints. These fingerprints were being used to generate the Pseudo ID which would ensure the ID is unique to the user whose record has been created on the system.

The figure below shows the scanner connected to the Offline system and having a fingerprint scanned and stored on the system database.

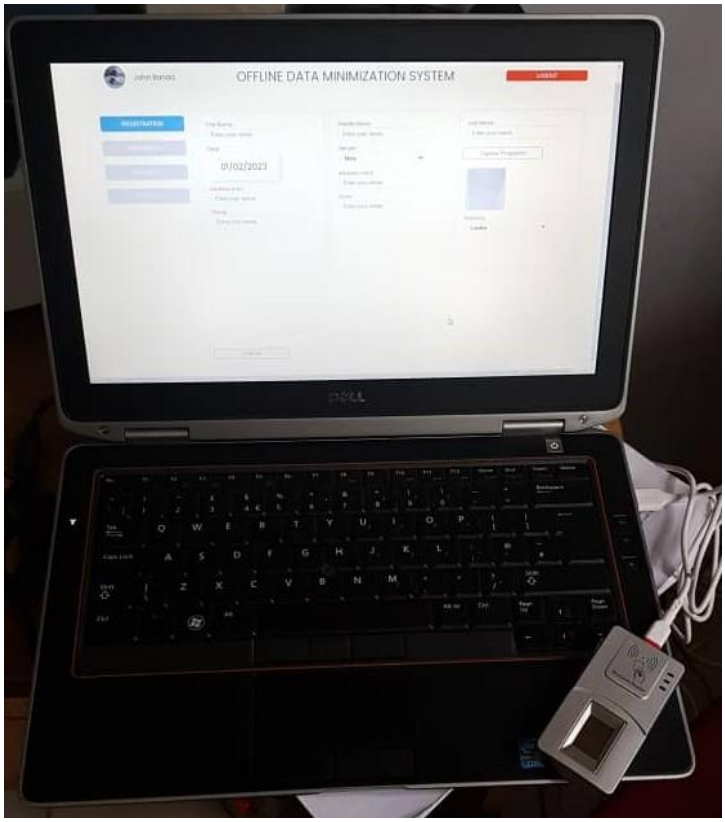


Figure 5. 9: Setup of the Fingerprint Scanner

Chapter Six

Prototype System Validation and Testing

6.0. Introduction

This chapter outlines the tests that were conducted on the prototype that was built to validate its conformance to the Model that was designed in chapter four. The three key components tested were as follows.

(i) Data protector: control of data rate and direction between the offline and online system

(ii) Offline System: hold all PII for customers.

(iii) Online System: offer KYC validation services using minimized pseudonymized data

(iv) Code generator: used to generate unique random electronic IDs

6.1. Concept Compatibility Tests

To ascertain whether the system would effectively protect PII, it is vital to understand the stages hackers usually undertake before the eventual exfiltration of data to their secure location or encryption of sensitive data for ransom. Thereafter, compare the system test results with the various exploits hackers use to breach security and determine whether such exploits can be prevented by the proposed system.

There are about eight (8) steps a hacker is likely to undertake to get to their target. The first is called **Reconnaissance**. This is the stage where the hacker will try to gather information about their target to understand their weaknesses and strengths. They will gather information such as the software, hardware, and associated versions the target is using. The information gathered can also include personal information. This stage takes longer than most of the other stages. Once all critical data has been gathered, they then proceed to the **intrusion** stage. At this point, the hacker tries to breach the perimeter security of the organization by using data they have collected such as compromised credentials for staff. They can also try to use malware through email delivery and so

on. Deceiving people is much easier than deceiving security software or hardware that has been configured in place [187].

Once the attacker has gained access to the environment; they will try and gain more ground through the **exploitation** of various tools available to them. Tools will help them understand the internal environment better and give them a better view of the system and application architecture and potential vulnerabilities to exploit. **Privilege escalation** follows exploitation. At this stage, the attacker tries to gain access to more systems and parts of the network by escalating the rights of an ordinary user account they might have compromised. They can achieve this by trying to capture administrator accounts using weak or known default passwords. Once this is attained, they start conducting **Lateral movements**. In this activity, they continue exploring additional systems, applications, data, and other critical resources they consider of great value to their mission[187].

So far, their activities would have been leaving some tracks behind. They would then start masking their tracks to avoid being detected. This process is called **Obfuscation**. They will clean log files and some even plant fake forensic evidence to mislead investigators once they discover the breach. At this point, the attackers can launch a **Denial of Service** (DoS) attack to prevent legitimate users from accessing the breached systems. Not all attackers choose this route. The last stage of the attack is **Data Exfiltration**. This is where the perpetrators copy and move data from the victim systems to a secure location only, they have access to. They can then analyse the data and seek ransom for it. At this stage, the attacks can even encrypt the local copy of the data they access and seek a ransom to release the data they would have stolen and externalized [187].

Tests were conducted by sending data in both directions. That is, data was sent from the offline system to the online system via the RMS as per the configuration outlined. Data was also sent from the online system to the offline system. Six scenarios were tested. In scenarios 1 and 2, both the Transmitting and receiving PINs were physically connected as shown in Figure 6.1. This was to determine whether data would be transmitted both ways if Full or half-duplex serial communication was setup and ultimately determine its suitability in the setup of the proposed solution. That is if data can be transmitted one-way.

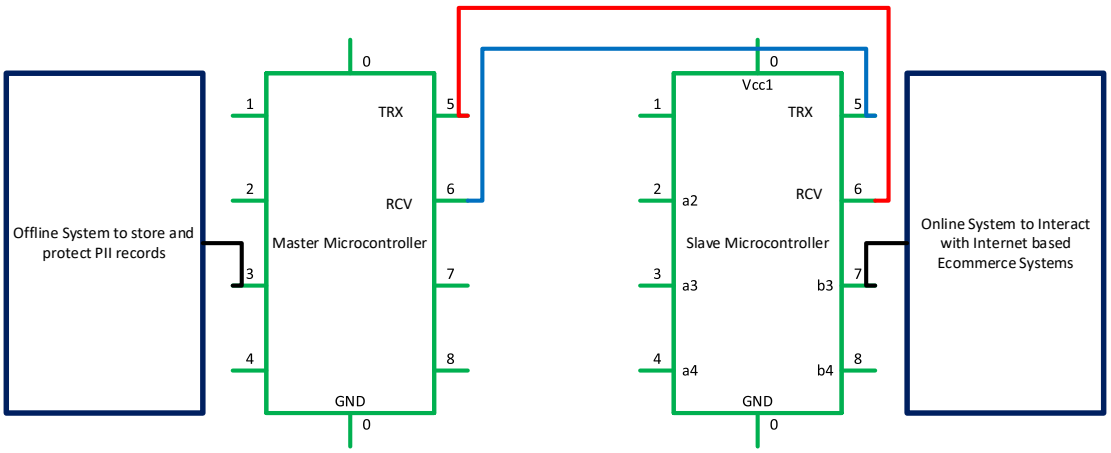


Figure 6. 1: Scenario 1 and 2 Experiment Setup

while in Scenarios 3 and 4, the cable connecting the Transmitting PIN for the Master microcontroller to the receiving PIN of the Slave Microcontroller was disconnected as shown in Figure 6.2. The main objective of the solution is to prevent access to sensitive data from someone who is online. That is, they would be connecting from the system on the slave microcontroller side. The test was to ascertain which way the data would flow and whether reconnaissance would be possible.

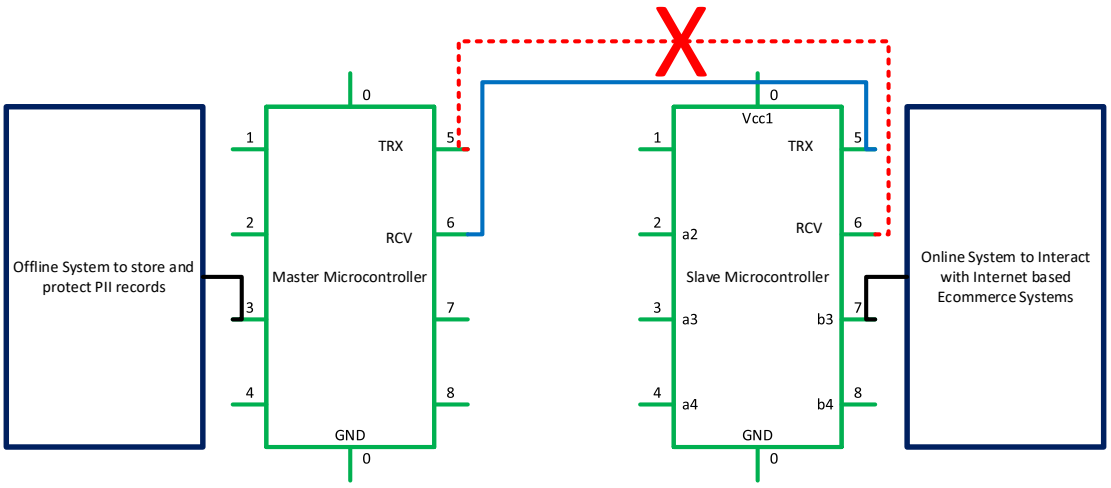


Figure 6. 2: Scenario 3 and 4 Experiment Setup

In scenarios 5 and 6, the receiving PIN of the Master microcontroller connected to the Transmitting PIN of the slave Microcontroller was disconnected as shown in Figure 6.3 Again, the test was to help find the most appropriate connection that would enable effective protection of sensitive data through multi-layered hardware solution and enforcing a one-way traffic flow, from offline to online.

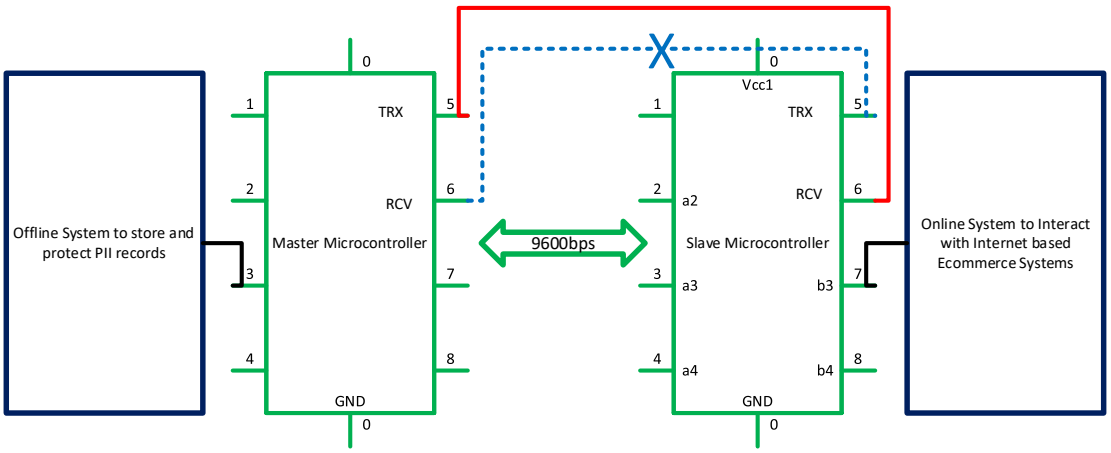


Figure 6. 3: Scenario 5 and 6 Experiment Setup

The Bandwidth between the two Microcontrollers was set at 9600bps. This speed can be adjusted as desired. The lower the speed, the more time will be required to send huge amounts of data hence the more frustrating to the hacker. The restriction is very critical to ensure there is no exfiltration of huge amounts of data by disgruntled internal staff.

Further tests were conducted to determine if it was possible to generate random electronic IDs that users could employ online to protect their privacy from being compromised by profiling their static Pseudo electronic IDs. In addition, for investigation purposes, it was also investigated to ascertain if it was possible to decipher the actual ID of the user if they engaged in fraudulent activities while using random IDs online. The Trusted Third Party was the only one to be able to decipher the actual identity of the perpetrator from the Random ID used online.

6.2. The capture of the End-to-end Tests

The screen shot below in Figure 6.4 shows the capture of KYC data on the Offline system. This includes the capture of fingerprints for the person being registered.

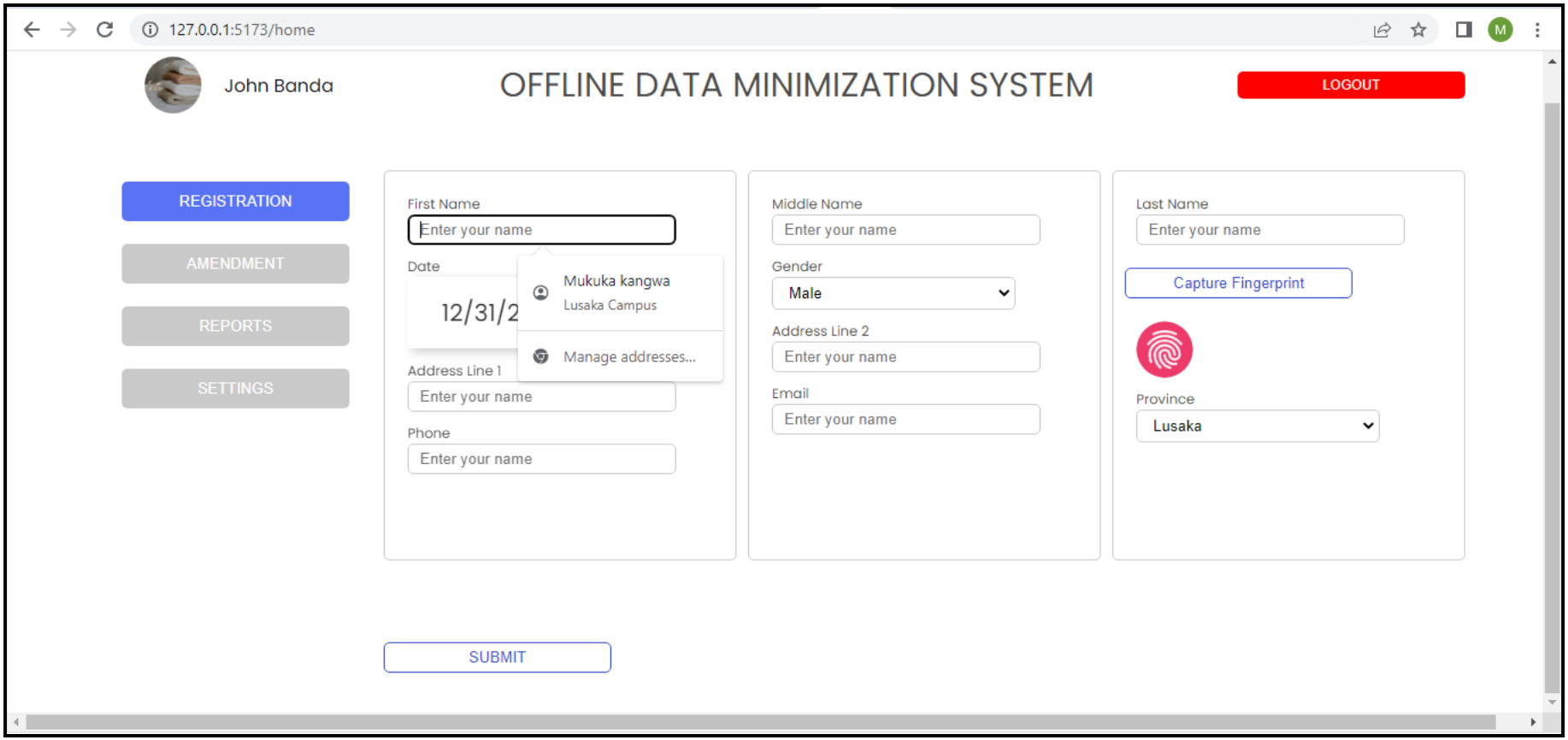


Figure 6. 4: Offline System-Interface for Capturing of KYC data (Including Fingerprint)

The screen shot that follows in Figure 6.5 is showing data that has been captured together with the fingerprints awaiting committing to the database:

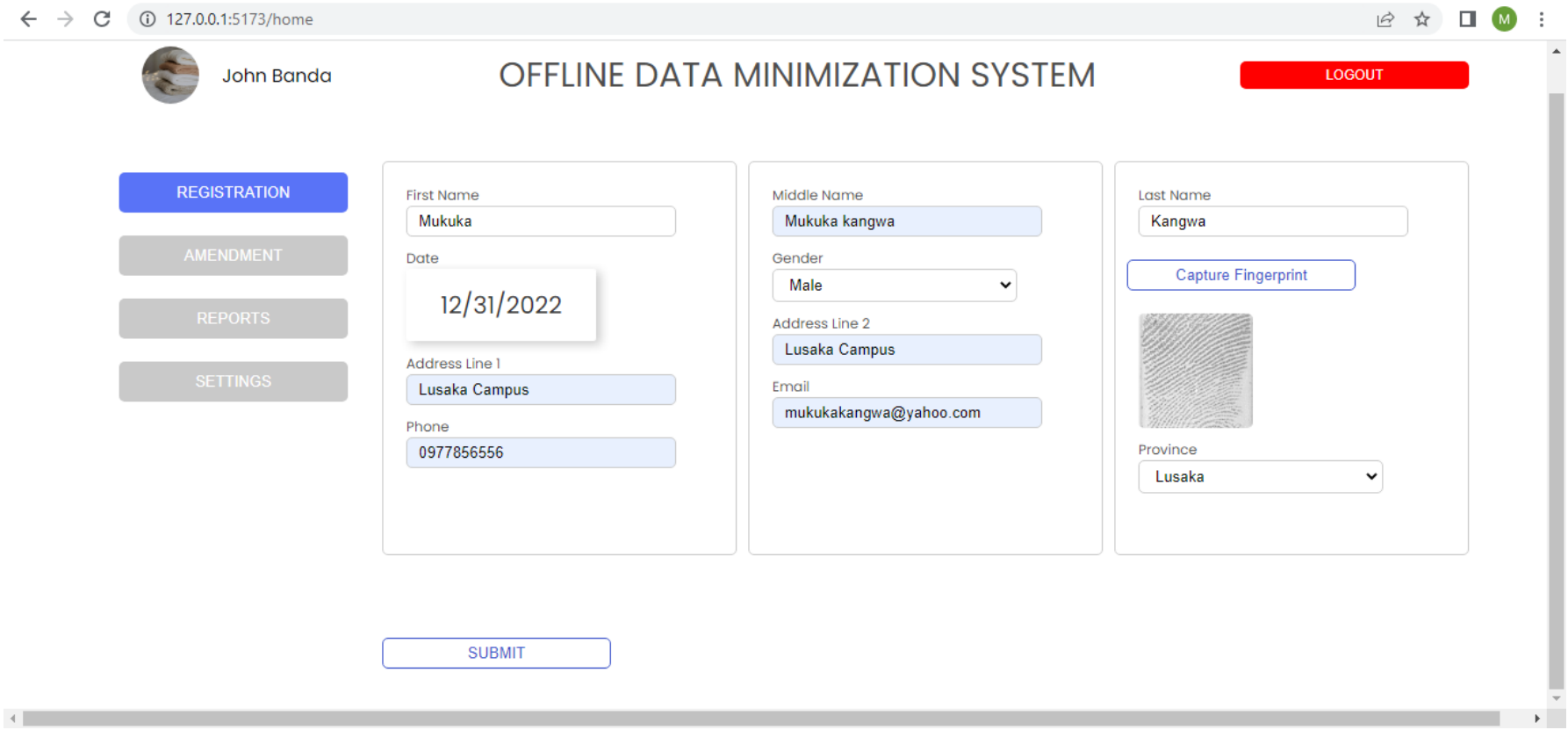


Figure 6. 5: Actual KYC Data Captured to be committed to the Database

After committing the data to the database the message in Figure 6.6. below is displayed when successful

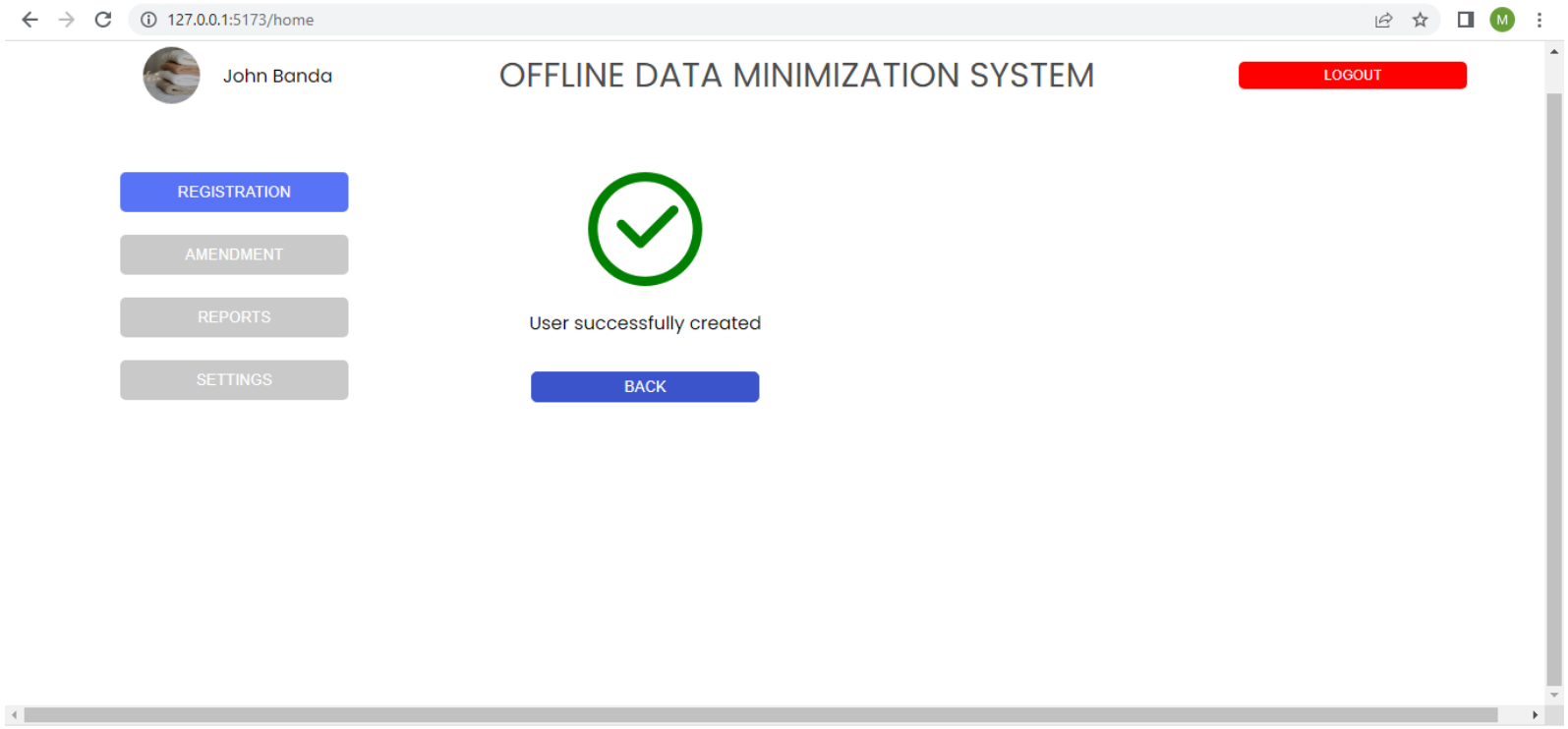


Figure 6. 6: Message Displayed after Successful Committal to the Database

6.3. Chapter Summary

It is very essential to thoroughly test a given system before it can be concluded whether the system meets its required objectives. In this case, the tests were being conducted to determine if the system would effectively protect PII for e-commerce users. The system can be extended to other forms of data that need stringent protection measures.

The tests conducted focused on determining how plausible it was to access sensitive data by a hacker who is online as most data exfiltration cases have occurred with the perpetrators accessing the data using online tools. The tests looked at several critical elements involved during the various stages of cyber-attacks. One critical element is reconnaissance. Tests were done to ascertain whether one could reach the system storing sensitive data from the online side unauthorized. Before any attack to exploit can be launched, an attacker needs to collect data about the target to determine the appropriate tools and attack methods to employ. Further, exploitation of the discovered weaknesses must take place for the hacker to gain access to the target systems. Once inside, then data can be stolen.

The tests looked at all the crucial exploitation stages and focused on preventing data exfiltration as well as data encryption by ransomware. The tests further looked at user privacy preservation while using online platforms.

Further to the tests was the need to determine the effectiveness of using a multi-layered hardware architecture to protect PII from external attacks as well as from Hardware Trojans.

Chapter Seven

Experiment Results

7.0 Introduction

The Solution Built was subjected to various tests to ascertain its efficacy. It was first tested for Technical Concept Compatibility and then tested for Data Leakage Vulnerability. An additional test was done to determine whether one could decipher an ID from a Random eID issued to a user if the need arose. Results for all the tests are presented in the sections that follow starting a summary of all the tests in Table 7.1.

7.1 Results for Technical Concept Compatibility Tests

Table 7.1 below gives a summary of the results that were obtained from the tests conducted using the prototype built as well as simulations conducted using Proteus software.

Table 7. 1: Results for Concept Compatibility Tests

No	Test |Scenario	Results	Desired Result	Overall Result	Status
1	Data Transmission from Offline System to Online System	Data was successfully sent and reached destination	Data should only flow in one direction	Pass	
2	Data Transmission from online to Offline System	Data could not be successfully sent	Data should only flow in one direction	Pass	
3	Bandwidth Limitation Effect	Sending 10GB estimated at more than 100days	Estimated time should be frustrating. Sending 10GB should take less than 1 hour under normal circumstances	Pass	
4	Generation of Random IDs	Generated Random ID with ebededed User ID	Always produce unique Random ID	Possible	
5	Decipher original static user ID	Generate Static User ID kept offline by KYC agency	Always produce unique static ID	Possible	
6	Connect to Master microntroller (Offline) from Slave controller (Online)	Not able to connect	Should not be able to connect	Pass	

Scenario 1 above indicates that data was able to flow from the offline system to the online system. This is the desired state as data should be able to flow in this direction to enable the transmission of Pseudo IDs generated for newly added users. Without this result being achieved, the proposed solution would be a failure as manual transfer of Pseudo IDs is not desired to prevent malware transmission and physical exchange of data media. The figure below shows the screenshot of the results from the experiment:

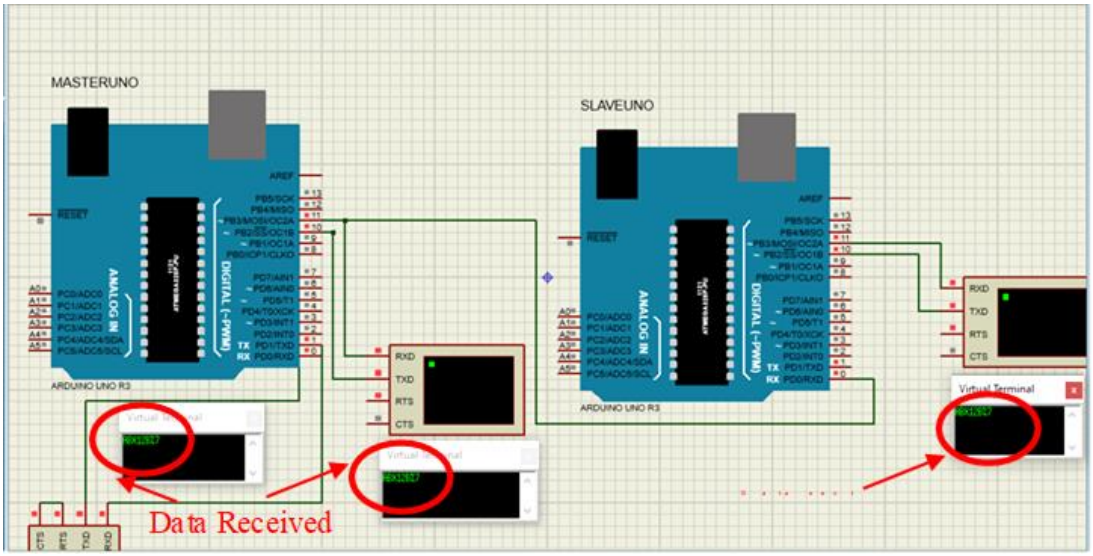


Figure 7. 1: Test 1 Result

Figure 7.2 below shows data monitored on the serial port being received after transmission. The port monitored was for the online system to determine if data from the offline would reach the intended destination.

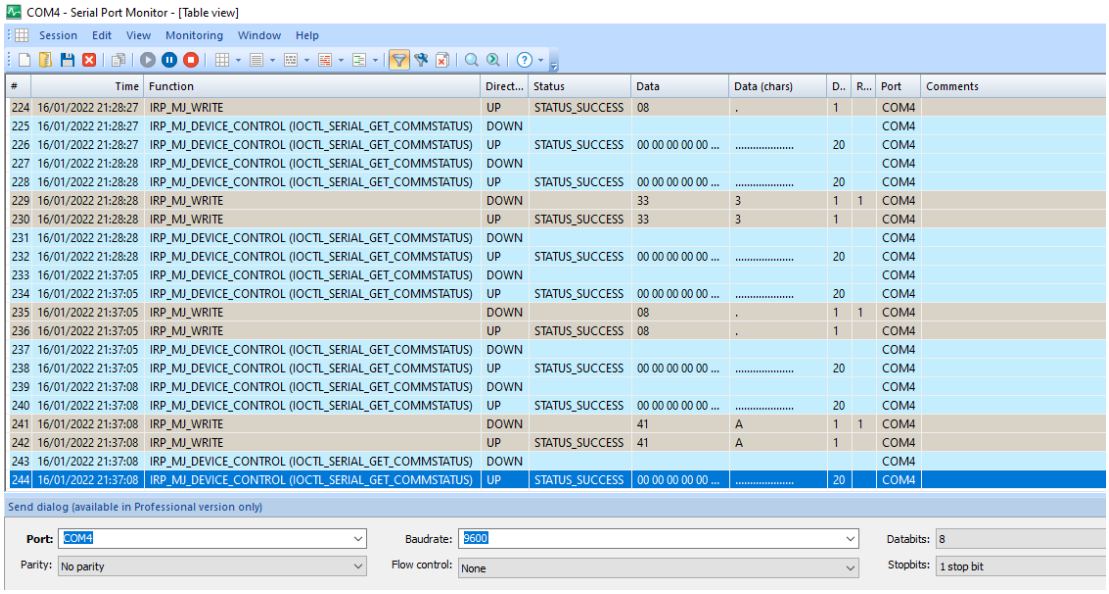


Figure 7. 2: Serial Data Monitored on Serial Port

Scenario 2 indicates that data could not be sent from the online system to the offline system. This is the desired result. No form of communication should take place from the internet to the offline system storing sensitive user data. This result enables our physical Hardware security system to protect data from unauthorized access by online users as well as malware attack such as ransomware. These attacks are normally

orchestrated by hackers who would be miles away launching the exploitation using their computers or other botnets. Preventing access to data from the internet would help stop such attacks from being successful. Blocking any form of access from the Internet to the offline system using hardware solutions achieves an effective solution that cuts across various system/application layers. That is the physical layer, network layer, application layer, and so on. Hence, regardless of the layer on which a hacker's tool operates, it is rendered ineffective as the possibility of the tool exploiting the weaknesses of the target system would have been curtailed since the prevention of access is physical. Figure 7.3 below shows that the data sent did not reach the intended target.

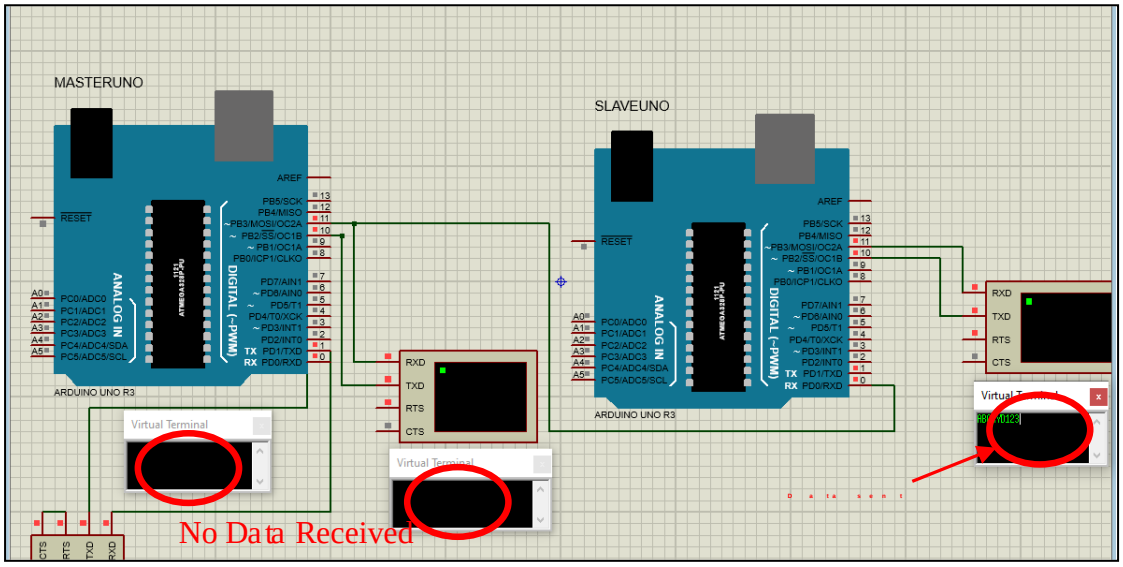


Figure 7. 3: Test 2 Result

Result 3 shows that restricting bandwidth would adversely affect the speed at which data can be transmitted. This is the desired result to prevent the transfer of huge amounts of sensitive data from the offline system to the online system. It would be frustrating for a hacker to wait for 100 days to transmit 10 GB of data. One of the objectives of securing systems is either preventing perpetrators from accessing the system or delaying them to an extent whereby the time they access the desired data, the data has become useless. For example, if it takes 10 years to transfer reasonable amounts of data, it becomes useless as the desired usage might not wait for 10 years for the exploit to materialize. The graph in Figure 7.4 below gives a feel of how long it would take to transmit various amounts of data across the RMS at the given speed.

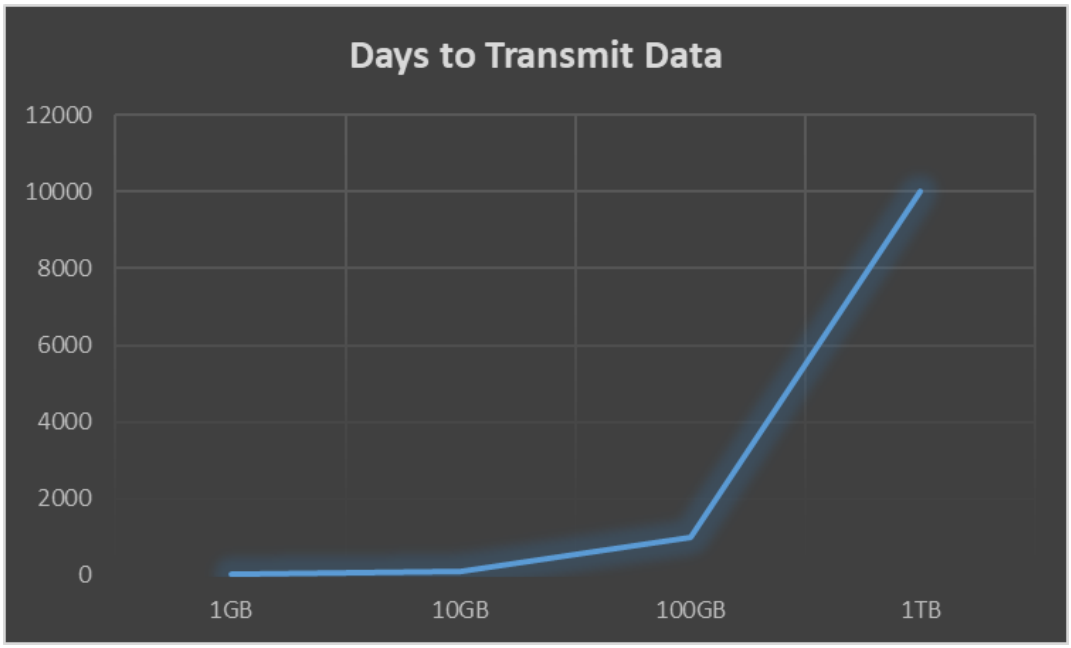


Figure 7. 4: Data Transmission Duration

Result 4 shows that Random IDs can be generated. This is very important for this solution to achieve its intended objectives. Using static IDs would easily give away the privacy of the user via profiling. A continuously changing ID would help hide the profile of the user and hence guard their privacy while online. It would be very difficult to profile a user who keeps changing their IDs when transacting online. The result is achieved via the use of existing solutions with minimal modification so that existing platforms can still be used when adopting the solution.

Result 5 is equally important. The main reason Pseudonymization was preferred to Anonymization, was to make it possible to decipher the original ID of the user from the Random ID if the need arises. Some users might take advantage of the fact that they are anonymous while transacting online. They might engage in fraudulent activities believing they cannot be traced. This result shows that despite being anonymous while online, the Agency could decipher the original ID if the need arose. This can help cab fraud by users knowing they would be traced if they misbehaved. It must be further stated that the Agency can decipher because they have additional information not accessible online hence protecting the privacy of the users. Moreover, the algorithms used to decipher the actual users are kept by the agency. Even if the algorithms were leaked to the public, the actual identifying data is kept on the offline system hence the

actual identity of the users would not be revealed thereby protecting and preserving the privacy of users.

Last, but not least, result 6 shows a very critical part of the solution. One cannot connect from the online microcontroller to the offline one. This is very vital as one of the main objectives of the solution is to prevent unauthorized access from internet users. Usually, hackers attack the internet. This goes to show that the RMS can protect our offline system despite the hardware security solution facing the internet being compromised. Multilayered hardware security comes into play.

7.2 Offline and Online System Backend Results

The screen shot below in Figure 7.5 is a display of what is happening in the backend of the system. It shows some of the elements captured as part of KYC data. Fingerprint, email address name, and so on.

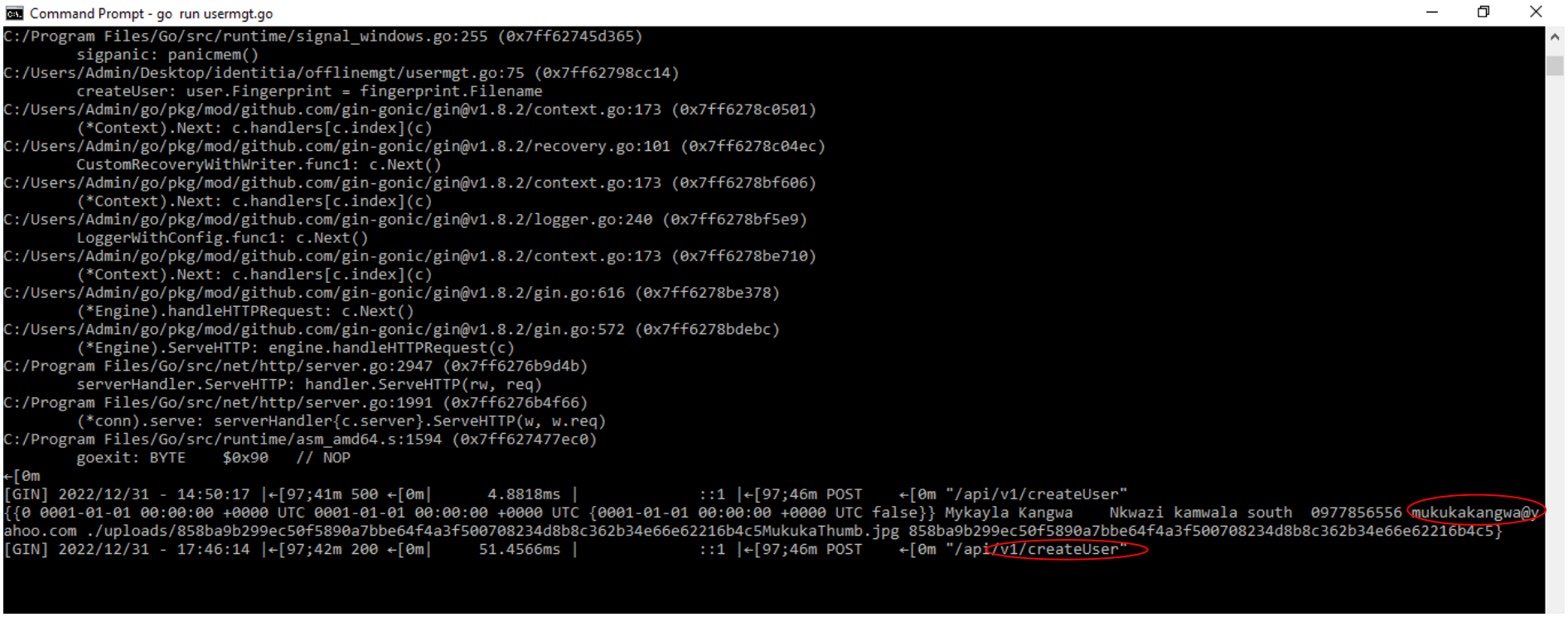


Figure 7. 5: Data being committed in the backend system

The screen shot below in Figure 7.6 shows the face of the code generator. This is the system used by users to generate their random IDs.

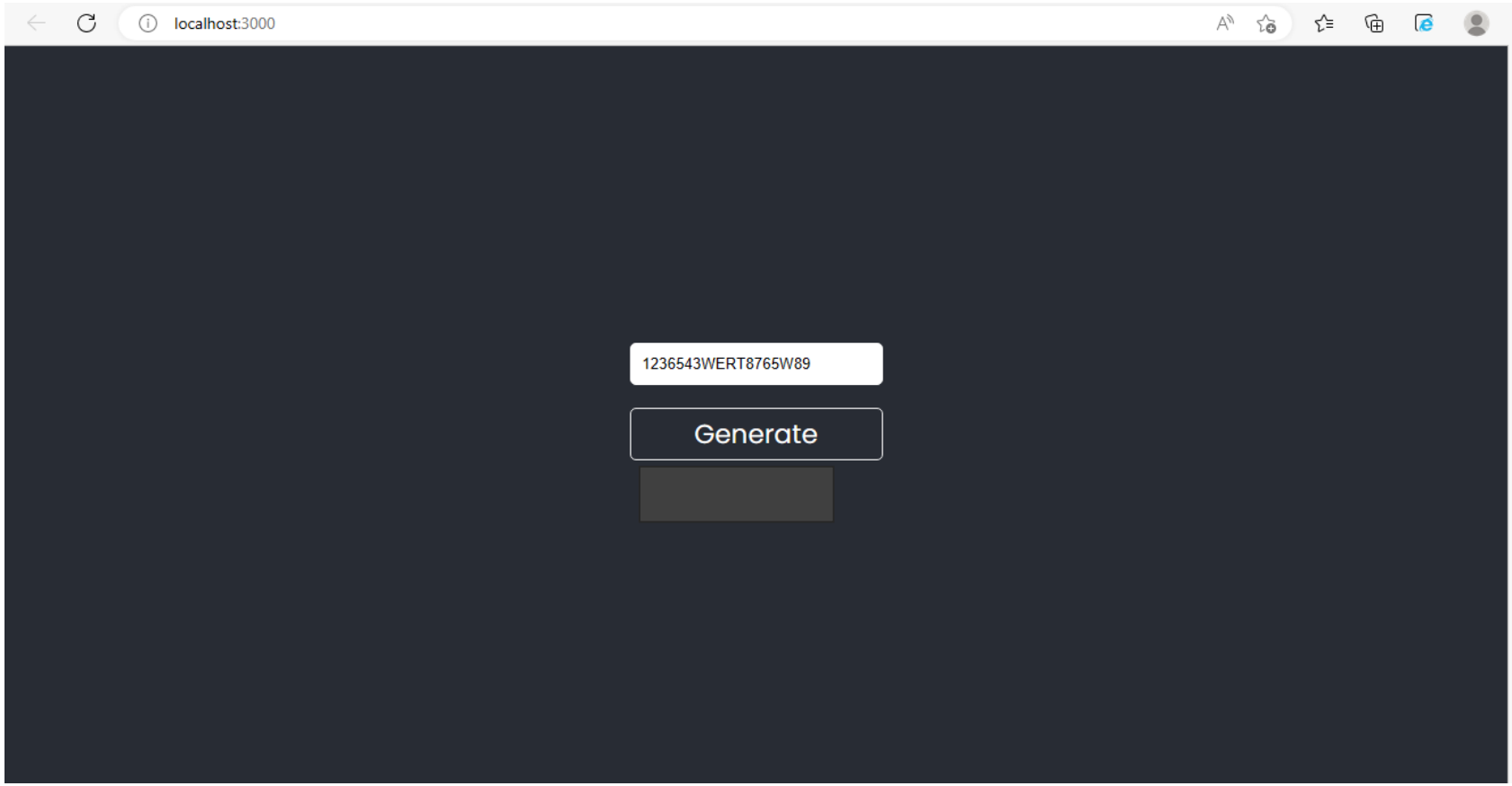


Figure 7. 6: Random ID Generator

Online API for User confirmation

The screen shot below in Figure 7.7 shows the online API system responding to user confirmation queries.

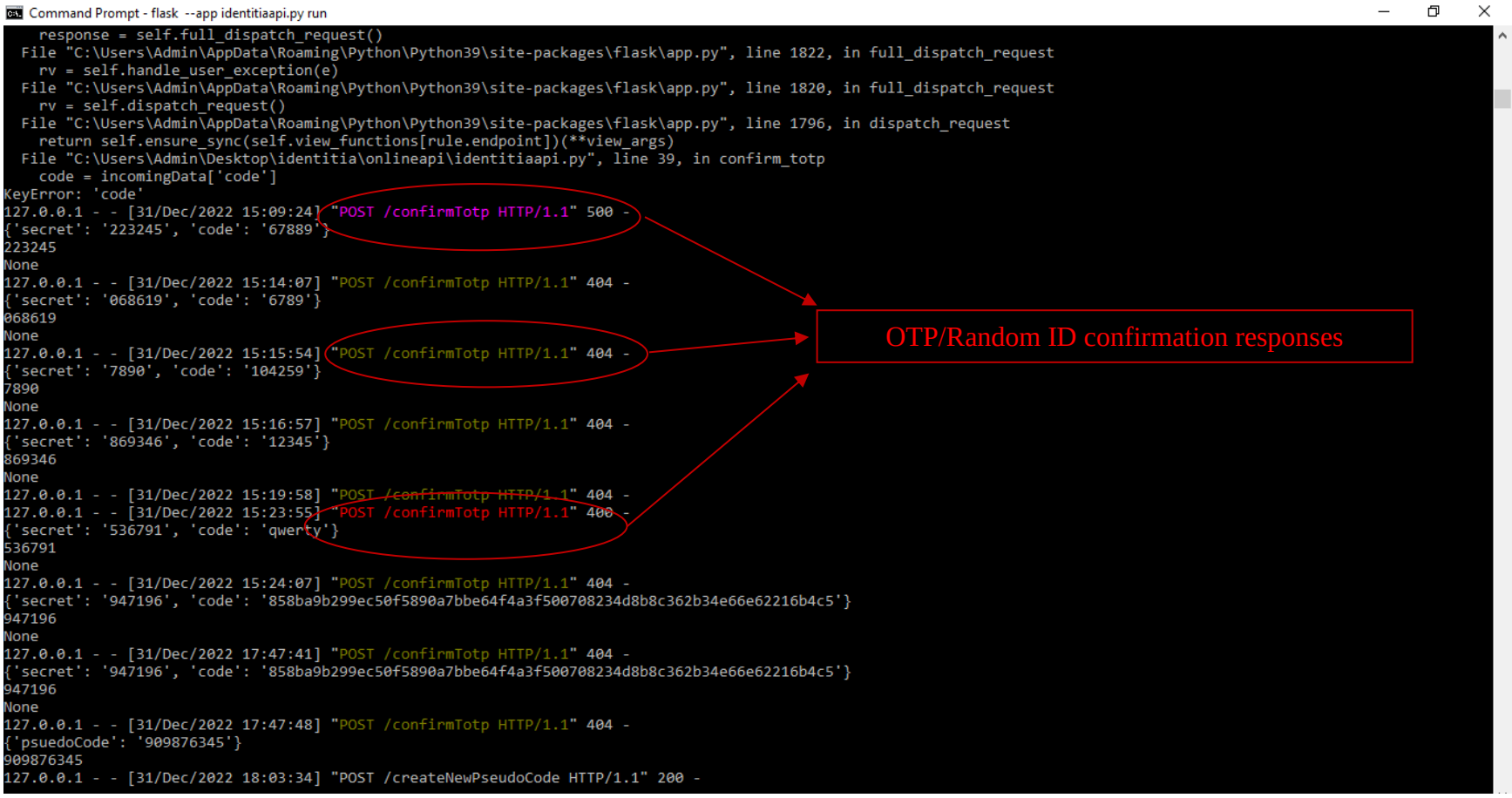


Figure 7. 7: Online API Responses

7.3 **Comprehensive Model Validation Results**

Table 7. 2: Results for the End-to-end Validation of the Prototype

Detailed System Test Results				
	Requirement-Registration System	**User Story**	**Expected Result**	**Actual Result**
1	Login page	I want to be able to logon to the system using a combination of static password and AI based Captcha	4. Land on a page that will request for a static password and username. 5. Ask user to prove they are human by presenting a Captcha challenge. 6. Grant access if provided respond to Captcha and password-username combination is correct	PASS Only provided User authentication as Captcha service would require the system to be online
2	Registration Portal	I would like to be able to register a new user onto the KYC system by capturing all the required details	2. Have a portal with all required details to be captured: j. First Name k. Surname l. Date of Birth m. Village and chief n. Physical/Home Address o. Capture live Photo of applicant. p. Capture reference letters and details of referee	PASS- Concept was demonstrated. Additional fields can be added when needed

Table 7. 3: Results for the End-to-end Validation of the Prototype Continued

			q. Capture contact details: Phone numbers, email addresses and state preferred mode of communication. r. Capture any other critical KYC data such as biometric data (fingerprint, face etc)	
3	Generation of Pseudo ID	The system should generate a universally unique electronic ID for the successful applicant	9. If all details are correct, register member and generate universally unique electronic ID and append to their record on the offline database. 10. Generate a Pseudo ID based on the electronic ID generated and send to the online System. 11. DO NOT append Pseudo ID to record on the offline database system	PASS- concept was demonstrated. Actual implementation can be varied depending on application
4	Online System Record creation	Once a new record is received from the Offline system create an online minimized account for the user with Pseudonymized data	12. Create a new account with the Pseudo ID as the key for the record. 13. Append scrambled Mobile Phone number for the user. 14. The scrambled Pone number must be stored under s filed called serial or anything other than phone number. The scrambling must replace numbers with letters and special characters not with numbers. 15. Create Pseudo email box for the user and communicate to user on site. 16. Issue user with key they will use to setup App for generating Random electronic IDs	PASS- Concept was demonstrated. Actual implementation can vary depending on the need

Table 7. 4: Results for the End-to-end Validation of the Prototype Continued

	Requirement-RMS System	User Story	Expected Result	Actual Result
1	Communication Mode	The system must setup in such a way that data will only flow in one direction at a restricted speed	6. Data should only flow from the offline system to the online system 7. Only the Pseudo IDs and the scrambled Phone numbers must be sent to the online system. 8. Microcontrollers are to be used for this purpose. 9. Bandwidth must be set to 9600bps. 10. The Microcontrollers must be set in serial mode half duplex	PASS Concept was demonstrated successfully
2	Multi-layered Hardware	The system must be setup with multiple hardware layers to enhance hardware security	7. The RMS must be built from two microcontrollers. 8. They must be connected via serial. 9. The bandwidth between them must be 9600bps or less. 10. They should be configured to communicate in half duplex mode. 11. The microcontroller connecting directly to the offline system must be configured as the Master while the one connecting to the online system must be configured as the slave. 12. They communication WIRE from the receiving offline side of the system to the sending side of the online system MUST NOT be connected	PASS System was built from two Arduino controllers. Can also be achieved using other controller types
3	Periodic connection resets	The connection between the Master microcontroller and slave must periodically reset	5. The connection between the Master and slave microcontrollers must reset every 8 hours. 6. If data transmission persists for more than 20 seconds, it must reset	This can be implemented in Real life

Table 7. 5: Results for the End-to-end Validation of the Prototype Continued

	Requirement- Random ID generating System	User Story	Expected Result	Actual Result
1	Provide an App to be used for Authentication	I would like to be able to generate Random IDs using an authentication app like Google Authenticator	6. The App must be based on TOTP using SHA 2 7. The App must be able to generate an OTP as a Random ID for the user using the combination of the standard RFC 6238 TOTP based on UNIX epoch time and the generated Pseudo ID sitting on the online system. 8. The Resulting modified TOTP (Random ID) can be delivered via the Mobile App, Phone, or pseudo email box. 9. The App must allow user to setup the account on the app using the key they were issued by the KYC agency. 10. The App should not require communication with server to generate Random ID.	PASS- An application was developed using React for concept validation
2	Random ID generation	The system must generate the Random ID based on the RFC 6238 standard	5. Generate the TOTP based on the RFC 6238 standard. 6. Combine the TOTP with the user's Pseudo ID 7. The resulting TOTP is the Random user ID to be delivered to the user's preferred channel. Recommended is the Mobile App 8. Create Algorithm that can decipher the Pseudo ID from the resulting TOTP	PASS
3	Algorithms to be created	I would like to be able to decipher the Pseudo ID from the random ID when need arise in case of fraud	3. Create Algorithm to only obtain the Pseudo ID from a random ID issued by a user to an online platform. 4. Create an Algorithm that will be able to generate the actual ID from the Pseudo ID.	PASS

7.4 Chapter Summary

The results obtained indicate that the proposed solution can effectively protect PII. This solution can be extrapolated to any other sensitive data-carrying system. Of course, the actual implementation might need some modification, using the same principle, depending on how the system operates and the services it renders.

The system needed to allow the transfer of data from the offline system to the online system to ensure there are no physical media used to transfer data from the offline system to the online system as this can result in leakage of data as well as the introduction of malware via external media. Further, it would be inefficient to be transferring pseudo equivalent IDs of newly created records onto the online platforms manually.

The blocking of data flow from the online system to the offline system was equally important. Most attacks ate started from the internet into the local systems. Blocking any form of data flow into the offline system helps to ensure no hack can access sensitive data and also prevents malware from being introduced that would compromise the sensitive data. The hacker requires physical access to the offline hardware to access the data hence making it almost unattainable and thereby protecting sensitive data.

Furthermore, the use of random IDs also helped in enhancing the protection of users while online. The constantly changing electronic user ID makes it difficult to profile the user hence making it difficult to compromise their privacy. Despite the user accessing online platforms using random IDs, the solution demonstrated that it is possible to identify the actual user when the need arises. What must be comforting to the users is that this is only possible by the platform owners supplying the random ID used to the Trusted Party and this need would only arise if the user is suspected to have been involved in fraudulent activities while online.

The capturing of Fingerprints and using them to generate Pseudo IDs greatly helped to enhance the privacy and security of the users. Fingerprints are unique universally hence the IDs created would remain unique to the user and would be very difficult to guess or spoof as the fingerprint input would be different from one user to another. To prevent

access to sensitive data via direct access by database administrators or even other staff that have authorized access to the system, the data is masked in the database. This helps prevent the leakage of data by internal staff.

The use of multi-layered hardware enhances security from the hardware Trojans. Even if some HT were hidden in the microcontroller, one would need access to a microcontroller to activate the HT. However, even if the HT didn't require physical or remote access to be activated, it would still be difficult for the HT to send data as the communication is physically configured to be one-way. In addition, there is a bandwidth limitation between the two microcontrollers forming the RMS.

Chapter Eight

8.0 Conclusion

The challenge of data leakage remains persistent despite several solutions being proposed and implemented to cab the vice. Almost every single day comes with new data breaches where sensitive PII is leaked into the wrong hands. Leakage of personal data puts the data owners at risk; their privacy is compromised, and they can be defrauded financially or even get harmed physically.

PII can be used to gain access to Bank accounts via platforms such as mobile Banking, Internet Banking, USSD, and many other electronic channels. In the spirit of improving service delivery, most service providers have decided to enable self-service on their electronic platforms. If, for example, someone forgets their password or account PIN, some platforms allow a user to recover their login credentials using their PII such as National Registration Number, Residential address, phone number, and so on. If one got access to this information, then there is a possibility that they can try to access the victims banking online accounts. Further, the residential address information can be used to track where the victim is found, and physical harm can be done to them if the perpetrator wished so.

The possibility of PII being leaked has contributed to some users being reluctant in adopting online platforms for conducting business transactions such as the purchasing and selling of goods and services. Lack of trust in online platforms affects their adoption in some countries such as those in developing regions.

The continuous cases of data leakages, even with several solutions in the market and deployed into real-life systems, clearly indicate more effective solutions and approaches are required. This study proposed a novel approach and solution to enhance user privacy and data protection for e-commerce users. The principles proposed are universal and can be applied to other data processing and storing systems.

The use of the multi-layered hardware by the solution was very critical to address both software and hardware data security challenges. The use of multiple hardware (in this case two microcontrollers) helped to address the prevalent HT that is usually ignored by software designers. Software designers assume the Hardware works the way it was

designed to operate. However, research has shown that many pieces of hardware are supplied with HT that can compromise the safety of data. Some testing does not address the HT problem introduced at the manufacturing stage. Hence the need to design solutions that will ensure that even if HTs are introduced during manufacture and are not picked up or discovered by independent testers, their destructive objectives do not materialize. Our design addresses this challenge via the use of multi-layered hardware.

The experiment results show that it is possible to protect PII from hackers by preventing any possibility of data being accessed. Since no online user can reach the offline system holding sensitive data, the system is more secure. Enhanced protection is achieved because no one would be able to access the offline system from the internet as the separation is physical. In addition, even if someone breached the security of the online system, they would need physical access to the offline side of the data protector to configure it to accept and allow the transfer of data to the offline system.

The restricted amount of data that can be sent via the data protector (RMS) is a huge deterrent to would-be data criminals as the time it would take would render the exercise futile. For example, it would take more than 100 days to transfer about 100 GB of data and hence would discourage the perpetrator. This is a small amount of data.

Furthermore, the use of the modified RFC6238 standard for the generation of random IDs makes it possible for the user to maintain their privacy while at the same time providing a possibility to trace a fraudster hiding behind being anonymous whenever the need arises. Random IDs help protect the user from profiling that would compromise their privacy.

The system was designed and built to ensure a fraudulent user can still be traced and identified if they were to misbehave. This is to promote the legitimate use of the privacy-enhancing platform. The Agency would help with the identification as they would be the only agency with identifying data on the user. The online platform service providers would only provide the Random IDs used at a given time

The use of TOR can further help achieve enhanced privacy for users. The modification of existing standards and platforms would help promote the adoption of the solution and existing platforms can be used with minimal modification to achieve enhanced security of PII.

The proposed identification format can be universally adopted as it takes into account the various countries that exist globally as well as the global population. The provision even looks at the possibility of new countries emerging from some countries splitting for whatever reasons.

Furthermore, the use of Fingerprints as one of the inputs into the generation of the Pseudo ID adds even more security as fingerprints are unique to each person hence making the individual Pseudo ID universally unique. In addition, the security of the data stored in the database is enhanced by the masking of the key data elements. This is to prevent physical access to data by database administrators who connect to the databases directly.

Successful implementation of the ODMS in real life would greatly help resolve the prevalent challenge of PII being leaked into the wrong hands now and then. The leakage of PII has resulted in some users suffering financial and reputational damage. Some users are discouraged from using e-commerce platforms because of the possibility of having their private information leaked. If the challenge is resolved and the users are made to feel comfortable using the online systems, more users will adopt e-commerce. The use of technology in performing transactions adds efficiency and reduces the cost of some of the services offered. For example, the cost of shipment for an audio CD can be transferred to the users as a saving.

The resolution of the PII leakage challenge is also expected to reduce the number of fraudulent activities performed online as the required data to commit such crimes will become scarce.

Table 8.0.1 below gives a summary of comparisons of various critical performance parameters the prototype built and other existing solutions. The summary indicates that the prototype improves the current systems.

Table 8.0.1 Summary of Comparisons with Other Similar Solutions

Comparisons with other Similar Solutions		
	Our Research	Others
1. Scalability	Yes-a universal ID format designed to cater for everyone worldwide	Mostly focused on countries of authors Use of Blockchain currently not scalable as the technology is resource intensive
2. Cost	Reasonable as it rides on existing infrastructure with very minimal changes. Cheaper in comparison to similar solutions in literature reviewed	Costly due to use of blockchain which is resource intensive
3. Compliance to standards and regulations	Complies to National and International regulations such as GDPR	comply
4. Reputation	Random IDs create a level of trust as users' privacy is not compromised by profiling of their ID	Static IDs (Pseudo or Real) can easily be profiled, and privacy compromised
5. Use of TTP Agency	Yes-Full data stored offline Minimized data stored online for real time transactions	Yes-Full data stored online in the cloud
6. Performance	• Same as the current solutions in use depending on connectivity and specifications of equipment used. • Faster. User is given service to generate random IDs that are time sensitive and request for service directly. The service provider confirms with TTP and TTP responds without asking user to confirm as the ID and user are known based on random ID provided. • Just seconds depending on network	• Dependency on Blockchain slows down the process as several nodes are needed to confirm transaction as complete. • Involves two cycles for one transaction. The user request for a service from provider, the provider ask the Static ID to confirm if user is genuine, then the static provider as user to confirm then responds to service provider who then responds to the user. Too long a process. • Distributed ledger use slows and complicates solution • Takes about 15minutes or more for Bitcoin ledger

8.1 **Research Questions**

Coming back to our research questions, the following were conclusions against each one of them:

(i) What model can be formulated to enhance the prevention of PII data leakage for e-commerce users?

A model employing Offline Data minimization and Online Pseudonymization coupled with multi-layered hardware using a random ID for user access to preserve user privacy was formulated.

(ii) How can the proposed model be implemented and validated for its effectiveness in preventing PII leakage?

The Model in (i) was validated by building a simulation for the Data Protector module. The module was tested using simulators and the results produced were positive as desired.

(iii) How effective is the formulated and implemented model in protecting PII and the privacy of its users?

From the various test results obtained, the prototype was able to protect the PII as well as preserve the privacy of the users.

8.2 Recommendations

Following the successful design and implementation of the solution to prevent PII leakage: the following recommendations are made:

(i) A Trusted Third Party be appointed as the sole issuer of National and by implication electronic IDs. This TTP will be responsible for the collection of KYC information required by various online service providers as well as other institutions such as Banks that collect KYC data and store it on their systems.

(ii) Other institutions to stop collecting PII from prospective customers. Instead, start confirming KYC compliance with TTP. The TTP would not surrender any PII for the user but only confirm or deny knowing the user based on the random ID or National ID supplied.

(iii) Let the TTP establish standards that would define the KYC information to be collected. This must be a consultative process with all key stakeholders to ensure data collected by the TTP meets the needs of various services.

(iv) The proposed global identity be implemented to ensure every human being can be represented electronically with non-repudiation.

(v) The Random ID solution be implemented to ensure the privacy of online service consumers.

8.3 **Future Works**

The research conducted helped to reveal areas that need further investigation, and these include the following:

(i) There is already a lot of PII out in cyberspace. There is a need to formulate and implement an effective way for individual users to clean the internet of their leaked PII. Our research was focused on how to protect PII and provide user privacy in e-commerce.

(ii) Further investigations must be done to determine the possibility of reducing the length of the Random eID to about 6 to 4 digits without losing user traceability in case of fraud.

(iii) Investigate ways this solution can be extended to social media without compromising the objective of social media; that is being able to socialize as it is currently happening with “real” identities.

References

[1] M. Kangwa, C. S. Lubobya, and J. Phiri, “Protection of personally identifiable Information and Privacy via the use of Hardware and Software,” vol. 0958, 2021.

[2] S. Mekhmonov and M. Temirkhanova, “Development of Electronic Commerce in the Republic of Uzbekistan.,” *Int. J. Financ. Manag. Econ.*, vol. 78, no. 10, pp. 39–45, 2020.

[3] R. Ahmed, “Ecommerce in Pakistan: Challenges & Opportunities,” *Proc. Eighteenth Wuhan Int. Conf. E-bus.*, pp. 592–601, 2019.

[4] K. Kumain, P. Chaudhary, and N. Joshi, “E-Commerce Security Issues and Role of AI: A Review,” *Int. J. Manag.*, vol. 11, no. 10, pp. 504–509, 2020.

[5] K. C. Laudon and C. G. Traver, *E-commerce business, technology and society*, vol. 17. 2022.

[6] C. Robinson, “Disclosure of personal data in ecommerce: A cross-national comparison of Estonia and the United States,” *Telemat. Informatics*, vol. 34, no. 2, pp. 569–582, 2017.

[7] N. A. Bt Mohd and Z. F. Zaaba, “A review of usability and security evaluation model of E-commerce website,” *Procedia Comput. Sci.*, vol. 161, pp. 1199–1205, 2019.

[8] S. W. Khan, “Cyber Security Issues and Challenges in E-Commerce,” *SSRN Electron. J.*, p. 2019, 2019.

[9] W. Yanyan and I. Engineering, “m nl ad in e e V by e th rsio is n fil O e is nly Bo m nl ad in e e V by e th rsio is n fil O e is nly,” vol. 8, no. 3, pp. 153–162, 2014.

[10] Z. Song, Y. Sun, J. Wan, L. Huang, and J. Zhu, “Smart e-commerce systems: current status and research challenges,” *Electron. Mark.*, vol. 29, no. 2, pp. 221–238, 2019.

[11] F. Prasser, K. A. Kuhn, and J. Eicher, “Flexible data anonymization using ARX — Current status and challenges ahead,” no. January, pp. 1277–1304, 2020.

[12] M. Yamac *et al.*, “Multi-level Reversible Data Anonymization via Compressive Sensing and Data Hiding,” vol. 14, no. 8, 2020.

[13] P. Prinetto and G. Roascio, “Hardware security, vulnerabilities, and attacks: A comprehensive taxonomy,” *CEUR Workshop Proc.*, vol. 2597, no. August, pp. 177–189, 2020.

[14] G. Van Blarkom, J. J. Borking, and J. G. E. Olk, “Handbook of privacy and privacy-enhancing technologies,” *Priv. Inc. Softw. ...*, pp. 42–50, 2003.

[15] C. Eyupoglu, M. A. Aydin, A. H. Zaim, and A. Sertbas, “An efficient big data anonymization algorithm based on chaos and perturbation techniques,” *Entropy*, vol. 20, no. 5, pp. 1–18, 2018.

[16] G. Bansal, F. M. Zahedi, and D. Gefen, “Do context and personality matter? Trust and privacy concerns in disclosing private information online,” *Inf. Manag.*, vol. 53, no. 1, pp. 1–21, 2016.

[17] G. Mazurek and K. Małagocka, “Are we down to zero-one code ? Perception of privacy and data protection in the context of the development of artificial intelligence,” *J. Manag. Anal.*, vol. 0, no. 0, pp. 1–21, 2019.

[18] A. Majeed, “Anonymization Techniques for Privacy Preserving Data Publishing : A Comprehensive Survey,” vol. 9, 2021.

[19] P. Brief, P. Legislation, and P. Brief, “Data Privacy in the Indonesian Personal Data Protection Legislation,” no. 7, 2021.

[20] M. Hasal, J. Nowaková, V. Snášel, L. Ogiela, K. A. Saghair, and H. Abdulla, “Chatbots : Security , privacy , data protection , and social aspects,” no. May, pp. 1–13, 2021.

[21] D. S. Vishal, “Emerging Roles of Artificial Intelligence in ecommerce,” *Int. J. Trend Sci. Res. Dev.*, vol. 4, no. 5, pp. 223–225, 2020.

[22] M. D. P. Frank A Cona, “" Digital Identity " Personal Dato,” US 2019 / 0333054 A1, 2019.

[23] X. Wang and X. Peng, “Research on data leak protection technology based on trusted platform,” *Open Autom. Control Syst. J.*, vol. 6, no. 1, pp. 919–926, 2014.

[24] J. Chicaiza, M. C. Cabrera-Loayza, R. Elizalde, and N. Piedra, “Application of data anonymization in Learning Analytics,” *ACM Int. Conf. Proceeding Ser.*, 2020.

[25] A. Ali, F. Jamil, M. Saleh, A. Muthanna, and M. Ammi, “AI-Enabled Cloud Security Based on Organized Identity System,” no. June, 2022.

[26] O. Lynskey, “Grappling with ‘ Data Power ’: Normative Nudges from Data Protection and Privacy,” vol. 189, pp. 189–220, 2019.

[27] T. Hoel and W. Chen, “Privacy and data protection in learning analytics should be motivated by an educational maxim — towards a proposal,” 2018.

[28] G. Chassang, “The impact of the EU general data protection regulation on scientific research,” *Ecancermedicalscience*, vol. 11, pp. 1–12, 2017.

[29] E. Di Minin, C. Fink, A. Hausmann, J. Kremer, and R. Kulkarni, “How to address data privacy concerns when using social media data in conservation science,” *Conserv. Biol.*, vol. 35, no. 2, pp. 437–446, 2021.

[30] G. Danezis *et al.*, *Privacy and Data Protection by Design - from policy to engineering*, no. December. 2015.

[31] N. Soudani, B. G. Raggad, and B. Zouari, “A formal design of secure information systems by using a formal secure Data Flow Diagram (FSDFD),” *Post-Proceedings 4th Int. Conf. Risks Secur. Internet Syst. Cris. 2009*, pp. 131–134, 2009.

[32] Z. A. Mohammed and G. P. Tejay, “Examining privacy concerns and ecommerce adoption in developing countries: The impact of culture in shaping individuals’ perceptions toward technology,” *Comput. Secur.*, vol. 67, pp. 254–

265, 2017.

[33] Q. Ji, “Study on Information Security Issues of E-Commerce,” *IOP Conf. Ser. Mater. Sci. Eng.*, vol. 452, no. 3, 2018.

[34] P. Jílková and P. Králová, “Digital Consumer Behaviour and eCommerce Trends during the COVID-19 Crisis,” *Int. Adv. Econ. Res.*, vol. 27, no. 1, pp. 83–85, 2021.

[35] M. Pejić-Bach, “Editorial: Electronic commerce in the time of covid-19 - Perspectives and challenges,” *J. Theor. Appl. Electron. Commer. Res.*, vol. 16, no. 1, p. I, 2021.

[36] J. Tagliabue *et al.*, *SIGIR 2021 E-Commerce Workshop Data Challenge*, vol. 1, no. 1. Association for Computing Machinery, 2021.

[37] S. Akter and S. F. Wamba, “Big data analytics in E-commerce: a systematic review and agenda for future research,” *Electron. Mark.*, vol. 26, no. 2, pp. 173–194, 2016.

[38] P. Voigt and A. von dem Bussche, *Introduction and ‘Checklist.’* 2017.

[39] H. R. Pawar and D. G. Harkut, “Classical and Quantum Cryptography for Image Encryption Decryption,” *Proc. 2018 3rd IEEE Int. Conf. Res. Intell. Comput. Eng. RICE 2018*, pp. 1–4, 2018.

[40] B. Hauer, “Data and information leakage prevention within the scope of information security,” *IEEE Access*, vol. 3, pp. 2554–2565, 2015.

[41] M. Naarttijärvi, “Balancing data protection and privacy – The case of information security sensor systems,” vol. 000, pp. 1–20, 2018.

[42] L. Coppolino, S. D’Antonio, G. Mazzeo, and L. Romano, “A comprehensive survey of hardware-assisted security: From the edge to the cloud,” *Internet of Things*, vol. 6, p. 100055, 2019.

[43] H. Mutlaq, A. Id, A. A. Norman, and B. H. Ahmed, “Privacy and data protection in mobile cloud computing : A systematic mapping study,” pp. 1–28,

2020.

[44] J. Reardon, D. Basin, and S. Capkun, “SoK: Secure data deletion,” *Proc. - IEEE Symp. Secur. Priv.*, pp. 301–315, 2013.

[45] J. Reardon, D. Basin, and S. Čapkun, “On secure data deletion,” *IEEE Secur. Priv.*, vol. 12, no. 3, pp. 37–44, 2014.

[46] F. Xhafa, X. Chen, and L. Barolli, “Editorial preface for the special issue ‘Advances in security, privacy and trust technologies,’” *J. Ambient Intell. Humaniz. Comput.*, vol. 6, no. 5, pp. 531–532, 2015.

[47] L. H. Wei, M. A. Osman, N. Zakaria, and T. Bo, “Adoption of e-commerce online shopping in Malaysia,” *Proc. - IEEE Int. Conf. E-bus. Eng. ICEBE 2010*, pp. 140–143, 2010.

[48] T. M. Shannon, “THE LIMITATIONS OF EUROPEAN DATA PROTECTION AS A MODEL FOR GLOBAL PRIVACY REGULATION,” vol. 3, no. L 119, pp. 20–25, 2020.

[49] M. Mourby *et al.*, “Are ‘pseudonymised’ data always personal data? Implications of the GDPR for administrative data research in the UK,” *Comput. Law Secur. Rev.*, vol. 34, no. 2, pp. 222–233, 2018.

[50] S. Schwerin, “Blockchain and Privacy Protection in Case of The European General Data Protection Regulation (GDPR): A Delphi Study,” vol. 1, no. 1, pp. 1–76, 2018.

[51] A. Mukhopadhyay, S. Chatterjee, D. Saha, A. Mahanti, and S. K. Sadhukhan, “Cyber-risk decision models: To insure IT or not?,” *Decis. Support Syst.*, vol. 56, no. 1, pp. 11–26, 2013.

[52] Y. Go, O. Satoshi, and A. Kazumaro, “Fast Integrity for Large Data,” *Softw. Perform. Enhanc. Encryption Decryption*, pp. 21–32, 2007.

[53] M. Cremonini, C. Braghin, and C. Agostino Ardagna, *Privacy on the Internet*. Elsevier Inc., 2013.

[54] G. Navarro-Arribas and V. Torra, “Preface,” *Stud. Comput. Intell.*, vol. 567, pp. 423–442, 2014.

[55] Shakila and B. Pasha, “Location Data,Personal Data Protection and Privacy in Mobile Device Usage: An EU Law Perspective,” University of Helsinki, 2018.

[56] W. J. Buchanan, Z. Kwecka, and E. Ekonomou, “A privacy preserving method using privacy enhancing techniques for location based services,” *Mob. Networks Appl.*, vol. 18, no. 5, pp. 728–737, 2013.

[57] A. Burman, “Will India ’ s Proposed Data Protection Law Protect Privacy and Promote Growth ? Will India ’ s Proposed Data Protection Law Protect Privacy and Promote Growth ? Anirudh Burman,” no. March, 2020.

[58] S. Murthy, A. Abu Bakar, F. Abdul Rahim, and R. Ramli, “A Comparative Study of Data Anonymization Techniques,” *Proc. - 5th IEEE Int. Conf. Big Data Secur. Cloud, BigDataSecurity 2019, 5th IEEE Int. Conf. High Perform. Smart Comput. HPSC 2019 4th IEEE Int. Conf. Intell. Data Secur.*, pp. 306–309, 2019.

[59] H. Liu, “Research on Feasibility Path of Technology Supervision and Technology Protection in Big Data Environment,” *Proc. - 2019 Int. Conf. Intell. Transp. Big Data Smart City, ICITBS 2019*, pp. 293–296, 2019.

[60] S. Alneyadi, E. Sithirasenan, and V. Muthukkumarasamy, “A survey on data leakage prevention systems,” *J. Netw. Comput. Appl.*, vol. 62, pp. 137–152, 2016.

[61] A. Muneer, R. S, and F. Z, “Data Privacy Issues and Possible Solutions in E-commerce,” *J. Account. Mark.*, vol. 07, no. 03, 2018.

[62] N. Gruschka, V. Mavroeidis, K. Vishi, and M. Jensen, “Privacy Issues and Data Protection in Big Data: A Case Study Analysis under GDPR,” *Proc. - 2018 IEEE Int. Conf. Big Data, Big Data 2018*, pp. 5027–5033, 2019.

[63] Y. Canbay, Y. Vural, and S. Sagiroglu, “Privacy Preserving Big Data Publishing,” *Int. Congr. Big Data, Deep Learn. Fight. Cyber Terror.*

IBIGDELFT 2018 - Proc., pp. 24–29, 2019.

[64] A. Yassine, A. A. Nazari Shirehjini, S. Shirmohammadi, and T. T. Tran, "Online information privacy: Agent-mediated payoff," *2011 9th Annu. Int. Conf. Privacy, Secur. Trust. PST 2011*, pp. 260–263, 2011.

[65] J. O. Ogbu and A. Oksiuk, "Information protection of data processing center against cyber attacks," *2016 3rd Int. Sci. Conf. Probl. Infocommunications Sci. Technol. PIC S T 2016 - Proc.*, no. August, pp. 132–134, 2017.

[66] G. Dhillon, T. Oliveira, and R. Syed, "Value-based information privacy objectives for Internet Commerce," *Comput. Human Behav.*, vol. 87, pp. 292–307, 2018.

[67] A. Azam, F. Qiang, and M. I. Abdullah, "Consumers' E-commerce acceptance model: Antecedents of trust and satisfaction constructs," *BEIAC 2012 - 2012 IEEE Business, Eng. Ind. Appl. Colloq.*, pp. 371–376, 2012.

[68] S. U. Bazai, J. Jang-Jaccard, and H. Alavizadeh, "Scalable, high-performance, and generalized subtree data anonymization approach for apache spark," *Electron.*, vol. 10, no. 5, pp. 1–28, 2021.

[69] R. Bild, K. A. Kuhn, and F. Prasser, "SafePub: A Truthful Data Anonymization Algorithm With Strong Privacy Guarantees," *Proc. Priv. Enhancing Technol.*, vol. 2018, no. 1, pp. 67–87, 2018.

[70] P. Mildenberger, *Basic Knowledge of Medical Imaging Informatics*. 2021.

[71] S. U. Bazai and J. Jang-Jaccard, "In-memory data anonymization using scalable and high performance rdd design," *Electron.*, vol. 9, no. 10, pp. 1–26, 2020.

[72] C. E. M. Jakob, F. Kohlmayer, T. Meurers, J. J. Vehreschild, and F. Prasser, "Design and evaluation of a data anonymization pipeline to promote Open Science on COVID-19," *Sci. Data*, vol. 7, no. 1, pp. 1–10, 2020.

[73] L. Bolognini and C. Bistolfi, "Pseudonymization and impacts of Big (personal/anonymous) Data processing in the transition from the Directive

95/46/EC to the new EU General Data Protection Regulation," *Comput. Law Secur. Rev.*, vol. 33, no. 2, pp. 171–181, 2017.

[74] K. N. Vokinger and D. J. Stekhoven, "Lost in Anonymization — A Data Anonymization Reference Classification Merging Legal and Technical Considerations Health Policy Portal," vol. 48, pp. 228–231, 2020.

[75] M. Tamper and A. Oksanen, "Anonymization Service for Finnish Case Law : Opening Data without Sacrificing Data Protection and Privacy of Citizens Anonymization Service for Finnish Case Law : Opening Data," 2018.

[76] J. Yoon, L. N. Drumright, and M. Van Der Schaar, "Anonymization through data synthesis using generative adversarial networks (ADS-GAN)," *IEEE J. Biomed. Heal. Informatics*, vol. 24, no. 8, pp. 2378–2388, 2020.

[77] T. Brekne, A. Årnes, and A. Øslebø, "Anonymization of IP traffic monitoring data: Attacks on two prefix-preserving anonymization schemes and some proposed remedies," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 3856 LNCS, pp. 179–196, 2006.

[78] S. Nimalaprakasan, S. Ramanan, B. A. Malalasena, K. Shayanthan, C. Gamage, and M. S. D. Fernando, "Privacy enhanced data management for an electronic identity system," *2009 Innov. Technol. Intell. Syst. Ind. Appl. CITISIA 2009*, no. July, pp. 358–363, 2009.

[79] R. Koch, M. Golling, L. Stiemert, and G. D. Rodosek, "Using Geolocation for the Strategic Preincident Preparation of an IT Forensics Analysis," *IEEE Syst. J.*, vol. 10, no. 4, pp. 1338–1349, 2016.

[80] H. Habibzadeh, T. Soyata, B. Kantarci, A. Boukerche, and C. Kaptan, "Sensing, communication and security planes: A new challenge for a smart city system design," *Comput. Networks*, vol. 144, pp. 163–200, 2018.

[81] C. Lee and G. Ahmed, "Improving IoT Privacy, Data Protection and Security Concerns," *Int. J. Technol. Innov. Manag.*, vol. 1, no. 1, pp. 18–33, 2021.

[82] M. Kangwa, C. S. Lubobya, and J. Phiri, "Prevention of Personally Identifiable

Information Leakage in E-commerce via Offline Data Minimisation and Pseudonymisation," *Int. J. Innov. Sci. Res. Technol.*, vol. 6, no. 1, pp. 209–212, 2021.

[83] Z. Bao, Q. Wang, W. Shi, L. Wang, H. Lei, and B. Chen, "When Blockchain Meets SGX: An Overview, Challenges, and Open Issues," *IEEE Access*, vol. 8, pp. 170404–170420, 2020.

[84] Z. II-Agure, A. Belsam, and C. Yun-ke, "The Semantics of Anomalies in IoT Integrated BlockChain Network," *IEEE*, pp. 144–146, 2019.

[85] P. Zhang, M. Alkubati, Y. Bao, and G. Yu, "Research advances on blockchain-as-a-service: architectures, applications and challenges," *Digit. Commun. Networks*, 2021.

[86] F. Ye, X. Dong, J. Shen, Z. Cao, and W. Zhao, "A Verifiable dynamic multi-user searchable encryption scheme without trusted third parties," *Proc. Int. Conf. Parallel Distrib. Syst. - ICPADS*, vol. 2019-Decem, pp. 896–900, 2019.

[87] N. Innab and A. Alamri, "The Impact of DDoS on E-commerce," *21st Saudi Comput. Soc. Natl. Comput. Conf. NCC 2018*, pp. 1–4, 2018.

[88] J. Zhan, X. Fan, L. Cai, Y. Gao, and J. Zhuang, "TPTVer: A trusted third party based trusted verifier for multi-layered outsourced big data system in cloud environment," *China Commun.*, vol. 15, no. 2, pp. 122–137, 2018.

[89] *IT Laws in the Era of Cloud Computing Nomos*. .

[90] M. Kaur and M. Mahajan, "Using encryption Algorithms to enhance the Data Security in Cloud Computing ," *Int. J. Commun. Comput. Technol.*, vol. 1, no. 2, pp. 130–133, 2019.

[91] A. Pansotra and S. P. Singh, "Cloud security algorithms," *Int. J. Secur. its Appl.*, vol. 9, no. 10, pp. 353–360, 2015.

[92] G. L. Prakash, M. Prateek, and I. Singh, "Data encryption and decryption algorithms using key rotations for data security in cloud system," *2014 Int. Conf. Signal Propag. Comput. Technol. ICSPCT 2014*, pp. 624–629, 2014.

[93] H. Zhang, F. Lou, H. Wang, and Z. Tian, “Research on Data Protection Based on Encrypted Attribute Access Control in Cloud Computing,” *Proc. - 2018 5th Int. Conf. Inf. Sci. Control Eng. ICISCE 2018*, pp. 450–453, 2019.

[94] A. Bhardwaj, G. V. B. Subrahmanyam, V. Avasthi, and H. Sastry, “Security Algorithms for Cloud Computing,” *Procedia Comput. Sci.*, vol. 85, no. Cms, pp. 535–542, 2016.

[95] P. Semwal and M. K. Sharma, “Comparative study of different cryptographic algorithms for data security in cloud computing,” *Proc. - 2017 3rd Int. Conf. Adv. Comput. Commun. Autom. (Fall), ICACCA 2017*, vol. 2018-Janua, pp. 1–7, 2018.

[96] W. Kuan Hon, C. Millard, and I. Walden, “The problem of ‘personal data’ in cloud computing: What information is regulated?-the cloud of unknowing,” *Int. Data Priv. Law*, vol. 1, no. 4, pp. 211–228, 2011.

[97] G. Kaur and M. Mahajan, “Analyzing Data Security for Cloud Computing Using Cryptographic Algorithms,” *Int. J. Eng. Res. Appl.*, vol. 3, no. 5, pp. 782–786, 2013.

[98] S. Nepal and M. Pathan, “Security, privacy and trust in cloud systems,” *Secur. Priv. Trust Cloud Syst.*, vol. 9783642385, pp. 1–459, 2013.

[99] P. Patil and B. Chiradeep, “Cloud Computing,” 2022. [Online]. Available: https://www.spiceworks.com/tech/cloud/articles/what-is-cloud-computing/. [Accessed: 15-Nov-2022].

[100] J. Zhang, B. Chen, Y. Zhao, X. Cheng, and F. Hu, “Data Security and Privacy-Preserving in Edge Computing Paradigm: Survey and Open Issues,” *IEEE Access*, vol. 6, no. c, pp. 18209–18237, 2018.

[101] J. Yang, “Computer Data Encryption System Based on Nonlinear Partial Differential Equations,” *Mob. Inf. Syst.*, vol. 2022, 2022.

[102] P. Kocher, R. Lee, G. McGraw, A. Raghunathan, and S. Ravi, “Security as a new dimension in embedded system design,” *Proc. - Des. Autom. Conf.*, pp.

753–760, 2004.

[103] D. Wendlandt, I. Avramopoulos, D. G. Andersen, and J. Rexford, “Don’t Secure Routing Protocols, Secure Data Delivery,” *5th ACM Work. Hot Top. Networks, HotNets 2006*, pp. 7–12, 2006.

[104] T. Locher, S. Obermeier, and Y. A. Pignolet, “When Can a Distributed Ledger Replace a Trusted Third Party?,” in *2018 IEEE International Conference on Internet of Things (iThings) and IEEE Green Computing and Communications (GreenCom) and IEEE Cyber, Physical and Social Computing (CPSCom) and IEEE Smart Data (SmartData)*, 2018, pp. 1069–1077.

[105] D. K. Aarthy, M. Aarthi, K. A. Farhath, S. Lakshana, and V. Lavanya, “Reputation-based trust management in cloud using a trusted third party,” *ICONSTEM 2017 - Proc. 3rd IEEE Int. Conf. Sci. Technol. Eng. Manag.*, vol. 2018-Janua, pp. 220–225, 2017.

[106] P. K. Jamshiya and D. M. Menon, “Design of a Trusted Third Party Key Exchange Protocol for Secure Internet of Things (IoT),” *Proc. Int. Conf. Inven. Commun. Comput. Technol. ICICCT 2018*, no. Icicct, pp. 1834–1838, 2018.

[107] N. S. Labib, M. R. Brust, and P. Bouvry, “Trustworthiness in IoT – A Standards Gap Analysis on Security , Data Protection and Privacy,” *2019 IEEE Conf. Stand. Commun. Netw.*, pp. 1–7, 2019.

[108] M. Suresh and M. Neema, “Hardware Implementation of Blowfish Algorithm for the Secure Data Transmission in Internet of Things,” *Procedia Technol.*, vol. 25, no. Raerest, pp. 248–255, 2016.

[109] R. Ande, B. Adebisi, M. Hammoudeh, and J. Saleem, “Internet of Things: Evolution and technologies from a security perspective,” *Sustain. Cities Soc.*, vol. 54, no. May, p. 101728, 2020.

[110] X. Teng, W. James B., and M. Potkonjak, “Security of IoT Systems: Design Challenges and Opportunities Teng,” pp. 417–423, 2014.

[111] D. Mazzei *et al.*, “A Blockchain Tokenizer for Industrial IOT trustless

applications," *Futur. Gener. Comput. Syst.*, vol. 105, pp. 432–445, 2020.

[112] P. Štarchoň and T. Pikulík, "GDPR principles in data protection encourage pseudonymization through most popular and full-personalized devices - mobile phones," *Procedia Comput. Sci.*, vol. 151, no. 2018, pp. 303–312, 2019.

[113] S. Goldwasser and G. N. Rothblum, "On Best-Possible Obfuscation," in *Theory of Cryptography*, 2007, pp. 194–213.

[114] I. You and K. Yim, "Malware Obfuscation Techniques: A Brief Survey," in *2010 International Conference on Broadband, Wireless Computing, Communication and Applications*, 2010, pp. 297–300.

[115] G. Sarada, N. Abitha, G. Manikandan, and N. Sairam, "A few new approaches for data masking," in *2015 International Conference on Circuits, Power and Computing Technologies [ICCPCT-2015]*, 2015, pp. 1–4.

[116] S. L. Ribeiro and E. T. Nakamura, "Privacy Protection with Pseudonymization and Anonymization in a Health IoT System: Results from OCARIoT," *Proc. - 2019 IEEE 19th Int. Conf. Bioinforma. Bioeng. BIBE 2019*, pp. 904–908, 2019.

[117] T. Neubauer and J. Heurix, "A methodology for the pseudonymization of medical data," *Int. J. Med. Inform.*, vol. 80, no. 3, pp. 190–204, 2011.

[118] G. M. Csányi, D. Nagy, R. Vági, J. P. Vadász, and T. Orosz, "Challenges and open problems of legal document anonymization," *Symmetry (Basel).*, vol. 13, no. 8, pp. 1–25, 2021.

[119] I. Goldberg, D. Wagner, and E. Brewer, "Privacy-enhancing technologies for the Internet," *Dig. Pap. - COMPCON - IEEE Comput. Soc. Int. Conf.*, pp. 103–109, 1997.

[120] K.-P. Wiedmann, H. Buxel, and G. Walsh, "Customer profiling in e-commerce: Methodological aspects and challenges," *J. Database Mark. Cust. Strateg. Manag.*, vol. 9, no. 2, pp. 170–184, 2002.

[121] M. S. Ackerman and D. T. Davis, "Privacy and security in E-commerce," *Market/Tržište*, vol. 21, no. 2, pp. 247–260, 2009.

[122] K. Kaur, I. Gupta, and A. K. Singh, “Data Leakage Prevention: E-Mail Protection via Gateway,” *J. Phys. Conf. Ser.*, vol. 933, no. 1, 2018.

[123] S. Yoshihama, T. Mishina, and T. Matsumoto, “Web-based Data Leakage Prevention,” *IWSEC ’10 Int. Work. Secur.*, no. June 2014, 2010.

[124] G. Katz, Y. Elovici, and B. Shapira, “CoBAn: A context based model for data leakage prevention,” *Inf. Sci. (Ny).*, vol. 262, no. June 2002, pp. 137–158, 2014.

[125] S. Alneyadi, E. Sithirasenan, and V. Muthukkumarasamy, “Detecting data semantic: A data leakage prevention approach,” *Proc. - 14th IEEE Int. Conf. Trust. Secur. Priv. Comput. Commun. Trust. 2015*, vol. 1, pp. 910–917, 2015.

[126] C. J. Chae, Y. J. Shin, K. Choi, K. B. Kim, and K. N. Choi, “A privacy data leakage prevention method in P2P networks,” *Peer-to-Peer Netw. Appl.*, vol. 9, no. 3, pp. 508–519, 2016.

[127] Y. Martín, “Methods and Tools for GDPR Compliance through Privacy and Data Protection Engineering,” *2018 IEEE Eur. Symp. Secur. Priv. Work.*, pp. 108–111, 2018.

[128] H. T. Tavani and J. H. Moor, “Privacy protection, control of information, and privacy-enhancing technologies,” *ACM SIGCAS Comput. Soc.*, vol. 31, no. 1, pp. 6–11, 2001.

[129] L. Wu, H. J. Cai, and H. Li, “SGX-UAM: A secure unified access management scheme with one time passwords via intel SGX,” *IEEE Access*, vol. 9, no. March 2012, pp. 38029–38042, 2021.

[130] H. Fahmy and N. Elkhateeb, “Proposed model for generation of one time password,” *Int. J. Comput. Sci. …*, no. July, 2018.

[131] M. Taufiq and D. Ogi, “Implementing One-Time Password Mutual Authentication Scheme on Sharing Renewed Finite Random Sub-Passwords Using Raspberry Pi as a Room Access Control to Prevent Replay Attack,” *Proc. - 2nd 2018 Int. Conf. Electr. Eng. Informatics, ICELTICs 2018*, pp. 13–

18, 2018.

[132] Z. Zheng, H. Cheng, Z. Zhang, Y. Zhao, and P. Wang, “An Alternative Method for Understanding User-Chosen Passwords,” *Secur. Commun. Networks*, vol. 2018, 2018.

[133] A. Oluwakemi Christiana, A. Noah Oluwatobi, G. Ayomide Victory, and O. Roseline Oluwaseun, “A Secured One Time Password Authentication Technique using (3, 3) Visual Cryptography Scheme,” *J. Phys. Conf. Ser.*, vol. 1299, no. 1, 2019.

[134] M. K. Sharma and M. J. Nene, “ Dual factor third-party biometric-based authentication scheme using quantum one time passwords † ,” *Secur. Priv.*, vol. 3, no. 6, pp. 1–18, 2020.

[135] E. Erdem and M. T. Sandikkaya, “OTPaaS-One time password as a service,” *IEEE Trans. Inf. Forensics Secur.*, vol. 14, no. 3, pp. 743–756, 2018.

[136] K. Pfeffer *et al.*, “On the usability of authenticity checks for hardware security tokens,” *Proc. 30th USENIX Secur. Symp.*, pp. 37–54, 2021.

[137] R. Tolosana, R. Vera-Rodriguez, and J. Fierrez, “BioTouchPass: Handwritten Passwords for Touchscreen Biometrics,” *IEEE Trans. Mob. Comput.*, vol. 19, no. 7, pp. 1532–1543, 2020.

[138] V. Landyshev, T. Blinovskaya, and D. Krakhmalev, “The practice of using one-time passwords in modern corporate information systems,” *E3S Web Conf.*, vol. 224, 2020.

[139] F. Alshanketi, I. Traoré, and A. Awad, “Multimodal mobile keystroke dynamics biometrics combining fixed and variable passwords,” *Secur. Priv.*, vol. 2, no. 1, p. e48, 2019.

[140] V. Shukla, A. Chaturvedi, and N. Srivastava, “A new one time password mechanism for client-server applications,” *J. Discret. Math. Sci. Cryptogr.*, vol. 22, no. 8, pp. 1393–1406, 2019.

[141] M. A. Hassan, Z. Shukur, and M. K. Hasan, “An Improved Time-Based One

Time Password Authentication Framework for Electronic Payments," *Int. J. Adv. Comput. Sci. Appl.*, vol. 11, no. 11, pp. 359–366, 2020.

[142] K. Bicakci, K. Ulker, Y. Uzunay, and H. T. S, "White-Box Implementations for Hash-Based Signatures and One-Time Passwords," pp. 1–17, 2002.

[143] D. M'Raihi, S. Machani, M. Pei, and J. Rydell, "Internet Engineering Task Force (IETF):Request for Comments: 6238," 2011. .

[144] S. Ma *et al.*, "An empirical study of SMS one-time password authentication in android apps," *ACM Int. Conf. Proceeding Ser.*, pp. 339–354, 2019.

[145] D. R. Chandran, "Use of AI Voice Authentication Technology Instead of Traditional Keypads in Security Devices," *J. Comput. Commun.*, vol. 10, no. 06, pp. 11–21, 2022.

[146] J. Obermaier and S. Tatschner, "Shedding too much light on a microcontroller's firmware protection," *11th USENIX Work. Offensive Technol. WOOT 2017, co-located with USENIX Secur. 2017*, 2017.

[147] J. Shahid, R. Ahmad, A. K. Kiani, T. Ahmad, S. Saeed, and A. M. Almuhaideb, "applied sciences Data Protection and Privacy of the Internet of Healthcare Things (IoHTs)," 2022.

[148] S. Kulkarni, R. M. Vani, and P. V. Hunagund, "FPGA based hardware security for edge devices in internet of things," *Proc. 5th Int. Conf. Commun. Electron. Syst. ICCES 2020*, no. Icces, pp. 1133–1138, 2020.

[149] C. Dong, J. Chen, W. Guo, and J. Zou, "A machine-learning-based hardware-Trojan detection approach for chips in the Internet of Things," *Int. J. Distrib. Sens. Networks*, vol. 15, no. 12, 2019.

[150] S. Singh and N. Singh, "Internet of Things (IoT): Security challenges, business opportunities & reference architecture for E-commerce," *Proc. 2015 Int. Conf. Green Comput. Internet Things, ICGCIoT 2015*, pp. 1577–1581, 2016.

[151] A. Syed and R. M. Lourde, "Hardware security threats to DSP applications in an IoT network," *Proc. - 2016 IEEE Int. Symp. Nanoelectron. Inf. Syst. iNIS*

2016, pp. 62–66, 2017.

[152] Y. Jin, “Introduction to hardware security,” *Electron.* , vol. 4, no. 4, pp. 763–784, 2015.

[153] K. M. Goertzel, “Integrated circuit security threats and hardware assurance countermeasures,” *CrossTalk*, vol. 26, no. 6, pp. 33–38, 2013.

[154] S. Sidhu, B. J. Mohd, and T. Hayajneh, “Hardware security in IoT devices with emphasis on hardware trojans,” *J. Sens. Actuator Networks*, vol. 8, no. 3, 2019.

[155] B. Jia, X. Zhang, J. Liu, Y. Zhang, K. Huang, and Y. Liang, “Blockchain-enabled Federated Learning Data Protection Aggregation Scheme with Differential Privacy and Homomorphic Encryption in IIoT,” vol. XX, no. XX, pp. 1–10, 2021.

[156] N. Samir *et al.*, “Energy-adaptive lightweight hardware security module using partial dynamic reconfiguration for energy limited internet of things applications,” *Proc. - IEEE Int. Symp. Circuits Syst.*, vol. 2019-May, pp. 1–4, 2019.

[157] A. Çelik *et al.*, “No 主観的健康感を中心とした在宅高齢者における健康関連指標に関する共分散構造分析Title,” *J. Mater. Process. Technol.*, vol. 1, no. 1, pp. 1–8, 2018.

[158] C. Gao, L. Luo, Y. Zhang, B. Pearson, and X. Fu, “Microcontroller based IoT system firmware security: Case studies,” *Proc. - IEEE Int. Conf. Ind. Internet Cloud, ICII 2019*, no. Icii, pp. 200–209, 2019.

[159] D. Strobel, D. Oswald, B. Richter, F. Schellenberg, and C. Paar, “Microcontrollers as (In)Security Devices for Pervasive Computing Applications,” *Proc. IEEE*, vol. 102, no. 8, pp. 1157–1173, 2014.

[160] X. Wang, C. Konstantinou, M. Maniatakos, and R. Karri, “ConFirm: Detecting firmware modifications in embedded systems using Hardware Performance Counters,” *2015 IEEE/ACM Int. Conf. Comput. Des. ICCAD 2015*, pp. 544–551, 2016.

[161] Y. Chen *et al.*, “Lightweight one-time password authentication scheme based on radiofrequency fingerprinting,” *IET Commun.*, vol. 12, no. 12, pp. 1477–1484, 2018.

[162] Z. Wang, L. Chen, S. Song, P. X. Cong, and Q. Ruan, “Automatic cyber security risk assessment based on fuzzy fractional ordinary differential equations,” *Alexandria Eng. J.*, vol. 59, no. 4, pp. 2725–2731, 2020.

[163] Y. sheng Liu, S. wei Wang, C. Xue, X. Y. Shao, and J. long Shao, “A simple and practical embedded software system architecture,” *Procedia Comput. Sci.*, vol. 166, pp. 78–83, 2020.

[164] E. Tattershall, C. Iddon, and J. Jensen, “Prototyping a secure, hierarchical internet of things system,” *Proc. - 11th IEEE/ACM Int. Conf. Util. Cloud Comput. Companion, UCC Companion 2018*, pp. 266–271, 2019.

[165] Anon, “IoT Architechture,” 2021. [Online]. Available: https://www.hiotron.com/iot-architecture-layers/.

[166] P. Marwedel, *Embedded Design System: Embedded Systems Foundations of Cyber-Physical Systems, and the Internet of Things*. 2011.

[167] J. B. Hou, T. Li, and C. Chang, “Research for Vulnerability Detection of Embedded System Firmware,” *Procedia Comput. Sci.*, vol. 107, no. Icict, pp. 814–818, 2017.

[168] D. Davidson, B. Moench, T. Ristenpart, and S. Jha, “{FIE} on Firmware: Finding Vulnerabilities in Embedded Systems Using Symbolic Execution,” in *22nd USENIX Security Symposium (USENIX Security 13)*, 2013, pp. 463–478.

[169] S. Skorobogatov, “How Microprobing Can Attack Encrypted Memory,” *Proc. - 20th Euromicro Conf. Digit. Syst. Des. DSD 2017*, pp. 244–251, 2017.

[170] M. R. Fadiheh, D. Stoffel, C. Barrett, S. Mitra, and W. Kunz, “Processor Hardware Security Vulnerabilities and their Detection by Unique Program Execution Checking,” *MBMV 2019 - 22. Work. "Methoden und Beschreibungssprachen zur Model. und Verif. von Schaltungen und Syst.*, pp.

85–86, 2019.

[171] A. Adomnicai, J. J. A. Fournier, and L. Masson, “Hardware security threats against bluetooth mesh networks,” *2018 IEEE Conf. Commun. Netw. Secur. CNS 2018*, pp. 1–9, 2018.

[172] I. Butun, A. Sari, and P. Österberg, “Hardware security of fog end-devices for the internet of things,” *Sensors (Switzerland)*, vol. 20, no. 20, pp. 1–28, 2020.

[173] A. Fasano *et al.*, “SoK: Enabling Security Analyses of Embedded Systems via Rehosting,” *ASIA CCS 2021 - Proc. 2021 ACM Asia Conf. Comput. Commun. Secur.*, no. Ii, pp. 687–701, 2021.

[174] J. Knechtel, “Hardware security for and beyond CMOS technology,” *Proc. Int. Symp. Phys. Des.*, pp. 115–126, 2021.

[175] F. Adelstein, M. Stillerman, and D. Kozen, “Malicious code detection for open firmware,” *Proc. - Annu. Comput. Secur. Appl. Conf. ACSAC*, vol. 2002-Janua, pp. 403–412, 2002.

[176] W. Hu, C. H. Chang, A. Sengupta, S. Bhunia, R. Kastner, and H. Li, “An Overview of Hardware Security and Trust: Threats, Countermeasures, and Design Tools,” *IEEE Trans. Comput. Des. Integr. Circuits Syst.*, vol. 40, no. 6, pp. 1010–1038, 2021.

[177] M. Bettayeb, Q. Nasir, and M. A. Talib, “Firmware update attacks and security for IoT devices survey,” *ACM Int. Conf. Proceeding Ser.*, 2019.

[178] R. Elnaggar and K. Chakrabarty, “Machine Learning for Hardware Security: Opportunities and Risks,” *J. Electron. Test. Theory Appl.*, vol. 34, no. 2, pp. 183–201, 2018.

[179] K. G. Liakos, G. K. Georgakilas, S. Moustakidis, P. Karlsson, and F. C. Plessas, “Machine Learning for Hardware Trojan Detection: A Review,” *5th Panhellenic Conf. Electron. Telecommun. PACET 2019*, pp. 1–6, 2019.

[180] R. Kastner and T. Huffmire, “Threats and Challenges in Recon fi gurable Hardware Security.”

[181] S. S. Ali, R. S. Chakraborty, D. Mukhopadhyay, and S. Bhunia, “Multi-level attacks: An emerging security concern for cryptographic hardware,” *Proc. - Design, Autom. Test Eur. DATE*, pp. 1176–1179, 2011.

[182] Z. Huang, Q. Wang, Y. Chen, and X. Jiang, “A Survey on Machine Learning against Hardware Trojan Attacks: Recent Advances and Challenges,” *IEEE Access*, vol. 8, pp. 10796–10826, 2020.

[183] K. Hasegawa, Y. Shi, and N. Togawa, “Hardware Trojan Detection Utilizing Machine Learning Approaches,” *Proc. - 17th IEEE Int. Conf. Trust. Secur. Priv. Comput. Commun. 12th IEEE Int. Conf. Big Data Sci. Eng. Trust. 2018*, pp. 1891–1896, 2018.

[184] A. Cui, M. Costello, and S. J. Stolfo, “When Firmware Modifications Attack : A Case Study of Embedded Exploitation,” *20th Annu. Netw. Distrib. Syst. Secur. Symp.*, 2013.

[185] T. A. Henzinger and J. Sifakis, “The embedded systems design challenge,” *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 4085 LNCS, pp. 1–15, 2006.

[186] S. Krstić, J. Yang, D. W. Palmer, R. B. Osborne, and E. Talmor, “Security of SoC firmware load protocols,” *Proc. 2014 IEEE Int. Symp. Hardware-Oriented Secur. Trust. HOST 2014*, pp. 70–75, 2014.

[187] I. Chomiak-Orsa, A. Rot, and B. Bartosz, “Artificial Intelligence in Cybersecurity: The Use of AI Along the Cyber Kill Chain,” *Artif. Intell. Cybersecurity Use AI Along Cyber Kill Chain*, pp. 406–416, 2019.

[188] N. Kaloudi and L. I. Jingyue, “The AI-based cyber threat landscape: A survey,” *ACM Comput. Surv.*, vol. 53, no. 1, 2020.

[189] R. Elankavi and R. Udayakumar, “Captcha as a graphical passwords - A new security primitive based on hard AI problems,” *Eurasian J. Anal. Chem.*, vol. 12, no. 4, pp. 93–100, 2017.

[190] H. Sun, X. Zhang, and B. Yao, “Construction of New Graphical Passwords

with Graceful-Type Labellings on Trees," *Proc. 2018 2nd IEEE Adv. Inf. Manag. Commun. Electron. Autom. Control Conf. IMCEC 2018*, no. Imcec, pp. 1491–1494, 2018.

[191] B. R. Maddipati, "Implementation of Captcha as Graphical Passwords For Multi Security," 2018.

[192] T. V. S. Vivek, V. N. Rajavarman, and S. R. Madala, "Advanced graphical-based security approach to handle hard AI problems based on visual security," *Int. J. Intell. Enterp.*, vol. 7, no. 1–3, pp. 250–266, 2020.

[193] V. Rana, "CaRP a graphical password : enhancing security using AI," pp. 675–677, 2018.

[194] F. Bergadano and I. Drago, "AI for Cybersecurity: from Adversarial Anomaly Detection to Intelligent Network Security Systems," *Ital-Ia 2022*, 2022.

[195] Anon, "Serial Data Transmission.," *IBM Tech. Discl. Bull.*, vol. 28, no. 7, pp. 2957–2958, 1985.

[196] V. Boed, "Serial communications," *Netw. Integr. Facil. Autom. Syst.*, no. Chapter 10, pp. 69–80, 1999.

[197] T. Tutorial, "Introduction to Serial Communication Technical Tutorial," *Interface*, 2002.

[198] A. Haldorai, A. Ramu, and S. Mohanram, *2nd EAI International Conference on Big Data Innovation for Sustainable Cognitive Computing*. 2019.

[199] N. van Dijk, A. Tanas, K. Rommetveit, and C. Raab, "Right engineering? The redesign of privacy and personal data protection," *Int. Rev. Law, Comput. Technol.*, vol. 32, no. 2–3, pp. 230–256, 2018.

[200] M. Al-Zubaidie, Z. Zhang, and J. Zhang, "PAX: Using pseudonymization and anonymization to protect patients' identities and data in the healthcare system," *Int. J. Environ. Res. Public Health*, vol. 16, no. 9, 2019.

[201] J. Samuel, J. Jaskolka, and G. O. M. Yee, "Leveraging External Data Sources

to Enhance Secure System Design," *2021 Reconciling Data Anal. Autom. Privacy, Secur. A Big Data Challenge, RDAAPS 2021*, 2021.

[202] F. Blix, S. A. Elshekeil, and S. Laoyookhong, "Data protection by design in systems development: From legal requirements to technical solutions," *2017 12th Int. Conf. Internet Technol. Secur. Trans. ICITST 2017*, pp. 98–103, 2018.

[203] I. Technology, I. T. Security, I. T. Security, and I. Layers, "3 IT-Security 3.1.," pp. 35–106, 2001.

[204] A. S. C. Alliance, "Embedded Hardware Security for IoT Applications," *A Smart Card Alliance Internet Things Secur. Counc. White Pap.*, no. December, 2016.

[205] H. Yang, L. Huang, C. Luo, and Q. Yu, "Research on Intelligent Security Protection of Privacy Data in Government Cyberspace," *2020 IEEE 5th Int. Conf. Cloud Comput. Big Data Anal. ICCCBDA 2020*, pp. 284–288, 2020.

[206] M. Colesky, J. H. Hoepman, and C. Hillen, "A Critical Analysis of Privacy Design Strategies," *Proc. - 2016 IEEE Symp. Secur. Priv. Work. SPW 2016*, pp. 33–40, 2016.

[207] T. Hoel, D. Griffiths, and W. Chen, "The Influence of Data Protection and Privacy Frameworks on the Design of Learning Analytics Systems CCS Concepts • Security and privacy→Privacy protections • General and reference→Design • Security and privacy→Social aspects of security and privacy • Secu," 2017.

[208] B. J. Koops and R. Leenes, "Privacy regulation cannot be hardcoded. A critical comment on the 'privacy by design' provision in data-protection law," *Int. Rev. Law, Comput. Technol.*, vol. 28, no. 2, pp. 159–171, 2014.

[209] P. Schaar, "Privacy by Design," *Identity Inf. Soc.*, vol. 3, no. 2, pp. 267–274, 2010.

[210] L. Sion *et al.*, "An architectural view for data protection by design," *Proc. - 2019 IEEE Int. Conf. Softw. Archit. ICSA 2019*, no. i, pp. 11–20, 2019.

[211] I. Gaidarski, “Using Big Data for Data Leak Prevention,” *2019 Big Data, Knowl. Control Syst. Eng.*, pp. 1–5, 2020.

[212] A. Romanou, “The necessity of the implementation of Privacy by Design in sectors where data protection concerns arise,” *Comput. Law Secur. Rev.*, vol. 34, no. 1, pp. 99–110, 2018.

[213] T. Suphakul and T. Senivongse, “Development of privacy design patterns based on privacy principles and UML,” *Proc. - 18th IEEE/ACIS Int. Conf. Softw. Eng. Artif. Intell. Netw. Parallel/Distributed Comput. SNPD 2017*, pp. 369–375, 2017.

[214] S. P. Skorobogatov, “Semi-invasive attacks-a new approach to hardware security analysis,” *Tech. report, Univ. Cambridge, Comput. Lab.*, no. 630, 2005.

[215] A. Skrop, “Data Leakage Detection Using Information Retrieval Methods,” no. c, pp. 74–78, 2014.

[216] E. Jardine, A. M. Lindner, and G. Owenson, “The potential harms of the Tor anonymity network cluster disproportionately in free countries,” *Proc. Natl. Acad. Sci. U. S. A.*, vol. 117, no. 50, pp. 31716–31721, 2020.

[217] M. Bernaschi, A. Celestini, M. Cianfriglia, S. Guarino, F. Lombardi, and E. Mastrostefano, “Onion under Microscope: An in-depth analysis of the Tor network,” pp. 1–19, 2021.

[218] R. Hansen, “First Glance: an Introductory Analysis of Network Forensics of Tor.,” *J. Digit. Forensics, Secur. Law*, no. c, pp. 105–120, 2013.

[219] M. Muir, P. Leimich, and W. J. Buchanan, “A Forensic Audit of the Tor Browser Bundle,” *Digit. Investig.*, vol. 29, pp. 118–128, 2019.

[220] M. Traudt, R. Jansen, and A. Johnson, “FlashFlow: A secure speed test for tor,” *Proc. - Int. Conf. Distrib. Comput. Syst.*, vol. 2021-July, pp. 381–391, 2021.

[221] I. Karunanayake, N. Ahmed, R. Malaney, R. Islam, and S. K. Jha, “De-Anonymisation Attacks on Tor: A Survey,” *IEEE Commun. Surv. Tutorials,*

vol. 23, no. 4, pp. 2324–2350, 2021.

[222] P. K. Masur, “How Online Privacy Literacy Supports Self-Data Protection and Self-Determination in the Age of Information,” vol. 8, no. 2, pp. 258–269, 2020.

[223] F. Fiedler, C. Dopmann, F. Tschorsch, and S. Lucia, “PredicTor: Predictive congestion control for the tor network,” *CCTA 2020 - 4th IEEE Conf. Control Technol. Appl.*, pp. 863–870, 2020.

[224] K. D. Watson, “The Tor Network: A Global Inquiry into the Legal Status of The Tor Network: A Global Inquiry into the Legal Status of Anonymity Networks Anonymity Networks Recommended Citation Recommended Citation THE TOR NETWORK: A GLOBAL INQUIRY INTO THE LEGAL STATUS O,” *Washingt. Univ. Glob. Stud. Law Rev.*, vol. 11, no. 3, p. 2012, 2012.

[225] B. Fabian, F. Goertz, S. Kunz, S. Müller, and M. Nitzsche, “Privately waiting - A usability analysis of the Tor anonymity network,” *16th Am. Conf. Inf. Syst. 2010, AMCIS 2010*, vol. 2, pp. 1427–1437, 2010.

[226] J. Juen, A. Johnson, A. Das, N. Borisov, and M. Caesar, “Defending Tor from Network Adversaries: A Case Study of Network Path Prediction,” *Proc. Priv. Enhancing Technol.*, vol. 2015, no. 2, pp. 171–187, 2015.

[227] E. Petagna, G. Laurenza, C. Ciccotelli, and L. Querzoni, “Peel the Onion: Recognition of Android Apps Behind the Tor Network,” *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 11879 LNCS, pp. 95–112, 2019.

[228] P. Winter, R. Ensafi, K. Loesing, and N. Feamster, “Identifying and characterizing Sybils in the Tor network,” *Proc. 25th USENIX Secur. Symp.*, pp. 1169–1185, 2016.

[229] J. Diaz, D. Arroyo, and F. B. Rodriguez, “Fair anonymity for the Tor network,” 2014.

[230] Pfitzmann, Andreas, and Marit Hansen. "A terminology for talking about privacy by data minimization: Anonymity, unlinkability, undetectability, unobservability, pseudonymity, and identity management." (2010).

[231] Garofalo, Giuseppe, Davy Preuveneers, and Wouter Joosen. "A Siamese Adversarial Anonymizer for Data Minimization in Biometric Applications." In 2020 IEEE European Symposium on Security and Privacy Workshops (EuroS&PW), pp. 334-343. IEEE, 2020.

[232] Aledhari, Mohammed, Marianne Di Pierro, and Fahad Saeed. "A Fourier-based data minimization algorithm for fast and secure transfer of big genomic datasets." In 2018 IEEE International Congress on Big Data (BigData Congress), pp. 128-134. IEEE, 2018.

Appendices

Appendix A: Published Papers

1. Enhanced Protection of Pseudonymized User Data via the Use of Multilayered Hardware Security
https://link.springer.com/chapter/10.1007/978-981-19-7138-9_9

International MultiConference of Engineers and Computer Scientists
↳ IMECS 2021: **Transactions on Engineering Technologies** pp 103–115 | Cite as

Enhanced Protection of Pseudonymized User Data via the Use of Multilayered Hardware Security

Mukuka Kangwa, Charles S. Lubobya & Jackson Phiri

Conference paper | First Online: 12 October 2022

Abstract

The use of Information Communication Technology (ICT) in various spheres of life has led to the need for service consumers to have an electronic presence in the cyber-space. The world has become a global village; one can buy goods and services thousands of miles away from the country thanks to technology. One can even access medical services remotely using telemedicine. ICT is also being used at Airports, Universities, and many other areas that require interaction. To be able to interact and use some of these online platforms, a user is normally required to provide some form of Identity for their electronic Id to be created and thereafter

2. Protection of personally identifiable Information and Privacy via the use of Hardware and Software
https://www.iaeng.org/publication/IMECS2021/IMECS2021_pp75-81.pdf

Proceedings of the International MultiConference of Engineers and Computer Scientists 2021
IMECS 2021, October 20-22, 2021, Hong Kong

Protection of personally identifiable Information and Privacy via the use of Hardware and Software

Mukuka Kangwa, Charles S. Lubobya, Jackson Phiri

Abstract: **This paper proposes a novel approach to enhance the protection of Personally Identifiable Information (PII) using a unique combination of hardware and software as well as the use of a One Time Password (OTP) algorithm based formulated through the modification of the RFC based Time-based One Time Password (TOTP) standard. The adoption of electronic channels for commerce has necessitated the need for enhanced protection of PII. For one to be granted to e-services one has to surrender part, if not, all of their PII hence making their personal data susceptible to leakage. To ascertain the effectiveness of the proposed solution, tests were conducted using various methods and tools such as Arduino microcontrollers, python programming language, Arduino programming platform and the Proteus Simulation software. Results from the experiments conducted demonstrate the effectiveness of the proposed solution in preventing the leakage of PII.**

Key Words: **Personally Identifiable Information, Data Privacy, One Time Password, Data Protection, Time-based One Time Password, Firmware and TOR**

I. Introduction

The Information Age has witnessed an unprecedented adoption of electronic channels in the delivery of services to consumers. Most providers of electronic services request users to submit Personally Identifiable Information (PII) in order to get access to their electronic services [1]. This has resulted in huge volumes of aggregated PII being collected by a number of service providers and thereby making that data vulnerable to leakage [2]. Leaked PII exposes the owner of the information to high risk such as financial fraud and physical harm. Despite a number of solutions having been formulated and implemented to address this challenge, the problem remains [3]. Several incidents have occurred where huge volumes of data has been leaked and privacy breached [4]. Data that is exposed to the internet, whether on the edge equipment like phones and tablets, or in the cloud is at risk hence the need to provide more effective protection methods [5].

This paper is a substantially revised version of the paper presented at the 18th International Conference for e-Business with the Digital Object Identifier (DOI) of 10.5220/0010576201160126. This paper proposes the use of an enhanced data protection approach together with Onion routing and the enhanced RFC6238 based TOTP to protect personal data and provide user privacy.

Mukuka Kangwa is a PhD Candidate at the University of Zambia, Great East Road Campus, Lusaka Zambia Email: mukukakangwa@yahoo.com
Charles S. Lubobya is the Head of Electrical Department at the University of Zambia, Great East Road Campus, Lusaka Zambia. Email: cslubobya@unza.zm
Jackson Phiri is a Senior Lecturer in the Computer Science Department at the University of Zambia, Great East Road Campus, Lusaka Zambia. Email: jackson.phiri@cs.unza.zm

II. Related Work

Several literature was reviewed to appreciate similar prior works by other authors. Frank and Michael patented a solution to help protect personal data. They proposed having a Trusted Party that provides static Identities (ID) to users. In addition, Block chain technology was to be used to protect the data. The diagram shown in Fig 1 below shows a summary of how the solution is to work; the user obtains an ID from the digital ID provider and submits it to the service provider as proof of identification. The service provider verifies with the ID provider if the user can be trusted and the response the ID provider returns determines whether or not a service will be offered to the user. Furthermore, an offline escrow is to be used for keeping the PII to be accessed via legally approved means. Pseudonymization (and not anonymization) is to be used to make it possible to trace a user when there is need [2].

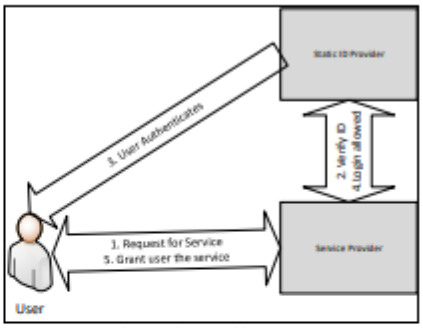
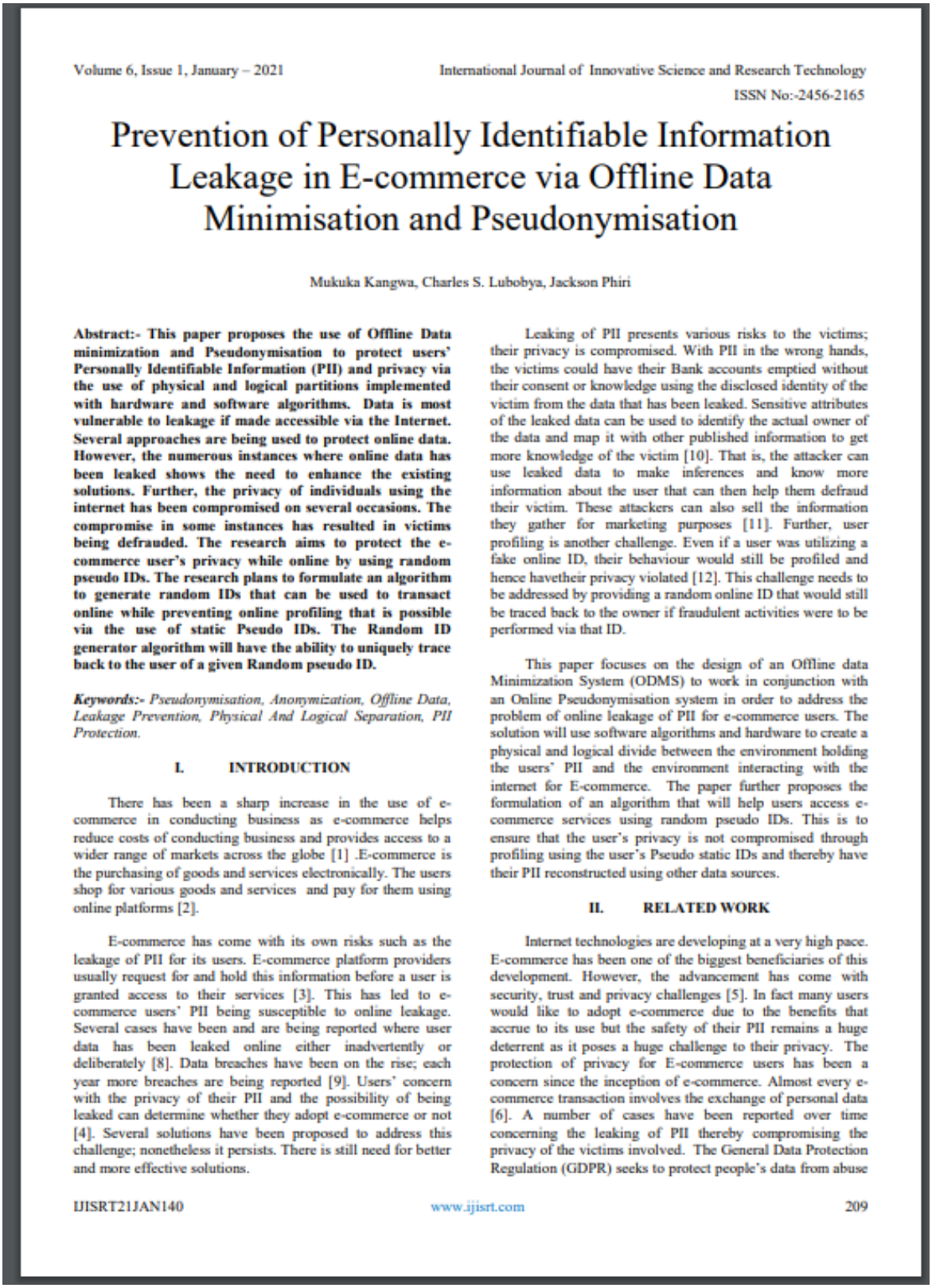


Fig 1: Static ID Provider Concept

The use of static electronic IDs is not adequate for providing privacy to the users as a static ID can be profiled thereby compromise the privacy of the user [1]. Furthermore, the use of Block chain technology might not be very feasible as the technology is currently resource intensive [6]. A global solution based on this proposal would consume a huge amount of resources for the proof of works to be used to protect data from being leaked or modified or even deleted.

ISBN: 978-988-14049-1-6
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

3. Prevention of Personally Identifiable Information Leakage in E-commerce via Offline Data Minimisation and Pseudonymisation
https://ijisrt.com/assets/upload/files/IJISRT21JAN140.pdf



Volume 6, Issue 1, January – 2021 International Journal of Innovative Science and Research Technology
ISSN No:-2456-2165

Prevention of Personally Identifiable Information Leakage in E-commerce via Offline Data Minimisation and Pseudonymisation

Mukuka Kangwa, Charles S. Lubobya, Jackson Phiri

Abstract:- This paper proposes the use of Offline Data minimization and Pseudonymisation to protect users' Personally Identifiable Information (PII) and privacy via the use of physical and logical partitions implemented with hardware and software algorithms. Data is most vulnerable to leakage if made accessible via the Internet. Several approaches are being used to protect online data. However, the numerous instances where online data has been leaked shows the need to enhance the existing solutions. Further, the privacy of individuals using the internet has been compromised on several occasions. The compromise in some instances has resulted in victims being defrauded. The research aims to protect the e-commerce user's privacy while online by using random pseudo IDs. The research plans to formulate an algorithm to generate random IDs that can be used to transact online while preventing online profiling that is possible via the use of static Pseudo IDs. The Random ID generator algorithm will have the ability to uniquely trace back to the user of a given Random pseudo ID.

Keywords:- *Pseudonymisation, Anonymization, Offline Data, Leakage Prevention, Physical And Logical Separation, PII Protection.*

I. INTRODUCTION

There has been a sharp increase in the use of e-commerce in conducting business as e-commerce helps reduce costs of conducting business and provides access to a wider range of markets across the globe [1] .E-commerce is the purchasing of goods and services electronically. The users shop for various goods and services and pay for them using online platforms [2].

E-commerce has come with its own risks such as the leakage of PII for its users. E-commerce platform providers usually request for and hold this information before a user is granted access to their services [3]. This has led to e-commerce users' PII being susceptible to online leakage. Several cases have been and are being reported where user data has been leaked online either inadvertently or deliberately [8]. Data breaches have been on the rise; each year more breaches are being reported [9]. Users' concern with the privacy of their PII and the possibility of being leaked can determine whether they adopt e-commerce or not [4]. Several solutions have been proposed to address this challenge; nonetheless it persists. There is still need for better and more effective solutions.

Leaking of PII presents various risks to the victims; their privacy is compromised. With PII in the wrong hands, the victims could have their Bank accounts emptied without their consent or knowledge using the disclosed identity of the victim from the data that has been leaked. Sensitive attributes of the leaked data can be used to identify the actual owner of the data and map it with other published information to get more knowledge of the victim [10]. That is, the attacker can use leaked data to make inferences and know more information about the user that can then help them defraud their victim. These attackers can also sell the information they gather for marketing purposes [11]. Further, user profiling is another challenge. Even if a user was utilizing a fake online ID, their behaviour would still be profiled and hence havetheir privacy violated [12]. This challenge needs to be addressed by providing a random online ID that would still be traced back to the owner if fraudulent activities were to be performed via that ID.

This paper focuses on the design of an Offline data Minimization System (ODMS) to work in conjunction with an Online Pseudonymisation system in order to address the problem of online leakage of PII for e-commerce users. The solution will use software algorithms and hardware to create a physical and logical divide between the environment holding the users' PII and the environment interacting with the internet for E-commerce. The paper further proposes the formulation of an algorithm that will help users access e-commerce services using random pseudo IDs. This is to ensure that the user's privacy is not compromised through profiling using the user's Pseudo static IDs and thereby have their PII reconstructed using other data sources.

II. RELATED WORK

Internet technologies are developing at a very high pace. E-commerce has been one of the biggest beneficiaries of this development. However, the advancement has come with security, trust and privacy challenges [5]. In fact many users would like to adopt e-commerce due to the benefits that accrue to its use but the safety of their PII remains a huge deterrent as it poses a huge challenge to their privacy. The protection of privacy for E-commerce users has been a concern since the inception of e-commerce. Almost every e-commerce transaction involves the exchange of personal data [6]. A number of cases have been reported over time concerning the leaking of PII thereby compromising the privacy of the victims involved. The General Data Protection Regulation (GDPR) seeks to protect people's data from abuse

4. Enhanced Protection of Ecommerce Users' Personal Data and Privacy using the Trusted Third-Party Model
https://www.scitepress.org/Papers/2021/105762/105762.pdf

Enhanced Protection of Ecommerce Users' Personal Data and Privacy using the Trusted Third Party Model

Mukuka Kangwa[a], Charles S Lubobya[1] and Jackson Phiri[2][b]
[1]Department of Electrical Engineering, University of Zambia, Lusaka, Zambia
[2]Department of Computer Science, University of Zambia, Lusaka, Zambia

Keywords: Personally Identifiable Information, Data Privacy, Electronic Services, Data Protection, Random Electronic Identity, Hardware and Software.

Abstract: The rapid adoption of electronic delivery of services by various electronic service providers such as ecommerce and e-governance services leaves the users of these services with no option but to adapt if they are to continue accessing their desired services. To access these services, very often one has to reveal some of their personal data in order to get registered on the platforms made available courtesy of the service providers. One person is likely to surrender their personal identifying data to several service providers hence making their aggregated data susceptible to leakage online. Despite several solutions already in use data leakage is still prevalent. Our research proposes and tests a method that aggregates personal identifying data and seeks to enhance its protection from leakage using a novel approach formulated from software and hardware. This paper outlines the design and explains in detail how the approach is expected to protect data. It further gives details of the results that were obtained from experiments conducted on the constructed key component of the proposed solution.

1 INTRODUCTION

The information Age has seen an unprecedented rise in the delivery of various services using electronic means. Most of the providers of electronic services such as ecommerce and e-governance require a person to submit some elements of their Personally Identifiable Information (PII) before they could grant that potential service consumer access to their electronic services (Kangwa, Lubobya, & Phiri, 2021). This has resulted into huge amounts of aggregated PII being collected across number of service providers and thereby making it vulnerable to intentional or inadvertent leakage (Patent No. US 2019 / 0333054 A1, 2019). Leaked PII puts the owner of the information at high risk; users can have their Bank account broken into, their privacy compromised and even pose physical threat to the victim. Despite a number of solutions having been formulated and implemented in order to address this challenge, the problem persists. With the advent of the Covid-19 Pandemic across the globe, more service providers have opted to use electronic means to deliver their services to their clients to minimize physical contact hence putting more PII at risk of being leaked. In fact a number of incidents have already occurred where data privacy has been breached (Hauer, 2015).

There are several methods and approaches such as cryptography that are being used to provide confidentiality and privacy to data (Pawar & Harkut, 2018). This paper proposes a simplified, yet effective method that can be employed to protect personal data. It builds on what other scholars have formulated to come up with a more effective approach.

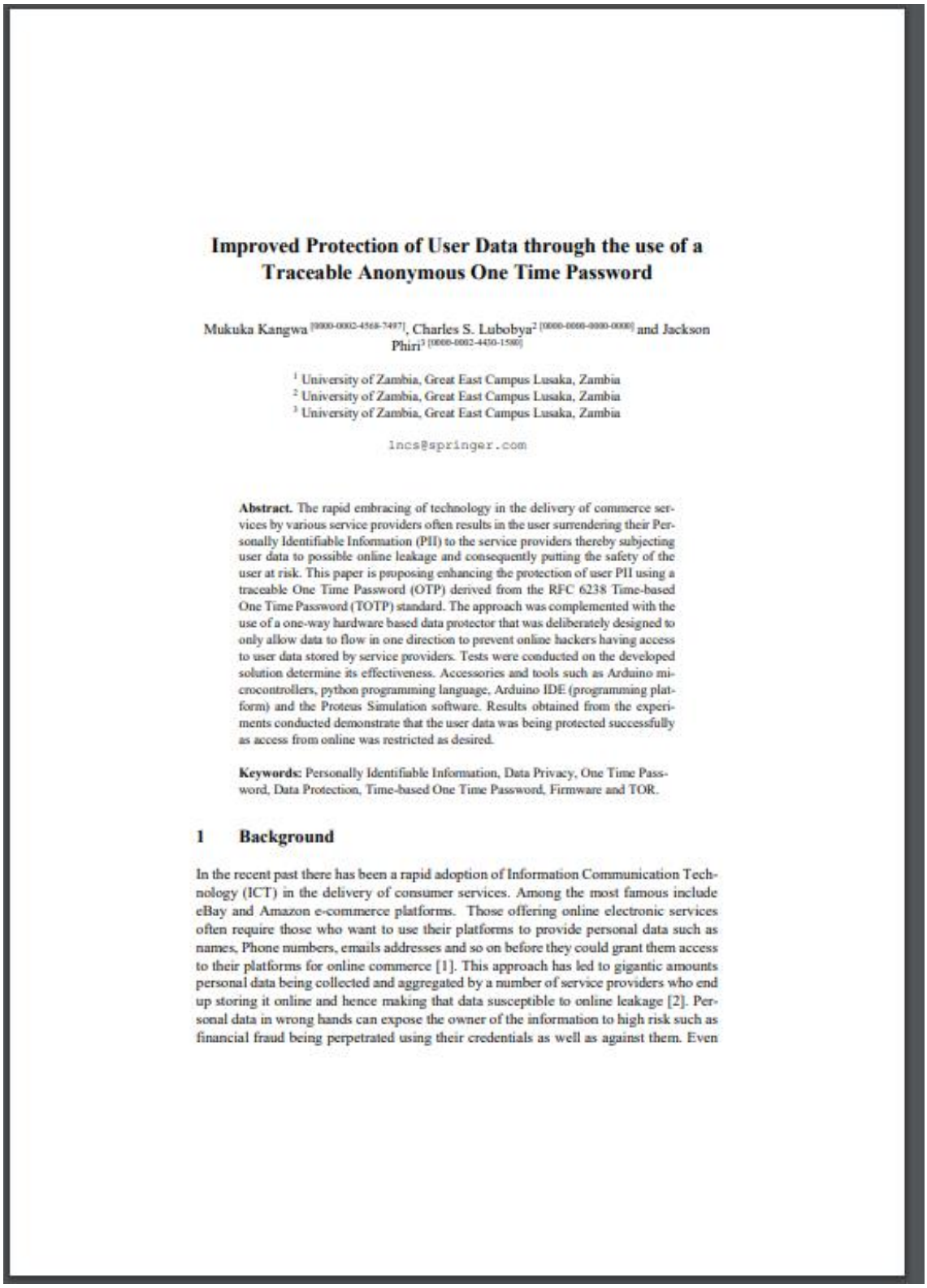
2 RELATED WORK

A number of scholars have proposed varying approaches to help protect personal data while allowing the owners access to online services. Frank and Michael proposed and patented a solution to help

[a] https://orcid.org/0000-0002-4568-7497
[b] https://orcid.org/0000-0002-4430-1580

Kangwa, M., Lubobya, C. and Phiri, J.
Enhanced Protection of Ecommerce Users' Personal Data and Privacy using the Trusted Third Party Model.
DOI: 10.5220/0010576201160126
In Proceedings of the 18th International Conference on e-Business (ICE-B 2021), pages 116-126
ISBN: 978-989-758-527-2

5. Improved Protection of User Data through the use of a Traceable Anonymous One Time Password
Accepted by Springer to be published as a Book Chapter in 2023



Improved Protection of User Data through the use of a Traceable Anonymous One Time Password

Mukuka Kangwa [0000-0002-4568-7497], Charles S. Lubobya[2] [0000-0000-0000-0000] and Jackson Phiri[3] [0000-0002-4450-1580]

[1] University of Zambia, Great East Campus Lusaka, Zambia
[2] University of Zambia, Great East Campus Lusaka, Zambia
[3] University of Zambia, Great East Campus Lusaka, Zambia

lncs@springer.com

Abstract. The rapid embracing of technology in the delivery of commerce services by various service providers often results in the user surrendering their Personally Identifiable Information (PII) to the service providers thereby subjecting user data to possible online leakage and consequently putting the safety of the user at risk. This paper is proposing enhancing the protection of user PII using a traceable One Time Password (OTP) derived from the RFC 6238 Time-based One Time Password (TOTP) standard. The approach was complemented with the use of a one-way hardware based data protector that was deliberately designed to only allow data to flow in one direction to prevent online hackers having access to user data stored by service providers. Tests were conducted on the developed solution determine its effectiveness. Accessories and tools such as Arduino microcontrollers, python programming language, Arduino IDE (programming platform) and the Proteus Simulation software. Results obtained from the experiments conducted demonstrate that the user data was being protected successfully as access from online was restricted as desired.

Keywords: Personally Identifiable Information, Data Privacy, One Time Password, Data Protection, Time-based One Time Password, Firmware and TOR.

1 Background

In the recent past there has been a rapid adoption of Information Communication Technology (ICT) in the delivery of consumer services. Among the most famous include eBay and Amazon e-commerce platforms. Those offering online electronic services often require those who want to use their platforms to provide personal data such as names, Phone numbers, emails addresses and so on before they could grant them access to their platforms for online commerce [1]. This approach has led to gigantic amounts personal data being collected and aggregated by a number of service providers who end up storing it online and hence making that data susceptible to online leakage [2]. Personal data in wrong hands can expose the owner of the information to high risk such as financial fraud being perpetrated using their credentials as well as against them. Even

Appendix B: The RFC 6238 Standard

“
Internet Engineering Task Force (IETF) D. M'Raihi
Request for Comments: 6238 Verisign, Inc.
Category: Informational S. Machani
ISSN: 2070-1721 Diversinet Corp.
M. Pei
Symantec
J. Rydell
Portwise, Inc.
May 2011

TOTP: Time-Based One-Time Password Algorithm

Abstract

This document describes an extension of the One-Time Password (OTP) algorithm, namely the HMAC-based One-Time Password (HOTP) algorithm, as defined in RFC 4226, to support the time-based moving factor. The HOTP algorithm specifies an event-based OTP algorithm, where the moving factor is an event counter. The present work bases the moving factor on a time value. A time-based variant of the OTP algorithm provides short-lived OTP values, which are desirable for enhanced security.

The proposed algorithm can be used across a wide range of network applications, from remote Virtual Private Network (VPN) access and Wi-Fi network logon to transaction-oriented Web applications. The authors believe that a common and shared algorithm will facilitate adoption of two-factor authentication on the Internet by enabling interoperability across commercial and open-source implementations.

Status of This Memo

This document is not an Internet Standards Track specification; it is published for informational purposes.

This document is a product of the Internet Engineering Task Force (IETF). It represents the consensus of the IETF community. It has received public review and has been approved for publication by the Internet Engineering Steering Group (IESG). Not all documents approved by the IESG are a candidate for any level of Internet

Standard; see Section 2 of RFC 5741.

Information about the current status of this document, any errata, and how to provide feedback on it may be obtained at http://www.rfc-editor.org/info/rfc6238.

This document describes an extension of the One-Time Password (OTP) algorithm, namely the HMAC-based One-Time Password (HOTP) algorithm, as defined in [RFC4226], to support the time-based moving factor.

1.2. Background

As defined in [RFC4226], the HOTP algorithm is based on the HMAC-SHA-1 algorithm (as specified in [RFC2104]) and applied to an increasing counter value representing the message in the HMAC computation.

Basically, the output of the HMAC-SHA-1 calculation is truncated to obtain user-friendly values:

HOTP(K,C) = Truncate(HMAC-SHA-1(K,C))

where Truncate represents the function that can convert an HMAC-SHA-1 value into an HOTP value. K and C represent the shared secret and counter value; see [RFC4226] for detailed definitions.

TOTP is the time-based variant of this algorithm, where a value T, derived from a time reference and a time step, replaces the counter C in the HOTP computation.

TOTP implementations MAY use HMAC-SHA-256 or HMAC-SHA-512 functions, based on SHA-256 or SHA-512 [SHA2] hash functions, instead of the HMAC-SHA-1 function that has been specified for the HOTP computation in [RFC4226].

2. Notation and Terminology

The key words "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT", "SHOULD", "SHOULD NOT", "RECOMMENDED", "MAY", and "OPTIONAL" in this document are to be interpreted as described in [RFC2119].

3. Algorithm Requirements

This section summarizes the requirements taken into account for designing the TOTP algorithm.

R1: The prover (e.g., token, soft token) and verifier (authentication or validation server) MUST know or be able to derive the current Unix time (i.e., the number of seconds elapsed since midnight UTC of January 1, 1970) for OTP generation. See [UT] for a more detailed definition of the commonly known "Unix time". The precision of the time used by the prover affects how often the clock synchronization should be done; see Section 6.

R2: The prover and verifier MUST either share the same secret or the knowledge of a secret transformation to generate a shared secret.

R3: The algorithm MUST use HOTP [RFC4226] as a key building block.

R4: The prover and verifier MUST use the same time-step value X.

R5: There MUST be a unique secret (key) for each prover.

R6: The keys SHOULD be randomly generated or derived using key derivation algorithms.

R7: The keys MAY be stored in a tamper-resistant device and SHOULD be protected against unauthorized access and usage.

4. TOTP Algorithm

This variant of the HOTP algorithm specifies the calculation of a one-time password value, based on a representation of the counter as a time factor.

4.1. Notations

- X represents the time step in seconds (default value X = 30 seconds) and is a system parameter.

- T0 is the Unix time to start counting time steps (default value is 0, i.e., the Unix epoch) and is also a system parameter.

4.2. Description

Basically, we define TOTP as TOTP = HOTP(K, T), where T is an integer and represents the number of time steps between the initial counter time T0 and the current Unix time.

More specifically, T = (Current Unix time - T0) / X, where the default floor function is used in the computation.

For example, with T0 = 0 and Time Step X = 30, T = 1 if the current Unix time is 59 seconds, and T = 2 if the current Unix time is 60 seconds.

The implementation of this algorithm MUST support a time value T larger than a 32-bit integer when it is beyond the year 2038. The value of the system parameters X and T0 are pre-established during the provisioning process and communicated between a prover and verifier as part of the provisioning step. The provisioning flow is out of scope of this document; refer to [RFC6030] for such provisioning container specifications.

5. Security Considerations

5.1. General

The security and strength of this algorithm depend on the properties of the underlying building block HOTP, which is a construction based on HMAC [RFC2104] using SHA-1 as the hash function.

The conclusion of the security analysis detailed in [RFC4226] is that, for all practical purposes, the outputs of the dynamic truncation on distinct inputs are uniformly and independently distributed strings.

The analysis demonstrates that the best possible attack against the HOTP function is the brute force attack.

As indicated in the algorithm requirement section, keys SHOULD be chosen at random or using a cryptographically strong pseudorandom generator properly seeded with a random value.

Keys SHOULD be of the length of the HMAC output to facilitate interoperability.

We RECOMMEND following the recommendations in [RFC4086] for all pseudorandom and random number generations. The pseudorandom numbers used for generating the keys SHOULD successfully pass the randomness test specified in [CN], or a similar well-recognized test.

All the communications SHOULD take place over a secure channel, e.g., Secure Socket Layer/Transport Layer Security (SSL/TLS) [RFC5246] or IPsec connections [RFC4301].

We also RECOMMEND storing the keys securely in the validation system, and, more specifically, encrypting them using tamper-resistant hardware encryption and exposing them only when required: for example, the key is decrypted when needed to verify an OTP value, and re-encrypted immediately to limit exposure in the RAM to a short period of time.

The key store MUST be in a secure area, to avoid, as much as possible, direct attack on the validation system and secrets database. Particularly, access to the key material should be limited to programs and processes required by the validation system only.

5.2. Validation and Time-Step Size

An OTP generated within the same time step will be the same. When an OTP is received at a validation system, it doesn't know a client's exact timestamp when an OTP was generated. The validation system may typically use the timestamp when an OTP is received for OTP comparison. Due to network latency, the gap (as measured by T, that is, the number of time steps since T0) between the time that the OTP was generated and the time that the OTP arrives at the receiving system may be large. The receiving time at the validation system and the actual OTP generation may not fall within the same time-step window that produced the same OTP. When an OTP is generated at the end of a time-step window, the receiving time most likely falls into the next time-step window. A validation system SHOULD typically set a policy for an acceptable OTP transmission delay window for validation. The validation system should compare OTPs not only with the receiving timestamp but also the past timestamps that are within the transmission delay. A larger acceptable delay window would expose a larger window for attacks. We RECOMMEND that at most one time step is allowed as the network delay.

The time-step size has an impact on both security and usability. A larger time-step size means a larger validity window for an OTP to be accepted by a validation system. There are implications for using a larger time-step size, as follows:

First, a larger time-step size exposes a larger window to attack. When an OTP is generated and exposed to a third party before it is consumed, the third party can consume the OTP within the time-step window.

We RECOMMEND a default time-step size of 30 seconds. This default value of 30 seconds is selected as a balance between security and usability.

Second, the next different OTP must be generated in the next time-step window. A user must wait until the clock moves to the next time-step window from the last submission. The waiting time may not be exactly the length of the time step, depending on when the last OTP was generated. For example, if the last OTP was generated at the halfway point in a time-step window, the waiting time for the next OTP is half the length of the time step. In general, a larger time-step window means a longer waiting time for a user to get the next valid OTP after the last successful OTP validation. A too-large

window (for example, 10 minutes) most probably won't be suitable for typical Internet login use cases; a user may not be able to get the next OTP within 10 minutes and therefore will have to re-login to the same site in 10 minutes.

Note that a prover may send the same OTP inside a given time-step window multiple times to a verifier. The verifier MUST NOT accept the second attempt of the OTP after the successful validation has been issued for the first OTP, which ensures one-time only use of an OTP.

6. Resynchronization

Because of possible clock drifts between a client and a validation server, we RECOMMEND that the validator be set with a specific limit to the number of time steps a prover can be "out of synch" before being rejected.

This limit can be set both forward and backward from the calculated time step on receipt of the OTP value. If the time step is 30 seconds as recommended, and the validator is set to only accept two time steps backward, then the maximum elapsed time drift would be around 89 seconds, i.e., 29 seconds in the calculated time step and 60 seconds for two backward time steps.

This would mean the validator could perform a validation against the current time and then two further validations for each backward step (for a total of 3 validations). Upon successful validation, the validation server can record the detected clock drift for the token in terms of the number of time steps. When a new OTP is received after this step, the validator can validate the OTP with the current timestamp adjusted with the recorded number of time-step clock drifts for the token.

Also, it is important to note that the longer a prover has not sent an OTP to a validation system, the longer (potentially) the accumulated clock drift between the prover and the verifier. In such

cases, the automatic resynchronization described above may not work
if the drift exceeds the allowed threshold. Additional authentication measures should be used to safely authenticate the prover and explicitly resynchronize the clock drift between the prover and the validator.

7. Acknowledgements

The authors of this document would like to thank the following people
for their contributions and support to make this a better specification: Hannes Tschofenig, Jonathan Tuliani, David Dix, Siddharth Bajaj, Stu Veath, Shuh Chang, Oanh Hoang, John Huang, and
Siddhartha Mohapatra.

”

From https://www.ietf.org/rfc/rfc6238.txt [143]

Appendix C: Ethical Clearance

THE UNIVERSITY OF ZAMBIA
DIRECTORATE OF RESEARCH AND GRADUATE STUDIES

Great East Road Campus | P.O. Box 32379 | Lusaka10101 | Tel: +260-211-290 258/291 777
Fax: (+260)-211-290 258/253 952 | **E-mail**: director.drgs@unza.zm | Website: www.unza.zm

APPROVAL OF STUDY

IORG No. 0005376
NASRECREC IRB No. 00006465

5th May, 2023

REF NO. NASREC-2023- MAY – 001

Mr. Mukuka Kangwa,
The University of Zambia,
School of Engineering,
P.O. Box 32379,
LUSAKA.

Dear, Mr.
Kangwa,

RE: " PREVENTION OF PERSONALITY IDENTIFIABLE INFPRMATION LEAKAGE IN E- COMMERCE USING OFFLINE DATA MINIMIZATION AND ONLINE PSEUDONYMISATION"

Reference is made to your protocol dated as captioned above. NASREC resolved to approve this study and your participation as Principal Investigator for a period of one year.

REVIEW TYPE	ORDINARY REVIEW	APPROVAL NO. NASREC-2023 MAY - 001
Approval and Expiry Date	Approval Date: 5th May, 2023	Expiry Date: 4th May, 2024
Protocol Version and Date	Version - Nil.	4th May, 2024
Information Sheet, Consent Forms and Dates	• English.	To be provided
Consent form ID and Date	Version - Nil	To be provided
Recruitment Materials	Nil	Nil
Other Study Documents	Questionnaire.	

Specific conditions will apply to this approval. As Principal Investigator it is your responsibility to ensure that the contents of this letter are adhered to. If these are not adhered to, the approval may be suspended. Should the study be suspended, study sponsors and other regulatory authorities will be informed.

CONDITIONS OF APPROVAL

- No participant may be involved in any study procedure prior to the study approval or after the expiration date.
- All unanticipated or Serious Adverse Events (SAEs) must be reported to NASREC within 5 days.
- All protocol modifications must be approved by NASREC prior to implementation unless they are intended to reduce risk (but must still be reported for approval). Modifications will include any change of investigator/s or site address.
- All protocol deviations must be reported to NASREC within 5 working days.
- All recruitment materials must be approved by NASREC prior to being used.
- Principal investigators are responsible for initiating Continuing Review proceedings. NASREC will only approve a study for a period of 12 months.
- It is the responsibility of the PI to renew his/her ethics approval through a renewal application to NASREC.
- Where the PI desires to extend the study after expiry of the study period, documents for study extension must be received by NASREC at least 30 days before the expiry date. This is for the purpose of facilitating the review process. Documents received within 30 days after expiry will be labelled "late submissions" and will incur a penalty fee of K500.00. No study shall be renewed whose documents are submitted for renewal 30 days after expiry of the certificate.
- Every 6 (six) months a progress report form supplied by The University of Zambia Natural and Applied Sciences Research Ethics Committee as an IRB must be filled in and submitted to us. There is a penalty of K500.00 for failure to submit the report.
- When closing a project, the PI is responsible for notifying, in writing or using the Research Ethics and Management Online (REMO),both NASREC
- and the National Health Research Authority (NHRA) when ethics certification is no longer required for a project.
- In order to close an approved study, a Closing Report must be submitted in writing or through the REMO system. A Closing Report should be filed when data collection has ended and the study team will no longer be using human participants or animals or secondary data or have any direct or indirect contact with the research participants or animals for the study.
- Filing a closing report (rather than just letting your approval lapse) is important as it assists NASREC in efficiently tracking and reporting on projects. Note that some funding agencies and sponsors require a notice of closure from the IRB which had approved the study and can

only be generated after the Closing Report has been filed.

- A reprint of this letter shall be done at a fee.

- All protocol modifications must be approved by NASREC by way of an application for an amendment prior to implementation unless they are intended to reduce risk (but must still be reported for approval). Modifications will include any change of investigator/s or site address or methodology and methods. Many modifications entail minimal risk adjustments to a protocol and/or consent form and can be made on an Expedited basis (via the IRB Chair). Some examples are: format changes, correcting spelling errors, adding key personnel, minor changes to questionnaires, recruiting and changes, and so forth. Other, more substantive changes, especially those that may alter the risk-benefit ratio, may require Full Board review. In all cases, except where noted above regarding subject safety, any changes to any protocol document or procedure must first be approved by NASREC before they can be implemented.

Should you have any questions regarding anything indicated in this letter, please do not hesitate to get in touch with us at the above indicated address.

On behalf of NASREC, we would like to wish you all the success as you carry out your study.

Yours faithfully,

Dr. Mususu Kaonda

VICE-CHAIRPERSON
THE UNIVERSITY OF ZAMBIA NATURAL AND APPLIED SCIENCES RESEARCH ETHICS COMMITTEE - IRB

cc: Director, Directorate of Research and Graduate Studies
Assistant Director (Research), Directorate of Research and Graduate Studies
Assistant Registrar (Research), Directorate of Research and Graduate Studies