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

Mukuka Kangwa[1][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

protect personal identifying data. In their proposal they have a Trusted Party that provides static Identities (ID) to users. In addition, they proposed the use of Block chain technology to protect personal data for the users. In the scenario shown in Figure 1 below, 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 is indeed genuine and the response the ID provider returns determines whether or not a service will be rendered to the user. Frank and Michael are proposing the use of an offline escrow to keep the identifying data to be accessed via legally approved means. They further submitted that pseudonymization rather than anonymization be used to make it possible to identify a user when need arise (Patent No. US 2019 / 0333054 A1, 2019).

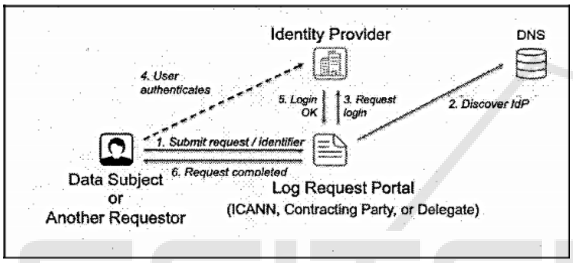


Figure 1: Proposed Approach by Frank and Michael (Patent No. US 2019 / 0333054 A1, 2019).

The implied use of static electronic IDs in the proposed approach above might not be sufficient to help maintain the privacy of the users as a static ID can be profiled easily and hence compromise the privacy of the owner while surfing online (Kangwa, Lubobya, & Phiri, 2021). Furthermore, the use of Blockchain technology might not be very feasible as the technology is currently resource intensive (Bao et al., 2020). To provide a global solution 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. The Blockchain technology will require some modifications to make it friendlier to the proposed approach. Lighter Blockchain-like implementation on a local scale might be more feasible than a peer-to-peer global approach that is currently wide spread.

Another key challenge with Blockchain technology is its lack of scalability due to its design.

The solution seeks to solve a global problem by providing access that is global hence scalability is very key to accommodate everyone who desires to use the proposed solution. Moreover, Blockchain faces some privacy and security challenges hence might provide a solution to one problem and introduce more challenges (Bao et al., 2020) . In addition, the time lag that the technology inherently experiences due to its design poses a huge challenge. When one node generates a transaction, a number of other nodes need to confirm and reach consensus before a transaction can be considered as complete. This results in delayed completion of transactions (II-Agure, Belsam, & Yun-ke, 2019). Further, its complexity 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. Perhaps build as a service where costs can be shared (Zhang, Alkubati, Bao, & Yu, 2021). This paper 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 for the purpose of issuing National IDs. It further proposes a solution to be employed to effectively prevent PII leakage by the appointed TTP and is a buildup of our earlier work in the paper "Prevention of Personally Identifiable Information Leakage in E-commerce via Offline Data Minimisation and Pseudonymization" (Kangwa et al., 2021).

3 TRUSTED AND UNTRUSTED THIRD PART MODELS

The leakage of PII has been going on for some time now. Ecommerce platforms normally hold client information online so that they can identify them before offering any service. A lot of user data is being held in the cloud by various service providers hence making that data susceptible to leakage (Ye, Dong, Shen, Cao, & Zhao, 2019). For example, e-bay, an Ecommerce platform, suffered a Distributed Denial of Service (DDoS) attack where their databases were scanned and client data was exposed (Innab & Alamri, 2018). This attack was possible because the databases were accessible online.

Fanghan et'al proposed a model that would control access to data using a system whose security was to be enhanced using encryption. The system was aimed at replacing a Trusted Party model as they believed that the Trusted Third Party (TTP) might not be so reliable to control access to user data on behalf of the user. They proposed a model that would give the data owner the power to control who accesses their data. Figure 2 below gives a summary of the model:

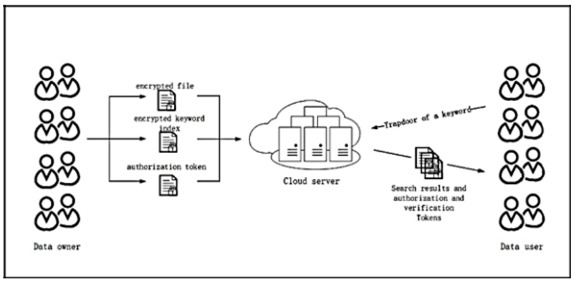


Figure 2: Model without Trusted Third Parties (Ye et al., 2019).

According the model, a cloud server would be required to keep some of the data. The user encrypts their documents before sharing and decides who can be given access by sharing their decryption keys with the people who need access (Ye et al., 2019). The fact that some data is kept in the cloud makes that data vulnerable to leakage. Even if it is encrypted, if the key shared finds itself in wrong hands, then the encrypted date, if accessed can be deciphered. In addition, it means several can be granted access to the data thereby increasing points of potential data breach. Moreover, 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 (Zhan, Fan, Cai, Gao, & Zhuang, 2018).

Aarthy et'al proposed a Trusted Third Party Model that was aimed at resolving the issues of trust by users of cloud services. Users had huge concerns over security and trust of cloud service providers 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 (Aarthy, Aarthi, Farhath, Lakshana, & Lavanya, 2017). Figure 3 below gives a summary of how their proposed model was expected to work.

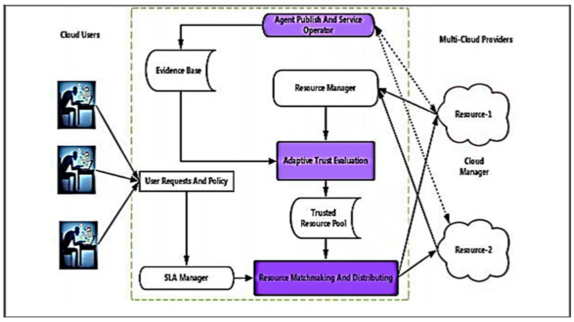


Figure 3: Trusted Third Party Model (Aarthy et al., 2017).

The idea was to have a party that both the service providers and potential users of the services can 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. 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.

Locher et'al, proposed the use of a distributed ledger to replace the use of a Trusted Third Party (TTP) (Locher, Obermeier, & Pignolet, 2018). As proposed by other authors earlier considered, the distributed ledger owes its security around the approval of any transaction to its requirements for consensus before any transaction can be confirmed. However, this approach results in delayed transaction completion as well as huge resources being required to make the technology operational (Bao et al., 2020). Locher et'al acknowledged that despite the distributed approach of using block chain technology, users still needed to trust each over (Locher et al., 2018). The aspect of trust is what the Trusted Third Party model aims to address hence the use of blockchain technology might have a limited application to certain use cases where trust might not be an immense requirement.

The Trusted Third Party Model was also tabled by Jamshiya et'al, to provide trust amongst Internet of Things (IoT) devices where security and trust establishment is a challenge. For these devices to connect to each other 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 to each other. Each IoT device is first connected to the TTP and assigned an id to be identified by. 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, off 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 (Jamshiya & Menon, 2018). Figure 4 below shows their proposed model:

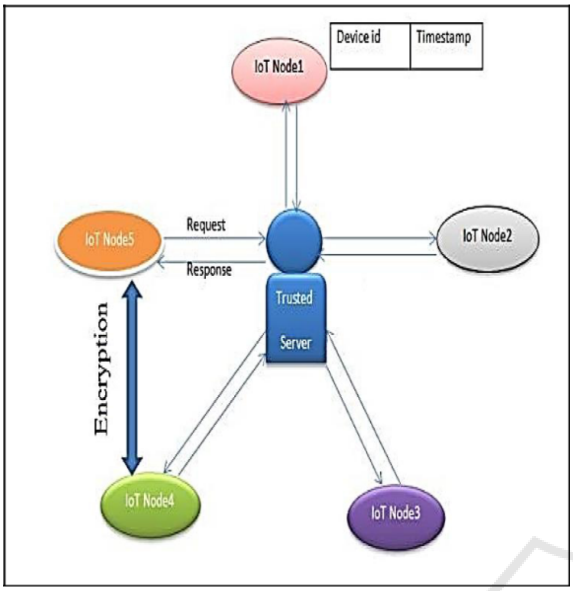


Figure 4: Trusted Third Party for IoT Devices (Jamshiya & Menon, 2018).

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 to each other in future.

More scholars were considered regarding the protection of PII and privacy of users while online. Peter et’al recognizes 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 pseudonymization of PII in addition to already existing data security techniques (Štarchoň & Pikulík, 2019). The additional measure indicates that the existing techniques applied by various service providers are not sufficient hence the data leakages and privacy violations that are experienced from time to time. Peter et’al defined pseudonymization as transforming data in such a way that the resulting data cannot be associated with the original owner without additional data. That is, if one was to stumble upon pseudonymized data, they should not be able to identify the owner without requiring additional information to fill in the blanks. The authors propose 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 (Štarchoň & Pikulík, 2019). Pseudonymization requires that, if legally demanded, the transformed data must be traceable to the real owner via the data processors. The challenge with having data scattered across various service providers is that, the probability of that information being leaked remains high as any of the service providers holding the data might be compromised.

Sergio et’al are of the view that the advancement in technology has resulted in the need for more effective techniques and solutions to provide security and privacy to personal and other sensitive information. They contend that current solutions might not be sufficient to meet the required levels of privacy and security demanded by regulations such as the European GDPR (Ribeiro & Nakamura, 2019). The team proposed the use of methods such as pseudonymization and anonymization. Pseudonymization was preferred to anonymization as they intended to use their solution for the protection of Health data for children. Pseudonymization provides a possibility of identifying the actual individual using additional information when need arise. Anonymization, on the other hand, alters data in such a way that it can no longer be traced back to the actual owner.

The team further proposed combining pseudonymization with other security techniques such as hashing of pseudo IDs, encryption of pseudonymized data and so on (Ribeiro & Nakamura, 2019). It must be noted that as long as data is kept online, the possibility of being leaked remains. There is 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.

4 SOLUTION DESIGN

As long as information is available online, the possibility of someone accessing that information without authorization remains despite the various mechanisms used to protect that information. Our considered view is that the best way to protect information is to make it unavailable to the online hackers hence the proposed design in Figure 5 below:

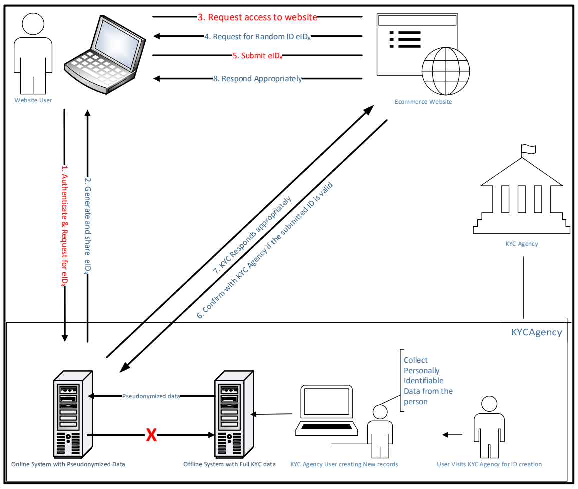

3. Request access to website
4. Request for Random ID eID$_R$
5. Submit eID$_R$
8. Respond Appropriately
Website User
Ecommerce Website
1. Authenticate & Request for eID$_R$
2. Generate and share eID$_R$
7. KYC Responds appropriately
6. Confirm with KYC Agency if the submitted ID is valid
KYC Agency
KYCAgency
Collect Personally Identifiable Data from the person
Pseudonymized data
X
Online System with Pseudonymized Data
Offline System with Full KYC data
KYC Agency User creating New records
User Visits KYC Agency for ID creation

Figure 5: 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 for the purpose of issuing National IDs.

The model demands that other service providers requiring KYC 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.

The approach will operate as follows:

4.1 User Registration

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

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 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 6 below gives a detailed flow of the registration process.

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 minimize 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 only system will be the automatic transmission of the Pseudo ID, eIDs, to the online system. The transmission will be 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 will automatically create a record with the eIDs as the primary key. It will then create an anonymous email box for the user. The mail box is to be used for delivering Random eIDr to the user. In addition, the emails will be automatically destroyed after a predetermined period to prevent formulation of the Key/algorithm being used for the generation of random IDs in case the online system with mail boxes 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.

4.2 Proposed Universally Unique Electronic ID (eID)

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

The eID will comprise 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:

XXX.XXXXXXXXXX

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 digits for 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 below:

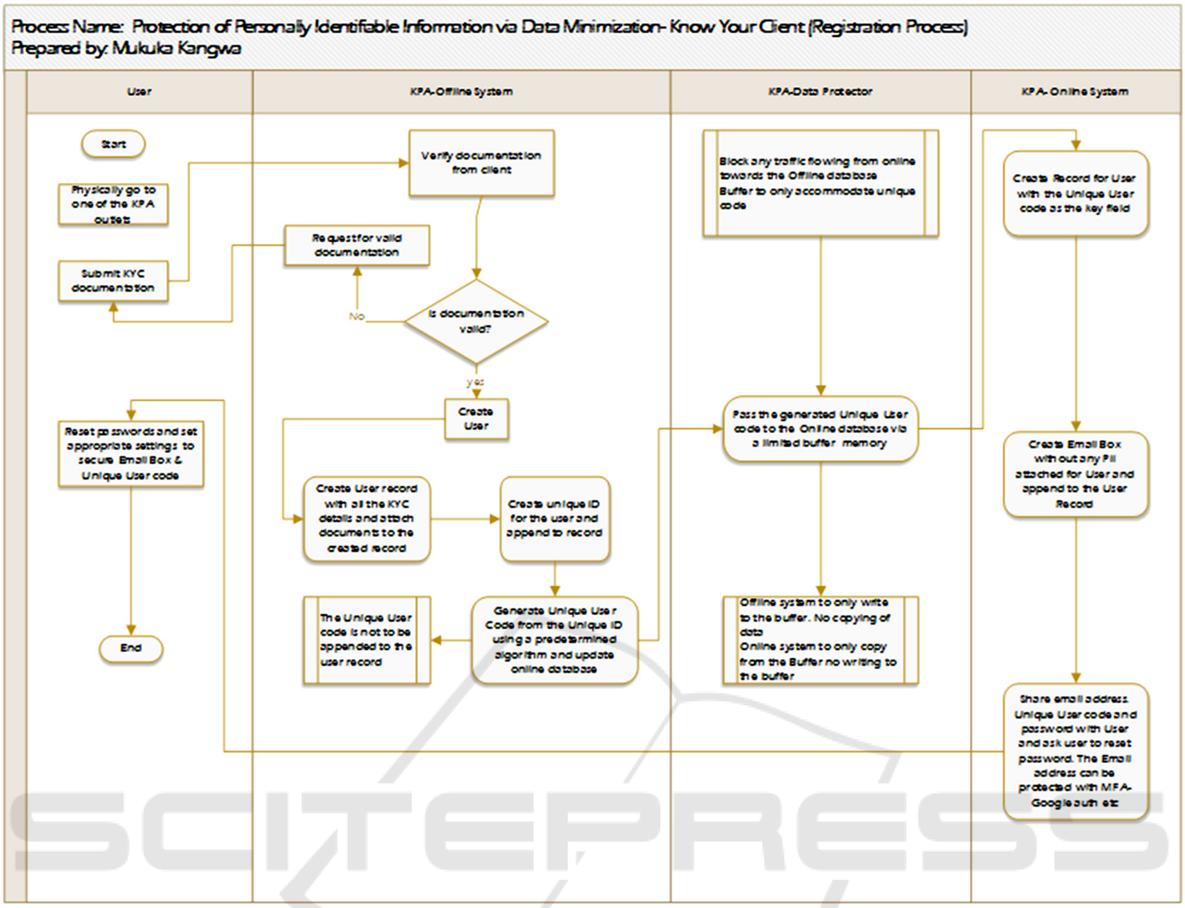
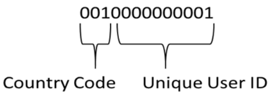

Process Name: Protection of Personally Identifiable Information via Data Minimization- Know Your Client (Registration Process)
Prepared by: Mukuka Kangwa
User
KPA-Offline System
KPA-Data Protector
KPA-Online System
Start
Physically go to one of the KPA outlets
Submit KYC documentation
Verify documentation from client
Request for valid documentation
Is documentation valid?
No
yes
Create User
Create User record with all the KYC details and attach documents to the created record
Create unique ID for the user and append to record
Generate Unique User Code from the Unique ID using a predetermined algorithm and update online database
The Unique User code is not to be appended to the user record
Block any traffic flowing from online towards the Offline database
Buffer to only accommodate unique code
Pass the generated Unique User code to the Online database via a limited buffer memory
Offline system to only write to the buffer. No copying of data
Online system to only copy from the Buffer no writing to the buffer
Create Record for User with the Unique User code as the key field
Create Email Box with out any PII attached for User and append to the User Record
Share email address, Unique User code and password with User and ask user to reset password. The Email address can be protected with MFA-Google auth etc
Reset passwords and set appropriate settings to secure Email Box & Unique User code
End

Figure 6: KYC Registration Process.



4.3 Data Protector Operations (RMS)

The Data protector that will safeguard the Personally Identifiable Information will connect the offline system to the online system and operate as follows:

Data exchange between the two systems will only flow in one direction; that is, it will only flow from the offline system towards the online system as depicted in Fig1. The aim of this restriction is to ensure that no one is able to 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 (Kangwa et al., 2021).

Figure 7 shows the summary of how the Data protector (Restricted Memory System) will be operating:

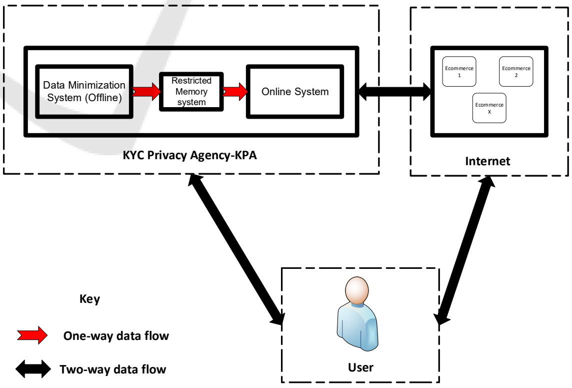


Figure 7: Data Protector Operations.

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 10gigabyte 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 few milliseconds as the strings would only constitute 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 disrupt any exploitive data transfer in session as legal sessions will be expected to only last few milliseconds.

(i) User Transaction with Ecommerce Sites.

The user will either access the KYC Agency to generate a universally unique random ID, eIDr, or first access an ecommerce site or 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 to 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.

User Logons on to the Website or App for the KYC Agency

Once authenticated, the user generates Random ID eIDr. The KYC system sends the ID, r, to 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 8 below gives a pictorial view of how the transaction will flow from the beginning to the end.

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

There is 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 tape must be encrypted.

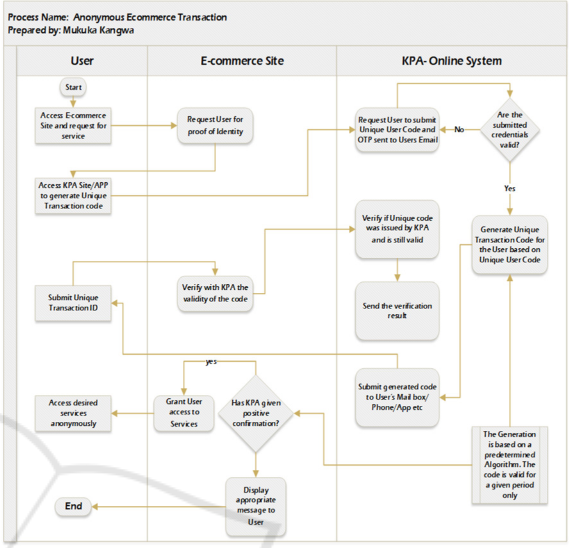

Process Name: Anonymous Ecommerce Transaction
Prepared by: Mukuka Kangwa
User
E-commerce Site
KPA- Online System
Start
Access E-commerce Site and request for service
Request User for proof of Identity
Access KPA Site/APP to generate Unique Transaction code
Request User to submit Unique User Code and OTP sent to Users Email
No
Are the submitted credentials valid?
Yes
Verify if Unique code was issued by KPA and is still valid
Generate Unique Transaction Code for the User based on Unique User Code
Submit Unique Transaction ID
Verify with KPA the validity of the code
Send the verification result
yes
Access desired services anonymously
Grant User access to Services
Has KPA given positive confirmation?
Submit generated code to User's Mail box/ Phone/App etc
The Generation is based on a predetermined Algorithm. The code is valid for a given period only
End
Display appropriate message to User

Figure 8: Anonymous Ecommerce Transaction.

5 EXPERIMENTAL METHODOLOGY

To ascertain the effectiveness of the proposed solution, tests were conducted using various methods and tools. Only the Data protector was built as other components proposed could be easily built using existing approaches. That's, components such as websites do exist already while the offline system for storing data is similar to other systems except it will be kept offline focus on the proposed algorithms for pseudonymization and creation of traceable random IDs.

The Data protector otherwise, known us Restricted Memory System (RMS), was built using the following components; two Arduino UNO microcontrollers, copper cables, serial ports, serial monitors, python programming language, Arduino UNO IDE and Proteus Simulation software.

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 limited amount of bandwidth. Two way communication would require two cables physically connected between the two devices communicating to each other. To ensure one way communication, one cable was physically removed as shown in Figure 9 below.

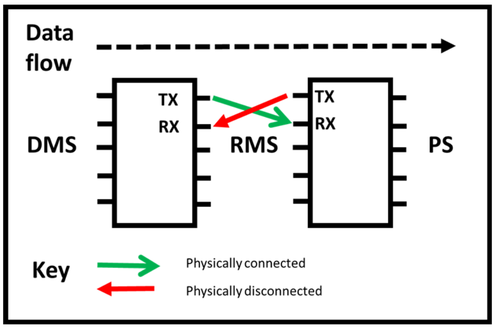


Figure 9: Simplified Circuit Diagram (Kangwa et al., 2021).

This would ensure that even if the online component wanted to communicate to the offline component, the communication would not be successful as there would be no means of reaching the other side.

The design uses two microcontrollers instead of one to control the security of the system. Hardware is susceptible to hardware Trojans. These viruses can be put into the hardware at 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 (Ali, Chakraborty, Mukhopadhyay, & Bhunia, 2011). One Arduino UNO 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 10 below:

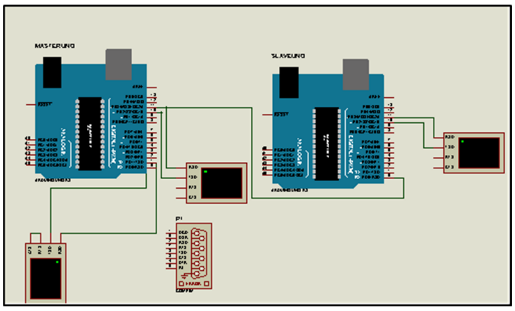


Figure 10: Physical Configuration of RMS.

Even if the Arduino 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 Arduino to change configurations and enable two-way communication thus making it impossible to achieve without having physical access. This is what makes the solution simple, yet very effective, if built and deployed as designed.

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 configuration outlined. Data was also sent from the online system to the offline system. Six scenarios were tested. In scenario 1 and 2, both the Transmitting and receiving PINs were physically connected while in Scenario 3 and 4, the cable connecting Transmitting PIN for the Master Arduino to the receiving PIN of the Slave Arduino was disconnected. In scenario 5 and 6, the receiving PIN of the Master Arduino connected to the Transmitting PIN of the slave Arduino was disconnected.

The Bandwidth between the two Arduinos 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.

6 RESULTS AND DISCUSSION

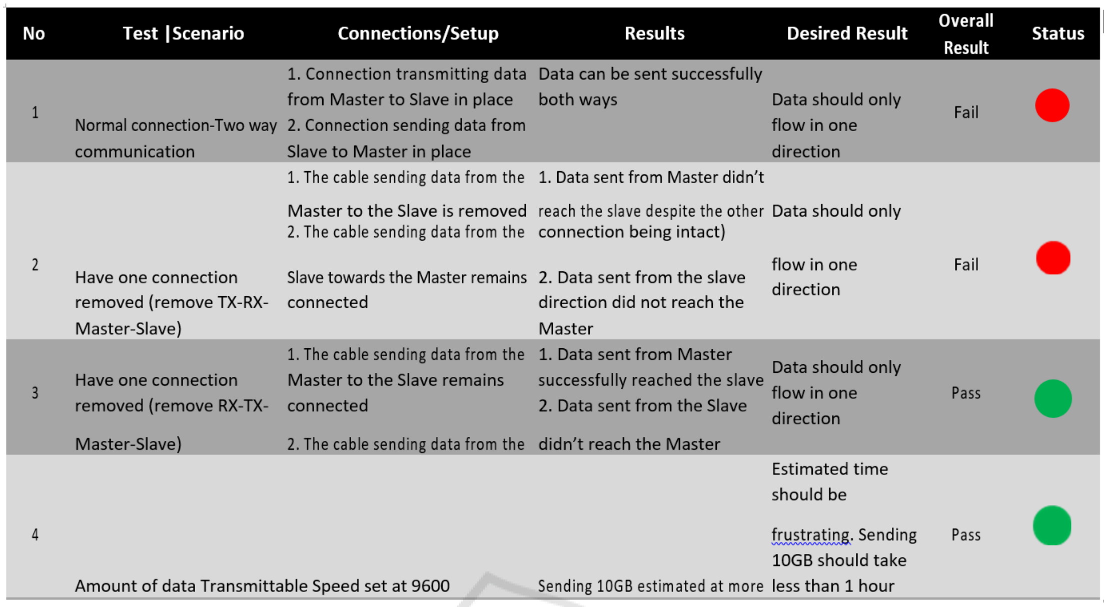
Table1 below gives a summary of results that were obtained from various scenarios.

In scenario 1 and 2 covered by Test number 1 in the results of Table 1, Data was successfully sent both ways with the Transmitting and receiving PINs connected correctly on the Master and Slave Arduinos.

In Scenario 1 and 2 confirmed by test 1, with both Transmitting and receiving PINs connected correctly on the Master and Slave Arduinos, data was successfully sent both ways. That is, data was able to flow from the offline system holding PII towards the online system susceptible to hacking and vice versa.

This was a fail as the objective was to ensure that data could only flow in one direction. That is, from the offline system towards the online system. In this scenario data successfully flowed in both directions hence the Data protector cannot protect PII by preventing access by users connecting from the Internet. It is vital that data cannot flow from the online system to the offline system even if bandwidth is restricted as malware such as ransomware can be created as a very small payload and yet cause serious damage to data once deployed into the offline system holding PII. The chances of one infecting the offline system, if they cannot access it from the internet, is very slim as they would need physical access to the offline system. Hence the test results for the first scenario renders the configuration undesirable and a fail.

Table 1.



No	Test |Scenario	Connections/Setup	Results	Desired Result	Overall Result	Status
1	Normal connection-Two way communication	1. Connection transmitting data from Master to Slave in place 2. Connection sending data from Slave to Master in place	Data can be sent successfully both ways	Data should only flow in one direction	Fail	🔴
2	Have one connection removed (remove TX-RX-Master-Slave)	1. The cable sending data from the Master to the Slave is removed 2. The cable sending data from the Slave towards the Master remains connected	1. Data sent from Master didn't reach the slave despite the other connection being intact) 2. Data sent from the slave direction did not reach the Master	Data should only flow in one direction	Fail	🔴
3	Have one connection removed (remove RX-TX-Master-Slave)	1. The cable sending data from the Master to the Slave remains connected 2. The cable sending data from the	1. Data sent from Master successfully reached the slave 2. Data sent from the Slave didn't reach the Master	Data should only flow in one direction	Pass	🟢
4	Amount of data Transmittable	Speed set at 9600	Sending 10GB estimated at more	Estimated time should be frustrating. Sending 10GB should take less than 1 hour	Pass	🟢

In the Scenario covered by test 2 with transmitting PIN on the Master Arduino not connected to the receiving PIN on the Slave Arduino while the receiving PIN on the Master Arduino remained connected to the transmitting PIN of the Slave Arduino, data could not flow in any direction.

In the scenario depicted in test 3, with the transmitting PIN of the Master Arduino connected to the transmitting PIN of the receiving Arduino, while the receiving PIN.

While the receiving PIN is disconnected from the transmitting PIN of the Slave Arduino, data was sent successfully from the Master Slave Arduino connecting the offline system towards the Slave Arduino connecting the Online system but data could not be sent from Slave Arduino connecting the online system towards the Master Arduino connecting the offline system.

With the speed/bandwidth of 9600bps configured between the two Arduinos, 10 GB of data would take about 103days to transfer across the Data protector from the offline system holding PII to the online system susceptible to hacking.

Test 2 resulted in data failing to flow in any direction. This too was a fail as the main objective of the proposed solution was to allow automatic creation of online pseudonymized records for users while preventing access to the PII data by users with access to the internet.

With data not flowing to the online system from the offline system, it would require another approach of transferring pseudo data matching records on the offline database to the online system. That approach might be manual hence introducing another risk where if data has to be moved using external media then malware can be introduced onto the offline system using that media. Therefore, this configuration is not desirable and was a fail.

Test 3 was successful as data could only flow in one direction. That is, data could only flow from the offline system towards the online system. This was the desired configuration as it would prevent hackers successfully accessing PII data sitting on the offline database. It would also prevent malware from being introduced from the online system to the offline system as it would corrupt the PII data sitting on the offline system. This result shows that it is possible to keep sensitive data "offline" while allowing real-time connection between the offline system and online system for the creation of corresponding records for the user to access online services anonymously once created on the KYC system.

The restriction of the bandwidth between the offline and online system to 9600bps ensured that only minimal data could pass across at any given time. For example to transmit 10GB of PII would take more than 100days. It is very possible to reduce the bandwidth further and increase how long it would take to transmit reasonable amount of data across the data protector component. The restricted bandwidth would discourage both external and internal disgruntled elements from attempting to steal PII sitting on the offline system. In addition, the valid data transmissions across the two systems are short bursts of few characters. The restriction helps prevent theft of sensitive PII by both internal and external parties.

It must be noted that for this solution to be effective, it must be used in-conjunction with other techniques such as the pseudonymizing of offline data before corresponding records are created online as outlined in Fig5. In addition, online data must be anonymous so that if the records are leaked, no identifying information would be part of the leaked records. Furthermore, random IDs must be used for the online system to ensure privacy of users is maintained.

In Our next paper will detail the algorithms to be used for the pseudonymization of PII as well as the generation of Random electronic IDs for anonymous use of electronic services such as ecommerce.

7 CONCLUSION

The experiment results show that it is possible to protect PII from hackers by not presenting any possibility of accessing the data regardless of the security configurations of the systems holding that data. The fact that no online user can reach the offline system holding sensitive data makes the system more secure. Enhanced protection comes in because no one would be able to access the offline system from the online system 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 transfer of data towards the offline system. The Restricted amount of data that can be sent via the data protector is a huge deterrent to would-be data criminals as the time it would take would render the exercise futile.

To make the proposed solution effective, it must be implemented as recommended in Fig5 as well as the detailed process flows in Fig6 and Fig8. The implementation of the proposed solution in the manner outlined would create a layered defence mechanism to protect PII as well provide privacy to the user. It would further, make it possible for authorities to trace users who would commit fraud online if need arises. The approach is good to the good elements and bad to the bad elements.

REFERENCES

Aarthy, D. K., Aarthi, M., Farhath, K. A., Lakshana, S., & Lavanya, V. (2017). Reputation-based trust management in cloud using a trusted third party. *ICONSTEM 2017 - Proceedings: 3rd IEEE International Conference on Science Technology, Engineering and Management*, *2018-Janua*, 220–225. https://doi.org/10.1109/ICONSTEM.2017.8261418

Ali, S. S., Chakraborty, R. S., Mukhopadhyay, D., & Bhunia, S. (2011). Multi-level attacks: An emerging security concern for cryptographic hardware. *Proceedings -Design, Automation and Test in Europe, DATE*, 1176–1179. https://doi.org/10.1109/date.2011.5763307

Bao, Z., Wang, Q., Shi, W., Wang, L., Lei, H., & Chen, B. (2020). When Blockchain Meets SGX: An Overview, Challenges, and Open Issues. *IEEE Access*, *8*, 170404–170420. https://doi.org/10.1109/access.2020.3024254

Coppolino, L., D'Antonio, S., Mazzeo, G., & Romano, L. (2019). A comprehensive survey of hardware-assisted security: From the edge to the cloud. *Internet of Things*, *6*, 100055. https://doi.org/10.1016/j.iot.2019.100055

Frank A Cona, M. D. P. (2019). *Patent No. US 2019 / 0333054 A1*. United States of America.

Hauer, B. (2015). Data and information leakage prevention within the scope of information security. *IEEE Access*, *3*, 2554–2565. https://doi.org/10.1109/ACCESS.2015.2506185

II-Agure, Z., Belsam, A., & Yun-ke, C. (2019). The Semantics of Anomalies in IoT Integrated BlockChain Network. *IEEE*, 144–146.

Innab, N., & Alamri, A. (2018). The Impact of DDoS on E-commerce. *21st Saudi Computer Society National Computer Conference, NCC 2018*, 1–4. https://doi.org/10.1109/NCG.2018.8593125

Jamshiya, P. K., & Menon, D. M. (2018). Design of a Trusted Third Party Key Exchange Protocol for Secure Internet of Things (IoT). *Proceedings of the International Conference on Inventive Communication and Computational Technologies, ICICCT 2018*, (Icicct), 1834–1838. https://doi.org/10.1109/ICICCT.2018.8473281

Kangwa, M., Lubobya, C. S., & Phiri, J. (2021). Prevention of Personally Identifiable Information Leakage in E-commerce via Offline Data Minimisation and Pseudonymisation. *International Journal of Innovative Science and Research Technology*, *6*(1), 209–212. Retrieved from https://scholar.google.com/scholar?hl=en&as_sdt=0%2C5&q=mukuka+kangwa&oq=

Locher, T., Obermeier, S., & Pignolet, Y. A. (2018). When Can a Distributed Ledger Replace a Trusted Third Party? *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)*, 1069–1077. https://doi.org/10.1109/Cybermatics_2018.2018.00197

Pawar, H. R., & Harkut, D. G. (2018). Classical and Quantum Cryptography for Image Encryption

Decryption. *Proceedings of the 2018 3rd IEEE International Conference on Research in Intelligent and Computing in Engineering, RICE 2018*, 1–4. https://doi.org/10.1109/RICE.2018.8509035

Ribeiro, S. L., & Nakamura, E. T. (2019). Privacy Protection with Pseudonymization and Anonymization in a Health IoT System: Results from OCARIoT. *Proceedings - 2019 IEEE 19th International Conference on Bioinformatics and Bioengineering, BIBE 2019*, 904–908. https://doi.org/10.1109/BIBE.2019.00169

Štarchoň, P., & Pikulík, T. (2019). GDPR principles in data protection encourage pseudonymization through most popular and full-personalized devices - mobile phones. *Procedia Computer Science*, *151*(2018), 303–312. https://doi.org/10.1016/j.procs.2019.04.043

Ye, F., Dong, X., Shen, J., Cao, Z., & Zhao, W. (2019). A Verifiable dynamic multi-user searchable encryption scheme without trusted third parties. *Proceedings of the International Conference on Parallel and Distributed Systems - ICPADS*, *2019-Decem*, 896–900. https://doi.org/10.1109/ICPADS47876.2019.00131

Zhan, J., Fan, X., Cai, L., Gao, Y., & Zhuang, J. (2018). TPTVer: A trusted third party based trusted verifier for multi-layered outsourced big data system in cloud environment. *China Communications*, *15*(2), 122–137. https://doi.org/10.1109/CC.2018.8300277

Zhang, P., Alkubati, M., Bao, Y., & Yu, G. (2021). Research advances on blockchain-as-a-service: architectures, applications and challenges. *Digital Communications and Networks*. https://doi.org/10.1016/j.dcan.2021.02.001