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].

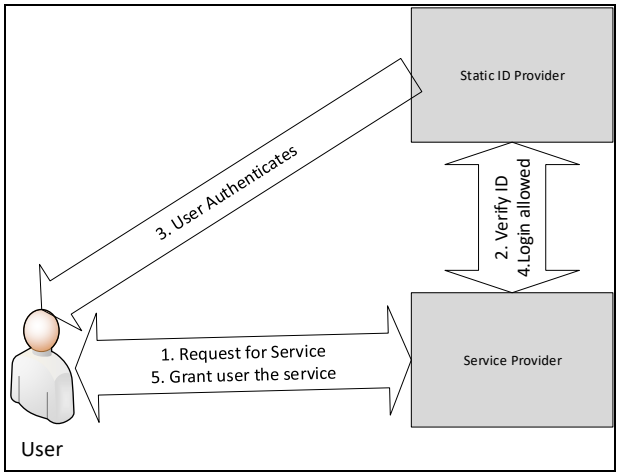

Fig 1: Static ID Provider Concept

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.

The Block chain technology requires some modifications to make it feasible.

Further, Block chain technology has an inherent scalability challenge due to its design. The solution proposed by Frank and Michael seeks to solve a global problem by providing access that is global hence scalability is very key to accommodate everyone who desires to employ 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 [6].

Another critical drawback with Block chain technology is time lag due to its design; 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 transactions delaying to complete [7]. 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. It can maybe be built as a service where costs can be shared [8].

Most e-commerce platforms hold 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 [9]. Attacks have been orchestrated against e-commerce platforms such as e-bay who suffered a Distributed Denial of Service (DDoS) attack where their databases were scanned and client data was exposed [10] . The attack was possible as the databases were accessible via the internet.

Fanghan et'al proposed a model 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 [9]. Even though the data is encrypted, if the key shared finds itself in wrong hands, then the encrypted data, if accessed can be deciphered. In addition, it means several user 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 [11].

Locher et'al, proposed the use of a distributed ledger to protect data [12]. The distributed ledger owes its security to the requirement of consensus being reached before any transaction is approved. This approach, nonetheless, results in delayed transaction confirmation as well as huge resources being required to make the technology operational [6]. Locher et'al acknowledged that despite the distributed approach of using block chain technology, users still needed to trust each other [12]. The aspect of trust is what the Trusted Third Party model aims to address.

More scholarly works were reviewed 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 in conjunction with existing data security techniques [13]. 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 of additional information. 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 [13].

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 [14]. 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 and encryption of pseudonymized data [14].

There is 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.

III. Solution design

As long as information is available online, it remains susceptible to leakage. The best way to protect information from leakage is to make it inaccessible online [15]. The proposed design in Fig. 2 below seeks to achieve that.

The use of a Trusted Third Party to hold PII demands that other service providers requiring KYC confirm with the KYC Agency if the requesting party is genuine. The KYC agency provides assurance without sharing the PII of the requesting party. This enables the requesting party to have access to services without sharing their PII.

The solution will operate as outlined below:

The user will first register with the KYC Agency in their country of residence by submitting their PII such as their National IDs, Residential address, Contact details like phone numbers and email addresses.

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 as depicted in Fig 2.

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

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

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.

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. When the online system receives the pseudo ID, eIDs, it will automatically create a record with the eIDs as the primary key.

The Data protector ((Restricted Memory System) 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 as depicted in Fig.3. 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 system hosting the sensitive data [1].

Furthermore, despite data being able to flow towards the online system from the offline system, to prevent huge amounts of data from being sent using the offline system, there is a bandwidth restriction imposed between the two systems. If, for example, 10gigabyte of data was to be sent from the offline system to the online system via a serial connection of 9600bps, it would take more than 100 days to complete the transfer. Sending of pseudo IDs would take few milliseconds as they only constitute few kilobytes per unique record created at any given time. The slow rate of data transfer would be a deterrent to a hacker. Moreover, the system would periodically reset the connection between the two hence disrupt any exploitive data transfer in session.

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 using an enhanced TOTP 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, eIDr, 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.

To protect the user from being profiled based on their location, sites visited, the identity of the device they are using and so on, Onion Routing (TOR) can be used optionally. TOR masks any identifying data about the user and hence help maintain their privacy [16]**.** The data hidden includes identifying information about the devices being used the client thereby addressing the challenge of profiling.

Fig.4 on page 6 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 periodically destroy emails containing random IDs after the predetermined validity period elapses.

IV. METHODOLOGY

Only the Restricted Memory System (RMS) was built as other components such as ecommerce sites could be built using existing solutions.

The Data protector (RMS) was put together using the following; two Arduino UNO microcontrollers, copper cables, serial ports, serial monitors, python programming language, Arduino UNO IDE and Proteus Simulation software.

Serial communication was preferred over parallel communication in building the RMS to ensure that data flows in one direction only. Two way communication would require two cables physically connected between the two devices. To enforce one-way communication, one cable was disconnected. This would ensure that even if the online component attempted to communicate to the offline component, the communication would fail.

Two microcontrollers were employed instead of one. This was to control the security of the system as hardware is susceptible to hardware Trojans. These viruses can be embedded into the hardware when being manufactured to deliberately leak information later. For the Trojans to be activated, one would need access to the hardware either physically or remotely [17]. 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 Fig. 5 below. 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.

Tests were conducted as follow: Data was sent from the offline system to the online system via the RMS. 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.

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

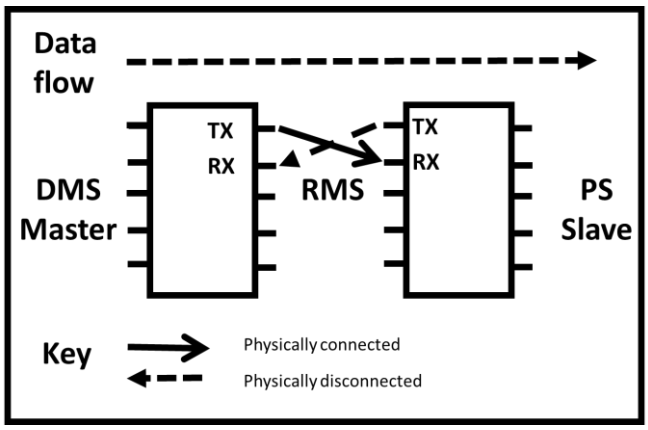

Fig 5 Physical Configuration of RMS

V. GENERATION OF RANDOM IDS

The generation of Random IDs was to be 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 [18]. Our objective was to make the random OTP act as an identifier when need arise without comprising the privacy of the user involved. The approach is to generate the password as proposed by the standard then modify the resulting password by passing it through a function that modifies it with an ID for that particular user as shown in Fig.6 below.

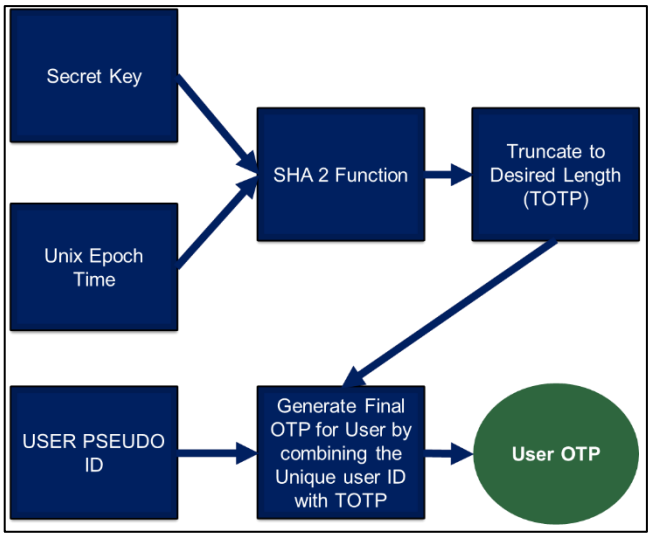

Fig 6 Generation of Random IDs (User OTP)

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 need arise.

Despite the use of random IDs, the IP address and other unique properties of the computer being used can help trace and profile user such as determining where they are connecting from, what sites they are accessing and so on. To help address this challenge the use Onion Routing (OR) is recommended. OR will mask the source of the requests to access services hence prevent online profiling. The diagram below shows how OR works.

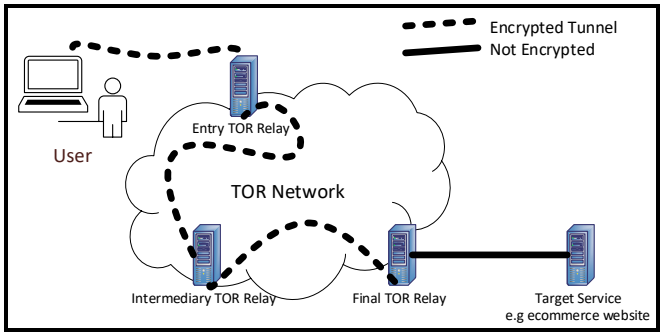

Fig 7 How a TOR Network Works

By masking the actual of the requests, the user is shielded from being identified online not only in terms of user name but also other profiling elements such as sites visited, services accessed, from where and so on. This helps further enhance user privacy

VI. RESULTS AND DISCUSSION

Table. 1 provides a summary of Results based on Experiments conducted.

In Test 1 data was successfully sent both ways with the transmitting and receiving PINs connected correctly on both the Master and Slave Arduino. That is, data was able to flow from the offline system holding PII to the online system susceptible to online hacking. 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 RMS cannot protect PII by preventing access by users connecting from the Internet. It is vital that data does not flow from the online system to the offline system even if bandwidth is restricted as malware can be created as a very small payload and yet cause serious damage to data once deployed into the system storing PII. The test results for the first scenario make the configuration undesirable.

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. 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

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

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 if data has to be moved using external media. This configuration is also not suitable.

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. 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. 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.

For this solution to be effective, online data must be anonymous so that if the records are leaked, no identifying information would be part of the leaked records. Furthermore, the use of the modified TOTP based on the RFC6238 standard to generate random IDs will help ensure the privacy of users is maintained and at the same time make it possible to retrieve the actual identity of the user if they were to abuse their anonymity while online.

VII. Conclusion

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 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. Furthermore, the use of the modified RFC6238 for the generation of random IDs makes it possible for the user to maintain their privacy while at the same time provide a possibility to trace a fraudster hiding behind being anonymous whenever need arise. The use of TOR can further help achieve enhanced privacy for users.

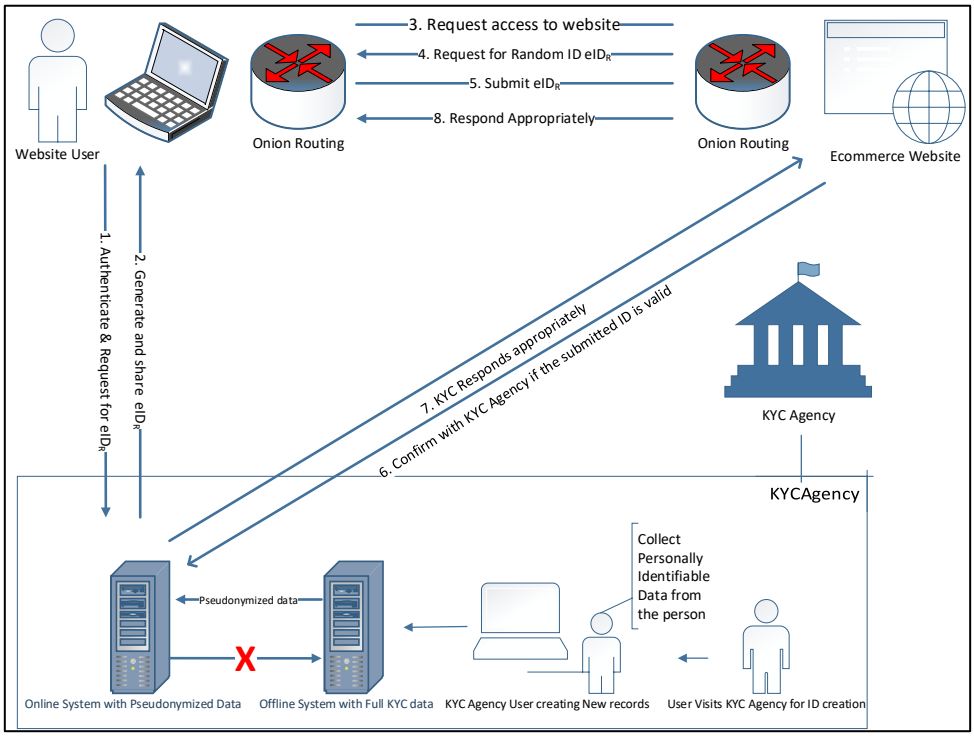

Fig 2: Know Your Customer Agency Operation

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

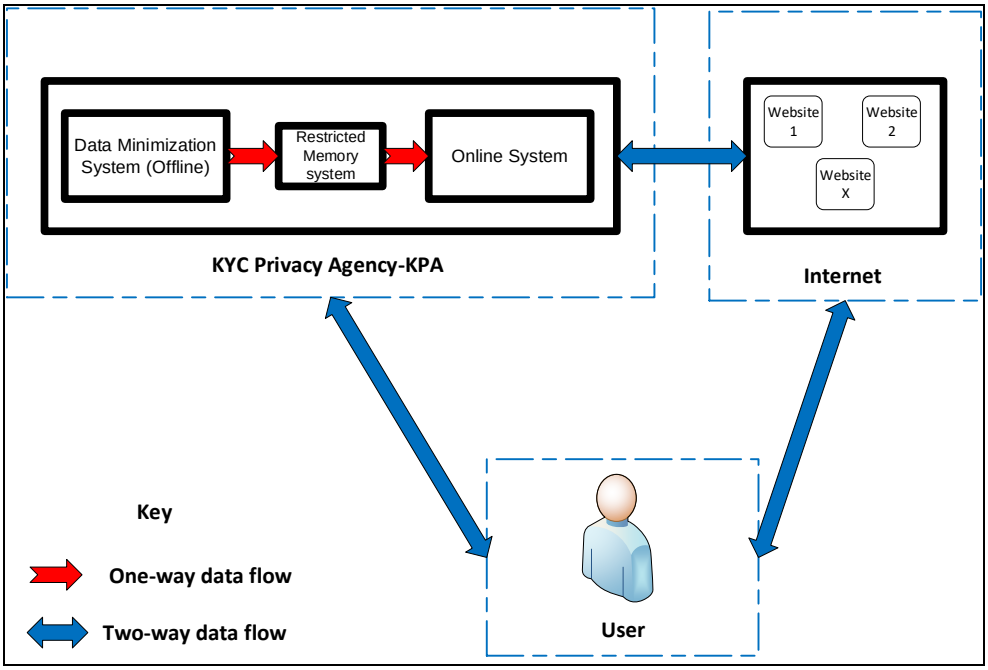

Fig 3: Data Protector Operations

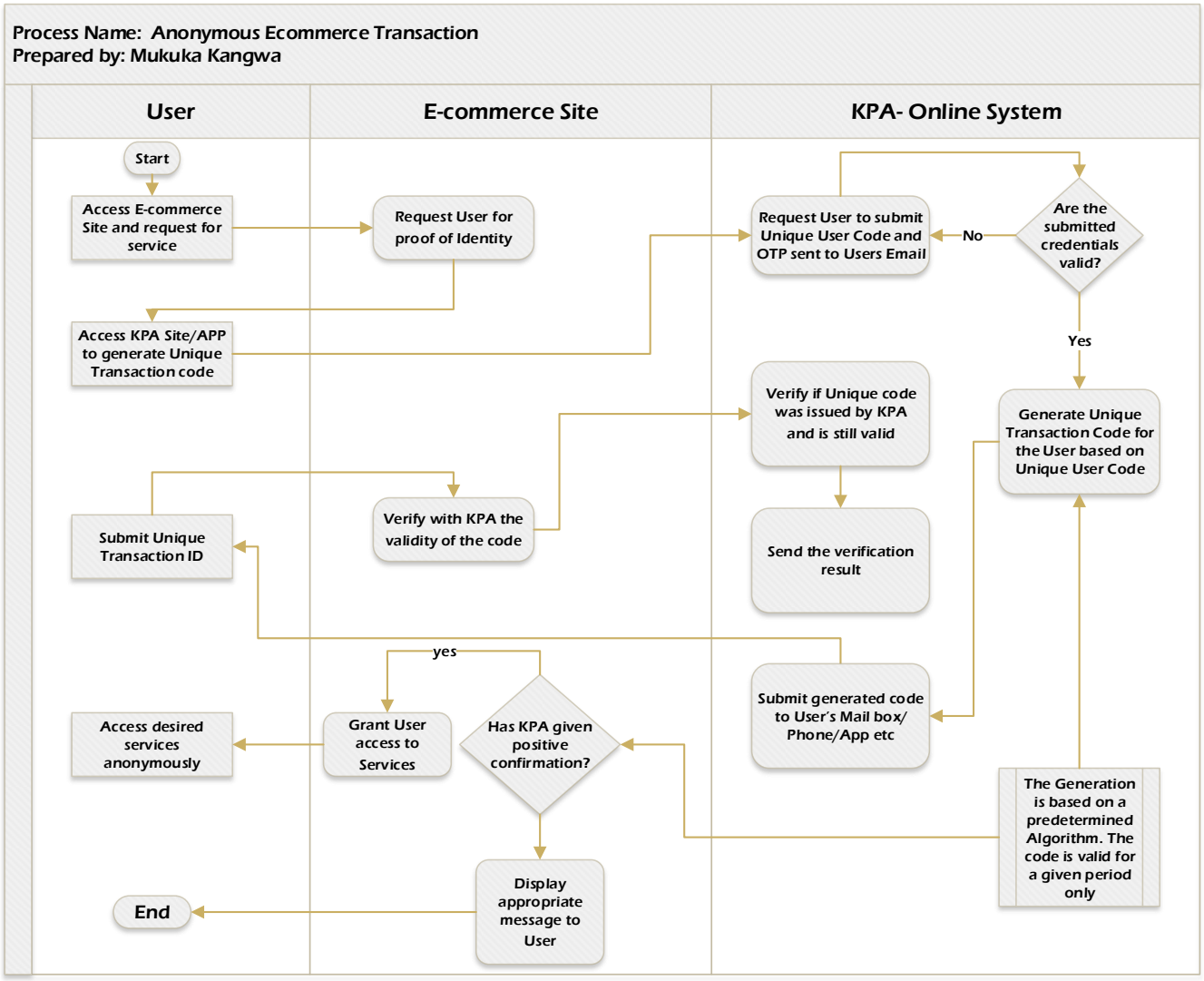

Fig 4: Anonymous Ecommerce Transaction

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

Table I: Experiment Results

No	Test |Scenario	Connections/Setup	Results	Result
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	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	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 Slave towards the Master is disconnected	1. Data sent from Master successfully reached the slave 2. Data sent from the Slave didn't reach the Master	Pass
4	Amount of data Transmittable	Speed set at 9600	Sending 10GB estimated at more than 100days	Pass
5	Generation of Random IDs	Use system based on RFC6238 Result further processed using user ID	Generate OTPs with embeded Unqiue user ID	Possible
6	Decipher user ID	Use reverse of functional to process TOTP using user ID	Should be able to generate User identifier only known to the KYC urgency	Possible

REFERENCES

[1] 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.

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

[3] 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.

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

[5] 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.

[6] 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.

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

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

[9] 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.

[10] 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.

[11] 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.

[12] 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.

[13] 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.

[14] 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.

[15] M. Kangwa., C. Lubobya., and J. Phiri., "Enhanced Protection of Ecommerce Users' Personal Data and Privacy using the Trusted Third Party Model," in *Proceedings of the 18th International Conference on e-Business - ICE-B*, 2021, pp. 116–126.

[16] 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.

[17] 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.

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

[19] P. Winter, "Enhancing Censorship Resistance in the Tor Anonymity Network Enhancing Censorship Resistance in the Tor Anonymity Network," Karlstad University, 2014.