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

by its handlers. GDPR proposes such methods as pseudonymisation 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 available[7].

Studies have been conducted on how best to protect ones PII while transacting using e-commerce platforms and various approaches have been proposed to achieve the desired user privacy. Inadequate security in e-commerce can be a huge deterrent to its adoption hence the need to find solutions that will provide sufficient protection and privacy[15].

A few sections below give some of the solutions that have been implemenetd or merely proposed and their associated gaps that our proposed solutions seek to address.

(i). Electronic ID (eID)

S. Nimalaprakasan et al [13] discussed the concept of an electronic ID (eID) which would mimic the National Identification Cards that are issued to citizens in each country for the purpose of identity. The eID would work like the conventional ID except that it would be electronic and get used to acquire electronic services such as having access to e-commerce and e-government services [13]. The challenge with the use of the eID solution is that users can be tracked and profiled using their eIDs and hence have their privacy violeted.

(ii). Pseudo ID System

S. Nimalaprakasan and team [13] 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 decreased depending on the service they are trying to access. The Pseudo name would be different from the real ideintity of the user but can be associated with the actual user details by the handlers of the data [13]. However, as long as the ID used, even though pseudo, is static, it is possible to profile the owner of the pseudo name and hence have their privacy compromised.

(iii). 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 for that transaction only. This approach addresses that challenge of profiling that comes with static Pseudo names. 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 untraceble [12]. However, this alone 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 and hence making it susceptible to hacking [14]. We seek to improve on this solution by building an actual algorithm and prototype that will generate Pseudo codes that will keep them traceable to the actual user so that incase of a fraud being committed by this user, they can be traced via the KYC agency that would have the PII for the user involved.

(iv). Data Leakage Prevention (DLP) Systems

Another approach to addressing leakage of data has been the development of Data Leakage Prevention Systems (DLPs). This class of technologies uses various algorithms to prevent leakage of sensitive data.For example, credit card data can be blocked from being sent or an alert is sent to appropriate personel. 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 actually slow down the processing speed of a system as it will have to scan, scrutinize and analyse every data element traversing it [8]. Furthermore, DLPs are usually inadequate to prevent leakages of sensitive data via Peer to peer (P2P) networks as these normally use various ports including the port used by the famous internet protocol http. It would therefore, be dificult to block various ports as that would affect several services [16].

III. PROPOSED SOLUTIONS (METHODS)

The system has been designed to use software and hardware as a way of logically and physically separating the PII from the online side of the system to prevent intentional or accidental leakage of PII. Further, the online system will use an algorithm to generate random IDs to provide online privacy while users access electronic services such as e-commerce.

The first step will be to put together an implementable solution that will use physical and logical algorithms to protect data as well as facilitate tracing back to the users involved if need arises. The design will then be implemented as a prototype that will in turn be subjected to various ICT security tests to ascertain the effectiveness of the proposed approach. That is, determine how susceptible to leakage data protected by the solution is as well as ascertain how ones privacy is protected while they are conducting online transactions.

System Design and Operations

The system will basically consist of three subsystems connected logically and physically. That is, the Data Minimisation System (DMS), RMS and the Pseudonymisation System (PS) as shown in Figure 2 below.

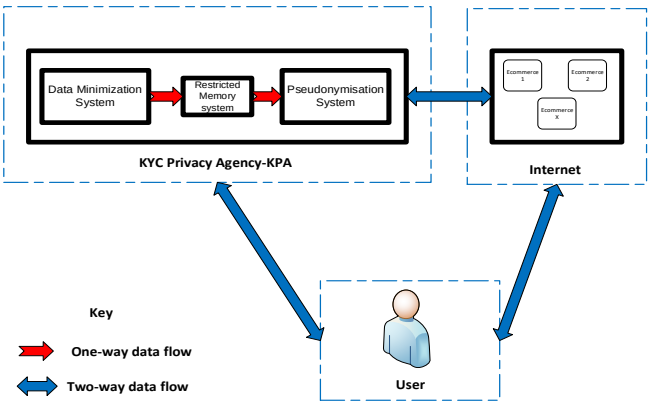

Figure 1. Block Diagram of a Data Minimization and Pseudonymisation System

The DMS is completely offline while the PS is online and interacts with the Users and the E-commerce sites. Information flow between the DMS and PS is one way towards the PS direction. The data that can pass from DMS to PS is restricted by the RMS with limited memory joining the two systems. This is to ensure that it is not possible to transfer huge amounts of data from the DMS, which is offline, to the PS which is online. The RMS ensures that data can only be written into its memory by the DMS and that the DMS cannot read data from its memory while the PS can neither write nor read data from its memory. Only the firmware program within the RMS can read the data in its memory and write in the memory of the PS. The RMS is not allowed to write in its memory nor read from the PS memory. Transfer of data can be through a worm that has been deliberately copied into the DMS. However, this is minimized by restricting data to only flow in one direction. This will be achieved by the use of the physically constructed RMS that will prevent the flow of data from the online section to the offline section by only connecting physical PINs that need to receive data and physically disconnecting PINs that transit as shown below.

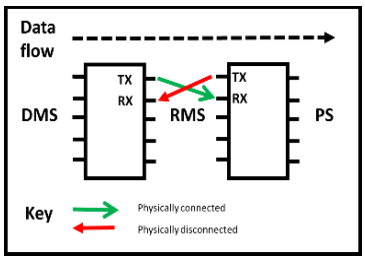

Figure 2: High Level Circuit Design of Proposed Solution

The physical seperation ensures that even if there was an attempt to transmit data in the wrong direction, the physical setup of the PINs wouldn't allow.

IV. MATERIALS AND METHODS

The proposed approach focuses on the use of Pseudonymisation as the basic principle on which the solution to protect e-commerce users' PII is being developed. Pseudonymisation makes it difficult to trace leaked data to the original owner if implemented properly [7]. The research seeks to design a model which, if implemented, would effectively protect users PII and privacy while allowing them to enjoy the services of e-commerce without compromising quality.

The research deliberately prefers to use Pseudonymisation to Anonymization in the design of the solution as 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 [7]. It is important to ensure that a user of e-commerce can be traced via relevant authorities when need arise [13]. Some users can commit fraud hence the need for them to be traceable. To address the issue of the third party holding user PII being untrustworthy, the study proposes the use of a KYC Privacy Agency (KPA) appointed by governments to use the proposed solution. This approach uses what is called the Trusted Model in as far as User Trust is concerned. The users must have trust in the appointed third party[14]. The Figure below depicts the concept of using a KPA.

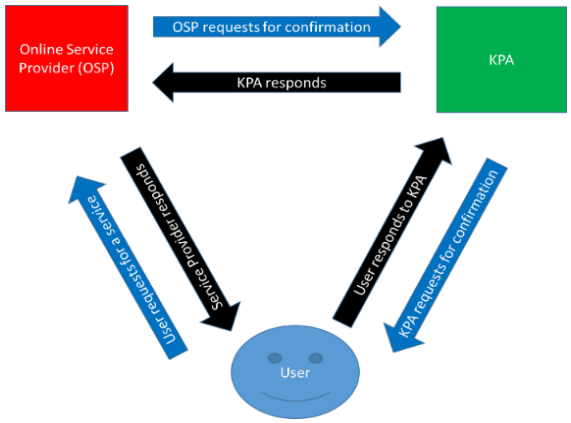

Figure 3: Use of KPA as a Trusted Third Party

An e-commerce service provider uses a third party to conduct Know-Your-Customer/Client (KYC) on the prospective users of their e-services. This solution, therefore, can be used by any institution offering e-services and require KYC confirmation to be done before granting a client access to their services. The institutions would confirm with the Third party if they know the prospective client.

The solution will be implemented as follows: The hardware component will actualized using Arduino circuit boards. The software component will be setup using Python language. To test the effectiveness of the implementation, Kali Linux will be used as it comes with adequate hacking tools that can be employed for exploiting vulnerabilities.

The assumption is that the hacking tools to be used will be effective enough to exploit any vulnerability on the system built so that the results can be relied upon to determine how effective the solution is.

The constraint with the proposed methodology is that tests will only be done for known vulnerabilities exploitable using the tools to be used such as Kali Linux and perhaps a few other tools to compare and contrast. The challenge with this is that some tools might pick certain vulnerabilities while others might not.

V. CONCLUSIONS

A successful implementation of the ODMS will greatly help resolve the prevalent challenge of PII being leaked into wrong hands every now and then and enhance the privacy of user conducting online transactions. The leakage of PII and online profiling has resulted into a number of users suffering financial and reputational damage. Some users are discouraged from using e-commerce because of the possibility of having their private information leaked and their privacy violated. If the challenge is resolved and the users are made to feel comfortable using the system, more users will adopt e-commerce. 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 eliminated to the benefit of the users.

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.

REFERENCES

[1]. 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.
[2]. 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.
[3]. 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.
[4]. 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.
[5]. 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.
[6]. 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.
[7]. 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.
[8]. S. Alneyadi, E. Sithirasenan, and V. Muthukkumarasamy, "A survey on data leakage prevention systems," *J. Netw. Comput. Appl.*, vol. 62, pp. 137–152, 2016.
[9]. A. Muneer, R. S, and F. Z, "Data Privacy Issues and Possible Solutions in E-commerce," *J. Account. Mark.*, vol. 07, no. 03, 2018.
[10]. 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.
[11]. 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.
[12]. M. S. Ackerman and D. T. Davis, "Privacy and security in E-commerce," *Market/Tržište*, vol. 21, no. 2, pp. 247–260, 2009.
[13]. 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.
[14]. G. Navarro-Arribas and V. Torra, "Preface," *Stud. Comput. Intell.*, vol. 567, pp. 423–442, 2014.
[15]. S. W. Khan, "Cyber Security Issues and Challenges in E-Commerce," *SSRN Electron. J.*, p. 2019, 2019.
[16]. 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.