ICCSE'21 International Conference on Contemporary approach on revolutionary Science and Engineering, April 9-10, 2021 ISBN: 978-81-910765-1-6

Secured E-Commerce Application for User Behaviour Analysis

[1] G.Anitha ,[2] M. Mohamed Riyaj Khan ,[3] S.Mukesh Kanna ,[4] K.Santhiya, [5] S.Saroja Devi.

[1]Final year UG Student , [2] Final year UG Student ,[3] Final year UG Student , [4] Final year UG Student, [5] AssistantProfessor

Indra Ganesan College Of Engineering , Tiruchirapalli , Tamil Nadu , India.

ABSTRACT

The explosion in the e-commerce industry which has been necessitated by the growth and advance expansion of Information technology and its related facilities in recent years have been met with adverse security issues thus affecting the industry and the entire online activities. This project examines the prevailing security threats e-commerce is facing which is predominantly known as cyber-crime and how computer related technology and facilities such as digital forensics can be adopted extensively to ensure security in online related business activities. Attacks on customer sensitive information has the adverse effect of decreasing the consumer faith on online transactions, which happens in e-commerce. Prevention is better than cure, good knowledge and understanding of the online threats can be used as a tool to avoid online attack. The main hindrance in the growth of e-commerce is cyber fraud and identity theft. Hackers are people who carry out the cybercrime; hence, poor security on e-Commerce web servers and use of the computer is a core issue to be resolved for the rapid growth of e-commerce. This paper provides directions for ecommerce security to improve customer confidence in e-commerce shopping.

Keywords: User Behaviour Analysis, Cyber Security, E-commerce,

I. INTRODUCTION

CYBER ATTACK IN E-COMMERCE

The emergence and popularization of the world wide web (WWW) has facilitated a drastic explosion in E-commerce and consequently has become an essential and indispensable constituent of business strategy and operations, hence the recent advancement in businesses, thus the main medium for trade and industry development in any given country and the world at large (Guan, Tan, & Hua, 2004). Due to this rapid growth in the e-commerce industry, many computer programmers and Information Security experts have directed their abilities to the development of applications that enhances computer assisted transactions. The upsurges in cyber-crime by unknown assailant who usually take the advantage the internet offers and in anonymity perpetrate heinous crime under the notion that they cannot be seen, most crimes are perpetrated with this belief system. (Perez, 2005). There are various crimes that are committed via the internet and its related technologies from denial of services, identity theft, Privacy violation and intrusion as well

 ISBN: 978-81-910765-1-6

as industrial and financial espionage, this study expounds on these crimes further in the following sections and in appendix 1. Perpetrators of these crimes do it for various reasons with the underlining fact of illegally making money for themselves among other reason.

CONSUMER MARKETING

Marketing is the art that persuades and provides customer satisfaction which is transformed to customer delight. The traditional marketing approaches have been replaced by the advent of latest technology wherein the buyers and the sellers meet at a virtual market through World Wide Web. This trend of product exchange had reduced the influence of middlemen over sales thereby gaining profit to the business and satisfaction to the consumers. India is ranked second with 462, 124,989 internet users in the year June 2017 and ASSOCHAM report (7th January 2016) the high internet penetration with high disposable income in tier II and tier III cities will lead approximately to 100 million transactions in 2020, in India which shows the rapid growth of internet.

ONLINE USER BEHAVIOUR

Online user behavior analysis studies how and why users of e-commerce platforms and web applications behave. It has been widely applied in practice, especially in commercial environments, political campaigns, and web application development. Data aggregation is one of the most critical operations in behavior analysis. Nowadays, the aggregation tasks for user data are outsourced to third-party data aggregators including Google Analytics, comScore, Quantcast, and StatCounter. While this tracking scheme brings great benefits to analysts and aggregators, it also raises serious concerns about disclosure of users' privacy. Aggregators hold detailed data of users' online behaviors, from which demographics can be easily inferred to protect users' privacy, government and industry regulations were established, Do-Not-Track [7], which significantly restricts the analysis of users' online behaviors [4]. To address the conflict between the utility of analysis results and users' privacy, much effort has been devoted to designing protocols that allow operations on user data while still protecting users' privacy. Unfortunately, existing schemes guarantee strong privacy at the expense of limitations on analysis. Most of them can only compute summation and mean of data over all users without filter or selection, i.e., overall aggregation. Some previous methods allow more complex computation. However, selective aggregation is one of the most important operations for queries on databases. It can be used to tell the difference among different user groups in a certain aspect. For instance, "select avg(income) from database group by gender".

II. LITERATURE SURVEY

A).PERSONALIZATION OF WEB SEARCH BASED ON PRIVACY PROTECTED AND AUTO-CONSTRUCTED USER PROFILE

Search engines are widely used to find huge amount of data on the web in a minimum amount of time. But sometimes it becomes difficult to the users to get exact result for given query. Personalized Web Search (PWS) provides the better search results for individual user needs and improves the quality of the search result based on the user profile. However, user's unwillingness to disclose their private information during search has become major issue in wide increase in PWS. This paper presents a scalable way for users to build rich hierarchical user profiles automatically and provide privacy for the user profiles. Extended User customizable Privacy-preserving Search (E-UPS) framework generalize the user profiles for each query according to user-specified privacy requirements. It will improve the search quality and hide the privacy contents existence in the user profile and gives protection against a typical model of privacy attack.

To address this, nowadays more and more online publishers launch their anti-ad blocker project, for example, RateMyProfessor, Wired, Forbes, and Digiday. By analyzing JavaScript codes on the websites, Rafique et al. (2016) found that anti-adblocking scripts were used by 16.3% of the 1,000 domains they crawled. They adapted similar strategies: they ban users who try to view their content with active ad blockers. If the website detects a user has an ad blocker, it would pop up a message to request the user to turn off or at least pause the ad blocker, i.e. whitelist, in order to view the content. If the user rejects to whitelist, s/he would be forbidden to access the content they intend to view. As "counter-ad blocking wall" prevents users to read the content without ads, it can result in loss of readers and the web traffic to their websites. Thus some publishers abandon their counter-ad blocking

 ISBN: 978-81-910765-1-6

strategies after conducting for a while. Some other publishers decide to provide incentive to encourage customers to whitelist them, for example, a "less-ads" promise. Lacking of good solutions, both publishers and advertisers have great interests in understanding the usage of ad blockers and the impact of "counterad blocking wall" to user behaviors, however, which is an open question. Thanks to the technological advancements in digital advertising, publishers and advertisers can analyze consumers' digital footprints at a more granular level, at a large scale easily. In this study, we employed a quasi-experiment framework in cooperation with Forbes Media and collected a large size of real-life data. With the abundant behavior data, we can study customers' usage of ad blockers and activities detailed. For example, we are able to measure environmental factors (like device, traffic origins) of each web visit, which is infeasible to obtain through traditional survey methods.

Our data set is collected in collaboration with Forbes Media, a large online publisher. We employed a quasi-experiment method that avoids the self-selection and other treatment selection biases. Forbes blocks ad-blocker users to access the site. We collected the data for one week in August 2016. During the experiment period, we use a JavaScript program to detect the existence of ad blockers. All users who enter the website for the first time would be forwarded to a welcome page. For non-ad blocker users, they can view the content directly after the welcome page. If the website detects an ad-blocker, the welcome page will pop up a message "Adblock Detected" to inform the current user to pause or turn off the ad blocker in order to view the content of that website. Once an ad blocker is disabled, users would receive a message "Thank you" and they are promised the "Adlight experience" for 30 days, with fewer ads, and ads that typically load faster. If a user rejects to disable the ad blocker, s/he will be prevented from viewing any online content in that website.

Therefore, the ads may be perceived to be more annoying to young people. Another possible reason is that young people are more familiar with computer technology. And they know well how to install ad blocker plugins or software. As time flies, more and more young users will replace old users who are not familiar with technology. It explains why the rate of ad blocker usage is keeping growing quickly. An interesting finding is that we found customers who older 65 have a slight higher usage of ad blockers compared with mid-age customers. A possible reason is that the influence of age to ad blocker usage is interacted with another variable, for example, age. Maybe most of the users older than 65 of Forbes are male instead of female. And thus with that curiosity about aged customers, we will study the interaction between age and other factors in the second phase.

The second phase is to study the interaction effects among user profile, browsing behavior patterns, device features on ad blocker

B). BIG DATA TECHNOLOGY AND ETHICS CONSIDERATIONS IN CUSTOMER BEHAVIOR AND CUSTOMER FEEDBACK MINING

The rapidly increasing attention to customer behavior and satisfaction by department stores and commercial companies and development in social media as well as online systems has promoted production and research in user engagement pattern recognition, user network analysis, topic detection from customer feedback, text-based sentiment analysis, etc. With the development of Internet of Things and social media network, ethics consideration has also playing an important role in application of big data, especially customer behavior and feedback mining. Our proposed novel system, which extracts the important topics or issues from Skype customer feedback sources and measures the emotion associated with those topics using Vibe metric, can be a good example in this area. Unlike other previous research, which has focused on extracting the user sentiments either globally or in separate topics, our work focuses on tracking the correlated emotional trajectories across all the important issues from Skype customer feedback over time. Moreover, it also provides a platform for both studying customer emotions and tracking how the emotions regarding different important topics correlatively change over time by leveraging unstructured textual customer feedback data and structured user activity telemetry data.

C) USER BEHAVIOR MINING ON LARGE SCALE WEB LOG DATA

This paper propose a web log mining-based network user behavior analysis scheme, which plays an important role in network structure optimization and website server configuration. Based on clustering and regression model, we studied the network user's visit model in a university by analyzing a large

 ISBN: 978-81-910765-1-6

amount of web log data which is collected from the university campus network. The data analyzing software was developed by VC and SQL Server 2000.

This paper discussed how to regressively analyze the log data which is collected from the web server in the university network center and processed by the software developed by ourselves. According to the theory of Internet broadcasting and storage (IBS)[2], which is proposed by Prof. Youping Li. The data resources for broadcasting and how to build Internet users' interest model play important roles in IBS. At present, there is a large amount of web data on the Internet. But we can not broadcast all the data from the broadcasting network. Through the analysis of user's behavior, we found that it's unnecessary to broadcast all the data. If we just broadcast the data which is interested by all users, it will help us improve the efficiency of users' Internet usage and alleviate the Internet overhead. So, all the users can get more personalized services.

We can mine and analyze the log data from the web server by using the technology of web data mining and behavior analysis of network data. Through this way, we can find out the information related to the user's behavior, which is valuable for studying the structure of the network.

DISADVANTAGES

- In existing system, the user behavior analysis will be done using item set mining.
- Thus the user privacy will be known to anyone who mines data from big data.
- The mining implements pattern based technique to mine the data.

III. PROPOSED SYSTEM

Growing efforts have been devoted to mining the abundant behavior data to extract valuable information for research purposes or business interests. However, online users' privacy is thus under the risk of being exposed to third-parties. The last decade has witnessed a body of research works trying to perform data aggregation in a privacy-preserving way. Most of existing methods guarantee strong privacy protection yet at the cost of very limited aggregation operations, such as allowing only summation, which hardly satisfies the need of behavior analysis. . The proposed system uses scheme PPSA and ECC, which encrypts users' sensitive data to prevent privacy disclosure from both outside analysts and the aggregation service provider, and fully supports selective aggregate functions for online user behavior analysis while guaranteeing differential privacy. We have implemented our method and evaluated its performance using a trace-driven evaluation based on a real online behavior dataset. Experiment results show that our scheme effectively supports both overall aggregate queries and various selective aggregate queries with acceptable computation and communication overheads.

ADVANTAGES

- The user privacy on the data mining pattern will be achieved up to 94%.
- Thus the unwanted add will be blocked from the website user's side.
- The search count can only be aggregated by the aggregator and provide to the analyst

MODULES DESCRIPTION

ENTITY REGISTER AND LOGIN

Here the entities are user, admin and intermediate where they will be the main users. The intermediate will be of three people authority, aggregator and the analyst. Here the analyst work is to get the count of the product searched. Here the data will be stored with the efficient storage procedures with the entity access with registration. Aggregator aggregates and views the data. Authority secures the data and the admin is the super user of each websites such that the user data are get stored.

USER SEARCH

Usually user buys the product with the online E- commerce websites. The product searched data are get stored in the database where extracting data from the data storage. The user behavior analysis is made here where the needs of the user are get known using the data searched. Here the user search can be serious problem since their behaviors are misused by the other website owners by the deceiving advertisements. Here the searched online data privacy is added for the user data with the privacy preserving scheme for the aggregators.

KEY GENERATION

 ISBN: 978-81-910765-1-6

The intermediary comprised of an aggregator and an authority, bridges clients and analysts. They are in charge of aggregating user data from clients and responding to queries of analysts. They provide both functionality and security. Clients are installed on the user side. They can be bundled with users' software that requires private analytics. Thus, it is reasonable to assume clients are trusted. A client collects a user's data, detects and removes outliers. Once the user gets online, the client sends encrypted data to the intermediary. Clients are not involved in the process of statistical aggregation. Here the data encryption is added using the ECC based encryption technology.

ECC ENCRYPTION

the

Elliptic-curve cryptography is an approach to public-key cryptography based on the algebraic structure of elliptic curves over finite fields. ECC requires smaller keys compared to non-EC cryptography to provide equivalent security

ALGORITHM

We assume that those who are going through this article will have a basic understanding of cryptography (terms like encryption and decryption). The equation of an elliptic curve is given as,

K=E(n)+P
(6.1)

E -> Elliptic Curve

P -> Point on the curve

n -> Maximum limit (This should be a prime number)

Key generation is an important part where we have to generate both public key and private key. The sender will be encrypting the message with receiver's public key and the receiver will decrypt its private key. Now, we have to select a number'd' within the range of 'n'. Using the following equation we can generate the public key

Q = d * P
(6.2)

d = The random number that we have selected within the range of (1 to n-1). P is the point on the curve.

'Q' is the public key and 'd' is the private key.

Let 'm' be the message that we are sending. We have to represent this message on the curve. This have in-depth implementation details. All the advance research on ECC is done by a company called certicom. Consider *'m'* has the point *'M'* on the curve *'E'*. Randomly select 'k' from [1 – (n-1)].

Two cipher texts will be generated let it be C1 and C2.

C1 = k*P
(6.3)

C2 = M + k*Q
(6.4)

C1 and C2 will be send.We have to get back message 'm' that was send to us,

M = C2 – d * C1
(6.5)

M is the original message that we have send.

SEARCH ENCRYPTION

The data are get encrypted using the ECC (Elliptical Curve Cryptography) algorithm**.** Here the data are get encrypted using a 256 bit hash code value system**.** The user analysis are get eradicated using this enhancement techniques so, that the data will be shown as a cipher text to the analyst. Selective aggregation literally refers to selecting the users who satisfy some conditions before aggregating their values, e.g., "the average amount of time online of all the male users". Herein, "male" is a condition to pick out target users. We suppose there is a centralized table T that contains attributes and collects users' answers to them. Attributes (denoted by att) can only be numeric, because non-numeric attributes cannot be directly aggregated. Boolean attributes are a special type of numeric attributes, to which users' answers are boolean values (0 or 1), e.g., "gender is male". A male user's answer would be 1. Most categorical attributes can be easily transformed into boolean attributes. For example, education level can be decomposed into several boolean attributes: "education level is bachelor", "education level is master", etc. Differential privacy is independent of the adversary's computational power and auxiliary information so it is a very strong guarantee.

SEARCH COUNT

Analysts are individuals or institutions that want to query about user data. An analyst sends a query Q to the intermediary and then receives a noisy

 ISBN: 978-81-910765-1-6

answer from it. Analysts are assumed to be semi-honest, trying to learn individual users' privacy. An analyst may collude with other analysts or make one single query multiple times. The analyst is the one who will be authorized by any of the organization who will analyze the data count and should make visualization. The analyst will get data from the aggregator who will aggregate the data and sold it to the analyst. Now the aggregator is the one who will search for the user searched contents and the searched content are analyzed. The user searching will be produced with the count value to the analyst and the analyst can only get the count value.

CONCLUSION

In this paper, we have described the challenges of making online user data aggregation while preserving users' privacy. Based on ECC homomorphism cryptosystem, we have designed the first system that is able to securely and selectively aggregate user data, making it practical in realistic data analytics. It guarantees strong privacy preservation by utilizing differential privacy mechanism to protect individuals' privacy. We have presented PPSA to evaluate aggregation selected by one boolean attribute, and extended it to aggregation selected by multiple boolean attributes and by one numeric attribute. Extensive experiments have shown that PPSA supports various selective aggregate queries with acceptable overhead and high accuracy. The mining patterns are get eliminated and the count will be only produced to the analyst. This helps to avoid the deception of unknown advertisements in website.

FUTURE ENHANCEMENT

behavior

For the further research the user based advertisement can be divided as malicious advertisements and normal advertisements. Here the data will classify using the spam reports with good accuracy algorithms and system.

REFERENCES

[1] Ben Paul Miroglio ,David Zeber ,Jofish Kaye ,Rebecca Weiss, The Effect of Ad Blocking on User Engagement with the Web, proceedings of world wide conference, pp: 813-821, 2018.

[2] Edyta Abramek, Technical and Social Reasons for Blocking Web Advertising in the Context of Sustainable Development of E-Business, 2019.

[3] Rasika M. Kaingade ; Hemant A. Tirmare, Personalization of Web Search based on privacy protected and auto-constructed user profile, International Conference on Advances in Computing, Communications and Informatics (ICACCI), 2015.

[4] Rishab Nithyanand, Sheharbano Khattak, Mobin Javed, Narseo Vallina-Rodriguez, Marjan Falahrastegar, Julia E. Powles, Emiliano De Cristofaro, Hamed Haddadi, Steven J. Murdoch, Ad-Blocking and Counter Blocking: A Slice of the Arms Race, Proceedings of the 6th USENIX Workshop on Free and Open Communications on the Internet (FOCI 2016), 2016.

[5] Shuai Zhao, Ad Blocking and Counter-Ad Blocking Ad Blocking and Counter-Ad Blocking: Analysis of Online Ad Blocker Usage, Americas Conference on Information Systems, 2018.

[6] Shun-Hua Tan ; Miao Chen ; Guo-Hai Yang, User behavior mining on large scale web log data, International Conference on Apperceiving Computing and Intelligence Analysis Proceeding, 2010.

[7] Tong Gan ; Fuhong Lin ; Changjia Chen ; Yuchun Guo ; Yi Zheng, User behaviors analysis in website identification registration, China Communications (Volume: 10 , Issue: 3 , March 2013)

[8] Xin Deng, Big data technology and ethics considerations in customer behavior and customer feedback mining, Big data technology and ethics considerations in customer behavior and customer feedback mining, 2017

[9] . Yuanying Peng ; Ke Yu, User

analysis of automobile websites based on distributed computing and sequential pattern mining, IEEE International Conference on Network Infrastructure and Digital Content (IC-NIDC), 2016.