ISSN No. 0976-5697

Volume 4, No. 8, May-June 2013

International Journal of Advanced Research in Computer Science

RESEARCH PAPER

Available Online at www.ijarcs.info

Data Security approach for Right to Information of Developer

Vivek Aggarwal
B.Tech, Department of Computer Science and
Engineering,College of Engineering Roorkee, India
vkwal28@ymail.com

Abhimanyu Bhatia
B.Tech, Department of Computer Science and
Engineering,College of Engineering, Roorkee, India
abhimanyu.bhatia92@gmail.com

Abstract: Nowadays Data Security is a major field in IT sector. Data Security doesn't just mean password protection, data hiding, encryption or adding additional firewalls -- it also means having complete information about your data i.e. where your data is kept and who all view it. This paper consists of an effective technique to know who is accessing your data. As the user 'X' opens a document on the network, the IP tracker starts and finds the IP of 'X' by which we can also find his location. This information is stored in a database and converted into pdf format when the developer wants to access the records. When a predefined limit of database entries is reached, an email containing the records (in pdf format) is sent to the developer and the database memory is flushed.

Keywords: Protection; encryption; firewall; access; IP tracker; database; entries

I. INTRODUCTION

Data security has been a leading issue in the Information Technology arena because as users we don't want anyone to hinder our privacy and as developers we don't want anyone to use our work as their own. Data Security does not only mean password protection, data hiding, encryption or adding additional firewalls it also means having complete information about your data i.e. where is your data kept and who all view it. According to a leading data security expert not knowing who uses what data and where it is kept is a major data security issue. You cannot secure your data without knowing in detail how it is accessed through the network. If we maintain a record of when and where data has been accessed over a network, it can be utilized to enhance data security.

This paper comprises of an effective technique to know the complete details about the users who are accessing your data. There are many social networking websites who follow it but still it is not accessible to the user whose page it is and moreover why only to the social networking websites, the provider of even a small amount of data also has the right to know that who is accessing their information. This technique comprises of an IP tracker linked to the developer's page [1]. The information generated by that tracker is stored in the database which is accessible by the registered user.

As the user opens a document on the network, the IP tracker starts. The goal of IP tracker is to know the path of an IP packet from its origin to its destination. Identifyingthe sources of packets is a significant step in securing our data. This tracker helps in finding the IP of the user by which we can find location, time of viewing, domainname, lastvisited, etc. [2].The information is stored in a database with a certain entry limit say 1000 entries. When the developer requests to see the related information a query according to the developer's URL runs and searches it in the database. The selected rows are converted into Portable document format (pdf) and displayed to the developer with a save or discard option [3]. If the developer saves the information then data from data base is deleted else there is no change in the database at that particular moment but as soon as the data limit is reached an email is sent and then information from data base is deleted[4].

II. PROBLEM DEFINITION

In today's era, our focus is turning towards security of our data to maintain the integrity, authenticity and confidentiality of data. To do so, our main aim will be towards making use of IP address of client to retrieve information about him in order to check for his authenticity. Along with the security issues, we will be focussing on how to maintain data redundancy and data integrity in order to reduce complexity and complicacy.

III. SOLUTION

As mentioned above, our main aim is to get IP address of user. We will track the IP address of user by making use of PHP code and will store that address and information retrieved from IP like location, last visit, session time, website visited and all sort of activities in a database.

Any website who wants to secure its data using this technique will have to register itself with our website and with this there will be an entry in another database for that website. This database will contain information about the registered websites like their URL, name of owner, his email id, phone number. This will help us to group the IP addresses tracked based on registered urls.

We can also provide a complete data set to any registered user related to him as per his query. Also to maintain data redundancy or integrity, we will mail the corresponding data set to that user in certain period of time and will remove that much data set from database.

Describing each feature separately as:

A. IP Tracking:

Internet Protocol address management (IPAM) is a means of planning, tracking, and managing the Internet Protocol address space used in a network. Most commonly, tools such as DNS and DHCP are used in conjunction as integral functions of the IP address management function, and true IPAM glues these point services together so that each is aware of changes in the other (for instance DNS knowing of the IP address taken by a client via DHCP, and updating itself accordingly). Additional functionality, such as controlling reservations in DHCP as well as other data aggregation and reporting capability, is also common. IPAM tools are increasingly important as new IPv6 networks are deployed with larger address pools, different sub netting techniques, and more complex 128-bit hexadecimal numbers which are not as easily human-readable as IPv4 addresses. IPv6 networking, mobile computing, and multihoming require more dynamic address management and are causing technical obsolescence of the early generations of IP address management (IPAM) software and spreadsheets used for address management.

IP trace back is a name given to any method for reliably determining the origin of a packet on the Internet. Due to the trusting nature of the IP protocol, the source IP address of a packet is not authenticated. As a result, the source address in an IP packet can be falsified (IP address spoofing) allowing for Denial Of Service attacks (DoS) or one-way attacks (where the response from the victim host is so well known that return packets [5] need not be received to continue the attack. The problem of finding the source of a packet is called the IP trace back problem. IP Trace back is a critical ability for identifying sources of attacks and instituting [6] protection measures for the Internet. Most existing approaches to this problem have been tailored toward DoS attack detection. Such solutions require high numbers of packets to converge on the attack path(s).

B. Data redundancy:

Data redundancy occurs in database systems which have a field that is repeated in two or more tables. For instance, in case when customer data is duplicated and attached with each product bought then redundancy of data is a known [7] source of inconsistency, since customer might appear with different values for given attribute. Data redundancy leads to data anomalies and corruption and generally should be avoided by design. Database normalization prevents redundancy and makes the best possible usage of storage [8.] Proper use of foreign keys can minimize data redundancy and chance of destructive anomalies. However, concerns of efficiency and convenience can sometimes result in redundant data design despite the risk of corrupting the data [9].

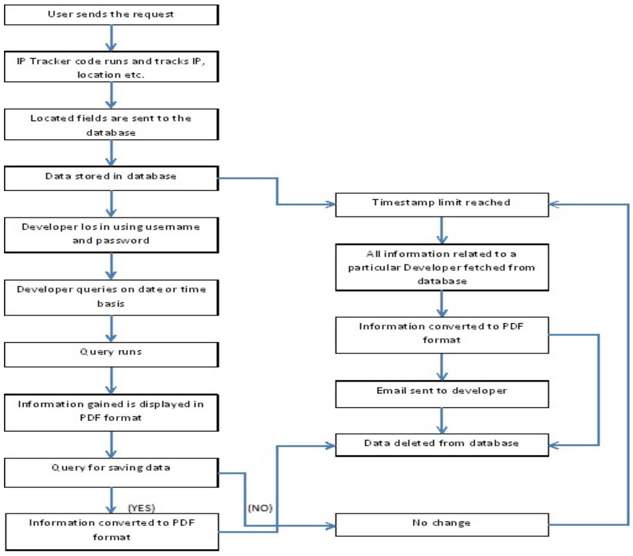
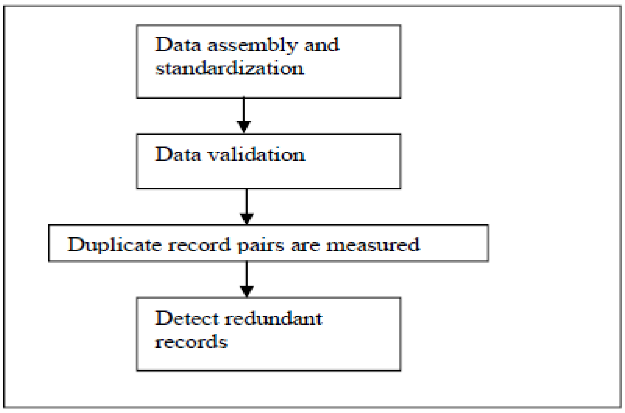

User sends the request
IP Tracker code runs and tracks IP, location etc.
Located fields are sent to the database
Data stored in database
Developer losin using username and password
Developer queries on date or time basis
Query runs
Information gained is displayed in PDF format
Query for saving data
(YES)
Information converted to PDF format
(NO)
Timestamp limit reached
All information related to a particular Developer fetched from database
Information converted to PDF format
Email sent to developer
Data deleted from database
No change

Figure: 1

C. Data Consistency:

Data consistency summarizes the validity, accuracy, usability and integrity of related data between applications and across an IT enterprise[10].This ensures that each user observes a consistent view of the data, including visible changes made by the user's own transactions and transactions of other users or processes[11]. Data Consistency problems may arise at any time but are frequently introduced during or following recovery situations when backup copies of the data are used in place of the original data.


Data assembly and standardization
Data validation
Duplicate record pairs are measured
Detect redundant records

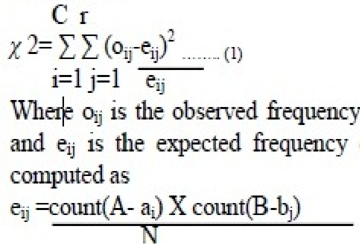
IV. ANALYSIS

This technique can be considered as the most user friendly approach because all the data which is required by the user is easily available to him at any time and if the developer forgets to check his data his/her data, it is sent to them via email. Thus his data is never lost [12].

Also if we focus towards complexity of tracking the IP address of the client, it is not very high. The IP address can be tracked within a maximum order of few nanoseconds.

Mathematical analysis of data redundancy is basically used for any two attributes, such analysis can measure how strongly one attribute implies the other, based on the available data. For the numerical attributes, we can evaluate the correlation between attributes A and B, by computing the correlation coefficient. This isWhere N is the number of tuples [13], ai and bi are the respective values of A and B in tuple i, Ā and □ are the respective mean values of A and B, where n is the number of tuples, A and B are respective mean values of A and B, and σA and σB Are the respective standard deviations of A and B. If the resulting value of equation is greater than 0, than A and B are positively correlated, σA and σB are the respective standard deviation of A and B, and Σ (AB) is the sum of the AB crossproduct. Note that $-1 \leq r_{A,B} \leq +1$ [14]. If rA,B is greater than 0, then A and B are positively correlated, meaning that the values of A increase as the values of B increase.

The higher the value, the stronger the correlation (i.e., the more each attribute implies the other attribute decrease. Note that correlation does not imply causality [15]. That is, if A and B are correlated, this does necessarily imply that A causes B or that B causes A. For example, in analyzing a demographic database, we may fine that attributes representing the number of hospitals and the number of car thefts in a region are correlated. Thisdoes not mean that one causes the other. Both are actually causallylinked to a third attribute, namely, population. For categorical (discrete) data, a correlation relationship between two attribute, A and B, can be discovered by χ 2 (chi-square) test. Suppose A has c distinct values, a1,a2,…ac. B has r distinct values, namely b1,b2,…br. The data tuples described A and B can be shown as a contingency table, with the c values of making up the column and the r values of B making up the rows, Let(Ai,Bj) denote the event that attribute A takes on value bj, that is, where (A=ai, B=bj). Each and every possible (Ai,Bj) joint event has its own cell in the table. The χ 2 value is computed as:



$$\chi 2= \sum_{i=1}^{C} \sum_{j=1}^{r} \frac{(o_{ij}-e_{ij})^2}{e_{ij}} \ldots\ldots (1)$$

Where o_{ij} is the observed frequency of the joint event (A_i,B_j) and e_{ij} is the expected frequency of (A_i,B_j), which can be computed as

$$e_{ij} = \frac{count(A- a_i) X count(B-b_j)}{N}$$

Where N is the number of data tuples, count (A= ai) is the number of tuples having value ai for A and count (B- bj) is the number of tuples having value bj for B. The sum in Equation is computed over all of the rXc cells. Note that the cells that contribute the most to the χ 2 value are those whose actual count is very different from that expected [16].

V. CONCLUSION

The IT Sector is the most growing and lucrative area in the world. As the amount of information increases, so do risk of misuse increases. Thus the security of the data is most important for the developer which can be achieved with help of our technique.

VI. REFERENCES

[1]. Nicolas Scandamis, Frantzis Sigalas & Sofoklis Stratakis, “The Case of Data Protection and the Criterion of Connexity” CEPS Challenge Programme, 2007, pp. 4-19.

[2]. Dorothy E. Denning and Peter J. Denning, ”Data Secutiry”, Computing Surveys, 2001, pp. 4-20.

[3]. Sharon Bolton & Matthew Woollard, “Strengthening Data Security: an Holistic Approach”, IASSIST, 2008, pp. 2.

[4]. Vishwa Gupta, Gajendra Singh and Ravindra Gupta, “Advanced Cryptography algorithm for improving data security”, International Journal of Advanced Research, 2012, pp. 2-4.

[5]. Ajit Singh and Upasna Jauhari, “Data Security by Preprocessing the Text with secret Hiding”, International Journal, 2012, pp. 3-10.

[6]. http://en.wikipedia.org/wiki/IP_traceback

[7]. http://en.wikipedia.org/wiki/Data_redundancy

[8]. Fiona Campbell, “Survivors of Redundancy”, Working Paper Series, 1999, pp. 5-13.

[9]. Chris Ding and Hanchuan Peng, “Minimum Redundancy Feature Selection from Microarray Gene Expression Data”, University Of California, 2005, pp. 4-7.

[10]. Kanat Tangwongsan, Himabindu Pucha, David G. Andersen, Michael Kaminsky, “Efficient Similarity Estimation for Systems Exploiting Data Redundancy”, Simest Infocom, 2010, pp. 2-7.

[11]. Dost Muhammad Khan, Nawaz Mohamudally, “Adaptability of Conventinal Data Minig Algorithms through Intelligent Mobile Agents in Modern Distributed Systems”, IJCSI, 2012, pp. 3-5.

[12]. Deyan Chen and Hong Zhao, “Data Security and Privacy Protection Issues in Cloud Computing”, International Conference on Computer Science and Electronics Engineering, 2012, pp. 2-3.

[13]. Vivek Waghamre, Dr. Ravindra Thool, “Issues Related to Security on Tree Structure Data”, IJCSI, 2012, pp. 4.

[14]. Sadaqat Ur Rehman, Muhammad Bilal, Basharat Ahmad, Khawaja Muhammad Yahya, Anees Ullah, Obaid Ur Rehman, “Different Cryptographic and Encryption Techniques Using Message Authentication Code in Wireless Sensor Networks”, IJCSI, 2012, pp. 3-5.

[15]. Majid Bakhtiari, Mohd Aizaini Maarof, “Weakness in RSA Cryptosystem”, IJCSI, 2012, pp. 2-3.

[16]. Mrs.C.Sumithiradevi, Dr.M.Punithavalli, “Detecting Redundancy in Biological Databases”, Global Journal, 2005, pp 2-4.