A Novel Approach using XMI Log Generation Cryptography and Mining Techniques for to Enhance Security of Data

Gurpreet Kaundal
Research Scholar,
Department of Computer Science and Engineering,
Lovely Professional University,
Phagwara, Punjab, India

Sheveta Vashisht
Assistant Professor,
Department of Computer Science and Engineering,
Lovely Professional University,
Phagwara, Punjab, India

ABSTRACT
Security is the major issue in IT sector because data of an organization is an important asset for them. The information can be there client or customer contacts or it may be the information about their accounts, in short that information is very confidential information that they never want to disclose. But because of security attacks an attacker i.e. Hackers can access this information. Securities are required for network data and the database data. The all information is stores in the databases and the log of this information is also generated by the server. In the log files all information of users is stored, so the security of the log files is required. The log contains the all information of the users such as IP address, Mac address, Session time, Operating System etc. The hacker can hack the log and steal all the users' information and edit it and upload on the server to avoid intrusion detection. So, in this research work proposes an idea for to prevent these types of attacks of the hacker. This work proposes the security of log files with the XMI file because the XMI file is a secured file. Retrieving the contents of XMI file is not easy because the XMI file contains the very big library of the java named XMI.Java. The log file data converting it into XMI file then encrypt the XMI file data again; it provides the more security of data. This makes the information more secure.

Keywords
XMI, Log Files, RAS, AES, Data mining

1. INTRODUCTION
Now days, Internet is a gold mine for those companies who realize the importance of the web mining because the Companies find a new and better way to do business. Companies want to know their own strength and weakness of their E-marketing effort on the web through continuous improvement and also check and understand their customer's profiles on the web [1]. E-commerce through internet i.e. the customers sales and buy their goods on the web through internet. There are many companies' websites on the web and the users can easily access the information of the company's. E- Commerce companies such ase-bay, Home 18, Jabong, OLX etc. These companies provide the customers online shopping through internet. At the home users can place the order online, within two or three day's user can get the thing which is ordered and for payment the users having number of options such as credit card payment, debit card and cash on delivery, these payment modes are depends on the users or customers. Online shopping is the time saving because user can place the order at home. But some disadvantages of is that the speed of the internet connection and on the network the attacker can attack on the data. The hacker can hack the account of the user or attack on the user account on the network. The hacker can also access the databases [2], in the databases all information of the users are stored and the server also generated the log files. In the log files all information of the users is stored i.e. IP address, Session etc. The information is easily hacked by hacker when security techniques are not applied on the log data [3]. To encrypt the log files then the hacker cannot easily access the information from the database and the log. In this paper, uses the XMI files for security of log files because the XMI is secured, the hacker cannot access the XMI code easily. The log file data again encrypt it its makes the data more secure [4]. The attacker cannot access the data of the log file data because it is encrypted two times and provides the double data level security of the log file data. This work proposes the security of log data files with the XMI file and the encryption algorithm. Retrieving the contents of XMI file is not easy because it required the very big library of java named XMI. Java and the decryption the encrypted data is not easy task because it requires the secret key. So, provides the double security of log file data. The main aim of generating an XMI file is to secure the log created by server. When the data is available publicly, then data needs security. Data mining techniques are used to find the patterns form the large amount of data [5]. In business, security of data is very important because data or information is very useful for business growth. So, data mining techniques are used to find the behavior of the user and then recognize the user's behavior according to it. The users are request to the server and the server gives the response back to the user, the server generates the log files of the users. All information of the users is stores in the log. So, it needs to provide the security of log data files. The web usage mining is used to find the common behavior of the users and then recognize the users according to it. The web usage mining includes the collection of data, preprocessing of data and then discover the patterns. By using the web usage mining to find the users behavior and then detects the fraud in the business. For business confidentiality and availability is very important, so security of data is needed for the growth of the business. In the network the data is transfers from one place to another place. Then the attacker can attack on the data and steal the information and misuse the information it affects the business. So, the encryption techniques are used to encrypt the data and then transfer from the network. If the attacker attacks on the data then attacker gets the only encrypted data. So, the data is secure. The web usage mining used for finding the users behavior; in business it is very necessary to find the behavior of the user because the user can be an attacker. So, find the behavior of the users on the bases of many parameters such as

users click on the websites link, according to it to find it is valid user or not. The data are transfer from one place to another through network, to transfer the encrypted data. So, the problem is formulated from these ideas, the proposed work is that to secure the log file data with XMI and use the encrypted algorithm for more security of data. The attacker cannot access the data from the log because log files are converted into XMI and then encrypt the data. This makes log file data more secure.

2. RELATED WORK

Singh A. *et al* (2013) described the three phases such as preprocessing, pattern discovery, and pattern analysis. These three phases are describing in details and also describe its current applications of the technique [6]. On the bases of data usage the web usage mining is classified in further such as web server data, application level data, application server data and the web server data. Basically web usage mining is used for loges, and also discovery the useful patterns which are generated by the web servers (client-server transaction). In which generated the data stored automatically in server such as access logs, agent logs and referrer logs, and also discussed the some approaches like data collection, knowledge discovery, data preprocessing and pattern analysis. So, in which analyzing the data and apply the knowledge to the users i.e. serve the commercial offerings.

Shaily G. *et al* (2013) described the security of the system; the main aim of the author is that to finding the common behavior of the visitor from the web log file of website. The web Usage mining involves the three phases such as pattern discovery, preprocessing and pattern analysis. The author used the DBSCAN clustering algorithm for finding the visitor's common behavior. In which the clustering quality is depends on the both similarity measure, and the commonly distance is measure the Euclidean, Manhattan and the Minkowski distance. In DBSCAN algorithm, it requires the two parameters such as eps and minpts to form the cluster require the minimum number of points [7]. The web usage mining with the DBSCAN cluster algorithm which having the used for finding the common properties and behavior or interest efficiently from the web log file. This method is used either personalization or business intelligence for find the group of users common properties.

Hoisl B. *et al* (2012) discussed the secure object flow that is passed in between the different participants in the service oriented architecture. For security of the object flows used the unified modeling language (UML), in the business process the data or objects are flows in between the different participants i.e. security of the data or object flow are more important. In which also describe the description of how to map the independent model with the specific software and also give the description howto integrate this approach with the eclipse modeling tools [8]. In which also discussed the process driven systems with confidentiality and integrity of the object flow, object flow is that to passed the object from one node to the another in the business process. So, for secure object flow in the business process of SOA is that to integrated the layers such as computation, platform models; the computation model provide the meta models for objects and the platform model provides the UML extension for secure model object flow. In future work, needs to extend the activities of the business framework which is providing the secure business process in distributed systems.

Sun W. *et al* (2010) discussed transformation engine which is based on the model driven architecture framework and also supports the transformation of the automatic tool from the RSA to USE [9]. In which uses the modeling tool and create the model with constraints (OCL) Object Constraint Language, and the modeling toolis IBM (Rational Software Architect (RSA), for the model validation uses the USE tool. The RSA supports the class diagrams for the UML Meta model and the activity diagram and the USE supports the only class diagram. The author proposed the transformation engine based on the MDS transformation and also called XMI2USE, and presents the USE metamodel with the mapping between the RSA and USE.

3. THE PROPOSED WORK

The information is stores in the database and its log is created by server, in the log file all information of user are stored. So, the security of log file is required, in the log files the noisy data is also present in the file.The hacker can easily access the log file and steal the all information of the users andmisuse the information of the user. Thecaneasily steal the log data file, change the fileinformation and upload to the server to avoid the intrusion detection. So, the security of log file is needed. In the log file, the data or information is stored in the text files for security of log files to convert the log files into other format or to encrypt the information of the log file because the hacker cannot watch and edit the information. This work proposes the security of log files with the XMI file because the XMI file is a secured file and retrieving the contents of XMI file is not easy. For converting the data of XMI to normal text requires a heuristic approach and algorithm and also required the very big library of java named XMI.Java. For Mining the XMI content and knowledge discovery uses the Fuzzy Soft K means clustering along with Decision tree. The main aim of generating an XMI file is to secure the log created by server. After that uses the encryption techniques for more security of data. In case security is breached by hacker can steal the log file, edit it and upload back to server to avoid the intrusion detection. But in this approach the hacker cannot access the XMI file content as decrypting the XMI content is far much difficult task and also difficult to decrypt the data which is encrypted because hacker do not have the key for decryption. So, provides the double data level security of the log file data. The log files are generated by the server, in the log files the noisy data are also generated with log files. The log files are generated in the form of XMI and also needs to remove the noisy data from the XMI file. With this approach provides the security to the log file information which is generated by the server. This makes data more secure.The main aim of generating an XMI file is to secure the log created by server. After that uses the encryption techniques for more security of data. This work provides the double data level security with in the XMI file.

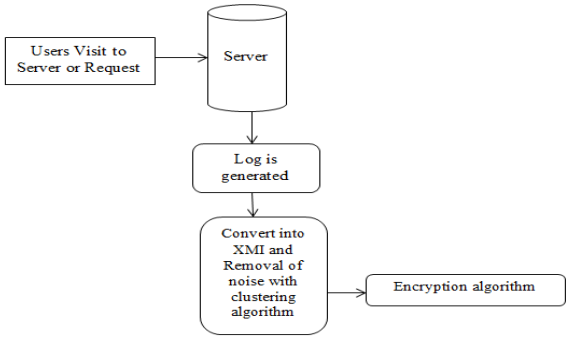

Figure 1: Overall Research Design

Here in this Figure1 shown that users are request to the server and the server response to the users. The user's data are stored in the database; log of the data is also generated by the server. The log contains the all information of the users such as IP address, Mac address, Session time, Operating System etc. with this some noisy data is also generated. The hacker can also hack the log and steal all the users' information and edit it and upload on the server to avoid intrusion detection. So, in this research work proposes an idea for to prevent these types of attacks of the hacker. This work proposes the security of log files with the XMI file because the XMI file is a secured file. Retrieving the contents of XMI file is not easy because the XMI file contains the very big library of the java named XMI.Java. Firstly, Log is generated in the form of XMI which is highly secured XMI file. The hacker cannot access this file because the decryption of the content of this file is far much difficult task. To provide the security of the log with the XMI and also remove the noisy data from the file with the help of Fuzzy Soft K mean algorithm. After that the encryption technique are applied on the XMI file data for more security and then data mining techniques are applied on the encrypted data but the data not mined because the data is encrypted. So, no one can mine the encrypted data with the data mining techniques. Its means data is more secure and the attacker cannot get the knowledge from the encrypted data.

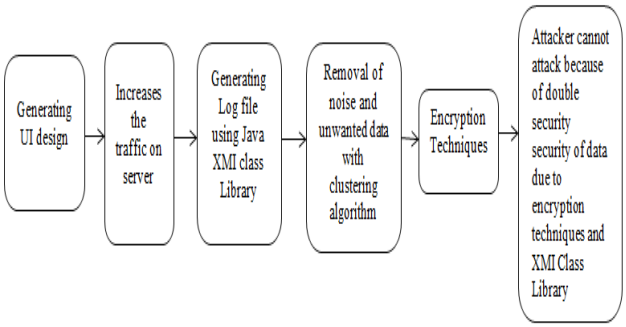

Figure 2: Work Flow Design

The data mining algorithms and techniques are uses for mining the information. Firstly, generate the log file in the XMI format and then encrypt the data. It makes the data more secure. The attacker can access the data, if the attacker can attack on it or hacker can hack the data but hacker or attacker cannot decrypt the data because hacker do not have the key, if the hacker decrypt the data by using other techniques or method then hacker cannot get the original message or data because the data is in XMI form and the decryption of XMI file is not easy task because it require big library of java. So, the data is still secure. This work provides the more security of data or also provides the double data level security. The main aim of this approach is that to provide the security of data. So, firstly log files are generated by the server then convert the log files into the XMI form with the help of XMI class library. It makes the data secure. After that applied the encryption algorithm on it for more security of data because the data mining techniques are applied on the encrypted data, there is not effect on it. The encrypted data cannot mine by using the data mining technique, so attacker cannot access the data from the log. This approach provides the two ways, first is that to convert the log file data into the XMI file, some unwanted and noisy data are also generated with it. So, needs to remove the unwanted and the noisy data from the XMI file and then also encrypt the data for more security. The attacker cannot attack on the data because the data is double encrypted. Then applied the data mining techniques on the encrypted data, there is no effect on the encrypted data. So, the data of the log files are more secure. The attacker cannot attack on the data, if the attacker can get the data from the network then attacker needs to decrypt it. If the attacker can decrypt the data and then attacker get the XMI file. The XMI file is more secure because XMI file require the very big library of java named XMI.java. The admin can do this because only admin having the key and the Java library of the XMI. The admin can decrypt the data with key, and then the data is still secure because of XMI file. The XMI file is not easy to decrypt it because it requires the very big library of the java named XMI.Java. So, the admin first decrypt the data with the help of key and then convert the XMI file into the simple text file with the help of the very large library of the java named XMI.Java and maintain all the information of the log of the users as shown in below Figure 3.

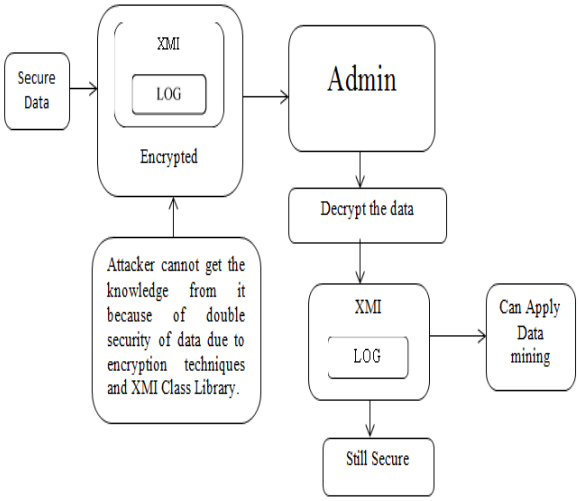

Figure 3: Double Security to Data

The main aim of this approach is that to provide the security of data. So, firstly log files are generated by the server then convert the log files into the XMI form with the help of XMI class library. It makes the data secure. After that applied the encryption algorithm on it for more security of data because the data mining techniques are applied on the encrypted data, there is not effect on it. The encrypted data cannot mine by using the data mining technique, so attacker cannot access the data from the log. This approach provides the two ways, first is that to convert the log file data into the XMI file, some unwanted and noisy data are also generated with it. So, needs to remove the unwanted and the noisy data from the XMI file and then also encrypt the data for more security. The attacker cannot attack on the data because the data is double encrypted. Then applied the data mining techniques on the encrypted data, there is no effect on the encrypted data. So, the data of the log files are more secure. The attacker cannot attack on the data, if the attacker can get the data from the network then attacker needs to decrypt it. If the attacker can decrypt the data and then attacker get the XMI file. The XMI file is more secure because XMI file require the very big library of java named XMI.java. So, the admin first decrypt the data with the help of key and then convert the XMI file into the simple text file with the help of the very large library of the java named XMI.Java and maintain all the information of the log of the users. But, the attacker cannot access the data because the message or data are encrypting double time. So, we are provides the double security of the data.

4. RESULTS AND DISCUSSIONS

4.1 AES (Advanced Encryption Standard)

AES algorithm is block cipher, the size of the block is uses in AES are 128 bits and the size of the keys is 18 bits, 192 bits and 256 bits [10]. The block size of the plaintext is 128 bits or 16 bytes and the key size can be 16, 24 or 32 bytes i.e. 128, 192 and 256 bits, the rounds depends on the key size (Stallings, 2011). The overall steps for AES are:

i. First the AES is not a feistel structure, the classic feistel structure is that the half of the block data is used to modify and other half block data is used to swap. In AES the entire block data is used as a single matrix during each round using substitutions and permutation.

ii. The key provided as input i.e. expanded into an array of w[i].

iii. In AES the different stages are used such as one is permutation and three for substitution, the four stages are:

4.1.1 Substitute bytes

For substitute uses the S-box and perform byte by byte substitution of the block.

4.1.2 Shift Rows

For shift rows uses the permutation.

4.1.3 Mix Columns

For mix columns use the arithmetic value of the matrix, in which two matrixes are multiple, with each other.

4.1.4 Add Round Key

In which simple perform the XOR operation and it uses the bitwise XOR operation of the current block.

i. The rounds are depends on the size of the key, firstly to add the round key withXOR operation then the first round starts.

ii. Substitute bytes by using the S-box, to perform the byte by byte substitution of the block.

iii. Then shift the rows, it is also depends on the rounds if the round is first then shift rows values one by one. The shift rows use the permutation.

iv. In Mix Columns, uses the arithmetic values of the matrix, multiple the two matrixes one matrix is existing and other are comes after performing the shift operation.

v. After that again add the round key, and the second step begins. So, the rounds are depends on the size of the key. If the key size is 128 bits or 16 bytes then the rounds are 10.

These steps are followed for encryption and for decryption the steps are same but inverse are used i.e. The inverse shift row, inverse sub bytes and inverse mix cols, and the rounds, steps are same.

4.2 RSA (Rivest-Shamir-Adleman)

RSA algorithm is used for to encrypt and decrypt the data; it is a block cipher scheme [11].In which the plaintext and ciphertext are encrypted and decrypted in blocks and the size of the block having value less than some number n, the size of the n is 1024 bits and 309 decimal (Stallings, 2011).Suppose the size of the block is i bits where $2^i < n <= 2^{i+1}$ and the plaintext is M block and Ciphertext is C block i.e.

$$C= M^e \bmod n$$

$$M= C^d \bmod n= (M^e)^d \bmod n= M^{ed} \bmod n$$

The sender and receiver must know the value of n and the sender knows the value of e, the receiver knows the value of d [12]. The public key is used for encryption i.e. PU= [e, n] and the private key for decryption i.e. PR= [d, n]. Some following steps are included in this algorithm for public key encryption, the steps are:

i. It is require to find the values of e, d, n such that the $M^{ed} \bmod n= M$ for all $M < n$.

ii. To calculate the $M^e \bmod n$ and $C^d \bmod n$ for all values of $M < n$.

iii. To determine the values of d given e and n, it is infeasible.

Steps to generate the key are:

i. First select the two numbers such as p, q and p, q numbers both are prime, p=! q.

ii. To calculate the value of n i.e. n= p*q.

iii. To calculate the value of Euler totient function i.e. FI (n) = (p-1) (q-1).

iv. Now, select any integer number e i.e. gcd (FI (n)) =1; 1 < e < FI (n).

v. To calculate the value of d i.e. $d= e^{-1} \pmod{FI (n)}$.

vi. For Encryption uses the public key i.e. PU= [e, n] and for Decryption uses the private key i.e. PR= [d, n].

4.3 Steps of Encryption

i. We first calculate the secret key using rsa method to find E which we will use for swapping the value in matrix

ii. The matrix have multiple integer number from where we select a Secret key which we pass in encryption method for convert the plaintext to cipher text

iii. How we get the secret key lets an example we use 5 as an E then we take all number in matrix form then find out 5th position of row and column then extract the number from that position.

iv. Then we use that number apply this steps till of 14 time to calculate the number which we will use as a secret key.

v. Then we take string convert all string into character array then take every character then convert into ascii further convert into binary format divide by 2 after that we use secret key convert into binary format then we apply xor gate to then,

vi. After calculate the xor data convert that data in decimal form then convert into string which is called ciphertext.

Decryption is reverse from 6, 5, 4 steps

For implantation we are used the NetBeans and WAMP Server. Firstly start the NetBeans and then start the wamp server. In NetBeans start the server, Right click on theServer1

and the users send the request to the server. The log of the every user is generated by the server. In Figure 4.1 shows the server starts and waiting for the users request.

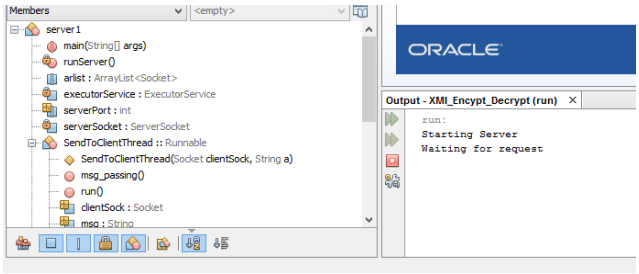

Figure 4: Start Server

Figure 5: Login Page

*Step 1:*The new user sign up enter the details and click on the register button then the request sends to the server. At the same time many users request to the server but user cannot sign up with same name. After click on the register button the request of the user go to the server as shown below:

Figure6: Sign UP and Registration

*Step2:*The users all information is present in the log, for security of the log the admin can convert the file into xmi and then encrypted it because of double security of data and only admin can decrypt the file and manage the details of the users. The attacker cannot access the information because the data is encrypted double times. Enter the details of the admin and then select the options according to it.

Figure 7: Administrator Login

The Admin enter the IP address, name and password and click on the Login button then the request again sends to the server as shown below:

After enter the name of the admin and then the other dialog box open as shown below if the admin wants to create the xmi file then click on the Click to Create XMI file of the Log Data and create the xmi file of the user's log. First save the xmi file with extension and it convert it into the xmi format as shown below:

Figure 8: Administrator Tasks

Admin wants to convert the simple file into the xmi file then click on the create xmi file of Log data and the save the file as shown below:

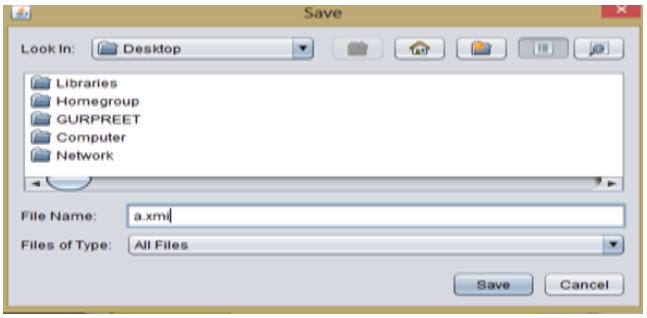

Figure 9: Save XMI File

This is the XMI file of the users log data:

Figure 10: XMI File

The 18 seconds time is consume by the system when the xmi file is created and the size of the xmi file is 2570 bytes and it takes the time 18 seconds.

Figure 11: Times Taken to XMI File

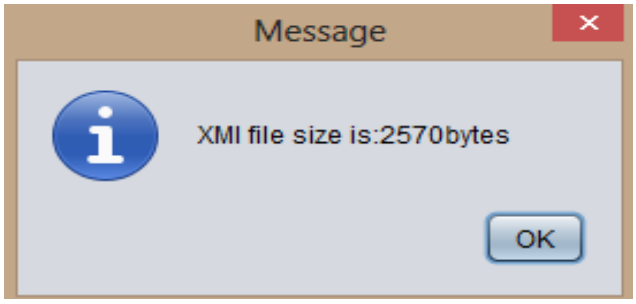

Figure12: Size of XMI File

*Step 3:*Admin can also encrypt the xmi file again with the encryption technique. So, select the xmi file encryption because of security. Click on the encryption button then select the xmi file and then save it. The saved file is encrypted, so provide the double security of data. The hacker cannot access the file and not steal the information of the users as shown below:

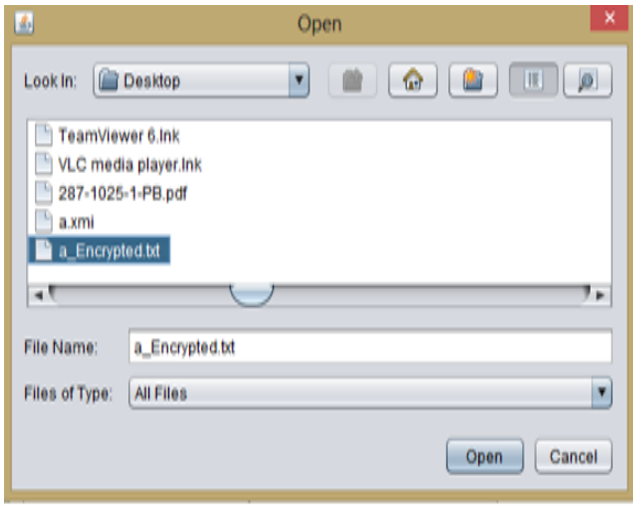

Figure 13: Save Encrypted File

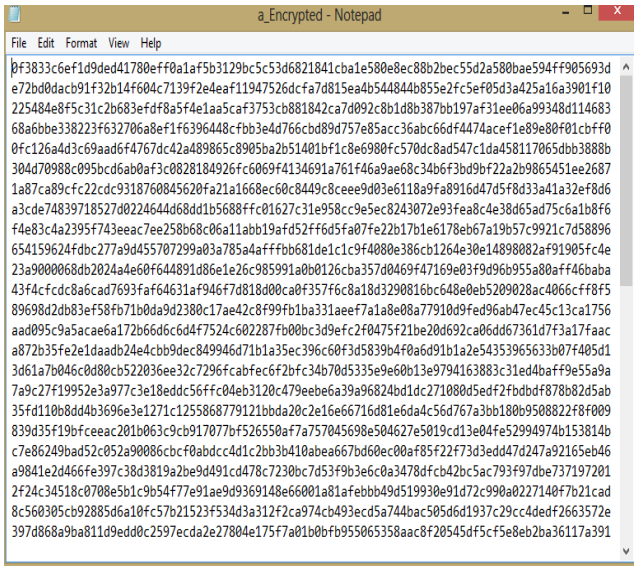

Figure 14: Encrypted File

Encrypt the xmi file again, Then the time is consumed by the system is that 26 Seconds as shown:

Figure 15: Times Taken to Encrypted File

*Step 4:*The only admin can decrypt the log file data because admin have the big java library of the xmi and also have the key. For the decryption, the admin click on the decryption button and select the encrypted file and then save it. The saved file is decrypted as Figure 4.15.

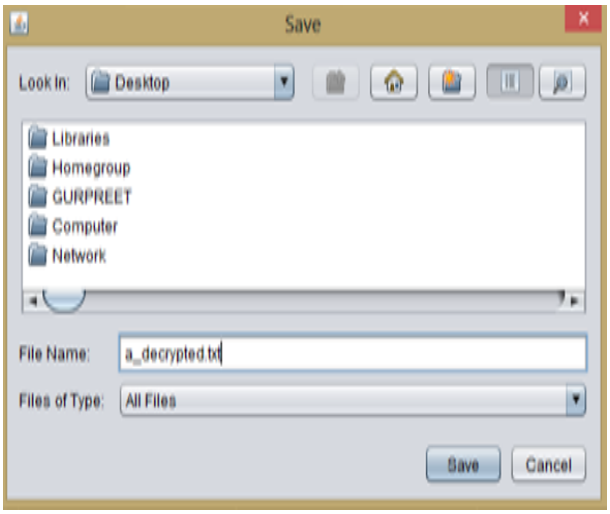

Figure 16: Save Decrypt File

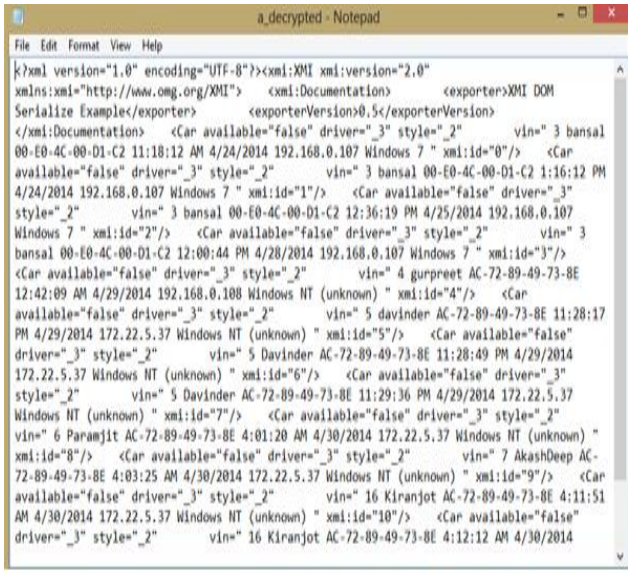

Figure17: Decrypted File

Now, the admin decrypt the encrypted file then the time is taken to decrypt the encrypted file is 21 seconds as shown Figure 4.17.

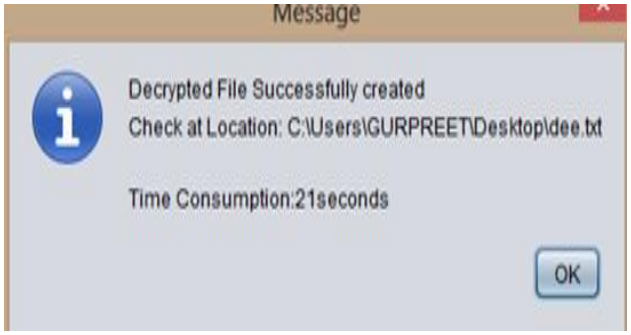

Figure 18: Times Taken to Decrypted File

The total Compilation time of the application is 237 seconds, this time is calculates from the starting point of the application when the admin create the xmi file to the admin decrypt the file and save it. Admin click on the compilation time button then the time is calculates from the starting point of the application i.e. time taken by the system when the process is executes and calculates the total compilation time before the process is completed as shown:

Figure 19: Total Time

*Step5:*The authorized user login and the log are generated by the server. Admin canconvert the log into the xmi form for security and then again encrypt the log data file. Admin also decrypt the log data file and maintain the user's information but the authorized user can not mine the log data or can't decrypt it. The user login and then mine the encrypted data but the data is not mined because data is encrypted as shown below:

Figure20: Mining

Now, Select the encrypted file for mining, the abc_encrypt.txt file is encrypt file. The user click on the button for mine the encrypted data but it is not be mined because the data is encrypted and then save the file. After that check the file the data is not mined or decrypt as shown below:

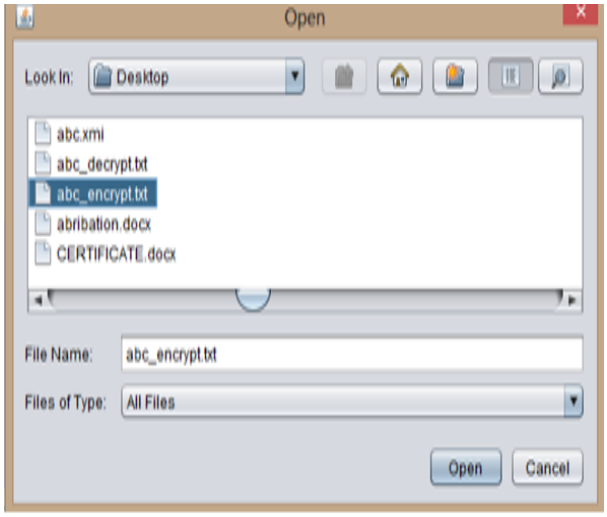

Figure 21: Select the File Encrypted File for Mining

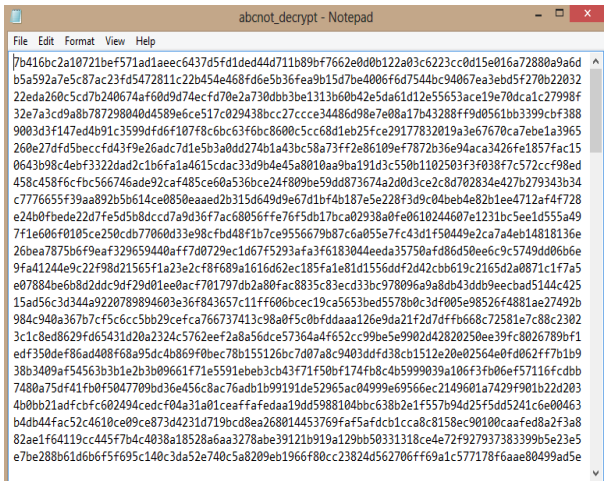

Figure 22: Not Mine the Encrypted File

5. CONCLUSION

In this Paper, discuss the security of the data, now a day's companies find the new and better way to do business. For business data is very important part because the data is used for take decisions and increase the growth of the business. E-Commerce is that the users buy and sale the goods through internet. The users can access the websites through internet, and the data is stored into the database, its log is also generated by the server. In the log all information of users are stored such as IP address, Session Time, MAC address etc. This work proposes the security of log files with the XMI file because the XMI file is a secured file and retrieving the contents of XMI file is not easy. For converting the data of XMI to normal text requires a heuristic approach and algorithm and also required the very big library of java named XMI.Java. For Mining the XMI content and knowledge discovery uses the Fuzzy Soft K means clustering along with Decision tree and also uses the encryption techniques AES and RAS for the security of the data.So, we are used the XMI file and encryption algorithm for security of log data file. It provides the double security of log file data. In future to improve the encrypted techniques and provides the more security of data. In this paper, security of data is enhanced on the data level, but for future work both file level and data level security can be provided to the log files or any other data file.

6. REFERENCES

[1] Wuliang Sun, Paul Grabow and Devon M. Simmonds (2013) "Web Mining: An Introduction",International Journal of Advanced Research in Computer Science and Software Engineering, pp. 683-687.

[2] Tomar Archana, Kumar Dubey Ashutosh and Richhariya Vineet (2011) "Novel Sensitive Information Preserving Mining (SIPM) Algorithm for Association Rule Mining in Centralized Database", IEEE, pp. 392-397.

[3] V.K.Deepa and J. Rexy R. Geetha (2013) "Rapid Development of Applications in Data Mining" IEEE-Proc. of International Conference on Green High Performance Computing.

[4] D. Karthikeswarant, V.M. Suresh and A. Javed Sultan (2012) "A Pattern Based Framework For Privacy Preservation Through Association Rule Mining", IEEE-International Conference On Advances In Engineering, Science And Management (ICAESM) , pp.816-821.

[5] Han, J., and Kamber, M. (2006*)* "Data Mining: Concepts and Techniques*",* Morgan Kaufmann, India, p. 6-10, 33-34.

[6] Arun Singh and Dheeraj Sharma (2013) **"**Web Usage Mining: Discovery of Mined Data Patterns and their Applications",International Journal of Computer Science and Management Research. 2(5), pp. 2423-2429.

[7] Shaily G. Langhnoja, Mehul P. Barot and Darshak B. Mehta (2013) "Web Usage Mining to Discover Visitor Group with Common Behavior Using DBSCAN Clustering Algorithm", International Journal of Engineering and Innovative Technology (IJEIT). 2(7), pp. 169-173.

[8] Bernhard Hoisl, Stefan Sobernig and Mark Strembeck (2012)"Modeling and enforcing secure object flows in process-driven SOAs: an integrated model-driven approach", Springer-Verlag, pp.816-821.

[9] Wuliang Sun, Eunjee Song, Paul Grabow and Devon M. Simmonds (2010) "XMI2USE: A Tool for Transforming XMI to USE Specifications",Springer, pp. 147-156.

[10] Kaur Gurtaptish, Malhotra Sheenam (2013) "A Hybrid Approach for Data Hiding using Cryptography Schemes", International Journal of Computer Trends and Technology (IJCTT).4(8), pp. 2917-2923.

[11] Stallings, W. (2011). Cryptography and Network Security. India: Dorling Kindersley Pvt. Ltd.

[12] D. Manivannan and R. Sujarani (2010) "Light Weight and Secure Database Encryption Using TSFS Algorithm", International Conference on Computing Communication and Networking Technologies (ICCCNT), pp. 1-7.