DESIGN & IMPLEMENTATION OF A SECURE BANKING MODEL USING FUZZY LOGIC AND CRYPTOGRAPHY

[a*]Meenakshi Bansal, [b]Dr. Dinesh Grover, [c]Dr. Dhiraj Sharma

[a]Assistant Professor, Computer Engineering Department, YCOE, Talwandi Sabo, Punjab, India
[b]Principal, Bajaj Engineering College, Gureh, Ludhiana
[c]Business School of Management Studies, Punjabi University, Patiala
**Corresponding Author Email: ermeenu10@gmail.com*

ABSTRACT

Day by day Privacy-preserving data mining using cryptography models become more important because of the security purpose to store personal data about users. Cryptography is used in secure transmission of data when we use online transmission process. Cryptography is the practice and study of such techniques for secure communication. Cryptography is used in security models of ATM, password protection in mobile phones and various online models for authentication process.

Keywords*—Data Mining; Security system; fuzzy logic; cryptography.*

1. Introduction

Due to the advances in information processing technology and the storage capacity, modern organisations collect a huge amount of data. For extracting hidden and previously unknown information from such huge data sets, the organisations rely on various data mining techniques. During the whole process of data mining these data often get exposed to several parties. If such a party has enough supplementary knowledge about an individual having a record in the data set, then the party can re-identify the record. This sensitive information stored about the individual can potentially be disclosed resulting in a breach of individual privacy. Therefore, we need techniques for protecting individual privacy while allowing data mining. Privacy preserving data mining is a novel research direction in data mining and statistical databases, where data mining algorithms are analyzed for the side-effects they incur in data privacy. The main consideration in privacy preserving data mining is twofold. First, sensitive raw data like identifiers, names, addresses and the like should be modified or trimmed out from the original database, in order for the recipient of the data not to be able to compromise another person's privacy. Second, sensitive knowledge which can be mined from a database by using data mining algorithms should also be excluded, because such knowledge can equally well compromise data privacy. In the era of information technology most banking related services are completed through internet. So internet banking came into role. It helps the customer to access various banking services like transferring money to others account, checking current balance along with checking the status of other financial transactions, anywhere anytime at their convenience. But the main challenge regarding internet banking is privacy or security. As data communicated on social network is always threat to be attacked by unauthorized persons**.** So it is necessary to build a secure banking architecture which ensures the privacy and integrity of the transactions along with maintaining the utility of the data.

During the whole process of data mining data often get exposed to several parties. If these parties get enough auxiliary knowledge about an individual having a record in the data set, then the party can re-identify the record. The main motive of privacy preserving is to secure the data from unauthorized or third party access. Privacy preserving data mining is to develop efficient frameworks and algorithms that can extract significant knowledge from a large amount of data without disclosure of any sensitive information [5] [13][14][15]. The breach between the data mining and data confidentiality is crammed up by the privacy preserving data mining [12]. This sensitive information stored about the individual can potentially be disclosed resulting in a breach of individual privacy. Therefore, we need some techniques for protecting individual privacy while allowing data mining.

The main objective in privacy preserving data mining is to develop algorithms for modifying the original data in some way, so that the private data and private knowledge remain private even after the mining process. The problem that arises when confidential information can be derived from released data by unauthorized users is also commonly called the "database inference" problem.

2. Important Terms

In our research, we are focusing on the vast matter of security and integrity of data. We will achieve this goal by implementing the combination of fuzzy logic and cryptography algorithms. Key terms used in our process

1) Data Pre-processing:- This component aims to collect, select, clean and transform data into the necessary form for association rule mining process [8].
2) Fuzzy Correlation Technique: - This technique helps to know the strength of relationship between two variables or fuzzy attributes.
3) Association rule mining: This component aims to discover association rules that satisfy given threshold for business process of the company and then find sensitive association rules needed to be hidden to protect their sensitive information. This component needs intervention of experts to decide which association rules are sensitive [1].
4) Fuzzy rule Pruning: - It is the method to remove the rules that are not needed. **Sensitive association rule hiding**: It aims to hide sensitive association rules from the rules that are mined in fuzzy rule pruning before sharing data to partners [16].
5) Pattern Matching: - Extracting previously unknown patterns from huge volume of data is the primary objective of any data mining algorithm. A technique in automated data analysis usually performed on a computer, by which a group of characteristic of an unknown object is compared with the comparable groups of characteristics of a set of known objects, to discover the identity or proper classification of the unknown object [17].
6) Cryptography: - "Authentication + Encryption + Certification Authority= Trust" [18]. A number of cryptographic techniques such as DES and RSA have emerged as efficient methods for secured data communication. However, they have their respective drawbacks. RSA requires very high computational cost, especially when the data is large. Thus, DES, an efficient algorithm, has been considered as an alternative to encrypt various attribute values with multiple keys of variable sizes. The main focus on the role of the cryptographic techniques for privacy preservation. We are tried to put some best cryptographic algorithm for secure sharing of the user interaction with the server [8] [15].

3. Related Work

A.Q. Ansari et.al (2007) described in his studies the impact of data mining and fuzzy on cyber security. Fuzzy Logic provides techniques for handling cognitive issues in the real world. Fuzzy logic provides membership function to fuzzify the data. With the introduction of e-commerce and e- governance applications as well as activity boom in cyber cafes, the pressure is on cyber security monitoring. Existing data mining solutions are not directly adaptable to support E-Discovery legal compliance process. The scope of fuzzy logic is to circumvent some problems in the cyber crime domain. The combination of Data Mining and Fuzzy Logic based techniques for dealing with Cyber Security issues in the present era by introducing cyber crime prevention practices.

Youwen Zhu et.al (2009) extended Privacy-preserving Add to Multiply Protocol to Privacy-preserving Adding to Scalar Product Protocol, which is more secure and more useful in the high security situations of Privacy-preserving Data Mining. In addition, this paper also explains the extended security PPAtMP to PPAtSPP and is more powerful in high security situations, and proposes a solution for the new privacy-preserving protocol. We can also enhance the practical development of SMC protocol to remove the problems in PPDM.

Jian Wang Yongcheng et.al (2009) discussed the problem of sharing sensitive data during communication because of attackers and hackers. In this authors propose to use the Randomized Response techniques to solve the DTPD problem. In randomized response is to scramble the data in such a way that the central place cannot tell with probabilities better than a pre-defined threshold whether the data from a customer contain truthful information or false information.

Maher Aburrous et.al (2009) had opinioned about how Fuzzy Data Mining (DM) Techniques can be an effective tool in assessing and identifying phishing websites for e-banking. In this e-banking phishing website model showed the significance importance of the phishing website two criteria's (URL & Domain Identity) and (Security & Encryption) in the final phishing detection rate result, taking into consideration its characteristic association and relationship with each others as showed from the fuzzy data mining classification and association rule algorithms. This phishing model also showed the insignificant trivial influence of the (Page Style & Content) criteria along with (Social Human Factor) criteria in the phishing detection final rate result. A major issue in using data mining algorithms is the preparation of the feature sets to be used.

Alaa Aref El Masri et.al (2009) provides in his studies, a solution that limits private data exposure to entities that already have it. This paper presents a new model for privacy assurance that limits private data sharing to those who already have it and avoids new unneeded exposure. The PrOPrP protocol achieves this goal by only allowing the user's bank to access the user's private data, which does not constitute new exposure since the bank already holds this information.

Luong The Dung et.al (2010) argues that the participant's privacy without loss of accuracy can be ensured by the privacy-preserving computation of frequencies of a tuple of values. This study proposed a method for privacy-preserving classification learning in two-dimension distributed data, which has not been investigated previously. Basically, this method is based on the ElGamal encryption scheme and it ensures strong privacy without loss of accuracy. The applicability of the method by applying it to design the privacy preserving protocol for some learning methods such as association rules mining and decision tree learning.

C. Ronchi et.al (2011) analyzes the existing deployments of Internet banking services from the perspective of the End User, whose main goal is completing the online transaction. The sole use on the client side of so-called "trusted" hardware devices will be discussed and shown to fall short of the requirements for truly secure Internet banking. A new metric for gauging the effectiveness of security software will be described and applied to measure the practical security of existing Internet banking systems. Finally, a number of guidelines will be provided for assuring that reasonable care is exercised in the design and deployment of Internet banking systems. In this paper, somewhat naïvely attempted to spur the development of a higher deontology for Internet banking.

4. Fuzzy

Classification and prediction are two forms of data analysis that can be used to extract models describing important data classes or to predict future data trends. While classification predicts categorical labels (classes), prediction models continuous-valued functions.

Preprocessing of the data in preparation for classification and prediction can involve data cleaning to reduce noise or handle missing values, relevance analysis to remove irrelevant or redundant attributes, and data transformation, such as generalizing the data to higher-level concepts or normalizing the data.

Predictive accuracy, computational speed, robustness, scalability, and interpretability are five criteria for the evaluation of classification and prediction methods. ID3, C4.5, and CART are greedy algorithms for the induction of decision trees. Each algorithm uses an attribute selection measure to select the attribute tested for each nonleaf node in the tree. Pruning algorithms attempt to improve accuracy by removing tree branches reflecting noise in the data. Early decision tree algorithms typically assume that the data are memory resident—a limitation to data mining on large databases. Several scalable algorithms, such as SLIQ, SPRINT, and RainForest, have been proposed to address this issue.

Naive Bayesian classification and Bayesian belief networks are based on Bayes, theorem of posterior probability. Unlike naïve Bayesian classification (which assumes class conditional independence), Bayesian belief networks allow class conditional independencies to be defined between subsets of variables.

A rule-based classifier uses a set of IF-THEN rules for classification. Rules can be extracted from a decision tree. Rules may also be generated directly from training data using sequential covering algorithms and associative classification algorithms. Back propagation is a neural network algorithm for classification that employs a method of gradient descent. It searches for a set of weights that can model the data so as to minimize the mean squared distance between the network's class prediction and the actual class label of data tuples. Rules may be extracted from trained neural networks in order to help improve the interpretability of the learned network.

A **Support Vector Machine (SVM)** is an algorithm for the classification of both linear and nonlinear data. It transforms the original data in a higher dimension, from where it can find a hyperplane for separation of the data using essential training tuples called support vectors.

Associative classification uses association mining techniques that search for frequently occurring patterns in large databases. The patterns may generate rules, which can be analyzed for use in classification.

Decision tree classifiers, Bayesian classifiers, classification by back propagation, support vector machines, and classification based on association are all examples of eager learners in that they use training tuples to construct a generalization model and in this way are ready for classifying new tuples. This contrasts with lazy learners or instance based methods of classification, such as nearest-neighbor classifiers and case-based reasoning classifiers, which store all of the training tuples in pattern space and wait until presented with a test tuple before performing generalization. Hence, lazy learners require efficient indexing techniques.

In **genetic algorithms**, populations of rules "evolve" via operations of crossover and mutation. Until all rules within a population satisfy a specified threshold. Rough set theory can be used to approximately define classes that are not distinguishable based on the available attributes. Fuzzy set approaches replace "brittle" threshold cutoffs for continuous-valued attributes with degree of membership functions. Linear, nonlinear, and generalized linear models of regression can be used for prediction.

Stratified *k*-fold cross-validation is a recommended method for accuracy estimation. Bagging and boosting methods can be used to increase overall accuracy by learning and combining a series of individual models. For classifiers sensitivity, specificity and precision are useful alternatives to the accuracy measure, particularly when the main class of interest is in the minority.

Fuzzy Logic in Cryptography:

Fuzzy logic[10][2] is a form of multi-valued logic derived from fuzzy set theory to deal with reasoning that is appropriate rather than precise. As mentioned above that the existing cipher has the disadvantage of frequency attack. So here Fuzzy logic is used to generate the knowledge based model with the help of which a fuzzy encryption table is created. By using this cipher text is generated which is close to the ideal line, so that we have the frequencies of all alphabets equal or very close to each other. Then it would be impossible for an attacker to proceed with the frequency attack. Here the attempt has made to make the cipher more secure with the help of Fuzzy logic.

5. CRYPTOGRAPHY

Secret Key Cryptography

With *secret key cryptography*, a single key is used for both encryption and decryption. As shown in Figure 1A, the sender uses the key (or some set of rules) to encrypt the plaintext and sends the cipher text to the receiver. The receiver applies the same key (or rule set) to decrypt the message and recover the plaintext. Because a single key is used for both functions, secret key cryptography is also called *symmetric encryption*. With this form of cryptography, it is obvious that the key must be known to both the sender and the receiver; that, in fact, is the secret. The biggest difficulty with this approach, of course, is the distribution of the key.

Secret key cryptography schemes are generally categorized as being either *stream ciphers* or *block ciphers*. Stream ciphers operate on a single bit (byte or computer word) at a time and implement some form of feedback mechanism so that the key is constantly changing. A block cipher is so-called because the scheme encrypts one block of data at a time using the same key on each block. In general, the same plaintext block will always encrypt to the same cipher text when using the same key in a block cipher whereas the same plaintext will encrypt to different cipher text in a stream cipher.

Stream ciphers come in several flavors but two are worth mentioning here. *Self-synchronizing stream ciphers* calculate each bit in the key stream as a function of the previous *n* bits in the key stream. It is termed "self-synchronizing" because the decryption process can stay synchronized with the encryption process merely by knowing how far into the *n*-bit key stream it is. Block ciphers can operate in one of several modes; the following four are the most important:

Electronic Codebook (ECB) mode is the simplest, most obvious application: the secret key is used to encrypt the plaintext block to form a cipher text block. Two identical plaintext blocks, then, will always generate the same cipher text block. Although this is the most common mode of block ciphers, it is susceptible to a variety of brute-force attacks.

Cipher Block Chaining (CBC) mode adds a feedback mechanism to the encryption scheme. In CBC, the plaintext is exclusively-ORed (XORed) with the previous cipher text block prior to encryption. In this mode, two identical blocks of plaintext never encrypt to the same cipher text.

Cipher Feedback (CFB) mode is a block cipher implementation as a self-synchronizing stream cipher. CFB mode allows data to be encrypted in units smaller than the block size, which might be useful in some applications such as encrypting interactive terminal input. If we were using 1-byte CFB mode, for example, each incoming character is placed into a shift register the same size as the block, encrypted, and the block transmitted. At the receiving side, the cipher text is decrypted and the extra bits in the block (i.e., everything above and beyond the one byte) are discarded.

Output Feedback (OFB) mode is a block cipher implementation conceptually similar to a synchronous stream cipher. OFB prevents the same plaintext block from generating the same cipher text block by using an internal feedback mechanism that is independent of both the plaintext and cipher text bit streams.

Data Encryption Standard (DES): The most common SKC scheme used today, DES was designed by IBM in the 1970s and adopted by the National Bureau of Standards (NBS) [now the National Institute for Standards and Technology (NIST)] in 1977 for commercial and unclassified government applications. DES is a block-cipher employing a 56-bit key that operates on 64-bit blocks. DES has a complex set of rules and transformations that were designed specifically to yield fast hardware implementations and slow software implementations, although this latter point is becoming less significant today since the speed of computer processors is several orders of magnitude faster today than twenty years ago. IBM also proposed a 112-bit key for DES, which was rejected at the time by the government; the use of 112-bit keys was considered in the 1990s, however, conversion was never seriously considered.

Public-Key Cryptography

Public-key cryptography has been said to be the most significant new development in cryptography in the last 300-400 years. Modern PKC was first described publicly by Stanford University professor Martin Hellman and graduate student Whitfield Diffie in 1976. Their paper described a two-key crypto system in which two parties could engage in a secure communication over a non-secure communications channel without having to share a secret key.

Generic PKC employs two keys that are mathematically related although knowledge of one key does not allow someone to easily determine the other key. One key is used to encrypt the plaintext and the other key is used to decrypt the cipher text. The important point here is that it does not matter which key is applied first, but that both keys are required for the process to.

RSA: The first, and still most common, PKC implementation, named for the three MIT mathematicians who developed it — Ronald Rivest, Adi Shamir, and Leonard Adleman. RSA today is used in hundreds of software products and can be used for key exchange, digital signatures, or encryption of small blocks of data. RSA uses a variable size encryption block and a variable size key. The key-pair is derived from a very large number, *n*, that is the product of two prime numbers chosen according to special rules; these primes may be 100 or more digits in length each, yielding an *n* with roughly twice as many digits as the prime factors. The public key information includes *n* and a derivative of one of the factors of *n*; an attacker cannot determine the prime factors of *n* (and, therefore, the private key) from this information alone and that is what makes the RSA algorithm so secure.

Hash Functions

Hash functions, also called *message digests* and *one-way encryption*, and are algorithms that, in some sense, use no key (Figure 1C). Instead, a fixed-length hash value is computed based upon the plaintext that makes it impossible for either the contents or length of the plaintext to be recovered. Hash algorithms are typically used to provide a *digital fingerprint* of a file's contents often used to ensure that the file has not been altered by an intruder or virus. Hash functions are also commonly employed by many operating systems to encrypt passwords. Hash functions, then, provide a measure of the integrity of a file.
Hash algorithms that are in common use today include:
Message Digest (MD) algorithms: A series of byte-oriented algorithms that produce a 128-bit hash value from an arbitrary-length message.

a. *MD2 (RFC-1319):* Designed for systems with limited memory, such as smart cards.
b. *MD4 (RFC 1320):* Developed by Rivest, similar to MD2 but designed specifically for fast processing in software. (MD4 has been relegated to historical status, per RFC 6150.)
c. *MD5 (RFC 1321):* Also developed by Rivest after potential weaknesses were reported in MD4; this scheme is similar to MD4 but is slower because more manipulation is made to the original data. MD5 has been implemented in a large number of products although several weaknesses in the algorithm were demonstrated by German cryptographer Hans Dobbertin in 1996 ("Cryptanalysis of MD5 Compress").

d. *Secure Hash Algorithm (SHA):* Algorithm for NIST's Secure Hash Standard (SHS). SHA-1 produces a 160-bit hash value and was originally published as FIPS PUB 180-1 and RFC 3174

Cryptography is the science of writing in secret code and is an ancient art; the first documented use of cryptography in writing dates back to circa 1900 B.C. when an Egyptian scribe used non-standard hieroglyphs in an inscription. Some experts argue that cryptography appeared spontaneously sometime after writing was invented, with applications ranging from diplomatic missives to war-time battle plans. It is no surprise, then, that new forms of cryptography came soon after the widespread development of computer communications. In data and telecommunications, cryptography is necessary when communicating over any untrusted medium, which includes just about *any* network, particularly the Internet.

Within the context of any application-to-application communication, there are some specific security requirements, including:

1. *Authentication:* The process of proving one's identity. (The primary forms of host-to-host authentication on the Internet today are name-based or address-based, both of which are notoriously weak.)
2. *Privacy/confidentiality:* Ensuring that no one can read the message except the intended receiver.
3. *Integrity:* Assuring the receiver that the received message has not been altered in any way from the original.
4. *Non-repudiation:* A mechanism to prove that the sender really sent this message.

Fuzzy logic[10][2] is a form of multi-valued logic derived from fuzzy set theory to deal with reasoning that is appropriate rather than precise. As mentioned above that the existing cipher has the disadvantage of frequency attack so here Fuzzy logic is used to generate the knowledge based model with the help of this a fuzzy encryption table is created by using this cipher text is generated which is close to the ideal line so that we have the frequencies of all alphabets equal or very close to each other then it would be impossible for an attacker to proceed with the frequency attack. Here the attempt has made to make the cipher more secure with the help of Fuzzy logic

6. Compared Cryptography Algorithms

The necessary background to understand the key difference between the compared algorithms.

Table1: Comparison of various Cryptography Algorithm

	DES	3DES	AES
Designers	IBM		Vincent Rijmen, Joan Daemen
First published	1977	1998	1998
Derived from	Lucifer	DES	Square
Successors	Triple DES, G-DES, DES-X, LOKI89, ICE	----------	Anubis, Grand Cru
Key sizes	56 bits	168, 112 or 56 bits	128, 192 or 256 bits
Block sizes	64 bits	64 bits	128 bits
Structure	Balanced Feistel network	Feistel network	Substitution-permutation network
Rounds	16	48 DES-equivalent rounds	10, 12 or 14 (depending on key size)

7. CONCLUSION

A new approach will be proposed for minimizing the attacks that occurs for various queries of database. In our approach we will consider privacy preservation of the query based system with the help of cryptography using Hyper Elliptic curve.

REFERENCES

1. *Aburrous, M., Hossain, M. A., Dahal, K., and Thabatah, F., (2009), "Modelling Intelligent Phishing Detection System for e-Banking using Fuzzy Data Mining," International Conference on Cyber Worlds IEEE, 37(12), pp. 265-272.*
2. *Ansari, A.Q., Patki, T., Patki, A.B., and Kumar, V., (2007), "Integrating Fuzzy Logic and Data Mining: Impact on Cyber Security," Fourth International Conference on Fuzzy Systems and Knowledge Discovery IEEE, 4, pp. 498-502.*
3. *Bhanumathi, S., and Sakthivel. (2013), "A New Model for Privacy Preserving Multiparty Collaborative Data Mining," International Conference on Circuits, Power and Computing Technologies IEEE, 3, pp. 845-850.*
4. *Dung, L., Bao, H.T., Binh, N., and Hoang, T.H., (2010), "Privacy preserving classification in Two-Dimension Distributed Data", Second International Conference on Knowledge and Systems Engineering IEEE, pp. 96-103.*
5. *Edwin, S.B., and Annie Portia, A., (2011), "Analysis on Credit Card Fraud Detection Methods," International Conference on Computer, Communication and Electrical Technology IEEE, 4(7), pp. 152-156.*
6. *Khelifi, A., Aburrous, M., Talib, M., and Shastry, P.V.S., (2013), "Enhancing Protection Techniques of E-Banking Security Services Using Open Source Cryptographic Algorithms," ACIS International Conference on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing IEEE, pp. 89-95.*
7. *Le, H.Q., and Arch-int, S., (2012), "A Conceptual Framework for Privacy Preserving of Association Rule Mining in E-Commerce," IEEE Conference on Industrial Electronics and Applications (ICIEA), pp. 1999-2003.*
8. *Masri, A., and Sousa, J.P., (2009), "Limiting Private Data Exposure in Online Transactions A User-based Online Privacy Assurance Model" International Conference on Computational Science and Engineering IEEE, pp. 438-443.*
9. *Ronchi, C., Khodjanov, A., Mahkamov, M., and Zakhidov, S., (2011), "Security, Privacy and Efficiency of Internet Banking Transactions" EISST Development Laboratory IEEE, pp. 216-222.*
10. *Wang, J., Luo, Y., Zhao, Y., and Le, J., (2009) "A Survey on Privacy Preserving Data Mining" First International Workshop on Database Technology and Applications IEEE, pp. 111-114.*
11. *Zhu, Y., Huang, L., Yang, W., Li, D., Luo, Y., and Dong, F., (2009), "Three New Approaches to Privacy-preserving Add to Multiply Protocol and Its Application" Second International Workshop on Knowledge Discovery and Data Mining IEEE, pp. 554-558.*
12. *Bonathu, R., Devaki, K., Reddy, R., Meghavath, D., Vijaya, G., (2014), "A Report of the Privacy in Data Mining: Speakers Survey", International Journal of Innovative Science and Modern Engineering (IJISME), ISSN: 2319-6386, Vol 2(4), pp: 4-6*
13. *Bertino, E., Fovino, I. and Provenza, L., (2005), "A Framework for Evaluating Privacy Preserving Data Mining*
14. *Algorithms", Data Mining and Knowledge Discovery, Vol 11(2), pp: 121–154.*
15. *Vaidya, J., and Clifton, C., (2004), "Privacy-Preserving Data Mining: Why, How, and When", IEEE computer society.*
16. *Kachwala, T. and Parmar, S., (2014), "An Approach for Preserving Privacy in Data Mining", International Journal of Advanced Research in Computer Science and Software Engineering, Vol 4(9), pp: 377-372.*
17. *Kruse, R., Nauck, D. and Borgelt, C., "Data Mining with Fuzzy Methods: Status and Perspectives", Department of Knowledge Processing and Language Engineering Otto-von-Guericke-University of Magdeburg*
18. *Tatusov, R.L., Altschul, S.F., Koonin, E.V., (1994), "Detection of conserved segments in proteins: iterative scanning of sequence databases with alignment blocks." Proceedings of the National Academy of Sciences of the United States of America, pp: 12091-12095.*
19. *Dhenakaran, S., and Kavinilavu, N., (2012), "a new method for encryption using fuzzy set theory", International Journal of Engineering Trends and Technology, Vol3 (3), pp: 320-326.*