International Journal of Database Theory and Application
Vol.8, No.4 (2015), pp.299-306
http://dx.doi.org/10.14257/ijdta.2015.8.4.30

Secure Multi-party Communication in Data-mining Applications

Anshu Chaturvedi[1], D.N. Goswami[2], Rishi Soni[3] and Brijesh Kumar Chaurasia[4]
[1]MITS, Gwalior, [2,3]Jiwaji University, Gwalior, [4]ITM University Gwalior
[1]anshu_chaturvedi@yahoo.co.in, [2]goswamidn@yahoo.com,
[3]rishisoni17@gmail.com and [4]brijeshchaurasia@ itmuniversity.ac.in

Abstract

Data mining extracts knowledge or patterns from a large amount of data. Secure communication is an issue of shared database applications. Fundamentally, secure multi party computation is to enable a number of networked parties to carry out distributed computing tasks on sensitive information. In this paper, we have proposed two techniques for multi party communication one is for third party assisted and another without third party. Third party assisted technique uses Group based approach and in case of without third party mechanism ECC based approach is used. Simulation results show that the time taken by the schemes are significantly less as compared to other schemes and this proves the efficacy of the proposed scheme which makes it viable for multi party communication in data mining applications.

Keywords: *Group based scheme, ECC, RSA, Multi party Communication and Trusted / No trusted third party.*

1. Introduction

Data mining is the process of extracting valuable, meaningful patterns and relationships that lie hidden within very large databases [1]. Data mining plays an important role in many applications such as business management, marketing analysis, medical analysis, criminal records [2], as well as in financial and scientific research [3]. In recent times data mining has gained immense importance as it paves way for the management to obtain hidden information and use them in decision-making [4]. While dealing with sensitive information it becomes very important to protect data against unauthorized access [5]. A key problem faced in this regard is the need to balance the confidentiality of the disclosed data with the legitimate needs of the data users [6]. In the application of multi party communication, there are a number of networked parties to carry out distributed computing tasks on private or sensitive data. Sensitive data may be disclosed or leaked in multiparty communication during sharing data for mining applications. So, maintaining confidentiality of data and preserving privacy in data mining are main issues of such applications. We have addressed confidentiality, authentication and traceability issues of security of data mining in this proposed work in the presence of third party and absence of third party. Group based scheme and ECC based schemes provides all necessary security issues for the two approached respectively.

1.1. Our Contribution

In this paper, we have worked at secure multi party computation over private or sensitive data shared application. The proposed heuristics is based on asymmetric cryptography techniques. The heuristics is client adaptive secure protocol, which depends on security level of application. If client required more security for his application he can use elliptic curve cryptography (ECC) [7] based key generation algorithm instead of RSA algorithm [8]. To provide confidentiality asymmetric encryption / decryption algorithm

ISSN: 2005-4270 IJDTA

[9] is recommended. We illustrate this idea using third party in case third party is trusted or non-trusted. The proposed heuristic is able to achieve confidentiality, authentication privacy preserving data mining. Moreover, it is also able to achieve traceability of un-trusted third party.

The rest of the paper is organized as follows. Section 2 describes the related work. Problem definition is given in section 3. Proposed scheme is presented in section 4. The scheme is evaluated through simulation and results are in section 5; section 6 concludes the work.

2. Related Works

There are several cryptographic approaches available for protection of data mining. Cryptographic schemes are based on symmetric and asymmetric approaches. These existing techniques fail to cope up with all security issues such as privacy, authentication, traceability and confidentiality etc. Multiple levels of privacy for data mining approach in a distributed environment to secure sensitive and private information is presented in [10]. The DES and RSA algorithm to preserve privacy in data mining for sharing the data with a trusted third party is used in [11]. In this technique, trusted third party is responsible for setting the keys, RSA algorithm is used forsecured sharing of the secret key and DES algorithm for encryption and decryption through secret key. The technique is divided in two phase. In the first phase, DES is used to encrypt the values of the private attributes and in the second to encrypt the secret key of the first phase. However, the only secret is single key so it may be compromised or leaked easily. The problem with this technique is that secret key is also online exchanged. Cryptographic role based access control approach to preserve privacy in data mining is discussed in [12]. The work provides privacy for two sets of objects sensitive and non sensitive. In this technique, privacy can be achieved using encryption where server first splits data into sensitive and non sensitive objects. Non sensitive objects are accessed by all the clients and only sensitive objects are encrypted by standard encryption technique. The cryptographic technique is used to store sensitive data. It provides access to the stored data based on an individual’s role to safeguard the data from privacy breaches. Other secure multi party computation is presented in [13] using cryptography for different kind of problems in data mining. RSA based public key cryptography [14] is used in [15] to achieve privacy in data mining. Public key cryptography uses two keys, one of them is public key and another is private key. Public key is used for encryption while private key is used for decryption. So, confidentiality and privacy is achieved by this approach. The advantage of this scheme is its ability to communicate with many numbers of sites without any modification. This scheme is also compared with many previous well known techniques. Multilevel privacy preservation algorithm using DES and RSA to achieve privacy in data mining is presented in [16]. In this work, RSA is used to generate keys (public and private key) by trusted third party. The data is encrypted using secret key in DES algorithm. The advantage of DES algorithm is that it can use keys of different length. So, the scheme provides adaptive privacy in which the length of key is based on the level of security required.

3. Problem Formulation

The problem dealt in the present work is related to the information sharing based application of data mining, where multiple parties may aggregate or share private data for the purpose of knowledge discovery. Disclosure or leaking of sensitive information during multiple parties’ communication is a critical issue of data mining. Such types of applications require secure heuristic for sharing the information across multiple parties. In general, secure communication particularly confidentiality along with authentication in privacy preserved data mining is needed. Hence, multi party communication is more vulnerable than two party communications due to the nature of communication in network. The key exchange, confidentiality, traceability and authentication issues over this shared communication in data mining are resolved in this paper.

4. Proposed Methodology

To achieve the confidentiality, authentication and traceability issues of security the cryptographic approach is most suited for multiparty communication in data mining applications. The proposed solution is efficient enough to preserve the security of data stored at several private parties and who agree to share or disclose the outcome of data mining computation. We have considered two cases to measure efficacy of proposed protocol: third party assisted and without third party.

Case 1: Group based Communication using third party assisted

In this section, the third party assisted group signature based scheme is proposed and presented. In this scheme, group members are fixed, so static group based scheme is required. Third party plays the role of group manager in this scheme. Group manager will be responsible to generate the group keys. Group key consists of public group keys and private keys for secure communication among them. Group manager distribute keys after authentication of each parties.

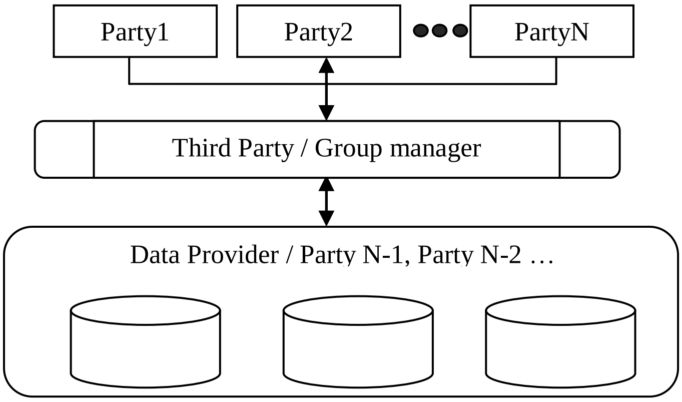

Figure 1. Secure Multiparty Communication using Group Based Scheme

Data provider and number of parties which are involved in multi party communication are the group members in this proposed scheme. Fig. 1 shows the group manager and group members in this scheme.

The formation of a group signature has five phases or protocols as in [17] and [18]. Key generation phase is *Setup* phase, new group member wants to join the group then he can use *Join* protocol, to sign the message *Sign* phase is used by each members of the group, after receiving this message group member will use the *Verify* protocol and to trace the signature if dispute occurs group manager will use the *Open* protocol.

The proposed third party assisted group based scheme is able to provide anonymity, where, actual signer cannot be identified from the group because two different signatures of the same group member are not and can not be linked. However, the group manager can always reveal if any dispute occurs to identify the group member(s) and finally no non member can forge and sign a message on behalf of the group. So, this scheme is most suited for secure group based communication whenever members are fixed.

Case 2: ECC based Communication using without third party assistance

In this section, to achieve confidentiality and secure communication ECC signature based on digital signature is discussed in [19]. We assume in the case ofAlice sends a message to Bob, Alice is party 1 and Bob is party 2 respectively. In this scheme, parties may generate the keys without involving third party. To convince Bob that the message does come from Alice, Alice needs to apply a digital signature for the message so that Bob can verify it by using Alice's public key [20]. Initially, Alice and Bob have to agree on a particular curve with base point P over the field $GF(p)$, and the order of$P$ is$q$. When Alice sends a message to Bob, she attaches a digital signature (r, s) generated by following steps

1. Alice and Bob agree on an elliptic curve E over a finite field F_q so the discrete logarithm problem is hard in$E(F_q)$. They also agree on a point $P \in 2E(F_q)$ such that the subgroup generated by P has large order (usually prime).
2. Alice chooses secret integer, a, and computes $P_a = aP$and then sends P_a to Bob.
3. Bob chooses secret integer, b, and computes $P_b = bP$ and then sends P_b to Alice.
4. Now Alice computes $aP_b = abP$ at one end and Bob computes $bP_a = abP$ at other end.
5. Alice and Bob make a common opinion on a method to extract a key from abP. (For example, use the last 256 bits of the x-coordinate.)

The only information the eavesdropper has, the curve E, the finite field F_q, and the points P, $aP$and bP. She will therefore need to solve: Diffie-Hellman problem for elliptic curves: Given P, aP and bP in $E(F_q)$compute. If Eavesdropper can solve discrete logs in $E(F_q)$ then she could use P and aP to find a. She could then compute $a(bP)$ to get abP, however, if E and F_q are chosen carefully then this is considered computationally infeasible. It is not known whether there is a way of computing abP without first solving a discrete log problem. All the elliptic curves are not ellipses. They are so named because of the fact that ellipses are formed by quadratic curves. Elliptic curves are always cubic and have a relationship to elliptic integrals in mathematics [21] where the elliptic integral can be used to determine the arc length of an ellipse. An elliptic curve in its "standard form" is described by

$$y^2 = x^3 + ax + b$$

For the polynomial $x^3 + ax + b$, the discriminant can be given as

$$D = -(4a^3 + 27b^2)$$

This discriminant must not become zero for an elliptic curve polynomial $\mathrm{x}^3 + \mathrm{ax} + \mathrm{b}$ to possess three distinct roots. If the discriminant is zero, that would imply that two or more roots have coalesced, giving the curves in singular form. It is not safe to use singular curves for cryptography as they are easy to crack. Due to this reason we generally take non-singular curves for data encryption. The advantage of ECC based signature scheme over RSA scheme is that ECC offers considerably greater security for a given key size. ECC with 160 bit key can provide the same level of security as RSA can provides with 1024 bit key. Elliptic curve cryptography has not only emerged as an attractive public key crypto-system for mobile/wireless environments but it also saves bandwidth. ECC is not easy to be understood by attacker, and hence not easy to be broken. In short asymmetric cryptography is demanding but looking at the cryptosystem for more security per bit, ECC is a better choice than RSA asymmetric key cryptosystem.

Case 2.1: ECC Based Proposed Algorithm for Secure Communication in Datamining

In this section, proposed ECC based algorithm for secure communication in datamining is presented. The proposed algorithm is applicable for multi party communication in datamining application. The use of ECC is also presented in [22]. An elliptic curve is a cubic equation of the form $E: y^2 + axy + by = x^3 + cx^2 + dx + e$, where $a, b, c\ and\ e$ are real numbers. The mathematical equation of ECC satisfies the form $E: y^2 = (x^3 + ax + b) mod\ p$ with $a, b \in F_p$ satisfying $(4a^3 + 27b^2) mod\ p \neq 0$. Here p is prime number of elliptic curve group. The key exchange between $Party_1$ and $Party_2$ using ECC is as follows:

2.1.1 Secure Key Exchange Algorithm

Step1:*$Party_1$ send public key to $Party_2$*

The user Party1 chooses a random integer r_1 as a private key, where $r_1 < n$ and compute the public key Q_1. Here $Q_1 = r_1 \times p$. Thus,Party1 send public key Q_1 to Party2.

Step2: *$Party_2$ send public key to $Party_1$*

The user Party2 chooses a random integer r_2 as a private key, where $r_2 < n$ and compute the public key Q_2. Here $Q_2 = r_2 \times p$. Thus,Party2 send public key Q_2 to Party1.

Both the users Party1 and Party2 can compute shared secret key K_1 and K_2 respectively.

$$K_1 = r_1 \times Q_2 = r_1 \times r_2 \times p$$

$$K_2 = r_2 \times Q_1 = r_2 \times r_1 \times p$$

Hence, $K_1 = K_2$ keyexchange is assumed symmetric key cryptosystem. ECC provides a smaller key size or half of the size of other public key cryptosystem along with faster computations.

2.1.2 Secure Data Exchange Algorithm

The proposed protocol may be used between client- server protocol and multi-party secure communication.

Step1:Party1 sends a REQUEST message to Party2. After agreeing Party2 sends RESPONSE message to Party1. The key exchange between two parties and among multi party will be as 2.1.1. After that Party1 sends messages using encryption algorithm.

Step2: Party1 sends an encrypted message to Party2 using shared key.

$$Party_1 \rightarrow E_{K_1}[message, t_0]$$

Here, E_{K_1} is the encryption function using key K_1,for the purpose of achieving confidentiality for message(s) containing sensitive informationand $t_0$is timestamp to achieve message integrity.

Step3: Party2 receives anencrypted message from Party1 using shared key to achieve original message by decrypted function.

$$Party_2 \rightarrow D_{K_2}[message, t_0]$$

Here, $D_{K_1}$is the decryption function using key K_2, to achieve original message.

The ECC based proposed scheme is able to achieve message integrity, confidentiality and secure communication among multi party used in data mining applications.

5. Results and Analysis

Generally it is seen that in data mining approach machine configuration is not an issue, rather generation of key and encryption and decryption process is cumbersome and takes order of milliseconds. This causes increased delay in the process. Therefore, to evaluate the performance of the proposed algorithm, the metric taken into consideration is delay. The delay here is calculated with the help of MIRACL [23] running on 2.50 *GHz* CPU and 2 *GB* RAM and Windows XP. The proposed protocol is able to provide confidentiality, authentication and traceability for multi party communication in data mining application. The protocols can work in both the conditions when third party is trusted or not trusted. We have used group based and ECC based approach to achieve confidentiality, authentication and non repudiation.

Table 1. Delays of 1024 Bit RSA Decryption

512 DSA 160 bit exponent	
Signature No precomputation	0.39 *ms*
Signature w. precomputation	0.08 *ms*
Verification	0.47 *ms*

Computational delays using RSA 1024 bit RSA decryption, 1024 bit DSA 16 bit component encryption / decryption and ECC based scheme is shown by Table1, Table 2 and Table 3 respectively.

Table 2. Delays of 1024 Bit DSA 160 Bit Component

DSA 160 bit component	
Signature No precomputation	1.27 *ms*
Signature w. precomputation	0.24 *ms*
Verification	1.53 *ms*

Table 3. Delays of 160 Bit Elliptic Curve Cryptography

160 bit GF(p) ECC	
ER 7433 iterations	1.35 *ms*
ED 6020 iterations	1.66 *ms*
EP 33800 iterations	0.30 *ms*

Verification delay of group signature based scheme is around 3.6 ms [17]. Results show that key length is 1024 RSA and DSA both techniques are equivalent secure to 160 bit ECC based technique for secure communication in datamining application because of computational power is not an issue in such applications. Results of Tables 1 - 3 show that ECC and group based approach is more suitable technique for online multi party communication in datamining applications.

6. Conclusion

In this paper, we addressed the security issues of multi party communication in data mining applications. Two techniques trusted third party assisted and without third party are proposed to tackle security issues. Furthermore, we propose group based and ECC based approach. The techniques are verified even when a large number of parties are involved. The proposed mechanism seems to be potentially capable of achieving the goal of obviating the need for multi party communication in data mining applications.

References

[1] M. S. Chen, J. Han and P. S. Yu, "Data mining: an overview from a database perspective", IEEE Transactions on Knowledge and Data Engineering, vol. 8, no. 6, **(1996)**, pp. 866–883.
[2] E. Magkos, M. Maragoudakis, V. Chrissikopoulos and S. Gritzalis, "Accurate and large-scale privacy-preserving data mining using the election paradigm", Data & Knowledge Engineering, vol. 68, **(2009)**, pp. 1224–1236.
[3] A. Agrawal, U. Thakar, R. Soni and B. K. Chaurasia, "Efficiency Enhanced Association Rule Mining Technique", International Conference on Parallel, Distributed Computing technologies and Applications (PDCTA), **(2011)**.
[4] V. S. Verykios, A. K. Elmagermld, E. Bertino, Y. Saygin and E. Dasseni, "Association Rule Hiding" IEEE Transactions on Knowledge and Data Engineering, vol. , no. 4, **(2004)**, pp. 447.
[5] S. L. Wang, Y. H. Lee, S. Billis and A. Jafari, "Hiding Sensitive Items in Privacy preserving Association Rule in Mining", IEEE International Conference on Systems, MAn and Cybermetics, **(2004)**.
[6] D. Jain, P. Khatri, R. Soni and B. K. Chaurasia, "Hiding Sensitive Association Rules without Altering the Support of Sensitive Item(s), "Advances in Computer Science and Information Technology, Networks and Communcations, vol. 84, **(2012)**, pp. 500-509.
[7] D. Hankerson, A. Menezes and S. Vanstone, "Guide to Elliptic Curve Cryptography", Springer, **(2004)**.
[8] S. Paine, "RSA Security's Official Guide to Cryptography", RSA Press, **(2001)**.
[9] A. J. Menezes, P. C. Van Oorschot and S. A.Vanstone, "Handbook of Applied Cryptography", CRC Press, **(1996)**.
[10] A. Y. Fuad and M. Sonajharia, "Multilevel Privacy Preserving in Distributed Environment using Cryptographic Technique", Proceedings of the World Congress on Engineering, **(2012)**.
[11] X. Zhou and X. F. Tang, "Research and Implementation of RSA Algorithm for Encryption and Decryption", The 6^{th} International Forum on Strategic Technology, **(2011)**.
[12] L. Vasudevan , S. E. DeepaSukanya and N. Aarthi, "Privacy Preserving Data Mining Using Cryptographic Role Based Access Control Approach", Proceedings of the International MultiConference of Engineers and Computer Scientists, **(2008)**.
[13] D. Bogdanov, M. Niitsoo, T. Toft and J. Willemson, "High-performance secure multi-party computation for data mining applications", International Journal Information Security, vol. 11, **(2012)**, pp. 403-418.
[14] W. Du and M. J. Attalah, "Secure multi-problem computation problems and their applications: A review and open problems", Proceedings of New Security Paradigms Workshop, **(2001)**.
[15] C. Clifton, M. Kantarcioglou, X. D. Lin and M. Y. Zhu, "Tools for privacy preserving distributed data mining", ACM SIGKDD Explorations, 4, no. 2, **(2002)**, pp. 28- 34.
[16] J. Vaidya and C. Clifton, "Privacy preserving association rule mining in vertically partitioned data", the 8th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, **(2002)**.
[17] B. K. Chaurasia, S. Verma and S. M. Bhasker, "Message broadcast in VANETs using Group Signature", Fourth International Conference on Wireless Communication and Sensor Networks, **(2008)**.
[18] D. Chaum and E. V. Heijst, "Group signatures", Proc. Advances in Cryptology - Eurocrypt, LNCS, Springer-Verlag, **(1991)**.
[19] D. Hankerson, A. Menezes and S. Vanstone, "Guide to Elliptic Curve Cryptography", Springer, **(2004)**.
[20] B. K. Chaurasia and S. Verma, "Secure Pay While On Move Toll Collection through VANET", Computer Standards & Interfaces, Elsevier, vol. 36, no. 2, **(2014)**, pp. 403-411.
[21] J. Tate, "The arithmetic of elliptic curves", Invent Math, vol. 23, **(1974)**, pp. 171-206.
[22] V. S. Miler, "Use of elliptic curves in cryptography", LNCS, Advances in Cryptology- Crypto-85: Proceedings, Springer Berlin, Heidelberg, **(1986)**.
[23] Shamus Software Ltd. MIRACL, "Multiprecision Integer and Rational Arithmetic C/C++ Library", Online Available at: http://indigo.ie/~mscott.

Authors

Brijesh Kumar Chaurasia is received his Ph.D. from Indian Institute of Information Technology, Allahabad, India in Privacy Preservation in Vehicular Ad hoc Networks. He is received his M.Tech. Computer Science and Engg. from D.A.V.V., Indore, India. He is an Professor at ITM University Gwalior, India.

Anshu Chaturvedi is Currently working as Reader in the Department of Computer Applications at Madhav Institute of Technology and Sciences, Gwalior. She has obtained her Ph. D. in 2009 in the area of Security in Adhoc Networks. Her research interests include Security in Adhoc Networks, Sensor Networks, Cloud Computing, Data Mining along with Privacy Preserving in Data Sharing. She is a life member of Computer Society of India and ISTE. She has more than ten years of experience in the academic field and almost eight years of experience in the research field. She has published several research papers in the International Journals and Conferences. She has been a reviewer for IEEE conference paper as well. She won Young Scientist Award by M. P. Council of Science and Technology in 2009.

D.N. Goswami is a Professor and Head in the School of Studies in Computer Science & Applications, Jiwaji University, Gwalior. He is currently holding the post of Director, School of Engineering, Jiwaji University as well. He has done Master of Computer Applications (1989) and Ph.D. in Computer Science (2004)from Jiwaji University, Gwalior. His research interests include Reliability, Software Engineering, Data Mining, Data Base Management Systems and computer Networks.

Rishi Soni is pursuing his Ph.D. from Jiwaji University, Gwalior India in Privacy Preservation Data mining. He is received his M.Tech. from SATI, Vidisha India.

