ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 2, February 2017

A Proposed Algorithm for Insider Collusion Attack on Privacy-Preserving Data Mining System by Non Homomorphic Encryption Methods

Rinku B. Kapdi, Hiral Agravat

M.E Student, Dept. of C.E., Noble College of Engineering, Junagadh, India

M.E Student, Dept. of C.E., Noble College of Engineering, Junagadh, India

ABSTRACT: In this paper, There is many types of threts are there for example Data owner,Insider,outside.From that a insider threat for privacy preserving for DKBDM distributed kernel based data mining for example distributed support vector machine. From all data breaching problem insider data attacks found most. Insider attacks name comes in top three central data violations. It mostly works on distribution of data mining and in this we will make design to protect our data against collaborative organizations. An untrustable system allow breaches to go without knowing and insider leak the data to the outsider and then outsider will get much more information from that data.On our solution we Are implementing global SVM classification model in that different parties will share their data to each other without disclosing to each other and we sketched vertically and horizontally data.

KEYWORDS: Insider, Outsider, breaches

I.INTRODUCTION

Insider attacks are arise from staff inside the company's enterprise not from the security errors of the system.Application of data mining mostly works on to store huge amount of data.in that data mostly it contains private and personal information thatswhy researchers mostly focused on dealing with privacy breaches.Support Vector Machine SVM is on of the prime area of research in privacy preserving.SVM is to map data into a higher dimensional feature by kernel tricks and also maintain archives with better mining results.State of the art privacy preserving scheme provide to securely merge kernels.And while transmission they encoded and hid the kernel values in a noisy mixtures.so that nobody can retrieve the original data.In that we used gram matrix computation.From the gram matrix we can computed different kernels.Here he issue is scalability it's a key challenge here.To make a gram matrix we want a dot product of every pair and key is communication cost.When the data is centralized, Our method generates the same SVM classification model.In our algorithm we quantify efficiency and security.in this we assume that each party does follow the proposed protocol correctly and does not collude. In that insider is key player with an attacker while sharing the data and from kernel value it can recover original data from SVM model. This is more realistic attack as its need to fetch few entries of data rather than entire database from an organization by this they can successfully fetch all the private data which is remaining.Her is the figure of different attack model in DKBDM.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 2, February 2017

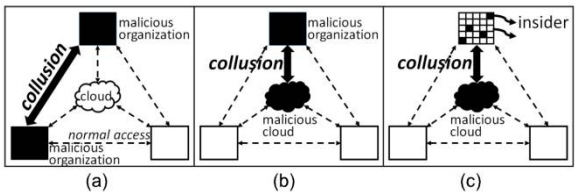

Fig 1.1 Different attack models in DKBDM [1]

II.RELETED WORK

For our knowledge In that insider is key player with an attacker while sharing the data and from kernel value it can recover original data from SVM model. This is more realistic attack as its need to fetch few entries of data rather than entire database from an organization by this they can successfully fetch all the private data which is remaining.

TYPES OF SVM DATA PARTITIONED
Vertically Partitioned Data: In vertical partitioned data parties collects difffrent data from the same set of entities.For example insurance company, a bank, and a health insurance company collect same type of data from same people.We can take example of a bank in that a bank keep record of account balance, average monthly deposit,etc.The car insurance company has right to get the data of types of car, accident claims, etc. The health insurance company has right to get the data of policy and medical information.From only local SVM model the global SVM model G can't built.So that we can't use use a local SVM model. The locally optimal coefficient computed on local data is different from the the globally optimal coefficient.

Horizontally partitioned data: In Horizontally partitioned of data from different data objects each party collect information which contains same features. For example different insurance company collect information about the customer such as name, age, gender,etc.which are same for all insurance company.In different banks they are collecting the data for their customer such as balance, gender, average monthly deposit,age,etc.which are same for all banks.and in horizontally partitioned, over each data pair we have to compute dot product so that we can securely compute the global gram matrix G.From all such method we are using secure dot product computation method.which is insecure or inefficient to be applied for gram matrix.To compute each scalar product it must run the protocol on every data pair, To secure and indeed use of of protocol scalar product protocol is useful.

III.FRAME WORK OF DATA ANALYSIS

For implementing this scheme there are many systems are available but among those we select that system which system use kernel values rather than original data such as by using securely merging kernels, vaidya propose a privacy preserving distribution SVM.In the threatscenario there are three players.

Data Owners Organization or Clients:In this organization has their own personal data and it can be trusted and they may take participate in distributed computing environment.

Insiders: Insiders are semi trusted as theses members are part of data owner's organization. And there is chances that they collude the organization's information to the attackers and they will not leak full information about the organization but some cotent of the data they will leak.

Outsider: The outsiders are not part of the organization we can not fully trust to this group as they are collude with insider.Some time while coordinates shares the different subset among them at that time the data mining server are coordinating and there is chances that it may act asif an outsider.this data mining server who are acting asif an outsider

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 2, February 2017

they may know the parameters of data mining but the data has been packed into a kernel format so it can't be access by outsider.

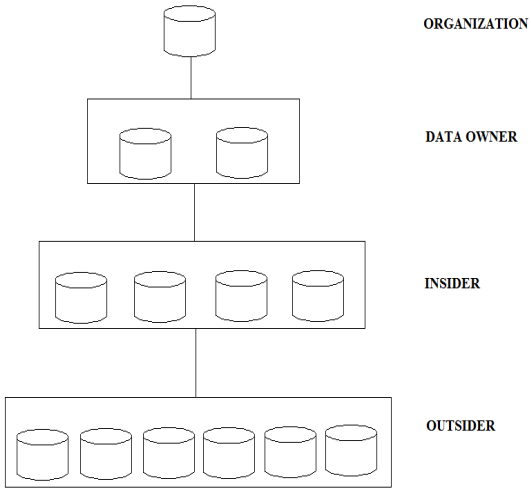

Fig 3.1 Players in Investigated threat

By seeing this figure I want to show you that suppose there is an organization and in organization there will be many important and private data.So there is possibility that someone will collude the data there are three types of data threat ono of them is Data owner meanes if there are more than one owner of data than there is chances that one of the owner will collude the information of data to the attackers and secondone is Insider,Insider is member of organizer he is not owner of this organizer but he is subpart of this institute he may collude the some information to the attackers and third one is outsider who are not part of this institute but because of isider collude some data to them they may now fetch more information about that data.So these are the players in the investigated threat.

THE STATE OF PRIVACY PRESERVING SVM SYSTEM[1]
Local Kernel matrix calculation
Transmitting of Local Kernel Matrix to the server.
Partial Weight calculation
Local Kernel Matrix Calculation Horizontally
In horizontally partitioned data there may be m × n data matrix A is there and A1 and a2 are part of them and k1 and K2 are m × m gram matrix of A1 and A2, respectively.So that K1=A1A1' and K2=A2A2'. So K is the gram matrix of A are as follows:
K = K1 + K2 = A1A1' + A2A2'
For brief describe suppose an (I,j)th element of k is xi·xj, In A xi and xj are ith and jthvalue.Suppose xi1 and xi2 are vectors of xi which are part of A.
So,
xi · xj = x1I · x1j + x2i · x2j
If we'll partitioned the A data matrix into r1, r2, r3 as drawn and there is gram matrix Ka of Hospital 1's Local gram matrix.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 2, February 2017

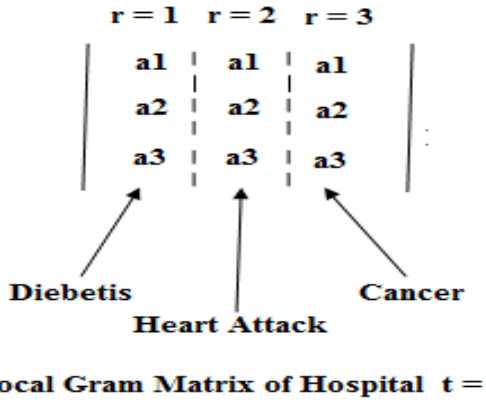

Fig 3.2 An example of 3×3 matrix

To build a global SVM model Organization use local data from vertical partitioned data in the SVM server from the high point of view,In vertically partitioned data records in entire set in full data matrix.there will be n records, and m features.A different hospital's data will be represented by a vertical column in data matrix.In equation (1) & (2) presents relationship of localkernel-merging theorem.two different patient data records will be presented by xi &xj.In the hospital r, xir is part of r data records of xi. In the hospital r, xjr is part of r data records of xj. As we have shown into the equation (3).and total number of hospitals are represented by z.

$$x_i^T x_j = x_i^{1T} x_j^1 + x_i^{2T} x_j^2 + \cdots + x_i^{zT} x_j^z \quad (1)$$

$$K_{ij}^{global} = K_{ij}^1 + K_{ij}^2 + \cdots + K_{ij}^z \quad (2)$$

$$K_{ij}^r = x_i^{rT} x_j^r \quad (3)$$

Here, we can see the euation of local gram matrix. By using his equation we can now get the local gram matrix and by merging all local gram matrix into 1 matrix in Horizontally we can get the gram matrix or we can say horizontally partitioned gram matrix. And after that we can merge all gram matrix into a matrix and we can get global gram matrix.

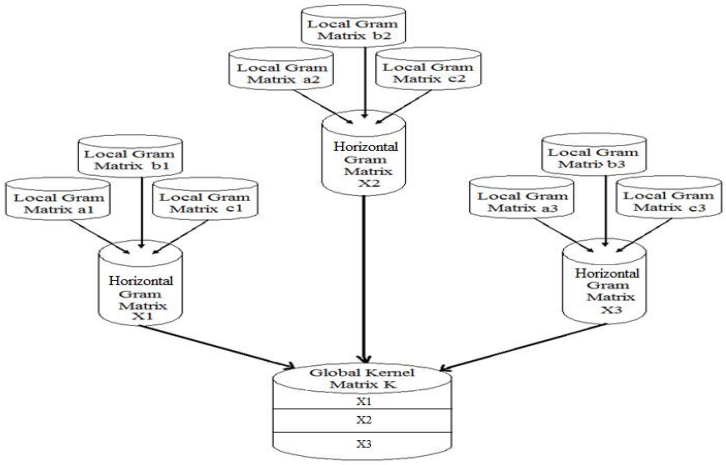

Fig 3.3 Flow Chart of Global Gram Matrix

Here we can see from the figure 3.3 that from the hospital 1 data we are getting the local gram matrix a1,b1,c1….etc

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 2, February 2017

As per that from hospital 2 and hospital 3 we can get the data a2,b2,c2…etc and a3,b3,c3…etcrespectively.and now from these local gram matrix we can apply the dot matrix to each row and apply horizontal partitioned data and we can get the value horizontal gram matrix x1,x2,x3 and from these gram matrix we can again apply the horizontal partitioned data and we can get the kernel global gram matrix K.

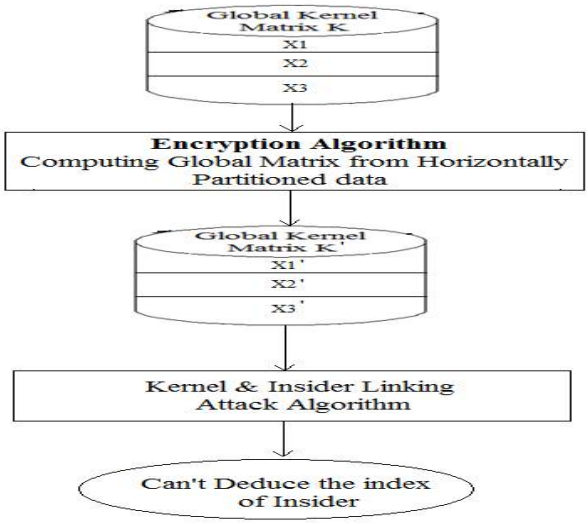

Fig 3.4 Flow Chart of Attack and Encryption

Now on the Global kernel matrix we can apply the encryption algorithm and after that we'll get encrypted x and its call x' and now if outsider will apply the algorithm of Kernel and Insider Linking algorithm then they may not find the index of got information and if they will not get the index of given data then they can not find more information of given data.

Here, from the flow chart we can come to know that by applying the dot vector on local gram matrix a1,b1 and c1 of hospital 1 we can get the gram matrix X1 and by applying the dot vector on local gram matrix a2,b2 and c2 of hospital 2 we can get the gram matrix X2, and by this method we can get X3 also. After that by merging the gram matrix X1,X2,and X3 we can get the Global kernel matrix K and now If outsider will find out the index of given data of insider then it can easily fetch out private data of the organization,here we are going to apply one encryption method by using Computing global gram matrix from horizontally partitioned data we can get more security and privacy. So, we can get the Global encrypted kernel matrix K' Now if attacker will attack on this data then he is not able to deduce index of insider.

Attack Algorithm-Kernel and Data Linking Algorithm[1]
Require: m × m kernel matrix KM, total m data records x1 ~xm, and total n insider's data s1~sn
1: for k = 1… n do
2: {Compute K1 and K2, where K1 is the kernel value of (sk , sp;p≠k;1≤p≤n), and K2 is the kernel value of (sk, sq; q≠k || q≠p; 1≤q≤n)}
3: Let KC1 = [], KC2 = [], l1 = 0, l2 = 0, IndexCand = [], Index = []
4: for for i = 1…m do //Search for values equal to K1and K2 in KM
5: for j = 1…m do
6: if KM(i, j) = K1 then
7: KC1(l1) = (i; j)

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 2, February 2017

8: else if (KMi; j) = K2 then
9: KC2(l2) = (i; j)
10: end if
11: end for
12: end for
13: for u = 1… max(l1) do //Apply Principle 1 & 2 to kernel lines
14: for v = 1 … max(l2) do
15: if KC1(u)[1] ≠ KC1(v)[1] & KC1(u)[2] = KC1(v)[2] then
16: if no element of the array IndexCand(k) =KC1(u)[2]then
17: Insert the element KC1(u)[2] into the array IndexCand(k)
18: end if
19: end if
20: end for
21: end for
22: end for
23: for k = 1… n do //Apply Principle 3 to kernel lines
24: if #element of IndexCand(k) = 1 then
25: Index(k) D theelementofIndexCand(k)
26: end if
27: end for
28: for k = 1… n do
29: if #element of IndexCand(k) > 1 then
30: Delete all elements of IndexCand(k) that has been assigned to
the other Index
31: Index(k) =a randomly chosen ele-ment from the remaining
elements of IndexCand(k)
32: end if
33: end for

There are three principle to for attackers,
These are as follows:
It's consider only vertical and horizontal kernel lines as there is only symmetrical property in the kernel matrix
For the same axis of the index, merge the kernel lines as its represent the same index
If the indices is representing the othe insider's data then remove the kernel lines.
To protect our data from the attackers we have to encrypt our data so they cannot fetch our data.

Computing Global Gram Matrix FromHorozontallyPartioned Data.
Require: A third party Q, who receives the gram matrix and creates the classifier
1: Q creates a new semantically secure homomorphic encryption system keypair {pk, sk}
2: Q sends the public key pk to all of the parties
3: for i = 1 . . .m do
4: for j = 1 . . .m do
5: {Compute the dot product of data point i with data point j }
6: for k = 1 . . . n do
7: Let Pa hold Aik and Pb hold Ajk
8: Pa computes mk = Epk (Aik , r), where r is a random nonce and sends it to Pb
9: Pb computes m' k = mAjk = Epk (Aik , r)Ajk = Epk (Aik □ Ajk , r')
where r 'is some number from the domain of r
10: end for
11: {The parties together compute $\prod_{k=1}^{n}$m'k}

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 2, February 2017

12: res = 1
13: for k = 1 . . . n − 1 do
14: The party that owns m'k computes res = res □ m'k and sends it to the party owning m'k+1
15: end for
16: The party owning m'n computes res = res □ m'n and sends it to Q
17: Q receives res = $\prod_{k=1}^{n}$m'k= Epk (Σk=1n Aik □ Ajk , r'')
18: Q decrypts this using sk to get the desired dot product
19: end for
20: end for

For any of the cases we can apply this general solution.and it's really very helpful for every data partitioned.we have shown you that when data is horizontally partitioned then how will we merge it.to generate gram matrix it's a key idea.We can also use upgraded version of scalar product in which it use homomorphicmethod.Secure public key is similar to homomorphic encryption method.But in this homomorphic encryption method it gives extra plus point that its gives two encryption E(A) & E(B) and there will be existence of E(A*B) So that we can get the results as E(A) * E(B) = E (A*B) as we can take * as addition or multiplication.Additivelyhomomorphic system is being mentioned earlier by the cryptosystems mentioned.By using this type of system it's become very easy to create scalar product protocol. The key is to note that Σk=1n xi · yi = Σk=1n (xi + xi +· · ·+ xi) (yi times).as all vectors are horizontally partitioned so each party have own xi encrypts and it send to the another party which is having corresponding yi.To transfer the product in encrypted form, additive homomorphic method will be used by this party now,To computed the the dot product its need sum of all products.

Now to compare data before applying the encryption method and after applying the encryption method.

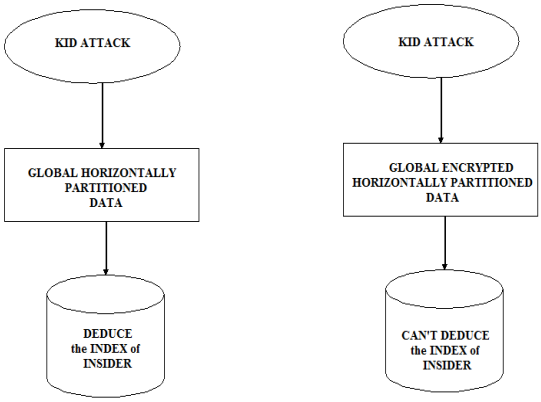

Fig 3.5 Comparision

Here we can compare our existing system and proposed system by that we can get the idea that before encryption the outsider can find the index of given data and can fetch more information of that but after incryption they can't find the index of any data.

IV. CONCLUSION

For privacy preserving SVM classification method we propose a scalable solution which is based on gram matrix.By assuming third party which is not trustable.In this we show that without disclosing any data or any information to eachother, how to compute secure global SVM model.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 2, February 2017

Future Work- Our proposed attack scheme is not only applicable to the vertically partitioned data and horizontally partitioned data but also applicable to arbitrarily partitioned data.For the everse from that kernel values we can take original data back.as its composed of two data vectors.and its store its value in th Kernel Matrix.

v. ACKNOWLEDGEMENT

Every thesis big or small is successful largely due to the effort of a number of wonderful people who have always given their valuable advice or lend a helping hand. I sincerely appreciate the inspiration, support and guidance of all those people who have been instrumental in making this thesis a success.I would like to express my deepest gratitude to my guide Prof. Daxa V. Vekariya for his unwavering support, collegiality and mentorship throughout this thesis. Apart from that his valuable and expertise suggestion during documentation of my report indeed help me a lot.I would like to extend my thanks to those who offered collegial guidance and support to make this thesis: Prof. Deep Patel , Prof. AshutoshAbhangi and Prof. YagneshParmar.And at last but not least, I would be grateful towards my parents and friends who had supported a lot and provided inspiration and motivation to go ahead in this project.

REFRENCES

[1] PETER SHAOJUI WANG, FEIPEI LAI, (Senior Member, IEEE), HSU-CHUN HSIAO, “Insider Collusion Attack on Privacy-Preserving Kernel-Based Data Mining Systems” Received April 18, 2016, accepted April 25, 2016, date of publication April 29, 2016, date of current version May 23, 2016.

[2] Amine Rahmani, Abdelmalek Amine, Reda Mohamed Hamou, “A Multilayer Evolutionary Homomorphic Encryption Approach for Privacy Preserving over Big Data” 2014 International Conference on Cyber-Enabled Distributed Computing and Knowledge Discovery

[3] W. R. Claycomb and A. Nicoll, ``Insider threats to cloud computing: Directions for new research challenges," in Proc. IEEE 36th Annu. Comput. Softw. Appl. Conf. (COMPSAC), Jul. 2012, pp. 387_394.

[4] Madhuri N.Kumbhar, Ms. Reena Kharat, “Privacy Preserving Mining of Association Rules on Horizontally and Vertically Partitioned Data: A Review Paper” 978-1-4673-5116-4/12/$31.00_c 2012 IEEE

[5] Fang Liu, Wee Keong Ng, Wei Zhang, “Encrypted SVM for Outsourced Data Mining” 2015 IEEE 8th International Conference on Cloud Computing

[6] S. Hartley, Over 20 Million Attempts to Hack into Health Database. Auckland, New Zealand: The New Zealand Herald, 2014.

[7] Lichun Li, Rongxing Lu, Senior Member, IEEE, Kim-Kwang Raymond Choo, Senior Member, IEEE, “1847 Privacy-Preserving-Outsourced Association Rule Mining on Vertically Partitioned Databases” IEEE TRANSACTIONS ON INFORMATION FORENSICS AND SECURITY, VOL. 11, NO. 8, AUGUST 2016

[8] P. Gaonjur and C. Bokhoree, ``Risk of insider threats in information technology outsourcing: Can deceptive techniques be applied?" in Proc. Int. Conf. Secur. Manage. (SAM), Las Vegas, NV, USA, Jun. 2006.

[9] G. B. Magklaras and S. M. Furnell, ``The insider misuse threat survey: Investigating IT misuse from legitimate users," in Proc. Austral. Inf. Warfare Secur. Conf., Perth, WA, Australia, 2004, pp. 1_9.

[10]Cloud Security Alliance (CSA). (2010). Top Threats to Cloud Computing,Version 1.0. [Online]. Available: https://cloudsecurityalliance.org/contact.

[11] S. Furnell and A. H. Phyo, ``Considering the problem of insider IT misuse," Austral. J. Inf. Syst., vol. 10, no. 2, pp. 134_138, 2003.

BIOGRAPHY

Rinku Biharidas Kapdi is a Master of Computer Engineering from Noble College of Engineering, Junagadh And she completed her Bachelor of Engineering from L.D. College of Engineering, Ahmedabad.