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 4, April 2017

Insider Collusion Attack on Privacy-Preserving Data Mining System by Non Homomorphic Encryption Methods

Rinku B. Kapdi, Hiral Agravat, Prof. DaxaVekariya

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

Professor at 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 4, April 2017

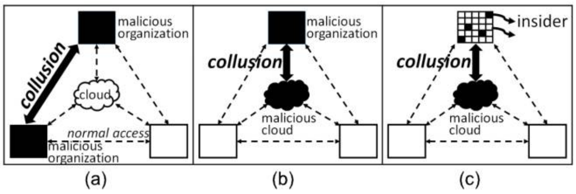

Fig 1.1 Different attack models in DKBDM [1]

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.LITRATURE SURVEY

3.1 LITERATURE SURVEY:

[1] Insider Collusion Attack on Privacy-Preserving Kernel-BasedData Mining Systems [1].
Authors: P.S. Wang
Method: Redusing the number of insider,Expanding the dimension of the data.

[2]A Multilayer Evolutionary Homomorphic Encryption Approach for Privacy Preserving over Big Data [2].
Authors: Amine Rahmani, Abdelmalek Amine, Reda Mohamed Hamou
Method: Homomorphicencryption,Evolutionary cryptography

[3]Encrypted SVM for Outsourced Data Mining [3].
Authors: Fang Liu, Wee Keong Ng, Wei Zhang
Methods: Fully Homomorphic Encryption

[4]Privacy-Preserving-Outsourced Association Rule Mining on Vertically Partitioned Database [4].
Authors: Lichun Li, Rongxing Lu, Senior Member, IEEE, Kim-Kwang Raymond Choo
Method: Substitution Cipher and Frequency Analysis,Cryptography Hash Function, Homomorphic Encryption

[5]Privacy Preserving Mining of Association Rules on Horizontally and Vertically Partitioned Data: A Review Paper[5].

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 4, April 2017

Authors: Lichun Li, Rongxing Lu, Senior Member, IEEE, Kim-Kwang Raymond Choo, Senior Member, IEEE, AnwitamanDatta, and Jun Shao

Method: For VPD:Cryptographytechniques,Scalar product protocol,Two party vector dot product computatio (T-VDC)

For HPD:Privacy Mining of Generic Basic Assosiation Rules, Hussein's Scheme

IV.IMPLEMENTATION ENVIRONMENT

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
for k = 1… n do
{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)}
Let KC1 = [], KC2 = [], l1 = 0, l2 = 0, IndexCand = [], Index = []
for for i = 1…m do //Search for values equal to K1 and K2 in KM
for j = 1…m do
    if KM(i, j) = K1 then
      KC1(l1) = (i; j)
   else if (KMi; j) = K2 then
     KC2(l2) = (i; j)
   end if
  end for
end for
 for u = 1… max(l1) do //Apply Principle 1 & 2 to kernel lines
     for v = 1 … max(l2) do
         if KC1(u)[1] ≠ KC1(v)[1] & KC1(u)[2] = KC1(v)[2] then
            if no element of the array IndexCand(k) =KC1(u)[2] then
         Insert the element KC1(u)[2] into the array IndexCand(k)
            end if
         end if
      end for
   end for
end for
for k = 1… n do //Apply Principle 3 to kernel lines
     if #element of IndexCand(k) = 1 then
         Index(k) D theelementofIndexCand(k)
     end if
end for
for k = 1… n do
    if #element of IndexCand(k) > 1 then
  Delete all elements of IndexCand(k) that has been assigned to the other Index
  Index(k) =a randomly chosen ele-ment from the remaining elements of IndexCand(k)
    end if
end for
```

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 4, April 2017

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.
Now If we'll encrypt the data while making Global Kernel matrix then outsider is not able to deduce the private data.
For that we are going to use Computing Global Gram Matrix from Horizontally Partitioned Data.
4.3 Step of Proposed Algorithm
For that we are going to use Computing Global Gram Matrix from Horizontally Partitioned Data.

Computing Global Gram Matrix from Horizontally Partitioned Data.
Require: Total m data points and n features split in some arbitrary fashion between k parties
Require: Data represented by matrix A; Abc represents the value of the cth feature of the bth point
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}
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

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 4, April 2017

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.

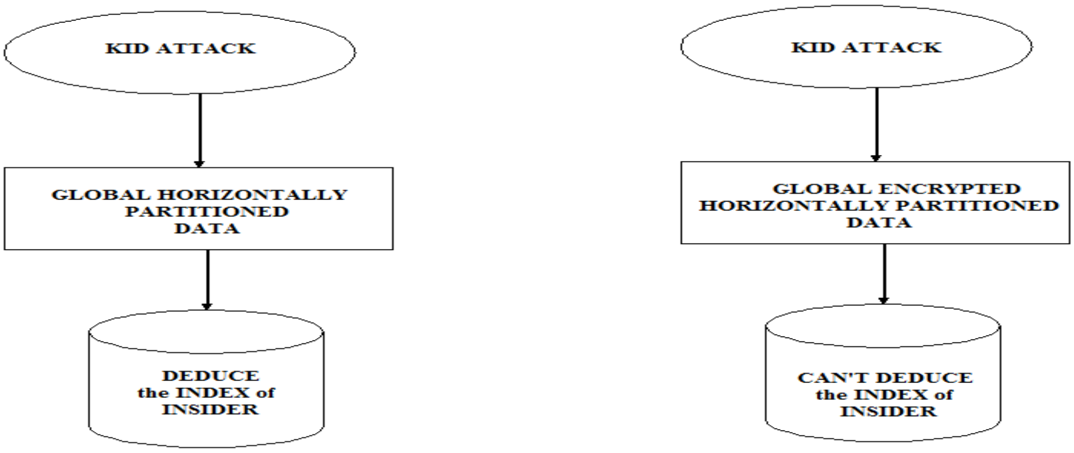

Figure 4.1 Comparision

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.
Existing homomorphic encryption schemes are generally asymmetric. In this paper, we propose a symmetric homomorphic encryption scheme (using only modular additions and multiplications), which is significantly more efficient than asymmetric schemes. The scheme supports many homomorphic additions and limited number of homomorphic multiplications, and comprises the following three algorithms:
Key generation algorithm KeyGen()
(s, q, p) $\leftarrow$KeyGen(λ)
The key generation algorithm KeyGen() is a probabilistic algorithm, which takes a security parameter λ as input and outputs a secret key SK = (s, q) and a public
parameter p. Both p and q are big primes, and p >> q.The bit length of q depends on the security parameter, and s is a random number from Z.
Encryption algorithm E()
E(SK,m, d) = sd(rq + m) mod p
The encryption algorithm E() is a probabilistic algorithm,which takes a secret key SK, a plaintext m $\in$ Fq and a parameter d as inputs. The algorithm outputs a ciphertext
c$\leftarrow$ E(SK,m, d). The parameter d is a small positive integer called ciphertext degree, and we say the ciphertext is a d-degree ciphertext. Let r denote a big random positive integer, and the bit length of r , |r |, satisfies |r| + |q| < |p|. We say r is the random ingredient of c.
The encryption of a plaintext m is denoted by E(m) for short.

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 4, April 2017

```
%% To encrypt the data use homomorphic encryption method
% Take veribles
p=996595253
q=996591151
r=1
s=7
d=3
for i=1:4
    for j=1:8
        uu1(i,j)=mod((r*q+uu(i,j)),p)
        uu1(i,j)=s*s*s*uu1(i,j)
    end
end
```

Figure 4.2 Encryption code to encrypt global gram matrix

```
uu1 =

   1.0e+11 *

    3.4183    2.9647    3.4183    1.9743    2.6984    1.9139    3.4183    3.4183
    2.4256    3.0491    3.0768    1.2666    2.7753    1.4491    3.2529    2.4570
    0.3870    3.4183    1.0773    2.9280    1.6715    0.5012    3.4183    2.9967
    1.7599    2.7209    0.0849    1.3377    1.5148    3.3378    2.6738    2.3053
```

Figure 4.3 Encrypted global gram matrix

After doing the encryption of thee data we'll again find the same value'2048' and this time we can't find the index of given value so, by doing this outsider can't find the index of given insider's data.

Decryption algorithm D()

$D(SK, c, d) = (c \times s^{-d} \bmod p) \bmod q$

The decryption algorithm D() is a deterministic algorithm,which takes a secret key SK, a ciphertext $c \in Fp$ and the ciphertext's degree d as inputs. The algorithm outputs a plaintext $m \leftarrow D(SK, c, d)$. Let s^{-d} denote the multiplicative inverse of sd in the field Fp. The correctness proof of the decryption algorithm is given below.

$D(SK, c, d)$

$= (c \times s^{-d} \bmod p) \bmod q$

$= ((sd(rq + m) \bmod p) \times s^{-d} \bmod p) \bmod q$

$= (rq + m) \bmod q$

$= m$

```
%%To decrypt the encrypted data
for i=1:4
    for j=1:8
        uu2(i,j)=uu1(i,j)/(s*s*s)
        uu2(i,j)=mod(uu2(i,j),p)
        uu2(i,j)=mod(uu2(i,j),q)
    end
end
```

Figure 4.4 Decryption code to decrypt global gram matrix

V.CONCLUSION AND FUTURE WORK

5.1 PRINCIPLE 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.

5.2 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 reverse from that kernel values we can take original data back.as its composed of two data vectors. and its store its value in Kernel Matrix.

VI. 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.