A FRAMEWORK FOR SECURITY IN DATA MINING USING INTELLIGENT AGENTS

Sharath Kumar J. and Maheswari N.
School of Computing Science and Engineering, VIT University Chennai, Vandalur, Kelambakkam Road, Tamil Nadu, India
E-Mail: sharathkumar.j@vit.ac.in

ABSTRACT

Nowadays it is possible to outsource data mining needs of a corporation to a third party. An establishment in the corporate world without much number or expertise in computational resources can outsource their mining needs. But the data as well as the association rules defined over it are the sole property of the company and thus privacy and security needs to be preserved. Also, partitioned databases are capable of simplifying the complexity of massive data as well as improving the overall performance of the system. In this paper we devise a scheme that ensures the privacy of data, incorporating database partitioning to ensure a highly efficient and secure system with new algorithm for privacy preservation. This is combination of L-diversity and P-sensitive technology. Agent Technology is also introduced in the given system. Different agents are used for different task like mining agent, data agent, task agent, user agent etc. and they communicate with each other and work together to provide a heuristic solution.

Keywords: partitioning, database, cipher, i- privacy, intelligent agents, Re-identification, L-diversity, p-sensitive.

INTRODUCTION

With new technologies having a model for IT services based on internet and big data centers, the importance of outsourced data and computing services are increasing day by day. With the data intensive nature of cloud computing, business intelligence as well as knowledge discovery services are those expected to be among the services externalized on the cloud. While achieving sophisticated analysis on bulk of business data is advantageous, the security issues cannot be overlooked. The main issue here is that the server is having access to private data of the owner and can gain knowledge from it. In other words the corporate privacy is at risk here. The existing privacy preserving techniques for personal privacy are modifying data using encryption, perturbation and generalization techniques, data mining algorithms with privacy techniques incorporated in it, randomized response techniques, heuristics based techniques. In this paper we device a technique in which each transformed item t is made indistinguishable from other items.

Main objective is to provide data privacy. In which l-diversity and p-sensitive these term are used [13]. Another term is agents. Agents are self learner that's why they are used to do this task so the heuristic solution given out [2]. Agent is defining as pieces of code that is situated in some environment and that is capable of autonomous action in this environment in order to meet its design objective. Agent having multiple properties like robust means recovers from failure, social this term related interact with other agent then reactive, which is define responds to change in its environment, autonomous means independent, we can say not controlled externally etc [4]. This is achieved by first encrypting the data and then grouping it.

Addition of noisy transactions is the method used to defend against frequency based attack models. We are integrating privacy preserving transparency techniques along with partitioned databases for increased privacy as well as efficient data retrieval.

In database partitioning, horizontal and vertical partitioning of database forms an integral part of the database design as it transforms the database into a more manageable and highly efficient system.

Horizontal partitioning: This partitioning divides the table, indexes or views horizontally into multiple smaller set of rows with lesser number than the original. Two widely used types of horizontal partitioning are hash and range partitioning. Horizontal partitioning is widely used as it makes database server management easier. Operations performed by DBA like backup and restore becomes easier if the partitions are aligned i.e. both the table and corresponding index is partitioned identically [9].

Vertical partitioning: This partitioning divides the table vertically into multiple tables each with lesser number of columns. There are mainly two types of vertical partitioning namely one through normalization where repeating attributes are eliminated and the other one is through row splitting by which original table is divided into tables which contain fewer columns [8].

Hybrid partitioning: In this data is partitioned first horizontally and then vertically or vice versa.

When we talk about main objective that is data privacy for this solution is l-diversity and p-sensitive used. Data is sensitive or secrete type of information. This two word having different meaning, secrete means data like ATM pin then password etc where as sensitive like any disease that particular person having like HIV or any health issue. So both types of data should protect from unauthorised person. Now a day's mainly focus on K-anonymity technique. But K-anonymity facing diversity and background knowledge problem [15]. To overcome this problem we combined l-diversity [13] and p-sensitive [14] technique. In l-diversity divide data into sensitive and no sensitive part. Sensitive data is key-attribute which gives the idea about particular user and non sensitive is other than sensitive attribute. Represent this attribute with verity of different arrangement [14]. Then add p-sensitive technique which is nothing but masking of sensitive data [13]. There are different ways to mask or modify data generalization, suppression, noise adding, substitution, and data swapping [13]. We mask data and also provide sensitive level of attribute like high, low, or medium. This new algorithm overcomes drawback of k-anonymity and other existing technique. This entire task can be done by intelligent agent [2].

RELATED WORK

Privacy preservation

Privacy preserved data mining is a widely researched topic since a lot of private data is being collected by a collector for the purpose of mining the data. The collector essentially does not protect the privacy of the data collected. So to preserve data privacy random perturbation techniques can be used. Many techniques have been developed so far to randomly perturb data so as to preserve data privacy. The perturbation is done in such a way that the mined patterns are almost identical to the patterns that could be mined from the original data.

The problem addressed in this paper is outsourcing of data mining. The main distinction from other data mining problems is that here both the data required for mining as well as the pattern that is mined should remain private. Also this paper looks into privacy preservation at the user side. Authentication and privacy preserving techniques are used to ensure high security of the underlying data.

Partitioning

Vertically partitioning a table splits it into more sub-tables each containing a subset of the columns. Vertical partitioning substantially reduces the amount of data to be scanned on query since many of these queries access only a small subset of columns present in the table. Sub-tables can be partitioned such that each of them contains distinct set of columns in them except for the key attributes. These key attributes are needed while reconstructing the original table from the sub-tables.

Horizontal partitioning can be done on tables or sub-tables, a non-clustered index or a view. This partitioning can be specified using a partitioned method that maps a given row in database to a partition number. All rows that are having the same partition number will be stored in the same partition. This single node horizontal partitioning can increase the performance as well as the manageability.

Two other kinds of partitioning methods are hash and range partition. In hash partitioning the partition number is generated using a hash function on the columns in a set of columns that are specified. It is defined by a row (C, n) where C is the set of attribute types and n is the number of partitions. For example, let S be the table to be partitioned with attribute types given as (c1 int, c2 float, c3 int, c4 date). The hash partition defined on it H=({int, float}, 10) will yield 10 partitions containing values after hash functions are applied first two columns c1 and c2 in each row of S [11].

In range partitioning, the partitions are formed on the basis of different range of values in an attribute. It is defined by a row (c, V) where c denotes the attribute type and V is a sorted sequence of values in the domain of c. For example, a range partitioning on S, R= (int, <10, 30>) when applied on column c1 divides the table into 3 partitions. The first partition will contain all rows with c1 column value less than 10, second partition contains value between 10 and 30 and the third partition contains value ranging above 30. Range partitioning is defined on a single column rather than a set of columns [12]. A hybrid partition can be formed by first range partitioning a table and then hash partitioning each range obtained.

When we are say data, it is mean by the information in structured format. Structure format means in the form of tables. In table rows and column are given. We can say tuple and attributes. In this attributes are unique [9]. We assumed that no two tuple pertain or contain same user information we can just create a link between private information and external information. Where external information refer as quasi identifier which contain same meaning data but not original data. When release of data is done it’s simply like to external data and release so it offers for privacy protection [12].

In our system we are using different agents for particular task. Like user agent, mining agent, task agent [2]. They are communicating with each other and then work together. Due to the adding agent in existing system lots of advantages are comes out over the existing system.

Work on k-anonymity and its drawback

K-anonymity is widely discussed because of its simplicity but k-anonymity not give guarantee of data privacy. In this session we will see drawbacks of K-anonymity and the existing work done on l-diversity and p-sensitive.

Table-1. Original table.

	Non-Sensitive			Sensitive
	Zip code	**Age**	**Nationality**	**condition**
1	120**	<30	*	Heart disease
2	120**	<30	*	Heart disease
3	120**	<30	*	Viral infection
4	130**	>=40	*	Cancer
5	130**	>=40	*	Heart disease
6	130**	>=40	*	Viral infection
7	120**	3*	*	Cancer
8	120**	3*	*	Cancer
9	120**	3*	*	Cancer

Table-2. 4-anonymus table.

Zip code	Age	Nationality	condition
12054	21	Russian	Heart disease
12068	23	Indian	Heart disease
12068	24	Japanese	Viral infection
13053	50	Russian	Cancer
13053	55	Japanese	Heart disease
13053	47	Indian	Viral infection
12054	37	Russian	Cancer
12068	36	Indian	Cancer
12068	35	Japanese	Cancer

Drawbacks of K-anonymity

In this session we discuss two simple attacks. Homogeneity attack and background knowledge attack [15].

Homogeneity attack

Let consider simple example, shriya and Krishna are neighbours. One day Krishna falls ill and he admitted in hospital. To seen ambulance shriya discover that Krishna is ill. And from following thing shriya come up with what disease Krishna is suffering from. Shriya discover from hospital record (Table-2) she knows one of them record having Krishna's information. She is Krishna neighbour so she knows his age is 37, and he is Indian male lives in 12068 zip codes. She knows Krishna's record between 7, 8 or 9. Every patient having some health condition so she conclude Krishna having cancer. It is nothing but lack of diversity in sensitive attribute. Diversity is defined as the repetition of multiple time occurrence of key attribute so it can be easy to find out particular information. This is homogeneity attacks.

Background knowledge attack

Let continue previous example. Shriya having another friend abhay who is admitted in same hospital where Krishna admitted. Both records are saved in Table-2. Shriya knows his age is 24 and zip code is 12068. Based on this information shriya come up with some records like 1, 2 and 3. Without knowing extra information related abhay, shriya cannot find out abhay disease. But she knows background knowledge related abhay so she concluded abhay having viral infection .so k-anonymity not protect against the background knowledge attack.

Drawback of l-diversity

In this we discuss related single l-diversity technique. In L-diversity, divide the data in sensitive and non-sensitive part is done [13]. Where the sensitive part is not released to anyone. It is send only etherised person. It is not provide sufficient privacy. If sensitive or key attribute is repeatedly occurs again and again then it becomes easily find out particular person record or information. The grouping of sensitive attribute with non sensitive attributes [13]. Let's see one example,

Table-3. Before l-diversity.

	Wine	Apple	But-ter	Ice-cream	Pregnancy test	HIV test
Abhay	X		X			
Krish	X		X			X
Nupur		X		X	X	
Dhanu		X	X			
millind	X		X	X		

In given example if Dhanu and Nupur both used pregnancy test then it become easy to find out attribute. It happens because sensitive attribute repeatedly occurs so it becomes drawback as shown Table-4.

To overcome both technique drawbacks we proposed new technology which is combination of l-diversity and p-sensitive. Let's see in next section.

Table-4. After l-diversity.

	Wine	Butter	Ice-cream	Apple
Abhay	X	X		
Krish	X	X		
millind	X	X	X	
Dhanu		X		X
Nupur			X	X

PROPOSED APPROACH

In the proposed system, the database is divided into smaller segments called partitions. The privacy of outsourced data is preserved using i-privacy scheme, to achieve this different agents are used.

Architecture

The architecture consists of both server and client side. The server here is the data mining server that mines the data send over to it. The user access of database is depicted in the second architecture. The steps taken for preserving privacy of database are encryption, grouping, adding some noisy data and finally shuffling or rearranging the indexes.

Architecture for data mining

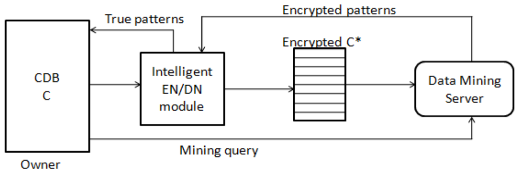


Figure-1. Privacy preserved data mining.

The system consists of an intelligent encryption and decryption module that transforms the items in the database, C, according to the scheme to an encrypted database C*. The module first encrypts some items in the database using substitution cipher method. In this method, a unit of the database item is replaced with another item. The transformation mapping that is made for each item is stored in a file so as to decrypt the returned result. This module after encryption, groups the items in the database depending on the total number of items in it. Then irrelevant items are detected and masked by this module. The indexes on each group are then jumbled or shuffled to obtain the transformed database. This encrypted database is send across to the mining server. On receiving a query, the server mines the encrypted data and returns back the result. The EN/DN module on receiving it decrypts and recovers the true patterns.

Architecture for user access

The central database is divided into smaller partitions using hybrid partitioning. The partitioned items are encrypted first using substitution method. The cipher texts are grouped based on number of records present. The indexes of these partitions are shuffled in a random order.

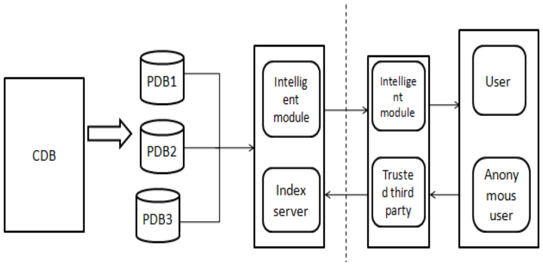


Figure-2. Authenticated user access.

The order of the shuffling is retained by the index server. When an authenticated user tries to access the database the intelligent module will generate a key based on the users request and send it to the users mobile. This code is the key to access the database for the user. The code generated by the intelligent module will consist of pointers to information like the access rights of the user; the index entry related the query raised the partition identifier etc. Aided by this code the intelligent agent retrieves the result and gets back to the user.

A. Partitioning

The central database C is divided into smaller partitions PDB_1, PDB_2… PDB_n. This can be done by horizontal, vertical or hybrid partitioning. Hybrid partitioning is done by first horizontally partitioning the database and then again vertically dividing each partition.

B. Encryption

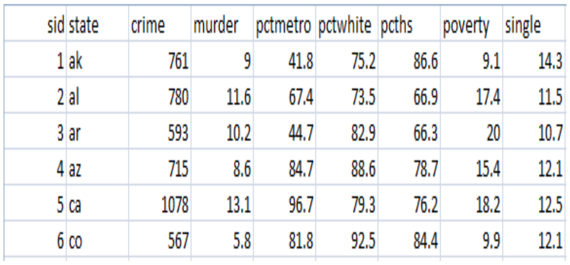
This paper introduces a new encryption scheme, i-privacy consisting of mainly four steps. In this scheme the CDB C when encrypted transforms to C*.The following are the steps:

a) Plain text is transformed to cipher text using substitution method.
b) When database is encrypted for mining purpose, some of the plain texts are transformed to cipher text using 1-1 substitution method.
c) All the database items are encrypted using substitution as it is stored in partitions for user access.
d) Items are grouped based on the total number of records present.
e) Duplicate records are added are to increase the noise in the database.
f) Indexed files are jumbled up in a random order.

The unwanted fields are masked by the intelligent agent before sending the encrypted database C* to the data

mining server. Also most sensitive data like credit card credentials or other private information will be eliminated before sending the database for data mining purpose. When it comes to user access, irrelevant data are masked based upon the query posed by the user.
For example, consider a sample dataset crime.

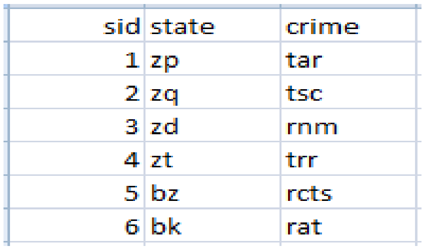
Table-5. Crime dataset.



sid	state	crime	murder	pctmetro	pctwhite	pcths	poverty	single
1	ak	761	9	41.8	75.2	86.6	9.1	14.3
2	al	780	11.6	67.4	73.5	66.9	17.4	11.5
3	ar	593	10.2	44.7	82.9	66.3	20	10.7
4	az	715	8.6	84.7	88.6	78.7	15.4	12.1
5	ca	1078	13.1	96.7	79.3	76.2	18.2	12.5
6	co	567	5.8	81.8	92.5	84.4	9.9	12.1

Applying cipher substitution on first two rows,
Plain text: a b c d e f g h i j k l m n o p q r s t u v w x y z
Cipher text: z e b r a y w x u p q t m j k l o d n c f i g s h t
Plain digit: 0 1 2 3 4 5 6 7 8 9
Cipher text: c r i m e r a t s n

Table-6. Encrypted dataset.



sid	state	crime
1	zp	tar
2	zq	tsc
3	zd	rnm
4	zt	trr
5	bz	rcts
6	bk	rat

C. Decryption

The mined result send back from the server is decrypted by the intelligent decryption module. First the noise added will be removed from the mined result. Then the masked data items are restored. The jumbled indexes are ordered with the aid of the file that stores the information about the cipher text, jumbling and the noise table. The cipher text item is converted back into plain text reversing the substitution to obtain the true patterns that are mined.

D. Grouping items

Items are grouped by considering the odd and evenly placed records. The number of items in each group is selected on the basis of number of records present. Let c_1,c_2… c_n be the set of cipher items then the grouping method groups as { c_1,c_3… c_k }, { c_2,c_4… c_{2k} } and so on. If the last group has elements less than k then it are merged with the previous group.

E. Constructing duplicate records

Noise table is formed specifying the noise N(c), corresponding to each cipher item c. The duplicate records are generated by first dropping all rows with noise value 0. The remaining is sorted in descending order of noise. Let c'_1, c'_2… c'_k be the sorted order obtained. Then the duplicate records generated are:

- $N(c'_1)$-$N(c'_2)$ instances of transaction $\{c'_1\}$.
- $N(c'_2)$-$N(c'_3)$ instances of transaction $\{c'_1, c'_2\}$.
- …..
- ….
- $N(c'_{m-1})$-$N(c'_m)$ instances of transaction $\{c'_1, c'_2.. c'_{m-1}\}$.
- $N(c'_m)$ instances of transaction $\{c'_1, c'_2… c'_m\}$.

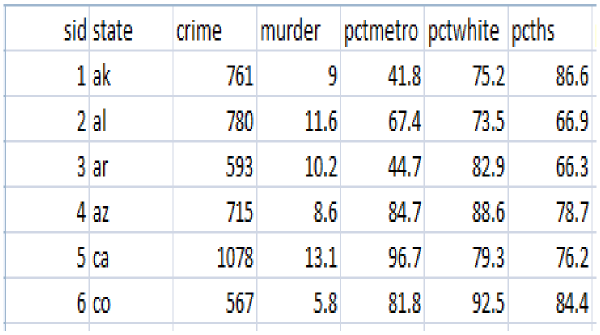
These are added to the database and then updated in the file managed by the intelligent module so as to retrace it as and when needed.

F. Shuffling Index and data masking

The index items on each partition are shuffled within a group in a random order. The ordering of the shuffling is stored by the index server. The intelligent module interprets the correct order and index when a authenticated user tries to access the database.

Data masking is done by the intelligent module. Based on the query given to the data mining server as well as on the query raised by the user some of the data which is sensitive or which are irrelevant on the context of the query are masked or removed from the dataset. These are stored in a file so that it can be restored when the mining server return the result or the user with write access updates and sends back the data.

Table-7. Masked response.



sid	state	crime	murder	pctmetro	pctwhite	pcths
1	ak	761	9	41.8	75.2	86.6
2	al	780	11.6	67.4	73.5	66.9
3	ar	593	10.2	44.7	82.9	66.3
4	az	715	8.6	84.7	88.6	78.7
5	ca	1078	13.1	96.7	79.3	76.2
6	co	567	5.8	81.8	92.5	84.4

For example consider the crime dataset; if a user, say magistrate, is accessing the database then the fields relevant to him will be the rate of crimes in the city. The field single, poverty will not be of much importance. The intelligent module masks this data and customizes the result according to the user trying to access the database.

L-diversity and P-sensitive

The proposed system having different advantages over the drawback of existing system.

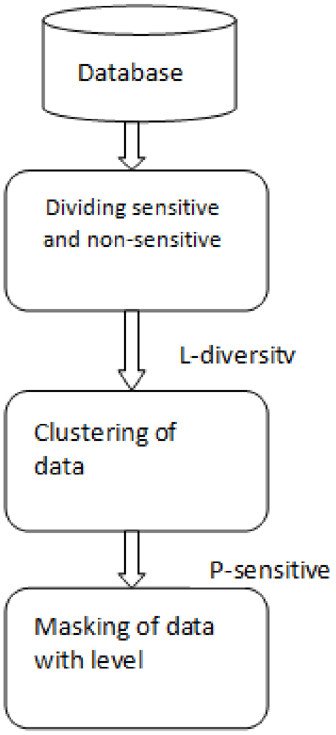

Database
Dividing sensitive and non-sensitive
L-diversity
Clustering of data
P-sensitive
Masking of data with level

Figure-3. Steps for algorithm.

L-diversity

The given Table is divide into two main part, one is sensitive attribute which contain information related disease, HIV, pregnancy test etc. The another part is non-sensitive part which contain attribute which are not sensitive like ice-cream, wine, butter etc shown in Table-4. Then clustering is done on attribute. Depends on nearest-Neighbour [6] and then rearranging the table as in Table-4 is done. This is first step of main algorithm.

L-diversity provides privacy even when information publisher does not know what kind of knowledge is possessed [2]. Value of sensitive attribute represented in each group. Let see,

Multiple sensitive attribute

if multiple sensitive attributes ate present like
S= {S1, S2.................., Sn}
And non-sensitive attribute {A1, A2...............An}
Then find out single and unique sensitive attribute from multiple and add remaining attribute.
A= {A1, A2........, An, S2, S3...............Sn}

C= {R1, R2......., Rn}

Attribute are A1, A2, A3...... Am
Where C is subset of some larger population of each tuple.

R [Ai] = Value of attribute for Ai for tuple R.

Quasi-identifiers- Set of non-Sensitive attributes of Table-6.

Let see algorithm for l-diversity

Input- Clustering C={R1,R2,........,Rt} of record in table
Dp: final diversity parameter where l>=1
Output- Clustering of data respective l-diversity.

Algorithm

a) Compute the Div (Ri) for all existing Ri € C.
b) Let Rm is clustering with minimal diversity in C.
c) If Div (Rm)>=l then
d) Output C and stop
e) End if
f) Compute the cost (Rt, Rm)for all Rt € C\{Rm}
g) Find cluster Rt € C\{tm} for minimal cost (Ri, Rm)
h) Remove Rt and Rm from C add to C cluster Rt U Rm.
i) Go to second step

This algorithm for l-diversity .in this way combination of two attribute is done depends on Quasi Identifier [5]. Next step is P-sensitive.

P-sensitive

The next step is masking of sensitive data. Masking is defined as the modification in sensitive attribute so it's difficult to find out original information. There are several ways for masking like generalization, suppression, substitution, adding noise etc. We have to use method in such way that the how much data change. It should not be too much so result is information lost and it should not be less which is causes fir risk of discloses. We are using generalization for masking in our system. Because it's simple to implement. Let see algorithm-
Input: The clustered data from l-diversity
Output: from of new cluster data with masking and level of sensitive attribute

Algorithm

a) If check wether masking required or not. If no goto last step.
Occurrence of sensitive only ones
b) Else mask the sensitive attribute <= {generating new table over old one MT->NT}
c) Check wether proper generalized or not.
d) If yes then stop
e) Else continue up to get well generalized table with respective sensitive attribute. Add the level of sensitive attribute (low/medium/high)
f) Stop

In this way the p-sensitive is done. And the final data we can get. In such way our algorithm works.

Agents

When we define agents it's not having any perfect definition. Its simply depends on that particular environment. When we related software then we can say that, an agent is a computer system that is situated in some environment, and that is capable of autonomous action in this environment in other to meet its design objective. In simple word it is software depends on situation its react. There are different types of agent .in our system we are using intelligent agents, which is taking decision on behalf of user on the basic of particular situation. Intelligent having different properties. Autonomous because they are independent and not control by other. They taking decision them self. Self learner, on the basic of history they are taking decision and react. Reactive, respond to change in its environment with respective time. They are flexible, robust, and social [2]. There are different advantages of agent. We are using multiagents so, its avoid single point of failure in system. Multiple agents work together for same purpose so fast execution and easily completion of task is done. Figure-4 shows Agents and Environment.

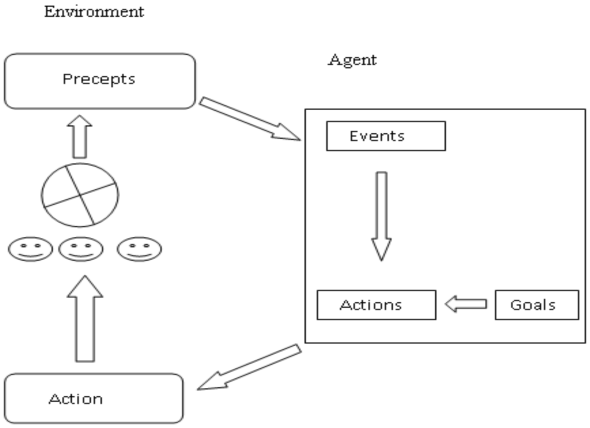


Figure-4. Agent and Environments.

There are different platform is used to perform agents service like ABLE (Agent Building Learning Environments) which is support java based but the disadvantages is this platform is unmainted. Because after 2005 in which it is introduce no updating is done so far. Another is FAMOJA (Framework for Agent- based Modeling with Java), JANUS, JIAC (Java-based Intelligent Agent Computer framework), JADE (Java agent Development Environment) [5]. As shown in Figure depends on the precepts agent take an action on the basis of what should be output required.

JADE

In our system we are using JADE platform. Which is support java language? JADE is open source software. Which is also support php? It is a middleware which develop for multi-agent application.
Let see how to create agent step by step.

1. Deciding on the agent type used in the application.
   1.1 Group functionalities into agent considering alternatives.
   1.2 Review coupling using agent acquaintance diagram and decide on a preferred grouping.
   1.3 Develop agent descriptors.
2. Describe the interaction between agent using interaction diagram and interaction protocol.
   2.1 Develop interaction dia.from scenarios.
   2.2 Generalize interaction dia. to interaction protocol.
   2.3 Develop protocol and message descriptors.
3. Design the overall system structure
   3.1 Identify the boundaries of the agent system and the interaction with often sub-system.
   3.2 Describe the precepts and action and the relationship between these and relevant agents.
   3.3 Define all shared data, both external persistent and internal shared data.
   3.4 Develop the system overview diagram

In this way agent are created.

METHODOLOGY

Administrator

Administrator main work is maintained all user detail and their code in their database. Also another task done by administrator is user allow the authenticate user to see original database and unauthorized user to suppress database and it is based on SMS Based authentication. In this simply six digits code is maintain.

So, meanwhile automatically both the confidentiality of the data and privacy of the user maintain by the administrator.

Registration

Registration is used for the trusted third party purpose so trusted third party registered with the database server. So, user can able to access the data, update the data and also retrieving the data from the database through the trusted third party only.

Surveyor

Surveyor is nothing but user who is going to change or update the database. Surveyor send request to the trusted third party. And the SMS based authentication is done then trusted third party communicate with administrator. And through trusted third party updating in data done.

User

User sends the request to the trusted third party then through administrator TTP provide SMS based authentication to user.

Description of architecture

Partitioning

The proposed architecture database divide into small part. We can say data partitioning is done. In our system we are using hybrid portioning [7]. This is defining as the first divide horizontal then the data again divide into vertically. In simple word perform horizontal portioning then perform vertical portioning. It is nothing but horizontal partitioning.

l-diversity plus P-sensitive

After portioning perform our main logic which is nothing but apply l-diversity and P-sensitive algorithm which is discussed in previous section. So it provides more security and data privacy. Working of logic is already discussed.

Intelligent agent

To do all this task we are using Intelligent agent [4]. The behavior and work of different agent is discussed in previous paper. We are using different agent for different task. Like mining agent, user agent, data agent etc as shown in figure, for this we are using JADE platform

User Login and multiprocessing module

In this login is given. Which is define as what kind of user is their so on the basis of user type the data or information disclose. There are multiple users access database same time. So multiprocessing is done.

The architecture of given system shown in Figure-5 let see how it works.

a) When user send a request to access database to TTP (Trusted third party) then on the basis of SMS based authentication it check authorisation.
b) If it is authorised then able to update the data base, if user wants to access particular data then send request through TTP.
c) Different agents are used mined the database and give solution for particular query .where agents are communicate with each other .the task is divided into different agents. The agents are data agent, mining agent, task agent, user agent etc. The all task is done by agents.
d) And the above define algorithm (l-diversity and p-sensitive) apply on database and provide generalized data to user.

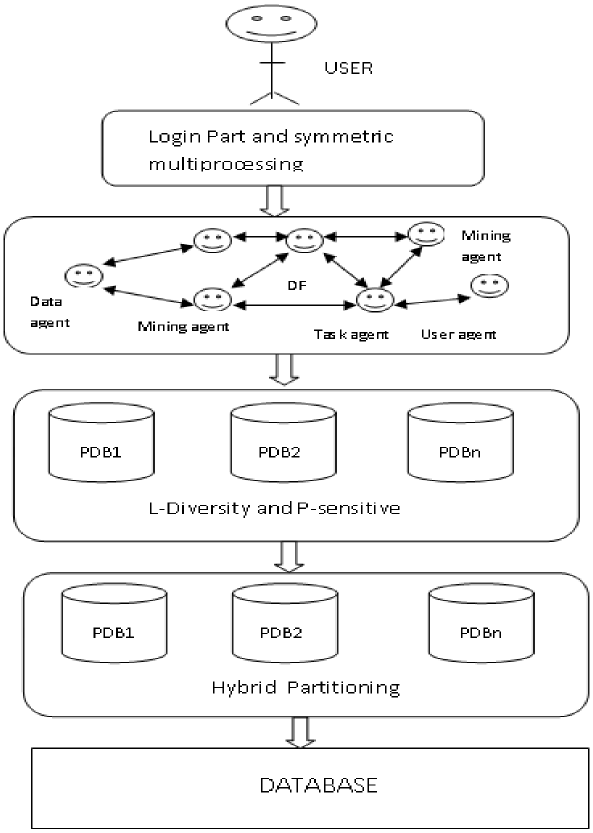


Figure-5. Architecture.

In this way the working of given system is worked. Let see how to create simple agent for display information

Algorithm for communication between different agents

a) When particular user send request then this request send to user agent.
b) On the basis of request user create a task agent.
c) Particular task agent sends the request to data mining agent.
d) Then data mining agent communicate with data agent which contain metadata related request.
e) On the basis of mining type specific mining is done.
f) And the final result as a response sends to task agent.
g) Through task agent it sends to user agent.
h) And finally user got reply to their query.

Description

There are different agents are used together do one specific task. So different agent doing different work. Let see work of each agent.

a. User agent

User agent is bridge between user and system. It allow user to communicate with system through send the request and response. User agent create task agent with respective the nature of task.

b. Task agent

Task agents are temporary created. When given request comes out then the particular task agent created. It is used to address specific request which come from user agent. Task agent is connected to data mining agent.

c. Data mining agent

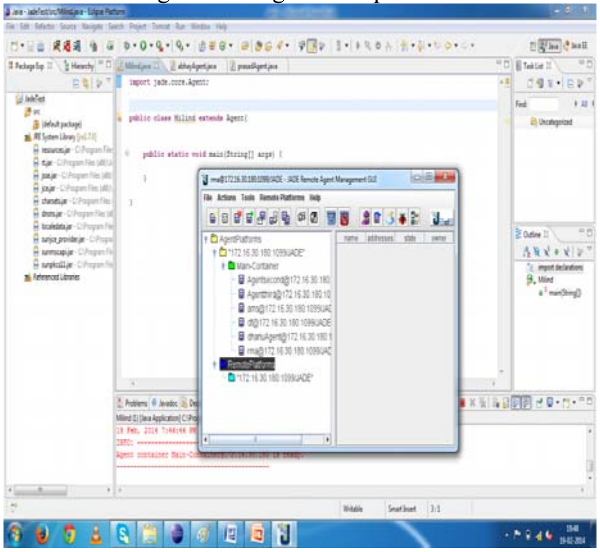
On the basis of mining specific mining is done. Data mining and data agent are work together. And the result is sending to task agent.

d. Data agent

Data agent holds the metadata. Mining and data agent work together and find out solution for given task. And send as a reply to task agent. In this way communication between agents is done.

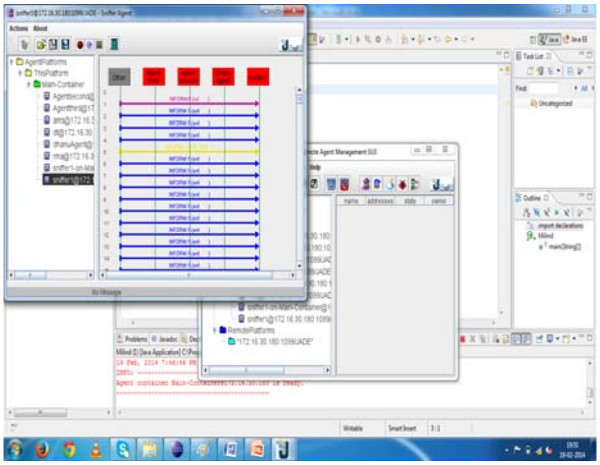
Snapshot

Agent running on Jade platform

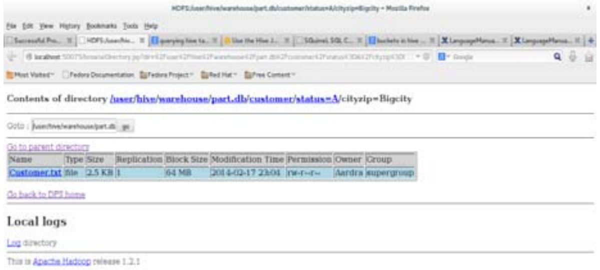


```
import jade.core.Agent;

public class Milind extends Agent{

    public static void main(String[] args) {

    }

}
```

Sniffing in different agent on same server



Partitioning folder



HDFS:/user/hive/warehouse/part.db/customer/status=A/cityzip=Bigcity - Mozilla Firefox
File Edit View History Bookmarks Tools Help
Most Visited
Fedora Documentation
Fedora Project
Red Hat
Free Content

Contents of directory /user/hive/warehouse/part.db/customer/status=A/cityzip=Bigcity

Goto : /user/hive/warehouse/part.db go

Go to parent directory

Name	Type	Size	Replication	Block Size	Modification Time	Permission	Owner	Group
Customer.txt	file	2.5 KB	1	64 MB	2014-02-17 23:04	rw-r--r--	Aardra	supergroup

Go back to DFS home

Local logs

Log directory

This is Apache Hadoop release 1.2.1

CONCLUSIONS

In this paper we propose a system that preserves privacy of data in outsourced data mining. Also, data partitioning is incorporated that improves the system efficiency. Though we are using Agent technology it's improve system performance. Database is made secure with authenticated access and also data anonymization by random shuffling.

This system provides data privacy, confidentiality of the data, secure data access, data integrity and access control. We combined two technique l-diversity and p-sensitive which provide the more data privacy. And also over come drawbacks of k-anonymity and only l-diversity successfully Integrating partitioning into the existing system further improves the manageability as well as security.

FUTURE WORK

As future work there is more scope in proposed system. On Handling multi sensitive attribute its need to work further work. We proposed simple way to handle it but still future work on that. Then the privacy and utility are connected to each other. This paper mainly focused on privacy .utility is not well understood.

REFERENCES

[1] 2002. Achieving k-anonymity privacy protection using generalization and suppression International Journal on Uncertainty, Fuzziness and Knowledge-based Systems. 10(5): 571-588.

[2] EMADS: An Extendible Multi-Agent Data Miner Kamal Ali Albashiri, Frans Coenen, and Paul Leng Department of Computer Science, The University of Liverpool, Ashton Building, Ashto Street, Liverpool L69 3BX, United Kingdom.

[3] User-side Personalization Considering Privacy Preserving in Cloud Systems Leila Sharifi Maryam Heidari Beisafar Department of Computer Engineering and Information Technology. 2013 27thInternational Conference on Advanced Information Networking and Applications Workshops.

[4] 2006. Development multiagent system with jade Nikolaos Spanoudakis, Pavlos Moraitis Published in

In: Applied Artificial Intelligence Journal, Taylor and Francis. 20(24): 251-273.

[5] 2012. A generalized Framework of Privacy Preservation in Distributed Data mining for Unstructured Data Environment. IJCSI International Journal of Computer Science Issues. 9(1, 2) ISSN (Online):1694-0814.

[6] 2013. Web Service Based Model for Inter-agent Communication in Multi-Agent Systems: A Case Study. International Journal of Computer Information Systems and Industrial Management Applications. ISSN 2150-7988, 5: 642-651.

[7] 2013. Cryptographic techniques for privacy preserving Data mining Agent Based Network Sniffer Detection. International Journal of Scientific and Research Publications. 3(4), ISSN 2250-3153.

[8] Fosca Giannotti, Laks V. S. Lakshmanan, Anna Monreale, Dino Pedreschi, and Hui (Wendy) Wang. 2013. Privacy-Preserving mining of association rules from outsourced transaction databases. IEEE systems journal. 7(3).

[9] B. Raghuraml, Jayadev Gyani. 2012. Privacy Preserving Associative Classification on Vertically Partitioned Databases. IEEE International Conference on Advanced Communication Control and Computing Technologies (ICACCCT).

[10] Jayanti Dansana, Raghvendra Kumar, Debadutta Dey. 2013. Privacy preservation in horizontally partitioned databases using randomized response technique. Proceedings of 2013 IEEE Conference on Information and Communication Technologies (ICT).

[11] F. Giannotti, L. V. Lakshmanan, A. Monreale, D. Pedreschi, and H. Wang. 2010. Privacy preserving data mining from outsourced databases. In: Proc. SPCC2010 Conjunction with CPDP. pp. 411-426.

[12] Sanjay agarwal, Vivek Narasayya, Beverly yang. 2004. Integrating vertical and horizontal partitioning into automated physical database design,” Microsoft research, SIGMOD 2004.

[13] ℓ-Diversity: Privacy beyond k-Anonymity, Department of Computer Science, Cornell University.

[14] Privacy Protection: p-Sensitive k-Anonymity Property Traian Marius Truta, Bindu Vinay Department of Computer Science, Northern Kentucky University.

[15] Achieving k-Anonymity Privacy Protection using Agents. International Journal of Computer Applications. 90(15): 25-30, March 2014. Published by Foundation of Computer Science, New York, USA. BibTeX.

[16] DATA PRIVACY on E-HEALTH CARE SYSTEM Abdullah Abdul rah man AlShwaier, Dr, Ahmed Zayed Emam Department of Information Systems King Saud University College of Computer and Information Sciences.