Achieving k-Anonymity Privacy Protection using Agents

Dhanashree Nanasaheb Kharde
School of Computing Science and Engineering
VIT University Chennai, India

Aadra Menon
School of Computing Science and Engineering
VIT University Chennai, India

Sharath Kumar J
School of Computing Science and Engineering
VIT University Chennai, India

ABSTRACT
Nowadays, in given network or system we have to share personal specific data. If unauthorized person read our personal data then it is simple break the privacy policy. and any one can read other personal data and its easy to figure out the particular person so indirectly violation in confidentiality also involves. so this paper on the achieving k-anonymity privacy protection using agents. where k-anonymity means given data is more general form so that the identification of particular user not so far easy. Agent technology is newly introduced with existing system. Different agents are used for different task like mining agent, data agent, task agent, user agent etc. and they are communicate with each other and work together. Its provide heuristic solution. The proposed system having several important properties so it can overcome drawback of existing system. like secure computation, confidentiality access, fast access, privacy preservation provide.

Keywords
Data anonymity, Intelligent agents, Data privacy confidentiality, Re-identification

1. INTRODUCTION
Now a days in network or given system we share personal data. And for effective result we have to share data. like cloud computing has many benefits so that’s why now days cloud is mostly used. But main objective is to provide data privacy. Using k-anonymity it becomes an easy approach. In which generalization and suppression these term are used[1]. Whereas 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]. Let’s see what is actual generalization and suppression. Generalization is defined as replace value or data with equivalent data. Which having same meaning but less importance and suppression is define as the value or particular person specification data never be release[1]. Data is presented in distributed format. As a distributed database so linking between database is their. So it becomes easy to identify that particular person[5].

Along with existing approach introduce new term which is agent technology. This is used to overcome drawback of existing system. Agent is define 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, its like not controlled externally etc[4]. Main objective that is data privacy for this solution is k-anonymity. a release of data in such way that if each release record has at least (k-1) other records also visible in the release value, whose value are indistinct over a special set of field called New-identifier[1]. In system maintain one data set called New-identifier.in which the data is maintain which is likely to appear in other know data set which is actual data.

This paper is formal presentation of k-anonymity using generalization and suppression using agent features in that. In this the data which is going to released is properly prepared data. Another is the information is used is truthful so this data used in fraud detection, in medical field for KDD [1]. For generalization and suppression used MinGen algorithm used which is used in existing system[1]. In this paper generalizing algorithm used which is slightly different than existing one. It gives maximum generalization with minimal distortion.

2. RELATED WORK
The basic concept related k-anonymity and its related terms like quasi-Identifier, generalization, suppression is very simple in nature. When particular term is their that term related meaning is clearly specifies[1]. In this section introduce each term as well as basics of agent technology also.

First and basic term is 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. Other name is tuple and attributes. in this attributes are unique[3]. Where no two tuple certain or contain same user information. Then 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 its simplylike to external data and release .so it offers for privacy protection [1].

In this paper different agents are used for particular task. like user agent, mining agent, task agent[2] they are communicate 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.

3. PROPOSED APPROACH
The proposed system provide different features. as already discuss like anonymous authentication, privacy preservation, confidential access, anonymity maintenance, data management, secure computation, data integrity and another one is access control[7]. In the given system user and surveyor register with the trusted third party. And trusted third party register with admin database[5]. Over the give user personalize database produce table that are k-anonyms with
minimal distortion. Let see one by one-actually what happen in that.

3.1 Generalization including suppression
As already discuss generalization is nothing but value is replace by less specific value. The original value converted into more general format. Lets take one example. VIT_Mtech_Computersci,VIT_Mtech_Cloud so its general form is VIT_Mtech*, let covert into generalized format using stripping of right most word. When VIT_Mtech_Compsci so its an ground ZIP domain. To achieve k-anonymity means converting ZIP codes into less informative. Where going to generalize last word or digit then its simply means converting Z0-Z1.10-09-1992 cover into 10-09-199* so, it is nothing but generalization of values.

\[ Z_3 = \{***\} \]
\[ Z_2 = \{\text{VIT}_*_*\} \]
\[ Z_1 = \{\text{VIT}_\text{Mtech}_*\} \]
\[ Z_0 = \{\text{VIT}_\text{Mtech}_\text{Mcs},\text{VIT}_\text{Mtech}_\text{Cloud},\text{VIT}_\text{Mtech}_\text{Bigdata}\} \]

\[ \text{Fig 1: Example for generalization including suppression} \]

When generalization of value is done, the new maximal value is suppressed value. And the height is increase by one. The given fig. 1 is example so, in this way generation of the data and maintain the New identifier table for each is done. when data releasing is done this data is used and after generalization and suppression nobody can access original database.

Lets take one example-

<table>
<thead>
<tr>
<th>Name</th>
<th>Birth</th>
<th>Gender</th>
<th>Zip</th>
<th>Branch</th>
</tr>
</thead>
<tbody>
<tr>
<td>Abhay</td>
<td>1991</td>
<td>M</td>
<td>02138</td>
<td>Cloud</td>
</tr>
<tr>
<td>Prasad</td>
<td>1991</td>
<td>M</td>
<td>02139</td>
<td>Computer</td>
</tr>
<tr>
<td>Dhanashree</td>
<td>1992</td>
<td>F</td>
<td>02141</td>
<td>Big data</td>
</tr>
<tr>
<td>Aardra</td>
<td>1992</td>
<td>F</td>
<td>02142</td>
<td>Computer</td>
</tr>
<tr>
<td>Avinash</td>
<td>1970</td>
<td>M</td>
<td>02138</td>
<td>Computer</td>
</tr>
<tr>
<td>Harshad</td>
<td>1970</td>
<td>M</td>
<td>02139</td>
<td>Big data</td>
</tr>
</tbody>
</table>

After generalization of a table.

<table>
<thead>
<tr>
<th>Name</th>
<th>Zip</th>
<th>Branch</th>
</tr>
</thead>
<tbody>
<tr>
<td>Abhay</td>
<td>02138</td>
<td>Student</td>
</tr>
<tr>
<td>Prasad</td>
<td>02139</td>
<td>Student</td>
</tr>
<tr>
<td>Dhanashree</td>
<td>02141</td>
<td>Student</td>
</tr>
<tr>
<td>Aardra</td>
<td>02142</td>
<td>Student</td>
</tr>
<tr>
<td>Avinash</td>
<td>02138</td>
<td>Student</td>
</tr>
<tr>
<td>Harshad</td>
<td>02139</td>
<td>Student</td>
</tr>
</tbody>
</table>

\[ \text{GT}[1,0] \]

<table>
<thead>
<tr>
<th>Name</th>
<th>Zip</th>
</tr>
</thead>
<tbody>
<tr>
<td>Student</td>
<td>0213*</td>
</tr>
<tr>
<td>Student</td>
<td>0213*</td>
</tr>
<tr>
<td>student</td>
<td>0214*</td>
</tr>
<tr>
<td>student</td>
<td>0214*</td>
</tr>
<tr>
<td>student</td>
<td>0213*</td>
</tr>
<tr>
<td>student</td>
<td>0213*</td>
</tr>
</tbody>
</table>

\[ \text{GT}[1,1] \]

<table>
<thead>
<tr>
<th>Name</th>
<th>Zip</th>
</tr>
</thead>
<tbody>
<tr>
<td>Abhay</td>
<td>0213*</td>
</tr>
<tr>
<td>Prasad</td>
<td>0213*</td>
</tr>
<tr>
<td>Dhanashree</td>
<td>0214*</td>
</tr>
<tr>
<td>Aardra</td>
<td>0214*</td>
</tr>
<tr>
<td>Avinash</td>
<td>0213*</td>
</tr>
</tbody>
</table>

\[ \text{GT}[1,2] \]

3.2 Agents
For defining agents, its not having any perfect definition. Its simply depends on that particular environment. When 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 an software depends on situation its react. There are different type of agent in this
paper intelligent agents are use, which is taking decision on behalf of user on the basis 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. In this paper multi agents are use 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. 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 unmaintained, because after 2005 in which it is introduce no updating is done so far. Another is FAMOJA (Framework for Agent-based Modelling with Java), JANUS, JIAC (Java-based Intelligent Agent Computer framework), JADE (Java agent Development Environment)[5]. as shown in fig depends on the precepts agent take an action on the basis of what should be output required.

3.3 JADE

In this paper JADE platform is used. Which is support java language ,JADE is an open source software, which is also support php. it is an middleware which develop for multi-agent application. it provide set of API’s.it having several advantages over other platform like it is easy to use[6]. It also provides set of container which is nothing but platform. JADE provide proper mechanism for authentication and verify the authentication of agents also allow to reuse their prepared system. JADE support for scalable property. The another advantages is it support for debugging and management also for monitoring. It also works in mobile environment. The latest version for JADE is 4.3.0.

3.4 Architecture for agent design

When start to think about agent development the particular step are required. if go through particular step by step then it become easy proper communication between multiple agents.

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 dia.
       And decide on a preferred grouping.
   1.3 Develop agent descriptors.

2. Describe the interaction between agent using interaction dia. & interaction protocol.
   2.1 Develop interaction dia. from scenarios.
   2.2 Generalise 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 dia.

In this way agent are created.

Fig 2: Agent and Environments

Fig 3: Phase and relationship in the agent design
4. METHODOLOGY
In this section explanation of architecture and methodology of system is given.

4.1 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 digit code is maintain.

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

4.2 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.

4.3 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.

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

4.5 Description of Architecture
There are certain steps for converting given database into generalized format. Let see steps and algorithm.

Steps-
In this paper algorithm is apply on small dataset. Like simple table,So\text{OT}(\{A_{x}...A_{y}\}); new-table. Ni={A_{1},....,A_{n}} where \{A_{1}....,A_{n}\} belongs to \{A_{x},....,A_{y}\}, where OT is original Table which contain original data and the NI is new Identifier which contain generalized with suppressed value related OT.MGT is k-minimal distortion of OT[N]. in step 1. in first step define original table contain it self generalized format. If it is so then it is k-minimal distortion. If not execute step 2. In step 2. first store all possible generalization of OT over NI into allgens. [2.2] store those generalization from allgens that satisfy the k-anonymity requirement into protected. [2.3] store the k-minimal distortion from protected into MGT. And |MGT|=1 [2.4] and last the func. Predefine () returns a single k-minimal distortion from MGT based on user-defined specification.

Algorithm-
Input-Original Table OT; New_Identified NI={A_{1},....,An}, K constraint, where i=1...n and predefined specifications.

Output-MGT , a minimal distortion of OT[N] with respect to k chosen according to the preference specification, assume |OT|=K.

Method-
1. if OT[N] satisfied k-anonymity requirement with respect to k then do
   1.1 MGT <=\{OT\} //OT is the soln

2. Else do
   2.1 allgen <=\{Ti:Ti is a generalization of OT over NI\}
   2.2 protected <=\{Ti:Ti belongs allgen \cap Ti satisfies k-anonymity of k\}
   2.3 MGT <=\{Ti:Ti belongs \cap there does not exist Tz belong protected such that \textit{prec}(Tz)>\textit{Prec}(Ti)\}
   2.4 MGT <=\textit{predefine}(MGT)//select predefined soln

3. Return MGT

This is algorithm for simple generalization with minimal distortion. This algorithm simply or quietly similar to minGen algorithm. This is used in existing system. This algorithm used on given database. So multi agents use to do this task.

Fig 4: Architecture of system

The architecture of given system shown in fig.4 let see how it works.

1. When user send a request to access database to TTP(Trusted third party) then on the basis of SMS based authentication it check authorisation.
2. If it is authorised then able to update the data base, if user wants to access particular data then send request through TTP.
3. 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.

4. And the above define algo apply on database and provide generalized data to user.

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

4.6 Pseudo code related Agents

4.6.1 Let see how to create simple agent for display information

Code-

```java
import jade.core.Agent;

public class AgentDhanu extends Agent {
    protected void setup() {
        System.out.println("Hello VIT Mtech Compsci. ");
        System.out.println("My name is "+getLocalName());
    }
}
```

Output-

```
% javac AgentDhanu.java
% java jade.Boot fred:AgentDhanu
IOR:00000000000001149444C3A464950412F4D54533A312E30000....
.... 0020501000100102000010109000000100010100
Agent container Main-Container@JADE IMTP://Dhanu-
Computer.local. is ready.
Hello VIT Mtech Compsci.
My name is fred
```

4.6.2 Program for sending and receiving message from agent

To simplify answering, Jade provides a method createReply() which creates a new message with the sender and receiver attributes switched and all other attributes set correctly. Generally, only the content and performative have to be modified before sending it back. The action method from pong.java, a modified Receiver agent which answers all messages with a Pong.

```java
public void action() {
    ACLMessage msg = receive();
    if (msg!=null) {
        System.out.println("- " + myAgent.getLocalName() + " <- " + msg.getContent());
        ACLMessage reply = msg.createReply();
        reply.setPerformative(ACLMessage.INFORM);
        reply.setContent(" Pong ");
        reply.send();
    }
    block();
}
```

4.6.3 Algorithm for communication between different agent

1. When particular user send request then this request send to user agent.
2. On the basis of request user create an task agent.
3. Particular task agent send the request to data mining agent.
4. Then data mining agent communicate with data agent which contain metadata related request.
5. On the basis of mining type specific mining is done.
6. And the final result as a response send to task agent.
7. Through task agent it send to user agent .
8. 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.

1. 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.

2. 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.

3. Data mining agent-

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

4. Data agent-

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

5. CONCLUSION

In this paper propose system ,which used to make the data privacy and confidentiality of the database. Server can allow only the authorized users to view the original database while maintaining their privacy, and in paper it shows how the data generalized the given data. The proposed security mechanism provides anonymous authentication, data privacy, confidentiality of the data, secure data access, data integrity
and access control. Using agent in existing system it becomes more effective on existing system.

6. FUTURE WORK
Now given algorithm is apply on centralized database. And as a future work we will apply on cloud. Also in future work we would like to introduce new algorithm which is gives more heuristic solution. In system basically focused on database side so as an future work user side privacy preservation also provide.

7. REFERENCES


