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

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 3, March 2015

Multilevel Anti-Discrimination Privacy Preserved Data Transmission

Naveena M.S[1], Merlin Shoerio[2]

M. Tech Student, Marian Engineering College, Trivandrum, Kerala, India[1]

Asst. Professor, Dept. of CSE, Marian Engineering College, Trivandrum, Kerala, India[2]

ABSTRACT: Most organizations will need to reveal their crucial data which includes the sensitive information that discloses one's identity during data transmission processes. To limit the access to such sensitive data, various privacy preservation techniques are applied based on the level of priority assumed. Discrimination is the unfair treatment of an individual or group based on their membership on a particular category. So, the decision attribute that leads to discrimination needs to be hided or transformed. This paper deals with the correlation of discrimination prevention and privacy preservation. By applying privacy preservation techniques, it can be shown that the discrimination prevention can be easily accomplished along with secure transmission of data to different levels of users. The privacy preservation methods such as generalization, bucketization, slicing, etc. are applied to the system based priority level. The multilevel anti-discrimination privacy preserving data transmission system used in this paper show the interwined concepts of both privacy preservation and discrimination prevention which leads to the efficient secure data transmission.

KEYWORDS: Discrimination Prevention, Privacy Preservation, Data Anonymization, Slicing.

I.INTRODUCTION

In fast growing technology, most organizations will share aggregated data for their business analytics and to provide services. For example, retailers will often share their customer's purchasing information to the product merchants for their effective advertising but they do not need to reveal customer's sensitive data which leads to discrimination. So, such valuable details should be preserved securely without any harm to the data. On the basis of sociological aspect, discrimination is the prejudicial treatment of an individual based on their membership in a particular group or category.

Privacy preserving of microdata is the release of aggregate information about data without leaking the individual information of the participants. Several microdataanonymization techniques [8] are developed so far. The most popular ones are generalization for k-anonymity, bucketization for l-diversity, slicing. The discrimination risks and the privacy should be tackled together to avoid information loss and to provide security while data transmission.

II. RELATED WORKS

Despite the wide technological development, the issue of anti-discrimination in data mining did not get as much attention until 2008. For massaging of data [4], the class labels of the most likely victims and the profiteers are changed. The modified data is then used for learning a classifier with no discrimination for future decision making. The drawback of this approach is that it is intrusive in nature.

Preferential sampling [5] changes the distribution of differential data objects for a particular dataset to be discrimination free. The data objects close to the decision boundaries are more prone to discrimination. So, the distribution of the borderline objects is changed to make it discrimination free. Drawback of this approach is low utility rate and minimum discrimination removal.In decision tree learning approach [6], the solution is based on the integration of discrimination awareness into the induction model process of a decision tree. This method gives high accuracy and low discrimination score, but the decision tree construction is complex.

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

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 3, March 2015

In [2], the crime intrusion is detected and only direct discrimination prevention is addressed. [2] is extended to [3] and it discovers indirect discrimination in databases. Here, only preliminary experimental proofs are given. A unified approach for discrimination prevention of direct as well as indirect discrimination, with finalized algorithms and all possible data transformation methods based on rule protection and/or rule generalization could be applied simultaneously or both at the same time was proposed in [1]. There is no efficient privacy preservation techniques explained which results in huge information loss.

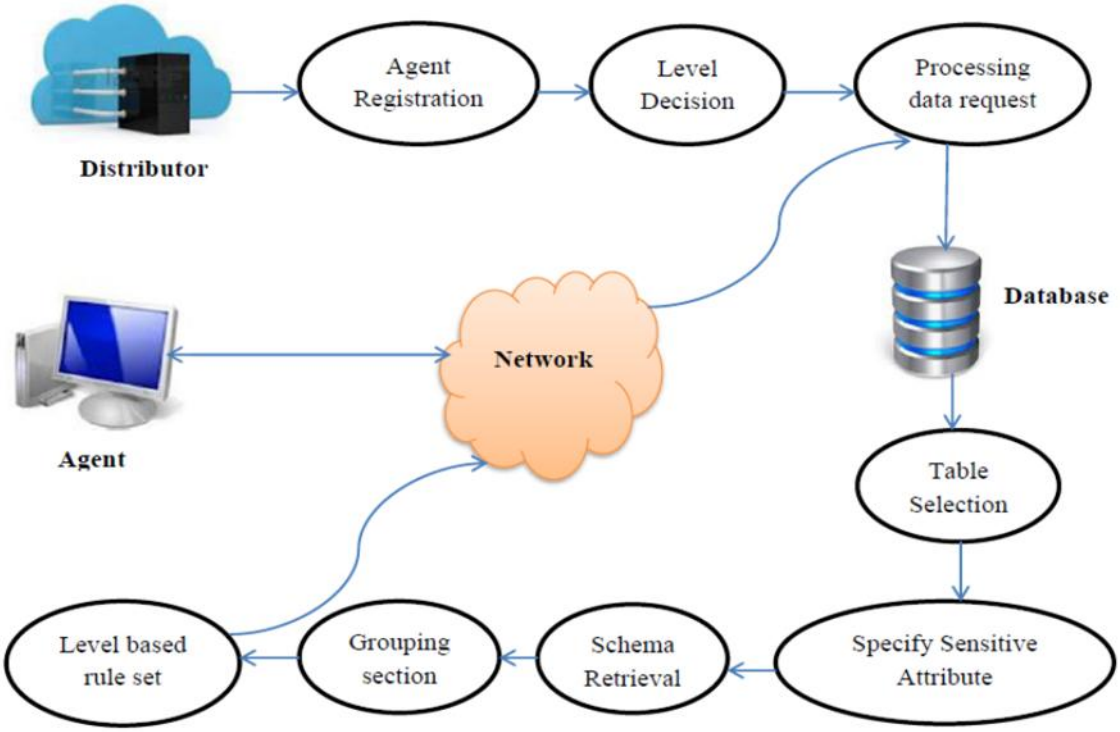

Fig.1. Multilevel Anti-Discrimination Privacy Preserved Data Transmission Architecture

III. MULTILEVEL ANTI DISCRIMINATION PRIVACY PRESERVED DATA TRANSMISSION

The correlation of discrimination prevention and privacy preservation is dealt in the proposed software. Here, the distributor will register the agent and he will decide the level of priority to be given to the agent. For each level, different privacy preservation rule sets are designed which determine the degree of data to be hided or preserved. The distributor will accept the data request from the registered agent only. Then, appropriate database requested by the agent is chosen and required table is selected. The sensitive attributes which lead to discrimination are specified thereby. Then, the schema is retrieved and the resultant structured data is subjected to grouping. Based on the sensitive attributes, the rules are classified into potentially discriminatory and potentially nondiscriminatory groups.

According to the level assigned by the distributor to each agent, the related rule set is applied. The rule sets are the various privacy preservation techniques designed on the basis of the degree of data to be hidden according to the type of agent. The transformed dataset is transmitted through the network and will reach to the agent who requests the dataset.

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

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 3, March 2015

A common anonymization approach is generalization which replaces quasi identifier values with values that are less specific but semantically consistent. So that, more records will have the same set of quasi identifier values. A considerable amount of information loss for high dimensional data is the major drawback of generalization.

A new notion of privacy was introduced in [8] called l-diversity. The records are sorted based on the occurrence of sensitive attributes. Then, group the similar records with set of buckets and analyse it. Combine the set of correlated attributes after diversity check is done. Bucketization does not prevent membership disclosure.

In slicing [7], after attribute partitioning, the highly correlated attributes are in same column. Then, column generalization is done thereby to protect membership disclosure. Tuple partitioning is done to generate the sliced table. This approach of data privacy is highly secure.

Multilevel Privacy Preservation (MLPP) Algorithm details the transformation of the original dataset to privacy preserved discrimination free dataset according to the level of the agent.

A.*MLPP (Multilevel Privacy Preservation Algorithm)*

Inputs: DB, $DR_{1...n,}$ $L_{1...n}$
Output: DB' (transformed data set)
1: **for** each DR_i
2: DB_s ←all records that completely supporting SA
3: **for** each L_i
4: **for** each $A_i \in DB_s$
5: **if** $A_i = SA_i$**then**
6: DB_c← all records completely supporting SA
7: **if**L($DR_i = L_i$) **do**
8: Select first record in DB_c
9: Modify class item of $DB_c$with R_i
10: DB' ← Recomputed DB
11: **while** (!DB_c(n))
12: **end do**
13: **end if**
14: **end if**
15: **end for**
16: **end for**
17: **end for**

The inputs for the MLPP algorithm is the original dataset (DB), the data requests for n agents ($DR_{1...n}$) and their level of privacy required ($L_{1...n}$). DB_s are the set of records that completely supporting the sensitive attributes (SA). $DB_c$are the corresponding set of records containing the column values which completely supports the sensitive attributes. DB' is the transformed privacy preserved dataset.

IV.PERFORMANCE ANALYSIS

The performance of the level based privacy preserved discrimination free data transmission software can be analysed on the basis of the security it provides. This software will provide high security without any information loss issue and can be used in any organization. According to the levels, the certain data can be made hidden which eliminates the discriminatory biases. As in [1], the performance analysis can be made on the German Credit Dataset.

V.CONCLUSION

The description of the multilevel anti-discrimination privacy preserved data transmission software can be concluded by pointing out the essentiality of the software. The proposed software is dynamic in nature and thus can be installed in any organization and can handle any databases. We can conclude that by preserving privacy of dataset itself, the discrimination can be avoided.

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

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 3, March 2015

REFERENCES

1. Sara Hajian, Josep Domingo-Ferrer, "A Methodology for Direct and Indirect Discrimination Prevention in Data Mining," IEEE Transactions on Knowledge and Data Engineering, Vol. 25, No. 7, July 2013.
2. S. Hajian, J. Domingo-Ferrer, and A.Martı´nezBalleste´,"Discrimination Prevention in Data Mining for Intrusion and Crime Detection," Proc. IEEE Symp. Computational Intelligence in Cyber Security (CICS '11), pp. 47-54, 2011.
3. Sara Hajian, Josep Domingo-Ferrer, and A. Martı´nez- Balleste´, "Rule Protection for Indirect Discrimination Prevention in Data Mining," Proc. Eighth Int'l Conf. Modeling Decisions for Artificial Intelligence (MDAI '11), pp. 211-222, 2011.
4. Faisal Kamiran and ToonCalders, "Classification without Discrimination", Proc. IEEE Second Int'l Conf. Computer, Control and Comm.(IC4 '09),2009.
5. Faisal Kamiran and ToonCalders, "Classification with no Discrimination by Preferential Sampling,"Proc. 19th Machine Learning Conf. Belgium and The Netherlands, 2010.
6. Faisal Kamiran, ToonCalders, and M. Pechenizkiy, "Discrimination Aware Decision Tree Learning," Proc. IEEE Int'l Conf. Data Mining (ICDM '10), pp. 869-874, 2010.
7. Tiancheng Li, Ninghui Li, Senior Member, IEEE, Jian Zhang, Member, IEEE, and Ian Molloy, "Slicing: A New Approach for Privacy Preserving Data Publishing", IEEE Transactions On Knowledge And Data Engineering, Vol. 24, March 2012
8. Ninghui Li, Tiancheng Li, Suresh Venkatasubramanian, "t-closeness:Privacy beyond k-anonymity and l-diversity", Proc. IEEE 23rdIn'l Conf. Data Engineering, (ICDE 2007).