Implement L-diversity By Using Generalization Algorithm

Miss Swati Abhimanyu Nase

Master of Engineering, Computer Science & Engineering ,CSMSS CHH , Shahu College Of Engineering , Aurangabad, Maharashtra ,India

---***---

Abstract : Database security is serious problem from several years .Security risk to database system includes unauthorized access to database. A release of data is said to have the k-anonymity property if the information for each individual contained in the release cannot be distinguished from at least k-1 individuals whose information also appear in the release. In Relational database, to avoid identity disclosure one record in table has same Quasi identifier at least k-1 record, to implement L-diversity .Background knowledge attack to access sensitive attribute possible by linking common Quasi identifier attribute in tables in relational database. This paper proposed generalization algorithm to implement L-diversity.

Keywords : L-diversity, Relational database, Quasi identifier ,k-anonymity, Background knowledge attack

I INTRODUCTION:

Consider following two tables containing Hospital Patient Data and Vote Registration Data. Hospital Patient Data contains fields Zip code , Age ,Sex ,Date of birth, Disease .Vote Registration Data contains field birth, Zip code , Age ,Sex ,Date of birth ,Name. Quasi identifier attribute except Name & Disease which can be link and leads to identity disclosure and leakage of sensitive information such as Disease .To avoid this Generalization technique used on relational database.

Table-1: Hospital Patient Data

Zip code	Age	Sex	Date of birth	Disease
10080	42	Female	9/9/68	Malaria
10150	48	Male	5/5/63	Flu
10270	56	Female	4/3/55	Heart attack
10350	60	Male	2/3/49	Cough &Cold
10430	65	Male	2/2/44	Anemia

Table-2: Vote Registration Data

Zip code	Age	Sex	Date of birth	Name
10080	42	Female	9/9/68	Jiya
10150	48	Male	5/5/63	Aman
10270	56	Female	4/3/55	Natasha
10350	60	Male	2/3/49	Brian
10430	65	Male	2/2/44	Akhil

By linking Quasi identifier attribute i.e. attribute which are common in both table viewer knows the personal sensitive information that Jiya has disease Malaria .To avoid this Generalization technique proposed in following table Zip code and Age are generalize instead of retrieve specific value.

Table-3: Generalization

Zip code	Age	Sex	Date of birth	Name
10000-10100	40-45	Female	9/9/68	Jiya
10101-10200	46-51	Male	5/5/63	Aman

10201-10300	52-57	Female	4/3/55	Natasha
10301-10400	58-62	Male	2/3/49	Brian
10401-10500	63-68	Male	2/2/44	Akhil

II LITERATURE SURVEY:

Protecting Respondents' Identities in Micro data Release, paper present generalization and suppression technique used for anonymity to preserve truthfulness of information. Minimal generalization used when data are not generalized more than necessary to provide k-anonymity. Algorithm for minimal generalization that satisfy prefer criteria. Quasi identifier a set of attribute in private table can linked with external information to re identify respondent to whom information refer. Each release data must be such that combination of values of quasi identifier match at least k individual. Generalization can be generated by dropping rightmost digit zip code; postal address can be generalized to street dropping number. In generalization five digit zip codes generalized to 4 digits & 3 digits. Value in private table can be substituted upon release with generalized value. Suppression is remove data from table so it can be released. An algorithm for minimal generalization is proposed in which lowest height among all generalization in return. It's having smaller number of generalization step. Both generalization and suppression technique important for protection of data [1]. L-Diversity: Privacy Beyond k-anonymity paper present to avoid attack on k-anonymous dataset L-diversity framework that gives stronger privacy guarantee .k anonymity can create group that leak information due to lack of diversity in sensitive Attribute k anonymity does not protect against at attack based on background knowledge .In Bayes optimal privacy modeling background knowledge as probability of distribution of attribute and use Bayesian interference technique to reason about privacy. Tool for reasoning privacy, theoretical principal of privacy, difficulties that need to be overcome to the arrive at practical definition of privacy, in privacy principle there is positive disclosure & negative disclosure. Adversary can correctly identifies value of sensitive attribute with high probability in positive disclosure. Adversary can correctly eliminate value of sensitive attribute. In uninformative principle little additional knowledge beyond background knowledge. Block is L-diverse if it contains at least well represented values of sensitive attributes. Implementing privacy preserving data publishing table preserve privacy then generalization of table also preserve privacy this is called as monotonicity property [2].

III GENERALIZATION ALGORITHM:

```
  1. Get attribute which is to be a generalized
   2. Initialize variable no to x value
   3. While (no%5! =0)
    {
     no++
     }
    return no
  4 While (x%5!=0)
   {
    x --
    }
 return x
5 Display value stored in variable no, x for getting maximum and minimum range
```

HOW IT WORKS:

Consider here age attribute is to be generalized. Initialize variable name no to user enter age value .Increment the value store in variable no till while loop condition not satisfy. When while loop condition fails it return value stored in no variable which is maximum value in range to be generalize. In fourth line decrement value of x till while loop condition satisfy & return value which is minimum value in range. In final step display value stored in variable name no and x for getting maximum and minimum range. In this algorithm to maintain interval of 5, the variable name no and x divided by 5.Depends on how much interval maintain in attribute which is to be generalize variable no and x divided by that number value.

IV CONCLUSION

By understanding problem definition in field of database security to , with reference to literature survey carried out , Generalization algorithm proposed to implement successfully implement concept of L-diversity.

REFERENCES

[1] P. Samarati , "Protecting Respondents' Identities in Micro data Release," IEEE Trans. Knowledge and Data Eng., vol. 13, no. 6, pp. 1010-1027, Nov. 2001.
[2] A. Machanavajjhala , D. Kifer , J. Gehrke , and M. Venkitasubramaniam ,"L-Diversity: Privacy beyond k-anonymity," ACM Trans. Knowledge Discovery from Data, vol. 1, no. 1, article 3,2007.