Preventing Inference to an Attribute Set due to Functional Dependencies

2Q-4

Somchai CHATVICHIENCHAI and Yahiko KAMBAYASHI
Faculty of Engineering, Kyushu University

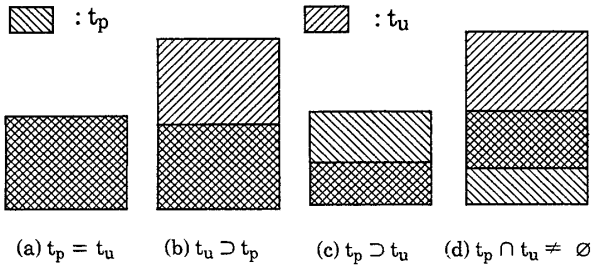
1. Introduction

This paper describes the concept of preventing some users from accessing an attribute set while an access to individual attributes of the set may be permitted. It seems that this kind of access control has not yet been provided in current relational database systems[1] (such as SQL/DS, INGRES) or studied in a Multilevel Relational Data Model proposed by Denning[2]. Although an access to the attribute set is protected, an inference to the attribute set may possibly be done by using knowledge of functional dependencies in the database. To protect such an inference, any access to other attribute sets needed to infer the attribute set must also be prevented. This paper applies the Chase Algorithm[3] to examine whether preventing access to those attribute sets is sufficient to protect inference to the prevented attribute set.

2. Basic Concept and Definitions

Types of disclosure of information of the protected attribute set are classified as follows:-

Let p be an attribute set to which an access made by user G is protected while an access to individual attributes is permitted; t_p be the set of tuples representing information of p; and t_u be a set of tuples inferred by user G.



Case (a) is protected by the approach we are proposing. Case (b), (c), and (d) will be discussed in Section 5.

Let $U = \{A_1, A_2, .., A_n\}$ be an universal scheme. We will denote a subset of U by .. X,Y,Z and will use concatenation for forming such subsets (thus A_iA_j stands for $\{A_i, A_j\}$). Let r be a relation for scheme U.

Definition 1: Access Prohibited Attribute Set (PHB) is an attribute set which an access to $\Pi_X(r)$ where $X \supseteq$ PHB is prohibited while an access to $\Pi_Y(r)$ where $Y \not\supseteq$ PHB may be permitted. The PHB defined for a user may be different from those of other users corresponding to application requirement.

Definition 2: FD-Compromise to PHB

Consider r, F, PHB, S where r: relation for scheme U, F: a set of functional dependencies in U, PHB: an access prohibited attribute set for user G, S: set of attribute sets that user G can access. There exists FD-Compromise to PHB with respect to S if there exists $\Pi_{PHB}(\Pi_{R_1}(r) * \Pi_{R_2}(r) * \cdots * \Pi_{R_k}(r)) = \Pi_{PHB}(r)$ with respect to F where $R_1 \cup R_2 \cup ... \cup R_k \supseteq$ PHB; $R_i \in S$ $(i=1, .., k)$; and * represents a natural join operator.

3. An Approach to prevent FD-Compromise to PHB

Refering to FD-Compromise to PHB, lets represent $\{R_1, R_2, .., R_k\}$ with Φ. In general, there may exist a set of $\Phi_1, \Phi_2, .., \Phi_m$ with respect to U. Let $\Phi_1 = \{R_1^{(1)}, R_2^{(1)}, .., R_{k_1}^{(1)}\}$, $\Phi_2 = \{R_1^{(2)}, R_2^{(2)}, .., R_{k_2}^{(2)}\}$, ..., $\Phi_m = \{R_1^{(m)}, R_2^{(m)}, .., R_{k_m}^{(m)}\}$.

Example (1) Consider U = ABCD; F = {A → B, B → C, D → B}; PHB = BC. Sets of attribute sets can be used to obtain $\Pi_{PHB}(r)$ are as follows:- {AB, AC}, {DB, DC}, {AD, AC, DB}, {AD, AB, DC}, {ABD, AC}, {ABD, DC}, {ACD, AB}, {ACD, DB}.

For every Φ_i where $(i=1, .., m)$, we need to prevent user G from accessing at least a member of Φ_i such that $\Pi_{PHB}(r)$ cannot be obtained by $\Pi_{PHB}(\Pi_{R_1}^{(i)}(r) * \Pi_{R_2}^{(i)}(r) * .. * \Pi_{R_{ki}}^{(i)}(r))$. That is, for each Φ_i (where $i = 1, .., m$), we will protect user G from acessing an attribute set $R_j^{(i)}$ (where $j = 1, .., k_i$) as well as PHB. Let IHB = $\{IH_1, IH_2, .., IH_p\}$ be a set of attribute sets such that:

$(\exists IH_k \in IHB)(\forall \Phi_i)(\exists R_j^{(i)} \in \Phi_i)$ s.t. $IH_k \subseteq R_j^{(i)}$,
and $(\exists IH_k \in IHB)$ s.t. $IH_k \subseteq$ PHB,
and $\forall k(1,2, .., p) \neg \exists l(1,2, .., p)$ s.t. $IH_k \supset IH_l$;
where $i=1,2, .., m$; $j= 1,2, .., k_i$

If all attribute sets in IHB are protected from being accessed by user G, there is no way for him to obtain $\Pi_{PHB}(r)$. Let us call IHB an FD-Compromise Inhibitor for PHB. In general, there may exist several IHBs for a PHB.

Example (2) According to example (1), a set of possible IHBs for PHB = BC are as follows:-

IHB_1 = { AB, DB, BC}, IHB_2 = {AB, DC, BC}, IHB_3 = {AC, DB, BC}, IHB_4 = {AC, DC, BC}, IHB_5 = {AB, DB, DC, BC}, IHB_6 = {AC, DC, DB, BC}, IHB_7 = {AB, AC, DB, DC, BC}

For simplicity, we will consider those IHBs in reduced form.

Definition 3: IHB = $\{IH_1, IH_2, .. , IH_p\}$ is in reduced form if there exists no IHB' = $\{IH'_1, IH'_2, .. , IH'_q\}$ such that

$p \geqq q$ and $\forall j(1,2,..,q)\ \exists i(1,2,..,p)\ IH'_j \supseteq IH_i$

According example (2), IHB_1, IHB_2, IHB_3, and IHB_4 are in reduced form.

Since the algorithm for finding a set of possible IHBs takes very much time to compute a set of possible IHBs, we proposed another algorithm that is used to examine whether given attribute sets is an FD-Compromise Inhibitor for a PHB. Futhermore, if the given attribute sets is an FD-Compromise Inhibitor but not in reduced form, the reduced form of the given attribute sets will be returned.

4. Algorithm for Checking FD-Compromise Inhibitor for PHB

Input: U = $\{A_1, A_2, .. , A_n\}$, a set of functional dependencies F, PHB, IHB = $\{IH_1, IH_2, .. , IH_p\}$.

Output: A decision whether IHB is a FD-Compromise Inhibitor for PHB. If IHB is a FD-Compromise Inhibitor for PHB, RDF_IHB which is a reduced form for IHB will be returned.

Method:

(1) Find ATTRS = $\{X \mid X \in 2^U - \cup_{i \in \{1,..p\}} SS(IH_i)\}$ where $SS(IH_i) = \{X \mid X \subseteq U, X \supseteq IH_i$ where $IH_i \in IHB\}$

(2) Find LARGE_ATTRS = $\{X \mid X \in ATTRS, \neg\exists Y \in ATTRS$ s.t. $Y \supset X\}$.

(3) Refering to the Chase Algorithm for Inference of Dependencies, construct a table with each column corresponding to each attribute of U and each row corresponding to each member of LARGE_ATTRS. We perform operations in the same way as the Chase Algorithm. After finishing the process, if we discover that some row contains $a_i a_j .. a_m$ each of which corresponds to each attribute in PHB, then IHB is not an FD-Compromise Inhibitor for PHB and the algorithm terminates here. If not, IHB is an FD-Compromise Inhibitor for PHB.

(4) From this step, RDF_IHB which is a reduced form for IHB will be computed. At first, find SS_IHB = $\{X \mid X \in \cup_{i \in \{1,..,p\}} SS(IH_i), X \subset$ PHB$\}$. Assign RD_IHB = $\emptyset$.

(5) Find SS_IH $\in$ SS_IHB where $\neg\exists$ SS_IH' $\in$ SS_IHB s.t. SS_IH $\supset$ SS_IH'. Set LARGE_ATTRS' = LARGE_ATTRS $\cup$ SS_IH. Perform operations in the same way as step (3) but using LARGE_ATTRS' instead of LARGE_ATTRS. After finishing the process, if we discover that some row contains $a_i a_j .. a_m$ each of which corresponds to each attribute in PHB, then do step (i). If not, do step (ii).

(i) SS_IHB = SS_IHB - $\{X \mid X \subseteq U, X \supseteq$ SS_IH$\}$, and proceed to step (6).

(ii) RD_IHB = RD_IHB $\cup$ SS_IH, SS_IHB = SS_IHB - SS_IH.

(6) If there exists a member in SS_IHB then repeatly do step (5). If not, find SS_IHB = $\cup_{i \in \{1,..,p\}} SS(IH_i)\}$ - RD_IHB, and RDF_IHB = $\{X \mid X \in$ SS_IHB, $\neg\exists Y \in$ SS_IHB s.t. $X \supset Y\}$

5. Protection of Other Types of Information Disclosure

Refering to Section (2), case (b) is permitted only if t_u is obtained by joining relations $R_1, R_2, .. , R_k$ where there does not exist JD*($R_1, R_2, .. , R_k$) satisfied by p over the union of set $R_1, R_2, .. , R_k$. Since t_u is much greater than t_p, it is very difficult for user G to distinguish which tuples associated with t_p. Case (c) must be protected since all tuples which user G can access are associated with t_p. Case (d) may be permitted since it is difficult for user G to identify which tuples associted with t_p.

6. Conclusion

To prevent users from obtaining information about an attribute set, we must consider inference to the attribute set due to functional dependencies in the database. An approach to prevent such inference is proposed. To apply this approach to a set of relation schemes instead of universal scheme, we must consider interaction problems between joining the relation schemes together and inference of the protected attribute set.

References

[1] Date C.J., An Introduction to Database Systems Vol I, 4th Edition, Addison-Wesley System, 1986, pp. 439-452

[2] Denning D.E. et.al., "A Multilevel Relational Data Model", Proc. 1987 Symp. on Security and Privacy, IEEE, pp. 220-234

[3] Ullman J.D., Principles of Database Systems, 2nd Edition, Computer Science Press, 1982, pp. 227-229