International Journal of Information, Control and Computer Sciences
ISSN: 2517-9942
Vol:3, No:7, 2009

A Comparative Study of Fine Grained Security Techniques Based on Data Accessibility and Inference

Azhar Rauf, Sareer Badshah, and Shah Khusro

Abstract**—**This paper analyzes different techniques of the fine grained security of relational databases for the two variables-data accessibility and inference. Data accessibility measures the amount of data available to the users after applying a security technique on a table. Inference is the proportion of information leakage after suppressing a cell containing secret data. A row containing a secret cell which is suppressed can become a security threat if an intruder generates useful information from the related visible information of the same row. This paper measures data accessibility and inference associated with row, cell, and column level security techniques. Cell level security offers greatest data accessibility as it suppresses secret data only. But on the other hand, there is a high probability of inference in cell level security. Row and column level security techniques have least data accessibility and inference. This paper introduces cell plus innocent security technique that utilizes the cell level security method but suppresses some innocent data to dodge an intruder that a suppressed cell may not necessarily contain secret data. Four variations of the technique namely cell plus innocent 1/4, cell plus innocent 2/4, cell plus innocent 3/4, and cell plus innocent 4/4 respectively have been introduced to suppress innocent data equal to 1/4, 2/4, 3/4, and 4/4 percent of the true secret data inside the database. Results show that the new technique offers better control over data accessibility and inference as compared to the state-of-the-art security techniques. This paper further discusses the combination of techniques together to be used. The paper shows that cell plus innocent 1/4, 2/4, and 3/4 techniques can be used as a replacement for the cell level security.

Keywords**—**Fine Grained Security, Data Accessibility, Inference, Row, Cell, Column Level Security.

I. Introduction

INFORMATION technology has revolutionized the ways of our dwelling life. A person sitting in a room is able to check online bank statement, items in a shop, pay his bills, reserve airline tickets, and book a hotel room. At one side the technology has added tones of comports and ease to our lives but on the other hand it has opened new ways of stealing secret information, and committing crimes via internet. Different tools and techniques have been using to protect data and stop potential leakage of information from intruders.

A. Rauf and S. Khusro are with Department of Computer Science University of Peshawar, Peshawar Pakistan.

S. Badshah is with Department of Statistics Islamia College (Chartered University) Peshawar, Pakistan

Databases, because of their critical role in the storing and retrieving of information have gained a tremendous importance in maintaining the secrecy of information. A database management system (DBMS) employs numerous techniques to protect data and make data available for routine work. Different vendors of database management systems offer way to protect data with the help of users' accounts, roles, views, and encryption. The Virtual Private Database (VPD) feature of Oracle 10g implements 'Column Relevance' and 'Column Filtering' techniques to protect data at fine levels of granularity [1]. Views and encryption features of MS SQL Server 2005 can be used to implement row and cell level security techniques [2]. Darryl offers a toolkit that includes a utility to implement cell and row level security in MS SQL Server 2005 [3]. All these state-of-the-art security techniques protect data in databases at instance level.

II. Fine-Grained Security Techniques

Protecting data in a database is more difficult as you have to make much data available to users and at the same time secure it from theft. This is why there has been various ways to protect data in a database as per the security and data availability requirements of an organization. Fine-grained security techniques are getting popular in database vendors as they allow a database administrator to secure data at data instance level in a table thus making much of data available to users for their routine work. The three flavors of fine grained security technique are row, cell, and column level security. These three techniques have their pros and cons. The row and column level security techniques suppress a large amount of innocent data along with the secret data that reduces data accessibility of a table [4]. On the other hand, cell level security does not suppress innocent data but it allows the possibility of data suppression inference as a suppressed cell carries secret data [4]. An intruder can easily generate useful results from the visible cells of the same row which may result information leakage. A new technique of cell plus innocent security has been introduced in this paper that controls such type of inference associated with cell level security by suppressing innocent data along with the secret data. Thus a suppressed cell may not necessarily contain secret data which reduces the level of confidence of an intruder about the secret data. The new technique is also cost efficient in terms of data accessibility and experiments show that the new technique

offers at least 30% or greater data accessibility as compared to row and column level security techniques.

This paper analyzes the three state-of-the-art fine-grained security techniques and the new introduced technique. The statistic models of analysis of variance (ANOVA) and Duncan are applied on the techniques. Security techniques are divided in different groups in terms of the two dependent variables i.e., Data accessibility and inference against the independent variable 'Total Number of Secret Cells'.

III. Research Results

This section gives an overview of the research results obtained from the experiment as stated in [4]. Sample data of 10000, 100,000 and 500,000 rows were generated using the TPC-H schema of wholesale-supplier model [30]. A suite of eight queries (four TPC-H Benchmark and four other queries) were selected to measure data accessibility and inference [4].

A. Data Accessibility

Data accessibility measures the proportion of readable cells available to a user after issuing queries from the suite 'Q'[1] to the database. It is a ratio between the total number of readable cells and the total number of fetched cells (readable and suppressed cells) returned by a query to the database. Data accessibility (D.A) of a query 'q_1' is:

$$\text{D.A.}\ q_1 = \frac{U_{q_1}}{N_{q_1}} \tag{1}$$

In the above equation 'U_{q1}' represents the total number of readable cells and 'N_{q1}' represents the total number of cells (readable and suppressed cells) fetched by the query. 'q_1' is one of the queries from the queries suite 'Q'.

Different patterns of randomly distributed secret cells were generated. Once a pattern of secret cells was generated, different security techniques were applied on the pattern and data accessibility was measured for each security technique. Data accessibility for 'n' queries in the suite 'Q' for a single pattern is:

$$\{\text{D.A.}_{q_n}\}_{p=1} = \frac{1}{n}\sum_{i=1}^{n} U_{qi} \Big/ N_{qi} \tag{2}$$

Data Accessibility is equal to 1 if there are no suppressed cells in the database.

Equation 2 shows the measure of data accessibility of a specific security technique for a single pattern only. This equation was generalized for 'k' patterns, and average data accessibility of a security technique was calculated as follows:

$$\{\text{Average D.A.}_{q_n}\}_{p=k} = 1/k \sum_{p=1}^{k} \{D.A._{q_n}\}_p$$

Or

$$\{\text{Average D.A.}_{q_n}\}_{p=k} = 1/k \sum_{p=1}^{k} \left\{ 1/n \sum_{i=1}^{n} U_{qi}/N_{qi} \right\}_p \tag{3}$$

where 'k' represents the total number of sample patterns and 'n' the total number of queries in suite 'Q'.

B. Inference

Cell level security leaves suppressed cells suspicious. An adversary could relate the unclassified data with other public information and generate useful results. The new technique of cell plus innocent security suppresses innocent cells in addition to the secret cells, which reduces the proportion of suspicious cells (True Secret Cells) among the suppressed cells. The adversary can no longer assume that a suppressed cell contains secret data.

This research took the 'Proportion of True Secret Cells' (ProbTSC) as a dependent variable against 'Total Number of Suppressed Cells', as an independent variable, to empirically measure the 'Inference' for different security techniques.

The number of suppressed cells returned by a query is the total number of cells returned minus the readable cells. So, the proportion of suppressed cells returned by 'n' queries for a single pattern can be computed from the data accessibility as follows:

$$\{\text{ProbSupp}_{q_n}\}_{p=1} = 1 - \{D.A._{q_n}\}_{p=1} \text{ Or}$$

$$\{\text{ProbSupp}_{q_n}\}_{p=1} = 1 - \left(1/n \sum_{i=1}^{n} U_{qi}/N_{qi} \right) \tag{4}$$

The above equation 4 is derived from equation 2.

For cell-level security with no innocent cells suppressed, ProbSupp is the proportion of true secret cells returned by a query. "I=0" indicates that no innocent cells are suppressed.

The new technique suppresses 'I' innocent cells in the database such that I > 0. In that case, ProbSupp includes all suppressed cells including the innocent ones. The ratio of returned True Secret Cell (TSC) to the total number of returned, suppressed cells for 'n' number of queries for a single pattern becomes:

$$\{\text{ProbTSC}_{q_n}\}_{p=1} = \frac{\left(\{\text{ProbSupp}_{q_n, I=0}\}_{p=1} * 1\right)}{\{\text{ProbSupp}_{q_n, I>0}\}_{p=1}} \tag{5}$$

Equation 5 measures 'Inference' of a specific security technique for a single pattern only. For cell-level security, the inference in this sense is equal to 1 indicating that each suppressed cell carries a secret value.

This equation was generalized for 'k' number of multiple patterns to calculate the Average Probability of True Secret Cells as follows:

$$\{\text{AvgProbTSC}_{q_n}\}_{p=k} = \frac{\left(\{\text{ProbSupp}_{q_n, I=0}\}_{p=k} * 1\right)}{\{\text{ProbSupp}_{q_n, I>0}\}_{p=k}} \tag{6}$$

In equation 6, 'k' represents the total number of sample patterns, 'n' the total number of queries in suite 'Q', and 'I'

[1] Suite 'Q' consists of four TPC-H and four other queries that were selected for this research.

the total number of innocent suppressed cells in the database. The study proved that increasing the number of innocent cells decreased the inference measured as ProbTSC.

C. Experiment Design

First of all the 10,000 rows size database was selected for the experiment. A pattern of secret cells was generated across the tables of the database after initially suppressing 4.0% cells of the database. These suppressed cells were considered as typical secret cells spread across the tables of the database. Following fine-grained security techniques were applied one by one on this pattern of secret cells.

1. Cell Level Security
2. Row Level Security
3. Column Level Security
4. Cell + Innocent ¼ Security
5. Cell + Innocent 2/4 Security
6. Cell + Innocent ¾ Security
7. Cell + Innocent 4/4 Security

It is noted here that the first three security techniques are the state-of-the-art security techniques currently offered by different database management systems vendors. The remaining four techniques are the new techniques introduced during the research work in [4]. Data accessibility was determined for each of the techniques.

Another pattern of 4.0% secret cells of the database was generated and data accessibility was recorded accordingly. The experiment was run the same way for 30 different patterns of secret cells but maintaining the percentage of secret cells constant to 4.0% of the database size. Table I shows these measures of Data accessibility associated with each security technique for 30 different patterns. Patterns of secret cells equal to 8.0%, 12%, 16%, 20%, 24%, 28%, 32%, 36%, 40%, 44%, and 48% of the database size were generated on the same fashion and average Data accessibility was recorded. Table II shows the average data accessibility associated with each of the security techniques for 12 different percentages of the secret cells' inside the database. Average inference was also calculated as per the inference formula defined in [4] and results are shown in Table III. In the next phase, sample databases of 100,000 and 500,000 rows were generated and average data accessibility and inference were recorded. Because of the time constraint, only patterns of secret cells equal to 4.0%, 8.0%, 12%,16%, 24%, and 44% of the database size were generated for the 100,000 rows size database. Similarly patterns of secret cells equal to 4.0% and 8.0% of the database size were taken for experiment for the 500,000 rows size database. Average measures of data accessibility and inference for the three different sizes of the database are shown in tables IV and V respectively.

IV. Methods

To check the level of data accessibility and inference using different techniques, Analysis of Variance (ANOVA) technique and Duncan method as a posthoc test are applied to data sets shown in Tables IV and V. Results of the models are given in Table VI for the two dependent variables – data accessibility and inference.

V. Findings

The ANOVA results in Table VI showed that the data accessibility (F+29.90, p=0.0001) was significantly different in different techniques. Similarly the inference obtained from different methods were also significantly different for using the same techniques (F=449.60, p=0.0001). Table VII classifies all security techniques on data accessibility using statistical method Duncan test. This test clearly shows that there is a difference between row and column level security techniques and they both differ from the cell and cell plus innocent security techniques. However, column 4 of Table VII showed that cell plus innocent 4/4, cell plus innocent 3/4, cell plus innocent 2/4, and cell plus innocent 1/4 techniques are not significantly different and form one group. There is no significant difference among these four techniques. In other words, cell plus innocent 1/4, 2/4, 3/4, and 4/4 techniques are interchangeable. Similarly, column 5 of Table VII shows that cell plus innocent 3/4, 2/4, 1/4, and cell level security techniques form another group. The grouping of these security techniques shows that their effect is almost same and we have the option to use one technique as a replacement for another technique.

Table VIII classifies security techniques based on inference. The Duncan test shows that there is no significant difference between row and column level security techniques. However, the remaining security techniques are significantly different from row and column level, including one another. Cell level security has the highest and column level security has the least vulnerability towards inference. The reason being that cell level security has the highest probability of inference, is that every suppressed cell carries secret data and an intruder can generate useful information from the visible cells of the same row. On the other hand, column level security technique suppresses a large number of innocent data along with the secret data in the same column and thus has less chances of inference. Duncan test

Fig. 1 combines the results of the data accessibility and inference for different security techniques discussed above. The figure shows that both data accessibility and inference decrease while moving down from cell to cell plus innocent to row and then column level security. Cell level security remains at the top for data accessibility and inference. It offers greatest data accessibility because the technique does not suppress innocent data. Only secret data is suppressed and that is why it does not start from the 100 percentage. In case of inference, it plots 100 percent results because there is a 100 percent probability in cell level security that the suppressed cell carries secret information. Moving forward, the four variations of the cell plus innocent security techniques i.e., 1/4, 2/4, 3/4, and 4/4 show a constant decrease in data accessibility and inference. The reason being that the techniques suppress a constant number of innocent cells which are equal to 1/4, 2/4, 3/4, and 4/4 of the total number of true

secret cells. Interestingly the cell plus innocent ¼ security technique shows exact percentage i.e., 80% for both inference and data accessibility which means that cell plus innocent¼ security technique offers 80% data accessibility and at the same time there is 80% chance of inference by an intruder.

The gap between data accessibility and inference increases in the case of cell plus innocent ¾, 2/4, and 4/4 security techniques. The greater the gap the better the results are which means greater rate of data accessibility and inference control. Hence the cell plus innocent 4/4 technique shows the greatest gap of inference and data accessibility among the four variations of cell plus innocent security techniques. This means that the technique can be used to control inference with greater rate while suppressing innocent data at fewer rates. It is obvious that cell plus innocent 4/4 security achieves 66% data accessibility and there is 50% chance of inference. There is a sharp decrease in data accessibility and inference in row and column level security techniques because the concerned techniques suppress a large number of innocent data along with the secret data. At one side it provides better results in terms of inference but on the contrary side data accessibility is greatly reduced.

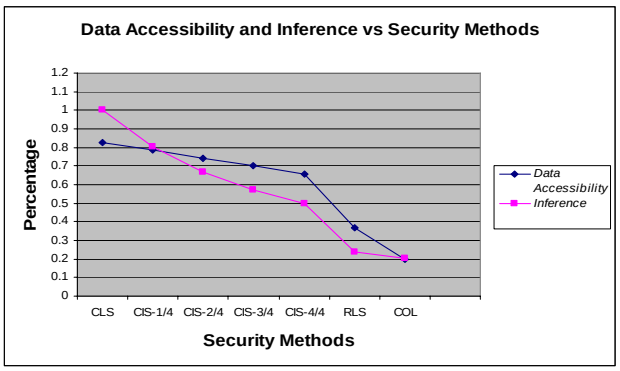

Fig. 1 Data Accessibility and Inference vs. Security Methods

VI. Related Work

Inference and multi-level security models have long been recognized problems. The basic SQL security model implements the policies of Discretionary Access Control [5]. The SeaView [6] project delivered a prototype for a multi-level secure relational database system (MLS-RSBMS) based on Mandatory Access Control. This model provides security at the finest level of granularity by labeling every item in the database including a data element, a tuple, a relation, or a database. Inference was investigated in statistical and general-purpose databases in the past. The problem has been studied at the database schema and instance levels. Researchers have identified different types of inference and proposed formalizations for the existence of inference. Denning [7] suggests database partitioning to detect the potential of an inference. Morgenstern [8] derived the inference relation from classical information theory. Su and Ozsoyoglu [9], [10], [11] studied the inference channels due to the functional and multilevel dependencies in databases. Meadows [12] explained the inference caused by value constraints. Our paper, studies inference caused by the data suppression. There has been recent work on computing k-anonymity to solve the data publishing problem [13], [14], [15]. Samarati [13] proposes the techniques of generalization and suppression for the anonymity of micro data. Researchers have similarly proposed different approaches to eliminate inferences in databases. MITRE research handles inference by modifying the responses of the query during query processing [16], [17], [18], [19]. Haigh [20], [21] detects inferences in databases with the help of auditing. One can monitor and analyze users' activities for possible inference violations. In this approach, a history can be kept of all queries made by a user. Whenever a user creates a query, the history is analyzed to determine whether the response to this query, correlated with the responses to earlier queries, could result in an inference violation. If a violation arises, the systems can take appropriate action e.g., abort the query. Ford Aerospace developed a knowledge-based tool called DataBase Inference Controller (DBIC) to detect and correct logical inferences [22]. DBIC was based on Probabilistic Knowledge Modeling to initiate procedures to calculate the probability of inference and identify violations. Jajodia [23] introduces the snapshot facility. According to this approach, there are situations in which it is possible to allow limited inferences (as in the case of U.S. Bureau of the Census). This method is useful in cases where the bandwidth of inference is so small that these violations do not pose any threat. This method works well in static databases only and has been used by the United States Bureau of the Census [24], [25], [26], [27]. Thomas Hinke [28] worked on preventing inference at Schema level by constructing a semantic relationship graph to check for the possibility of an inference. But this technique is restricted to the schema level and cannot detect inference in case of "many' association cardinalities [29].

VII. Conclusion

This paper is a comparative study of the row, cell, and column level security techniques for data accessibility and inference. A new technique – cell plus innocent security is introduced that provides better security from inference as compared to cell level security and greater data accessibility as compared to row and column level security techniques. The new technique is an intermediate way that offers a choice to a database administrator to select a security technique as per the data accessibility and security requirements of an organization. Statistical methods showed that the cell level security, cell plus innocent 1/4, cell plus innocent 2/4, and cell plus innocent 3/4 security techniques have strong relationship with each other and can be used as a replacement for one another. The security techniques of cell plus innocent 1/4, 2/4, 3/4, and 4/4 make another group, and its significantly different in the case of inference from another group of techniques i.e. row level security and column level security.

TABLE I
DATA ACCESSIBILITY (D.A.) WITH 4.0% TRUE SECRET CELLS FOR 10,000 ROWS SIZE DATABASE

PATTERN	D.A. CLS	D.A. CELL + INNOCENT ¼	D.A. CELL + INNOCENT ½	D.A. CELL + INNOCENT 3/4	D.A. CELL + INNOCENT 1	D.A. RLS	D.A. ColLS
A1	0.962971473	0.955942924	0.948780522	0.939649162	0.924382291	0.74415626	0.177579365
A2	0.967862129	0.961188751	0.952020005	0.940236664	0.932964946	0.790109087	0.220982143
A3	0.966647325	0.959750329	0.950454051	0.942261025	0.934195102	0.782025531	0.205357143
A4	0.958796877	0.952060015	0.944924223	0.933958984	0.926278728	0.713303761	0.175843254
A5	0.967195981	0.959209136	0.952623159	0.94368613	0.934335654	0.780158953	0.226190476
A6	0.962936258	0.956308733	0.943848749	0.933385535	0.924189151	0.750429124	0.174107143
A7	0.965478559	0.954601545	0.945610563	0.938994821	0.933678345	0.763617338	0.207093254
A8	0.962547048	0.955323358	0.946871727	0.940970012	0.930755384	0.768963713	0.161954365
A9	0.970799996	0.957175684	0.947004143	0.940887653	0.933376684	0.807408164	0.189732143
A10	0.96499717	0.957093114	0.950257775	0.939059684	0.932172057	0.765566448	0.191468254
A11	0.964854637	0.957923527	0.951870139	0.942381235	0.930573492	0.766682839	0.191468254
A12	0.959967683	0.948970454	0.941485635	0.93380213	0.923150623	0.746201174	0.175843254
A13	0.97038718	0.96400437	0.954446884	0.946951117	0.939236767	0.804359475	0.220982143
A14	0.96415626	0.955668477	0.948278978	0.938602834	0.93035311	0.756018403	0.193700397
A15	0.963399852	0.952742853	0.945296604	0.935591144	0.929037853	0.748112651	0.207093254
A16	0.967524651	0.960924294	0.954147254	0.947221952	0.939607937	0.788735354	0.207589286
A17	0.971149094	0.96394884	0.956924957	0.950090789	0.941128226	0.805910775	0.205357143
A18	0.965517845	0.957796833	0.951063005	0.943281804	0.932873638	0.76293892	0.207093254
A19	0.969020418	0.960639838	0.951631821	0.942788692	0.934416561	0.791006263	0.220982143
A20	0.964575129	0.951475126	0.94284356	0.936708229	0.927342178	0.754663823	0.193204365
A21	0.971996635	0.965445842	0.954544095	0.946432824	0.938211549	0.811480732	0.220982143
A22	0.964200158	0.953960404	0.9480764	0.940787511	0.933740798	0.756749436	0.189732143
A23	0.968333032	0.961401879	0.95318952	0.94613278	0.938207706	0.782346127	0.207093254
A24	0.96598921	0.95794437	0.947091303	0.941039377	0.933369719	0.786719954	0.212797619
A25	0.968357901	0.952833619	0.945561879	0.934108528	0.92694418	0.783798753	0.207589286
A26	0.962580482	0.955785456	0.948709882	0.936160974	0.929566995	0.742681453	0.144593254
A27	0.959328891	0.949707625	0.941509799	0.932780426	0.925688661	0.723180529	0.165426587
A28	0.966710767	0.959689399	0.948526231	0.938487876	0.929176662	0.775894754	0.207093254
A29	0.969744699	0.963962059	0.956300684	0.947871812	0.939666242	0.799089126	0.209821429
A30	0.968411716	0.959380217	0.953892824	0.945459577	0.938244737	0.787432946	0.193700397
AVG{D.A. q_n} $_{p=1}$	0.965881302	0.957428636	0.949259546	0.940659043	0.932228866	0.771324729	0.19708168

TABLE II
AVERAGE DATA ACCESSIBILITY (D.A.) FOR 10,000 ROWS SIZE DATABASE

PATTERN	Suppressed Cells	AVG (D.A. CLS)	AVG (D.A. C+IS ¼)	AVG (D.A. C+IS ½)	AVG (D.A. C+IS ¾)	AVG (D.A. C+IS 1)	AVG (D.A. RLS)	AVG (D.A. ColLS)
A	5000 (4.0% of database)	0.965881302	0.957428636	0.949259546	0.940659043	0.932228866	0.771324729	0.19708168
B	10000 (8.0% of database)	0.93146457	0.91563741	0.899102046	0.881642653	0.864551413	0.588526699	0.168047288
C	15000 (12.0% of database)	0.900743132	0.876019266	0.850401848	0.825596471	0.8000089	0.461602966	0.157093254
D	20000 (16.0% of database)	0.86176648	0.828268181	0.794908632	0.761018638	0.728543331	0.335372411	0.134408069
E	25000 (20.0% of database)	0.830987661	0.788832559	0.746926809	0.705009633	0.662723818	0.254198315	0.131861772
F	30000 (24.0% of database)	0.798075202	0.748492029	0.698910277	0.648155985	0.596890949	0.19323702	0.122718254
G	35000 (28.0% of database)	0.767059358	0.708778292	0.649661023	0.592257098	0.533992182	0.149444309	0.123817791
H	40000 (32.0% of database)	0.73423229	0.668309118	0.603201705	0.53690072	0.470813016	0.10430607	0.119593254
I	45000 (36.0% of database)	0.701540134	0.624431463	0.545295137	0.467853754	0.394003006	0.0910489	0.106746032
J	50000 (40.0% of database)	0.66529489	0.576878291	0.496327282	0.413517762	0.328892403	0.048923613	0.104662698
K	55000 (44.0% of database)	0.630449127	0.539531386	0.449792699	0.359967851	0.268679005	0.03881629	0.090426587
L	60000 (48.0% of database)	0.599713092	0.494835325	0.402631804	0.301587349	0.202995504	0.025177182	0.098412698
$\{\text{Avg D.A. } q_n\}_{p=30}$		0.78226727	0.72728683	0.673868234	0.619513913	0.565360199	0.255164875	0.129572448

TABLE III
AVERAGE INFERENCE FOR 10,000 ROWS SIZE DATABASE

PATTERN	Suppressed Cells	AVG Inference CLS	AVG Inference C+IS ¼	AVG Inference C+IS ½	AVG Inference C+IS ¾	AVG Inference C+IS 1	AVG Inference RLS	AVG Inference ColLS
A	5000 (4.0% of database)	1	0.801447137	0.672416094	0.574960359	0.503439976	0.14920152	0.042493361
B	10000 (8.0% of database)	1	0.81239125	0.6792549	0.579055136	0.505988517	0.166561062	0.082378997
C	15000 (12.0% of database)	1	0.800582997	0.663489935	0.569121901	0.496306426	0.184356268	0.117755456
D	20000 (16.0% of database)	1	0.804938307	0.674009451	0.578428036	0.509228677	0.207986431	0.159698254
E	25000 (20.0% of database)	1	0.800371206	0.667839759	0.572941893	0.50110962	0.226618339	0.194683674
F	30000 (24.0% of database)	1	0.802856453	0.670646596	0.573904313	0.500918541	0.250290114	0.230170978
G	35000 (28.0% of database)	1	0.799873895	0.66490073	0.571292941	0.499864235	0.273868771	0.265858675
H	40000 (32.0% of database)	1	0.80125118	0.669780374	0.573889275	0.502218911	0.296717105	0.301869234
I	45000 (36.0% of database)	1	0.794688149	0.656381514	0.560860606	0.492510472	0.32835635	0.334126549
J	50000 (40.0% of database)	1	0.791037431	0.664528964	0.570699483	0.498735391	0.351922427	0.373831303
K	55000 (44.0% of database)	1	0.802553881	0.671657524	0.577394235	0.505319655	0.384474757	0.40629032
L	60000 (48.0% of database)	1	0.792388952	0.670084062	0.573138111	0.502239209	0.410625295	0.443980197
$AVG\{\text{Inference}_{qn}\}_{p=30}$		1	0.80036507	0.668749159	0.572973857	0.501489969	0.269248203	0.24609475

TABLE IV
AVERAGE DATA ACCESSIBILITY FOR THE THREE DIFFERENT SIZES OF THE DATABASE

DB Rows	Secret Cells	D.A.CLS	D.A.C+I1/4	D.A.C+I2/4	D.A.C+I3/4	D.A.C+I4/4	D.A.RLS	D.A.ColLS
10,000	(4.0% of DB size)	0.97	0.96	0.95	0.94	0.93	0.77	0.20
10,000	(8.0% of DB size)	0.93	0.92	0.90	0.88	0.86	0.59	0.17
10,000	(12.0% of DB size)	0.90	0.88	0.85	0.83	0.80	0.46	0.16
10,000	16.0% of DB size	0.86	0.83	0.79	0.76	0.73	0.34	0.13
10,000	20.0% of DB size	0.83	0.79	0.75	0.71	0.66	0.25	0.13
10,000	24.0% of DB size	0.80	0.75	0.70	0.65	0.60	0.19	0.12
10,000	28.0% of DB size	0.77	0.71	0.65	0.59	0.53	0.15	0.12
10,000	32.0% of DB size	0.73	0.67	0.60	0.54	0.47	0.10	0.12
10,000	36.0% of DB size	0.70	0.62	0.55	0.47	0.39	0.09	0.11
10,000	40.0% of DB size	0.67	0.58	0.50	0.41	0.33	0.05	0.10
10,000	44.0% of DB size	0.63	0.54	0.45	0.36	0.27	0.04	0.09
10,000	48.0% of DB size	0.60	0.49	0.40	0.30	0.20	0.03	0.10
100,000	4.0% of DB size	0.97	0.97	0.96	0.95	0.95	0.82	0.32
100,000	8.0% of DB size	0.95	0.93	0.92	0.90	0.89	0.65	0.31
100,000	12.0% of DB size	0.92	0.90	0.88	0.86	0.84	0.55	0.31
100,000	16.0% of DB size	0.89	0.87	0.84	0.81	0.78	0.44	0.30
100,000	24.0% of DB size	0.84	0.80	0.76	0.72	0.68	0.29	0.29
100,000	44.0% of DB size	0.71	0.63	0.56	0.49	0.41	0.11	0.28
500,000	4.0% of DB size	0.98	0.97	0.96	0.96	0.95	0.83	0.33
500,000	8.0% of DB size	0.94	0.93	0.92	0.90	0.89	0.64	0.30

TABLE V
AVERAGE INFERENCE FOR THE THREE DIFFERENT SIZES OF THE DATABASE

DB Rows	Suppressed Cells	InfCLS	InfC+IS1/4	InfC+IS2/4	InfC+IS3/4	InfC+IS4/4	InfRLS	InfColLS
10,000	(4.0% of DB size)	1	0.80	0.67	0.57	0.50	0.15	0.04
10,000	(8.0% of DB size)	1	0.81	0.68	0.58	0.51	0.17	0.08
10,000	(12.0% of DB size)	1	0.80	0.66	0.57	0.50	0.18	0.12
10,000	16.0% of DB size	1	0.80	0.67	0.58	0.51	0.21	0.16
10,000	20.0% of DB size	1	0.80	0.67	0.57	0.50	0.23	0.19
10,000	24.0% of DB size	1	0.80	0.67	0.57	0.50	0.25	0.23
10,000	28.0% of DB size	1	0.80	0.66	0.57	0.50	0.27	0.27
10,000	32.0% of DB size	1	0.80	0.67	0.57	0.50	0.30	0.30
10,000	36.0% of DB size	1	0.79	0.66	0.56	0.49	0.33	0.33
10,000	40.0% of DB size	1	0.79	0.66	0.57	0.50	0.35	0.37
10,000	44.0% of DB size	1	0.80	0.67	0.58	0.51	0.38	0.41
10,000	48.0% of DB size	1	0.79	0.67	0.57	0.50	0.41	0.44
100,000	4.0% of DB size	1	0.80	0.68	0.57	0.49	0.14	0.04
100,000	8.0% of DB size	1	0.80	0.66	0.56	0.49	0.16	0.08
100,000	12.0% of DB size	1	0.81	0.65	0.55	0.48	0.17	0.11
100,000	16.0% of DB size	1	0.81	0.67	0.58	0.50	0.19	0.15
100,000	24.0% of DB size	1	0.80	0.67	0.58	0.51	0.23	0.23
100,000	44.0% of DB size	1	0.80	0.67	0.57	0.50	0.33	0.41
500,000	4.0% of DB size	1	0.80	0.66	0.57	0.50	0.15	0.04
500,000	8.0% of DB size	1	0.82	0.67	0.59	0.52	0.16	0.08

TABLE VI
DESCRIPTIVE STATISTICS, ANOVA RESULTS AND 95% CONFIDENCE INTERVAL FOR DATA ACCESSIBILITY AND INFERENCE

		N	Mean	Std. Error	95% Confidence Interval for Mean	
Data Accessibility	CLS	20	.82906780	.027156688	.77222820	.88590740
	CIS-1/4	20	.78616580	.034192829	.71459938	.85773221
	CIS-2/4	20	.74384959	.040704764	.65865354	.82904564
	CIS-3/4	20	.70121011	.047506553	.60177775	.80064247
	CIS-4/4	20	.65856678	.054277168	.54496336	.77217020
	RLS	20	.36893089	.062098725	.23895677	.49890502
(F = 29.90, p=0.0001)	COL	20	.19946966	.020438918	.15669051	.24224881
	Total	140	.61246581	.024456965	.56411004	.66082157
Inference	CLS	20	1.00000000	.000000000	1.00000000	1.00000000
	CIS-1/4	20	.80211405	.001361810	.79926375	.80496435
	CIS-2/4	20	.66781001	.001549836	.66456617	.67105386
	CIS-3/4	20	.57233208	.001627441	.56892580	.57573835
	CIS-4/4	20	.50077973	.001746617	.49712402	.50443544
	RLS	20	.23797563	.019109422	.19797915	.27797211
(F = 449.60,	COL	20	.20478830	.030232477	.14151100	.26806561
p=0.0001)	Total	140	.56939997	.023145531	.52363714	.61516280

TABLE VII
DATA ACCESSIBILITY

Duncan Test

Security Technique	N	Subset for alpha = .05			
	1	2	3	4	5
COL	20	.19946966			
RLS	20		.36893089		
CIS-4/4	20			.65856678	
CIS-3/4	20			.70121011	.70121011
CIS-2/4	20			.74384959	.74384959
CIS-1/4	20			.78616580	.78616580
CLS	20				.82906780
Sig.		1.000	1.000	.057	.057

Means for groups in homogeneous subsets are displayed.
Harmonic Mean Sample Size = 20.000.

TABLE VIII
INFERENCE

Security Technique	N	Subset for alpha = .05					
	1	2	3	4	5	6	1
COL	20	.20478830					
RLS	20	.23797563					
CIS-4/4	20		.50077973				
CIS-3/4	20			.57233208			
CIS-2/4	20				.66781001		
CIS-1/4	20					.80211405	
CLS	20						1.00000000
Sig.		.086	1.000	1.000	1.000	1.000	1.000

REFERENCES

[1] Oracle Virtual Private Database. "An Oracle Database 10g Release 2 White Paper", June 2005.
[2] Art Rask, Don Rubin, Bill Neumann. "Implementing Row and Cell Level Security in Classified Databases Using SQL Server 2005", MS SQL Server Technical Center, April 2005.
[3] Darryl "SQL Server 2005 Label Security Toolkit," Published November 16, 2006. URL: http://blogs.msdn.com/publicsector/archive/2006/11/16/sql-server-2005-label-security-toolkit.aspx
[4] "A Tradeoff Analysis between Data Accessibility and Inference Control for Row, Column, and Cell Level Security in Relational Databases," Azhar Rauf / Carol Keene, Bo I. Sanden, Elaine Waybright. Doctoral Thesis Published in January 2007 by Colorado Technical University.
[5] Dieter Gollmann, Computer Security, Copyright 1999 by John Wilsey and Sons Ltd., ISBN 0 471 97844 2
[6] Teresa F. Lunt, Dorothy E. Denning, Roger R. Schell, Mark Heckman, William R. Shockley, "The SeaView Security Model," IEEE Transactions on Software Engineering. VOL 16, NO 6. June 1990
[7] D. E. Denning. "A Preliminary Note on the Inference Problem in Multilevel Database Management Systems," In Proceedings of the National Computer Security Center Invitational Workshop on Database Security, June 1986
[8] Mathew Morgenstern, "Security and Inference in Multilevel Database and Knowledge-Base Systems," ACM 1987
[9] T. Su. Inferences in Database. Ph.D. Dissertation, Department of Computer Engineering and Science, Case Western Reserve University, August 1986
[10] T. Su and G. Ozsoyogiu. "Data Dependencies and Inference Control in Multilevel Relational Database Systems." In Proceedings of the IEEE Symposium on Security and Privacy, pp. 202-211, April 1987
[11] T. Su and G. Ozsoyoglu. "Multi-valued Dependency Inferences in Multilevel Relational Database Systems," In Database Security III: Status and Prospects, eds. D.L. Spooner and C. Landwehr, pp. 293-300, NorthHolland, Amsterdam, 1990
[12] C. Meadows and S. Jajodia. "Integrity Versus Security in Multi-Level Secure Databases." In Database Security: Status and Prospects, ed. Carl E. Landwehr, NorthHolland, Amsterdam, pp. 89-101, 1988
[13] Pierangela Samarati, "Protecting Respondents' Identities in Microdata Release," IEEE Trans. Knowl. Data Eng. 13(6): pp. 1010-1027 (2001).
[14] Kristen LeFevre, David J. DeWitt, Raghu Ramakrishnan, "Incognito: Efficient Full-Domain K-Anonymity", SIGMOD Conference 2005:pp. 49-60
[15] Ashwin Machanavajjhala, Daniel Kifer, Johannes Gehrke, Muthuramakrishnan Venkitasubramaniam, "L-diversity: Privacy beyond k-anonymity," TKDD 1(1): (2007)
[16] T. F. Keefe, M.B. Thuraisingham, and W.T. Tsai. "Secure Query Processing Strategies." In IEEE Computer. Vo. 22, #3, pages 63-70, March 1989
[17] M.B. Thuraisingham. "Security Checking in Relational Database Management Systems Augmented with Inference Engines," In Computers and Security, Vol. 6, pp. 479-492, 1987
[18] M.B. Thuraisingham, W. Tsai, and T. Keefe. "Secure Query Processing Using AI Techniques." In Proceedings of the Hawaii International Conference on Systems Sciences. January 1988
[19] M.B. Thuraisingham. "Towards the Design of a Secure Data/Knowledgebase Management System," In Data and Knowledge Engineering Journal. Vol. 5, #1, March 1990
[20] J.T. Haigh, R.C. O'Brien, P.D. Stachour, and D.L. Toups. "The LDV Approach to Security." In Database Security, III: Status and Prospects, ed. D.L. Spooner and C. Landwehr, North-Holland, Amsterdam, pp. 323-339, 1990
[21] J.T. Haigh, R.C. O'Brien, and D.J. Thompson. "The LDV secure relational DBMS model". In S. Jajodia and C Lawnwehr, editors, Database Security IV: Status and Prospects, pages 265-279. North Holland, 1991
[22] L.J. Buczkowski, and E.L. Perry. "Database Inference Controller Draft Top-Level Design." Ford Aerospace, July 1989
[23] S. Jajodia. "Aggregation and Inference Problems in Multilevel Secure Systems," In Proceedings of the 5th Rome Laboratory Database Security Workshop, June 1992
[24] L. S. Cox, S. McDonald, and D. Nelson. "Confidentiality Issues at the United States Bureau of the Census." In Journal of Official Statistics, Vol 2, No. 2, pp. 135-160, 1986
[25] L. S. Cox. "Practices of the Bureau of the Census with the Disclosure of Anonymized Microdata." In Forum der Bundesstatistik, pp. 26-42, 1987
[26] L. S. Cox. "Modeling and Controlling User Interface." In Database Security: Status and Prospects, ed. Carl E. Landwehr, North-Holland, Amsterdam, pp. 167-171, 1988
[27] D. E. Denning. "Cryptography and Data Security," Addison-Wesley, Reading, MA, 1982
[28] T.H. Hinke. "Inference Aggregation Detection in Database Management Systems." In Proceedings of the IEEE Symposium on Research in Security and Privacy, pp. 96-106, April 1988
[29] NCSC (National Computer Security Center) Technical Report, Volume 1/5, Library No. S-243,039, May 1996
[30] Transaction Processing Performance Council (TPC) Benchmark TM-H, Decision Support Standard Specification Revision 2.3.0, 1993-2005.