Query Evaluation Using Paging Module to Secure DB in SCPU for Privacy and Data Confidentiality

Mrs. Uma Maheshwari[1], Dr.P.Balakumar[2], Tamizharasan Senguttuvan[3] and Udayashankar Ramamoorthy[4]

[1] Assistant Professor, Department of Computer Science and Engineering

[2]Assosiate Professor,Department of Computer Science and Engineering

[3]Student, Department of Computer Science and Engineering

[4]Student, Department of Computer Science and Engineering

[1234] Prince Dr. K.Vasudevan College of Engineering and Technology, Chennai-600 048, India.

ABSTRACT: **Most information systems and business applications that have been built nowadays have a web frontend and they need to be available universally to clients, employees and partners all around the world, as this digital economy has its severe growth it is becoming more and more prevalent in the global economy. These web applications, that are capable of accessing from anywhere around the globe, it has also become exposed widely that any existing security vulnerability will most probably be uncovered and they are being exploited by hackers. Trusted DB is an outsourced database prototype that allows clients to execute SQL queries with privacy and under regulatory compliance constraints without having to trust the service provider. Trusted DB achieves this by leveraging server-hosted tamper-proof trusted hardware in critical query processing stages. SQL Queries allow attackers to access unauthorized data (read, insert, change or delete), gain access to privileged database accounts. In this paper, we thus propose to make trusted hardware a first-class citizen in the secure data management arena**.

Keywords: *SCPU (Secure Co-Processor), MD-5(Message digest 5), Data Privacy, Data Confidentiality.*

1. INTRODUCTION:

Numerous Scenarios of data leaks in the recent years have made the Clients to place their valuable data's in a third party provider without the assurance of providing the privacy and data confidentiality [19],[6]. Data privacy means the data owned by an individual will never be disclosed to anyone else. Whereas, the data confidentiality refers to the ability to share the sensitive data among a community of users. Thus it is tough to provide both the privacy and data confidentiality at the same instance of the time. Existing researches have made the way to provide the privacy and data confidentiality independently but not as a whole. We introduce Trusted DB [6], an outsourced database prototype that allows clients to execute SQL queries[2], with privacy and under the regulatory compliance of constraints by leveraging the server-hosted and tamper-proof trusted hardware in the critical stagesof query processing and thereby removing the limitations on the type of supported queries.

This provides the safety of the database from the hackers (intruders, insiders and administrators)[3].privacy is always a bit easier to implement rather than implementing the data confidentiality[5]. It is a tough task to implement both the security attributes to the database. There occur several problems with the modern day database that the hackers can hack a database by Bypass login[19] i.e. they hack the user login details by cracking the username or password and entering into the user database. Also, the administrator can also view the database which can be of a danger to the client data. There is a possibility of data leaks. This can only be controlled by providing the data confidentiality and the privacy at the same rate.

We have introduced the concept of placing the SCPU (secure co-processor)[6],outside the main database thus it doesn't have any limitation to the memory space of the database along with which the Concept of paging module is implemented for the ease of the database query parsing. The main objective is to keep database in a well secured manner under serious SQL injection attacks and to analyze the principle of SQL attacks. It provides safety methods to both users as well as administrators. The contributions of this paper are 1.) Provide privacy and data confidentiality to the Clients 2.) Reducing the process of hacking by the hackers to a certain extent. 3.) Providing the Transfer Secure Schema to the data.

2. PRELIMINARIES:

The problem of the hacking can be controlled by converting the username and password into the 16bit digit by using the MD-5 algorithm [13], whose output is converted into raw data and this raw data stored into the database so that the chance of bypassing the user account can be nullified.Hackers mainly target the code of the database where the username or the password of an user can be easily hacked this can degrade the database security and it can be rectified by saving the username and password not as such instead they are stored in the converted manner[9]. At the position of providing the information to a database there is a possibility of modifying the database or deleting its data's by code injection. This Possibility of modifying the database by SQL code injection can also be avoided by using the Bind variable method[13]. Whereas, by the usage of DBMS-Assert,Each Query Is Processed Word By Word Which Avoids Running Various Function.

Recent problems with the administrator viewing the data and compressing the privacy can be rectified by using the concept of wrapping[15], to provide the secure transfer of data and also to secure the data storage.

3. RELATED WORKS:

SQL injection attack still stands as an old reliable for attackers seeking to break into commercial databases.

SQL insertion is still out there for one easy inclination. It works!" says executive of IT safety and risk approach for Tripwire. As extended as there are so many weak Web applications with databases full of monetizable in sequence behind them, SQL injection attacks will persist."

3.1. WE'RE MAINTAINING JUICY TARGETS:

If the data in the SQL database is encrypted, and the encryption input is session anywhere else, that's a form of distributed security. Part of this issue is that SQL injection attacks[8],commonly provide way very priceless pay dirt. If organization did an improved job for reducing the target's attractiveness, then it would greatly reduce their incentives for those attackers to use that attack vector.

3.2. AT LEAST GO LEAST PRIVILEGE:

SQL injection attacks stay a defender desired because so various organizations create it a sure bet for attackers to drop deeper into the network following a winning injection attack. They do this by weakness to follow the rule of least privilege within application accounts.

3.3. SECURE CO PROCESSOR:

A secure co-processor is a dedicated chip which at earlier stages was implemented as a slave along with a processor, now a day they are used as a single processor that's acts as a master. This SCPU[6], provides us the flexibility to provide the data confidentiality to the database. The Trusted Database is an implementation of a secure coprocessor that provides the process trusted computing to ordinary PCs by enabling a secure environment. They are less power consuming and here the multiple cards can be used in parallel where their hardware chips provide fast hardware implementation of the algorithms[.

3.4. PAGING MODULE:

The paging module is used for the process of navigation of the page from the home page. This paging module is used to drag the pages from the database once the query is provided. Thus this enables us to place the SCPU outside the database and this also causes us to increase the data confidentiality to the data.

3.5. TYPES OF ATTACKERS:

There are several types attackers involved in attacking the database which damages the integrity of the database[9]. The main aim is to delete or modify the values in the database and causing problems to clients who have placed there valuable data in an third party server. The variable hackers are Intruders who are the external to the database whereas the insiders are those people who are present in the same database and cause damage to the database. The last type of hacking can be performed by the administrator i.e. the one who governs the database can also cause the damage to the database.

4 .PROBLEM FORMULATIONS:

In this section, we have formally defined the security primitives used in this paper to provide the ease of architecture a thereby enabling to provide the privacy and data confidentiality to the database.

4.1 MESSAGE DIGEST-5 ALGORITHM (MD5):

MD5 is a hash function which was developed by Rivest [12]; it generates a 128 bit long message and followed by a hash value. It is based on

the principle of Merkle-Damgard construction. This MD5 is not, like the MD-4 which is the forerunner of MD5, which is an iterating three-round hash function, but this has been expanded by a round among all other things. Thereby it said that the MD5 function has slightly decelerated in contrast to the MD4 function [3].

In summary, we may say that the essential differences between the two versions are, on the one hand, where the expansion of the compression function from 3 to 4 passageways and, on the other hand, where the addition of an additive constant per step and not hitherto 2 additive constants. Of course, still there are much more differences which, however, need a lot of understanding about the matter and they are not reasonable to be mentioned here. we have implemented this concept of MD5 in our paper for providing the 16bit digit output of numerical values.

Algorithms	***Key size/hash size(bits)***	***Extrapolated Speed (Kbytes/sec.)***	***PRB Optimized (Kb/sec)***
TEA	128	700	-
DES	56	350	7746
Triple-DES	112	120	2842
IDEA	128	700	4469
RSA	512	7	-
SHA	160	750	25162
MD5	128	1740	62425

Table1.Comparison of Performance of MD5

Initially, the message is raised to the multiple of 512 through the concept of padding. where the message should be congruent with 448, and modulo 512. Hence, this can be emanated from this that the message has always exactly 64 bit and is also a bit smaller than a multiple of 512 of exactly these 64 bit. Primarily, 1 bit and then an appropriate 0-bit sequence and the original message as 64-bit coded number are appended. Then, a cache consisting of the 4 registers A, B, C and D, where each of them having 32 bit, is initialized and setup. Now the first 512-bit block runs through the following four rounds consisting of 16 operations:

f (X,Y,Z) = X and Y or not (X) and Z

g (X,Y,Z) = X and Y or X and Z or Y and Z

h (X,Y,Z) = X or Y or Z

i (X,Y,Z) = Y or (X not (Z))

MD5 is an algorithm that is used to verify data integrity through the creation of a 128-bit message digest from data input (which may be a message of any length) that is claimed to be as unique to that specific data as a fingerprint is to the specific person. MD5, which was initially developed by Professor Ronald L. Rivestfrom MIT, which is intended for use along with digital signature applications, it requires that a large amount of files must be compressed by a secure method before that is being encrypted along with a code of secret key, under the public key cryptosystem. MD5 is currently a standard for conversion, Internet Engineering Task Force (IETF) Request for Comments (RFC) 1321.

According to this standard, it is "computationally infeasible" that two messages that have been provided as an input to the MD5 algorithm could have the same output as they have same message digest, or a false message can be created through apprehension of this message digest. MD5 is the third algorithm i.e. third message digest algorithm created by Rivest. All three (the others are MD2 and MD4) have their similar structures, but MD2 was optimized only for 8-bit machines, in comparison with the other two later formulas, which are all optimized for the 32-bit machines. The MD5 algorithm is obviously an extension of MD4, where the critical review is found to be fast, but not possibly they are absolutely secure. In comparison with others, MD5 is not quite as fast as the MD4 algorithm, but they offer much more assurance of data security.

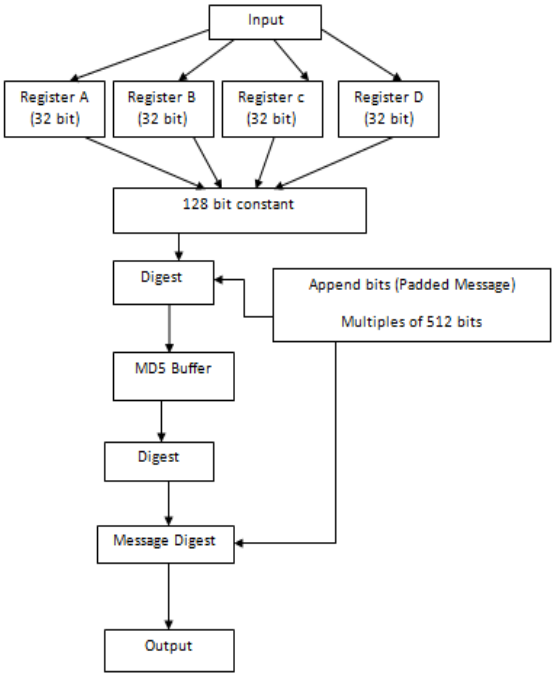

Figure 1.MD-5 Algorithm Structure

Step1: Appending Padding Bits:

At this phase the original message is padded so that the length is congruent to 448 of modulo512.Following are the Rules:

1) Original message is always padded with one bit "1" first.
2) Zero or more bits "0" are padded to bring the length of the message up to 64 bits fewer than multiple of 512.

Step 2: Append length

Here initially, 64 bits is appended along end of the padded message to indicate the length of the original messages in bytes. Following are the rules

1) Length of original message is converted to its binary format of 64 bits.
2) Break it into 2 words (32 bits each)
3) Lower order is appended first followed by second.

Step 3: Initializing MD5 Buffer

It requires 128 bit buffer with a specific initial value.

Buffer divided into 4words

A->0*67452301

B->0*EFCDAB89

C->0*98BADCFE

D->0*10325476

Step 4: Processing Message in 16 word blocks

For each input block 4 round of operations with 16 rounds in each round are performed

F(X, Y, Z) = (X AND X) OR (NOT XAND Z)

G(X, Y, Z) = (X AND Z) OR (Y AND NOT Z)

H(X,Y, Z) = X XOR Y XOR Z

I(X,Y,Z) = Y XOR (X OR NOT Z)

Algorithm: for K=1 to N

AA = A

BB =B

CC = C

DD =D

(X [0], X [1]……..X [15] =M [K])

A = A+AA

B =B+BB

C =C+CC

D =D+DD

T[1, 2 ….64] Array of special constants

T[i] = int (abs (sin ())*2*32)

M [1, 2 …N], Block of padded & Appended message

R1 (a, b, c, d, x, s, i)

Round 1: a = b+ ((a +f (b, c, d)+X+T[i]) <<S)

R2 (a, b, c, d, x, s, i)

Round 1: a =b+((a+G(b,c,d)+X+T[i])<<S)

R3 (a, b, c, d, x, s, i)

Round 1: a = b+((a+H(b,c,d)+X+T[i])<<S)

R4 (a, b, c, d, x, s, i)

Round 1: a=b+((a+I(b,c,d)+X+T[i])<<S)

Step 5: Content in Buffer words

A, B, C, D are returned in sequence with lower order byte first.

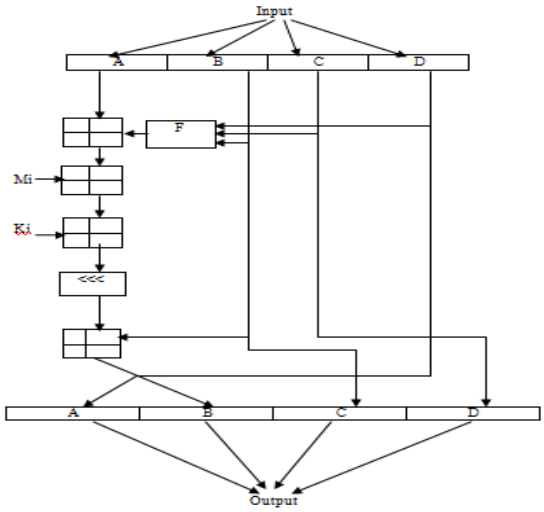

Figure 2.One MD5 Operation

The MD5 consists of a number of 64 of these operations, which are grouped under four rounds of 16 operations. Where F is an on linear function. M_i denotes a block of 32-bit input message and K_i denotes a constant of32-bit, where they are different for each other operations .Here the $\lll_s$ denotes the rotation of the left bit by S places; S varies with each operation.$\boxplus$Denotes an addition modulo 2^{32}.

4.2 RAWDATA:

Raw data (also known as primary data) is a term which is used for data collected from a source. These raw data has not been subjected to processing or any other manipulation, and are they also referred to as the primary data. This raw data can be input to a computer program and also used in manual procedures such as analyzing the statistics from a survey[1]. This term can also refer to the binary data which are present on the electronic storage devices such as hard disk drives (also referred to as low-level data).

In computing, raw data may have the following attributes: They might possibly contain errors, which are not validated in different (colloquial) formats; they might not be coded or unformatted; and also suspect, requiring the confirmation or citation. For example, the data input sheet may contain dates as raw data in many forms: "29th January 1994", "29/01/1994", "29/1/94", "29 Jan", or "today". Once when captured, these raw data could be processed stored in a normalized format, perhaps the Julian date, so as to be easier for computers and humans to interpret them during later processing.

Raw data (also called "sourcey" data or "eggy" data) are the data input to processing. Where a distinction is made at sometimes in between data and information to the effect. That this information is the end product of data processing. Raw data which has undergone processing are referred to as "cooked" data. Although these raw data has the potential to become "information,". The extraction, organization, and sometimes analysis and also the formatting for presentation are required for this to occur.

4.3 DBMS ASSERT:

SQL injection is a code injection technique that takes advantage of loose coding of database applications.

First Order Attack: Here The User enters injection code and gets a different result. Second Order Attack: The User injects code into the database, when it is run for the next time someone else will display that data. Blind or Inference: Here no information's are presented directly, although it is possible to infer the information based on the repeated results and also based on loose error trapping. These Repeated tests also allow information to be gathered. Whether it is the first number of the credit card. Compounded: Here the process of SQL Injection is used in conjunction with other techniques in order achieve a specific goal. For example, the goal may be to create a Denial of Service (DOS) attack.

4.4 WRAPPING:

The PL/SQL Wrapper process converts n PL/SQL source code into an intermediate form of the object code. Thus by hiding the application internals, this Wrapper prevents the Misuse of the application by other developers. Exposures of the algorithms to business competitors are also avoided. Wrapped code is as portable as source code [7]. The PL/SQL compiler recognizes it and loads the wrapped compilation units automatically without any prior information. Other advantages include Platform independence, Dynamic loading, and Dynamic binding, strict dependency checking, Normal importing and exporting.

4.5 BIND VARIABLE:

Bind variables are one of those Oracle Concepts that experts frequently cite as being key to Application Performance. Once a query is submitted the Oracle first checks in the shared pool to see whether the statement has been submitted before. If already executed the execution plan is retrieved and the SQL is executed.

If not executed earlier the Oracle has to parse the statement, work with various plans and finds an optimal path. Bind variable is created using VARIABLE command. They are referenced in PL/SQL by typing COLON followed immediately by the name of the variable.

5. CONSTRUCTION:

Providing both privacy and data confidentiality which is considered to be a tough task paved the way to create the new Architecture for the database.

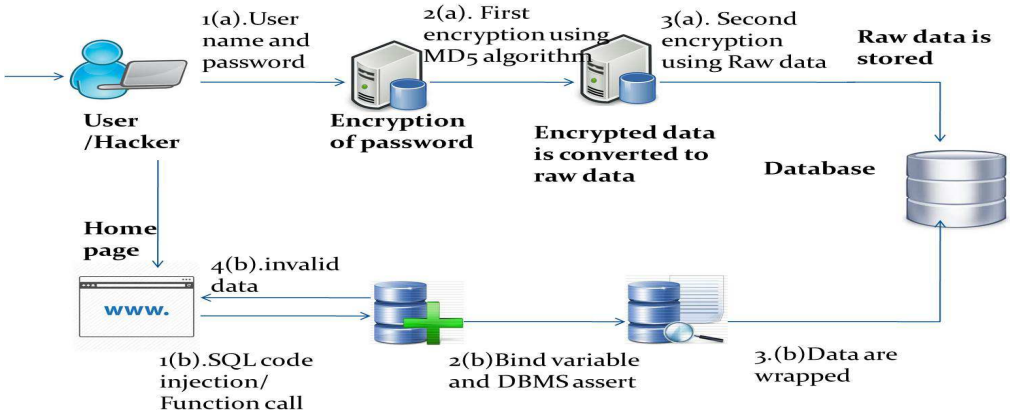

Figure 3. Architecture model of the Trusted DB including the Conversion of encrypted passwords into raw data

By construction of this database a Trusted DB is developed. The attributes in the database are always classified into public and private[10], where the public attributes are kept for the user purpose and the private attributes are encrypted and they can be decrypted only by the SCPU[7]. Here the DB is placed outside the SCPU hence it is not bounded to memory limitations and it keeps on Expanding with the data.

Paging module is used to pull the pages which are demanded by the SCPU for Query processing. Where Initially the DB is stored with data. Here the SENSITIVE keyword is used for Encryption. Query optimizer selects the best plan for execution by estimating the execution cost. The Dispatcher is used to dispatch the queries.

The initial phase of this model takes place with the user entering the values for the username and password. Once when he is completed entering the values the process of Encryption begins where the encryption takes place in two steps initial level of encryption begins with the conversion of the user entered values into a 16bit digit where the MD-5 algorithm is used for this purpose and this provides the output where the secondary phase of encryption is done at this position where this output is then converted into an alphanumeric value by using the raw data conversion methodology [9]. This converted value is then stored in the database. This end value which is stored in the database is encrypted and it is secured it cannot be accessed by any person. This process of encryption ends with this stage.

Once when a user enters the homepage by typing the URL, he will then provide the information which is necessary for login authentication; this process is continued by checking with the stored encrypted data in the database after which the authentication is provided. Following the authentication the users enters the main page where there is a possibility of SQL Code injection or function call injection can be performed. But this can be avoided by using the bind variable method where incase of code injection it will return a statement of invalid approach. Even though if they are a bit difficult to break this protection in case of exceptions this is overcome by using the concept of DBMS Assert[18], which provides and check up for the words using a word by word approach. Thus in case of finding any SQL queries in this, it will reply with an invalid statement. It is very much difficult to attack this database as we have the data wrapped [15]. Even the DBA doesn't have a chance to operate this code and also we have implemented a two database concept where a primary DB is kept as such and all other transactions are carried out with the help of the Secondary DB. Thus they involve in providing a transfer security to the database.

6. CONCLUSION:

We have provided Architecture for Trusted DB where both privacy and data confidentiality are involved at the same ratio and this architecture provide the advantage of avoiding the unauthorized access to any application and also addition of the SQL statements and the calling of an oracle function are also avoided by implementing this architecture. Thus they also provide the security to the database from both the hackers and also from the administrator of the database.

Thus involving this concept of using the SCPU as a master and not as an slave and also placing this SCPU out of the database provides comfort as there are no restrictions of memory usage.

7. FUTURE ENHANCEMENT:

As this concept of Trusted DB is now implemented in oracle 10g. In the near future this concept of Trusted DB can also be implemented in the cloud using the oracle 12c. It might provide better results for enabling privacy and data confidentiality to the database at low cost in the upcoming years. This might come in handy for the organizations who are building up their databases in cloud at that time.

Though the protection of database in a small area of network has been achieved using these concepts, it should be achieved in a wide area of network. The SQL protection as well as recovery of the information should be achieved in an easier manner. The disk space required for storing the data should be partly reduced. The information storage capacity of a database system should be enhanced without leaking out necessary information of a user. The information should be wrapped in a secured manner so that no one can access it. Likewise this database protection system must be enhanced in the nearly future.

References:

1. B. Bhattacharjee, N. Abe, K. Goldman, B. Zadrozny, C. Apte, V.R. Chillakuru, and M. del Carpio, "Using Secure Coprocessors for Privacy Preserving Collaborative Data Mining and Analysis," Proc. Second Int'l Workshop Data Management on New Hardware (DaMoN '06), 2006.
2. E. Mykletun and G. Tsudik, "Aggregation Queries in the Database-as-a-Service Model," Proc. 20th IFIP WG 11.3 Working Conf. Data and Applications Security, pp. 89-103, 2006.
3. R.A. Popa, C. Redfield, and N. Zeldovich, "Cryptdb: Protecting Confidentiality with Encrypted Query Processing," Proc. 23rd ACM Symp. Operating Systems Principles (SOSP '11), 2011.
4. V. Ganapathy, D. Thomas, T. Feder, H. Garcia-Molina, and R. Motwani, "Distributing Data for Secure Database Services," Proc. Fourth Int'l Workshop Privacy and Anonymity in the Information Soc. (PAIS '11), pp. 8:1-8:10, 2011.
5. Y. Chen and R. Sion, "To cloud or Not to Cloud?: Musings on Costs and Viability," Proc. Second ACM Symp. Cloud Computing (SOCC '11), pp. 29:1-29:7, 2011.
6. S. Bajaj and R. Sion, "TrustedDB: A Trusted Hardware Based Outsourced Database Engine," Proc. Int'l Conf. Very Large Data Bases (VLDB), 2011.
7. E. Mykletun and G. Tsudik, "Incorporating a Secure Coprocessor in the Database-as-a-Service Model," Proc. Innovative Architecture on Future Generation High-Performance Processors and Systems (IWIA '05), pp. 38-44, 2005.
8. N. Anciaux, M. Benzine, L. Bouganim, P. Pucheral, and D. Shasha, "GhostDB: Querying Visible and Hidden Data Without Leaks," Proc. 26th Int'l ACM Conf. Management of Data (SIGMOD), 2007.
9. L. Bouganim and P. Pucheral, "Chip-Secured Data Access: Confidential Data on Untrusted Server," Proc. 28th Int'l Conf. Very Large Data Bases (VLDB '02), pp. 131-141, 2002.
10. G. Aggarwal, M. Bawa, P. Ganesan, H. Garcia-Molina, K. Kenthapadi, R. Motwani, U. Srivastava, D. Thomas, and Y. Xu, "Two Can Keep a Secret: A Distributed Architecture for Secure Database Services," Proc. Conf. Innovative Data Systems Research (CIDR), pp. 186-199, 2005.
11. TPC-H Benchmark, http://www.tpc.org/tpch/, 2013.
12. MD5 Code Conversion technique http://nsfsecurity.pr.erau.edu/crypto/md5.html
13. Bind Variable technology implementation http://docs.oracle.com/cd/A81042_01/DOC/sqlplus.816/a75664/ch34.htm.
14. FIPS PUB 140-2, Security Requirements for Cryptographic Modules, http://csrc.nist.gov/groups/STM/cmvp/standards.html#02, 2013.
15. Wrapper classes for the primitive types in JAVA www.cs.rit.edu/~rlaz/cs2-20082/slides/WrapperClasses.pdf
16. IBM 4764 PCI-X Cryptographic Coprocessor, http://www-03.ibm.com/security/cryptocards/pcixcc/overview.shtml, 2007.
17. SCPU Validating the high performance csrc.nist.gov/nissc/1999/proceeding/papers/p16.pdf

18. DBMS Assert conceptual theory Implementation http://oracle-base.com/articles/10g/dbms_assert_10gR2.php
19. Sumeet Bajaj and RaduSion "TrustedDB: A Trusted Hardware-Based Database with Privacy and Data Confidentiality".

Mrs.UmaMaheshawari.B, Assistant Professor, Department Of Computer Science in Prince Dr.K.Vasudevan College Of Engineering And Technology ponmar, Chennai. She has completed M.Tech in SRM university. she is interested in the research areas such as Mobile adhoc network

Dr.P.BalaKumar, Associate Professor, Department of Computer Science in Prince Dr.K.Vasudevan College Of Engineering And Technology ponmar, Chennai. Has completed his Ph.D in the year 2011.He is interested in areas Networking and Cloud computing.

TamizharasanSenguttuvan is currently pursuing his B.E degree (final year) in Computer Science and Engineering from Prince Dr. K. Vasudevan college of Engineering and Technology in 2015.

UdayashankarRamamorthy is currently pursuing his B.E degree (final year) in Computer Science and Engineering from Prince Dr. K. Vasudevan college of Engineering and Technology in 2015.