ISSN No: 2348-4845
International Journal & Magazine of Engineering, Technology, Management and Research
A Peer Reviewed Open Access International Journal

OutSourced database Method execute SQL queries with privacy and Client Data Confidentiality

V. Archana
MTech Student
St.Mary's Group of Institutions,Hyderabad
Deshmukhi(v), Pochampally(M), Nalgonda(dist)

V. Goutham
Assistant Professor
St.Mary's Group of Institutions,Hyderabad
Deshmukhi(v), Pochampally(M), Nalgonda(dist)

Abstract: ***Outsourcing is the buzzword since last few years, as more and more cost and quality conscious businesses all over the world are turning to destinations like India for outsourcing their non-core business processes. Recent trend towards database outsourcing the information, as well as concerns and laws leading data confidentiality, have lead to huge importance in enabling secure database services. Preceding approaches to enabling such a service have been based on data encryption, causing a large overheading query processing. In TrustedDB is an outsourced database prototype that allows clients to execute SQL queries with confidentiality and under regulatory compliance constraints without having to trust the service provider. TrustedDB achieves this by leveraging server-hosted tamper-proof trusted hardware in dangerous query processing stages. TrustedDB does not perimeter the query expressiveness of supported queries.***

Keywords: ***Data Confidentiality, security, privacy, special-purpose hardware, Map-reduce, recursion.***

Introduction:

Outsourcing refers to the way in which companies entrust the processes of their business functions to external vendors. Any business process that can be done from an offshore location can be outsourced. This includes functions like transaction processing, payroll and order and inventory management to name a few. The most obvious and visible benefit relates to the cost savings that outsourcing brings about. You can get your job done at a lower cost and at better quality as well. Due to the difference in wages between western countries and Asia, the same kind of work that is done over there can be done in India at a fraction of the cost. Outsourcing your business processes would free your energies and enable you to focus on building your brand, invest in research and development and move on to providing higher value added services. Outsourcing eliminates the need for investment in infrastructure as the outsourcing partner takes the responsibility of the business processes and hence develops infrastructure for the same.

Security, Privacy and Confidentiality is a major concern for Outsourcing. Encryption is a well—recognized method for preserving the One of the basic, apparently confidentiality of sensitive information. Intrinsic, limitations of this technique is that an information system functioning with encrypted data can at most store or retrieve the data for the user; any more complex operations appear to require that the data be decrypted before being operated on. This limitation follows from the choice of encryption functions used, however, and although there are some truly inherent limitations on what can be accomplished, we shall see that it appears likely that there exist encryption functions which permit encrypted data to be operated on without preliminary decryption.

Presented research addresses several such security aspects, including access privacy and searches on encrypted data. In most of these efforts data is encrypted before outsourcing. Once encrypted however, inherent limitations in the types of primitive operations that can be performed on encrypted data lead to fundamental expressiveness and practicality constraints. Recent theoretical cryptography results provide hope by proving the existence of universal homeomorphisms, i.e., encryption mechanisms that

ISSN No: 2348-4845
International Journal & Magazine of Engineering, Technology, Management and Research
A Peer Reviewed Open Access International Journal

allow computation of arbitrary functions without decrypting the inputs. Ideas have also been proposed to leverage tamper-proof hardware to privately process data server-side, ranging from smartcard deployment in healthcare, to more general database operations. Generally trusted hardware is impractical because of its performance limitations and higher acquisition costs.

Computation inside secure processors is orders of magnitude cheaper than any equivalent cryptographic operation performed on the provider's unsecured common server hardware, instead of the overall greater acquisition cost of secure hardware.

This is so because cryptographic overheads (for cryptography that allows some processing by the server) are extremely high even for simple operations, a fact rooted not in cipher implementation inefficiencies but rather in fundamental cryptographic hardness assumptions and constructs. TrustedDB on the other hand utilizes secure, tamper resistant hardware such as the IBM 4764/5 cryptographic coprocessors deployed on the service provider's side to implement a complete SQL database processing engine. The TrustedDB design provides strong data confidentiality assurances.it does not limit query expressiveness.

TrustedDB enables the SCPU to transparently access external storage while preserving data confidentiality. Client queries are pre-processed to identify sensitive components to be run inside the SCPU. Non-sensitive operations are operated on to the untrusted host server. It greatly increase the performance and Minimize the cost of transactions.

The expense and performance limitations of trusted hardware, the costs of running TrustedDB are comparatively lower than any (existing or) cryptography mechanisms. The TrustedDB design gives strong data confidentiality assurances. It does not limit query expressiveness. The contributions of this work as follows

(i) the introduction of new cost models and insights that explain and quantify the advantages of deploying trusted hardware for data processing.
(ii) the design, development, and evaluation of TrustedDB, a trusted hardware based relational database with full data confidentiality.

Related Work:

In paper Map-Reduce Extensions and Recursive Queries Foto N. Afrati Vinayak Borkar ,Michael Carey discover the problems of implementing recursion, especially recursive queries, on a computing cluster. Author begin with the data-volume cost model, in which evaluate different algorithms for executing queries on a cluster, whether recursive or not. We have argued that extending the mapreduce model of cluster computing to recursive jobs requires considerable work.Thought must be given to the underlying algorithms,including the possibility that nonlinear recursions offer sub stantial advantages in this environment. Conversion of linear to nonlinear recursions is easy in some cases, such a transitive closure, but may be impossible in others. Major drawback is where exploring nonlinear recursion is possible and minimization of the data-volume cost for cases other than TC where the conversion is possible.

In paper Practical Server Privacy with Secure Coprocessors, S.W. Smith and D. Safford,2001 secure coprocessors provide a safe haven in which to execute code and carry out high speed cryptography. Authors are interested in providing a root secure retrieval service, using current secure coprocessor technology.In this paper model consists of a single server that has a number of secure coprocessors, and that provides a query service to a database containing records.Each record is stored as a unit on some suitable high-performance, but not necessarily secure, storage medium separate from the coprocessors. The major drawback of this is to reasonable performance of this scheme for large record sizes is the limitations imposed by the production level COTS devices: the components support sufficiently fast transfer and

ISSN No: 2348-4845
International Journal & Magazine of Engineering, Technology, Management and Research
A Peer Reviewed Open Access International Journal

encryption, but the configuration and firmware (as shipped) do not.

In Two Can Keep a Secret: A Distributed Architecture for Secure Database Services, G. Aggarwal, M. Bawa, P. Ganesan, H. Garcia-Molina, K. Kenthapadi, R. Motwani, U. Srivastava, D. Thomas, Y. Xu propose a new, distributed architecture that allows an organization to outsource its data management to two untrusted servers while preserving data privacy. To allow the client to partition its data across two, (and more generally,any number of) logically independent database systems that cannot communicate with each other. The partitioning of data is performed in such a fashion as to ensure that the exposure of the contents of any one database does not result in a violation of privacy. The client executes queries by transmitting appropriate subqueries to each database, and then piecing together the results at the client side.

EXISTING SYSTEM:

Existing research addresses several such security aspects, including access privacy and searches on encrypted data. In most of these efforts data is encrypted before outsourcing. Once encrypted however, inherent limitations in the types of primitive operations that can be performed on encrypted data lead to fundamental expressiveness and practicality constraints. Recent theoretical cryptography results provide hope by proving the existence of universal homeomorphisms, i.e., encryption mechanisms that allow computation of arbitrary functions without decrypting the inputs. Unfortunately actual instances of such mechanisms seem to be decades away from being practical

DISADVANTAGES OF EXISTING SYSTEM:

- Trusted hardware is generally impractical due to its performance limitations and higher acquisition costs. As a result, with very few exceptions, these efforts have stopped short of proposing or building full - fledged database processing engines.
- Computation inside secure processors is orders of magnitude cheaper than any equivalent cryptographic operation performed on the provider's unsecured server hardware, despite the overall greater acquisition cost of secure hardware.

PROPOSED SYSTEM:

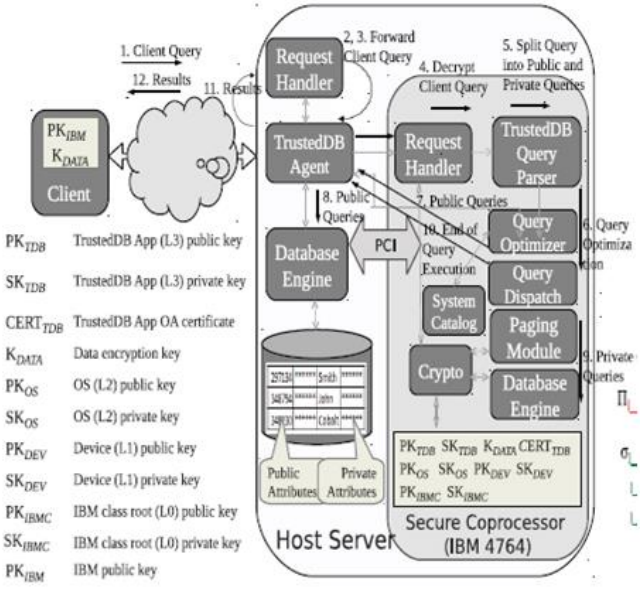
we posit that a full-fledged, privacy enabling secure database leveraging server-side trusted hardware can be built and run at a fraction of the cost of any (existing or future) cryptography-enabled private data processing on common server hardware. We validate this by designing and building TrustedDB, a SQL database processing engine that makes use of tamperproof cryptographic coprocessors such as the IBM 4764 in close proximity to the outsourced data. Tamper resistant designs however are significantly constrained in both computational ability and memory capacity which makes implementing fully featured database solutions using secure coprocessors (SCPUs) very challenging. TrustedDB achieves this by utilizing common unsecured server resources to the maximum extent possible. E.g., TrustedDB enables the SCPU to transparently access external storage while preserving data confidentiality with on-the-fly encryption. This eliminates the limitations on the size of databases that can be supported. Moreover, client queries are pre-processed to identify sensitive components to be run inside the SCPU. Non-sensitive operations are off-loaded to the untrusted host server. This greatly improves performance and reduces the cost of transactions.

ADVANTAGES OF PROPOSED SYSTEM:

1. The introduction of new cost models and insights that explain and quantify the advantages of deploying trusted hardware for data processing,
2. the design, development, and evaluation of TrustedDB, a trusted hardware based relational database with full data confidentiality, and
3. Detailed query optimization techniques in a trusted hardware-based query execution model.

ISSN No: 2348-4845
International Journal & Magazine of Engineering, Technology, Management and Research
A Peer Reviewed Open Access International Journal

SYSTEM ARCHITECTURE:



MODULES:

1. Query Parsing and Execution
2. Query optimization process
3. System Catalog
4. Analysis of Basic Query Operations

MODULES DESCRIPTION:

Query Parsing and Execution

In the first stage a client defines a database schema and partially populates it. Sensitive attributes are marked using the SENSITIVE keyword which the client layer transparently processes by encrypting the corresponding attributes:

CREATE TABLE customer (ID integer primary key, Name char (72) SENSITIVE, Address char (120) SENSITIVE);

(1) Later, a client sends a query request to the host server through a standard SQL interface. The query is transparently encrypted at the client site using the public key of the SCPU. The host server thus cannot decrypt the query. (2) The host server forwards the encrypted query to the Request Handler inside the SCPU. (3) The Request Handler decrypts the query and forwards it to the Query Parser. The query is parsed generating a set of plans. Each plan is constructed by rewriting the original client query into a set of sub-queries, and, according to their target data set classification, each sub-query in the plan is identified as being either public or private. (4)The

Query Optimizer then estimates the execution costs of each of the plans and selects the best plan (one with least cost) for execution forwarding it to the dispatcher.(5) The Query Dispatcher forwards the public queries to the host server and the private queries to the SCPU database engine while handling dependencies. The net result is that the maximum possible work is run on the host server's cheap cycles. (6) The final query result is assembled, encrypted, digitally signed by the SCPU Query Dispatcher, and sent to the client.

Query optimization process:

At a high level query optimization in a database system works as follows.

(i) The Query Plan Generator constructs possibly multiple plans for the client query.

(ii) For each constructed plan the Query Cost Estimator computes an estimate of the execution cost of that plan.

(iii) The best plan i.e., one with the least cost, is then selected and passed on to the Query Plan Interpretor for execution.

The query optimization process in TrustedDB works similarly with key differences in the Query Cost Estimator due to the logical partitioning of data mentioned above.

System Catalog:

Any query plan is composed of multiple individual execution steps. To estimate the cost of the entire plan it is essential to estimate the cost of individual steps and aggregate them. In order to estimate these costs the Query Cost Estimator needs access to some key information. E.g., the availability of an index or the knowledge of possible distinct values of an attribute. These sets of information are collected and stored in the System Catalog. Most available DBMS today have some form of periodically updated System Catalog.

ISSN No: 2348-4845
International Journal & Magazine of Engineering, Technology, Management and Research
A Peer Reviewed Open Access International Journal

Analysis of Basic Query Operations:

The cost of a plan is the aggregate of the cost of the steps that comprise it. In this section we present how execution times for a certain set of basic query plan steps are estimated.

Conclusion: Conclusion: Queries on encrypted data, Propose division of data into secret partitions and rewriting of range queries over the original data in terms of the resulting partition identifiers. This balances a trade-off between client and server-side processing, as a function of the data segment size, the propose using tuples-level encryption and indexes on the encrypted tuples to support equality predicates. The main contribution here is the analysis of attribute exposure caused by query processing leading to two insights. The attribute exposure increases with the number of attributes used in an index, and the exposure decreases with the increase in database size. Range queries are processed by encrypting individual Bþ Tree nodes and having the client, in each query processing step, retrieve a desired encrypted Bþ Tree node from the server, decrypt and process it. However, this leads to minimal utilization of server resources thereby undermining the benefits of outsourcing. Trusted DB, operates in an un trusted server model, where sensitive data are protected, both on disk and during processing. Data that are encrypted on disk but processed in clear (in server memory), compromise privacy during the processing interval. The disclosures risks in such solutions are analyzed also propose a new query optimizer that takes into account both performance and disclosure risk for sensitive data. Individual data pages are encrypted by secret keys that are managed by a trusted hardware module. The decryption of the data pages and subsequent processing is done in server memory.

References:

[1] Sumeet Bajaj, Radu Sion "TrustedDB: A Trusted Hardware based Database with Privacy and Data Confidentiality" - IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, VOL. 26, NO. 3, MARCH 2014

[2]. TPC-H Benchmark, http://www.tpc.org/tpch/, 2013.

[3]. IBM4764 PCI-X CryptographicCoprocessor,http://www03.ibm.com/security/cryptocards/pcixcc/overview.shtml, 2007.

[4]. 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 SecureDatabase Services," Proc. Conf. Innovative Data Systems Research(CIDR), pp.186-199, 2005.

[5]. Iliev and S.W. Smith, "Protecting Client Privacy with TrustedComputing at the Server," IEEE Security and Privacy, vol. 3, no. 2,pp.20-28, Mar./Apr. 2005.

[6]. M. Bellare, "New Proofs for NMAC and HMAC: Security WithoutCollision-Resistance," Proc. 26th Ann. Int'l Conf. Advances inCryptology, pp. 602-619, 2006.

[7]. B. Bhattacharjee, N. Abe, K. Goldman, B. Zadrozny, C. Apte, V.R.Chillakuru, and M. del Carpio, "Using Secure Coprocessors forPrivacy Preserving Collaborative Data Mining and Analysis,"Proc. Second Int'l Workshop Data Management on New Hardware(DaMoN '06), 2006.

[8]. M. Canim, M. Kantarcioglu, B. Hore, and S. Mehrotra, "BuildingDisclosure Risk Aware Query Optimizers for Relational Databases,"Proc. VLDB Endowment, vol. 3, nos. 1/2, pp. 13-24, Sept.2010.

[9]. FIPS PUB 140-2, Security Requirements for Cryptographic Modules,http://csrc.nist.gov/groups/STM/cmvp/standards.html#02,2013.

About Authors:

V. Archana