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

Performance and Cost Evaluation of an Adaptive Encryption Architecture for Cloud Databases

Mrs.Anipidi Kalyani
M.Tech Student,
Department of CSE,
KLR College of Engineering and Technology.

Mr.R.Adinarayana
Assistant Professor,
Department of CSE,
KLR College of Engineering and Technology.

Mrs.S.S.Madhavi
Associate professor,
Department of CSE,
KLR College of Engineering and Technology.

ABSTRACT:

A Cloud database management system (CDBMS) is a distributed database that delivers computing as a service instead of a product. Improving confidentiality of information stored in cloud database .It is an important contribution to cloud database. Data encryption is the optimum solution for achieving confidentiality. In some native method, encrypt the whole database through some standard encryption algorithm that do not allow the any sql operation directly on the cloud. This formal solution affected by workload and cost would make the cloud database service inconvenient. We propose a novel architecture for adaptive encryption of public cloud database. Adaptive encryption allow any sql operation over encrypted data. The novel cloud database architecture that uses adaptive encryption technique with no intermediate servers. This scheme provides cloud provider with the best level of confidentiality for any database workload. We can determine the encryption and adaptive encryption cost of data confidentiality from the research point of view.

Keywords:

Cloud database, confidentiality, encryption, adaptively, cost model.

I.INTRODUCTION:

The Database as a Service (DBaaS) [1] is a novel paradigms through which cloud providers offer the possibility of storing data in remote databases. The main concerns that are preventing the diffusion of DBaaS are related to data security and confidentiality issues [2]. Hence, the main alternative seems the use of cryptography, which is an already adopted solution for files stored in the cloud, but that represents an open issue for database operations over encrypted data. Cloud computing is the distribution of computing as a service somewhat than a product, whereby shared resources, software, and information are providing to computers and additional devices as a utility (like the electricity grid) over a network. Cloud computing offers computation, software, data access, and storage services that do not necessitate end-user information of the physical location and configuration of the system that transports the services.Counterparts to this concept can be drawn with the electricity grid, in which end-users ingest power without demanding to recognize the component devices or infrastructure requisite to offer the service. Cloud computing is dissimilar from hosting services and assets at ISP data center.

It is all almost computing systems are rationally at one place or virtual resources making a Cloud and user community retrieving with intranet or Internet. So, it means Cloud could reside in-premises or off premises at service provider location. There are kinds of Cloud computing like: 1. Public clouds 2. Private Clouds and 3. Inter-clouds or Hybrid Clouds In general there are two utmost common methods of network virtualization are protocol-based virtual networks (such as VLANs, VPNs, and VPLSs) and virtual networks that are deal with virtual devices (such as the networks linking virtual machines inside a hypervisor). In repetition, both forms can be used in combination. VLANs (Virtual LANs) are rational LAN's (Local Area Networks), grounded on physical LAN's. A VLAN can be shaped by partitioning a physical LAN into multiple logical LAN's (subnets) using a VLAN ID. On the other hand, numerous physical LAN's can task as a single logical LAN.

The segregated network can be on a single router, or multiple VLAN's can be on multiple routers just as multiple physical LAN's would be. A VLAN can be on a VPN. A VPN (Virtual Private Network) contains of numerous distant end-points (typically routers, VPN gateways of software clients) combined by some kind of tunnel over additional network, typically a third party network. Two such end points organize a 'Point to Point Virtual Private Network' (or a PTP VPN). Linking more than two end points by insertion in place a mesh of tunnels produces a 'Multipoint VPN'. A VPLS (Virtual Private LAN Service) is a particular type of Multipoint VPN.VPLS are distributed into Transparent LAN Services (TLS) and Ethernet Virtual Connection Services. A TLS directs what it receives, so it delivers geographic parting, but not VLAN sub netting. An EVCS adds a VLAN ID, so it delivers geographic parting and VLAN sub netting.

2. LITERATURE SURVEY:

Although data encryption seems the most intuitive solution for confidentiality, its application to cloud database services is not trivial, because the cloud database must be able to execute SQL operations directly over encrypted data without accessing any decryption key. An initial solution presented in [5] is based on data aggregation techniques [6] that associate plaintext metadata to sets of encrypted data. However, plaintext metadata may leak sensitive information and data aggregation introduces unnecessary network overheads. The use of fully homomorphic encryption [9] would guarantee the execution of any operation over encrypted data, but existing implementations are affected by huge computational costs to the extent that the execution of SQL operations over a cloud database would become impractical.

This approach is quite original because related papers evaluate the pros and cons of porting scientific applications to a cloud platform, such as [4] focusing on specific astronomy software and a specific cloud provider (Amazon), and [3] presenting a composable cost estimation model for some classes of scientific applications. Adaptive encryption architecture that is founded on an intermediate and trusted proxy. This tenant's component, which mediates all the interactions between the clients and a possibly untrusted DBMS server, is fine for locally distributed architecture. In the existing cost of cloud computing is computed by analyzing the cost of cloud computing from a provider's perspective.

The execution of SQL operations over encrypted data suffer from performance limits or require the choice of which encryption scheme must be adopted for each database column and SQL operations.

3.RELATED WORK:

Improving the confidentiality of information stored in cloud databases represents an important contribution to the adoption of the cloud as the fifth utility because it addresses most user concerns. Our proposal is characterized by two main contributions to the state of the art: architecture and cost model.Although data encryption seems the most intuitive solution for confidentiality, its application to cloud da-tabase services is not trivial, because the cloud database must be able to execute SQL operations directly over encrypted data without accessing any decryption key. Na¨ıve solutions encrypt the whole database through some standard encryption algorithms that do not allow any SQL operation directly on the cloud. As a conse-quence, the tenant has two alternatives for any SQL operation: downloading the entire database, decrypting it, executing the query and, if the operation modifies the databases, encrypting and uploading the new data; decrypting temporarily the cloud database, executing the query, and re-encrypting it. The former solution is affected by huge communication and computation overheads, and costs that would make the cloud data-base services quite inconvenient; the latter solution does not guarantee data confidentiality because the cloud provider obtains decryption keys. The right alternative is to execute SQL operations directly on the provider obtains the decryption key. An initial solution in This proposal is based on data aggregation techniques [8], that of encrypted data to allow data retrieval. However, plaintext information and data aggregation introduces unnecessary network.

4.ARCHITECTURE DESIGN:

The proposed system supports adaptive encryption methods for public cloud database service, where dis-tributed and concurrent clients can issue direct SQL operations. By avoiding an architecture based on one [10] or multiple intermediate servers between the clients and the cloud database, the proposed solution guarantees the same level of scalability and availability of the cloud service. Figure 1 shows a scheme of the proposed architecture where each client executes an encryption engine that manages encryption operations.

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

This software module is accessed by external user applications through the encrypted database interface. The proposed architecture manages five types of information.

•plain data is the tenant information;
•encrypted data is stored in the cloud database;
•plain metadata represent the additional information that is necessary to execute SQL operations on encrypted data;
•encrypted metadata is the encrypted version of the metadata that are stored in the cloud database;
•master key is the encryption key of the encrypted metadata that is distributed to legitimate clients.

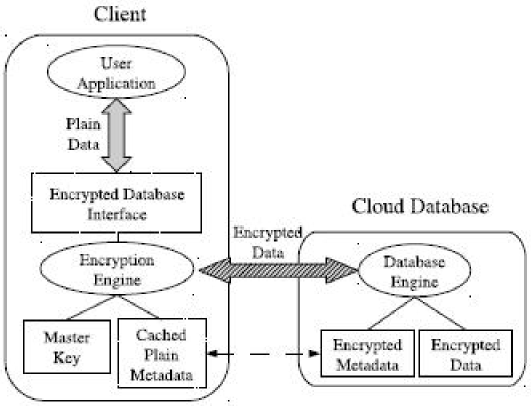

Fig. 1: Encrypted cloud database architecture

All data and metadata stored in the cloud database are encrypted. Any application running on a legitimate client can transparently issue SQL operations (e.g., SE-LECT, INSERT, UPDATE and DELETE) to the encrypted cloud database through the encrypted database interface. Data transferred between the user application and the encryption engine are in plain format, whereas infor-mation is always encrypted before sending it to the cloud database. When an application issues a new SQL operation, the encrypted database interface contacts the encryption engine that retrieves the encrypted metadata and decrypts it through the master key. In order to improve performance, the plain metadata are cached locally by the client as a volatile information. After obtaining the metadata, the encryption engine is able to execute the SQL operation on encrypted data, and then to decrypt the results. The results are returned to the user application through the encrypted database interface. As in related literature, the proposed architecture guarantees data confidentiality in a security model in which: the network is untrusted; tenant users are trusted, that is, they do not reveal information about plain data, plain metadata, and the master key; the cloud provider administrators are defined semi-honest or honest-but-curious [19], that is, they do not modify tenant's data and results of SQL operations, but they could be interested in accessing tenant's information stored in the cloud database. The remaining part of this section describes the adaptive encryption schemes (Section 3.1), the encrypted metadata stored in the cloud database (Section 3.2), and the main operations for the management of the encrypted cloud database (Section 3.3).

4.1 Adaptive encryption schemes:

We consider SQL-aware encryption algorithms that guarantee data confidentiality and allow the cloud database server to execute SQL operations over encrypted data. As each algorithm supports a specific subset of SQL operators, we refer to the following encryption schemes.
•Random (Rand): it is the most secure encryption (IND-CPA) [20], [21] because it does not reveal any information about the original plain value. It does not support any SQL operator, and it is used only for data retrieval.
•Deterministic (Det): it deterministically encrypts data, so that equality of plaintext data is preserved. It supports the equality operator.
•Order Preserving Encryption (Ope) [12]: it preserves in the encrypted values the numerical order of the original unencrypted data. It supports the compar-ison SQL operators (i.e., =, <, ≤, >, ≥).
•Homomorphic Sum (Sum) [13]: it is homomorphic with respect to the sum operation, so that the mul-tiplication of encrypted integers is equal to the sum of plaintext integers. It supports the sum operator between integer values.
•Search (Search): it supports equality check on full strings (i.e., the LIKE operator).
•Plain: it does not encrypt data; it is useful to support all SQL operators on non confidential data.

5.COST ESTIMATION OF CLOUD DATABASE SERVICES:

We consider a tenant that is interested in estimating the cost of porting its database to a cloud platform. This porting is a strategic decision that must evalu-ate confidentiality issues and the related costs over a medium-long term. For these reasons, we propose a model that includes the overhead of encryption schemes and variability of database workload and cloud prices. The proposed model is general enough to be applied to the most popular cloud database services, such as

Amazon Relational Database Service [23], EnterpriseDB [24], Windows Azure SQL Database [25], and Rackspace Cloud Database [26].

5.1 Cost model:

The cost of a cloud database service can be estimated as a function of three main parameters:
Cost = f (T ime, P ricing, Usage) (1)
where:
•Time: identifies the time interval T for which the tenant requires the service.
•Pricing: refers to the prices of the cloud provider for subscription and resource usage; they typically tend to diminish during T [27].
•Usage: denotes the total amount of resources used by the tenant; it typically increases during T .

In order to detail the pricing attribute, it is important to specify that cloud providers adopt two subscriptionpolicies: the on-demand policy allows a tenant to pay-per-use and to withdraw its subscription anytime; the reservation policy requires the tenant to commit in ad-between billing costs depending on resource usage and reservation costs denoting additional fees for commitment in exchange for lower pay-per-use prices [28]. Billing costs are billed periodically to the tenant every billing period TB. Moreover, if the tenant uses the reservation policy, the cloud provider requires the payment of the reservation cost at the beginning of each reservation period TR. An example of the relationship among T (three years), TR (one year) and TB (one month) is represented in Figure 4.

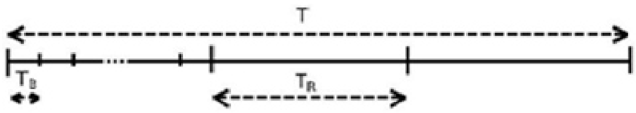

Fig. 4: Example of relationship among estimation (T), reservation (T_R) and billing (T_B) periods.

6.PERFORMANCE EVALUATION:

This section aims to verify whether the overheads of adaptive encryption represent an acceptable compromise from the performance point of view for guaranteeing data confidentiality in cloud database services. To this purpose, we design a suite of performance tests that allow us to evaluate the impact of encryption and adap-tive encryption on response time and throughput for different network latencies and for increasing numbers of concurrent clients.

The TPC-C standard benchmark is used as the workload model for the database services. The experiments are carried out in Emulab [34], which provides us with a set of machines in a controlled environment. Each client machine runs the Python client prototype of our architecture on a pc3000 machine hav-ing a single 3GHz processor, 2GB of RAM and two 10,000 RPM 146GB SCSI disks. The server machine hosts a da-tabase server implemented in PostgreSQL 9.1 on a d710 machine having a quad-core Xeon 2.4 GHz processor, 12GB of RAM and a 7,200 RPM 500GB SATA disk. Each machine runs a Fedora 15 image. The current version of the prototype supports the main SQL operations (SELECT, DELETE, INSERT and UP-DATE) and the WHERE clause expressions. We consider three TPC-C compliant databases having ten warehouses and a scale factor of five.

- Plaintext (PLAIN) is based on plaintext data.
- Encrypted (ENC) refers to a statically encrypted database where each column is encrypted at design time through only one encryption algorithm.

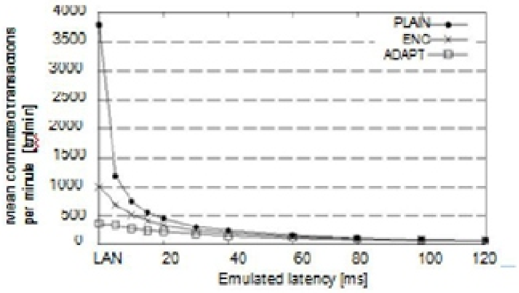

Fig. 5: TPC-C throughput with 5 clients

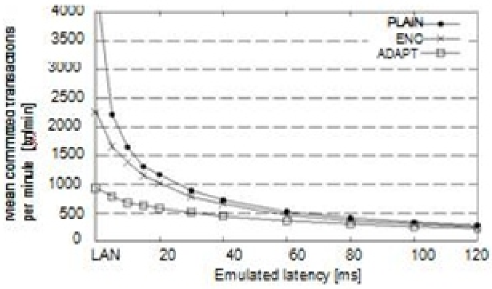

Fig. 6: TPC-C throughput with 20 clients

7.1. COST EVALUATION:

In this section we demonstrate the feasibility of the proposed cost model by applying it in the case of PLAIN, ENC and ADAPT configurations for real cloud data-base services.

We initially validate the usage estimation methodology presented in Section 4.3. We then analyze the variations of costs for different cloud providers and resource usages. We finally evaluate tenant's costs over a mid-term period equal to three years by considering realistic resource usage increments and price reductions.

7.2 Validation of the usage estimation:

To validate the usage estimation model, we perform several experiments by using the TPC-C benchmark.First we validate the storage estimation model. We de-ploy nine TPC-C compliant databases of three different sizes: 1, 5 and 10 warehouses (the number of warehouses is the TPC-C parameter that influences database size). For each size, we generate three database configurations: PLAIN, ENC and ADAPT. Results are summarized in Table 2. Estimated storage of PLAIN, ENC and ADAPT are calculated by using the analytical model presented in Section 4.3. For each estimated value, we report the esti-mation error with respect to the measured database size. Errors are expressed as a percentage. We observe that the proposed model always overestimates the database size. However, errors show that estimations are close to measured sizes. For PLAIN databases, the error is always below 2%, while for ENC and ADAPTS databases the error is always between 5% and 6%.

W	Estimated Storage [MB] (Error %)		
	PLAIN	ENC	ADAPT
1	99 (1.0)	187 (5.6)	273 (6.6)
5	453 (1.6)	859 (5.4)	1270 (6.3)
10	894 (1.5)	1698 (5.3)	2516 (6.2)

TABLE 2: Validation of storage overhead due to encryp-tion of TPC-C compliant databases.

	Network Usage [Bytes]		Error
	Estimated	Measured	
ENC	13175	13329	-1.2%
ADAPT	13671	13862	-1.4%

TABLE 3: Estimation of outgoing network due to data-base encryption over a TPC-C workload.

Now we validate the network estimation model. We de-ploy PLAIN, ENC and ADAPT TPC-C compliant databases of 10 warehouses. We observe that network consumption is invariant with respect to the number of warehouses, because it only depends on encryption and query workload. We measured the network usage of the PLAIN database, and we obtain an average of 7162 Bytes per transaction. By using Equation (8), we estimate np = k • 548. Hence, we determine k = 13.07. Then we use this value of k to determine the estimated network usage of ENC and ADAPT configurations. We compare these values with the experimentally measured network usages. Results are summarized in Table 3. Estimations are quite accurate, since we achieve errors of −1.2% and −1.4% for the ENC and ADAPT configurations, respectively. The validation demonstrates the efficacy of the pro-posed analytical usage estimation methodology in the TPC-C workload. Costs estimations proposed in the fol-lowing sections are based on the same usage estimations.

7.3 Analysis of cloud database costs:

We analyze cloud database costs with respect to different cloud provider offers and different storage and network usages. We consider a billing period equal to one month, and 24/7 availability (730 uptime hours per month).We initially estimate the monthly costs of a cloud database service in the PLAIN, ENC and ADAPT con-figurations with respect to a plaintext storage usage of 100 GB and a plaintext network usage of 100 GB. In Table 4, we report the results for the following cloud instances: Small, Large, and High Memory: Double Extra Large from Amazon RDS [28]; Premium P1 and Premium P2 from SQL Azure [31].

8. CONCLUSIONS:

There are two main tenant concerns that may prevent the adoption of the cloud as the fifth utility: data confi-dentiality and costs.

This paper addresses both issues in the case of cloud database services. These applications have not yet received adequate attention by the academic literature, but they are of utmost importance if we con-sider that almost all important services are based on one or multiple databases. We address the data confidentiality concerns by proposing a novel cloud database architecture that uses adaptive encryption techniques with no intermediate servers. This scheme provides tenants with the best level of confidentiality for any database workload that is likely to change in a medium-term period. We investigate the feasibility and performance of the proposed architecture through a large set of experiments based on a software prototype subject to the TPC-C standard benchmark. Our results demonstrate that the network latencies that are typical of cloud database environments hide most overheads related to static and adaptive encryption.

10.REFERENCES:

1.R. Buyya, C. S. Yeo, S. Venugopal, J. Broberg, and I.Brandic, “Cloud computing and emerging it platforms:Vision, hype, and reality for delivering computing as the5th utility,” Future Gener-ation Computer Systems, vol. 25, no. 6, pp. 599–616, 2009.

2.A. Boldyreva, N. Chenette, and A. O’Neill, “Order-preserving encryption revisited: Improved security analysis and alterna-tive solutions,” in Proc. Advances in Cryptology – CRYPTO 2011. Springer, Aug. 2011.

3.P. Paillier, “Public-key cryptosystems based on composite degree residuosity classes,” in Proc. Advances in Cryptology – EURO-CRYPT99. Springer, May 1999.

4.D. Song, D. Wagner, and A. Perrig, “Practical techniques for searches on encrypted data,” in Proc. IEEE Symposium on Security and Privacy., May 2000.

5.L. Ferretti, F. Pierazzi, M. Colajanni, and M. Marchetti,“Security and confidentiality solutions for public cloud database services,” in Proc. Seventh Int’l Conf. EmergingSecurity Information, Systems and Technologies, Aug. 2013.

6.A. Greenberg, J. Hamilton, D. A. Maltz, and P. Patel,“The cost of a cloud: research problems in data center networks,” SIGCOMM Comput. Commun. Rev., vol. 39, no. 1, pp. 68–73, Jan. 2008.

7.L. Popa, S. Ratnasamy, G. Iannaccone, A. Krishnamurthy, and Stoica, “A Cost Comparison of DataCenter Network Architec-tures,” in Proc. ACM Int’l Conf.Emerging Networking Experiments and Technologies, 2010.

8.R. N. Calheiros, R. Ranjan, A. Beloglazov, C. A. DeRose, and Buyya, “Cloudsim: a toolkit for modeling.

Author’s Details:

Mrs.Anipidi Kalyani, M.Tech Student, Department of CSE, KLR COLLEGE OF ENGINEERING AND TECHNOLOGY.

Mr.R.Adinarayana, working as an Asst professor in the Department of Computer Science and Engineering, KLR COLLEGE OF ENGINEERING AND TECHNOLOGY, JNTUH, Hyderabad.

Mrs.S.S.Madhavi, working as an Associate professor in the Department of Computer Science and Engineering, KLR COLLEGE OF ENGINEERING AND TECHNOLOGY, pursuing her Ph.D. in Computer Science and Engineering from JNTUH, Hyderabad. Her research area is Big Data.