

Literature Survey on Building Confidential and Efficient Query Processing Using RASP Methodology

Shreen Sumayya A. G¹, Aravind Gosh², Rajakumari³

^{1,2,3} Dept of CSE, Bharath University, Chennai, India

Abstract: Hosting data query services with the deployed cloud computing infrastructure increase the scalability and high performance evaluations with low cost. However, some data owners might not be interested to save their data in the cloud environment because of data confidentiality and query processing privacy should be guaranteed by the cloud service providers. Secured Query should be able to provide very high efficiency of query processing and also should reduce in-house workload. In this paper we proposed RASP data perturbation techniques which combine various objectives like random noise injection, dimensionality expansion, efficient encryption and random projection, henceforth RASP methodology is also used to preserve multidimensional ranges. KNN – R algorithm used to work with RASP range for processing KNN queries. The experimental result of our project carried out to define realistic security and threat model approaches for improved efficiency and security.

Keywords: KNN Query Services, Privacy Range Query, OPE (Order Preserving Encryption), OBDA (Ontology Based Data Access), and CRT (Cryptographic Transformation).

1. INTRODUCTION

Hosting data and intensively querying the process in the cloud environment because there is the unique content of scalability and low cost services providers. The services owners need to pay for the amount of time accessing the service providers, this methodology is an very high attractive feature, henceforth providing workloads within house infrastructure, however the service providers are able to lose the control over the information in the cloud services providers like IBM, Microsoft etc can be able to make a database which is difficult to detect and prevent cloud infrastructure. While there is a need for new approaches for preserving confidentiality and query privacy, therefore we should be able to provide high meaning process without slow query process.

Data owners should use the cloud environment for maintaining in-house infrastructure, so there should be an intricate relationship between query privacy confidentiality and compatibility, economically of using the cloud. Complexity for constructing the query services dramatically increases efficient processing of query in the real time, Query privacy, and data confidentiality. Random Space Perturbation (RASP) is a new method we proposed in this paper for developing range query and KNN query services in the cloud. The basic idea for the Random Space Perturbation is to transform the dataset with multidimensional analysis with a combination of random project, random noise, expansion of dimensionality and preserving order of encryption therefore utility of processing query ranges has been preserved, this transformation of the data carried out securely in the form of polyhedral in the Random Space Perturbation (RASP) data space. The key components used in the RASP framework includes unique combination of RASP perturbation with OPE (Order Preserving Encryption). The proposed techniques used to minimize workload of in-house processing by the high query precision and perturbation with low cost.

The layout of the paper is as follows. In section 2, we address the above mentioned techniques and also give a brief on the literature being reviewed for the same. Section 3, presents a comparative study of the various research works explored in the previous section. Lastly, we concluded in section 4 and section 5 is provided references.

2. RELATED WORKS

In this paper [1] Random Space Perturbation (RASP) method used to host query servicing in the cloud environment RASP used to meet CPEL criteria like efficient processing of query, Query Privacy, very low – in house workload and high confidentiality data transformation of the basic requirements for the low – house workload used to benefit the cloud computing and makes very accurate, efficient processing of query by the quality of query measure services. The Perturbation of RASP has been the unique combination of random projection, noise injection, expansion of dimensionality and OPE (Order Preserving Encryption) queries ranged to preserve perturbed space which allows efficient query processing.

In this paper [2] the main components like Optique's architectural descriptions gives out the information for the query processing for the end users in order to pose the queries mapping management techniques and ontology methodology are used for the installation. Components like query transformation used to rewrite the users queries in the cloud data servers. Processing components and distributed query optimization used to execute transformation of query. Ontology Based Data Access (OBDA) used for the execution of variable scalable access of data efficient query for the project to provide generic type of architecture.

In this paper [3] cheaper commodity machines and very large amount of scalable database for the storage purpose therefore cloud technology has been implemented for solutions. In the cloud data storage customer targeted for the data which has been outsourced (i.e the data which is not stored and not retrieved any sort of the information from customers own servers). From customers point of view the data should be accessed quickly and immediately without no time when the users has given the request and also the data stored on the cloud server for the efficient and effective transformation of query optimization. This paper helps to study about detail description of available query optimization techniques for the retrieval of data from the cloud storage.

In this paper [4] Ranging the queries is the most frequently used for the outline data analytics process which providers very much expensive for the data owners accessing cloud server. With the huge and high dimensionality of services like cloud computing became outsourced for the large amount of data from the database services providers hence it maintain effective and efficient amount of range query services, because of the outsourced service, the data owner who access the cloud infrastructure and data rich storage applications. Mostly encryption based approaches for cloud access linear scanning devices for the whole database which is usually in appropriated for the online data analytics for the huge amount of database, hence very few encryption focuses in the vulnerable prior of knowledge.

In this paper [5] content sharing and collaborative services allows the data owner's for sharing the private spatial data (eg. Some important key points displayed in the internet, tagging various business data) with the authentication and authorized users of the community. Spatial transformation of data done by the Enhanced HSD and Cryptographic Transformation (CRT), which offeres high range of query processing, confidentiality and increases communication between cloud server and local host. Data owner can choose any type of transformation technique which matches the requirements, trading between query efficiency and data privacy acquired in the cloud computing.

In [6] the authors described a newly developed data management technique is used in the cloud environment for fencing various cloud services providers, but most of the cloud servers are designed in a structured data management for accessing host system management configuration. SQL query language cannot directly support the traditional database and newly relational database methods; therefore a newly structured language has been created called RDMS (Relation Database Management System) for overall query performances. This paper gives detail information about efficiency of storage data in the cloud data management system. Efficient algorithm helps to address the operation of query processing a structured data.

In [7] the authors study about that cloud computing will grow for the accounting of horizontal scalability of different types of virtual machines which access over cloud server to the single virtual machine, hence there need and required of application software for scale – up and scale – down process which includes pay – for – use licensing model. Newly

developed infrastructure is needed for building up longer running of virtual machines over cloud infrastructure. This paper helps us to know about operational cost for matching performances analysis and rewarding energy proportionality by making some part of idle portions of the memory, network shall act in lower power mode of transformation.

In this paper [8] data perturbation a popular technique used in the cloud computing has been detailed studied data perturbation used for accessing privacy data mining methodology. The major part of data perturbation is to balance privacy and quality performance. Geometrical analysis of data perturbation consists of noise additions, translation perturbation and accessing multidimensional dataset which is able to provide details about privacy guarantee for modeling classification analog. This paper studies about preserving the accuracy model for the different types of classification models which includes kernel, different types of operating system and virtual system management.

In this paper [9] Encryption has been well – established making or creating technology which can be used for protecting sensitive types of data. Order – Preserving Encryption Scheme (OPES) used in the various types of numeric data by comparing encrypted cloud data, the newly updated value added without any changes in the encryption values of other service providers. Standard database values has been encrypted and easily integrated with database in the existing system. This paper shows OPES performance where query processing is deployed in the production of cloud environment.

In this paper [10] rapidly advancing technology used in the network, internet technologies have been widely desperate and used emerging cloud technology like “Software as the Service” for supporting industrial needs. “Database as a Service” used to create various type of application like creating the database, storing the data in the database and retrieving the data from cloud server from any part of the world at any time over the internet. Several types of challenges issued for the data privacy. This paper gives details about two types of privacy issues: 1.The owner of the database should make agreement with cloud service providers in order to protect thefts and hacking of the messages from the outsiders. 2. The data has to be protected from the cloud service providers because they are also agency whom the data owners can trust with a point.

3. COMPARTIVE STUDY

We have analyzed the various research works on several parameters and presented their comparison in the table below.

Table1. COMPARISON OF VARIOUS RESEARCH WORKS

| Sl. No | TITLE | AUTHOR | ISSUES | METHOD USED | TOOLS/LANG | ADVANTAGE/ DISADVANTAGE |
|--------|--|--|---|--|---|---|
| 1 | Building Confidential and Efficient Query Services in the Cloud with RASP Data Perturbation. | Huiqi Xu, Shumin Guo, and Keke Chen | VOL. 26, NO. 2, February 2014. | RASP (Random Space Perturbation). OPE (Order Preserving Encryption). CPEL Criteria. | Query Privacy. Cloud Computing. Data Transformation tools. | Advantages: 1. Host query processing in the cloud. 2. Very low – in house work load had been created. Disadvantages: 1. Arranging the queries in the preserved leads to the query processing. |
| 2 | Distributed Query Processing on the Cloud: the Optique Point of View | H. Kllapi, D. Bilidas, I. Horrocks, Y. Ioannidis, E. Jiménez-Ruiz, E. Kharlamov, M. Koubarakis, D. | <i>France. May 26-30, 2013</i> | Query Ontology Data Access (OBDA) Mapping management. Ontology Method. | Optique’s architectural tool Query optimization. | Advantages: 1. Query processing for the end user has been created successfully. 2. Queries can be rewritten in the cloud server using query transformation 3. Distributed query optimization used to transform query process. |

| | | | | | | |
|---|--|---|--|--|---|---|
| | | Zheleznyakov. | | | | Disadvantage: 1. Generic type of architecture used to access different type of variable access. |
| 3 | An Optimistic Approach for Query Construction and Execution in Cloud Computing Environment | P. Ravinder Rao, S. V. Sridhar, V. Ramakrishna | IJARCSS, vol. 3, Issue 5, May 2013 Issn: 2277 128X | Cloud data storage Cloud technology Cloud data server | Commodity Machines Scalable database Query optimization technique | Advantage: 1. Very large amount of cloud storage devices has been installed. 2. Data has been accessed quickly. 3. Stored data can be accessed in an efficient and effective way. Disadvantage: 1. Data which is not stored in the cloud server cannot be accessed by data owners. 2. Data has been retrieved without any cost of time when user given the request at any time from any part of the world. |
| 4 | Cloud Based SQL Query Processor. | Atul Phad, Swapnil Patil, Sujeet Purane, Vineet Patil | ISSN 2278-4721, Vol. 2, Issue 4 (February 2013), | Data owners accessing method Data service providers method Encryption based approach technique | Online data analytic tools Outsourcing method Cloud server | Advantage: 1. Data has been outsourced with the high dimensionality of services. 2. A very high range of data queries has been executed 3. Data has been encrypted based linear scanning devices. 4. Very rich storage of the data in cloud environment. Disadvantage: 1. Cloud processing is very much expensive. 2. Very few amount of encryption solution will be focused in the extracting prior knowledge. |
| 5 | Rasp: Efficient multidimensional range query on attack-resilient encrypted databases. | K. Chen, R. Kavuluru, and S. Guo | ACM Conference on Data and Application Security and Privacy, 2011 | Content sharing and collaborative services method Spatial transformation technique Cloud server and local host transformation method | Enhanced HSD Cryptographic Transformation (CRT) | Advantages: 1. Data owners able to share private spatial database like tagging, spamming. 2. Authentication and authorized user community has been developed. 3. Very high range of query processing has been communicated between the local host and cloud server. 4. Confidentiality has been increased. Disadvantage: 1. Trading the data between query efficiency and privacy data somewhat difficult during the networking congestion. |
| 6 | Enabling search services on outsourced private spatial data | M. L. Liu, G. Ghinita, C. S. Jensen and P. Kalnis | The International Journal of on Very Large Data Base, vol. 19, no. | Structured data management method Host system management | Data management technique Cloud | Advantages: 1. Newly developed data management technique has been created. 2. Enhanced various cloud service providers. |

| | | | | | | |
|----|---|--|---|---|---|--|
| | | | 3, 2010 | configuration RDMS query processing language | environment Cloud server and local host accessing tool. | 3. RDMS structured language been implemented. 4. Efficient algorithm used for the cloud environment. Disadvantages: 1. SQL query language cannot directly implement. 2. Traditional database and relational database cannot support data management technique. |
| 7 | ESQP: An Efficient SQL Query Processing for Cloud Data Management | Jing Zhao, Xiangmei Hu and Xiaofeng Meng | ACM 978-1-4503-0380-4/10/10 | Operational cost method Matching performance analysis method Scale – up and scale down process. | Java, Information Tools. Cloud Infrastructure tool Software Application | Advantages: 1. Horizontal scalability has been accessed by the virtual Machine. 2. Application software has been licensed for the usage. 3. Building up a longer running cloud services and virtual machine. 4. Performance increased. Disadvantage: 1. Only single virtual machine accessed through the cloud server. |
| 8 | Above the clouds: A berkeley view of cloud computing. | M. Armbrust, A. Fox, R. Griffith, A. D. Joseph, R. K. and Andy Konwinski, G. Lee, D. Patterson, A. Rabkin, I. Stoica, and M. Zaharia | Technical Report, University of Berkerley, 2009 | Data perturbation method Privacy data mining method Privacy Gurantee | Translation pertubation Multidimensiona l dataset Classification analog Virtual machine | Advantages: 1. Data perturbation is the popular method. 2. Accessing privacy data mining. 3. Balanced data privacy and quality performance. 4. Multidimensional dataset has been implemented. 5. Different types of operating system and virtual system management have been clubbed. Disadvantages: 1. Geometrical analysis causes noise addition. |
| 9 | Towards attack-resilient geometric data perturbation | K. Chen, L. Liu, and G. Sun | SIAM Data Mining Conference, 2007 | Sensitive data type method Standard database value Encrypted value Query processing method | Order – Preserving Encryption Scheme (OPES) Encrypted cloud data. Production performance tools. | Advantages: 1. Sensitive data types have been protected. 2. Various types of numeric data has been compared and encrypted in cloud. 3. Query processing in the cloud environment 4. Encrypted value services. Disadvantage: 1. Encryption of data takes slow process. 2. Identifying standard database takes time. |
| 10 | Order preserving encryption | R. Agrawal, J. Kiernan, | Proceedings of ACM SIGMOD | Software as the service Database as the | Network Topology | Advantages: 1. Internet technologies has been implemented in cloud. |

| | | | | | | |
|--|-------------------|-----------------------|------------------|---------------------------------------|--|--|
| | for numeric data. | R. Srikant, and Y. Xu | Conference, 2004 | service method Privacy Issues. | Internet Cloud server Cloud services providers Data owner | 2. Software as the service imported for the industrial needs. 3. Database as the service has been created. 4. Retrieving the data from the cloud server at any time. Disadvantage: 1. Data owners cannot trust cloud services providers all the time. |
|--|-------------------|-----------------------|------------------|---------------------------------------|--|--|

4. CONCLUSION

Cloud computing infrastructures are popularly used by peoples now a day. By using cloud users can save their cost for query services. The proposed RASP method with range query and kNN query is mainly used to perturb the data given by the owner and saved in cloud storage it also combines random injection, order preserving encryption and random noise projection and also it has contains CPEL(Data Confidentiality, query Privacy, Efficient query processing, and Low in house processing cost) criteria in it. By using the range query and kNN query user can retrieve their data's in secured manner and the processing time of the query is minimized. Case study is done on the related subject to improve the effect of query.

REFERENCES

- [1] Huiqi Xu, Shumin Guo, and Keke Chen, "Building Confidential and Efficient Query Services in the Cloud with RASP Data Perturbation", IEEE Transactions on knowledge and data engineering, VOL. 26, NO. 2, February 2014.
- [2] H. Kllapi, D. Bilidas, I. Horrocks, Y. Ioannidis, E. Jiménez-Ruiz, E. Kharlamov, M. Koubarakis, D. Zheleznyakov, "Distributed Query Processing on the Cloud: the Optique Point of View", In OWL Experiences and Directions Workshop (OWLED 2013). Montpellier, France. May 26-30, 2013.
- [3] P. Ravinder Rao, S. V. Sridhar, V. Ramakrishna, "An Optimistic Approach for Query Construction and Execution in Cloud Computing Environment", in IJARCSS, vol. 3, Issue 5, May 2013 Issn: 2277 128X
- [4] Atul Phad, Swapnil Patil, Sujeet Purane, Vineet Patil, "Cloud Based SQL Query Processor", in International Journal Of Engineering And Science, Issn: 2278-4721, Vol. 2, Issue 4 (February 2013), Pp 01-04.
- [5] K. Chen, R. Kavuluru, and S. Guo, "Rasp: Efficient multidimensional range query on attack-resilient encrypted databases," in ACM Conference on Data and Application Security and Privacy, 2011, pp. 249–260.
- [6] M. L. Liu, G. Ghinita, C. S. Jensen and P. Kalnis, "Enabling search services on outsourced private spatial data", The International Journal of on Very Large Data Base, vol. 19, no. 3, 2010.
- [7] Jing Zhao, Xiangmei Hu and Xiaofeng Meng, "ESQP: An Efficient SQL Query Processing for Cloud Data Management", in Proceedings of the second international workshop on Cloud data management. New York, NY, USA: ACM 978-1-4503-0380-4/10/10, Pages 1-8.
- [8] M. Armbrust, A. Fox, R. Griffith, A. D. Joseph, R. K. and Andy Konwinski, G. Lee, D. Patterson, A. Rabkin, I. Stoica, and M. Zaharia, "Above the clouds: A berkeley view of cloud computing," Technical Report, University of Berkerley, 2009.
- [9] K. Chen, L. Liu, and G. Sun, "Towards attack-resilient geometric data perturbation," in SIAM Data Mining Conference, 2007.
- [10] R. Agrawal, J. Kiernan, R. Srikant, and Y. Xu, "Order preserving encryption for numeric data," in Proceedings of ACM SIGMOD Conference, 2004.