Secure and Effective Spatial database Query Processing on the Cloud system

Mr. Ajay S. Sonawane
School of Computer Science and Engineering
Sandip University, Nashik, Maharashtra

Dr. Bhushan Chaudhari
School of Computer Science and Engineering
Sandip University, Nashik, Maharashtra

Abstract:
Cloud storage is a service model in which data is maintain, manage and backup remotely and made available to users over a network. Data Sharing is beneficial activity in cloud memory. The main challenge is maintaining data privacy with respect to untrusted parties i.e., cloud service provider, as well as providing relevant query results in real-time to authenticated users and it shows how to protectively, effectively scalable data is shared with others in cloud database. Existing approaches either compromise Confidentiality, security and privacy of the data or suffer from high communication cost between the server and the user. To overcome this problem, the proposed system uses dual transformation and encryption over spatial data. User fire encrypted queries on encrypted database present on a cloud server and encrypted results are returned to the user. The result is decrypted at the user end. Along with the database query, database user roles are mentioned which provides restricted database access to the user. The proposed system provides encryption level as well as database level security and requires only one round of communication for query execution.

Keywords: Cloud database, Spatial Databases, Data Encryption, Query Processing,

I. Introduction

In day to day life, large amount of data is generated in every industry. To manage this data requires high storage cost and computational devices. To manage this data at user end is difficult task with respect to the costing. To avoid this extra cost overhead many organizations and industry outsource their data to the third party data service provider. The service provider manages flat files as well as all types of databases.

Third party data service provider provides dynamic storage facility to the user. Using this facility user can upload high volume of data and increase the storage space as per the requirements. As the whole data is managed by the third party service provider, data security, confidentiality and integrity issues arises. Data Security requires the complete data existence. Whereas the data integrity assures that data is present in the uploaded form and not altered by any unauthorized user. The data confidentiality makes sure that data is not read by any un-trusted party.

To provide solution to the data confidentiality a cryptographic solution is provided. For database, data encryption can be provided but there is a problem of basic database operations. On encrypted database user cannot perform basic operations and lose the primary control.
Spatial database is database containing geometric information such as points, lines, polygon. The location based services uses spatial database. Navigational systems creates large amount of spatial information. This spacial database is outsourced to the third party cloud service provider. Location based systems are used by every user in his day to day life and do not want to disclose the location information to the third party.

For uploading such spatial database to the third party server, the database content should be encrypted. The encryption key should be kept at the user end to provide data confidentiality. The encrypted databases do not provide any underlying information and hence service provider is unable to perform any computational operations. For data retrieval, user needs to download the whole database, apply decryption and then user will be able to fire query on database. But this is not applicable for real time applications because large amount of data is present in a database and communication cost will be higher for each database query.

To avoid high communication cost, the queries need to be processed at the service provider's end by keeping the user data confidentiality. For this there a Homomorphic encryption scheme that encrypts the database data and provide operations on encrypted data.[2]. User can fire a query that will be entirely executed at the service provider's end over encrypted data. But such Homomorphic encryption scheme is inefficient and requires high computational overhead. Hence such scheme will not be applicable to the real life applications.

A secure spatial database service is required that preserves the user data confidentiality. A query processing over such spatial database is required that provide efficient query execution at the cloud end without burdening the end user. Along with the query execution the database query security need to be provided to the user to restrict the query access.

II. Literature Survey

Large amount of data is generated in variety of applications. To manage this data at user end is highly difficult task. At the very first time, Hacigumus et al. [3] proposed a technique to outsource the data to the third party service provider.

A. Anti-Tamper Hardware

An external hardware is proposed to provide security to the database outsourced to the external third party server [3]. This is middleware device or tamper-proof device. This device provides support in query processing. This device performs encryption and decryption of transmitted messages. Damiani et al. proposed searchable encryption technique at the server end. The tamper proof device is present at server end and an AES encryption scheme is applied for data encryption and decryption [4]. End user builds a b-tree over 1-dimensional data values and encrypts the record at node level. But it is impractical to have separate trusted device at server end for every authorized user [5].

B. Symmetric Cryptography Schemes

Yiu et al. [6] proposed R*-tree based cryptographic solution for 2-dimensional spatial data. R*-tree structure indexes the database and encrypts each node using AES encryption scheme. Service provider sends encrypted root node to the authorized user. The authorized user decrypted the descending tree nodes with secured encryption key. But this structure is used for static dataset and not able to handle dynamic updates.

Kim et al. [7] proposed a technique based on Hilbert-curve transformation. At the user end 2-dimensional spatial data is transformed in to 1-dimensional values followed by AES encryption. The whole encrypted file is transferred to the server. For each query processing the encrypted file is downloaded at the authorized user end and query get processed after decryption and re-instantiating the data in 2-dimensional space. But it is impractical to download whole encrypted data file every time. The data transfer is time consuming and network resource consuming.

C. Preserving Location Data Privacy

A data transformation technique is proposed by Yiu et al.[6]. The spatial data coordinates are transformed in different space using 3 different techniques. Hierarchical Space Division (HSD), Error-Based Transformation (ERB) and hybrid of HSD and ERB. These are shear and rotation based transformation techniques. No encryption technique is used.

D. Privacy and Integrity Guarantee

Ku et al. [11] provides a technique for data privacy and integrity. For data privacy encryption technique is used whereas for data integrity probabilistically replication method is used. For data privacy Hilbert-curve technique is proposed. Tian et al. [12] proposed index modification method for the standard Hilbert curve.

E. Partitioned Indexing Methods

To provide security as well as efficiency to the data owner indexing scheme is used. Wang et al. [8] proposed a r-tree based indexing scheme. Hierarchical encrypted index mechanism is used in this scheme with asymmetric encryption. This scheme uses leaf Minimum Bounding Rectangle to hide data ordering.

A. Talha, I. Kamel and Z. Aghbari [1] proposed 2-dimensional spatial dataset privacy protection over third party data services. This technique uses Hilbert-curve technique with AES encryption. The query execution is processed at server end and encrypted search result is returned to the user. The data privacy is preserved but user access right are not provided for user specific data access.

Following study represents detailed description of Hilbert-curve technique and AES encryption.

F. Hilbert-curve technique:

Hilbert-curve is space filling curve. It converts multidimensional data to 1 dimensional data. The hilbert curve passes portioned the whole space without making line intersection. The converted data preserves the spatial proximity of the original space. The multidimensional data do not follow order of dimensions. The requirements of conversion are

1. To create single point for every point in a space and
2. Preserve distance between those points i.e. nearby points present in space should be mapped near on the Hilbert curve.

The Hilbert curve has clustering property. It is widely used for dimensionality reduction and clustering.[9][10]. Recursive Hilbert curve construction provides better granularity. Following figure 1 represents the first three hilbert curve orders.

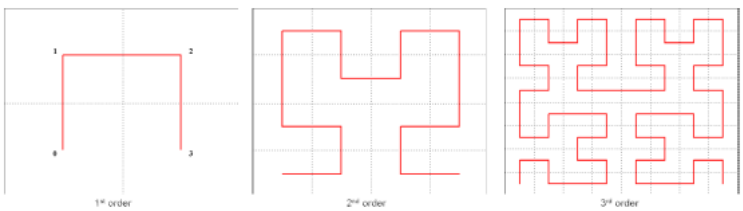

Figure 1: Hilbert Curve Orders in 2 Dimensions

Based on the Hilbert curve, Spatial points are indexed. Let N X N grid as a single cell. While iterating the grid, i^{th} grid is partitioned into 2^i X 2^i blocks having size 2n-1 X 2n-1. Each iteration replaces the grid by block with block size 2^{n-i-1} X 2^{n-i-1}. Then for each point, Hilbert cell values are assigned based on the generated curve. The grid is spanned according to the curve using the Hilbert Space Key (HSK) [11]. The HSK = {x0, y0, Θ, o}, where (x0, y0) is the curve's starting point, Θ is the curve's orientation and o

is the curve order. This is a one way transformation of data points i.e multidimensional data can be converted to single dimension value using HSK but form single HSK point, original dimensions cannot be retrieved. [7]

G. Advanced Encryption Standard

Advanced Encryption Standard [5] is a block cipher that is the NIST standard for symmetric key encryption, fast and highly secure. AES-256 is used here with a key size of 256 bits and encryption is done in 14 rounds. The AES uses the same key K to encrypt and decrypt the data. Also, deterministic AES allows equality comparisons to be made on the encrypted data.

III. Methodology

3.1 Preliminary:

1. Hilbert Space-Filling Curve:

Space-Filling Curves are used to convert multidimensional data to the one dimensional data. There is no order in multi-dimensional space. The spatial points contain multidimensional data. Hilbert curve is continuous curves that provides the spatial proximity of the original space. Hilbert curve uses clustering properties.[9][10]

2. Advanced Encryption Standard:

Advanced Encryption Standard is a block cipher. It uses symmetric key encryption technique. In symmetric encryption same key used for encryption and decryption. AES-256 uses 256 bit key for encryption and performs 14 rounds.[5]

3.2 System architecture:

Following figure 3.1 shows the system architecture. System has 3 entities: cloud, owner and Authorized user. The end users i.e. owner and Authorized user connects to the cloud server with http connection.

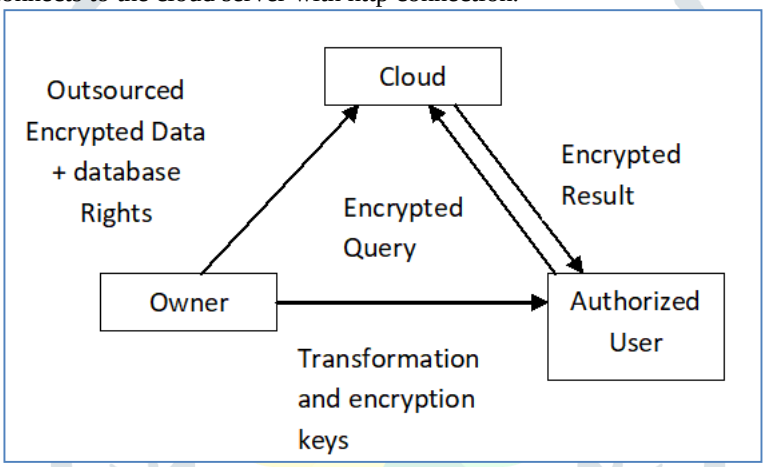

Figure 2: System Architecture

3.3 System Description:

The system includes 3 main entities: data owner, cloud i.e. service provider and authorized user. Data owner has spatial dataset containing 2-D points D = (d1, d2, …, ds). The dataset is normalized to convert the space unit square $[0; 1]^2$. These datapoints are transformed and then encrypted at the owner end and outsourced to the untrusted third party cloud server. Owner provides transformation and encryption keys to the authorized user. User can fire a query to the cloud service provider. Service provider executes query and returns the encrypted result to the authorized user. User decrypts the data and transform to the original space and then view the results.

Hilbert space-filling curve technique is applied to transform two dimensional data points in to one dimensional data. The data owner generates the Hilbert cell values and applies AES 256 symmetric encryption technique to hide the information. Authorized user converts 2 D point query to the 1 D Hilbert indices. This converted integer set is then encrypted and send the encrypted query to the cloud service provider. The cloud executes the query and returns the encrypted result. The result is then decrypted and original data points are extracted using Hilbert cell values.

IV. System Algorithm:

Algorithm 1: Hilbert Packet List Construction:

Input:
H: Hilbert key,
D={d1,d2,..,dn}:Spatial Data points,
K: Packet size,
Ek: Encryption Key
Output:
HPL = {p1,p2,..pp} where pi = [Ps,Pe,Pc] where Ps = Start Hilbert index of packet, Pe = End Hilbert index of packet, Pc= Spatial data points in the packet
Processing:

1. C= ø , Initialize Hilbert cell value set
2. For all d in D
3. Normalize d
4. Compute cx and add to C

5. End for
6. Sort packets in C in ascending order
7. For all c in C
8. Ps = Cx
9. While size of Ps<K
10. Add dj to Pc
11. End while
12. Pe = cy
13. End for
14. Encrypt Ps,Pe,Pc using encryption key Ek

Algorithm 2 Spatial Range Query

Input: Rectangle query: [(cx0,cy0), (cx1,cy1)]
HPL: Hilbert packet List
Output: Data points belonging to query R= {r1,r2,..rn}
Processing:

1. Q: Generate list of Hilbert cell present in query [(cx0,cy0), (cx1,cy1)]
2. QR={},Generate empty query list
3. For all q in Q
4. QR = Qr U Ek(q)
5. Sort QR in ascending Order
6. For all P in HPL
7. While QRi<Ps
8. if QRi Ɛ [Ps, Pe] then
9. R = R U Pc, Update R
10. Return R to Authorized User

V. Conclusion:

This works aims to provide data confidentiality to the end user for outsourced spatial 2D data on the third party cloud server. The system provides data encryption and encrypted query execution at the server end along with the database access rights to the authorized users. This system provides secure data sharing and access using third party cloud server with minimum communication cost. For spatial dataset Hilbert curve transformation is used.
The location information contains road traversal information shared using multiple lines or some area can be defined with polygon. In Future more than 2 dimensional data sharing can be performed.

REFERENCES

[1] Ayesha M. Talha, Ibrahim Kamel and Zaher Al Aghbari, "Facilitating Secure and Efficient Spatial Query Processing on the Cloud," in IEEE Transactions on Cloud Computing, pp. 1-1 July 2017.
[2] C. Gentry et al., "Fully homomorphic encryption using ideal lattices." in STOC, vol. 9, 2009, pp. 169–178.
[3] H. Hacig¨um¨us, B. Iyer, and S. Mehrotra, "Providing database as a service," in 18th International Conference on Data Engineering, 2002. Proceedings. IEEE, 2002, pp. 29–38.
[4] E. Damiani, S. Vimercati, S. Jajodia, S. Paraboschi, and P. Samarati, "Balancing confidentiality and efficiency in untrusted relational dbmss," in Proceedings of the 10th ACM conference on Computer and Communications Security. ACM, 2003, pp. 93–102.
[5] N. F. Pub, "197: Advanced encryption standard," Federal Information Processing Standards Publication, vol. 197, pp. 441–0311, 2001.
[6] M. L. Yiu, G. Ghinita, C. S. Jensen, and P. Kalnis, "Enabling search services on outsourced private spatial data," The VLDB Journal, vol. 19, no. 3, pp. 363–384, 2010.
[7] H.I. Kim, S.T. Hong, and J.W. Chang, "Hilbert-curve based cryptographic transformation scheme for protecting data privacy on outsourced private spatial data," in 2014 International Conference on Big Data and Smart Computing (BIGCOMP). IEEE, 2014, pp. 77-82.
[8] P. Wang and C. V. Ravishankar, "Secure and efficient range queries on outsourced databases using r-trees," in 2013 IEEE 29th International Conference on Data Engineering (ICDE). IEEE, 2013, pp. 314-325.
[9] B. Moon, H. V. Jagadish, C. Faloutsos, and J. H. Saltz, "Analysis of the clustering properties of the hilbert space-filling curve," IEEE Transactions on Knowledge and Data Engineering, vol. 13, no. 1, pp. 124-141, 2001.
[10] M. F. Mokbel, W. G. Aref, and I. Kamel, "Analysis of multi-dimensional space-filling curves," GeoInformatica, vol. 7, no. 3, pp. 179-209, 2003.
[11] W.S. Ku, L. Hu, C. Shahabi, and H. Wang, "Query integrity assurance of location-based services accessing outsourced spatial databases," in Advances in Spatial and Temporal Databases. Springer, 2009, pp. 80-97.
[12] F. Tian, X. Gui, P. Yang, X. Zhang, and J. Yang, "Security analysis for hilbert curve based spatial data privacy-preserving method," in 2013 IEEE 10th International Conference on High Performance Computing and Communications & 2013 IEEE International Conference on Embedded and Ubiquitous Computing. IEEE, 2013, pp. 929-934.