ISSN:0975-9646

Reshma Sultana N et al, / (IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 5 (3) , 2014, 4576-4581

Efficient Bucketization Technique for Multidimensional Range Queries over Encrypted Metering Data for Smart Grid

Reshma Sultana N[#1], Girish[*2]

[#]*II Year M.Tech ,Department of Computer Networking Engineering,*
The National Institute of Engineering,
Manadavady Road,Mysore-570008, INDIA
[*]*Associate Professor, Department of Computer Engineering & Applications,*
The National Institute of Engineering,
Manadavady Road,Mysore-570008, INDIA

***Abstract*— The term "Smart grid" can be described as the next generation electric power system. The metering data should be continuosly examined for valid testing and this is one of the challenge for our smart grid. In this paper we analyze the problem of supporting multidimensional range queries on encrypted metering data. The problem is motivated by the applications of secure data outsourcing where a residential user may store his data on a cloud server in an encrypted form and want to execute queries using server's computational capabilities. When financial auditing is needed, an authorized requester can send its range query tokens to cloud server to retreive the metering data. The solution approach is to compute a secure indexing tag of the data by applying bucketization(data partitioning) which prevents the cloud server from learning exact values but still allows server to check if a record satisfies the query predicate. Queries are evaluated in an approximate manner where the returned set of records may contain some false positives. In this scheme we can achieve the data confidentiality & query privacy because here only an authorized requester can be able to obtain the query results. Also this approach can significantly reduce communication & computation costs.**

***Keywords*— Range query, smart grid, privacy,encrypted data, metering data, outsourcing.**

I. INTRODUCTION

The term "smart grid" which are also known as "Modern grid" , "Intelligent grid", "GridWise" was introduced first in 2005 [1]. Before to smart grid Investigations revealed that the failure was due to load imbalance and lack of effective real-time diagnosis. Indeed, because electricity cannot be easily stocked, load must be matched by the power supply and transmission capacity in the electric power grid.

Recently, the concept of smart grid has emerged and been recognized as the next generation electric power system. Traditional grid is featured with centralized oneway transmission and demand-driven response. Smart grid combines traditional grid and information and control technologies. It allows decentralized two-way transmission and reliability- and efficiency driven response, and aims to provide improved reliability, sustainability, cosumer involvement and security. Smart grid is a system that not only supplies energy to end users but also allows the end users to contribute their energy back to the grid in future. Because of this reason we can call this smart grid as a two way communication device.The use of this robust two way communication increases the efficiency and reliability of Fig 1. The Conceptual architecture for Smart Grid power delivery and usage [2].The use of this smart gid commnication sytems is more nowadays to collect real-time metering data which is present at the Control center through a reliable communication network [3] as shown in Fig 1.

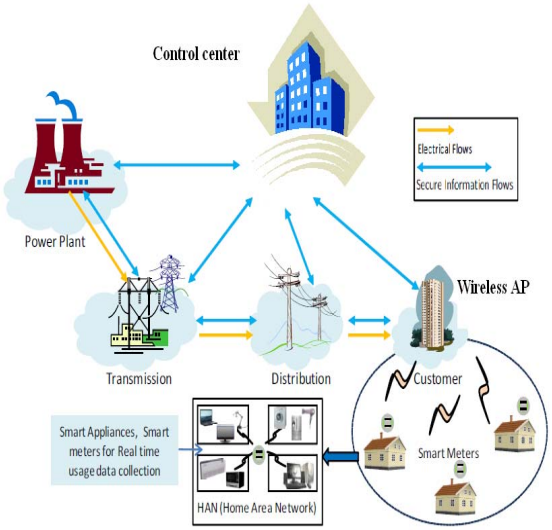

Fig. 1 The conceptual architecture of smart grid

In the smart grid, smart meters will be deployed in various residential places. These smart meters acts as two way communication devices [4][5] as discussed above. The job of these smart meters is to record periodically the metering data and report the same to wireless access point(AP). Then this wireles access point then collects the metering data and forward that to the control center. This Metering data should be audited periodically to present the billing and pricing statements fairly [6]. Various authorized requesters, market analysts will query these smart grid information systems for financial auditing, analysis, accounting and tax related activities [7]. If the information hacked by hackers the data confidentiality will no longer be preserved. For a given bucketization scheme, we derive cost

and disclosure-risk metrics that estimate client's computational overhead and disclosure risk. To keep the sensitive information private we should also achieve data confidentiality and query privacy in financial audit for smart grid.

However because the population is in increasing order this Metering data is increasing to a larger extent [8]. It is impossible for the control center to handle such a large amount of data alone. Also without any security aspects we were storing the data in control center. If any person in the network who is technically stronger means then he/she can login into the application and altered the information easily. Therefore we should relieve the burden of control center and also provide some security aspects so that our data will be in safe . In this approach, residential users can store their data on cloud servers and execute computation and queries using server's computational capabilities [9]. We cannot trust cloud servers because these might be untrusted and they intensionally share the sensitive information with other parties.Therefore in financial audit for smart grid achieving data confidentiality is very much important.

In addition , privacy is much more important in finanial auditing [10]. In particular, data from a single house would reveals the activities of a person who resides in a home , e.g.,when the individual person is in home and washing clothes [3] by washing machine. If an attacker can query these data, then data privacy might be destructed. Therefore, We should protect data confidentiality and privacy and we should enable only authorized users to query the sensitive data.

The requester , who manages the data query for financial auditing should frequently query this metering data by using either data ranges or geographic regions etc. and also if he query is sensitive then the requester should not exposed that to servers. Operating range queries with guaranteed query privacy is very much important for smart grid.

In this paper , we propose a Bucketization technique for multidimensional range query over encrypted metering data for smart grid communications. By the introduction of bucketization technique we addresses the problems regarding the data confidentiality and privacy. The main contributions of this paper are two-fold.

- Firstly , the data are first partitioned into buckets and encrypt the data with randomly generated session key and public key of control center followed by assignment of bucket-id as the tag for each data item in the bucket. A query posed by the requester is then translated suitably before issuing it to the cloud server who can evaluate it using only the information in the index tags corresponding to the data items.
- Secondly, the main goal here is to minimize the risk of disclosure and to acieve data confidentiality and query privacy and to reduce the communication and computation overhead and to shorten the response time.

The remainder of this paper is organized as follows. In section II, we investigate the related works. In section III, we introduce our system model, security requirements and our design goals. Then in section IV, we review preliminaries. In section V, we present performance evaluation. Finally, in section VI, we conclude this paper.

II. RELATED WORKS

A. Security and Privacy in Smart Grid

Authors P.sakarindr and N.Ansari [10] states that Security and privacy are critical to the development of wireless networks, especially in real-time financial auditing. Li [11] reviews the cyber security and privacy issues in smart grid and discusses some security aspects and privacy solutions for smart grid. The smart grid interpretability panel-cyber security working group [6] presents some guidelines for smart grid cyber security, including security strategy, architecture, and high-level requirements. Lu et al. [3] realizes a multi-dimensional data aggregation approach based on the homomorphic Paillier cryptosystem. Compared with the one dimensional data aggregation methods, EPPA can significantly reduce computational cost and significantly improve communication efficiency, satisfying the real-time high-frequency data collection requirements in smart grid communications. Li et al. [12] propose an authentication scheme based on merkle tree for smart grid. Acs and Castelluccia [13] proposed differentially private smart metering to exploit the privacy-preserving aggregation technique of time-series data in smart meters. X.Liang. [7] propose a usage-based dynamic pricing with privacy preservation for smart in a community environment, which enables the electricity price to correspond to the electricity usage in real time. Mi Wen.[27] Propose a Privacy preserving range query scheme over encrypted metering data for smart grid where with the help of Hidden vector encryption (HVE) data confidentiality and query privacy can be achieved. In short , few works on the query especially range query over encrypted data for smart grid which is very much significant in data auditing.

B. About Range Query

One of the popular studied approach for encryption is public key encryption with keyward search (PEKS) [14]. Constructing a PEKS is related to identity-based encryption (IBE) and this approach can protect user's data and query privacy. One of the most popular PEKS scheme called Searchable Encryption Scheme for Auction (SESA) [15] is suitable only for equality checks. Range query over the encrypted data with numeric attributes is much more difficult and in case of these we cannot achieve both data and query privacy at a time.

There are four solution categories that have been developed for range queries. Order- preserving encryption(OPE), bucketization (bucket), Hidden vector encryption (HVE) and special data structure traversal.

An order-preserving symmetric encryption (OPE) [16] scheme is a deterministic symmetric encryption scheme whose encryption algorithm produces ciphertexts that preserve numerical ordering of the plaintexts. This allows

direct translation of range predicate from the original domain to the domain of the ciphertext. However, R.agarwal [17] states that the coupling distribution of plaintext and ciphertext domain might be exploited by hackers to guess the hope of corresonding plaintext for a given ciphertext.

In the bucketization method, the data are first partitioned into buckets by using distributional properties of the datasets[18] and the bucket-id is set as the tag for each data item in the bucket. A query posed by the client is then translated suitably before issuing it to the server who can evaluate it using only the information in the index tags corresponding to the data items.

In the Hidden vector Encryption(HVE) approach [19], two vector over attributes are associated with ciphertext and token. Under the predicate translator, the ciphertext matches the token if and only if the two vectors are component wise equal. Several HVE schemes [20], [21], [22] have been proposed in literatures. All of them use bilinear groups equipped with bilinear maps, and each constructs a proper method to hide attributes in an encrypted vector. However, it is expensive to compute exponentiation and pairing in a composite-order group. A new HVE scheme proposed by Jong [20] which works well in prime order groups, but this scheme requires shorter token size and fewer pairing computations. But unfortunately Jong scheme failed in the smart grid applications where the data are high in dimension, variety or both.

Last category of a solution for range query is special data structrure traversal. Recently, Shi et al. [23] propose a searchable encrption scheme that supports multidimentional range queries over encrypted data (MRQED) to address the privacy concerns related to the sharing of network audit logs. This scheme allows a network gateway to encrypt summaries of network flows before submitting them to an untrusted repository. When network intrusions occur then it is the responsibility of an authority to release a key to an auditor, allowing the auditor

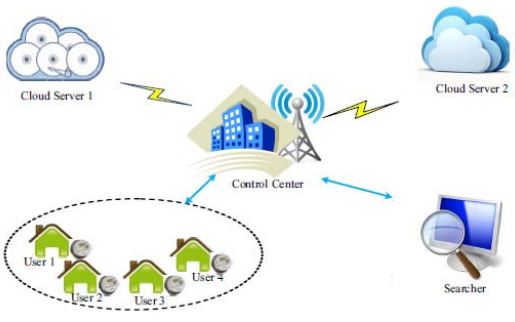

Fig. 2 Proposed system model

to decrypt flows whose attributes (e.g., source and destination addresses, port numbers, etc.) fall within specific ranges. Apart from network audit logs, this scheme can also be used for financial audit logs, medical privacy, untrusted remote storage, etc. In particular, this scheme enables investors to trade stocks through a broker in a privacy preserving manner. Financial audit logs contain sensitive information about financial transactions. Our MRQED scheme allows financial institutions to release audit logs in encrypted format. When necessary, an authorized auditor can obtain a decryption key from a trusted authority. With this decryption key, the auditor can decrypt certain transactions that may be suspected of fraudulent activities. However, the privacy of all other transactions are preserved.

III. System Model, Security Requiremnts And Design Goal

In this section, we formalize the system model, and identify the security requirements and our design goals.

A. System Model

Our focus here is how to outsource metering data from residential user to control center in an encrypted form and to operate multidimensional range queries over that encrypted metering data with control center(CC). Fig 2. Which consist of Control center (CC), searcher s, two cloud servers:CS1 and CS2 and residential users or data owners U={u1,u2,u3,….,un}.

First the smart meters will be placed in various residential areas and those smart meters will record metering data . Here the data owner is the residential user , who will partition the data first into data items (buckets) and encryting that metering data with the hep of randomly generated session key and encrypts again with the help of public key of control center(CC) before outsourcing the data o cloud servers. Later bucket-id is set as the (index) tag for each encrypted data item in the bucket wit the help of bucketization technique. In our system model there are two cloud servers CS1 and CS2. Cloud server 1(CS1) stores data ciphertexts and Cloud server 2(CS2) stores the key ciphertexts and indexes. We cannot trust these servers because it might share the sensitive information intentionally with third parties. We should assume that either of these servers might be compromised and controlled by an adversary. But the adversary cannot control both cloud servers. We can call our control center as the trusted proxy which can help not only the residential users to deposit their metering data to cloud servers but also generate query tokens for authorized requesters to get the data from the cloud servers. Then the authorized requester can get the session key from the cloud server 2(CS2) by depositing tokens which he got from control center.

The Control center consists of two main components:

- ciphertext forwarder and
- query translator.

Both of these components works within a secure environment. To protect the query privacy, the requester's query needs to be translated into two tokens, so that cloud server 2(CS2) can evaluate this query without seeing its original value.

B. Security Requirements

We identify the security requirements for our approach . In our system model, Control center is the only trustable component, and the data owners (users) U={u1,u2,u3,….,un} are also honest. However it is common that eavesdroppers or attackers attack the database

on cloud servers to steal the reports of individual users which contains sensitive information. Also these attackers can also launch some kind of active attacks to damage the data and query privacy. Therefore in order to prevent eavesdroppers from learning the sensitive data, the following security requirements should be satisfied in case of range query applications for smart grid.

- Data confidentiality: The residential user can either use symmetric or asymmetric encryption techniques to encrypt the metering data before outsourcing the same to cloud servers and should prevent the unauthorized users, including eavesdroppers and cloud servers for accessing the data which is sensitive.
- Data privacy: If the sensitive information which belongs to residential user's is able to access by unauthorized requesters , then data privacy will get violated. Therefore in our proposed scheme we should allow only authorized requesters to audit the data for financial purpose. Thus, only the authorized requester can be able to decrypt the encrypted metering data with the private key of contro center.
- Query privacy: As requesters intensionally or unintensionally prefer to keep their queries from being exposed to others. Therefore, the biggest deal is to hide their queries into tokens to protect the query privacy.

C. Designing Goal

To enable effective range query over encrypted metering data under our system model, our design goal is to develop a range query scheme over encrypted data for smart grid, and to achieve the security of the data and efficient range query as follows.

- Our proposed scheme should provide guaranteed security requirements. Smart grid is a device that does not consider the security. As a result of this residential user's data could be disclosed, and real time metering reports could be stealed by eavesdroppers. Therefore our proposed scheme should achieve the data confidentiality and data privacy and also the query privacy.
- Our proposed scheme should achieve performance efficiency. Also this scheme should reduce communication and computation costs.

IV. PRELMINARIES

In this section, we briefly describe the basic definations related to bucketization, which serves as the basis of our proposed scheme.

A. *Bucketization*

Hacigumus et al. [26] were the first ones to propose the bucketization-based data representation for query processing in an untrusted environment. It is simply a data partitioning step followed by assignment of a random (index) tag to each bucket effectively making every element within a bucket indistinguishable from another.When a query is issued by the requester, it is first determined which buckets intersect the query using the index tag stored on the client (a small amount of information) and all contents of the intersecting buckets are retrieved from the cloud server.

One important point to note that is this bucketization will oftenly suffers from unavoidable side-effects regarding privacy loss because labels (bucket id-s) disclose some information about the cleartext. Hore [24] and A.Ceselli [25] analyze and estimate the loss of privacy due to bucketization. These results show that, although some degree of privacy is invariably lost , only very limited information can be deduced from encrypted tuples and associated labels stated by A.Ceselli in "Modelling and assessing inference exposure in encrypted databases [25].

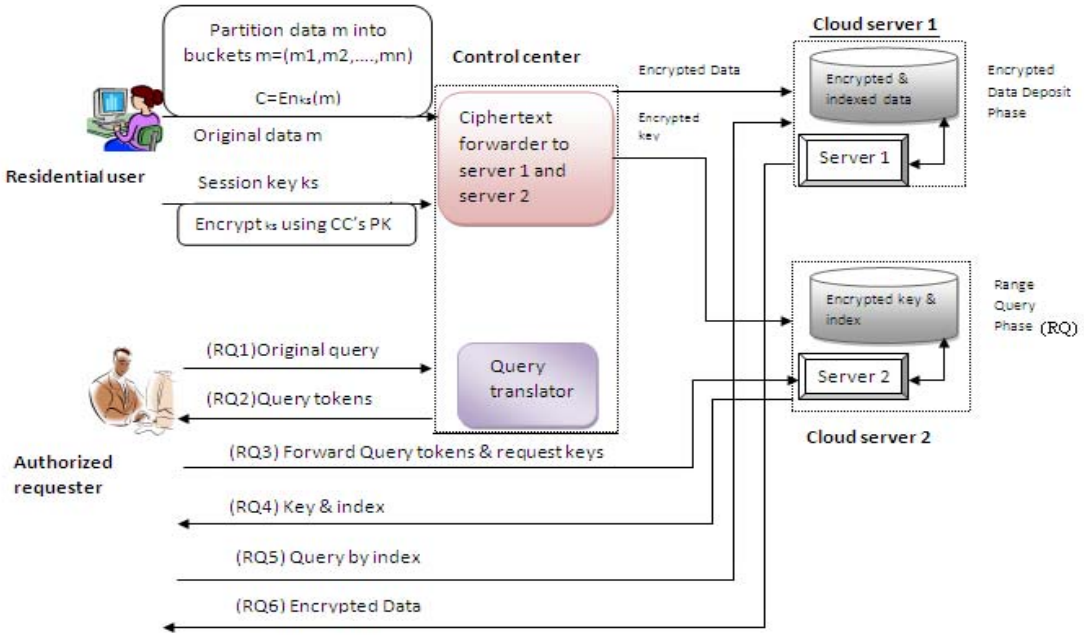

Fig. 3 Data query procedures

The main procedures of how the range query operates on the encrypted data in smart grid communications is illustrated in fig 3. The Control center consist of 2 main components.

a) Data forwarder

b) Query translator

The first half of the figure represnts a phase called Encrypted Data deposit phase (ED).

The second half of the figure represents a phase called Range Query phase (RQ).

In the ED phase, before outsourcing his/her data, the residential user or the data owner first partition the data into buckets. After partitioning the data , the residential user *ui* encrypts the data *m* into a ciphertext C with the help of randomly generated session key *ks.*Then residential user requests public key from trusted control center(CC) and encrypt the data again by that public key. Then the residential user assigns a bucket-id as the (index) tag to each data item in the bucket and forwards the encrypted indexed data and keys to Control center(CC) And CC forwards encrypted & indexed data ciphertexts to cloud server 1(CS1) and encrypted key index to cloud server 2(CS2).

As shown in fig.3 when an authorized requester S posts a range query, the query translated to query tokens by Control center's private key or by mapping functons before issuing it to the control server who can evaluate it using only the information in the index tags corresponding to the data items. Then the requester deposit those tokens to cloud server 2(CS2) and request for keys. Then the Control center (CS2) first determined which buckets intersect the query token using the index tag stored on the requester. Corresponding to that bucket, CS2 releases session key *ks* to the requester. Requester then forwards that session key *ks* and requests for encrypted data to cloud server 1(CS1). After getting the encrypted data from cloud server 1(CS1) requester can decrypt that encrypted data with the help of private key of Control center(CC).

V. Performance Evaluation

Multi-dimensional range query over encrypted data(MRQED) [23] is an encryption scheme that addresses the privacy issues regarding the network audit logs. However this scheme can also used for financial audit logs and medical privacy. An authority holding a master key can issue a search capability to an authorized party, allowing it to decrypt data entries whose attributes fall within specific ranges, while the privacy of other data entries is preserved.

In the Hidden vector encryption (HVE) technique [19], two vector over attributes are associated with ciphertext and a token. At a high level, the ciphertext matches a token if and only if the two vectors are component-wise equal. Also data confidentiality and query privacy will also get preserved.

In this paper, we can enhance the performance of Mi Wen [27] by the construction of bucketization method for range query where we minimized the risk of disclosure while keeping the cost of client's query below a certain user-specified threshold value. Data bucketization is set up as a cost based optimization problem where buckets are created so as to minimize the average number of false positives per query. One of the biggest advantage of bucketization framework is that expressive and complex queries can be evaluated relatively efficiently. Further, the implementation of bucket-based query evaluation is often much simpler than cryptographic protocols and also do not require any specialized algorithms to be executed on the server side. Finally, bucketization can actually be composed with many of the cryptographic techniques to enhance the confidentiality properties of searchable encryption schemes.

VI. Conclusion

In this paper we constructed a bucketization based range query predicate to realize the range query on encrypted metering data. This scheme allows residential users to deposit their metering data on cloud servers in encrypted form and allows to execute range queries by using server's computational capabilities. Authorized requester who will get authenticate first based on some ID-based authentication scheme will get the authorized tokens and can obtain the session keys to retreive the metering data within specific query ranges. This scheme successfully achieve data confidentiality and query privacy also get preserved. Performance evaluation shows that our proposed scheme can significantly reduce communication and computation overhead and also the response time. Future plan is to extend our aproach to support ranked range query with security and privacy preservation.

Acknowledgment

We thank Girish, an Associate professor, NIE, Mysore for his support and assistance throughout this project.

References

[1] Stefanos Stefanidis and Antonios Pitarokoilis, "Wireless Systems in Smart Grids" Project report for Wireless Systems (TSKS03) version 0.3.

[2] R. Zeng, Y. Jiang, C. Lin, and X. Shen, "Dependability analysis of control center networks in smart grid using stochastic petri nets," *IEEE Transactions on Parallel and Distributed Systems*, vol. 23, no. 9, pp. 1721–1730, 2012.

[3] R. Lu, X. Liang, X. Li, X. Lin, and X. Shen, "EPPA: An efficient and privacy-preserving aggregation scheme for secure smart grid communications," *IEEE Transactions on Parallel and Distributed Systems*, vol. 23, no. 9, pp. 1621–1631, 2012.

[4] C. Lo and N. Ansari, "Alleviating solar energy congestion in the distribution grid via smart metering communications," *IEEE Transactions on Parallel and Distributed Systems*, vol. 23, no. 9, pp. 1607–1620, 2012.

[5] ——, "Decentralized controls and communications for autonomous distribution networks in smart grid," *IEEE Transactions on Parallel and Distributed Systems*, vol. 4, no. 1, pp. 66–77, 2013.

[6] The Smart Grid Interoperability Panel-Cyber Security Working Group, "Nistir 7628 guidelines for smart grid cyber security: Smart grid cyber security strategy, architecture, and highlevel requirements," http://csrc.nist.gov/publications/nistir/ir7628/nistir-7628 vol1.pdf, August 2010.

[7] X. Liang, X. Li, R. Lu, X. Lin, and X. Shen, "UDP: Usage-based dynamic pricing with privacy preservation for smart grid," *IEEE Transactions on Smart Grid*, vol. 4, no. 1, pp. 141–150, 2013.

[8] R. Yu, Y. Zhang, S. Gjessing, C. Yuen, S. Xie, and M. Guizani, "Cognitive radio based hierarchical communications infrastructure for smart grid," *IEEE Network*, vol. 25, no. 5, pp. 6–14, 2011.

[9] C. Wang, Q. Wang, K. Ren, and W. Lou, "Privacy-preserving public auditing for data storage security in cloud computing," in *Proc. The IEEE International Conference on Computer Communications (INFOCOM' 10)*, 2010, pp. 1–9.

[10] P. Sakarindr and N. Ansari, "Security services in group communications over wireless infrastructure, mobile ad hoc, and wireless sensor networks," *IEEE Wireless Communications*, vol. 14, no. 5, pp. 8–20, 2007.

[11] X. Li, X. Liang, R. Lu, X. Shen, X. Lin, and H. Zhu, "Securing smart grid: cyber attacks, countermeasures, and challenges," *IEEE Communications Magazine*, vol. 50, no. 8, pp. 38–45, 2012.

[12] H. Li, R. Lu, L. Zhou, B. Yang, and X. Shen, "An efficient merkle tree based authentication scheme for smart grid," *IEEE Systems Journal*, to appear.

[13] G. Acs and C. Castelluccia, "I have a dream!(differentially private smart metering)," in *Proc. the 13th international conference on Information hiding*. Springer, 2011, pp. 118–132.

[14] D. Boneh, G. Di Crescenzo, R. Ostrovsky, and G. Persiano, "Public key encryption with keyword search," in *Proc. Advances in Cryptology (Eurocrypt'04)*, 2004, pp. 506–522.

[15] M. Wen, R. Lu, J. Lei, H. Li, X. Liang, and X. Shen, "SESA: An efficient searchable encryption scheme for auction in emerging smart grid marketing," *Security and Communication Networks*, to appear.

[16] A. Boldyreva, N. Chenette, and A. *O*,Neill, "Order-preserving encryption revisited: Improved security analysis and alternative solutions," in *Proc. Advances in Cryptology (CRYPTO'11)*. Springer, 2011, pp. 578– 595

[17] R. Agrawal, J. Kiernan, R. Srikant, and Y. Xu, "Order preserving encryption for numeric data," in *Proc. ACM international conference on Management of data (SIGMOD'04)*, 2004, pp. 563–574.

[18] B. Hore, S. Mehrotra, M. Canim, and M. Kantarcioglu, "Secure multidimensional range queries over outsourced data," *The VLDB Journal*, vol. 21, no. 3, pp. 333–358, 2012.

[19] D. Boneh and B. Waters, "Conjunctive, subset, and range queries on encrypted data," in *Proc. Theory of Cryptography Conference (TCC'07)*, 2007, pp. 535–554.

[20] J. Park, "Efficient hidden vector encryption for conjunctive queries on encrypted data," *IEEE Transactions on Knowledge and Data Engineering*, vol. 23, no. 10, pp. 1483–1497, 2011.

[21] J. Katz, A. Sahai, and B. Waters, "Predicate encryption supporting disjunctions, polynomial equations, and inner products," in *Proc. Advances in Cryptology (EUROCRYPT'08)*. Springer, 2008, pp. 146–162.

[22] V. Iovino and G. Persiano, "Hidden-vector encryption with groups of prime order," in *Proc. Pairing-Based Cryptography (Pairing'08)*. Springer, 2008, pp. 75–88.

[23] E. Shi, J. Bethencourt, T. Chan, D. Song, and A. Perrig, "Multidimensional range query over encrypted data," in *Proc. the IEEE Symposium on Security and Privacy (SP'07)*, 2007, pp. 350–364.

[24] B. Hore, S. Mehrotra, and G. Tsudik, .A privacypreserving index for range queries,. in *International Conference on Very Large Databases (VLDB)*, 2004.

[25] A. Ceselli, E. Damiani, S. Vimercati, S. Jajodia, S. Paraboschi, and P. Samarati, Modeling and assessing inference exposure in encrypted databases,. In *ACM Transactions on Information and System Security*, vol. 8, pp. 119.152, 2005.

[26] Hacıgümü¸s, H., Iyer, B., Li, C., Mehrotra, S.: Executing sql over encrypted data in database service provider model. In: SIGMOD (2002).

[27] Mi Wen, Rongxing Lu, Kan Zhang, Jingsheng, Xiaohui Liang, "PaRQ: A Privacy-preserving Range Query Scheme over Encrypted Metering Data for Smat Grid.