Available Online at www.ijcsmc.com

International Journal of Computer Science and Mobile Computing

A Monthly Journal of Computer Science and Information Technology

ISSN 2320–088X

IJCSMC, Vol. 4, Issue. 3, March 2015, pg.221 – 225

RESEARCH ARTICLE

DATA PRESERVING AND MINING USING RASP AND K-NN TECHNIQUES

Kumar Abhishek,
BE CSE- 4th year
abhishekatifet@gmail.com
IFET college of Engineering
Villupuram, Tamilnadu.

D.Ramesh,
BE CSE- 4th year
ramesh.da93@gmail.com
IFET college of Engineering
Villupuram, Tamilnadu

A. Kalaiselvi
M.Tech (Assistant Professor)
kalaiselvi270189@gmail.com
IFET college of Engineering
Villupuram, Tamilnadu.

ABSTRACT - ***The random space perturbation (RASP) approach to constructing practical range query and knearest-neighbor (kNN) query services in the database. The RASP perturbation is designed in such a way that the queried ranges are securely transformed into polyhedra in the RASP-perturbed data space, which can be efficiently processed with the support of indexing structures in the perturbed space. The basic idea is to randomly transform the multidimensional data sets with a combination of order preserving encryption, dimensionality expansion, random noise injection, and random project, so that the utility for processing range queries is preserved The RASP kNN query service (kNN-R) uses the RASP range query service to process kNN queries. The attacks on data and queries under a precisely defined threat model and realistic security assumptions. The proposed approach will address all the four aspects of the CPEL criteria and aim to achieve a good balance on them. Domain generation algorithm is being used for finding attacker or data or content modifier in database and it also blocks the account of user if the data is modified. The data owner exports the perturbed data to the database. Meanwhile, the authorized users can submit range queries or kNN queries to learn statistics or find some records.***

Index Terms— DG algorithm, kNN query, RASP technique

I. INTRODUCTION

The random space perturbation (RASP) approach to constructing practical range query and knearest-neighbor (kNN) query services in the database. The proposed approach will address all the four aspects of the CPEL criteria and aim to achieve a good balance on them. The basic idea is to randomly transform the multidimensional data sets with a combination of order preserving encryption, dimensionality expansion, random noise injection, and random project, so that the utility for processing range queries is preserved. The RASP perturbation is designed in such a way that the queried ranges are securely transformed into polyhedral. We also introduce the framework for constructing the query services with the RASP perturbation. We describe the algorithm for transforming queries and processing range queries., the range query service is extended to handle kNN queries. When describing these two services, we also analyze the attacks on the query privacy. The data owner exports the perturbed data to the database. Meanwhile, the authorized users can submit range queries or kNN queries to learn statistics or find some records.

II. RELATED WORKS

A. User Interface Design

The User Interface Design plays an important role for the user to move login the Application. This module has created for the security purpose. In this login page we have to enter user name and password, it will check username and password, if valid means directly go to home page, invalid username or password means show the error message and redirect to registration page. So we are preventing from unauthorized user entering into the login page to user page. It will provide a good security for our project.

A. Range-Query Processing

Based on the RASP perturbation method, we design the services for two types of queries: range query and kNN query. This section will dedicate to range query processing. We will first show that a range query in the original space can be transformed to a polyhedron query in the perturbed space, and then we develop a secure way to do the query transformation.

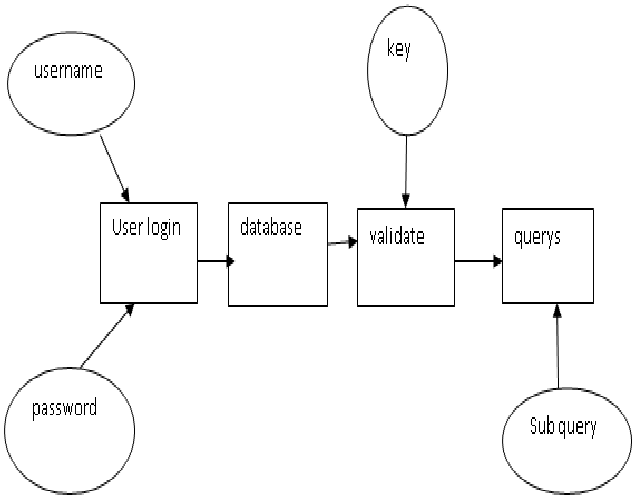
B. KNN Query Processing

The original distance-based kNN query processing finds the nearest k points in the spherical range that is centered at the query point.

The basic idea of our algorithm is to use square ranges, instead of spherical ranges, to find the approximate kNN results, so that the RASP range query service can be used.

In this query is given in the nearest neighbour of words, assuming a word as student, then user will put query as; ‘stu’ instead of typing full word.

By using this technique and it will search to the nearest neighbour of letter ‘stu’ and finally it will search the word starting with these three letter from the database.



C. Server Protecting Data

Server protect the original data from attackers they provide random space data for user the user will use the data without loss in original data from server. Any attacker to corrupt the data random space data will be loss so the server what happen to delete the corrupt data and insert new clone of the original data from database. Server protect the original data from attackers.

D. Data Confidentiality Analysis

As the threat model describes, attackers might be interested in finding the exact original data records or estimating them based on the perturbed data. Once the attacker is revocation that node will be eliminate from server. The server to analysis the user attribute and which user to use application in our server to maintain the user attribute and secure. And to relieve the attackers in service.

B. QUERY TECHNIQUES AND ALGORITHM

A. K nearest-neighbor (kNN-R).

The range-query-based kNN processing with 2D data. The Inner Range is the square range that contains at least k points, and the Outer Range encloses the spherical range that encloses the inner range. The outer range surely contains the kNN results but it may also contain irrelevant points that need to be filtered out. The sphere in between the outer range and the inner range covers all points with distances less than the radius r. Because the inner range contains at least k points, there are at least k nearest neighbors to the query points with distances less than the radius r. Therefore, the k nearest neighbors must be in the outer range.

B. Domain Generation Algorithm:

DGA is used to protect original data in database by blocking the attacker account and providing multiple path name of data when it attacked by a attackers. This algorithm is used to find the attacker or an user who modifies the content in the data base. It finds the user and it also blocks the account of the user permanently and maintains the information about the user in admin side. To read the data, user has to create an account in database by specifying there identification. If an authorized user tries to modify the data or to delete the data from database. This algorithm immediately blocks the account of the user and the content will be deleted in the database. Path name is being changed when data is modified and recovered, this all are performed automatically using this algorithm. To recover the content, now admin has to login his account and recover the data which is being deleted. In this way, the admin will be knowing the particular authorized user who tries to modify the content. **Domain generation algorithm** (DGA) are algorithms seen in various families of malware that are used to periodically generate a large number of domain names that can be used as rendezvous points with their controllers. Generating of
Domain name of attacker by its actual time of attacking and it will store to database in admin side.

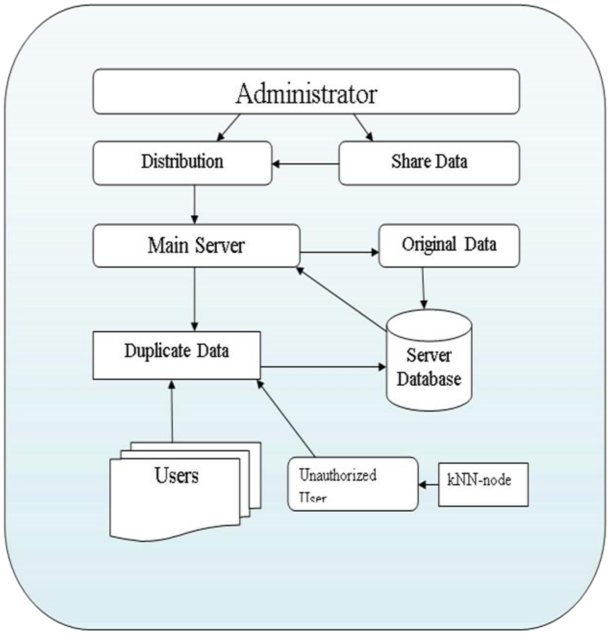
```
*// def generate_domain(year, month, day):
    """Generates a domain by the current date"""
    domain = ""

    for i in range(16):
        year = ((year ^ 8 * year) >> 11) ^ ((year & 0xFFFFFFF0) << 17)
        month = ((month ^ 4 * month) >> 25) ^ 16 * (month & 0xFFFFFFF8)
        day = ((day ^ (day << 13)) >> 19) ^ ((day & 0xFFFFFFFE) << 12)
        domain += chr(((year ^ month ^ day) % 25) + 97)

    return domain //*
```

The large number of potential rendezvous points makes it difficult for law enforcement to effectively shut down botnets since infected computers will attempt to contact some of these domain names every day to receive updates or commands. Every time an attacker wants to communicate with their malware, they choose a strike-time and a register the domain corresponding to that strike-time 24 hours before the time is hit. It can also combine words from a dictionary to generate domains using a web service through an web API.

C. System Architecture



The purpose of this architecture is to extend the proprietary database servers to the public database, or use a hybrid private-public database to achieve scalability and reduce costs while maintaining confidentiality. The trusted parties and the untrusted parties. The trusted parties include the data/service owner, the in-house proxy server, and the authorized users who can only submit queries. The data owner exports the perturbed data to the database. Meanwhile, the authorized users can submit range queries or kNN queries to learn statistics or find some records. The untrusted parties include the curious database provider who hosts the query services and the protected database. The RASP-perturbed data will be used to build indices to support query processing and if any authorized user is changing the content in database, by using the Domain generation algorithm, it identifies the user who modifies the data and it also blocks the account of user permanently.

D. CONCLUSION

RASP perturbation is a unique composition of OPE, dimensionality expansion, random noise injection, and random projection, which provides unique security features. It aims to preserve the topology of the queried range in the perturbed space, and allows to use indices for efficient range query processing. With the topology-preserving features, we are able to develop efficient range query services to achieve sub linear time complexity of processing queries. We then develop the kNN query service based on the range query service. The security of both the perturbed data and the protected queries is carefully analyzed under a precisely defined threat model. We also conduct several sets of experiments to show the efficiency of query processing and the low cost of in-house processing. Formally analyze the leaked query and access patterns and the possible effect on both data and query confidentiality. Further improve the performance of query processing for range query service. The security of both the perturbed data and the protected queries is carefully analyzed under a precisely defined threat model. The scheme is resilient to replay attacks. A writer whose attributes and keys have been revoked cannot write back stale information. It should be ensured that users must not have the ability to access data, even if they possess matching set of attributes.

E. REFERENCES

[1] R. Agrawal, J. Kiernan, R. Srikant, and Y. Xu, “Order Preserving Encryption for Numeric Data,” Proc. ACM SIGMOD Int’l Conf. Management of Data (SIGMOD), 2004.

[2] M. Armbrust, A. Fox, R. Griffith, A.D. Joseph, R.K. Andy Konwinski, G. Lee, D. Patterson, A. Rabkin, I. Stoica, and M. Zaharia, “Above the database: A Berkeley View of Cloud Computing,” technical report, Univ. of Berkerley, 2009.

[3] J. Bau and J.C. Mitchell, “Security Modeling and Analysis,” IEEE Security and Privacy, vol. 9, no. 3, pp. 18-25, May/June 2011.

[4] S. Boyd and L. Vandenberghe, Convex Optimization. Cambridge Univ. Press, 2004.

[5] N. Cao, C. Wang, M. Li, K. Ren, and W. Lou, “Privacy-Preserving Multi-Keyword Ranked Search over Encrypted Cloud Data,” Proc. IEEE INFOCOMM, 2011.

[6] K. Chen, R. Kavuluru, and S. Guo, “RASP: Efficient Multidimensional Range Query on Attack-Resilient Encrypted Databases,” Proc. ACM Conf. Data and Application Security and Privacy, pp. 249-260, 2011.

[7] K. Chen and L. Liu, “Geometric Data Perturbation for Outsourced Data Mining,” Knowledge and Information Systems, vol. 29, pp. 657- 695, 2011.

[8] K. Chen, L. Liu, and G. Sun, “Towards Attack-Resilient Geometric Data Perturbation,” Proc. SIAM Int’l Conf. Data Mining, 2007.

[9] B. Chor, E. Kushilevitz, O. Goldreich, and M. Sudan, “Private Information Retrieval,” ACM Computer Survey, vol. 45, no. 6, pp. 965-981, 1998.

[10] R. Curtmola, J. Garay, S. Kamara, and R. Ostrovsky, “Searchable Symmetric Encryption: Improved Definitions and Efficient Constructions,” Proc. 13th ACM Conf. Computer and Comm. Security, pp. 79-88, 2006.

[11] N.R. Draper and H. Smith, Applied Regression Analysis. Wiley, 1998.

[12] H. Hacigumus, B. Iyer, C. Li, and S. Mehrotra, “Executing SQL over Encrypted Data in the Database-Service-Provider Model,” Proc. ACM SIGMOD Int’l Conf. Management of Data (SIGMOD), 2002.

[13] T. Hastie, R. Tibshirani, and J. Friedman, The Elements of Statistical Learning. Springer-Verlag, 2001.

[14] B. Hore, S. Mehrotra, and G. Tsudik, “A Privacy-Preserving Index for Range Queries,” Proc. Very Large Databases Conf. (VLDB), 2004.

[15] Z. Huang, W. Du, and B. Chen, “Deriving Private Information from Randomized Data,” Proc. ACM SIGMOD Int’l Conf. Management of Data (SIGMOD), 2005.