ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 6, Issue 1, January 2018

Privacy Preserving By Third Party Auditing Using Frequent Itemset In Data Mining With Md5 Algorithm

D.J. Hani Mary Sheniha[1], G. Krithiga[2], S. Kalaivani[3],

Assistant Professor, Department of Computer Science and Engineering, Jeppiaar SRR Engineering College, Padur, Chennai, Tamil Nadu, India

B.E, Department of Computer Science and Engineering, Jeppiaar SRR Engineering College, Padur, Chennai, Tamil Nadu, India

B.E, Department of Computer Science and Engineering, Jeppiaar SRR Engineering College, Padur, Chennai, Tamil Nadu, India

ABSTRACT: This frequent itemset mining is promising to carry this computation intensive mining process. In supermarket the amount of work also transferred the approximate mining computation into the exact computation, where such methods not only improve the accuracy also aim to enhance the efficiency. In this paper, we propose a new framework for enforcing privacy in frequent itemset mining, where in supermarkets itemset data are both collected and mined in an encrypted form. We specifically design three secure frequent itemset mining protocols on top of this framework. To guarantee data privacy and computation efficiency, we adopt two different homomorphic encryption schemes and design a secure and effective comparison scheme. Our first protocol achieves more efficient mining performance while our second protocol provides a stronger privacy guarantee. In order to further optimize the performance of the second protocol, we leverage a minor trade-off of privacy to get our third protocol.

KEYWORDS: Data Security, Data Integrity, Data Protection, Association rules, Data Mining.

I. INTRODUCTION

In this instance of the problem, the data owner (the target database) seeks to minimize the load on the source databases. Security concerns prevent source databases to allow access to their transaction logs or to alter their database schema or to add triggers. Techniques that are more conducive to the model pro-posed in this work are described in these are all Apriori based. The problem of maintaining association rules in a centralized environment is studied. Maintenance of association rules in a distributed environment is studied.

However, incremental techniques to mine association rules with privacy protection in a fully dynamic centralized database have not been researched in the past. Nor does the literature contain work on incrementally mining privacy preserving association rules using target-only resources.

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 6, Issue 1, January 2018

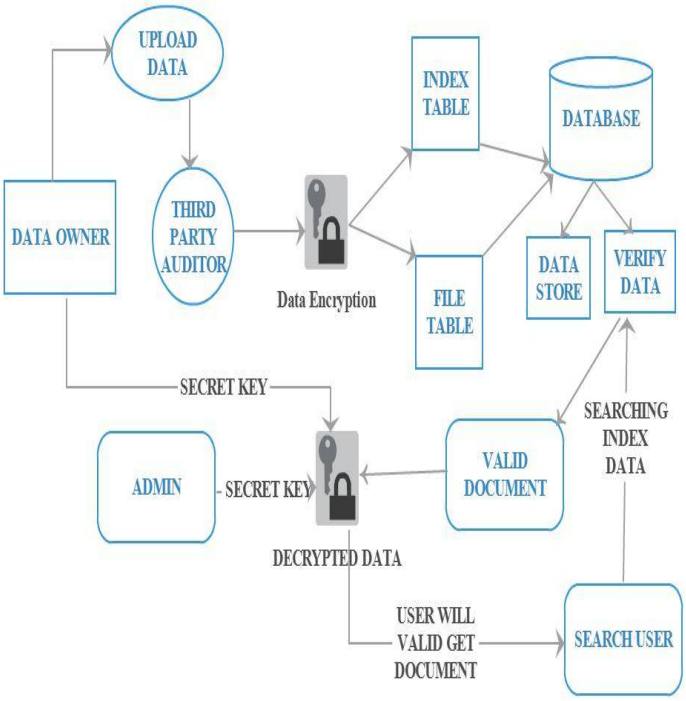


Fig.1 Proposed System Architecture

Issues in Past System and its Resolution

In the past approaches, with the single exception of, past work on privacy preserving association rule mining has focused on one-time mining. And even though the authors in study private and incremental mining, they do so for non-quantitative data. To date, no work has been published on incrementally mining association rules with privacy protection when the data upon which mining occurs is quantitative and subject to change.

We study this problem against the backdrop of supply chain management where a data owner's operational data is scattered over multiple source sites, but collected at a single target site. Consider for example, large grocery or apparel chains in the United States, such as Giant or Land's End with several retail outlets (source sites/databases), but a single centralized data warehouse (target site/database), that outsource the task of mining to a consulting firm (a data miner). The existing approaches contains lots of disadvantages, some of them are listed below:

(i) We believe that this problem has significant relevance in many application areas such as retail, health and finance.

(ii) There are many discrete wavelet transforms. The simplest one is the wavelet transform.

(iii) The averages and differences are designated as wavelet coefficients of the trans-formed data with averages being referred to as approximations and differences the details.

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 6, Issue 1, January 2018

In the proposed system, the structure consists of the clients, data miner and the storage section. The clients, who own the data, will outsource and store their encrypted data to the untrusted storage server. The adversary cannot distinguish between two test sequences of values as long as the sequence has the same itemset.

The security of our scheme must not leak anything besides the bits for deduplication itemsets choosing in the raw itemsets. The correctness properity required of the scheme is straightforward equality should finally return only the right tags that exist at the server and the equality test algorithm will finally output the correct result. Finally we need to prove that the client will get the right itemsets. We provide the security and theoretical performance analysis for the proposed schemes, and implement the equality-testing schemes with the existing itemsets and provide the performance evaluation of our proposed systems with different itemsets. The proposed approaches contains lots of advantages, some of them are listed below:

(i) To enhance the security of itemset and protect the data confidentiality transforming the predictable itemsets into an unpredictable itemsets.

(ii) In their system, a third party called key server is introduced to generate the file tag for duplication check.

(iii) Addressed the key management issue in block-level deduplication by distributing these keys across multiple servers after encrypting the files.

II. SYSTEM IMPLEMENTATIONS

The proposed system is composed based on several modules. All are listed and summarized below in detail.

A. File Integrity

In this module, the data owner will store the file to the storage server. The server will use pairing computation over the whole database to realize the equality test. The equality test algorithm will finally output the correct result. If the same data owner or clients sent the same file containing the same information then storage will not accept the file .

Techniques and Algorithms Used for File Integrity:

(i) Advanced Hardening Server Algorithm, and

(ii) File Integrity Checker

Implementation Steps:

Step-1: Creates snapshot of files : a hashed signature for each file (advanced encryption standard)

Step-2: After an attack, compares the file integrity with encryption standard

Step-3: This allows system administrator to determine which files were changes

Step-4: Tripware is a file integrity checker for linux/unix, windows etc

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 6, Issue 1, January 2018

B. Third Party Validation

In this module, the third party role is to validate the file integrity. The file integrity will be validating by the dataset containing in the file. Since the data items and the deduplication decision datasets are randomly generated, the storage time of each data item is usually not the same each time.

Techniques and Algorithms Used for Third Party Validation:

(i) MD5

Suppose a b-bit message as input, and that we need to find its message digest.

Implementation Steps:

Step-1: Append padded bits: – The message is padded so that its length is congruent to 448, modulo 512. • Means extended to just 64 bits shy of being of 512 bits long. – A single "1" bit is appended to the message, and then "0" bits are appended so that the length in bits equals 448 modulo 512

Step-2: Append length A 64 bit representation of b is appended to the result of the previous step. The resulting message has a length that is an exact multiple of 512 bits.

Step-3: Initialize MD Buffer A four-word buffer

(A,B,C,D) is used to compute the message digest. – Here each of A,B,C,D, is a 32 bit register.

Step-4: Process message in 16-word blocks.

Step-5: The message digest produced as output is A, B, C, D. – That is, output begins with the low-order byte of A, and end with the high-order byte of D.

C. File Encryption

To preserve the security in datasets, the file is encrypted and then the file is stored to the storage server. So if any authorized user request the file, the storage will initially declined their request or else if any user get the file, the files is encrypted and it is not in a readable format.

Techniques and Algorithms Used for File Encryption:

(i) MD5

Suppose a b-bit message as input, and that we need to find its message digest.

Implementation Steps:

Step-1: Append padded bits: – The message is padded so that its length is congruent to 448, modulo 512. • Means extended to just 64 bits shy of being of 512 bits long. – A single "1" bit is appended to the message, and then "0" bits are appended so that the length in bits equals 448 modulo 512

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 6, Issue 1, January 2018

Step-2: Append length A 64 bit representation of b is appended to the result of the previous step. The resulting message has a length that is an exact multiple of 512 bits.

Step-3: Initialize MD Buffer A four-word buffer (A,B,C,D) is used to compute the message digest. – Here each of A,B,C,D, is a 32 bit register.

Step-4: Process message in 16-word blocks.

Step-5: The message digest produced as output is A, B, C, D. – That is, output begins with the low-order byte of A, and end with the high-order byte of D.

D. Deduplication Verification

In this module, the file is found out to be a duplicated or replicated then the third party will disposed it as a deduplication and will not accept the file. In this way we can overcome from the deduplication.

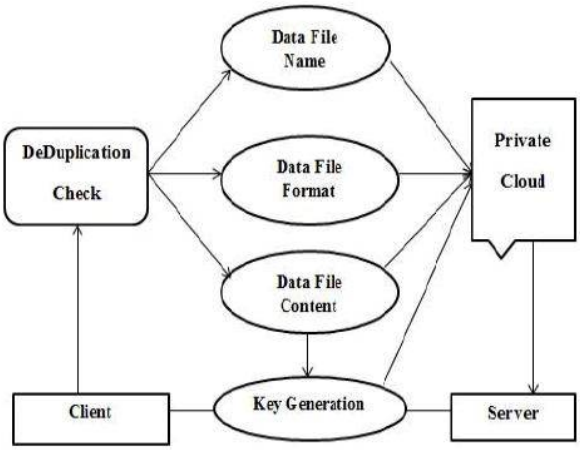


Fig.2 Deduplication Verification

E. File Retrieved by User

The file upload by the owner will be requested by the user. The verifier will verify the file and if it is unique then it will be sent to the user.

III. LITERATURE SURVEY

In the year of 2010, the authors "M. Ahluwalia, A. Gangopadhyay, and Z. Chen" proposed a paper titled "Preserving Privacy in Mining Quantitative Association Rules", in that they described such as: Information mining is the way toward separating concealed examples from information. With the blast of information at a colossal rate, information mining is fundamental to remove helpful data. Affiliation run mining is a strategy for discovering connection connections among expansive arrangement of information things. An administer is portrayed as delicate if its exposure hazard is over a specific certainty esteem. Delicate principles ought not be revealed to people in general, as

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 6, Issue 1, January 2018

they can be utilized to derive touchy information and give leverage to the business contenders. Methods for concealing affiliation rules are constrained to twofold things. However, true information comprises of quantitative esteems.

In this paper, a strategy to cover up fluffy affiliation govern is proposed, in which, the fuzzified information is mined utilizing changed apriori calculation keeping in mind the end goal to extricate leads and distinguish touchy standards. The touchy guidelines are covered up by diminishing the help estimation of Right Hand Side (RHS) of the run the show. A system for mechanized age of participation work is likewise proposed. Test aftereffects of the proposed approach exhibit productive data stowing away with least reactions.

In the year of 2010, the authors "M. Ahluwalia, R. Gupta, A. Gangopadhyay, Y. Yesha, and M. McAllister" proposed a paper titled "Target-Based Database Synchronization", in that they described such as: Synchronizing source and target databases is an essential errand in numerous database applications. There are examples when the synchronization of source and target databases must be driven from the objective's side and include no progressions to the source's pattern or triggers. We depict a calculation for such synchronization. Our calculation bunches tuples into allotments and thinks about hashes of coordinating source and target parcels before synchronizing just those segments whose hashes don't coordinate. The hash examinations diminish the quantity of tuples that must be traded when the source and target are about synchronized as of now.

Two variations of full replication that vary on locking methodologies are utilized as benchmarks. Experimental outcomes demonstrate that our technique beats both when there are few changes to the database and outflanks push level locking when less than 70% of the segments are changed.

In the year of 2011, the authors "M. Ahluwalia, A. Gangopadhyay, Z. Chen, and Y. Yesha" proposed a paper titled "Target-Based Privacy Preserving Association Rule Mining", in that they described such as: We consider an uncommon case in affiliation govern mining where mining is led by an outsider over information situated at a focal area that is refreshed from a few source areas.

The information at the focal area is very still while that streaming in through source areas is in movement. We force a few restrictions on the source areas, with the goal that the focal target area tracks and privatizes changes and an outsider mines the information incrementally. Our outcomes indicate high productivity, security and exactness of standards for little to direct updates in vast volumes of information. We trust that the structure we create is consequently appropriate and important for safely mining huge information.

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 6, Issue 1, January 2018

IV. EXPERIMENTAL RESULTS

The following figure illustrates the Registration page of the proposed system design.

Fig.3 Registration Page Design

The following figure illustrates the User Login Page design.

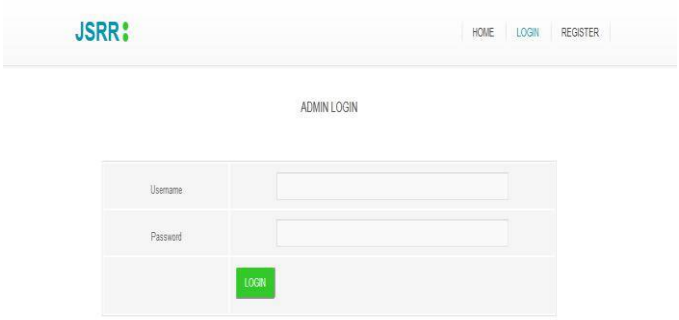


Fig.4 User Login Page

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 6, Issue 1, January 2018

The following figure illustrates the File Verification Page of the proposed system design.

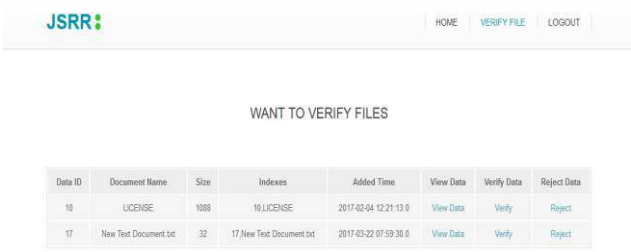


Fig.5 File Verification Page

The following figure illustrates the View File Option in the proposed system design.

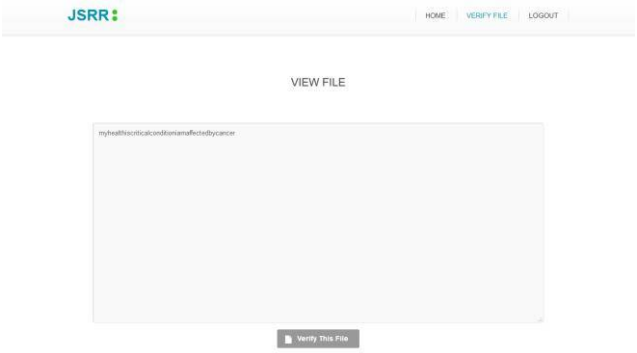
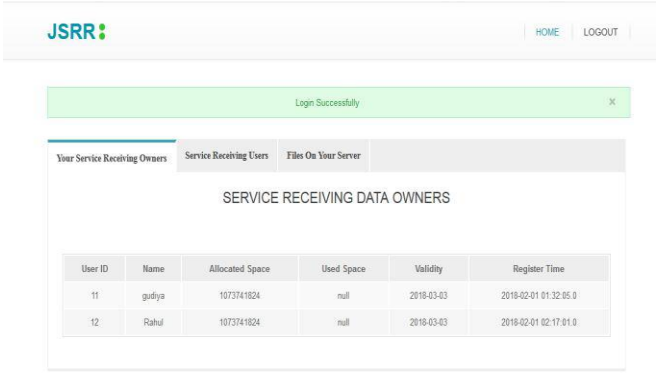


Fig.6 File View Option

The following figure illustrates the Service Receiving Data Owner Details.



ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 6, Issue 1, January 2018

V. CONCLUSION

Past work on protection saving quantitative affiliation manages mining utilizing discrete wavelet change has been set in a situation where the accumulation of information isn't transient and not subject to change. Since information is at times static and very still, that approach has minimal down to earth importance. We propose an answer for incrementally mine quantitative affiliation leads safely in a dynamic situation where the recognition of progress in the first information is started and controlled by a database other than the one in which changes begin. Tests indicate 90% - 100% precision of the tenets for most datasets notwithstanding when half of the first information experiences a change. Examinations demonstrate that the proposed TB-PIRU calculation outflanks different methodologies in protecting the two information security and governs and is extremely productive for 10% of changes in the first information, when this information is expansive in measure. We give a heuristic investigation of security for numeric information. A future report is required to direct a formal investigation of security.

REFERENCES

[1] W. K. Wong, D. W. Cheung, E. Hung, and H. Liu, "Protecting privacy in **Fig.7 Service Receiving Data Owner Details** incremental maintenance for distributed association rule mining," PAKDD'08: Proceedings of the 12th Pacific-Asia conference on Advances in knowledge discovery and data mining, 2008.
[2] M. Ahluwalia, A. Gangopadhyay, and Z. Chen, "Preserving Privacy in Mining Quantitative Association Rules," International Journal of Information Security and Privacy, 2010.
[3] Ahluwalia, R. Gupta, A. Gangopadhyay, Y. Yesha, and M. McAllister, "Target-Based Database Synchronization," presented at the 25th ACM Symposium on Applied Computing, Sierre, Switzerland, 2010
[4]M. Ahluwalia, A. Gangopadhyay, Z. Chen, and Y. Yesha, "Target-Based Privacy Preserving Association Rule Mining," presented at 26th ACM Symposium on Applied Computing, Tai-Chung, Taiwan, 2011.
[5]D. W. Cheung, S. D. Lee, and B. Kao, "A general incremental technique for maintaining discovered association rules," Proc. 5th Int. Conf. Database Systems Advanced Applications, pp. 1-4, 1997.
[6]A. Evfimevski, R. Srikant, R. Agrawal, and J. Gehrke, "Privacy preserving mining of association rules," 8th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD'02), pp. 217 - 228, 2002.
[7]J.-L. Lin and J. Y.-C. Liu, "Privacy preserving itemset mining through fake transactions," in Proceedings of the 2007 ACM symposium on Applied computing. Seoul, Korea: ACM Press, 2007, pp. 375-379.
[8]S. Rizvi and J. R. Haritsa, "Maintaining Data Privacy in Association Rule Mining," VLDB, pp. 682-693, 2002.
[9]Z.-Y. Chen and G.-H. Liu, "Quantitative Association Rules Mining Methods with Privacy-preserving," Sixth International Conference on Parallel and Distributed Computing, Applications and Technologies, 2005. PDCAT 2005., pp. 910-912, 2005.
[10] J. S. Vaidya and C. Clifton, "Privacy preserving association rule mining in vertically partitioned data," 8th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 639 - 644, 2002.
[11] M. Kantarcioglu and C. Clifton, "Privacy-preserving distributed mining of association rules on horizontally partitioned data," IEEE Transactions on Knowledge and Data Engineering, vol. 16, pp. 1026-1037, 2004.
[12] C. C. Aggarwal and P. S. Yu, "A Condensation Approach to Privacy Preserving Data Mining," 9th International Conference on Extending Database Technology, pp. 183-199, 2004.