**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**
**IMPACT FACTOR: 5.258**

***IJCSMC, Vol. 5, Issue. 3, March 2016, pg.71 – 75***

# Verifying Result Correctness of Outsourced Frequent Item set Mining using Rob Frugal Algorithm

Author Name 1: Meenambigai.D
Department of CSE,
IFET College of Engineering,
Villupuram

Author Name 2: Vijayalakshmi.S
Associate Professor, CSE
IFET College of Engineering,
Villupuram

**ABSTRACT:** The server is untrusted and tries to escape from verification by using its prior knowledge of the outsourced data. Outsourcing data mining computations to a third-party service provider (server) offers a cost-effective option, especially for data owners (clients) of limited resources. The increasing ability to generate vast quantities of data presents technical challenges for efficient data mining. Cloud computing provides a natural solution for the data-mining-as-a-service (DM aS) paradigm. Informally, frequent item sets refer to a set of data values (e.g., product items) whose number of co-occurrences exceeds a given threshold. Frequent item set mining has been proven important in many applications such as market data analysis, networking data study, and human gene association study. The Rob Frugal method is implemented with a strategy for incremental maintenance of the synopsis against updates consisting of appends and dropping of old transaction batches. This techniques provides bidirectional encryption of client and server which protects against the forging the content of communication and helps to prevent man in the middle attack.

## I. INTRODUCTION

Data mining (sometimes called data or knowledge discovery) is the process of analyzing data from different perspectives and summarizing it into useful information - information that can be used to increase revenue, cuts costs, or both. Data mining software is one of thenumber of analytical tools for analyzing data. It allows users to analyze data from many different dimensions or angles, categorize it, and summarize the relationships is identified. Technically, data mining is the process of finding correlations or patterns among dozens of fields in large relational databases.Data mining (the analysis step of the "Knowledge Discovery in Databases" process, or KDD), an interdisciplinary subfield of computer science, the computational process of discovering patterns in large data sets involving methods at the intersection of artificial intelligence , statistics,

 

and database systems. The overall goal to decide upon marketing strategies for their product. They can use data to compare and contrast among competitors. Data mining interprets its data into real time analysis that can be used to increase sales, promote new product, or delete product that is not value-added to the company.

**Pattern mining** is a data mining method that involves finding existing patterns in data. In this context patterns often means association rules. The original motivation for searching association rules came from the desire to analyze supermarket transaction data, that is, to examine customer behavior in terms of the purchased products. For example, an association rules "beer ⇒ potato chips (80%)" states that four out of five customers that bought beer also bought potato chips. In the context of pattern mining as a tool to identify terrorist l of the data mining process is to extract information from a data set and transform it into an understandable structure for further use.

In this paper, I have proposed the Rob Frugal method. It will investigate encryption schemes that can resist such privacy vulnerabilities. It also interested in exploring how to improve the Rob Frugal algorithm to minimize the number of spurious patternsthis is based on 1–1 substitution ciphers for items and adding fake transactions to make each cipher item share the same frequency as≥k−1 others. It makes use of a compact synopsis of the fake transactions from which the true support of mined patterns from the server can be efficiently recovered. It also proposed a strategy for incremental maintenance of the synopsis against updates consisting of appends and dropping of old transaction batches. Previous research has shown that frequent item set mining can be computationally intensive, due to the huge search space that is exponential to data size as well as the possible explosive number of covered frequent item sets. Therefore, for those clients of limited computational resources, outsourcing.

Data mining uses information from past data to analyze the outcome of a particular problem or situation that may arise. Data mining works to analyze data stored in data warehouses that are used to store that data that is being analyzed. That particular data may come from all parts of business, from the production to the management. Managers also use data mining to decide upon marketing strategies for their product. They can use data to compare and contrast among competitors. Data mining interprets its data into real time analysis that can be used to increase sales, promote new product, or delete product that is not value-added to the company.

## II. LITERATURE SURVEY

C. S. Yeo [1] in this paper identifies various computing paradigms promising to deliver the vision of computing utilities; defines Cloud computing and provides the architecture for creating market-oriented Clouds by leveraging technologies such as VMs; provides thoughts on market-based resource management strategies that encompass both customer-driven service management and computational risk management to sustain SLAoriented resource allocation; presents some representative Cloud platforms especially those developed in industries along with our current work towards realizing market-oriented resource allocation of Clouds by leveraging the 3rd generation Aneka enterprise Grid technology; reveals our early thoughts on interconnecting Clouds for dynamically creating an atmospheric computing environment along with pointers to future community research; and concludes with the need for convergence of competing IT paradigms for delivering our 21st century vision.

**W. K. Wong[2]** proposed that Outsourcing association rule mining to an outside service provider brings several important benefits to the data owner. These include (i) relief from the high mining cost, (ii) minimization of demands in resources, and (iii) effective centralized mining for multiple distributed owners. On the other hand, security is an issue; the service provider should be prevented from accessing the actual data since (i) the data may be associated with private information, (ii) the frequency analysis is meant to be used solely by the owner. This paper proposes substitution cipher techniques in the encryption of transactional data for outsourcing association rule mining. After identifying the non-trivial threats to a straightforward one-to-one item mapping substitution cipher, here I have propose a more secure encryption scheme based on a one-to-n item mapping that transforms transactions non-deterministically, yet guarantees correct decryption. I have developed an effective and efficient encryption algorithm based on this method. Our algorithm performs a single pass over the database and thus is suitable for applications in which data owners send streams of transactions to the service provider. A comprehensive cryptanalysis study is carried out. The results show that our technique is highly secures with a low data transformation cost.

**L. Qiu[3]** data mining plays an important role in decision making. Since many organizations do not possess the in-house expertise of data mining, it is beneficial to outsource data mining tasks to external service providers. However, most organizations hesitate to do so due to the concern of loss of business intelligence and customer privacy. In this paper, we present a Bloom filter based solution to enable organizations to outsource their tasks of mining association rules, at the same time, protect their business intelligence and customer privacy. Our approach can achieve high precision in data mining by trading-off the storage requirement.

 

**C. Clifton [4]** Privacy preserving data mining – getting valid data mining results without learning the underlying data values – has been receiving attention in the research community and beyond. It is unclear what privacy preserving means. This paper provides a framework and metrics for discussing the meaning of privacy preserving data mining, as a foundation for further research in this field.

## III. PROPOSED SYSTEM

This techniques provides bidirectional encryption of client and server which protect against the forging the contents of the communication and helps to prevent man in the middle attack. An attack model was generated based on following criteria such as based on assumption that the service provider (who can be an attacker) is semi honest in the sense that although he does not know the details of the encryption algorithm, he can be curious and thus can use his background knowledge to make inferences on the encrypted transactions. It has been assumed that the attacker always returns (encrypted) item sets together with their exact support. Rob Frugal algorithm helps to provide privacy for the database on server, who ships the data to client for association rule mining.

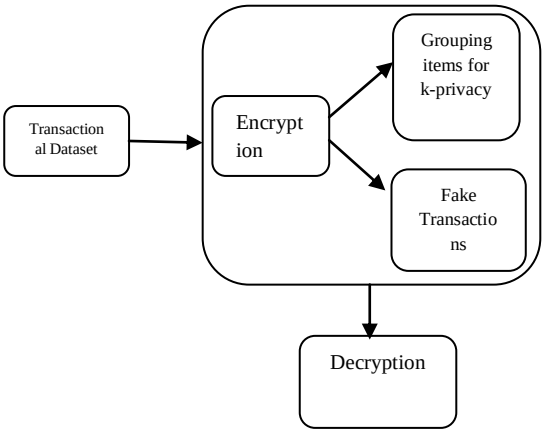


**Fig 1: Overall Architecture Diagram**

### Dataset Collection and Encryption

A dataset (or data set) is a collection of data. Dataset is collected from Belgium retail market dataset. It contains the (anonym zed) retail market basket data from an anonymous Belgian retail store. The data are provided 'as is'. Basically, any use of the data is allowed as long as the proper acknowledgment is provided and a copy of the work is provided to Tom Brijs.

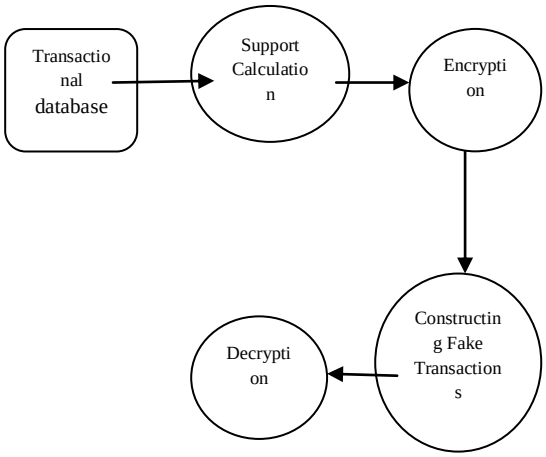


**Fig 2: Encryption and Decryption**

 

The supermarket store carries 16,470 unique SKU's, but some of them only on a seasonal basis. In total, 5,133 customers have purchased at least one product in the supermarket during the data collection period. A dataset is encrypted by using Homomorphic encryption. Homomorphic encryption is an encryption scheme which transforms a TDB D into its encrypted version D∗.

**Grouping Items for k-Privacy**

Given the items support table, several strategies can be adopted to cluster the items into groups of size k. It starts from a simple grouping method. I assume the item support table is sorted in descending order of support and refer to cipher items in this order as e1, e2, etc.

Assume e1, e2 ... en is the list of cipher items in descending order of support (with respect to D), the groups created are {e1,..., ek}, {ek+1,... ,e2k}, and so on. The last group, if less than k in size is merged with its previous group.

Given the fact that the support of the items strictly decreases monotonically, the grouping is optimal among all the groupings with the item support table sorted in descending order of support. This means, it minimizes ||G||, the size of the fake transactions added, and hence the size ||D∗||.

**Constructing Fake Transactions**

Given a noise table specifying the noise N(e) needed for each cipher item e, I generate the fake transactions as follows. First, I drop the rows with zero noise, corresponding to the most frequent items of each group or to other items with support equal to the maximum support of a group. Second, I sort the remaining rows in descending order of noise. This method yields a minimum number of different types of fake transactions that equal the number of cipher items with distinct noise. This observation yields a compact synopsis for the client of the introduced fake transactions. The purpose of using a compact synopsis is to reduce the storage overhead at the side of the data owner who may not be equipped with sufficient computational resources and storage, which is common in the outsourcing data model.

In order to implement the synopsis efficiently, I use a hash table generated with a minimal perfect hash function. Minimal perfect hash functions are widely used for memory efficient storage and fast retrieval of items from static sets. A minimal perfect hash function is a perfect hash function that maps n keys to n consecutive integers, usually [0 ...n − 1].

**Decryption**

When the client requests the execution of a pattern mining query to the server, specifying a minimum support threshold σ, the server returns the computed frequent patterns from D∗. Clearly, for every item set S and its corresponding cipher item- set E, I have that supp D(S) ≤ supp D∗ (E). For each cipher pattern E returned by the server together with supp D∗ (E), the E/D module recovers the corresponding plain pattern S. It needs to reconstruct the exact support of S in D and decide on this basis if S is a frequent pattern. To achieve this goal, the E/D module adjusts the support of E by removing the effect of the fake transactions.

## IV. SYSTEM IMPLEMENTATION

The system output is mainly based on privacy preserving method. It will be evaluated using robfrugal algorithm. The evaluation of data can be done by using calculating support value and arrange that in a descending order. Then the data have been split up into two groups to preserve the data from the third party server. Now the data have been encrypted and stored in a server side. Multiplecompanies have access the server. So, the data will be disclosed. For that purpose, client encrypts its data and stores it in a server in some other format. Based on the mining queries server conducts mining and sends encrypted pattern to the client. Finally client decrypts the encrypted pattern and gets true support of the original transactions.

## V. CONCULSION AND FUTURE WORK

The project involved the problem of (corporate) privacy-preserving mining of frequent patterns (from which association rules can easily be computed) on an encrypted outsourced TDB. I assumed that a conservative model where the adversary knows the domain of items and their exact frequency and can use this knowledge to identify cipher items and cipher item sets. I have proposed an encryption scheme, called Homomorphic, which is based on 1–1 substitution ciphers for items and adding fake transactions to make each cipher item share the same frequency. It makes use of a compact synopsis of the fake transactions from which the true support of mined patterns from the server can be efficiently recovered. I have also proposed a strategy for incremental maintenance of the synopsis against updates consisting of appends and dropping of old transaction batches.

 

Currently, our privacy analysis is based on the assumption of equal likelihood of candidates. It would be interesting to enhance the framework and the analysis by appealing to cryptographic notions such as perfect secrecy. Moreover, our work considers the cipher text-only attack model, in which the attacker has access only to the encrypted items. It could be interesting to consider other attack models where the attacker knows some pairs of items and their cipher values.

**REFERENCES**

[1] R. Buyya, C. S. Yeo, and S. Venugopal, "Market-oriented cloud com- puting: Vision, hype, and reality for delivering it services as computing utilities," in Proc. IEEE Conf. High Performance Comput. Commun., Sep. 2008, pp. 5–13.

[2] W. K. Wong, D. W. Cheung, E. Hung, B. Kao, and N. Mamoulis, "Security in outsourcing of association rule mining," in Proc. Int. Conf. Very Large Data Bases, 2007, pp. 111–122.

[3] L. Qiu, Y. Li, and X. Wu, "Protecting business intelligence and customer privacy while outsourcing data mining tasks," Knowledge Inform.Syst., vol. 17, no. 1, pp. 99–120, 2008.

[4] C. Clifton, M. Kantarcioglu, and J. Vaidya, "Defining privacy for data mining," in Proc. Nat. Sci. Found. Workshop Next Generation Data Mining, 2002, pp. 126–133.

[5] I. Molloy, N. Li, and T. Li, "On the (in)security and (im)practicality of outsourcing precise association rule mining," in Proc. IEEE Int. Conf. Data Mining, Dec. 2009, pp. 872–877.

[6] F. Giannotti, L. V. Lakshmanan, A. Monreale, D. Pedreschi, and H. Wang, "Privacy-preserving data mining from outsourced databases," in Proc. SPCC2010 Conjunction with CPDP, 2010, pp. 411–426.

[7] R. Agrawal and R. Srikant, "Privacy-preserving data mining," in Proc. ACM SIGMOD Int. Conf. Manage. Data, 2000, pp. 439–450.

[8] S. J. Rizvi and J. R. Haritsa, "Maintaining data privacy in association rule mining," in Proc. Int. Conf. Very Large Data Bases, 2002, pp. 682– 693.

[9] M. Kantarcioglu and C. Clifton, "Privacy-preserving distributed mining of association rules on horizontally partitioned data," IEEE Trans. Knowledge Data Eng., vol. 16, no. 9, pp. 1026–1037, Sep. 2004.

[10] B. Gilburd, A. Schuster, and R. Wolff, "k-ttp: A new privacy model for large scale distributed environments," in Proc. Int. Conf. Very Large Data Bases, 2005, pp. 563–568.

 