© 2016 IJSRSET | Volume 2 | Issue 2 | Print ISSN : 2395-1990 | Online ISSN : 2394-4099
Themed Section: Engineering and Technology

Privacy Preserving Mining Association Rule from Outsourced Transaction

Thiruvenkatasamy, Deverajsamy. S, Dhanraj. M

Department of Computer Science and Engineering, Anna University, Anna University, Chennai, India

ABSTRACT

Reproached by developments such as cloud computing, there has been considerable recent interest in paradigm of data mining-as-a-service. The company (data owner) lacking in expertise or computational resources can outsource it's to the third revelry service worker (server). However, both items and association rules of outsourced database are considered private property of corporation (data owner). To protect corporate privacy, data owner transforms its data and ships it to server, sends mining queries to server, and recovers true patterns from extracted patterns received from server. In this paper, we study problem of outsourcing association rule mining task within the corporate privacy-preserving outline. We recommend an attack model based on background knowledge and devise the scheme for privacy preserving outsourced mining. Our scheme ensures that each transformed item was indistinguishable, w.r.t. attacker's background knowledge, from at minimum k-1 other transformed items. Our comprehensive experiments on the very large and real transaction database establish that our systems are effective, scalable, and protect privacy.

Keywords: Rule Mining Task, Outsourced Mining, Direct Mail Marketing, Privacy Preserving

I. INTRODUCTION

Today there was large amount of data proceed in every day from different sources. That large amount of data stored in different database. These data store in storage devices in from of row data. Data mining was process of discovering interesting pattern and knowledge from large amount of data. Following Example where data mining techniques are used are Direct mail marketing, bio informatics, praise fraud correction, text investigation and market basket analysis. Extracting knowledge from row data, There Database Amish Desai, Computer Science and Engineering was some technique to deal with security .Privacy preserving in data mining was one of technique that deal with security of knowledge that extracted by data mining technique. There are various Data Mining Tasks: Cataloging Bunching Association Rule Mining Sequential Pattern Mining Regression.

II. METHODS AND MATERIAL

A. Related Works

Data mining was process of gathering information about user specific data, also called knowledge discovery, on internet. Problem with data mining output was that it also discloses some information, which was considered to be isolated and particular. Effortless access to such personal data causes the peril to individual privacy [9]. Recent research in area of privacy preserving data mining has considerate effort to determine the trade-off between privacy and need for knowledge discovery, which was necessary in order to improve decision-making processes and other human activities. PPDM cope with problem of learning accurate models over aggregate data, while protecting privacy at level of individual records. Main purpose of privacy preserving data mining was to design competent frameworks and algorithms that can extract relevant knowledge from the

large amount of data without revealing of any sensitive information [9]. It protects complex material by as long as sterile catalog of original database on internet or the process was used in such the way that private numbers and sequestered knowledge remain private even after mining process. It was PPDM due to which assistances of data pulling out be enjoyed, without compromising privacy of concerned individuals.

Association Rule Mining: Association law withdrawal one of task of data mining. Association rule mining was important field to under privacy preserving data mining. R. Agrawal was first proposed basic concept of Association rule mining. Association rule was basically using concept of IF-THEN relationship among different data. Following example of shows concept of Association rule. "If consumer acquisition the mainframe , then he/she was 85% likely to also buying protection ". Analysis of above example that laptop was somewhat related to anti-virus because every time customer buy the computer then he/she buy anti-virus. Association rule was used for market basket analysis. Let 1= {II, I2, ..., In} be the set of item. Let D be the database of transactions where each transaction T was the set of item such that T belongs to I.

For every transaction was associated to an identifier, called TID. the transaction T was contain the if and only if the belongs to T. An association rule was applied of form A-B. Where AQ, BQ. And AB belongs to <1>. Every association rule must be satisfy two contain support and confidence. Support of rule A-B was transaction database that contain support count of AUB. Support for rule (A-B) can be calculated using below formula in (1).

IAUBI support (A-B) belongs to D where D was total number of transaction in transaction database. Confidence of rule A-B was transaction database that contain the also contain B. confidence for rule (A-B) can be calculated using below formula in (2). Confidence

Association Rule Hiding: Association rule hiding was one technique to PPDM (Privacy Preserving Data Mining). Association rule hiding methodology aim was to sanitize original data. so it may be applied to following condition: (1) sanitized database was not reveal any sensitive rules. (2) Sanitized database was mining of all non-sensitive rules. (3) Sanitized database was not add any new rules, not present in database D. Association rule hiding was depend on support and confidence of rule, There was two way to hide any rule (i) Decrease support up to certain threshold. (ii) Decrease confidence up to certain threshold.

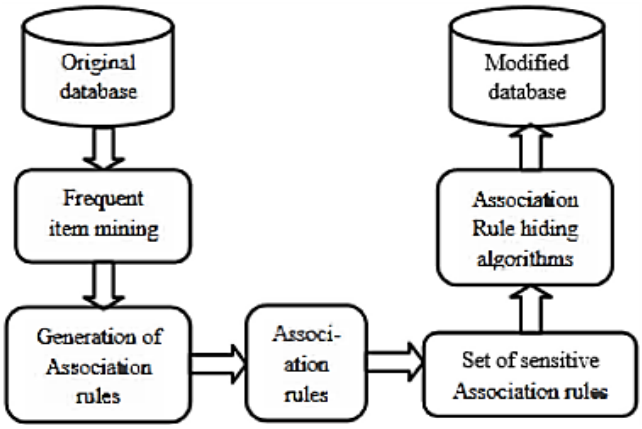


Figure 1. Basic Architecture of rule hiding

B. Existing System

There are many methodologies used for maintaining privacy in transaction database. Before developing tool it was necessary to determine time factor, economy and company strength. Once these things are satisfied, ten next steps are to control which effective system and language can be used for developing tool. Once programmers start building tool programmers need lot of external support. these support can be obtained from senior computer operator, since book or from websites. Before building system above consideration r taken into account for developing proposed system. This section provides background to research through the review of some of literature on privacy. literature review was focused on those areas central to scope of these research. It was an almost customary feature of many analysis of privacy to begin with the disclaimer about inherent difficulty of defining exactly what 'privacy' was and disaggregating its various dimensions. It was something that was taken for granted and most people would have the sense of what privacy was but have difficulty putting it into words. Concept and meaning of privacy has long been debated by philosophers, community geniuses, theoretical notaries and other members. All explanations, to some extent, are based on assumptions about individualism and about distinction between realms of civil society and state. However, many gloss

over essential cultural, class-related and gender differences. Works on secrecy tends to give students an overwhelming sense that privacy was the deeply contested notion, which repeatedly varies according to situation and environment. Conferring to Bennett and Raab (2003), in Western culture, modern claim to privacy and contemporary justification for information privacy as the public policy goal was derived from the notion of the boundary between individual and other individuals, and between individual and state. these notion of secrecy rests on the paradigm of society as comprising comparatively autonomous personalities and on notions of differences between privacy claims and interests of different individuals. According to John Stuart Mill there should be certain 'self-regarding' actions of private apprehension, contrasted with 'other-regarding' actions to community interest and regulation. Shils argued that privacy was essential for strength of American pluralistic democracy because it bolsters boundaries between competing and countervailing centers of power. Dr Alan Westin, the foremost speculative reinforced importance of privacy for liberal democratic societies – in contrast to totalitarian regimes: A balance that certifies strong castles of single and group privacy and limits both disclosure and surveillance was the prerequisite for liberal democratic societies. Democratic society relies on publicity as the control over government, and on privacy as the shield for group and individual life. Westin also addresses specific functions that privacy plays. It promotes autonomy of association. It shields studentship and science from unnecessary interference by government. It permits use of the secret ballot and protects chosen by election process by dismal government following of the citizen's past voting record. It imprisons unsuitable laws conduct such as perverse search and seizure. It also serves to shield those institutions, such as press, that operate to keep government accountable.

Privacy has also been defined comprehensively: Privacy was the concept related to solitude, secrecy, and autonomy, but it was not synonymous with these terms; for beyond purely descriptive aspects of privacy as isolation from company, curiosity, and influence of others, privacy implies the normative element: correct to limited control of contact to private realms… right to privacy asserts sacredness of person;… any invasion of privacy constitutes an offence against rights of personality – against individuality, dignity, and freedom. Privacy can be divided into following facets Territorial privacy – concerning setting of confines on invasion into domestic and other environments such as workplace or public space.

- Privacy of person – these was concerned with protecting the person touching undue interventions such as physical searches and drug testing, and data that encroach upon his or her moral sense;
- Privacy of communications, covering security and privacy of mail, telephones, email and other forms of communication;
- Privacy in information context – these deals with gathering, collecting and selective broadcasting of personal material such as glory data and medical registers.

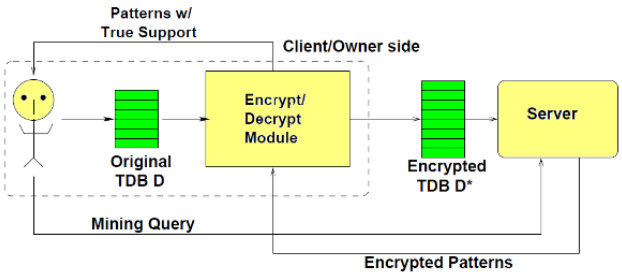
The homily on privacy as the policy question has largely focused on info privacy and it was these facet of privacy that these research project will focus on. In these sense, privacy can be defined as "the entitlement of folks, groups or institutions to determine for themselves when, how and to what range statistics about them was linked to others." However, rise to prominence of Internet transportations and e-commerce has led to disclosure of roads (and transmission) attracting more attention and concern. increased concern with discretion of communications has caused some mix-up between meanings of information privacy and information security and terms are often used interchangeably. As Clarke noted (as cited in Bennett & Raab, 2003), term 'privacy' was used by some people to refer to security of data or safekeeping of data during program as shelter against various jeopardizes, such as data being edited or modified by unauthorized persons. These aspects, however, are only the small fraction of deliberations within arena of 'information privacy'. That is, data refuge was the obligatory but not sufficient condition for info disclosure. An organization might keep personal information it collects highly secure, but if it should not be collecting that information in first place, individual's information privacy rights are clearly violated.

III. RESULTS AND DISCUSSION

Proposed Work

The particular problem attacked in our paper was outsourcing of pattern mining within the corporate privacy-preserving framework. the key distinction

between these problem and abovementioned PPDM problems was that, in our setting, not only underlying data but also mined fallouts are not projected for partaking and must remain private. In particular, when server possesses background facts and deportments attacks on that origin, it should not be able to guess correct candidate item or item set corresponding to the given cipher item or item set with the probability above the given threshold. We proposed to solve these problem by using k-privacy, i.e.,each item in subcontracted dataset should be hazy from at least k − 1 items regarding their support. In these paper, our goal was to devise an encryption scheme which enables formal privacy guarantees to be proved, and to validate these model over large-scale, real-life transaction database.



This work was to devise encryption schemes such that formal privacy guarantees can be proven against attacks conducted by server using background acquaintance, while protection there source requirements under control.
System Models

The Pattern Mining Task: The reader was assumed to be familiar with basics of association rule mining. We let I = i1, ..., in be set of items and D = t1, ..., tm the transaction database (TDB) of transactions, each of which was the set of items. We denote support of an item set S ⊆ I as supp D(S) and frequency by freq D(S). Memory, freq D(S)= supp D(S)/|D|. For each item i, supp D(i) and freq D(i) denote respectively individual support and frequency of i. function supp(D) projected over items, was also called *item support table* . well-known frequent pattern mining problem: given the TDB D and the support threshold σ, find all item sets whose support in D was at least σ. In these papers, we confine ourselves to study of the (corporate) discretion antibacterial outsourcing agenda for common pattern withdrawal.

Privacy Model: We let D denote original TDB that owner has. To protect identification of individual items, owner applies an encryption function to D and transforms it to D∗, encrypted database. We refer to substances in D as *natural items* and items in D∗ as *cipher items*. term item shall mean plain item by default. notions of plain item sets, plain dealings, plain decorations, and their cipher counterparts are defined in obvious way. We use I to denote set of plain items and E to refer to set of cipher items.

Attack Model: The waiter or an impostor who gains admission to it may have some background gen using which they can on encrypted database D∗. We generically refer to of these agents as an *attacker*. We adopt the conservative model and assume that attacker knows exactly set of (plain) items I in original transaction database D and their true supports. We assume provision benefactor (who can be an attacker) is ***semi-honest*** in sense that although he does not know details of our encryption process, he can be enquiring and thus can use his related knowledge to make interpretations on encrypted transactions. We also assume that attacker always earnings (encrypted) item sets calm with their exact support. Data owner (i.e., corporate) considers true uniqueness of: every encryption item, every cipher contract, and every cipher everyday pattern as cerebral property which should be protected. We consider following attack Model

• ***Item-based attack:*** semi honest service provider can attack owners data depend upon single item identity.
• ***Set-based attack:*** service provider attack owner's data depend upon many item identities. In these method attacker can easily attacks data correctly but they can't practice that facts because that data's are in ciper text form. Data owners are using separate E/D Module.

Encryption: In this section, we introduce encryption scheme, which transforms the TDB D into its encrypted version D∗. Our scheme was parametric w.r.t. k > 0 and consists of three main steps: (1) using substitution ciphers for each plain item; (2) using the specific item k-grouping method; (3) using the method for adding new fake transactions for attaining k-privacy.
The built fake businesses are added to D(once items are replaced by cipher items) to form D∗, and transmitted to server.

Decryption: When shopper wishes execution of the pattern mining query to server, specifying the minimum support threshold σ, server returns computed frequent patterns from D∗. Clearly, for every item set S and its consistent cryptogram item set E, we have that supp D(S) ≤ supp D_(E).For each cipher pattern E returned by server together with supp D_(E), E/D module recovers corresponding plain pattern S. It needs to reconstruct exact support of S in D and decide on this basis if S was the frequent pattern. To achieve this goal, E/D module adjusts support of E by removing effect of fake transactions. Supp D(S) = supp D_(E)−supp D_\D(E). these follows from fact that support of an item set was additive over the disjoint union of transaction sets. Finally, pattern S with adjusted support was kept in output if supp D(S) ≥ σ. calculation
of supp D \ D(E) was performed by E/D module using synopsis of fake transactions in D∗ \ D.

IV. CONCLUSION

We proposed the protocol for secure withdrawal of overtone rules in horizontally distributed databases that improves significantly upon current leading protocol in terms of privacy and efficiency. One of main ingredients in our proposed protocol was the novel secure multi-party protocol for computing union (or intersection) of private subsets that each of interacting players holds. Another ingredient was the protocol that test inclusion of element held by one player in the subset to another. That protocol exploits fact that underlying problem was of interest only when number of players was greater than two. One research problem that these study suggests. Namely, to devise an efficient protocol for inequality verifications that uses existence of the semi honest third party. Such the protocol might enable to further improve upon communication and computational costs of second and third stages of protocol. Other research problems that these study suggests was implementation of techniques presented here to problem of distributed association rule mining in vertical setting problem of mining generalized association rules, and problem of subgroup discovery in horizontally partitioned data

V. REFERENCES

[1] R. Agrawal and R. Srikant, "Fast Algorithms for Mining Association Rules in Large Databases," Proc 20th Int'l Conf. Very Large Data Bases (VLDB), pp. 487-499, 1994.

[2] R. Agrawal and R. Srikant, "Privacy-Preserving Data Mining," Proc. ACM SIGMOD Conf., pp. 439-450, 2000.

[3] D. Beaver, S. Micali, and P. Rogaway, "The Round Complexity of Secure Protocols," Proc. 22nd Ann. ACM Symp. Theory of Computing (STOC), pp. 503-513, 1990.

[4] M. Bellare, R. Canetti, and H. Krawczyk, "Keying Hash Functions for Message Authentication," Proc. 16th Ann. Int'l Cryptology Conf. Advances in Cryptology (Crypto), pp. 1-15, 1996.

[5] A. Ben-David, N. Nisan, and B. Pinkas, "FairplayMP - the System for Secure Multi-Party Computation," Proc. 15th ACM Conf. Computer and Comm. Security (CCS), pp. 257-266, 2008.

[6] J.C. Benaloh, "Secret Sharing Homomorphisms: Keeping Shares of the Secret Secret," Proc. Advances in Cryptology (Crypto), pp. 251-260, 1986.

[7] J. Brickell and V. Shmatikov, "Privacy-Preserving Graph Algorithms in Semi-Honest Model," Proc. 11th Int'l Conf. Theory and Application of Cryptology and Information Security (ASIACRYPT), pp. 236-252, 2005.