JCSE International Journal of Computer Sciences and Engineering Open Access
Research Paper Vol.-7, Issue-2, Feb 2019 E-ISSN: 2347-2693

Hiding Sensitive Itemsets at Multiple Support Thresholds without Side Effects

Surendra H[1*], Mohan H S[2]

[1, 2]Dept. of Information Science & Engineering, SJB Institute of Technology, Bangalore, India

[*]Corresponding Author: surendra.h@gmail.com, Tel.: +91-9611277995

DOI: https://doi.org/10.26438/ijcse/v7i2.115120 | Available online at: www.ijcseonline.org

Accepted: 13/Feb/2019, Published: 28/Feb/2019

***Abstract*—** Mining Frequent pattern is a common technique of data mining and used as a preliminary step to mine association rules. Some frequent patterns are sensitive as they may disclose confidential information to adversaries and needs to be hidden in the data before sharing. Many of the existing techniques hide sensitive itemsets at a single sensitive support threshold. Also, these techniques generate various side effects and suffer from unexpected information loss. In this paper, a novel approach to hide sensitive itemsets at multiple sensitive support thresholds is proposed. The database is modeled as a set of closed itemsets which are selectively sanitized to hide sensitive itemsets. The proposed Recursive Pattern Sanitization algorithm for Personalized Itemsets Hiding (RPS-PIH) sanitizes the closed itemsets to hide sensitive itemsets at multiple sensitive support thresholds without generating any side effects. The sanitized model represents privacy preserved patterns of the database which may be shared to the third party for further data analysis without disclosing private information. Experimental results indicate that the proposed approach is efficient in hiding sensitive itemsets at multiple sensitive support thresholds. The effectiveness of the proposed approach is measured using popular metrics for side effects and information loss. The proposed approach is effective in reducing information loss and eliminating the generation of side effects compared with existing state-of-the-art techniques.

***Keywords*—** Itemset Hiding, Multiple Support Threshold, Privacy Preserved Data Publishing (PPDP), Personalized Privacy Preservation, Pattern Sharing, Pattern Sanitization, Sensitive Knowledge

I. INTRODUCTION

Data mining has emerged as a vital area of knowledge for business and research in recent years. Data mining is the process by which valuable information and patterns in the data are mined which otherwise invisible in the data. The patterns that are mined may also contain confidential information which is not intended to be exposed by the data owners or data collectors. Clifton et al. [1, 2] are among the first to publish effects of data mining on privacy and security of the data and proposed schemes to disallow discovery of sensitive knowledge. Privacy-preserving data mining techniques [3-5] help in protecting sensitive information, yet allow the data analyst to mine for useful information from the data.

Frequent itemset mining is a commonly used technique for finding frequent patterns. It is also a preliminary stage before association rule mining. Frequent itemsets and association rules reveal the relationship between itemsets in a transactional database [6–9]. Some itemsets in the data can contain confidential and sensitive information which are called sensitive itemsets. Sensitive itemsets need to be protected from disclosing by hiding them during the data mining process. The process of hiding sensitive information from the database is called data sanitization. Itemset hiding is the process of hiding itemsets below a specified support threshold in a transactional database.

The general problem of itemset hiding was introduced in [10] and indicated it as NP-hard. Due to the complex nature of the problem, various itemset hiding methodologies have been proposed. Heuristic-based approaches [11–18] selectively change transactions of the database to hide sensitive itemsets. Border revision-based approaches [19–22] revise the border between frequent and infrequent itemsets to hide sensitive itemsets. Exact approaches [23, 24] articulate the itemset hiding as a constraint satisfaction problem. Solution to the applied constraint is solved using linear programming techniques. Some hybrid approaches [25–27] are also proposed which take the advantages of different techniques to improve the efficiency of itemset hiding. However, there is no perfect resolution present due to the intrinsic trouble of side effects with hiding itemsets directly in the database. Many significant contributions have been made to keep down the side effects and information loss, but none of them could eliminate side effects. Many existing techniques address hiding sensitive itemsets below a single support threshold. Also, these techniques do not support specifying separate support threshold to the sensitive itemsets based on their sensitive levels.

We propose to hide sensitive itemset using a model-based approach proposed in our previous work [28]. A novel Recursive Pattern Sanitization (RPS) algorithm was also proposed in [28] to hide sensitive itemsets below single minimum support threshold without side effects. In this paper, an extension to the RPS algorithm named Recursive Pattern Sanitization for Personalized Itemset Hiding (RPS-PIH) algorithm is proposed. The proposed approach models the given transactional database as a set of closed itemsets. The sanitization process of hiding sensitive itemsets is applied to the model, i.e., closed itemsets as an alternative of transactions in the database which eliminates the generation of side effects. Unless the RPS algorithm, separate sensitive support thresholds may be appointed to each sensitive itemsets and the proposed RPS-PIH algorithm hides sensitive itemsets at their specified sensitive support thresholds in the model without generating side effects.

This paper is organized as follows; Section 2 briefs related work on different itemset hiding techniques which hide sensitive itemsets at multiple support thresholds. Section 3 gives important definitions and preliminary information. The proposed methodology is described in Section 4. The experimental results are discussed in Section 5. Section 6 concludes the paper.

II. Related Work

Most of the existing itemset hiding techniques allow specification of a single support threshold below which all given sensitive itemsets to be hidden. The single support threshold does not allow hiding of sensitive itemsets at multiple support threshold based on their sensitive levels. Also, these techniques suffer from various side effects.

Gkoulalas-Divanis, A. and Verykios, V.S [29] proposed border-revision based exact hiding methodology to hide itemsets without side effects. The problem of hiding sensitive itemsets is formulated as a constraint satisfaction problem (CSP), and the exact solution is found using integer programming technique. The proposed technique does not have any side effects on the data. However, the algorithm is designed to hide sensitive itemsets below a single support threshold. Also, the binary integer programming is computationally intensive.

Ahmet Cumhur Ztrk and Belgin Ergen Bostanolu [30] proposed Pseudo Graph-Based Sanitization (PGBS), a graph-based technique which represents the transactions as Pseudo Graph. The scanning operation to hide itemsets in the original database is performed on the Pseudo Graph data structure instead of the actual database. The algorithm hides sensitive itemsets at multiple support threshold levels. They extended this technique to hide the itemsets in an incremental environment dynamically [31] which still uses PGBS to hide the sensitive itemsets. PGBS is used to identify the transaction containing sensitive itemsets, which produce minimal side effects when sanitized. Since the sanitization is performed by altering the database transactions, both PBGS and dynamic version of it suffers from side effects and unintentional information loss.

In our previous work, a model-based approach is proposed where the database is transformed as a set of closed itemsets. These closed itemsets make a model of the transactions in the database. The Recursive Pattern Sanitization (RPS) algorithm recursively hide sensitive itemsets by reducing the support of sensitive itemsets and their supersets in the model below the given support threshold. The side effects are treated by adding non-sensitive subsets of sensitive itemsets back into the model.

III. Preliminaries

Consider a transaction database containing a set of transactions T = (t_1, t_2, …, t_n) Where t is a transaction. The size of the database is denoted as n. Let every transaction is a subset of I = (i_1, i_2, …, i_m) Where I is the set of distinct items i. The support of an itemset P denoted by σ(P) is the frequency of occurrence of that itemset in the given transactional database.

An itemset P is said to be frequent in a given transactional database if its support σ(P) is greater than the given minimum support threshold σ_{min}. Otherwise, it is identified as infrequent. The major drawback of frequent itemsets is that for a given transactional database, the number of frequent itemsets generated can be very large for a lesser σ_{min} leading to itemset explosion.

Closed itemsets provide a compressed form of frequent itemsets by pruning redundant frequent itemsets without loss of information. An itemset is closed if its support is not the same as any of its immediate superset.

Consider, a set of sensitive itemsets S = {(s_1,$\sigma_{SST\text{-}s1}$), (s_2,$\sigma_{SST\text{-}s2}$), …, (s_3,$\sigma_{SST\text{-}s3}$)}, Where s_i is the sensitive itemset and $\sigma_{SST\text{-}si}$ is the sensitive support threshold below which the sensitive itemset s_i to be hidden. The role of the proposed RPS-PIH algorithm is to hide the sensitive itemsets at their specified support threshold $\sigma_{SST\text{-}si}$ without generating any side effects. A sample set of sensitive itemsets and their sensitive support threshold as specified by the data curator is shown in Table 1.

Table 1: Sample Sensitive Itemsets with their Sensitive Support Threshold (SST)

Sensitive Itemset	$\sigma_{SST\text{-}si}$
(A, B)	0.33
(A, D, E)	0.2
(E, F)	0.45

I. METHODOLOGY

In the proposed approach, the first step is to mine closed itemsets from the transactions of the database and build the model. The model is generated by specifying a base minimum support threshold σ_{model} to eliminate infrequent itemsets. Then, the proposed Recursive Pattern Sanitization for Personalized Itemset Hiding (RPS-PIH) is applied on the model to hide sensitive itemsets below their specified support threshold $\sigma_{SST\text{-}si}$ where $\sigma_{SST\text{-}si}$ is always greater than σ_{model}.

A. Recursive Pattern Sanitization for Personalized ItemsetHiding (RPS-PIH)

Apriori principle states that all supersets of an infrequent itemset are also infrequent. So, if an itemset is sensitive, then its supersets having support greater than given sensitive support threshold are also sensitive and must be hidden. However, the other subsets of these supersets may be non-sensitive and need to preserve their original support in the model during the sanitization process.

RPS-PIH effectively hides the sensitive itemsets and their sensitive supersets below the specified sensitive support threshold without modifying the original support of all other non-sensitive itemsets, thus not generating any side effects. The side effects are eliminated by recursively adding non-sensitive subsets of the sensitive itemsets into the model with their original support and, removing the sensitive itemset until their support reduces below the specified sensitive support threshold. The RPS-PIH algorithm uses the minimum support threshold specified for the sensitive itemset under sanitization. The sensitive support threshold can be different for different sensitive itemsets defined by the user based on their sensitive levels. The RPS-PIH is presented in Algorithm 1.

The CFIs in the model are arranged as groups of k-itemset, i.e., all 1-itemsets in a group, all 2-itemsets in another group and so on. These CFI groups are indexed by k so that it is easier and faster to locate the sensitive itemsets during the scanning process.

Algorithm 1: Recursive Pattern Sanitization for Personalized Itemset Hiding (RPS-PIH) Algorithm

Input: Closed Itemsets (Model) and Sensitive Itemsets S = $\{(s_1,\sigma_{SST\text{-}s1}), (s_2,\sigma_{SST\text{-}s2}), \ldots, (s_3,\sigma_{SST\text{-}s3})\}$

Output: Closed itemsets with sensitive itemsets hidden at their respective support threshold (Sanitized Model)

```
 1:  procedure MAIN(Model, S)
 2:    SortedS = SortByItemsetSize(S, Ascending)
 3:    LeastSISize = GetItemsetWithLeastSize(SortedS)
 4:    for k = LeastSISize to MaxItemsetSize(Model) do
 5:      CIList = GetCIsOfSize(Model, k)
 6:      for all CI in $CI_{List}$ do
 7:        $\sigma_{ci} = \sigma(CI)$
 8:        if $\sigma_{ci} \geq \sigma_{SST\text{-}si}$ then
 9:          if Match(CI, SortedS) = true then
10:            RecursiveHiding(CI, $\sigma_{ci}$)
11:            RemoveItemset(Model, CI)

12:  procedure RECURSIVEHIDING(CI, $\sigma_{ci}$)
13:    LeastSizeSI=GetLeastSizeSI(CI, SortedS)
14:    for all item in LeastSizeSI do
15:      $SI_{subset}$=GetSubsetExcludeItem(LeastSizeSI, item)
16:      if Match($SI_{subset}$, SortedS) = true then
17:        RECURSIVEHIDING($SI_{subset}$,$\sigma_{ci}$)
18:      else if Model.Find($SI_{subset}$, $\sigma_{ci}$) = false then
19:        AddItemsetToModel(SIsubset,$\sigma_{ci}$)
```

For a given sensitive itemset and its sensitive support threshold, first, the sensitive itemset is searched in the model considering its size. If not found then itemsets having their size greater by one compared to sensitive itemset (its immediate supersets) are searched. This process is repeated until the sensitive itemset is found either in its group or in any higher itemset size group (line 2 to 6). An itemset which is a sensitive itemset or the superset of the sensitive itemset is known as candidate itemset. If the candidate itemset is found and its support is more than the specified sensitive support threshold, then Recursive Hiding function is called passing the candidate itemset to hide it safely (line 7 to 10).

The recursive hiding function generates subsets of the candidate itemset and recursively checks these subsets for any possible containment of other sensitive itemsets. If the subsets contain other sensitive itemsets, then the recursive hiding function is recursively called on these subsets until all non-sensitive subsets of the candidate itemset are added back into the model with the original support (line 12 to 19).

After all non-sensitive itemsets are added to the model, the candidate itemset (sensitive itemset or its superset whose support is greater than specified sensitive support threshold) is removed from the model (line 11).

This process is repeated for all sensitive itemsets for their specified different sensitive support threshold. Since all non-sensitive itemsets with their original support are recursively added to the model, there are no side effects generated. Also, the loss of support information is limited to the inherent loss of support information incurred due to the reduction of the support of sensitive itemsets below their specified sensitive support threshold.

II. Results and Discussions

The model generation is done by extending the CHARM [32] algorithm implementation in SPMF open source data mining library. Though the proposed approach can be compared with many existing techniques that produce side effects, the performance of the proposed RPS-PIH technique is compared with Pseudo Graph-Based Sanitization (PGBS) [30] which is a recent technique known for hiding itemsets at multiple support thresholds.

The algorithm is tested on both real and synthetic datasets shown in Table 2. The retail, connect and chess datasets are popular benchmarked anonymized data used for studying frequent itemset and association rule mining algorithms. The mushroom dataset was prepared by Roberto Bayardo [33]. T10I4D100K and T40I10D100 are synthetic datasets generated by IBM Almaden quest research group. BMS1 and BMS2 are another synthetic datasets used in KDD-CUP 2000 competition.

Table 2: Characteristics of Datasets

Dataset	No. of Transactions	Distinct Items Count	Minimum Transaction Length	Maximum Transaction Length
retail	88162	16470	1	76
mushroom	8124	119	23	23
connect	67557	129	43	43
chess	3196	75	37	37
T10I4D100K	100000	870	1	29
T40I10D100K	100000	942	4	77
BMS1	59602	497	1	267
BMS2	77512	3340	1	161

The set of sensitive itemsets selected as test data are the highly frequent itemsets in their respective datasets. Their sensitive support threshold is selected by reducing their original support count by 10% to 25% for testing purposes. In practice, these sensitive itemsets can be any itemset that the data curator needs to protect from disclosure and sensitive support threshold varies from less than their original support to σ_{model} of its model.

The performance of the approaches considered is measured using side effects and information loss. Among different metrics used to measure the efficiency of itemset hiding techniques, Hiding Failure (HF), False Negative (FN) and False Positive (FP) metrics are used to evaluate the RPS-PIH technique.

- The Hiding Failure provides the measure of the efficiency of hiding sensitive itemsets and is defined as the number of sensitive itemsets whose support is still more than specified sensitive support threshold in the sanitized model.

- False Negative and False Positive assess side effects on the non-sensitive itemsets.

    - False Negative is the number of non-sensitive itemsets whose support is decreased from its original support in the sanitized model making it infrequent from frequent.

    - False Positive is the number of non-sensitive itemsets whose support has been increased from its original support in the sanitized model making it frequent from infrequent.

From the experiment, it is observed that both RPS-PIH and PGBS methods do not produce Hiding Failure and False Positive. However, PGBS suffers from False Negative, and RPS-PIH does not produce False Negatives.

The information loss is calculated by the difference in support of itemsets before and after sanitization as suggested in [34] and is given by Eq. 1.

$$IL(D,D`) = \frac{1}{\sum_{i=1}^{n} f_d(i)} \times \sum_{i=1}^{n} |f_d(i) - f_{d`}(i)| \qquad (1)$$

Where i is the frequent itemset, $f_D(i)$ is the support count of itemset i in the original model and $f_{D`}(i)$ is the support count of itemsets in the sanitized model. The total number of frequent itemsets in the database is denoted by n.

The side effect and information loss in percentage in the proposed RPS-PIH compared with PGBS technique is given in Table3.

From the experimental results presented in Table 3, it is ascertained that for any size of sensitive itemsets, the RPS-PIH produce no side effects. Since RPS-PIH generates no side effects, the information loss is limited to the amount of support reduced in sensitive itemsets during the hiding process. So, the information loss in RPS-PIH is minimal compared to PGBS technique. So, it is evident that the proposed RPS-PIH algorithm effectively hides sensitive itemsets at their respective specified sensitive support threshold. Also, this technique prevents the generation of any side effects and limits the information loss to minimal. The RPS-PIH technique is better than PGBS in hiding itemsets at multiple support thresholds.

Table 3: Side Effects and Information loss in PGBS and RPS-PIH techniques

Dataset/Model	**Sensitive Itemsets**	**False Negatives**		**% Information Loss**	
		PGBS	***RPS-PIH***	***PGBS***	***RPS-PIH***
retail $\sigma_{model} = 0.0005$	10	5866	0	19.799	2.383
	30	6840	0	22.711	4.246
rmushroom $\sigma_{model} = 0.1$	10	7075	0	5.195	0.019
	30	3755	0	41.585	0.064
connect $\sigma_{model} = 0.85$	10	27704	0	85.645	0.486
	30	3567	0	85.65	0.733
chess $\sigma_{model} = 0.7$	10	7950	0	70.645	0.06
	30	4398	0	75.872	0.113
T10I4D100K $\sigma_{model} = 0.00035$	10	4877	0	0.53	0.086
	30	11560	0	0.301	0.375
T40I10D100K $\sigma_{model} = 0.011$	10	1143	0	2.053	0.1
	30	6007	0	3.321	0.25
BMS1 $\sigma_{model} = 0.00085$	10	1984	0	31.379	0.752
	30	1833	0	38.908	1.804
BMS2 $\sigma_{model} = 0.0015$	10	829	0	50.74	0.71
	30	648	0	61.28	1.72

III. Conclusion and Future Scope

With the collection of enormous personal data through online social networking, e-commerce and mobile application services, the risk of disclosing private and sensitive information to adversaries have also increased. Data mining techniques find interesting patterns from the data which are otherwise hidden in the data. These patterns may contain sensitive information which should not be disclosed during the mining process to protect privacy. Itemset hiding is a technique to hide sensitive itemsets in the data. Existing techniques hide sensitive itemsets below a single minimum support threshold and also suffer from undesired side effects and information loss. To address this issue, a model-based approach is proposed which hides sensitive itemsets at multiple support thresholds without generating any side effects. The proposed recursive pattern sanitization with personalized itemsets hiding (RPS-PIH) algorithm sanitizes closed itemsets in the model to reduce the support of sensitive itemsets below their specified minimum support threshold. Since non-sensitive subsets of the sensitive itemsets are added back to the model during sanitization, there are no side effects produced. Also, the information loss due to hiding is also minimal. The empirical results indicated that the efficiency of the proposed technique in hiding sensitive itemsets outperforms the PGBS technique.

However, a limitation of the proposed system requires modeling of the data as closed itemsets which may add extra processing time to the overall data analysis process for large and dense datasets.

For future research, the RPS-PIH algorithm shall be designed to run parallel in the distributed computing environment to hide sensitive itemsets are multiple support thresholds in Big Data.

References

[1] Clifton C, Marks D, *"Security and privacy implications of data mining"*. In Proceedings of the 1996 ACM SIGMOD International Conference on Management of Data, pp 15-19, 1996.

[2] Clifton C, Kantarcioglu M, Vaidya J, *"Defining privacy for data mining"*. National Science Foundation Workshop on Next Generation Data Mining (WNGDM), pp 126-133, 2002.

[3] Agrawal R and Srikant R, *"Privacy-preserving data mining"*.In Proceedings of the ACM SIGMOD, ACM: 439-450, 2000.

[4] S. Sathyamoorthy, "*Data Mining and Information Security in Big Data*", International Journal of Scientific Research in Computer Science and Engineering, Vol.5, Issue.3, pp.86-91, 2017.

[5] G. Pannu, S. Verma, U. Arora, and A. K. Singh, "*Comparison of various Anonymization Technique*", International Journal of Scientific Research in Network Security and Communication, Vol.5, Issue.6, pp.16-20, 2017

[6] Agrawal R, Srikant R, *"Fast algorithms for mining association rules in large databases"*. In Proceedings of the20th International Conference on Very Large Databases, pp. 487-499, 1994.

[7] Bodon F, *"A fast APRIORI implementation"*. Workshop Frequent Itemset Mining Implementations (FIMI03), vol.90, pp. 56-65, 2003.

[8] Brijs, T., Swinnen, G., Vanhoof, K., Wets, G., *"Using association rules for product assortment decisions: a case study"*. In Knowledge Discovery and Data Mining, pp.254-260, 1999.

[9] Zheng Z, Kohavi R, Mason L, *"Real world performance of association rule algorithm"*. In Proceedings of 7thACM-SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 401-406, 2001.

[10] Atallah M, Bertino E, Elmagarmid A, Ibrahim M, Verykios VS, *"Disclosure limitation of sensitive rules"*. In Workshop on Knowledge and Data Engineering Exchange, pp. 45-52, 1999.

[11] Pontikakis E D, Tsitsonis A A, Verykios V S, *"An experimental study of distortion-based techniques for association rule hiding"*. In Proceedings of the 18th Conference on Database Security (DBSEC 2004), pp. 325-339, 2004.

[12] Dasseni E, Verykios V, Elmagarmid A, Bertino E, *"Hiding association rules by using confidence and support"*. In Proceedings of the 4th International Workshop on Information Hiding, IHW, pp. 369-383, 2001.

[13] Oliveira S R M, Zaiane O R, *"Privacy-preserving frequent itemset mining"*. In Proceedings of the IEEE International Conference on Privacy, Security and Data Mining, CRPIT, pp. 43-54, 2002.

[14] Oliveira S R M, Zaiane O R, *"Protecting sensitive knowledge by data sanitization"*. In Proceedings of the Third IEEE International Conference on Data Mining (ICDM2003), pp. 211-218, 2003.

[15] Verykios V S, Emagarmid A K, Bertino E, Saygin Y, Dasseni E, *"Association rule hiding"*. IEEE Transactions on Knowledge and Data Engineering, 16(4), pp. 434-447, 2004.

[16] Wu Y H, Chiang C M, Chen A L P, *"Hiding sensitive association rules with limited side effects"*. IEEE Transactions on Knowledge and Data Engineering, 19(1), pp.29-42, 2007.

[17] Aniket Patel, Patel Shreya, Kiran Amin, "*A Survey on Heuristic Based Approach for Privacy Preserving in Data Mining*", International Journal of Scientific Research in Computer Science and Engineering, Vol.5, Issue.5, pp.21-25, 2017.

[18] Mohnish Patel, Aasif Hasan and Sushil Kumar, "*A Survey: Preventing Discovering Association Rules For Large Data Base*", International Journal of Scientific Research in Computer Science and Engineering, Vol.1, Issue.3, pp.35-38, 2013

[19] Moustakides G V, Verykios V S, *"A max-min approach for hiding frequent itemsets"*. In Proceedings of the 6th IEEE International Conference on Data Mining (ICDM 2006), pp. 502-506, 2006.

[20] Leloglu E, Ayav T, Ergenc B, "Coefficient-based exact approach for frequent itemset hiding". In eKNOW2014: The 6th international conference on information, process, and knowledge management, pp. 124-130, 2014.

[21] Sun X, Yu PS, *"A border-based approach for hiding sensitive frequent itemsets"*. In Proceedings of the Fifth IEEE International Conference on Data Mining (ICDM2005), pp. 426-433.

[22] Sun X, Yu PS, *"Hiding sensitive frequent itemsets by a border-based approach"*. Computing Science and Eng.,1(1), pp. 74-94, 2007.

[23] Gkoulalas-Divanis A and Verykios VS, *"An integer programming approach for frequent itemset hiding"*. In Proceedings of the 15th ACM International Conference on Information and Knowledge Management, CIKM, pp.748-757, 2006.

[24] Menon S, Sarkar S, Mukherjee S, *"Maximizing accuracy of shared databases when concealing sensitive patterns"*. Info. Sys. Research 16, 3, pp. 256-270, 2005.

[25] Kantarcioglu M, Jin J, Clifton C, *"When do data mining results violate privacy?"* In Proceedings of the 10th ACMSIGKDD international conference on knowledge discovery and data mining (KDD04), pp. 599-604, 2004.

[26] Elias C. Stavropoulos, Vassilios S. Verykios, and Vasileios Kagklis. "A transversal hyper-graph approach for the frequent itemset hiding problem." Knowledge and Information Systems 47, 3, pp. 625-645, 2016.

[27] Akbar Telikani and Asadollah Shahbahrami. *"Optimizing association rule hiding using combination of border and heuristic approaches"*. Applied Intelligence 47, 2, pp. 544-557, 2017.

[28] Surendra H and Mohan H S, *"Hiding sensitive itemsets without side effects"*. Applied Intelligence.10.1007/s10489-018-1329-5, 2018.

[29] Gkoulalas-Divanis A, Verykios VS, *"Hiding sensitive knowledge without side effects"*. Knowledge and Information Systems20(3), pp. 263-299, 2009.

[30] Ahmet Cumhur ztrk and Belgin Ergen Bostanolu, *"Itemset Hiding under Multiple Sensitive Support Thresholds"*. In Proceedings of 9th International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management (IC3K 2017) – vol 3: KMIS, pp, 222-231, 2017.

[31] Ztrk, Ahmet Cumhur, Belgin Ergen, *"Dynamic Itemset Hiding Algorithm for Multiple Sensitive Support Thresholds"*.IJDWM 14.2, pp 37-59, 2018.

[32] Mohammed J. Zaki and Ching-Jui Hsiao, *"CHARM: An efficient algorithm for closed itemset mining"*. In Proceedings of International Conference on Data Mining, pp. 457-473, 2002.

[33] Bayardo R, *"Efficiently mining long patterns from databases"*. In Proceedings of the 1998 ACM-SIGMOD International Conference on Management of Data (SIGMOD98), pp 85-93, 1998.

[34] Bertino E, Lin D, Jiang W, *"A Survey of Quantification of Privacy Preserving Data Mining Algorithms"*. In Aggarwal C.C., Yu P.S. (eds) Privacy-Preserving Data Mining. Advances in Database Systems, vol 34. Springer, Boston, MA, 2008.

Authors Profile

Mr. Surendra H received the B.E degree in electronics and communications engineering, and M.Tech degree in computer science and engineering from Visveswaraya Technological University, Belgaum, India in the year 2004 and 2013 respectively. He is currently pursuing Ph.D. in the Department of Information Science and Engineering of SJB Institute of Technology, Bengaluru, India. He was a software engineer with Ingersoll Rand Engineering and Technology Center, Bengaluru. His interests are data science, big data, and information privacy.

Mr. Mohan H S received the Bachelor's degree in computer science and engineering from Malnad College of Engineering, Hassan, India in the year 1999, M.Tech in computer science and engineering from Jawaharlal Nehru National College of Engineering, Shimoga, India in the year 2004 and Ph. D. in computer science & engineering from Dr. MGR University, Chennai, India. He is working as a Professor and Head in the Department of Information Science and Engineering at SJB Institute of Technology, Bengaluru, India. He is having a total of 19 years of teaching experience. His area of interests is Networks Security, Image processing, Data Structures, Computer Graphics, Finite Automata, and Formal Languages and Compiler Design. He has obtained the Best Teacher award for his teaching during the year 2008 at SJB Institute of Technology, Bengaluru, India. He has published and presented papers in journals, international and national conferences.