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

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 6, June 2016

Improving Frequent Itemset Mining by Bifurcating Transactions

Aakash P Joshi, Prof. P. M. Kamde

PG Student, Department of Computer Engineering, Sinhgad College of Engineering, Pune, India

Associate Professor, Department of Computer Engineering Sinhgad College of Engineering, Pune, India.

ABSTRACT: There is need to protect the data and there has been huge rising curiosity in differential private data-mining algorithm. Differential privacy plans to give intends to amplify the exactness of questions from measurable databases while minimizing the odds of recognizing its records. Frequent itemset mining (FIM) is one of the most primary problems in data mining. In the existing system a differentially private FIM algorithm proposed which is based on the FP-growth algorithm mentioned as PFP-growth (Private- FP). The PFP-algorithm made up of two phases, pre-processing and mining phase. In the Pre-processing phase, smart splitting method is used to transform the database. In the mining phase, a run-time estimation method is used to measure the actual support of itemsets in the original database. But in smart splitting method, splitting transitions are sent one by one to the mining phase requires high memory and time to the system. Therefore in this paper the PFP-growth (Parallel-FP) is proposed which overcomes the memory and time problem. It uses the Thread instead of splitting and thread executes the operation in parallel.

KEYWORDS: Data-mining, information loss, transaction splitting and frequent itemset mining.

I. Introduction

Frequent itemset mining plays an essential role in many data mining tasks that try to find interesting patterns from databases. The mining of data is one of the most popular problems of all these. The distinguishing proof of sets of things, items, side effects and qualities which regularly happen together in the given database, can be seen as a standout amongst the most essential undertakings in Data Mining.

The original motivation for searching frequent itemset came from the need to analyze so called supermarket transaction data that is to examine customer behavior. In existing system, private FP-growth (PFP-development) algorithm presented which comprised of a preprocessing phase and a mining phase. In the preprocessing phase, Database is changed to restrict the length of exchanges. The preprocessing stage is limited to user specified limits and should be performed just once for a given database. They argue to insist on this type of limit, lengthy transaction should be split i.e. from sub-transactions rather than truncate. For this a Smart Splitting technique is used to split the transactions.

In the Mining phase, the operation is performed on transformed dataset which is output of preprocessing phase. In this phase, Run-Time-Estimation and Dynamic Reduction technique is used. Run-Time-Estimation technique applied to find out information loss during mining phase. For maintain the privacy needs to add some amount of noise in the transactions. For finding the actual support of transactions (Original Database) and final support of transaction, Run-Time-Estimation is used. Dynamic Reduction technique is used to remove the noisy items in the transaction at the final stage i.e. in the mining phase after performing the run-Time-Estimation.

In Smart Splitting technique, after splitting the long transactions into sub-transactions it sends that transaction and their sub-transactions to mining phase one-by-one. So it takes high time to complete an operation and occupies memory to store all transactions for mining phase. To overcome this, we propose the Parallel FP-Growth algorithm. The PFP consists of five different steps, (Sharding, Parallel counting, Grouping Items, Parallel FP-Growth and Aggregation). It performs the operation by dividing the dataset using the Sharding method and assigns to the different-different threads and parallel Mining of the itemsets is done. In existing system we send the truncated translation one-by-one. It overcomes this by sending transactions in parallel. At the end of step grouping done, FP-tree algorithm is generated. In aggregation step, gather all the items sets again to generate the frequent patterns.In this paper, further we will see: Section II talks about related work studied till now on topic. Section III includes current implementation details, introductory definitions.

II. RELATED WORK

In this section discuss existing work done by the researchers for text mining process.In paper [1], to investigate the matter of planning a differentially private FIM algorithm presented. Private FP-growth (PFP-growth) algorithm was proposed that consisted of a preprocessing phase and a mining part. Within the preprocessing phase, to improve the utility-privacy tradeoff, conceived a wise rending methodology to remodel the information. In the mining part, a run-time estimation methodology planned to offset the knowledge loss incurred by group action splitting. Moreover, by using the downward closure property, dynamic reduction methodology applied to dynamically cut back the quantity of noises other to ensure privacy throughout the mining method. Formal privacy analysis and the results of intensive experiments on real datasets show that our PFP-growth algorithm is time-efficient and can achieve each sensible utility and privacy.

In paper [2], explained about parallel FP-Growth algorithm. The formula was predicated on a completely unique knowledge and computation distribution theme that eliminates communication among computers and makes it attainable for us to specify the formula with the Map-Reduce model. Experiments can be done on a vast dataset will give outstanding scalability of this formula. To form the formula suitable for mining internet knowledge, that are sometimes of long tail distribution, a tendency is to designed this formula to mine top-k patterns associated with every item, instead of wishing on a user specified worth for international smallest support threshold. Demonstration is done such that PFP is effective in mining associations and Webpage associations to support query recommendation or connected search.

In paper [3], it showed that a k-anonymity dataset permits solid attacks unpaid to lack of diversity within the sensitive attributes. Author introduced ℓ-diversity, a framework that gives stronger privacy guarantees. There square measure many avenues for future work. First is, to extend initial concepts for handling multiple sensitive attributes, and to develop ways for continuous sensitive attributes. Second, though privacy and utility square measures duals of alternative privacy, it has received far more attention than the utility of a broadcast table. As a result, the concept of utility wasn't well-understood.

In [4], author displayed two new algorithms, Apriori and Apriori-Tid, for finding all significant affiliation rules between things in a substantial database of exchanges. Author contrasted these algorithms with the beforehand known algorithms, the AIS and SETM algorithms. Author exhibited trial results, appearing that the proposed algorithms dependably outflank AIS what's more, SETM. The execution hole expanded with the issue measure, and ran from an element of three for little issues to more than a request of greatness for huge issues.

In paper [5] Sequential data is progressively utilized in many application. Publishing sequential data is of basic importance to the advancement of these applications. However, as appeared by the re-identification attacks on the AOL and Netix datasets, discharging sequential data might posture extensive dangers to individual privacy. Differential privacy is one of the main models that can be utilized to give such ensures. Our approach is to design tree structure and set of novel techniques depends on the Markov assumption for minimizing the added noise . The distributed n - grams are valuable for some reasons. Besides, we add to an answer for producing an synthetic database, which empowers a more wider spectrum of data analysis tasks. Broad tests on real-life datasets show that our methodology significantly outperforms the state-of-the-art techniques.

A. System Overview

The following figure 1 shows the architectural view of the proposed system. The description of the system is as follows:

In the proposed system initially input is a Dataset file. Input is given to the Parallel FP-Growth (PFP) algorithm .For the Database DB,PFP utilizes three Map-Reduce stages to parallelize PF-Growth.

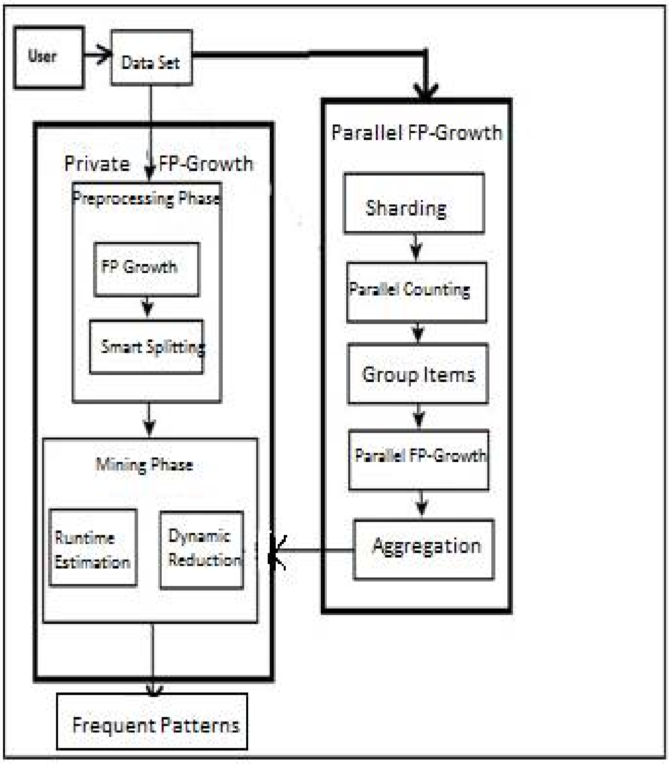


Figure 1: System Architecture

B. Algorithm

Proposed System Algorithm
Algorithm 1: The Parallel Counting Algorithm

Input: Dataset.
Output: Divided dataset.

Process:
Step 1: Input the Dataset.
Step 2: Sharding the dataset in number of shard.
Step 3: Assign each shard to each different Thread.
Step 4: Perform parallel counting.
Step 5: Grouping Items.

Algorithm 2: The Parallel FP-Growth Algorithm

Input: Divided dataset.

Output: Frequent Itemset in each thread.

Process:
Step 1: Generate the FP-tree
Step 2: Generate the header table

Step 3: Create a Mapper.
Step 4: Generate Hash Table *H* from G-List.
Step 5: Reducer

Algorithm 3: The Aggregating Algorithm

Input: Frequent Itemset in each thread.
Output: Frequents Patterns.

Process:
Step 1: Getting all threads.
Step 2: Create a Mapper
Step 3: Generate Hash Table *H* from G-List
Step 4: Reducer

C. PROPOSED ALGORITHM

System S is represented as S = {I, F, N, U, L, T, K, P, R, D, O }

Input:
I = {D, N} Where D = Sparse datasets: Retail Dataset a

N = {n1, n2, n3, n4 , n5}

Where, N is the set of Noise and n1, n2, n3, n4 , n5 are the noise sample.

Process:

- F = FP Growth algorithm
- Noise
  N = n1
  Where N is a noise and n1 is a noise sample.

- Undirected Weighted Graph U = { u1, u2, u3,....,un }

Where U is represent as a graph and u1, u2, u3, ...,un
Number of vertices from the graph

- L= Louvain Method
- T= CR-Tree

$$\Delta Q = \left[\frac{\sum_{in} + k_{i,in}}{2m} - \left(\frac{\sum_{tot} + k_i}{2m} \right)^2 \right] - \left[\frac{\sum_{in}}{2m} - \left(\frac{\sum_{tot}}{2m} \right)^2 - \left(\frac{k_i}{2m} \right) \right]$$

Where,
ΔQ = used to find the Modularity of node in CR-Tree.

$\sum_{in}$ = total of the considerable number of weights of the links inside the community i is moving into.

$\sum_{tot}$ = som of number of weights of the links to nodes in the community.

k_i = degree of i.

k_{i, in} = total of the weights of the links in the middle of i and different nodes in the community.

and

m = whole of the weights of all links in the networks.

- Splitting K = { k1, k2, k3, ...kn }

Where K be a set of Splitting where, k1, k2, k3,......,kn are the number of split tree.

- P= Parallel FP-Growth
- R= Run-time Estimation

$$\omega_a = \text{avg\_supp}(\varpi, i) = \int_{\omega'=\varpi-5}^{\omega'=\varpi+5} \Pr(\omega' \mid \varpi) avg(\omega', i)$$

For calculating Average Support of ItemSet ,we have

$$\omega_m = \text{max\_supp}(\varpi, i) = \int_{\omega'=\varpi-5}^{\omega'=\varpi+5} \Pr(\omega' \mid \varpi) \max(\omega', i)$$

For calculating Maximum Support of Item Set, we have,

- D= Dynamic Reduction

Output:

O = Reduce noise and Privacy with frequent Patterns

D. Experimental Setup

The system is built using Java framework (version jdk 8) on Windows platform. The Netbeans (version 8.1) is used as a development tool. The system doesn't require any specific hardware to run; any standard machine is capable of running the application.

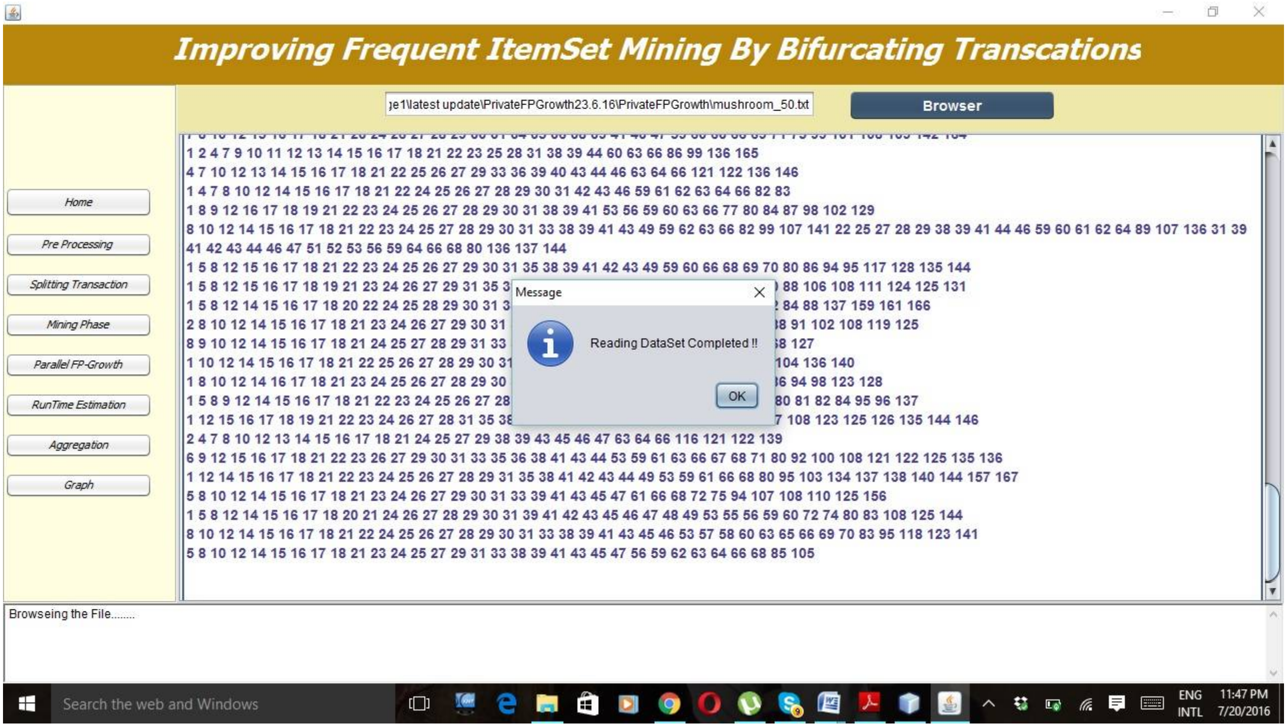
IV. RESULT AND DISCUSSION

a. DataSet

Sparse datasets: Retail Dataset. This dataset contain numeric value for items. Contain 200 transaction of different length.

b. Results

In this paper we find the frequent item-sets. In existing system, we send the truncated transactions one-by-one. In proposed, mining is done parallel. It's sparse the time as well as memory.



```
e1\latest update\PrivateFPGrowth23.6.16\PrivateFPGrowth\mushroom_50.txt
```

```
1 2 4 7 9 10 11 12 13 14 15 16 17 18 21 22 23 25 28 31 38 39 44 60 63 66 86 99 136 165
4 7 10 12 13 14 15 16 17 18 21 22 25 26 27 29 33 36 39 40 43 44 46 63 64 66 121 122 136 146
1 4 7 8 10 12 14 15 16 17 18 21 22 24 25 26 27 28 29 30 31 42 43 46 59 61 62 63 64 66 82 83
1 8 9 12 16 17 18 19 21 22 23 24 25 26 27 28 29 30 31 38 39 41 53 56 59 60 63 66 77 80 84 87 98 102 129
8 10 12 14 15 16 17 18 21 22 23 24 25 27 28 29 30 31 33 38 39 41 43 49 59 62 63 66 82 99 107 141 22 25 27 28 29 38 39 41 44 46 59 60 61 62 64 89 107 136 31 39
41 42 43 44 46 47 51 52 53 56 59 64 66 68 80 136 137 144
1 5 8 12 15 16 17 18 21 22 23 24 25 26 27 29 30 31 35 38 39 41 42 43 49 59 60 66 68 69 70 80 86 94 95 117 128 135 144
1 5 8 12 15 16 17 18 19 21 23 24 26 27 29 31 35 3                                    88 106 108 111 124 125 131
1 5 8 12 14 15 16 17 18 20 22 24 25 28 29 30 31 3                                    84 88 137 159 161 166
2 8 10 12 14 15 16 17 18 21 23 24 26 27 29 30 31                                     8 91 102 108 119 125
8 9 10 12 14 15 16 17 18 21 24 25 27 28 29 31 33                                     8 127
1 10 12 14 15 16 17 18 21 22 25 26 27 28 29 30 31                                    104 136 140
1 8 10 12 14 16 17 18 21 23 24 25 26 27 28 29 30                                     6 94 98 123 128
1 5 8 9 12 14 15 16 17 18 21 22 23 24 25 26 27 28                                    80 81 82 84 95 96 137
1 12 15 16 17 18 19 21 22 23 24 26 27 28 31 35 38                                    7 108 123 125 126 135 144 146
2 4 7 8 10 12 13 14 15 16 17 18 21 24 25 27 29 38 39 43 45 46 47 63 64 66 116 121 122 139
6 9 12 15 16 17 18 21 22 23 26 27 29 30 31 33 35 36 38 41 43 44 53 59 61 63 66 67 68 71 80 92 100 108 121 122 125 135 136
1 12 14 15 16 17 18 21 22 23 24 25 26 27 28 29 31 35 38 41 42 43 44 49 53 59 61 66 68 80 95 103 134 137 138 140 144 157 167
5 8 10 12 14 15 16 17 18 21 23 24 26 27 29 30 31 33 39 41 43 45 47 61 66 68 72 75 94 107 108 110 125 156
1 5 8 12 14 15 16 17 18 20 21 24 26 27 28 29 30 31 39 41 42 43 45 46 47 48 49 53 55 56 59 60 72 74 80 83 108 125 144
8 10 12 14 15 16 17 18 21 22 24 25 26 27 28 29 30 31 33 38 39 41 43 45 46 53 57 58 60 63 65 66 69 70 83 95 118 123 141
5 8 10 12 14 15 16 17 18 21 23 24 25 27 29 31 33 38 39 41 43 45 47 56 59 62 63 64 66 68 85 105
```

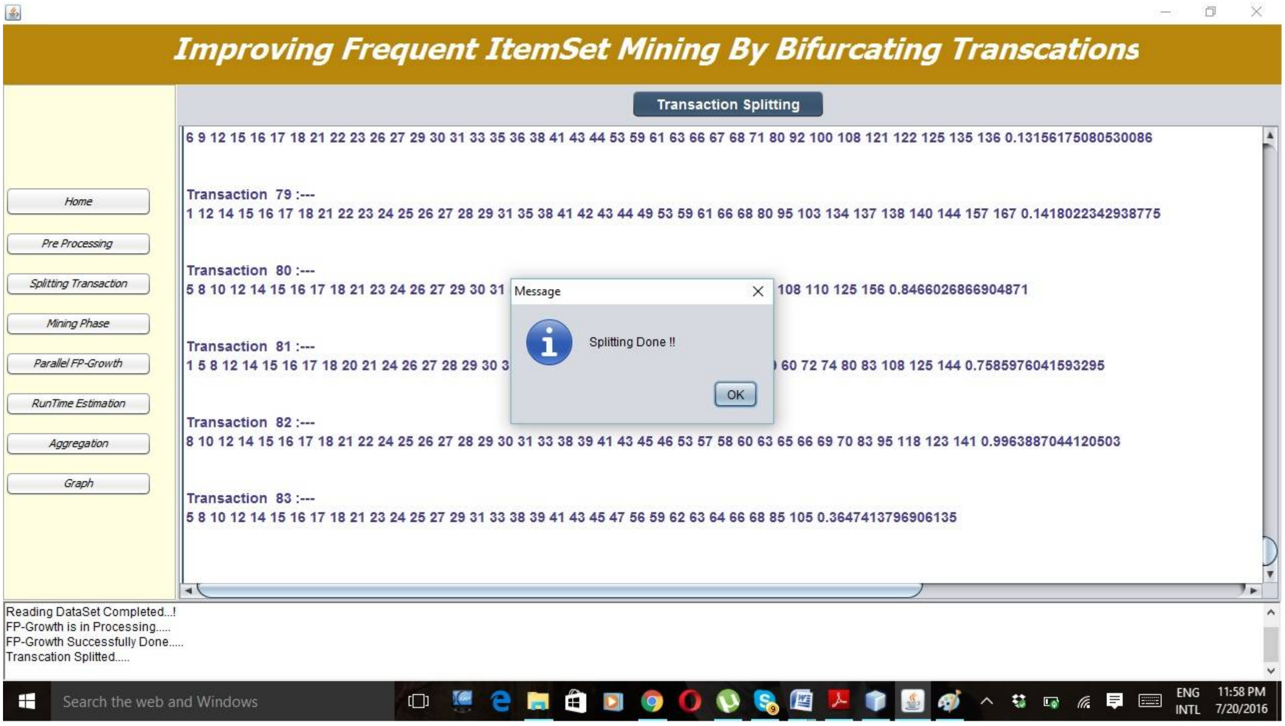
Message

Reading DataSet Completed !!

OK

```
Browseing the File........
```

Figure 2: Reading Dataset



```
6 9 12 15 16 17 18 21 22 23 26 27 29 30 31 33 35 36 38 41 43 44 53 59 61 63 66 67 68 71 80 92 100 108 121 122 125 135 136 0.13156175080530086


Transaction  79 :---
1 12 14 15 16 17 18 21 22 23 24 25 26 27 28 29 31 35 38 41 42 43 44 49 53 59 61 66 68 80 95 103 134 137 138 140 144 157 167 0.1418022342938775


Transaction  80 :---
5 8 10 12 14 15 16 17 18 21 23 24 26 27 29 30 31                                     108 110 125 156 0.8466026866904871


Transaction  81 :---
1 5 8 12 14 15 16 17 18 20 21 24 26 27 28 29 30 3                                     60 72 74 80 83 108 125 144 0.7585976041593295


Transaction  82 :---
8 10 12 14 15 16 17 18 21 22 24 25 26 27 28 29 30 31 33 38 39 41 43 45 46 53 57 58 60 63 65 66 69 70 83 95 118 123 141 0.9963887044120503


Transaction  83 :---
5 8 10 12 14 15 16 17 18 21 23 24 25 27 29 31 33 38 39 41 43 45 47 56 59 62 63 64 66 68 85 105 0.3647413796906135
```

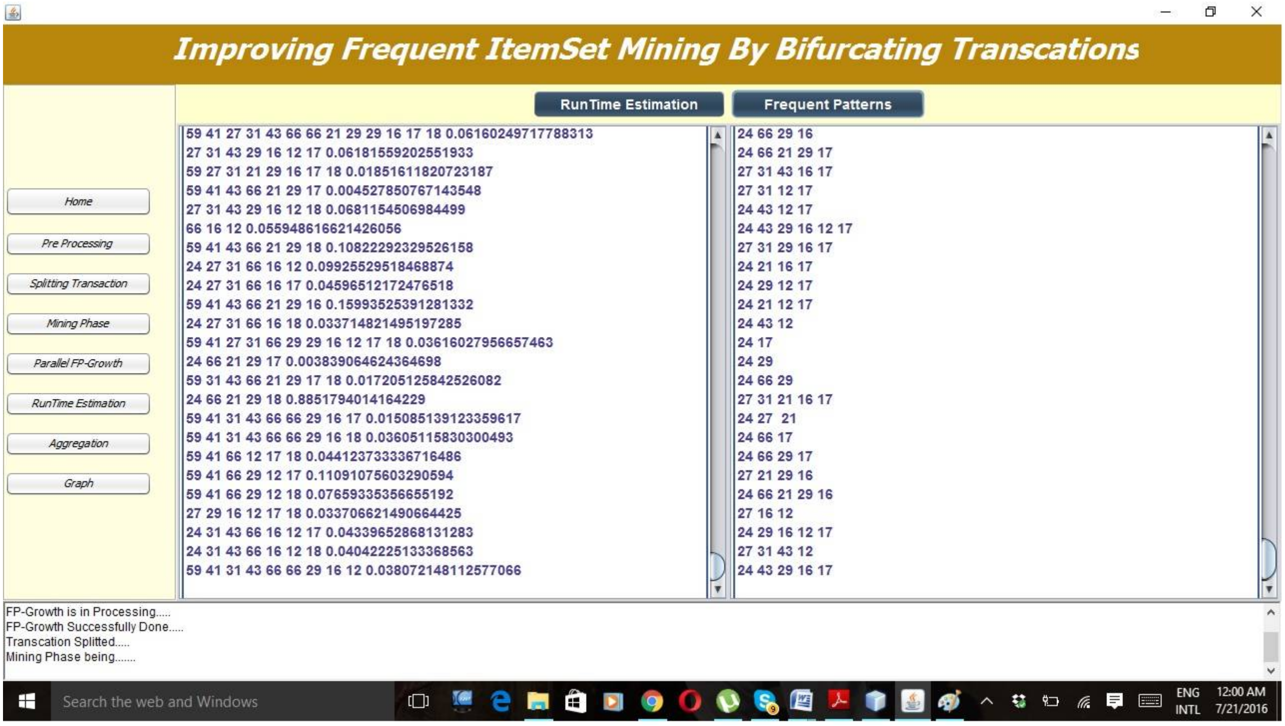
Message

Splitting Done !!

OK

```
Reading DataSet Completed...!
FP-Growth is in Processing.....
FP-Growth Successfully Done.....
Transcation Splitted.....
```

Figure 3: Splitting done in existing System

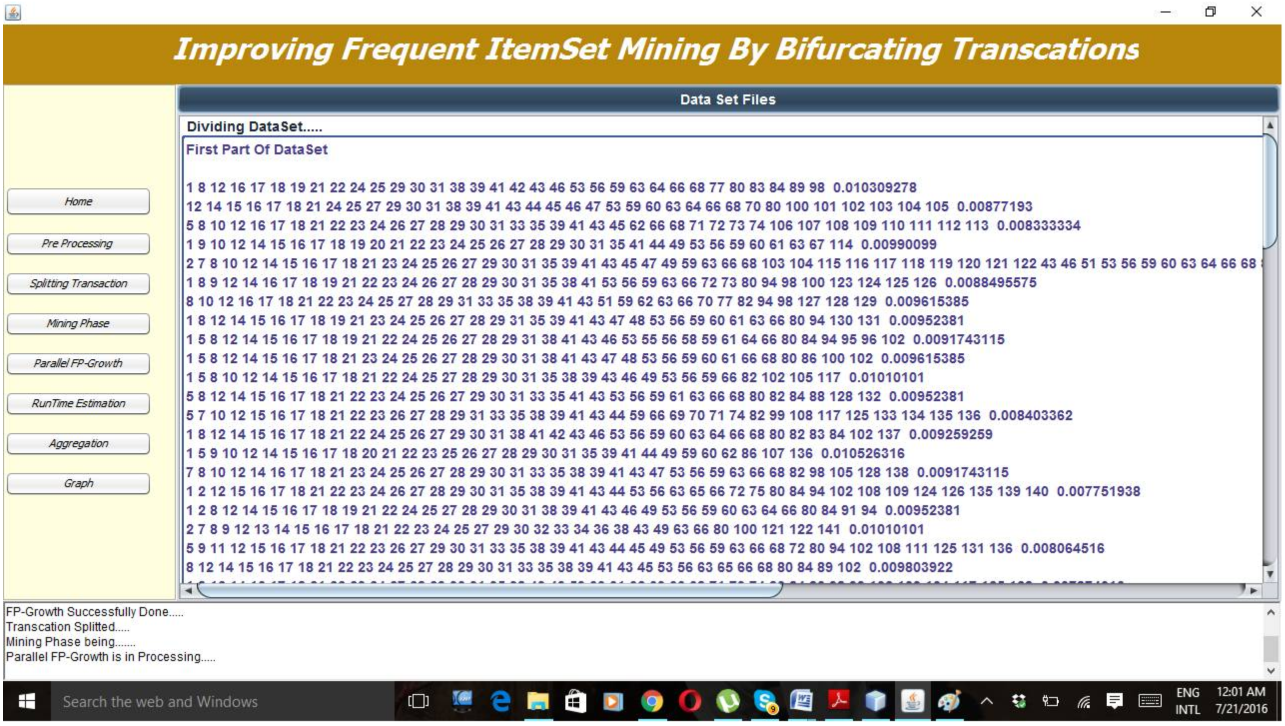


```
59 41 27 31 43 66 66 21 29 29 16 17 18 0.06160249717788313
27 31 43 29 16 12 17 0.0618155920255193 3
59 27 31 21 29 16 17 18 0.01851611820723187
59 41 43 66 21 29 17 0.004527850767143548
27 31 43 29 16 12 18 0.0681154506984499
66 16 12 0.05594861662142605 6
59 41 43 66 21 29 18 0.10822292329526158
24 27 31 66 16 12 0.09925529518468874
24 27 31 66 16 17 0.04596512172476518
59 41 43 66 21 29 16 0.15993525391281332
24 27 31 66 16 18 0.033714821495197285
59 41 27 31 66 29 29 16 12 17 18 0.03616027956657463
24 66 21 29 17 0.003839064624364698
59 31 43 66 21 29 17 18 0.017205125842526082
24 66 21 29 18 0.8851794014164229
59 41 31 43 66 66 29 16 17 0.015085139123359617
59 41 31 43 66 66 29 16 18 0.03605115830300493
59 41 66 12 17 18 0.04412373333671648 6
59 41 66 29 12 17 0.11091075603290594
59 41 66 29 12 18 0.07659335356655192
27 29 16 12 17 18 0.033706621490664425
24 31 43 66 16 12 17 0.04339652868131283
24 31 43 66 16 12 18 0.04042225133368563
59 41 31 43 66 66 29 16 12 0.038072148112577066
```

```
24 66 29 16
24 66 21 29 17
27 31 43 16 17
27 31 12 17
24 43 12 17
24 43 29 16 12 17
27 31 29 16 17
24 21 16 17
24 29 12 17
24 21 12 17
24 43 12
24 17
24 29
24 66 29
27 31 21 16 17
24 27  21
24 66 17
24 66 29 17
27 21 29 16
24 66 21 29 16
27 16 12
24 29 16 12 17
27 31 43 12
24 43 29 16 17
```

```
FP-Growth is in Processing.....
FP-Growth Successfully Done.....
Transcation Splitted.....
Mining Phase being.......
```

Figure 4: Runtime estimation method

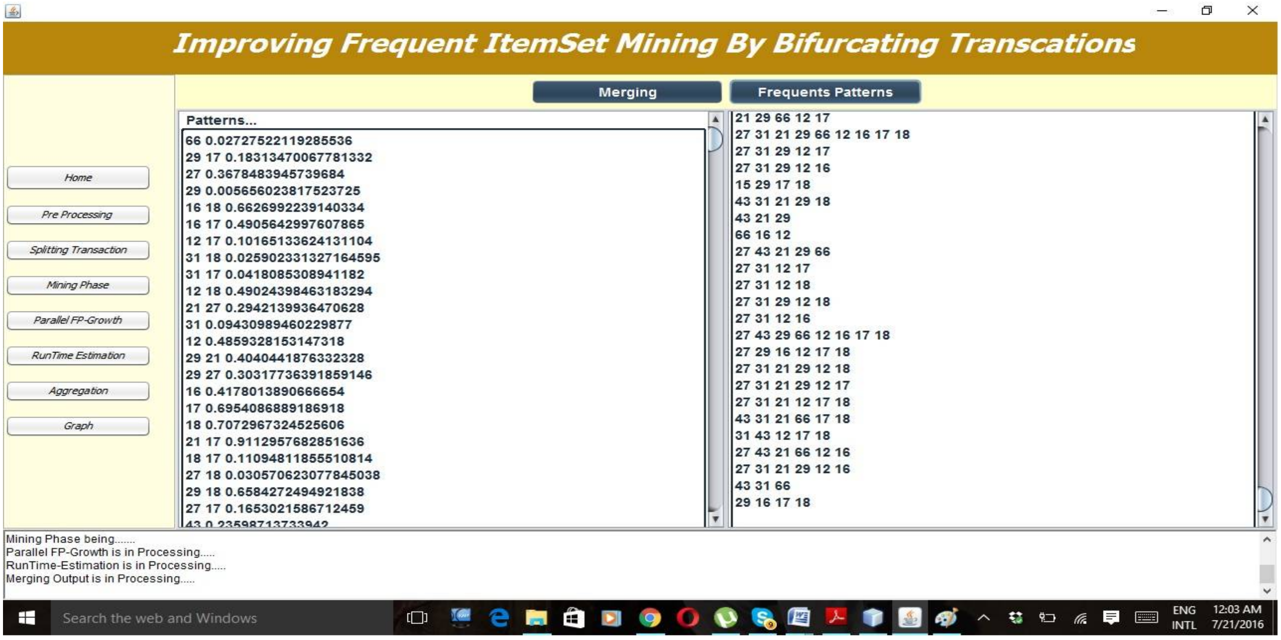


```
Dividing DataSet.....
First Part Of DataSet

1 8 12 16 17 18 19 21 22 24 25 29 30 31 38 39 41 42 43 46 53 56 59 63 64 66 68 77 80 83 84 89 98  0.010309278
12 14 15 16 17 18 21 24 25 27 29 30 31 38 39 41 43 44 45 46 47 53 59 60 63 64 66 68 70 80 100 101 102 103 104 105  0.00877193
5 8 10 12 16 17 18 21 22 23 24 26 27 28 29 30 31 33 35 39 41 43 45 62 66 68 71 72 73 74 106 107 108 109 110 111 112 113  0.008333334
1 9 10 12 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 35 41 44 49 53 56 59 60 61 63 67 114  0.00990099
2 7 8 10 12 14 15 16 17 18 21 23 24 25 26 27 29 30 31 35 39 41 43 45 47 49 59 63 66 68 103 104 115 116 117 118 119 120 121 122 43 46 51 53 56 59 60 63 64 66 68 8
1 8 9 12 14 16 17 18 19 21 22 23 24 26 27 28 29 30 31 35 38 41 53 56 59 63 66 72 73 80 94 98 100 123 124 125 126  0.0088495575
8 10 12 16 17 18 21 22 23 24 25 27 28 29 31 33 35 38 39 41 43 51 59 62 63 66 70 77 82 94 98 127 128 129  0.009615385
1 8 12 14 15 16 17 18 19 21 23 24 25 26 27 28 29 31 35 39 41 43 47 48 53 56 59 60 61 63 66 80 94 130 131  0.00952381
1 5 8 12 14 15 16 17 18 19 21 22 24 25 26 27 28 29 31 38 41 43 46 53 55 56 58 59 61 64 66 80 84 94 95 96 102  0.0091743115
1 5 8 12 14 15 16 17 18 21 23 24 25 26 27 28 29 30 31 38 41 43 47 48 53 56 59 60 61 66 68 80 86 100 102  0.009615385
1 5 8 10 12 14 15 16 17 18 21 22 24 25 27 28 29 30 31 35 38 39 43 46 49 53 56 59 66 82 102 105 117  0.01010101
5 8 12 14 15 16 17 18 21 22 23 24 25 26 27 29 30 31 33 35 41 43 53 56 59 61 63 66 68 80 82 84 88 128 132  0.00952381
5 7 10 12 15 16 17 18 21 22 23 26 27 28 29 31 33 35 38 39 41 43 44 59 66 69 70 71 74 82 99 108 117 125 133 134 135 136  0.008403362
1 8 12 14 15 16 17 18 21 22 24 25 26 27 29 30 31 38 41 42 43 46 53 56 59 60 63 64 66 68 80 82 83 84 102 137  0.009259259
1 5 9 10 12 14 15 16 17 18 20 21 22 23 25 26 27 28 29 30 31 35 39 41 44 49 59 60 62 86 107 136  0.010526316
7 8 10 12 14 16 17 18 21 23 24 25 26 27 28 29 30 31 33 35 38 39 41 43 47 53 56 59 63 66 68 82 98 105 128 138  0.0091743115
1 2 12 15 16 17 18 21 22 23 24 26 27 28 29 30 31 35 38 39 41 43 44 53 56 63 65 66 72 75 80 84 94 102 108 109 124 126 135 139 140  0.007751938
1 2 8 12 14 15 16 17 18 19 21 22 24 25 27 28 29 30 31 38 39 41 43 46 49 53 56 59 60 63 64 66 80 84 91 94  0.00952381
2 7 8 9 12 13 14 15 16 17 18 21 22 23 24 25 27 29 30 32 33 34 36 38 43 49 63 66 80 100 121 122 141  0.01010101
5 9 11 12 15 16 17 18 21 22 23 26 27 29 30 31 33 35 38 39 41 43 44 45 49 53 56 59 63 66 68 72 80 94 102 108 111 125 131 136  0.008064516
8 12 14 15 16 17 18 21 22 23 24 25 27 28 29 30 31 33 35 38 39 41 43 45 53 56 63 65 66 68 80 84 89 102  0.009803922
```

```
FP-Growth Successfully Done.....
Transcation Splitted.....
Mining Phase being.......
Parallel FP-Growth is in Processing.....
```

Figure 5: Dataset Division in Proposed Parrallel FP growth Algorithm.



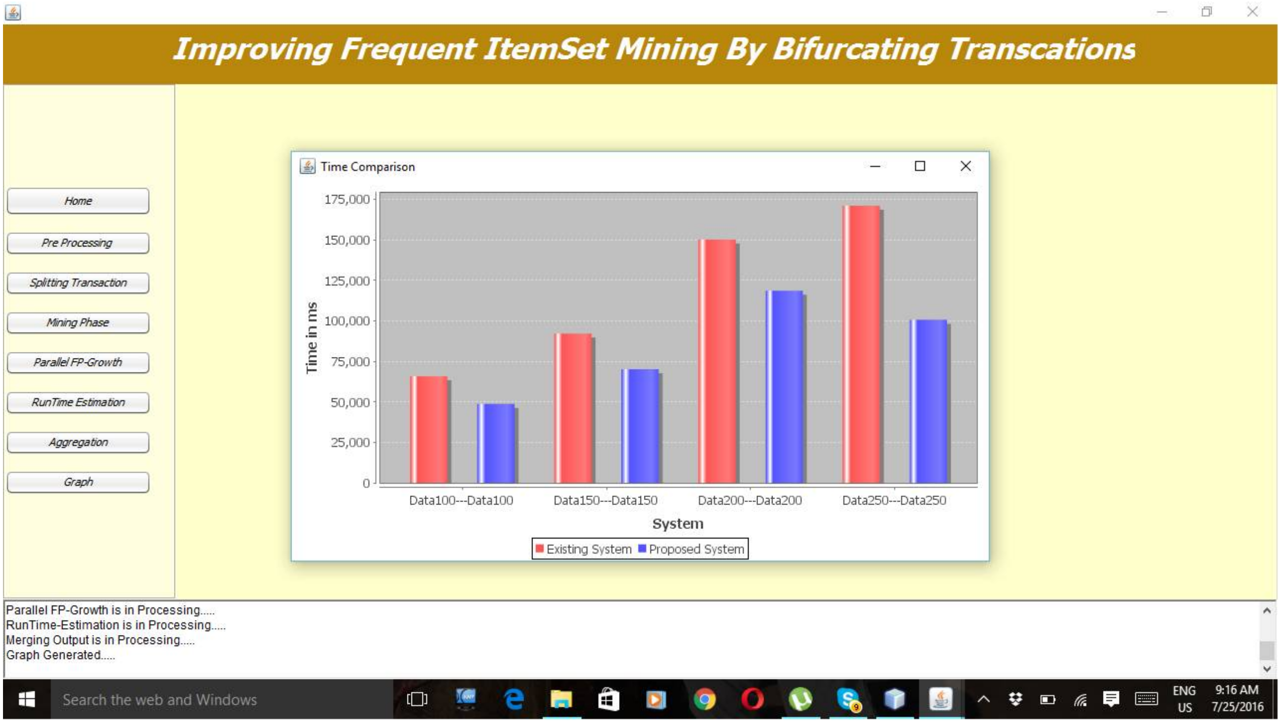
```
Patterns...
66 0.02727522119285536
29 17 0.18313470067781332
27 0.3678483945739684
29 0.005656023817523725
16 18 0.6626992239140334
16 17 0.4905642997607865
12 17 0.10165133624131104
31 18 0.025902331327164595
31 17 0.0418085308941182
12 18 0.49024398463183294
21 27 0.2942139936470628
31 0.09430989460229877
12 0.4859328153147318
29 21 0.4040441876332328
29 27 0.30317736391859146
16 0.4178013890666654
17 0.6954086889186918
18 0.7072967324525606
21 17 0.9112957682851636
18 17 0.11094811855510814
27 18 0.030570623077845038
29 18 0.6584272494921838
27 17 0.1653021586712459
43 0.23598713733942
```

```
21 29 66 12 17
27 31 21 29 66 12 16 17 18
27 31 29 12 17
27 31 29 12 16
15 29 17 18
43 31 21 29 18
43 21 29
66 16 12
27 43 21 29 66
27 31 12 17
27 31 12 18
27 31 29 12 18
27 31 12 16
27 43 29 66 12 16 17 18
27 29 16 12 17 18
27 31 21 29 12 18
27 31 21 29 12 17
27 31 21 12 17 18
43 31 21 66 17 18
31 43 12 17 18
27 43 21 66 12 16
27 31 21 29 12 16
43 31 66
29 16 17 18
```

```
Mining Phase being.......
Parallel FP-Growth is in Processing.....
RunTime-Estimation is in Processing.....
Merging Output is in Processing.....
```

Figure 6: Aggregation and final frequent patterns.

For different value of threshold, estimated result show's that, the memory and time require for the proposed system is less than existing system.



```
Parallel FP-Growth is in Processing.....
RunTime-Estimation is in Processing.....
Merging Output is in Processing.....
Graph Generated.....
```

Figure 7: Time Comparison graph.

Conclusion and Future scope

In this paper, we explore the issue of planning a differentially private FIM algorithm. We propose our parallel FP-growth (PFP-growth) algorithm, which consists of a five different steps. PFP-algorithm is time-productive and require less memory.

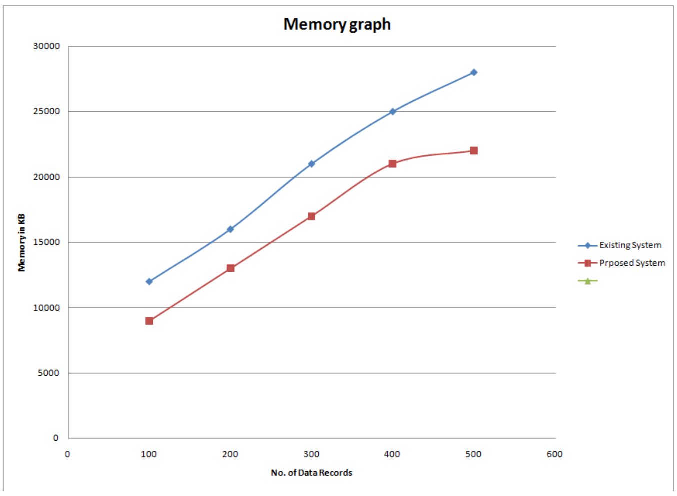


Figure 8: Memory Comparison graph.

REFERENCES

[1]Sen Su, Shengzhi Xu, Xiang Cheng, Zhengyi Li, and Fangchun Yang, "Differentially Private Frequent Itemset Mining via Transaction Splitting", IEEE Transactions on Knowledge and Data Engineering,DOI10.1109/ TKDE.2015.2399310 ,2015.

[2] Haoyuan Li, Yi Wang, Dong Zhang, Ming Zhang, Edward Chang," PFP: Parallel FP-Growth for Query Recommendation ", Google Beijing Research, Beijing, 100084, China.

[3] Ashwin Machanavajjhala ,Johannes Gehrke, Daniel Kifer,Muthuramakrishnan Venkitasubramaniam, "ℓ-Diversity: Privacy Beyond k-Anonymity", Department of Computer Science, Cornell University.

[4] Rakesh Agrawal, Ramakrishnan Srikant, "Fast Algorithms for Mining Association Rules", IBM Alma den Research Center 650 Harry Road, San Jose, CA 95120

[5] R. Chen, G. Acs, and C. Castelluccia, "Differentially Private Sequential Data Publication via Variable-Length n-Grams," Proceedings of the 2012 ACM conference on Computer and communications security. ACM , 2012.

[6] L. Bonomi and L. Xiong, "A Two-Phase Algorithm for Mining Sequential Patterns with Differential privacy," CIKM, 2013.

[7] E. Shen and T. Yu, "Mining Frequent Graph Patterns withDifferential Privacy," in KDD, 2013.

[8]M. Hay, V. Rastogi, G. Miklau, and D. Suciu. "Boosting the Accuracy of Differentially Private Histograms through Consistency". VLDB, 3(1):1021-1032, 2010.

[9] X. Xiao, G.Wang, and J. Gehrke. "Differential privacy via Wavelet Transforms", IEEE Transaction on Knowledge. Data Eng., 23(8):1200-1214, 2011.

[10]Chen Zeng and Jin-yi Cai, "On Differentially Private Frequent Itemset Mining", Proceedings of VLDB Endowment, Vol 6, NO. 1, 26th august- 30th 2013.

11]Raghav Bhaskar, Srivatsan Laxman, "Discovering Frequent Patterns in Sensitive Data" ACM KDD, 2010, Washington DC.

[12]Ninghui Li, W. Qardaji, Dong Su, Jianneng Cao,"PrivBasis: Frequent Itemset Mining with Differential Privacy " Proceedings of the VLDB Endowment, Vol 5,No.11, August 27th-31st 2012 Istanbul Turkey.

BIOGRAPHY

Aakash Joshi is currently pursuing M.E(Computer Networks) from Department of Computer Engineering, Sinhgad College of Engineering, Pune,Savitribai Phule Pune University, Pune, Maharashtra, India-411041. He received his B.E (Computer Engineering) Degree in 2013 from Trinity College of Engineering,Pune, Savitribai Phule Pune University, Pune, Maharashtra,
India -411048. His area of interest is Data Mining and Information Security.

Prof. Pravin M. Kamde is a Associate Professor in the Department of Computer Engineering at Sinhgad College of Engineering,Pune,India. He has received his degree of B.E.(Computer Science and Engineering) from SGGS College of Engineering and Technology, Nanded,Marathwada University of Aurangabad in 1993,M.E.(Computer Science and Engineering) in 2004 from Walchand College of Engineering, Sangli, Shivaji University. He has 13 Internal Journal Publications, 7 International Conferences and 14 National Conferences publication. his research interests include Content-Based Image and video Retrievaal, Web Multimedia Mining and Image Processing.