RESEARCH ARTICLE OPEN ACCESS

A Study on Call Data Records in Data Centres: Incoming Calls and Outgoing Calls

V.Sathyendra Kumar [1], Dr.A.Muthukumarvel [2], C.Sasidhar[3], P.Kavitha[4]

1. (Research Scholar, Department of MCA, Bharath Institute of Higher Education & Research (Declared as Deemed to be University), Chennai-600 073, Tamil Nadu, Phone: 09985666531, e-mail: vsk9985666531@gmail.com)
2. Dean, Faculty of Arts & Science, Professor, Bharath Institute of Higher Education & Research (Declared as Deemed to be University),Chennai-600 073, Tamil Nadu, India,Phone:09500137273, e-mail:muthu14673@gmail.com.
3. (Assistant Professor, Department of MCA, Annamacharya Institute of Technology & Sciences(Autonomous), Rajampet, Kadapa(Dt.), Andhra Pradesh, Phone: 09848779497, e-mail: csasidhar.aits@gmail.com)
4. (Assistant Professor, Department of MCA, Annamacharya Institute of Technology & Sciences(Autonomous), Rajampet, Kadapa(Dt.), Andhra Pradesh, Phone: 09703489363, e-mail: kavithapaleti@gmail.com)

Abstract:

With the quick development of knowledge in Mobile Communication the difficult issue for providing services to their users. Frequent Pattern mining are often utilized in the sector of mobile communications for mining of cluster patterns from mobile user movement information, for client behavior prediction, for predicting future location of a mobile user for location primarily based services and for mining patterns helpful for mobile commerce. improvisation data processing technology, some helpful information are often discovered from network traffic information and invasive behavior and traditional behavior are often established then the abnormal acts are going to be detected from the time period information. In mobile databases, to handle voluminous quantity of user's records and notice the amount of frequent patterns mixtures and hidden data. The information goes up exponentially and needs simulated machine time. The most important downside is that to access the user's record and to filter the foremost important frequent item sets and to accelerate the process and scale back error rate in classification result. With this regards it's attainable to maximize the out turn and minimize the service time for the user. To scale back the load on network traffic within the information data centres.

Keywords—**Pattern Mining; Call Data Records; Network Traffic; Data Centres**

I. INTRODUCTION

Mobile Call Data Record (MCDR) is rich and mostly possesses its own complexity that needs massive analysis and certain level of intelligence in computation. The effect of genetic factors to databases is considered never ending story and always need further research. Among the most waited results would be the cause and effect of genetic factors to the users call data. Normally the study is conducted in different factors and currently, it is gaining a lot of attention by many researchers. Due to the importance of accessing and efficient method to process the users call data through experiment should be dome to develop good algorithms in processing the high dimensional data set in data centres. Due to the slow nature of the data collection where the mobile device is able to collect information without any interactions with the user the information could potentially provide us insights of user's day-to-day routines. The goal of this paper is to utilize the data collected from the mobile phone and survey answers gathered from the users to infer individual's levels of socialization, mental focus capacity and physical activity level and utilize effective data visualization to present the data. This paper aims to explore the relationships between different hints and inspection responses from the participants. Therefore, the aim of this paper is to express a method for specifying the period of holding the users data in data centres. The accessing rate and the probability of processing a record in network throughput and the number of requests that can be processed per unit time shall be the metrics for evaluating the performance of methods under consideration.

II. RELATED WORK

Utilizing a large data set that is collected over a long duration of time for learning about participants behaviors of call data to the network to build and it needs more memory for storing the data items and existing system use the formulas to calculate the maximal weight of each transaction. Collecting such data is not only costly but also required enormous amount of effort and time. Moreover, it is also hard to retain a large number of research participants over a long period of time when the data collection requires researchers to seek information from the participants actively. As the result, the throughput of data samples in research that utilize this kind of practice is often very low. However, given the recent advancement of the mobile phone technology, mobile phones are always continuously collecting information from the users and this requires no active interaction with the users for their input of the information. Furthermore, we are able to leverage on mobile phone's iniquitousness to obtain data that has very high sampling rate.

OBJECTIVES

The Main objectives of this paper is :

- To formulate a method specifying the period of holding the users data in high network traffic data centres.
- To improve the probability of processing a record, network throughput and the number of requests that can be processed per unit time.
- To improve the poor performance in detecting attacks in high network traffic.
- It is required to maximize the throughput and minimize the processing time in the data centres for mobile communications.
- The proposed system will exists its robustness and scalability in high network data centres.

METHODOLOGY

The Methodology carried out in my work comprises of the following steps:

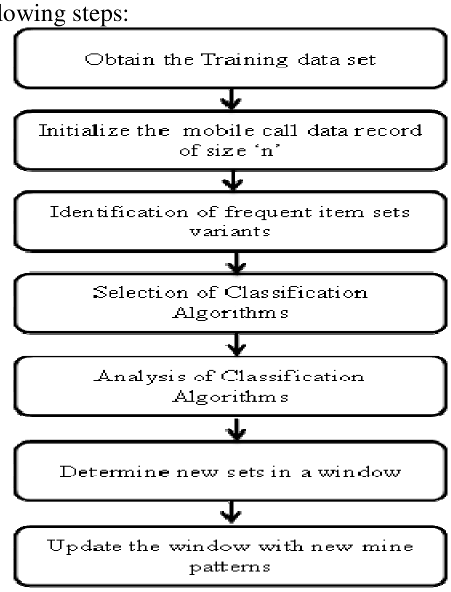


Fig 1.1 Process Analysis of CRD Dataset

III. MOBILE CALL DATA RECORD SYSTEM

The primary aim of mobile phone data collection system in network analysis is to observe group interactions, but not every phone call is made with the same group purpose. A Mobile Call Data Record (MCDR) is a collection of information for each call that is processed by the mobile service provider. In the *table 1.1* the mobile calls might be for business drives, some might be accidental calls, some nodes may be call centers that call a large number of people, and all such interactions are present in MCDRs. In short, Mobile CDRs are noisy datasets. It is sometimes useful to represent a mobile call network by an undirected network, arguing that communication during a single phone call goes both ways, and set the weight of the link as the sum of the weights from both directions.

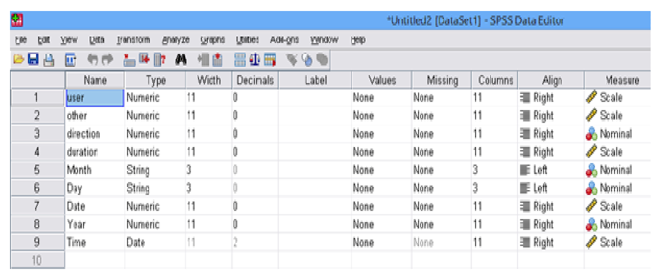
The Mobile Call Data Record is produced form the usage pattern defined in the profile of the mobile user. The mobile CDRs dataset is as follows:

Variables	Description
cust_from	The unique mobile number of the customer to whom should call
cust_to_other	The unique mobile number of the customer to whom should receive
direction	to identify the type of the call
duration	duration of calls in seconds
call_date	The date in the format dd/mm/yyyy from the beginning of the call.
call_time	The date in the format hh::mm::ss from the beginning of the call

Table 1.1 Mobile Call Data Record Variables

IV. SURVEY DATA

In this experiment, the mobile call data was analyzed on a regular base on their mobile phone. The surveys would be sent out to databases at random time. The survey response rates can vary for each mode and are affected by the accepts of the survey design *(cust_from, cust_to,direction,duration,date and time)*. In the recent years most of the surveys have been faced with the declining response rates. The survey for mobile data call record for table 1.1 bias with SPSS is as



	Name	Type	Width	Decimals	Label	Values	Missing	Columns	Align	Measure
1	user	Numeric	11	0		None	None	11	Right	Scale
2	other	Numeric	11	0		None	None	11	Right	Scale
3	direction	Numeric	11	0		None	None	11	Right	Nominal
4	duratior	Numeric	11	0		None	None	11	Right	Scale
5	Month	String	3	0		None	None	3	Left	Nominal
6	Day	String	3	0		None	None	3	Left	Nominal
7	Date	Numeric	11	0		None	None	11	Right	Scale
8	Year	Numeric	11	0		None	None	11	Right	Nominal
9	Time	Date	11	2		None	None	11	Right	Scale
10										

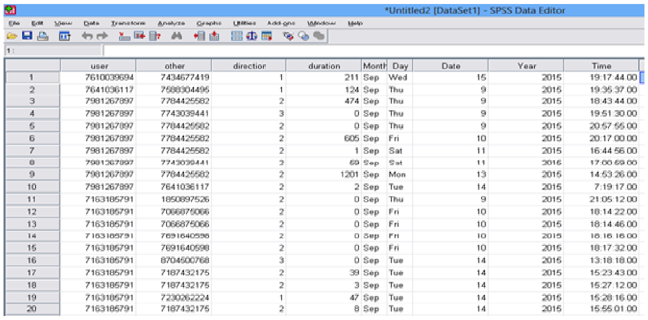
Fig 1.2 Mobile CRDs dataset

Implementation of Mobile Call Data Analysis in SPSS

Data View

The Data View is a Spreadsheet which holds rows and columns, The Data Can be entered in the Data View Sheet Either Manually or the data can be imported from the data file, SPSS can read the data file in any of the one format, which can be Excel, Plain text files or relational (SQL) Databases, Before importing the excel sheet change the excel file format to (.xls).

In SPSS (The Statistical Package for the Social Sciences) software has been developed by IBM and it is widely used to analyses data and makes forecasts based on specific gatherings of data. The inferences of the results are fairly obvious and are statistically valid. Using the software, one can conduct a series of revisions quickly and effectively. The Mobile CDRs are analyzed in frequency table. The frequency analysis is a descriptive statistical method that shows the number of incidences of each response chosen by the respondents. When using frequency analysis, SPSS Statistics can also calculate the mean, median and mode to help users analyze the results and draw conclusions. The analysis of mobile call record data for 3000 records is evaluated. The part of the data we collected is shown below:



	user	other	direction	duration	Month	Day	Date	Year	Time
1	7610039694	7434677419	1	211	Sep	Wed	15	2015	19:17:44.00
2	7641036117	7588304495	1	124	Sep	Thu	9	2015	19:35:37.00
3	7981267897	7784425582	2	474	Sep	Thu	9	2015	18:43:44.00
4	7981267897	7743039441	3	0	Sep	Thu	9	2015	19:51:30.00
5	7981267897	7784425582	2	0	Sep	Thu	9	2015	20:57:55.00
6	7981267897	7784425582	2	605	Sep	Fri	10	2015	20:17:00.00
7	7981267897	7784425582	2	1	Sep	Sat	11	2015	16:44:56.00
8	7981267897	7743039441	2	69	Sep	Sat	11	2015	17:00:59.00
9	7981267897	7784425582	2	1201	Sep	Mon	13	2015	14:53:26.00
10	7981267897	7641036117	2	2	Sep	Tue	14	2015	7:19:17.00
11	7163185791	1850897526	2	0	Sep	Thu	9	2015	21:05:12.00
12	7163185791	7066875066	2	0	Sep	Fri	10	2015	18:14:22.00
13	7163185791	7066875066	2	0	Sep	Fri	10	2015	18:14:46.00
14	7163185791	7691640598	2	0	Sep	Fri	10	2015	18:16:16.00
15	7163185791	7691640598	2	0	Sep	Fri	10	2015	18:17:32.00
16	7163185791	8704600768	3	0	Sep	Tue	14	2015	13:18:18.00
17	7163185791	7187432175	2	39	Sep	Tue	14	2015	15:23:43.00
18	7163185791	7187432175	2	3	Sep	Tue	14	2015	15:27:12.00
19	7163185791	7230262224	1	47	Sep	Tue	14	2015	15:28:16.00
20	7163185791	7187432175	2	8	Sep	Tue	14	2015	15:55:01.00

Fig 1.3 Mobile CRDs dataset in data view

(a) Frequencies

The Frequencies procedure can produce summary measures for categorical variables in the form of frequency tables, bar charts, or pie charts.

direction

		Frequency	Percent	Valid percent	Cumulative Percent
Valid	1	3165	24.3	24.3	24.3
	2	8040	61.7	61.7	86.0
	3	1824	14.0	14.0	100.0
	Total	13029	100.0	100.0	

Table 1.2 Frequency table w.r.t. direction

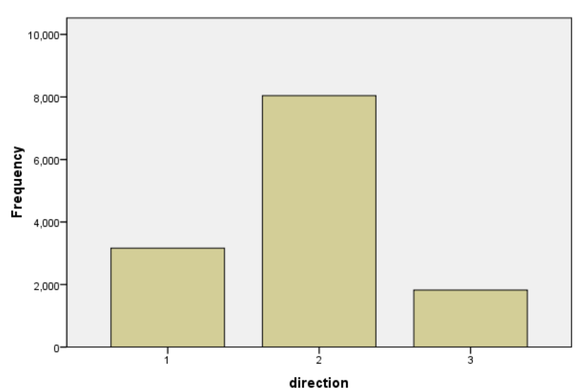


Fig 1.4 Frequencies for No. of calls and direction

(b) Crosstabs

Call observing and record applications used by mobile communications generates extremely large amount of call detail records (CDRs) in real-time, and companies constantly need to control from this data to boost throughput. The volume of the calls and data taken by the call observing applications are so large that it impossible to manually analyze and conclude the behavior of the network. By using the crosstab the aggregate values are used to extract the relationship between different variables of catergorical data. The *table 1.3* shows the proportion of cases in subgroups for direction of mobile calls data record in the data centres. Crosstabs are widely used in quantitative analysis in market research and surveys. The crosstabs are used to identify the new trends. Further, they help to identify the type of response of a particular category of users. This often helps us in understanding the effects and action that will have on a category of our target segment. By using the crosstab construction and analysis of mobile CDR data is usually dome through statistical packages in SPSS. While SPSS provides the flexibility of choosing whichever variable you want on which axis, it allows only different categories in the column variable without the problem of wrapping the data. However there is no limit on the number of categories for the dependent variable. In SPSS allows calculating a display the percentages for all the variables.

Input	Active Dataset	Mobile CDR
	Filter	<none>
	Weight	<none>
	Split File	<none>
	N of Rows in Working Data File	13029
Missing Value Handling	Definition of Missing	User-defined missing values are treated as missing.
	Cases Used	Statistics for each table are based on all the cases with valid data in the specified range(s) for all variables in each table.
Syntax		CROSSTABS /TABLES=Date BY direction /FORMAT=AVALUE TABLES /CELLS=COUNT COLUMN /COUNT ROUND CELL.
Resources	Processor Time	00:00:00.094
	Elapsed Time	00:00:00.028
	Dimensions Requested	2
	Cells Available	174762

Table 1.3 Cross tabulation for date and direction

Analysis: By performing Crosstabs through SPSS, derived missing value handling statistics for each value are based on all the cases with valid data in the specified range(s) for all the variables in the table. So we conclude that Mobile CDRs data satisfaction is dependent of the users call data per day.

In the above table 1.3 shows the Crosstabs that contains a cell for every arrangement of categories in the two variables.

- ✓ Inside each cell is the number of cases that fit that particular arrangement of responses.
- ✓ SPSS can also report the row, column, and total percentages for each cell of the table.

Because Crosstabs creates a row for each value in one variable and a column for each value in the other, the procedure is not suitable for *continuous* variables that assume many values. Crosstabs is designed for discrete variables--usually those measured on nominal or ordinal scales. To evaluate the robustness of the classifier, the normal methodology is to perform cross tabulation on the classifier. For the present study dataset is divided into the trained set. The syntax of cross tabulation for the dataset is given below:

CROSSTABS/TABLES=Date BY direction,
FORMAT=AVALUE TABLES
CELLS=COUNT COLUMN
COUNT ROUND CELL.

Date * direction Crosstabulation

			Direction			Total
			1	2	3	
Date	1	Count	99	344	63	506
		% within direction	3.1%	4.3%	3.5%	3.9%
	2	Count	78	193	69	340
		% within direction	2.5%	2.4%	3.8%	2.6%
	3	Count	87	230	44	361
		% within direction	2.7%	2.9%	2.4%	2.8%
	4	Count	76	200	43	319
		% within direction	2.4%	2.5%	2.4%	2.4%
	5	Count	78	207	48	333
		% within direction	2.5%	2.6%	2.6%	2.6%
	6	Count	102	238	51	391
		% within direction	3.2%	3.0%	2.8%	3.0%
	7	Count	44	174	35	253
		% within direction	1.4%	2.2%	1.9%	1.9%
	8	Count	88	164	32	284
		% within direction	2.8%	2.0%	1.8%	2.2%
	9	Count	88	280	69	437
		% within direction	2.8%	3.5%	3.8%	3.4%
	10	Count	126	304	63	493
		% within direction	4.0%	3.8%	3.5%	3.8%
	11	Count	129	396	97	622
		% within direction	4.1%	4.9%	5.3%	4.8%
	12	Count	92	294	51	437
		% within direction	2.9%	3.7%	2.8%	3.4%
	13	Count	114	254	72	440
		% within direction	3.6%	3.2%	3.9%	3.4%
	14	Count	141	340	38	519
		% within direction	4.5%	4.2%	2.1%	4.0%
	15	Count	105	244	41	390
		% within direction	3.3%	3.0%	2.2%	3.0%
	16	Count	98	221	58	377
		% within direction	3.1%	2.7%	3.2%	2.9%
	17	Count	141	340	84	565
		% within direction	4.5%	4.2%	4.6%	4.3%
	18	Count	137	337	114	588
		% within direction	4.3%	4.2%	6.2%	4.5%
	19	Count	92	229	62	383
		% within direction	2.9%	2.8%	3.4%	2.9%
	20	Count	113	246	56	415
		% within direction	3.6%	3.1%	3.1%	3.2%
	21	Count	93	197	52	342
		% within direction	2.9%	2.5%	2.9%	2.6%
	22	Count	98	185	70	353
		% within direction	3.1%	2.3%	3.8%	2.7%
	23	Count	126	239	64	429
		% within direction	4.0%	3.0%	3.5%	3.3%
	24	Count	144	314	53	511
		% within direction	4.5%	3.9%	2.9%	3.9%
	25	Count	94	263	72	429
		% within direction	3.0%	3.3%	3.9%	3.3%
	26	Count	109	331	74	514
		% within direction	3.4%	4.1%	4.1%	3.9%
	27	Count	85	276	56	417
		% within direction	2.7%	3.4%	3.1%	3.2%
	28	Count	79	211	50	340
		% within direction	2.5%	2.6%	2.7%	2.6%
	29	Count	156	322	60	538
		% within direction	4.9%	4.0%	3.3%	4.1%
	30	Count	92	263	47	402
		% within direction	2.9%	3.3%	2.6%	3.1%
	31	Count	61	204	36	301
		% within direction	1.9%	2.5%	2.0%	2.3%
Total		Count	3165	8040	1824	13029
		% within direction	100.0%	100.0%	100.0%	100.0%

Table 1.4 Cross tabulation for a month with the count% within the direction per day-wise results

V. CONCLUSIONS

This investigation is used to understand the components which are affecting the subscriber intention for switching service provider in India. The continuous work and research has done by facility providers and researchers for providing precise and efficient information. The need is to handle the received data carefully so that the users could get relevant and exact information. There is an almost no exploration accessible to research for the components affecting versatile client's expectation for exchanging between specialist organizations in India. The level of accomplishment of goals is referred to as viability and the same should be proficient using insignificant assets. The discoveries of this investigation add to a larger comprehension of the connection between course of the calls and date of rate with call information

REFERENCES

1) Zahra Jadidi, V.Muthukukumaraswamy, Mansour Sheikhan, "Flow-Based Anomaly Detection Using Neural Network Optimized with GSA Algorithm", 2013 IEEE 33rd International Conference on Distributed Computing Systems Workshops,pp.76-81.
2) Sofianita Mutalib, Shuzlina Abdul-Rahaman, Azlinah Mohamed "Towards Mining Frequent Patterns in Genome Wide Association", 2013 IEEE 16th International Conference on Computational Sciences and Engineering.
3) AJM Abu Afza, Mohammed Uddin "Intrusion Detection Learning Algorithm through Network Miniing", 2014 IEEE 16th International Conference on Computer and Information Technology.
4) Brijesh Sharma and Huma Gupta, "A Design and Implementation of Intrusion Detection System by using Data Mining", 2014 IEEE 4th International Conference on Communication Systems and Network Technologies.
5) Rahul Anil Ghatage , "Frequent Pattern Mining Over Data Stream Using Compact Sliding Window Tree & Sliding Window Model", 2015 International Research Journal of Engineering and Technology(IRJET)Vol,2 Issue:04, pp.217-223.
6) Karakoulas G. and Shawe-Taylor J. (1998). Optimizing classifiers for imbalanced training sets. In: *Proceedings of Neural Information Processing Workshop*, NIPS'98, pp. 253-259.
7) Flouvat, F. 2005. "Experimental study of frequent itemsets datasets".
8) Roma Mitra Debnath, Ravi Shankar,"Benchmarking telecommunication service in India: An application of data envelopment analysis", Benchmarking: An International Journal, Vol. 15 Iss: 5, pp.584 – 598, 2008
9) S.K.Tanbeer, C. F. Ahmed, B.-S. Jeong, & Y.-K. Lee. "Sliding window-based frequent pattern mining over data streams", *Information Sciences*, vol. 179(22), pp. 3843-3865, 2009.
10) J.H. Chang, W.S. Lee, "estMax: Tracing Maximal Frequent Itemsets Instantly over Online Transactional Data Streams", *IEEE Transactions on Knowledge and Data Engineering*, vol.21 (10) pp.1418-1431, 2009.
11) K.-F. Jea & C.-W. Li, "Discovering frequent itemsets over transactional data streams through an efficient and stable approximate approach, Expert Systems with Applications", 36(10), 2009, pp. 12323–12331.
12) Mahnoosh Kholghi, Mohammadreza Keyvanpour, "An Analytical Framework for Data Stream Mining Techniques Based on Challenges and Requirements", International Journal of Engineering Science and Technology (IJEST) ISSN: 0975-5462 Vol. 3 No. 3 Mar 2011.
13) Nori, F., Deypir, M., Sadreddini, M. H. and. Hashemi, M. H. "A sliding window based algorithm for frequent closed itemset mining over data streams", The Journal of Systems and Software, 2013, 86 615– 623.
14) http://www.oracle.com/technology/products/bi/odm/index.html
15) http://www.sqlserverdatamining.com/ssdm