An Efficient Protocol for Predicting User Behaviour using Call Detail Data

B.Praveena[1] R.Indra[2] Patel Sumesh[3]
[2]Assistant Professor
[1,2]Department of Computer Applications
[1,2]Shrimati Indira Gandhi College, Tiruchirapalli [3]Value Tech

Abstract— Wireless Mobile devices has led to the explosion of the computing world. This huge information can be utilized by observing the mobile network's behavior. The big data thus generated in the mobile networks enable us to gain useful insights into the user patterns by using big data analysis techniques and computing techniques. The proposed model introduces uses data set information, data analysis techniques to categorize the data set in the network communication into two types, one is user based and the other one is network oriented. The essential data is then computed into by the telecommunication operators who are facing the tremendous challenge to provide satisfactory service to mobile users with varying QoS requirement. Thus by including high volume media transmission, huge amount of machine to machine(M2M) connectivity the data is summarized and reviewed to form temporal and spatial analysis to data mining and statistical test. The call measurement and call detail record, respectively, to understand the base station behavior. The users behavior is revealed and predicted by comparing the base stations location and real-world map and determine the behavior of the users.

Key words: Mobile Networks, Wireless Networks, Traffic Analysis, Spatial Correlation

I. INTRODUCTION

The mobile communication networks have revolutionized the way work is done and some of the major issues concerning networks include network-oriented issues, user-oriented issues, and data-oriented issues. The huge quantum of data generated in mobile devices has been found to be indispensable and has both social and economic values. Therefore to support the increasing demand, it's necessary to investigate the possibilities that huge quantum of mobile data offers us. This leads one to spatial and temporal dynamics in mobile networks which contribute to the understanding of mobile user behaviors in such mobile networks. Thus with the help of modern information technology, collection and analysis of large-scale mobile network communication traffic information is possible in todays world. This brings forth many empirical studies and significant outcomes. The focus of this study is to study one such efficient model protocol by collecting data and predicting user behaviour. Such a framework should be anonymous while processing the query in networks. The proposed model is safe and secure methods are provided apart from providing Location Based Services. Also computationally the overheads are low and the efficiency is maximum.

II. RELATED WORK

A.Palaios et al [1] have analyzed that communication networks produce huge amount – quantum of data when serving the mobile users. Some like spectrum measurements [2], device status report and etc, but this topic is beyond this paper's scope. U. Paul et al [2] in "Understanding traffic dynamics in cellular data networks," have analyzed the network traffic from the base stations point of view and have found significant temporal and spatial variations which provide vital insights into parameters like pricing, protocol, design, resource and spectrum used. Opechowski, Z et al [3] investigated call blocking and handover failure probabilities of the Telecel GSM network and made an analysis which compared measured data from base station "Parede" in the south of Lisbon. Aggarwal et al [4] contend that data management solutions for information privacy must restore controls in the individual's hands. Y.-A. de Montjoye et al [5] presented important design frameworks dedicated to protect the individual's privacy. M. Michalopoulou, et al [6] proposed optimal pricing mechanism, novel network design and unique protocol design for spectrum allocation and energy saving plan for mobile phone users with good services. X. ZUO et al [7] suggested detailed studies on aggregated voice traffic which are neither plentiful nor relevant and included hot spots detection based on Erlang data and traffic analysis. E. Nan et al [8] investigated Mobile telephone traffic data and suggested that it can be used to characterize resource usage where subscriber behaviours in mobile networks, and some basic knowledge of spatial and temporal dynamics of data traffic has been captured.

III. PROPOSED

The spatial locations of base stations in different clusters are plotted here. Normally Base stations in a particular cluster are mainly located in three places, one in the neighboring campuses, second by comparing location of the base stations in map. So the behavior of cluster it is believed may represent traffic distribution of students in university campuses.

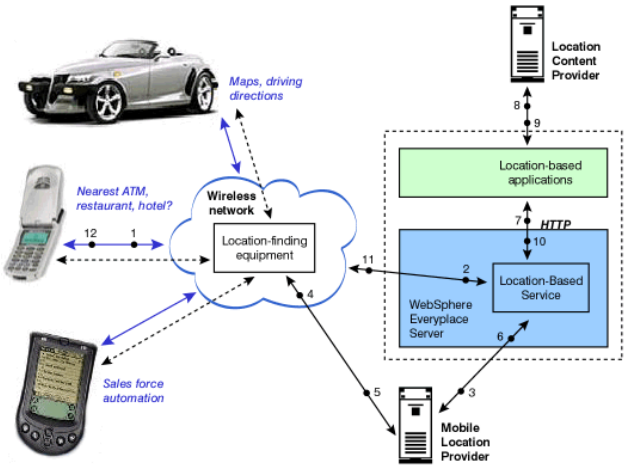

Fig. 1: Spatial Locations of base stations in different clusters

There are mainly 4 parts that base stations in cluster 2 locate at, but among them no direct relationship has been found.

One can only infer that these parts to be official areas because the traffic is high only in working hours, and further verification will be further work. The problem of auto spatial correlation is successfully addressed and the regions location of the base stations having positive correlation and the places in the positive correlation concentrated region, but still negative correlation to the base stations. Comparatively the volume of voice traffic in base stations is very small, while the traffic volume is very high in other base stations. It is imperative that there exists a positive correlation among many related or closely located base stations leading to a larger global spatial correlation.

IV. EVALUATION

The availability of correlations among base stations with varying traffic volumes leads us to the fact that neighbouring nodes can be identified on the basis of spatial correlation. The focus here in case of cellular optimization is to balance the loads of the nearby or proximity base stations. If not balanced in a network one base station will have abnormally high load while its neighbour nodes have only small traffic. This directly leads to computational overheads, route inefficiency and bandwidth resource wasting. The proposed protocol model reduces overheads, increase accuracy and efficiency by spatial positioning and the results and findings are discussed below.

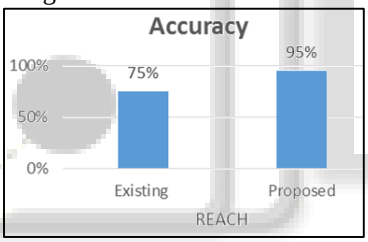

Fig. 1: Chart Showing Accuracy

First by verifying the long-range time-correlation of the call arrivals at the base station, it is found that call arrivals in a minute are uncorrelated to the number of call arrivals in another minute in short-term.6

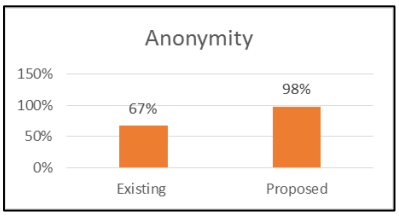

Fig. 2: Chart Showing Security

Second, it is found that correlation of time to call arrivals is governed by the arrival time and the present location of base stations.

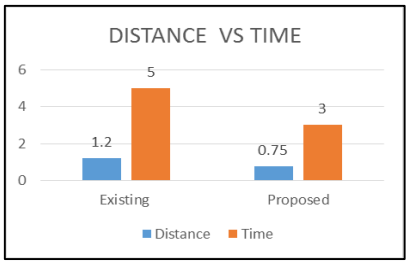

Fig. 3: Chart Showing Distance vs Time

The ratios of call holding time to call inter-arrival time were found to be modeled as a function between distance and anonymity.

V. CONCLUSION

The proposed network location based query uplink service offers MAP aware WLAN network-based anonymization and processing framework, the system for query processing in road networks. The system also relies on a mobile user network that satisfies reciprocity and guarantees mobile users. This is to identify the queries characteristics that affect subsequent processing, and qualitatively provide answers to the node. Then the proposed query uplink provides techniques that exploit these characteristics to give results which are consistent with the user needs. First user privacy is preserved and the query is not made to the other nodes in the group, second it achieves low overheads in computational and communication costs, and third most importantly responses are very quick. Such easy and simple deployable models, requiring only basic network operations are the future. The project can be further enhanced to include mobile web services and also include the services to the full extreme. If implemented as a mobile web service the providers can enhance the limits to the maximum output in location based services along with encryption services.

REFERENCES

[1] A. Palaios, J. Riihijarvi, O. Holland, A. Achtzehn, and P. Mahonen,“ Measurements of spectrum use in london: exploratory data analysis and study of temporal, spatial and frequency domain dynamics,” in Dynamic Spectrum Access Networks (DYSPAN), 2012 IEEE International Symposium on. IEEE, 2012, pp. 154–165

[2] Z. Opechowski and L. M. Correia, “Analysis of traffic distributions in gsm,” in Microwaves, Radar and Wireless Communications. 2000. MIKON-2000. 13th International Conference on, vol. 2. IEEE, 2000, pp. 390–394

[3] U. Paul, A. P. Subramanian, M. M. Buddhikot, and S. R. Das, “Understanding traffic dynamics in cellular data networks,” in INFOCOM, 2011 Proceedings IEEE. IEEE, 2011, pp. 882–890 267–270, 2012.

[4] A. Palaios, J. Riihijarvi, O. Holland, and P. Mahonen, “A week in London: Spectrum usage in metropolitan London,” in Personal Indoor and Mobile Radio Communications (PIMRC), 2013 IEEE 24th International Symposium on. IEEE, 2013, pp. 2522–2527.

[5] Y.-A. de Montjoye, C. A. Hidalgo, M. Verleysen, and V. D. Blondel, “Unique in the crowd: The privacy bounds of human mobility,” Scientific reports, vol. 3, 2013

[6] X. ZUO and Y. ZHANG, “Detection and analysis of urban area hotspots based on cell phone traffic,” Journal of Computers, vol. 7, no. 7, pp. 1753–1760, 2012

[7] H. Jiawei and M. Kamber, “Data mining: concepts and techniques,” San Francisco, CA, itd: Morgan Kaufmann, vol. 5, 2001.

[8] M. L. Goldstein, S. A. Morris, and G. G. Yen, “Problems with fitting to the power-law distribution,” The European Physical Journal B-Condensed Matter

and Complex Systems, vol. 41, no. 2, pp. 255–258, 2004.

[9] R. Tibshirani, G. Walther, and T. Hastie, "Estimating the number of clusters in a data set via the gap statistic," Journal of the Royal Statistical Society: Series B (Statistical Methodology), vol. 63, no. 2, pp. 411–423, 2001.

[10] D. Yin, S. Zhang, W. Zhou, and Y. Zheng, "Time-correlation analysis of gsm telephone traffic in dense population district," Sixth International Conference on Wireless Communications and Signal Processing, Hefei, China, Oct. 23-25, 2014.

[11] A. Pattavina and A. Parini, "Modelling voice call inter-arrival and holding time distributions in mobile networks," Performan. Challenges Efficient Next Generat. Networks, Proc. 19th Int. Teletraf. Congr, pp. 729–738, 2005.

