OPEN ACCESS INTERNATIONAL JOURNAL OF SCIENCE & ENGINEERING

A CASE STUDY ON BIG DATA ANALYTICS IN MOBILE CELLULAR NETWORKS

N. Siva Priya[1], Dr.N.Shenbagavadivu[2]
Research Scholar, Department of Computer Applications[1]
Assistant Professor, Department of Computer Applications[2]
University College of Engineering BIT Campus, Anna University, Trichy, India
nsiva_priya@yahoo.co.in[1],kshenthu@gmail.com[2]

Abstract: As the population of mobile users is increasing day by day, the data generated by the mobile cellular networks increases drastically. These data seems to be high in terms of velocity, variety and value. For efficient use of mobile cellular networks these data need to be analyzed by an effective technique. In this paper we review the various methods of analyzing data generated by mobile cellular networks. We aim to introduce the general background of data generated by mobile cellular networks and review certain technologies related to this. In this study we are about to discuss various analysis methods and the case studies involving mobile big data

Keywords: *Big Data Analytics, Mobile cellular networks, Hadoop*

I INTRODUCTION

Recent decades have witnessed tremendous increase in data in terms of size, speed, variety, value and veracity (called 5 V's of big data). The term of big data is mainly used to describe this enormous datasets. The big data is comprised of masses of unstructured data which also requires real time analysis. If these datasets are effectively organized and managed many useful and in depth knowledge can be obtained which leads to finding solutions for various unsolved issues.

1. Mobile big data

Android Apps has provided more than 650,000 applications, covering nearly all categories. Such huge mass of data and numerous applications call for mobile analysis, but also bring about a few challenges. The unique characteristics of mobile data are found to be Mobile sensing, moving flexibility, noise, and a large amount of redundancy. Mobile phones are now useful for building and maintaining communities, and these communities with geographical locations and communities based on different cultural backgrounds and interests with their growing number of users and improved performance. Mobile phones can support rich interaction at anytime and anywhere while the traditional network communities or SNS communities are in short of online interaction among members, and the communities are active only when members are sitting before computers. Mobile communities are defined as that a group of individuals with the same hobbies (i.e., health, safety, and entertainment, etc.) gather together on networks, meet to make a common goal, decide measures through consultation to achieve the goal, and start to implement their plan. The recent study in wireless sensor networks and mobile phones has lead to various mobile applications like real time health tracking etc., medical data from sensors seem to be of different characteristics in terms of attributes, time and space relations, as well as physiological relations, etc.

In addition, such datasets involve privacy and safety protection. In Garg et al. introduce a multi-modal transport analysis mechanism of raw data for real-time monitoring of health. Under the circumstance that only highly comprehensive characteristics related to health are available, Park et al .in examined approaches to better utilize. Application has been developed by researchers of Norway university for smart phones, which analyzes paces when people walk and uses the pace information for unlocking the safety system.

Apart from online health tracking applications many other applications can be developed from the analysis of big data generated by mobile cellular networks. Much useful information can be obtained by geolocating the mobile phones, recording phone calls and so on which will be helpful for the operators to provide better performance to their customers. Many new customers' friendly applications can be developed by analyzing this enormous datasets. From these applications the mobile operators would be able to provide better services to their customers thereby improving their revenue. Apart from this many real time applications could also be developed which will be helpful to the mobile users.

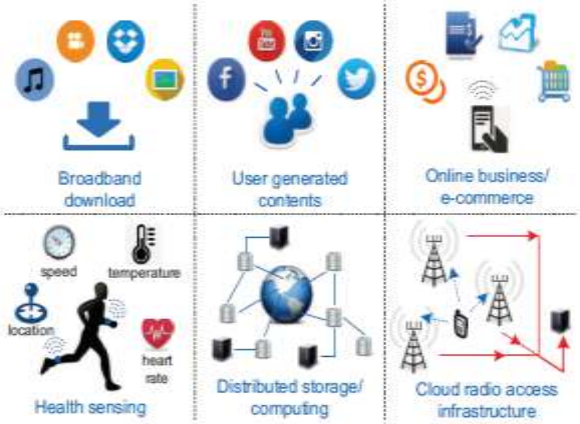

Figure 1. Some Example Sources Of Wireless Big Data Traffic

2. Need for analysis

Wireless cellular networks have witnessed tremendous advances in recent decades. Due to ever increasing mobile applications, mobile cellular networks have become both generators and carriers of massive data. These data are generated while geo-locating mobile devices, recording phone calls, and capturing mobile applications' activities. These enormous data should be paid much attention, for efficient use of the mobile cellular networks and to increase the revenue of the mobile cellular operators. When compared to traditional data analytics Big data analytics would be an efficient method in analyzing such enormous unstructured data. Traditional data analytics prove to be inadequate while encountering data involved in mobile cellular network. Big data analytics deals with both structured and unstructured data while only structured data is dealt in traditional data analytics. In making real-time decisions traditional data analytics proves to be inadequate. Traditional data analytics fails in such cases such as to improve the performance of mobile cellular networks and to increase the revenue of its operators. While these insightful data is being analysed by big data analytics while traditional data analytics fail to do so. In mobile cellular networks the complete data of the customers is scattered in various business department big data analytics is capable of collecting these data and extract useful information from these data while the traditional data analysis concentrates only on specific department.. The big data analytics helps the mobile operators in making dynamic and autonomous decision rather than traditional data analytics.

II BIG DATA ANALYTICAL TOOLS FOR WIRELESS NETWORK TRAFFIC DATA:

This study discusses some commonly used algorithms for analyzing wireless network traffic data and that are analysed and exploited by specially designed learning units (LUs) installed at both the BSs and CUs.

1) Stochastic Modeling: Stochastic modeling methods use probabilistic models to capture the explicit features and dynamics of the data traffic [2]. Some commonly used stochastic models may include: order-K Markov model, hidden Markov model, geometric model, time series, linear/nonlinear random dynamic systems, etc. Markov models and Kalman filters are widely used to predict user mobility and service requirements. The collected user data are often used for parameter estimation of stochastic models, such as estimating the transition probability matrix of a Markov chain.

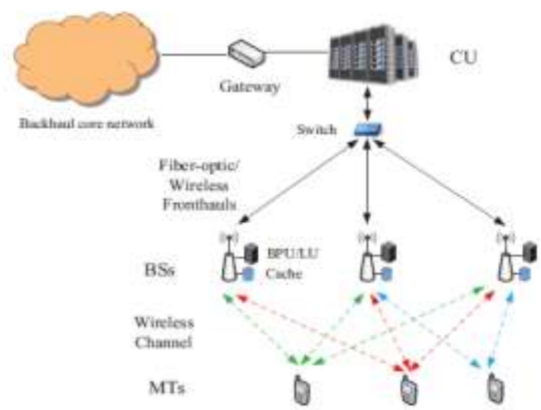

Figure 2. A Hybrid Cu-Bs Processing Network Structure.

2) Data mining: Data mining focuses on exploiting the implicit structures in the mobile data sets. Also taking the mobility prediction problem as an example, individual user's mobility pattern could be extracted and discovered by finding the most frequent trajectory segments in the mobility log. Prediction could be made accordingly by matching the current trajectory to the mobility profile. Clustering is another useful technique to identify the different patterns in the data sets. It is widely used in context-aware mobile computing, where a mobile user's context and behavioral information, such as sleeping and working, are identified from wireless sensing data for providing context-related services.

3) Machine learning: The main objective of machine learning is to establish functional relationship between input data and output actions, thus achieving auto-processing capability for unseen patterns of data inputs. Among the many useful techniques in machine learning applied to wireless communications, classification (determining the type of input data) and regression analysis (data fitting) are two common methods, whose applications include context identification of mobile usage and prediction of traffic levels (classification), or fitting the distributions of trajectory length, mobile user location, and channel holding times (regression). Besides, reinforcement learning, such as Q-learning , is useful for taking proper real-time actions to maximize certain long-term rewards. A typical example is making the handoff and admission control decision (action), given the current traffic load (state) and incoming new requests (event), in which the reward could be evaluated against the reduction of dropped calls or failed connections.

4) Large-Scale Data Analytics: Wireless big data poses many challenges to the aforementioned conventional data-analytical methods due to its high volume, large dimensionality, uneven data qualities, and the complex features therein. To improve signal processing efficiency, one can combine the following complexity reduction techniques with the conventional data analytical tools for large-scale data processing.

- Distributed optimization algorithms, such as primal/dual decomposition and alternating direction method of multipliers (ADMM), are very useful to decouple large-scale statistical learning problems into small sub problems for parallel computations so as to relieve both the computational burden at the CU and the bandwidth pressures to the front haul/backhaul links.
- Dimension reduction methods are useful to reduce the data volume to be processed while capturing the key features of big data. Among various methods, principle component analysis (PCA), along its many variants, is the mostly used method today. In addition, tensor decomposition methods are also popular in mobile data processing, which seek to approximately represent a high-order multi way array (tensor) as a linear combination of outer products of low-order tensors. By doing so, the hardware requirement and cost for storing the high-order arrays of mobile data could be reduced.
- Other advanced learning methods could be used to handle incomplete or complex data sets. Interesting examples include active learning, which deals with partially labeled data set; online learning for responding in real-time to sequentially received data; stochastic learning that makes a decision periodically in each time interval; and deep learning for modeling complex behaviors contained in a data set.

III CASE STUDIES OF BIG DATA ANALYTICS IN MOBILE CELLULAR NETWORKS

The following case studies are discussed as an illustration of introducing big data analytics into mobile cellular networks. Here the study focuses on improving network performance and deriving valuable insights. The application areas are studied to cover different scenarios, from current deployed mobile cellular networks to upcoming 5G, from network operational optimization.

a. Big traffic data

Due to the widespread use of mobile internet, the volume of traffic data increases at an unprecedented rate. The internet acts as a carrier of the traffic data. The cellular operators are responsible to manage the network resource appropriately to balance network load and optimize network utilization. Traffic monitoring and analyzing is an elementary but essential part for network management, enabling performance analysis and prediction, failure detection, security management, etc. In the context of big traffic data traditional approaches to monitor and analyze the traffic data seem, however, straightforward and inadequate. In, the interrelationship between big data and software-defined networking (SDN) has been studied. Liu et al. [3] proposed a novel large-scale network traffic monitoring and analysis system based on a Hadoop platform. The system is practically deployed in a commercial cellular network with 4.2 Tbytes input volume every day. The evaluation results indicate that the proposed system is capable of processing big network-generated data and revealing certain traffic and user behavior phenomenon. Since understanding the traffic dynamics and usage condition is of significance for improving network performance, the topic of traffic characteristics becomes a hot focus. In [5], the authors investigated three features of network traffic, namely network access time, traffic volume, and diurnal patterns from the perspective of device models. The traffic characteristics from the perspective of service providers were revealed in [10]. Another angle from operating systems was introduced in [11]. All the above results are beneficial for cellular network operators to make corresponding adjustments for network capacity management and revenue growth.

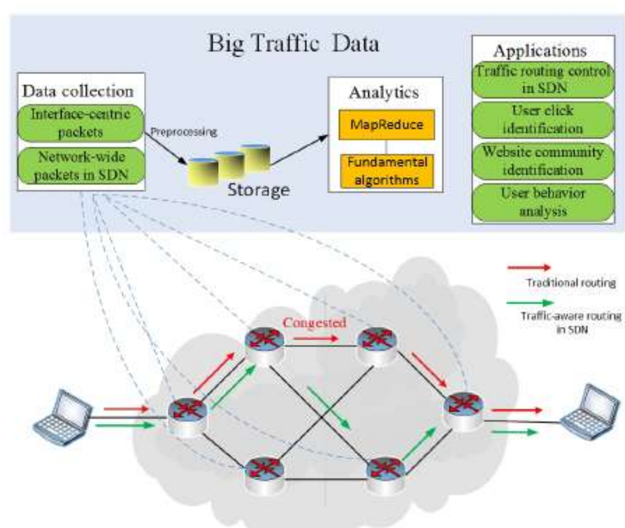

Figure 3. Big Traffic Data

b. Big location data

Location data analysis is informative since human activities are based on locations. The location based big data arises from GPS sensors, Wi-Fi, Bluetooth through mobile devices, have become precious strategic resources. These location data would provide support for government administration, such as public facility planning, transportation system constructions, demographic trends, risk warnings for crowed people, rapid emergence responses, crime hot spots analysis, etc. It can also gain amazing business insights, such as mobile advertising and marketing. In this an end-to-end Hadoop-based system was developed with a number of functional algorithms operated on call record details (CRDs). With the information about subscribers' habits and interests, it is capable of providing invaluable information about when, where & how a category of individuals (e.g., sports fans) move.

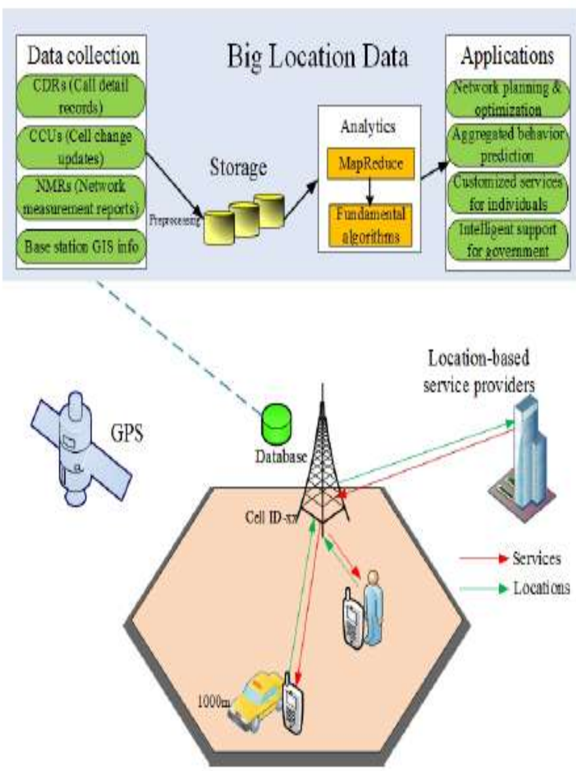

Figure 4. Big Location Data

c. Big radio waveforms data

Zhang and Qiu [8] used large random matrices as building blocks to model the big data arising from a 5G massive MIMO system that is implemented using software-defined radios. They exploited the fact that all data processing is done at CPU so all the modulated waveforms are stored at the RAMS or at the hard drives. On the other hand, big data analytics based on the random-matrix theory is applied to the collected data from their test bed, where a mobile user communicates with the massive MIMO base station while moving.The experimental results can estimate the user's moving speed, whether motionless, at an early constant speed, at as low speed or at a higher speed. This analytics is also implemented to reflect the correlation residing in the transmitted signals. These applications validate the fact that the massive MIMO system is not only a communication system, but also a massive data platform which can brings tremendous values through big data analytics.

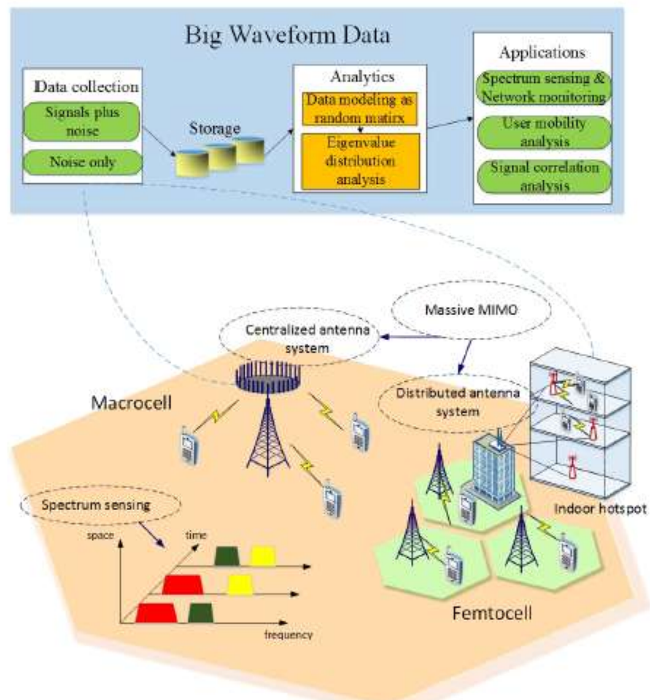

Figure 5. Big Radio Waveform Data

d. Big signaling data

In mobile cellular networks, the transmission of voice and data is accompanied by control messages, which are termed as signaling. The signaling works according to the predefined protocols and ensure the communication's security, reliability, regularity and efficiency. Signaling monitoring plays an important role in appropriate allocation of network resources, improving the quality of network services, real-time identifying network problems, and etc. With the rapid development of various mobile cellular networks, the volume of signaling data grows tremendously and the traditional signaling monitoring systems have too many problems to deal with. This discussion describes a signaling data monitoring and analyzing system architecture with big data analytics. This architecture mainly consists of three components: data collecting, data analyzing and applications. In data collection, various signaling protocols are copied from multiple network interfaces without interrupting normal operations. Afterwards, these copies are gathered and filtered through the protocol processor and then sent to the analyzer. In the analyzer, the data is processed using various algorithms, such as decomposition, correlation analysis, etc. Finally, the analysis results can be used by various applications. For example, Celibi et al. analyzed the BSSAP messages from A interface in a Hadoop platform to identify handovers from 3G to 2G. The simulation results show that the identified 3G coverage holes are consistent with the drive test results.

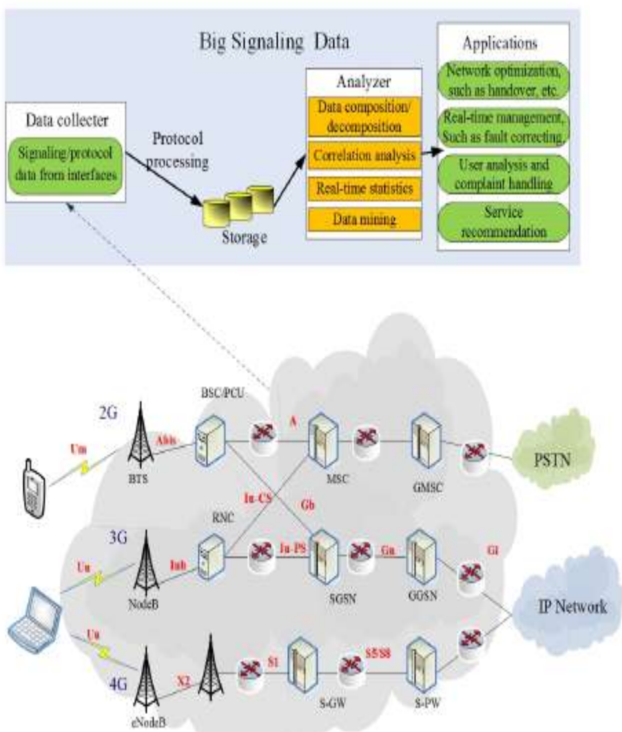

Figure 6. Big Signaling Data

e. Big Heterogeneous Data

One critical task of big data analytics in mobile cellular networks is the integration of very heterogenous data correlation mining in massive database. Data sources are rich in types such as data rate, packet drop, mobility, etc. Different base stations host these data over time. They need be aggregated across space and time to obtain big data analytics. For example, for cyber security, there are many different heterogeneous sources, such as ''numerous distributed packet sniffers, system log files, SNMP traps and queries, user profile databases, system messages, and operator commands.'' Essentially, data fusion is a technique to make overall sense of data from different sources that commonly have different data structures.

IV CONCLUSION

This study has been done to explore various methods of big data analytics in mobile cellular networks. We have studied various case studies involved in the generation of big data in mobile cellular networks like big traffic data, big location data, big radio waveform data, big signalling data and big heterogenous data and also the challenges involved in them. This study will be useful to carry out the work in future.

REFERENCES

[1]C. Liang, F. R. Yu, and X. Zhang, "Information-centric network function virtualization over 5G mobile wireless networks," *IEEE Netw.*, vol. 29,no. 3, pp. 68_74, May/Jun. 2015.

[2] C. Liang and F. R. Yu, "Wireless network virtualization: A survey, some research issues and challenges," *IEEE Commun. Surveys Tuts.*, vol. 17,no. 1, pp. 358_380, Mar. 2015.

[3] J. Liu, F. Liu, and N. Ansari, "Monitoring and analyzing big traffic data of a large-scale cellular network with Hadoop," *IEEE Netw.*, vol. 28, no. 4,pp. 32_39, Jul./Aug. 2014.

[4] S. Bi, R. Zhang, Z. Ding, and S. Cui, "Wireless communications in the era of big data," *IEEE Commun. Mag.*, vol. 53, no. 10, pp. 190_199, Oct. 2015.

[5] J. Liu, N. Chang, S. Zhang, and Z. Lei, "Recognizing and characterizing dynamics of cellular devices in cellular data network through massive data analysis," *Int. J. Commun. Syst.*, vol. 28, no. 12, pp. 1884_1897, Aug. 2015.

[6] X. He, Q. Ai, R. C. Qiu, W. Huang, and L. Piao, "A big data architecture design for smart grids based on random matrix theory," *IEEE Trans. Smart Grid*, to be published. [Online].Available:http://arxiv.org/pdf/1501.07329.pdf

[7]X. He, Q. Ai, R. C. Qiu, W. Huang, and L. Piao, "A big data architecture design for smart grids based on random matrix theory," *IEEE Trans. Smart Grid*, to be published. [Online].Available:http://arxiv.org/pdf/1501.07329.pdf

[8]L. Jun, L. Tingting, C. Gang, Y. Hua, and L. Zhenming, "Mining and modeling the dynamic patterns of service providers in cellular data network based on big data analysis," *China Commun.*, vol. 10, no. 12, pp. 25_36, Dec. 2013.

[9] J. Yang, S. Zhang, X. Zhang, J. Liu, and G. Cheng, "Characterizing smartphone traffic with MapReduce," in *Proc. IEEE WOCC*, Jun. 2013,pp. 1_5.

[10] L. Deng, J. Gao, and C. Vuppalapati, "Building advertising and marketing," in *Proc. IEEE1st Int. Conf. Big Data Comput. Service Appl. Big Data Service)*,Mar. 2015, pp. 256_266.

[11] V. Kolar *et al.*, "People in motion: Spatio-temporal analytics on call detail records," in *Proc. IEEE COMSNETS*, Jan. 2014, pp. 1_4.

[12] R. Zuech, T. Khoshgoftaar, and R. Wald, "Intrusion detection and big heterogeneous data: A survey" *J. Big Data*, vol. 2, no. 3, pp. 1_41,Dec. 2015.

[13] D. M. B. Ying and H. T. Mouftah, ``Sink privacy protection with minimum network traf_c in WSNs," *Ad Hoc SensorWireless Netw.*, vol. 25, nos. 1_2, pp. 69_87, 2015.
[14] M. Chen, S. Mao, and Y. Liu, "Big data: A survey," *Mobile Netw. Appl.*, vol. 19, no. 2, pp. 171_209, Apr. 2014.
[15] Y. Fei, V.W. S.Wong, and V. C. M. Leung, "Efficient QoS provisioning for adaptive multimedia in mobile communication networks by reinforcement learning," *Mobile Netw. Appl.*, vol. 11, no. 1, pp. 101_110, Feb. 2006.
[16] O. F. Celebi *et al.* "On use of big data for enhancing network coverage analysis," in *Proc. ICT*, Casablanca, Morocco, May 2013, pp. 1_5.
[18] L. Cui, F. R. Yu, and Q. Yan, ``When big data meets software-defined networking: SDN for big data and big data for SDN," *IEEE Netw.*, vol. 30, no. 1, pp. 58_65, Jan./Feb. 2016.
[19] D. Kreutz, F. Ramos, P. E. Veríssimo, C. E. Rothenberg, S. Azodolmolky, and S. Uhlig, ``Software-de_ned networking: A comprehensive survey," *Proc. IEEE*, vol. 103, no. 1, pp. 14_76, Jan. 2015.
[20] Q.Yan and F.Yu, ``Distributed denial of service attacks in software-defined networking with cloud computing," *IEEE Commun. Mag.*, vol. 53, no. 4, pp. 52_59, Apr. 2015.
[21] Q. Yan, F. R. Yu, Q. Gong, and J. Li, ``Software-defined networking (SDN) and distributed denial of service (DDoS) attacks in cloud computing environments: A survey, some research issues, and challenges," *IEEE Commun. Surveys Tuts.*, vol. 18, no. 1, pp. 602_622, Jan. 2016.
[22] Y. Cai, F. R. Yu, C. Liang, B. Sun, and Q. Yan, ``Software defined device-to-device (D2D) communications in virtual wireless networks with imperfect network state information (NSI)," *IEEE Trans. Veh. Technol.*, to be published, doi: 10.1109/TVT.2015.2483558.

BIOGRAPHY

N.Siva Priya holds a B.E., in Electronics and Communication Engineering from Anna University Chennai and M.E in communication Systems from Anna University Chennai. She is currently pursuing her Ph.D (full time) in Information and Communication Engineering from Anna University Chennai.

Dr. N.Shenbagavadivu graduated in Computer Science from Bharadidasan University, Trichy holds MCA and M.Phil from the same university. She pursuded her Ph.D in Computer Science from Mother Theresa university, Kodaikanal. She is currently working as an Assistant Professor in Department Of Computer Applications at Anna University (BIT campus) Trichirapalli. She had published over 15 international journals and 15 years of teaching experience.