On Analyzing Indian Cellular Traffic Characteristics for Energy Efficient Network Operation

Karunakaran Kumar, Archit Gupta, Rushabh Shah*, Abhay Karandikar, Prasanna Chaporkar
Department of Electrical Engineering,
Indian Institute of Technology Bombay, Mumbai - 400076, India
{kkumar, architgupta93, karandi, chaporkar}@ee.iitb.ac.in
*rushabh@ticet.res.in

***Abstract*—Recent proliferation of mobile devices and high market demand have pushed power consumption of cellular networks to high levels in India. At the same time, the marginal gains to telecom operators for providing services have dwindled. Thus, a gap is slowly building up in the demand and supply of telecom services. The effect is adverse in urban areas where the demand for throughput and other load handling capabilities are high. While the BSs are setup to meet the peak QoS demands, vast opportunities exist to save operational and energy cost when loads are low. For the traffic patterns of one of India's leading telecom service providers, we show that such opportunities do exist in a systematic fashion and can be tapped to lower the operational cost. We further show that these opportunities can be predicted reliably and discuss a possible scheme to cut down on both energy and operational cost.**

I. Introduction

Recently there has been a significant interest to reduce the energy consumption of mobile cellular networks. The reasons for this are threefold. First, it has an environmental impact – In 2011, a study stated that telecommunication networks alone posed a Green House Gas (GHG) footprint of 0.20 Gigatonne (Gt) CO_2, accounting for 22% of the entire Information and Communication Technology (ICT) footprint and 0.41% of the global GHG emissions (49 Gt CO_2). This is expected to grow to 0.30 Gt CO_2 by the year 2020, increasing by 50% in less than a decade, much faster than the global GHG emmissions (projected at 55 $GtCO_2$) [1]. Secondly, the rate at which the demand is increasing – it is estimated that the ICT energy consumption is doubling every 4-5 years [2], [3] and with the recent proliferation of smart phones, data-centers etc., and the increase in penetration of services, this trend can only get worse. Finally, from the operators perspective, it is not just about environmental responsibility but also about the Operational Expenditure (OpEx). Out of the total OpEx, a major chunk goes into the electricity bill, out of which it is estimated that the Radio Access Network (RAN) accounts for about 60% [4]. OpEx optimization is even more relevant in a country like India, where the low Average Revenue per User (ARPU) puts pressure on operator's profitability.

Among other things, a typical network design involves determining Base Station (BS) locations, frequency allocation etc. to meet the Quality of Service (QoS) parameters such as the Call Drop Ratio (CDR). The network design is usually over-provisioned and more BS and resources are typically alloted to cater to the peak utilization in dense urban areas. However, the traffic experienced by a BS would not be the same throughout the day, while the BSs are fully operational. It has been observed that the energy consumed by a BS is *fairly* independent of the carried load [4] and thus it consumes a similar amount of energy even when it is completely under utilized.

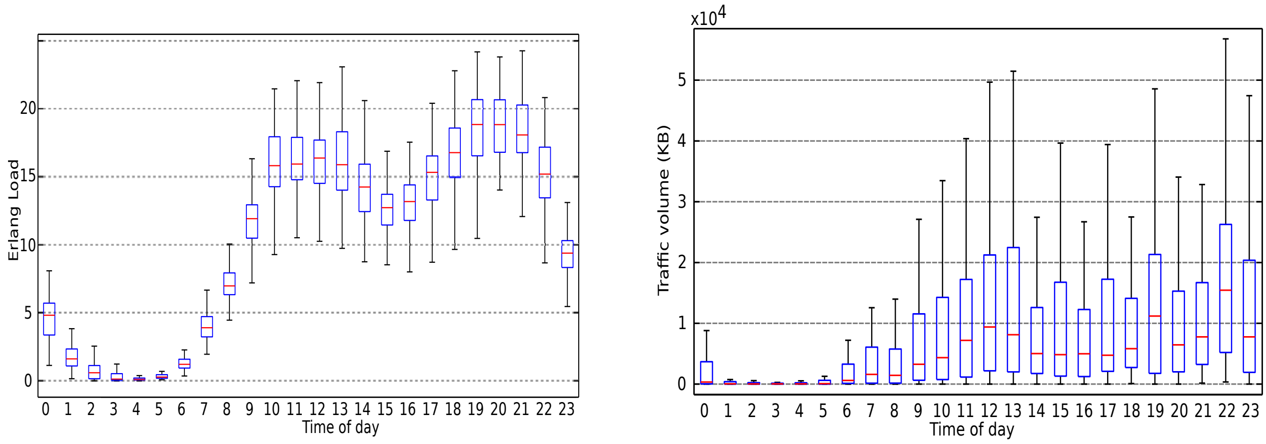
A lot of effort has been made to cut energy consumption in every aspect of the RAN – efficient power amplifiers, remote radio head, natural cooling, cell breathing, use of renewable sources, etc [5]. In this paper, we focus on load based operation of networks and inter operator RAN sharing with emphasis on the Indian perspective.

The rest of the paper is organized as follows: in Section II and III we present some results of analysis carried out on hourly traffic history at a few BS which were obtained from one of the leading telecom operator in India. For load dependent operation of RAN, it is essential to have a mechanism to reliably predict the traffic conditions (both voice and data) so that we can selectively switch off certain BSs appropriately. Several studies have been carried out in this context and usually the traffic conditions are measured by hourly average Erlang load for voice traffic, and hourly data volume for data traffic [6]. In [7] the authors consider the voice traffic to be a cyclo-stationary Gaussian process and use this assumption for prediction. In this paper, we make use of Auto Regressive Moving Average (ARMA) modeling for predicting the traffic.

In Section IV, we present an analysis that brings out the fact that it is indeed possible to shutdown some of the BSs while still maintaining coverage. For this as well, we gathered some information from the operator like geo-position of BSs, the azimuthal orientation of antennas etc., in a certain region of a city (Lucknow, Uttar Pradesh). In [3], the authors make a similar study of cost savings of a single operator as well as multi-operator RAN shared network, however in the Indian context, we need to consider the hand-off problem in inter-operator RAN sharing which we discuss in Section V, followed by conclusion. To the best of our knowledge, this is the first study based on real-life data collected from an Indian telecom operator.

II. Traffic analysis

The historical traffic data that we have obtained from the operator comprises of hourly Erlang load and hourly data volume on a certain number of BSs for a period of around 4 months. We first consider the voice traffic. The traffic behaviour clearly demonstrates periodic variation over the



(a) (b)

Fig. 1. Box plot of voice traffic (a) and data traffic (b), at different times of day for data collected from a leading telecom operator in India. The box, embedded marker and the surrounding lines indicate the standard deviation, median and 75-percentile marks respectively of observed load at that hour for a certain BS. Outliers not shown.

days. Figure 1a) shows the variation of Erlang load over a single day for a typical BS. This periodic pattern has been observed at all the BSs except for slight variations like, the time of day when traffic was higher etc. For example, peak load at a certain BS near a bus terminal was observed around 7:00 PM, whereas for BSs located in residential areas, it was around 9:00 PM. For all the BSs however, the load early in the morning was invariably very low and exhibited very little variance, as can be observed in Figure 1a.

Due to this inherent periodicity of the traffic, we assume that it is cyclo-stationary in nature. Such assumption is often made in practice whenever there is such a periodic component and has been used in literature for example in Electrical Load forecasting which shares a similar property (see [7], [8]). Figure 2 depicts the Empirical Correlation Coefficients of the load at a certain BS. In obtaining this, the weekends and holidays have been excluded, the mean has been adjusted, and the following notion of ergodicity has been assumed,

$$\mathbb{E}\, X_i X_j = \sum_{k=0}^{N} x_{24k+i} x_{24k+j} / N,$$

where X_i and X_j denote the load at the i^{th} and j^{th} hours of day respectively and N denotes the number of days. From Figure 2, it can be noted that a significant amount of correlation exists between X_i and X_{24+i} (i.e., previous day's load at same time) and that with X_{i-1}, X_{i-2}, etc., (i.e., adjacent hour's load), except during early hours. This pattern has been observed on all the BSs as well.

Data traffic on the other hand is quite 'fuzzy'. However, during the early hours, the load has a lower variance and mean (see Figure 1b) and for the rest of the times, the load has a high variance. Moreover, there is no significant correlation between X_i and X_j for $i \neq j$ for most of the BSs. A few BSs demonstrate correlations similar to that for voice traffic, however, magnitude of correlation coefficient is low. Note that for the purpose of load dependent BS operation, the hourly data volume *might* not be completely relevant, we need to capture the hourly throughput requirement instead, which we would like to investigate further.

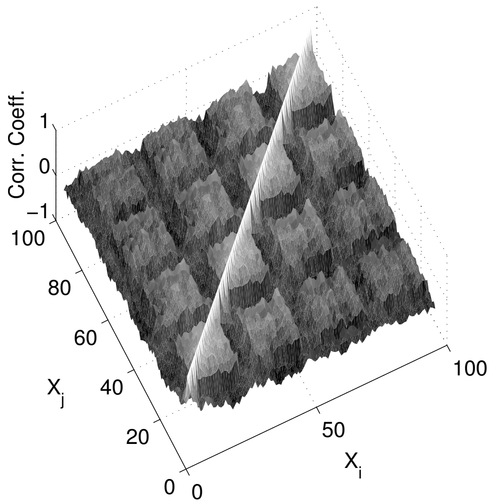


Fig. 2. Empirical Correlation Coefficient between X_i and X_j for $0 \leq i, j \leq 95$ of voice traffic

III. Voice Traffic Prediction

The periodicity in voice traffic has been visualized by its partial autocorrelation (PACF). For a time series, with its tth give by X$_t$, the partial autocorrelation of lag k, say α_k, is the autocorrelation between X_t and X_{t+k} after removing the linear dependence on the intermediate terms. The ACF as evaluated empirically in the previous section (Figure 2) is, however, not the most appropriate metric for modeling the time series as it does not account for these linear dependencies. Hence, we evaluated the PACF to evaluate the true correlation amongst data points. For our case, there happens to be an invariably strong dependence near lags of 24 hours and spread over a computationally small order (2-3 days in the past).

This trend in historic data is one of the key motivating factors to model the voice traffic as an ARMA process with a seasonality of 24 hours. Modeling the series by applying either

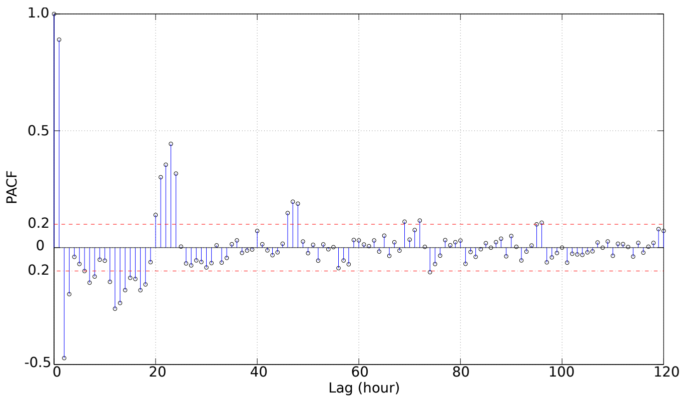


Fig. 3. Partial Auto Correlation Function (PACF) of voice traffic based on the collected data

first order differencing or Seasonal ARMA (SARMA) yields satisfactory reliability. Interestingly though, and as can be seen in Figure 1a, the variance in Erlang load, when measured over around 4 months, also varies significantly over the day. In fact, modeling the load at particular time for different days as a time series ($\{X_i, X_{i+24}, X_{i+48}, ...\},\ 0 \leq i \leq 23$) indicates several other properties that can be exploited to selectively improve the prediction accuracy at certain hours of the day (especially early morning and late night).

We thus proceed with modeling the Erlang load at each hour as an ARMA process. In order to smoothen the resultant time series and to account for the correlation between loads observed at consecutive hours of the same day, we model the residue of the actual load and the output of the aforementioned ARMA processes (Seasonal Load) as another ARMA process. This approach serves multiple purposes for the network optimization process.

- Selectively improve the prediction accuracy for time intervals where we do want to optimize the network operation.
- Reduce the computational complexity of data prediction (Unlike the previous approach, PACF falls off after a few lags).

A. Results

Modeling the time series as mentioned above, we observe an average 17.9% error in the prediction over a period of 5 days (averaged over 140 BSs) when measuring error as the absolute error from the actual values as a fraction of the total load. A histogram depicting average errors recorded in the load values is presented in Figure 5. While predictions for more than $2/3^{rd}$ of the BSs resulted in an average prediction error less than 20%, there were a few outliers where the total load itself was very low (peak load $<$ 10 Erlang) making reliable prediction infeasible. Moreover, if we consider a variation of 2 times the standard deviation from the predicted value on the conservative side, the actual load lies below the 95% confidence limit in approximately 90% of the test cases.

The results for a particular BS have been shown in Figure 4. As can be noted from the figure, the actual load lies very close to the predicted value at hours when we wish to reduce the

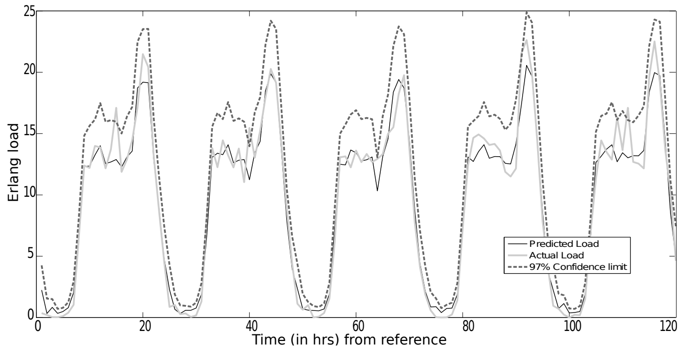


Fig. 4. Prediction results for a BS for 120 hours along with a conservative, 97% confidence limit on the predicted value

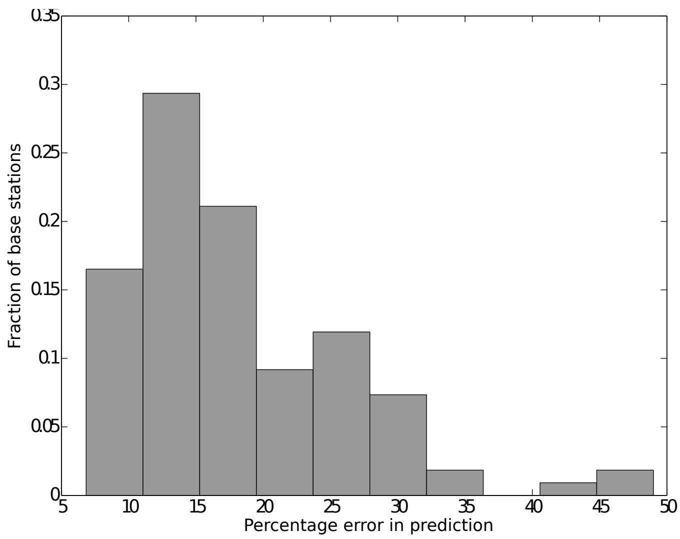


Fig. 5. Average prediction Error Percentage after performing predictions on 140 BSs for which we obtained the data

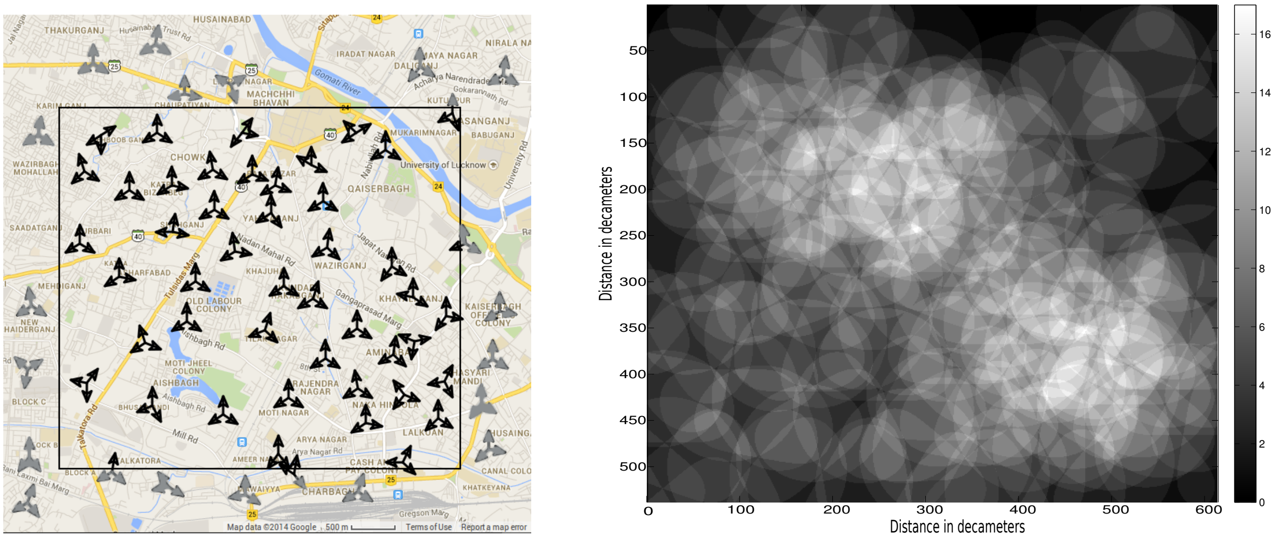
operation cost (early morning and late night). Similar trends were observed for most other BSs as well.

While the former observation indicates that the prediction will not cause a violation in the QoS constraints, the latter gives a precise information of the network state when operation cost can be reduced.

IV. Single operator BS on/off

We now consider the possibility of switching off BSs – we admit the possibility of switching off any sectorial antenna (referred to as *nodes* henceforth) independent of the others, without sacrificing on coverage. Figure 6a depicts the locations of 66 BSs in a 6km $\times$ 5km region of the city of Lucknow and orientations of their sectorial antennas (nodes) as obtained from the operator. A simulation environment was developed to obtain network behaviour for different scenarios. The details of the simulator are as follows:

- transmit power of each node when switched on is 10Watt,
- radiation pattern of each node is $\cos(\theta)$, $\theta \in [-\pi/2, \pi/2]$, where $\theta = 0$ is it's azimuthal orientation,



(a) (b)

Fig. 6. BS locations and sectorial antenna orientations as obtained from the operator, which were used for the simulation are given in (a). The redundancy present in the network is illustrated in (b)

- receive power threshold is set to -95dBm,
- path loss factor is chosen as 4.5 to get a conservative estimate of the coverage areas.

Figure 6b illustrates the ‘redundancy’ present in the network when all the nodes are fully operational, i.e., the number of nodes which can provide service to a given spatial location. As we can deduce from the plot, a given spatial point can be served by many nodes (by 6 nodes on an average).

There are a total of 198 nodes (66×3) in the given region. Out of these, we restrict the 63 nodes (shown grey in Figure 6a) located along the edges of the region to be in the ‘on’ state and we consider rest of the 135 nodes for optimization. We try to minimize the number of nodes (out of these 135 nodes) to be switched ‘on’, subject to the constraint that the received signal strength at each spatial (grid) point is at least -95dBm. A rectangular array of grid points spaced 10m from each other were considered for computing received signal strength. The objective function is,

$$f(\mathbf{x}) = \begin{cases} \sum_i \mathbf{x}_i, & \text{if atleast 99\% grid points are served,} \\ \infty, & \text{otherwise} \end{cases}$$

where, $\mathbf{x}$ is a vector of length 135, and $\mathbf{x}_i \in \{0,1\}$ denotes the state of node i (off/on). There are 2^{135} possible states which must be searched to get the optimal state of operation. We use Simulated Annealing [9] algorithm for the optimization.

Simulated annealing is a randomized search algorithm which must be iterated multiple times in order to determine the optimal solution. Upon convergence, it was found that only 52 of the considered 135 nodes were needed to maintain 99% coverage. The results from Section III indicate that the traffic load with 97% confidence is extremely low during the early hours (i.e., 1AM to 6AM) as compared to the rest of the day’s traffic. Therefore, during these hours, the relevant QoS metric is the signal quality which we are able to maintain with minimal number of active nodes.

V. Inter Operator Sharing

The above simulations capture the redundancies of only a single operator. However, typically there are multiple operators operating in the same region each of which would have a similar coverage profile. Often, the operators choose to co-locate their BSs in order to save on the capital expenditure needed for the setup as well as on rent. We also obtained the number of co-tenants of a few BS sites and it was found that almost 85% of the BSs are co-located by other operators and thus we may expect them to have similar coverage regions and be able to serve users when the host operator switches off certain BSs.

Given the predictability of the voice traffic, especially during the early hours, it is unnecessary for all the operators to run their BSs with full capacity. An operator can switch off some of the BS and let some other co-located operator to handle the traffic and pay them accordingly for the provided service.

A. Policy Regulations

Multi-operator RAN sharing can lead to significant savings and thus, we can envision a situation where all the operators in a telecom circle[1] collaborate and operate their RAN in a collective fashion. However, in India it is not permitted to do so as of now. As per Indian telecom regulations, inter operator handoffs are not permitted, i.e., if a call is originated in a region served only by say operator 1, then if the user moves to a region served only by operator 2, the call cannot be handed over to operator 2 and thus it must be dropped. This would degrade the service quality experienced by the users.

[1]A telecom circle is a certain large geographical region within which an operator is permitted to provide service.

VI. Conclusion

While demand for energy is bound to rise, efficient strategies for network operation are a way forward to meet the operational demands. In this paper, we have presented methods to reliably predict the network behavior which is one of the key steps in deciding the optimal operation strategy. The results obtained from our experiments on real-life cellular traffic data demonstrate the feasibility of a load dependent operational strategy for the network. Besides predicting the network state (voice traffic), we have also demonstrated that in an Indian urban setup, alternate strategies can be deployed to significantly save the operation cost.

Our future work would be directed towards expanding the scope of the proposed model in Section IV to incorporate other QoS metrics like CDR, Minimum Data Throughput, etc., using insights gained from the analysis presented in this paper. We would also like to look at other distributed approaches for optimizing the network operation cost to reduce the computational complexity and improve the scalability.

Acknowledgements

We thank the telecommunication operator for providing the data, and insights on state of policy regulation. We would also like to thank Ms R. Deepika (summer intern at our research lab) for her insights and contribution to the data analysis.

References

[1] "Gesi smarter 2020: The role of ict in driving a sustainable future," Global e-Sustainability Initiative, Tech. Rep., Dec. 2012. [Online]. Available: http://gesi.org/SMARTer2020

[2] G. Fettweis and E. Zimmermann, "Ict energy consumption – trends and challenges," in *Proceedings of the 11th International Symposium on Wireless Personal Multimedia Communications*, vol. 2, no. 4, 2008, p. 6.

[3] E. Oh, B. Krishnamachari, X. Liu, and Z. Niu, "Toward dynamic energy-efficient operation of cellular network infrastructure," *Communications Magazine, IEEE*, vol. 49, no. 6, pp. 56–61, 2011.

[4] G. Auer, V. Giannini, C. Desset, I. Godor, P. Skillermark, M. Olsson, M. Imran, D. Sabella, M. Gonzalez, O. Blume, and A. Fehske, "How much energy is needed to run a wireless network?" *Wireless Communications, IEEE*, vol. 18, no. 5, pp. 40–49, October 2011.

[5] Z. Hasan, H. Boostanimehr, and V. Bhargava, "Green cellular networks: A survey, some research issues and challenges," *Communications Surveys Tutorials, IEEE*, vol. 13, no. 4, pp. 524–540, Apr. 2011.

[6] U. Paul, A. Subramanian, M. Buddhikot, and S. Das, "Understanding traffic dynamics in cellular data networks," in *INFOCOM, 2011 Proceedings IEEE*, April 2011, pp. 882–890.

[7] R. Cavalcante, S. Stanczak, A. Eisenblätter, T. Kürner, D. Montvila, R. Nowak, and S. Stefanski, "Voice and traffic prediction in gsm and umts cells with real network measurements," European Cooperation in Science and Technology, Tech. Rep., 2013. [Online]. Available: http://www.ic1004.org/uploads/Meetings/VII%20MC%20-%20Ilmenau/Workshop/Papers/Cavalcante_1.pdf

[8] D. J. Leith, M. Heidl, and J. V. Ringwood, "Gaussian process prior models for electrical load forecasting," *Probabilistic Methods Applied to Power Systems*, pp. 112–117, 2004.

[9] P. Bremaud, *Markov chains: Gibbs fields, Monte Carlo simulation, and queues*. Springer, 1999.