

A Study of LTE Network Performance based on Data Analytics and Statistical Modeling

Sanqing Hu¹, Ye Ouyang¹, Yu-Dong Yao¹, M. Hosein Fallah¹, Wenyuan Lu²

¹Stevens Institute of Technology, NJ, USA

Email: {shu1, youyang, yu-dong.yao, hfallah}@stevens.edu

²VPIsystems, NJ, USA

Email: derek.lu@vpisystems.com

Abstract—Various simulation based approaches are developed to study wireless network performance and capacity traditionally. However, these simulation based approaches has its own limitations in supervised and unsupervised learning. Since big data techniques become more available, using big data techniques to understand the huge amount of the LTE network measurements and diagnosis data, evaluate and predict the LTE network capacity and performance becomes a very promising approach. This paper investigates the LTE network performance from the aspect of data analytics and statistical modeling. In this paper, we develop a methodology of data analytics and modeling to evaluate LTE network performance based on traffic measurements and service growth trends. A relational algorithm is developed to obtain the relationship between LTE network resources and LTE KPIs and a forecasting algorithm is developed to trend the network resource consumptions based on traffic and service growth. The numerical results show a high accuracy, robustness, and reliability of the proposed methodology.

Index Terms—Big Data Analytics, Statistical Modeling, LTE, Network Performance, Network Capacity, Network Resource

I. INTRODUCTION

In recent years, with the exponential increasing of smart devices and mobile data services, a large amount of new subscribers and a significant revenue growth have been brought to mobile operators. However, since the wireless network capacity is a finite resource, the data tsunami and signaling storm are overwhelming the wireless networks nowadays. To prevent network resources from being unexpectedly consumed, it requires mobile operators to leverage appropriate strategies of capacity management to meet the ever-increasing network capacity requirements.

The capacity management in wireless networks refers to a process of ensuring the wireless network is properly sized to meet traffic demand and to consume network resources in a most efficient way as much as possible. Due to mobile data tsunami, mobile operators have suffered unexpected network overloading to various extents over the past years. Traditional approaches to wireless capacity planning consider one primary factor only which is the subscriber count. A generic function is defined as future capacity for traffic demand = current capacity + incremental capacity proportional to subscriber growth. Before smart phone appeared, devices behaved in a

relatively similar manner since applications running on them were voice service and simple data services only. The traffic model based upon that service scenario is stable and trendable given a period of time. When dimensioning networks with a target at forecasting network capacity trends, mobile operators used to focus on the overall traffic transmitted through the wireless pipes where they also consider coverage and interferences.

Due to smart phones with diversified traffic patterns and user behaviors, the traditional method for network resource planning, based on service traffic growth only, has failed to solve the dilemma between network capacity and Capital Expenditure (CAPEX). Hence, a scientific approach becomes necessary to forecast network resources and dimension network capacity. In particular, the method should rely on a joint projection of dynamic changes in the traffic model where the subscriber count, smart phone user behaviors, and service patterns are taken into account in a timely manner.

This paper introduces a systematic methodology of data analytics and modeling to evaluate Long Term Evolution (LTE) network capacity based upon traffic measurements and service trends [1]. In this paper, we propose a novel methodology to trend the dynamic change in the traffic model using the historical network measurements and diagnosis data, which is due to the diversified services patterns and user behaviors. Such dynamic changes caused by the diversities are then reflected in the change of network traffic, which is ultimately carried by network equipped capacity. Furthermore, the capacity is supported by end-to-end network equipped resources, whose physical threshold (upper and lower bound) is determined by how different types of traffic generated by the diversified services and devices are effectively utilized on the software and hardware of the network equipment. A traffic model is developed with parameters that can represent the statistical characteristics of the network traffic and network resources. The systematic method should be auto-tunable in computing network capacity to reflect the dynamic changes in the traffic model. The method should be an incremental learning process to identifying the pain points in network capacity management and perform accurate network planning.

This paper is organized as follows: Section II presents the related works for wireless network performance evaluation;

Section III presents the proposed novel qualitative modeling for LTE network performance evaluation; the relational algorithm and forecasting algorithm is proposed in Section IV; Section V shows the numerical results to validate this study; and finally Section VI draws conclusion.

II. RELATED WORKS

Wireless network performance, service quality, and capacity management have been studied by many researchers in both academia and industry. Most of the existing research on 2G, 3G, or LTE networks is performed by various simulation approaches.

The simulation based approach to study performance and service quality for 3G UMTS network has been widely accepted and recognized in academia. [2] leveraged simulation to design and evaluation of three possible UMTS-WLAN interworking strategies, [3] proposed a tree-like UMTS terrestrial network where they developed two types of heuristic algorithms to ensure a low magnitude of traffic loss in such a network topology. One of their algorithms solves the problem by modifying the tree-topology, while others expand the network by inserting additional links. Their approach showed how to find a good compromise between topology refinement and network expansion in the case of realistic network scenarios. [4] presented a new content-based, non-intrusive quality of experience (QoE) prediction model for low bi-rate and resolution (QCIF) H.264 encoded videos and to illustrate its application in video quality adaptation over UMTS networks. They argued that the success of video applications over UMTS networks very much depends on meeting the QoE requirements of users. In their study, they performed the simulation via NS2 to demonstrate the effectiveness of the proposed adaptation scheme, especially at the UMTS access network which is a bottleneck. [5] performed an analysis on uplink and downlink air interface capacity. They defined canonical capacity as the maximum number of concurrent users of that service and define the network total load as a function of this parameter and the number of concurrent users. A number of case studies are also presented to compare the results of utilizing the proposed framework with the results of network simulation. This comparison demonstrates the accuracy and usefulness of the proposed approach for network design and dimensioning.

From UMTS to LTE, simulation is considered as a smooth continuation as the most typical approach to study network performance, quality, and capacity related problems. Below are a few typical research on LTE network performance, service quality, and capacity using simulation based algorithms and tools. For example, [6] conducted a typical simulation for LTE networks at system level. A LTE system level simulator was offered in Matlab to evaluate the downlink shared channel of LTE SISO and MIMO networks using Open Loop Spatial Multiplexing and Transmission Diversity transmit modes. [7] conducted a simulation analysis on S1

interface of LTE networks to study the throughput behavior. The network structure and topology are constructed via OPNET. A few different traffic scenarios are simulated in the network to investigate the traffic behavior in S1 interface. [8] presented a traffic-light-related approach to autonomous self-optimization of tradeoff performance indicators in LTE multi-tier networks. Introducing a low-complexity interference approximation model, the related optimization problem is formulated as a mixed-integer linear program and is embedded into a self-organized network operation and optimization framework. The optimization procedure is carried out considering time-variant optimization parameters that are automatically adapted with respect to changes in the network. Simulation-based evaluation of representative case studies demonstrates applicability and the benefit potential of our overall concept. [9] studied LTE cell selection process to determine the cells that provide service to each mobile station. They optimized these processes with a target at maximizing the utilization of current and future cellular networks. The simulation results indicate that the algorithms obtain up to 20 percent better usage of the network's capacity, in comparison with the current cell selection algorithms. [11] and [12] conducted a simulation analysis on UMTS and LTE networks to study the throughput behavior. The network structure and topology are constructed via Opnet. A few different traffic scenarios are simulated in the network to investigate the traffic behavior in wireless networks.

III. QUALITATIVE MODELING TO EVALUATE LTE NETWORK PERFORMANCE

The big data analytics leverages true measured data to study a given problem in the wireless networks. The measured data from network side or from users faithfully reflects the network and user behaviors, and the analysis result is more reliable and trustworthy. The big data analytics is normally a supervised-learning (can be unsupervised also) or posterior learning process, using the measured data to infer a certain pattern behind the data and reveal unseen situations in a reasonable way. In order to solve a given problem on network performance and capacity in UMTS or LTE networks, the following steps should be performed using the approach of data analytics and modeling:

- 1) *Determine the target network resource / KPI to be evaluated:* in data analytics, this step is to confirm the target objective, as the dependent variable in an unknown function to be derived by an appropriate algorithm. Some typical network resources at air interfaces that can represent wireless capacity, network coverage quality, and interferences in LTE and UMTS networks are Downlink Power, which is also known as Transmitted Carrier Power (TCP), Received Total Wideband Power (RTWP), Downlink and Uplink Channel Elements, Orthogonal Variable Spreading

Factor (OVSF) Utilization, Reference Signal Receive Power (RSRP), Received Signal Strength Indicator (RSSI), Received Signal code Power (RSCP), and CPU Utilization if any.

- 2) *Determine the representation of the input feature for the learned function:* this step is to conduct feature selection in which appropriate independent variables are selected. Typically the selected input variables are transformed into an independent feature vector, which contains the information gain that is descriptive of the object. In our case, those input variables (features) are network counters and KPIs showing high correlation to target network resources to be predicted. The number of features cannot be too large because of the curse of dimensionality; but should contain sufficient information gain, which is also known as entropy, to accurately predict the target network resources.
- 3) *Data set:* the overall data set is split into training data set and test set. Training set is a set of data to discover potentially predictive relationships between independent variables and dependent variables, which are target network resources in our case. The independent variables and dependent variables in the training set are used together with a supervised or unsupervised learning method to train the relationships between the response value, which is a target network resource, and the predictors, which are the independent variables. The training process is to fit a model that can be used to predict the target network resource from the independent variables selected in step 2. The test set is a set of data with same independent variables and dependent variables that is independent of the training set, but that follows the same probability distribution as the training data. It is used to assess the strength and utility of the predictive relationship derived by the training set. If a model fit to the training set also fits the test set well, minimal over-fitting has taken place. If the model fits the training set much better than it fits the test set, over-fitting is likely the cause.
- 4) *Model development, validation, and test:* this step serves to determine structure of the model, the learned model, and corresponding learning algorithm. Several algorithm candidates are selected by domain knowledge to train the relationship between target network resource and network KPIs. While performing machine learning to evaluate wireless network capacity, 3 primary steps should be followed.

- a) *Training phase:* present training data from "gold standard" and train the model, by pairing the input (network KPIs) with expected output (target network resource). Each algorithm candidate has its own

parameter options (the number of layers in a Neural Network, the number of trees in a Random Forest, etc). For each of your algorithms, we must pick one option. That's why we have a validation set for us to determine the best option in each algorithm.

- b) *Validation/test phase:* now there is a collection of algorithms to predict target network resource. A best algorithm should be determined. That's why we need a test set. We may want to pick the algorithm that performs best on the validation set, which is good. But, if not measuring the top-performing algorithm's error rate on the test set and just going with its error rate on the validation set, then we have blindly mistaken the "best possible scenario" for the "most likely scenario." Thus this step is to estimate how good the model has been trained and to estimate model properties (mean error for predicting network resource, classification errors in classifying network features etc.)
- c) *Application phase:* apply the freshly-developed model to the new data and get the predictive network resources. Since we normally don't have any reference value of target network resource in this type of data, we can only speculate about the accuracy of predicted network resource using the results of the validation phase.

IV. FORECASTING LTE NETWORK PERFORMANCE

A qualitative schematic relation between a given LTE network KPI and its associated network resources is illustrated in Figure 1. In the figure, the X axis denotes the LTE network resource that is being consumed and the Y axis indicates the measured KPI. The LTE KPI as a function of LTE network consumed resources is split in four areas, separated by three dotted lines in Fig. 1.

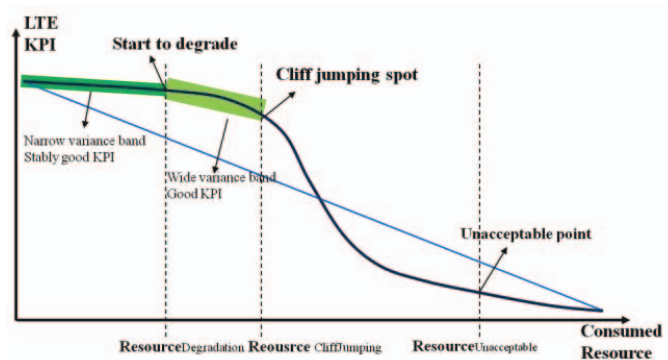


Fig. 1. Relational model for network KPI and resource

Zone 1: Constant good KPI - consider a vanishing distribution of LTE network KPI, the KPI is first equivalent

to the reference point at very beginning where almost no network resource is consumed which means the LTE network is extremely light loaded.

Zone 2: Sinking KPI- at a particular point or in a certain range, the formal quasi-perfect KPI scenario cannot be maintained any more.

Zone 3: Cliff jumping KPI-as the LTE network KPI continues to degrade and once it exceeds a particular threshold, such a quasi-flat relation between KPI and consumed resource starts to collapse.

Zone 4: Unacceptable KPI- as soon as the KPI drops at another particular point or a certain range, the KPI value behind that point or range is no longer acceptable.

we assume that the change of LTE network KPI depends on the current level of KPI ($\alpha \cdot \text{KPI}$), given a certain amount of resource (β) to be changed. The relation can be represented by the first step in Eq. (1).

$$KPI_{LTE} = A * \exp^{-B * Resource} + C, \quad (1)$$

Therefore, KPI is a function of n influence factors in terms of resource 1 to n .

$$\begin{aligned} KPI_{LTE} &= coeff_1 * (A_1 * \exp^{-B_1 * Resource} + C_1) \\ &+ coeff_2 * (A_2 * \exp^{-B_2 * Resource} + C_2) \\ &+ \dots + coeff_n * (A_n * \exp^{-B_n * Resource} + C_n) \\ &= \sum_{i=1}^n coeff_i * (A_i * \exp^{-B_i * Resource} + C_i) \end{aligned} \quad (2)$$

To obtain the optimal coefficient, the sigmoid model is used to derive the relation between the KPI and the resource indicators. In this model, the cliff jumping spot can be derived by letting 2nd order derivative equal to 0, in which the point at 2nd derivative changes sign. In addition, the smooth area in Zone 1 and 2 can also be fitted by sigmoid curve. The standard sigmoid function, which is represented by

$$y = \frac{1}{1 + \exp^{-x}}, \quad (3)$$

However, the sigmoid curve of the cliff jumping model may not be as perfect as the standard format. So let the S-Curve function as Eq. (4)

$$KPI_{LTE} = f(resource) = \frac{A}{1 + \exp^{B * Resource - C}}, \quad (4)$$

where A denotes the max value of KPI, B denotes the curvature of the sigmoid curve, C denotes inflection point of the sigmoid curve.

Then, the Eq. 2 can be rewritten as

$$\begin{aligned} KPI_{LTE} &= coeff_1 * \left(\frac{A_1}{1 + \exp^{B_1 * Resource - C_1}} \right) \\ &+ coeff_2 * \left(\frac{A_2}{1 + \exp^{B_2 * Resource - C_2}} \right) \\ &+ \dots + coeff_n * \left(\frac{A_n}{1 + \exp^{B_n * Resource - C_n}} \right) \\ &= \sum_{i=1}^n \frac{A_i}{1 + \exp^{B_i * Resource - C_i}} \end{aligned} \quad (5)$$

The forecasting model is depicted in Fig. 2. In the figure, the trend component T reflects how user behaviors, rate plan policy, and change of user numbers impact LTE network traffic and network resource consumption in a long term. The trend component is given by

$$T_t = f(X_k * Slope_k), \quad (6)$$

and the fitting algorithm is given by

TABLE I
FITTING ALGORITHM FOR TREND COMPONENT

1. Divide the time series of KPI $X(t)$ to m pieces;
2. For piece $K = 1$, fit a line with $Slope_K$ and starting point X_k ;
3. For piece $K = 2$ to m , fit a line with $Slope_K$ and starting point of Line K = the last fitted point of Line $K-1$;
4. Minimize the mean error rate between fitted value and true value in each piece $K = 1$ to m .

Seasonality component represents the periodical changes of traffic in a given period. Wireless traffic or resource consumption normally has a weekly period (7 days), which means the traffic of any two adjacent 7 days are highly correlated [10].

To forecast seasonality component, first we need to obtain the length of each period. By now the length of period is

TABLE II
FITTING ALGORITHM FOR DECIDING THE SEASONALITY PERIOD LENGTH

1. Assume the length of time series for a given LTE KPI is L . Assume the possible period length $i = 1$ to $L/2$. In each i , there are j pieces;
2. For i from 1 to $L/2$, compute the variance for each piece j in each i ;
3. Sum the variance of each piece j in each period i , $\sigma_{i_within}^2$. In other words, sum the variance for each period within the same i , which is within-group-sum-of-error;
4. For each i from 1 to $L/2$, assume p equal to the number of sample points in each piece for each i . We have $p = L/i$;
5. For each i from 1 to $L/2$, construct p sets of sample data, each of which includes all the samples in the same position in the p pieces ($q = 1 \dots p$). Compute the variance of each p ;
6. Sum the variance of each piece p in each period i , $\sigma_{i_between}^2$;
7. For each i from 1 to $L/2$, select the i with minimized $\sigma_{i_between}^2 \sigma_{i_within}^2$.

obtained. The final step is to derive the value of seasonality

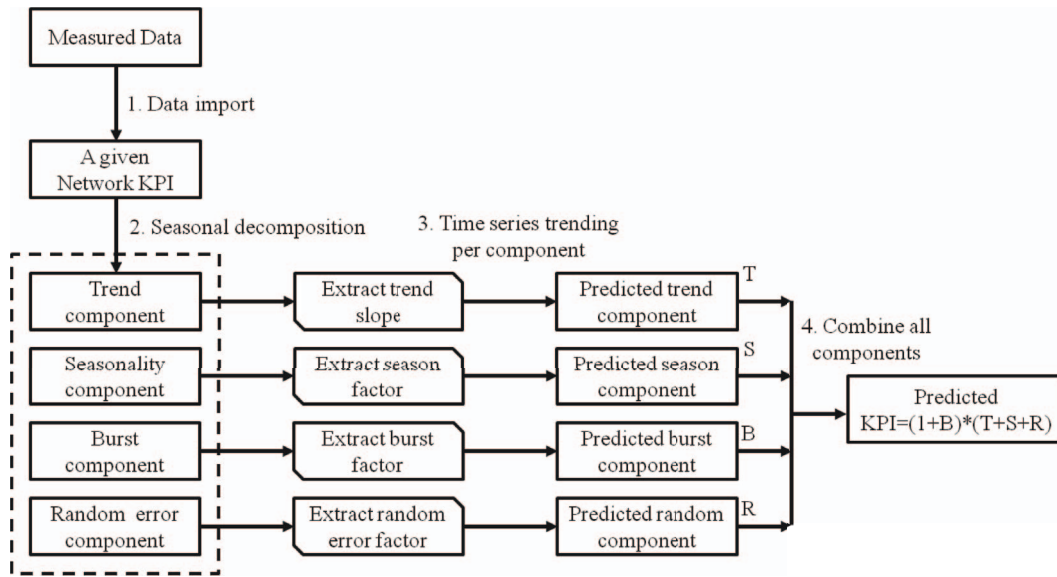


Fig. 2. Forecasting model for network resource and KPI

(period) component per period. The value of seasonality component in each position q ($q = 1 \dots p$) is given by the mean value of data points at the same position q in the p sample sets.

Burst component B indicates a significant change from normal trend which is caused by external known or unknown factors. Those known factors to be happening are predictable such as gambling day, festivals, game day etc. The unknown factors which are unpredictable may be caused by some random events with small probability. For example in a short period many users make calls simultaneously which generate huge traffic shortly.

Random component R can further be decomposed to stationary time series $RS(t)$ and white noise $RN(t)$.

V. NUMERICAL RESULTS

To evaluate the performance of the proposed relational and forecasting model, we select the RRC connection setup success rate as a sample KPI to deploy simulations. The RRC connection establishment is used to make the transition from RRC Idle mode to RRC Connected mode. UE must make the transition to RRC Connected mode before transferring any application data, or completing any signaling procedures. The RRC connection establishment procedure is always initiated by the UE but can be triggered by either the UE or the network. The signaling for RRC connection establishment is shown in Fig. 3. The entire procedure is completed using only RRC signaling. A 3-way handshake is used to move the UE into RRC connected mode.

This KPI describes the ratio of all successful RRC establishments to RRC establishment attempts for UTRAN network, and is used to evaluate UTRAN and RNC or cell admission capacity for UE and/or system load. It's a primary

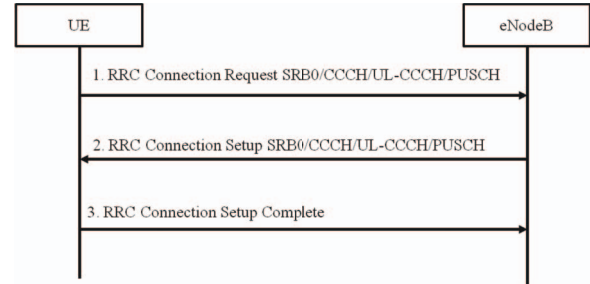


Fig. 3. Procedure of RRC connection setup

KPI to evaluate LTE accessibility performance. Moreover, we select the following typical LTE capacity/resource indicators:

- Average number of connected users per cell
- DL physical resource blocks usage
- UL physical resource blocks usage
- Physical DL control channel usage
- Paging resource usage
- Average number of active users in the UL buffer per cell
- Average number of active users in the DL buffer per cell
- Physical random access channel usage.

In the simulation, we are using real AT&T and Verizon historical data. The numerical result is shown in Table. III. In the table, the Mean Absolute Percentage Accuracy (MAPA) in training set ranges from 82.1% to 88.0% with a mean value 85.0%, which indicates the fitted values of the resource indicators in their respective training set shows very low variance between true and fitted values. The test set's MAPA ranges from 71.6% to 79.4% with a mean value of 75.4%, which is not far from the mean value in training set. This indicates that the reliability of the model is very good because the MAPA does not drop much from training to test set.

TABLE III
FITTING ALGORITHM FOR TREND COMPONENT

Resource Indicators	MAPA Training Set	MAPA Test Set	Goodness of fit
Average number of connected users	0.870	0.794	0.825
Downlink PRB Usage	0.863	0.787	0.810
Uplink PRB Usage	0.830	0.728	0.790
PDCCH Usage	0.838	0.758	0.802
Paging Resource Usage	0.845	0.722	0.781
Average number of active users in the downlink buffer	0.880	0.775	0.811
PRACH Usage	0.821	0.716	0.759
Mean value	0.850	0.754	0.797

Moreover, the goodness of fit of all resource indicators are above 75.9%, which is also an optimal result showing that the independent variables in terms of those 4 components can well explain the variance of each resource indicator.

The relational model shows the MAPA is 88.19% in training set and 84.50% in test set. The forecasting model indicates the MAPA is 89.79% in training and 84.23% in test set. As a whole, any given LTE KPI is obtained through step 1 to 4 where relational model and forecasting model are utilized. So the MAPA of overall analytical model will be 79.19% for training and 71.17% for test set.

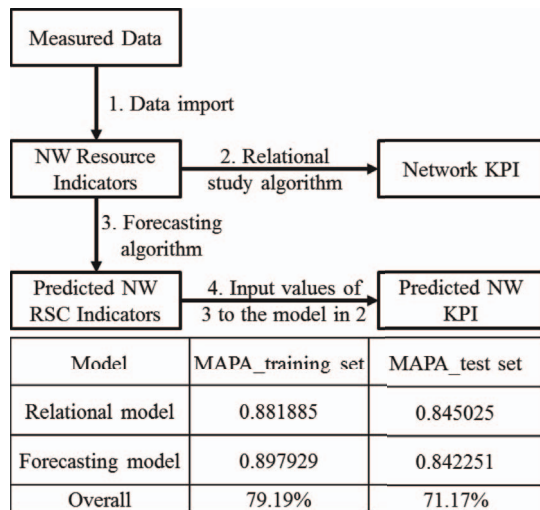


Fig. 4. Overall MAPA results

VI. CONCLUSIONS

This paper investigates the methodology for evaluating the LTE network capacity and performance from the aspect of big data analytics and statistical modeling. In this paper, we develop a novel mode to evaluate LTE network performance based on traffic measurements and service growth trends. A relational algorithm is developed to obtain the relationship between LTE network resources and LTE KPIs and a forecasting algorithm is developed to trend network resource consumptions based on traffic and service growth. To evaluate the validity of this study, case studies have been done using data gathered from AT&T and Verizon. The numerical results

show that the proposed methodology has a high accuracy, robustness, and reliability.

Moreover, this model can be a generalized model to study the performance of other networks such as UMTS, GPRS, and CDMA for the reason that the relationship between a network KPI and its associated network resource indicators in any given networks can be derived based upon the relational model. Since UMTS and CDMA are still mainstreams among mobile operators globally, it may be of more significant value to mobile operators if "downgrading" this model to study 3G and even 2G network performance and capacity issues.

REFERENCES

- [1] Ye Ouyang, M. Hosein Fallah, Sanqing Hu, Yong Ren, Yirui Hu, Zhichang Lai, Mingxin Guan and Wenyuan Lu, "A Novel Methodology of Data Analytics and Modeling to Evaluate LTE Network Performance," *2014 IEEE Wireless Telecommunications Symposium*, 2014.
- [2] Shiao-Li Tsao and Chin-Ching Lin, "Design and evaluation of UMTS-WLAN interworking strategies," *2002 IEEE 56th Vehicular Technology Conference*, 2002.
- [3] Szlovencsak, A., Godor, I., Harmatos, J., and Cinkler, T., "Planning reliable UMTS terrestrial access networks," *IEEE Communications Magazine*, vol.40, no.1, 2002.
- [4] Khan, A., Lingfen Sun, and Ifeakor, E., "QoE Prediction Model and its Application in Video Quality Adaptation Over UMTS Networks," *IEEE Transactions on Multimedia*, vol.14, no.2, 2012.
- [5] Navaie, K., and Sharafat, A. R., "A framework for UMTS air interface analysis," *Journal of Electrical and Computer Engineering*, vol.28, no.3, 2003.
- [6] Ikuno, J.C., Wrulich, M., and Rupp, M., "System Level Simulation of LTE Networks," *2010 IEEE 71st Vehicular Technology Conference*, 2010.
- [7] Ouyang, Ye, "Traffic Dimensioning and Performance Modeling of 4G LTE Networks," *Doctor Dissertation of Stevens Institute of Technology*, 2012.
- [8] Engels, A., Reyer, M., Xiang Xu, Mathar, R., Jietao Zhang, and Hongcheng Zhuang, "Autonomous Self-Optimization of Coverage and Capacity in LTE Cellular Networks," *IEEE Transactions on Vehicular Technology*, vol.62, no.5, 2013.
- [9] Amzallag, D., Bar-Yehuda, R., Raz, D., and Scalosub, G., "Cell Selection in 4G Cellular Networks," *IEEE Transactions on Mobile Computing*, vol.12, no.7, 2013.
- [10] Li Wang, Zengzhi Li, and Chengqian Song, "Network traffic prediction based on seasonal ARIMA model," *Fifth World Congress on Intelligent Control and Automation*, 2004.
- [11] Ye Ouyang and Fallah, M.H., "A performance analysis for UMTS packet switched network based on multivariate KPIs," *Wireless Telecommunications Symposium*, 2010.
- [12] Ouyang, Ye and M. Hosein Fallah, "An Analysis of Traffic and Throughput for UMTS Packet Core Networks," *International Journal of Interdisciplinary Telecommunications and Networking*, vol.2, no.2, 2010.