ACCEPTED FROM OPEN CALL

SPATIAL MODELING OF THE TRAFFIC DENSITY IN CELLULAR NETWORKS

DONGHEON LEE, SHENG ZHOU, XIAOFENG ZHONG, AND ZHISHENG NIU, TSINGHUA UNIVERSITY
XUAN ZHOU, ZHEJIANG UNIVERSITY
HONGGANG ZHANG, ZHEJIANG UNIVERSITY, UEB AND SUPÉLEC

ABSTRACT

Modeling and simulation of a cellular network typically assumes that the target area is divided into regular hexagonal cells and mobile stations (MSs) are uniformly scattered in each cell. This implies a statistically uniform distribution of traffic load over space, but in reality the spatial traffic distribution is highly non-uniform across different cells, which calls for actual spatial traffic models. In this article, we first present the analysis of traffic measurements collected from commercial cellular networks in China, and demonstrate that the spatial distribution of the traffic density (the traffic load per unit area) can be approximated by the log-normal or Weibull distribution depending on time and space. Then we propose a spatial traffic model which generates large-scale spatial traffic variations by a sum of sinusoids that captures the characteristics of log-normally distributed and spatially correlated cellular traffic. The proposed model can be directly used to generate realistic spatial traffic patterns for cellular network simulations, such as performance evaluations of network planning and load balancing.

INTRODUCTION

Due to the emergence of variety of mobile devices and their applications, the volume of mobile traffic carried by cellular networks has been growing rapidly. Cisco [1] reported that global mobile data traffic in 2012 was over twelve times greater than that in 2000 and forecasted that global mobile data traffic will increase 13-fold between 2012 and 2017, reaching 11.2 exabytes per month by 2012. In addition, mobile traffic has been diversifying from voice to multimedia, among which video traffic will account for two-thirds of the global mobile data traffic by 2017.

In order to prepare for such an exabyte mobile traffic era, network operators have been forced to search for solutions to substantially enhance network capacity with limited spectrum and energy resource. Traffic distribution over space shows considerable geographical disparity and varies hour by hour which is dependent upon people's daily activity. The resulting dynamics of network traffic bring on challenges and provide the motivation to develop a network architecture that can dynamically adapt to the variation in a cost-effective manner. Our previous work about TANGO (Traffic-Aware Network Planning and Green Operation) has suggested that network should be planned and dynamically operated according to the non-uniformly distributed nature of mobile traffic for improving energy efficiency [2]. For this purpose, analyzing and characterizing mobile traffic are indeed crucial.

Network traffic model is regarded as a large-scale traffic model which represents spatial-temporal variations of the aggregate traffic load over a large area gathered from base stations (BSs) or base station controllers (BSCs) [3, 4]. Knowledge about large scale space-time dynamics of network traffic will provide an opportunity to yield substantial improvements in performance, and this is specifically valuable from the network provider's point of view.

For example, network planning, which significantly affects installation (Capex) and operational (Opex) expenses of cellular networks, begins with grasping the spatial distribution of traffic demand. Cell breathing [5] allows the dynamic adjustment of cell coverage for enhancing network capacity where traffic loads are unevenly distributed over different cells.

In this article, we concentrate upon analyzing spatial traffic distributions from traffic measurements in commercial cellular networks and aim at providing a spatial model of network traffic. The spatial distribution of the cell traffic, which is the aggregate traffic load actually served by a BS within a specific time interval, has been studied in the literature for 2G networks [6, 7] and 3G networks [4]. Authors in [6] found that voice traffic in different cells of GSM networks can be described by a log-normal distribution. Authors in [7] found that data traffic loads in different cells of GPRS/EDGE networks can be approximated by log-normal mixtures. However, traffic distribution in different cells does not indicate the real spatial traffic distribution, because the traffic data is gathered from BSs with different coverage areas. Reference [4] provides comprehensive analysis works on traffic dynamics of 3G data networks, and regarding spatial traffic distribution, the results also showed that traffic distributions in different cells are highly uneven. Majority of prior studies focused on analyzing statistics and characteristics of cellular traffic but did not provide any mathematical model which can be used to simulate and to evaluate spatial traffic variations.

In order to understand the real spatial distribution of traffic demand, the traffic density which is defined as traffic demand per unit area should be considered instead of the cell traffic. Our previous work [8] proposed a spatial modeling of Scalable, Spatially correlated, and Log-normally distributed Traffic (SSLT), which captures inhomogeneous nature of spatial traffic distributions in cellular networks. The model is capable of generating diverse spatial patterns of random traffic demand in a target area by controlling its parameters. On the other hand, in order to validate our spatial traffic model, we collected traffic data from the EDGE/GPRS networks installed at one of the major provinces in China with 4 million subscribers. We found that the spatial distribution of the traffic density is highly skewed and the log-normal or Weibull distribution can be used to approximate it. In order to fit it accurately, the mixture distribution such as log-normal mixtures is required. The cell traffic, the traffic density, and the spatial correlation are evaluated, and especially for representing the spatial correlation, the measure of coherence distance is newly introduced. The measured statistics of traffic density are reflected in the spatial modeling of SSLT.

To the best of our knowledge, our work is the first trial to present a systematic representation of spatial pattern of network traffic based on statistical analysis of real measurements. Also, we have provided the required statistical parameters in the model. It is expected that the model can be utilized for realistic simulations of cellular networks.

The Measurements of Spatial Traffic Distribution

We collected traffic records including voice and data traffic from EDGE/GPRS networks during November 15 to December 3, 2012. The target area of the measurement is an area of 160 × 180 km which includes all types of areas (urban, rural, and etc.). There are about 5763 cells (21987 sectors) in the target area. Cell types include macro-cells as well as small cells, such as micro- and pico-cells, which are deployed inside or outside buildings and their coverage diameters vary from a few dozen meters (m) to a few kilometers (km).

We obtained the data of voice traffic measured every five minutes and that of data traffic measured every hour. The cell traffic is defined by the aggregated traffic load of all the sectors in the same cell during a certain time interval. Thus the cell traffic of data traffic is the aggregated traffic volume that each BS actually transmitted during a one-hour interval and measured in the unit of bytes. Since the voice traffic measurements show similar tendency to data traffic, we would merely present data traffic measurements in this article.

Exactly modeling the spatial distribution of real traffic demand, which also varies in time domain, requires massive raw data such as user locations and traffic volume of each device at every moment, which is impracticable to obtain from the commercial cellular networks. Instead we consider the traffic density, which can be easily calculated by using the information from BSs. It is defined as the cell traffic of a BS divided by the coverage area of the BS. As the actual area of cell coverage is difficult to measure, we obtained the area of Voronoi cells [9] drawn by using the location of BSs. It is noteworthy that the traffic density (byte/km^2) can be changed to the density of data rate requirements (b/s/km^2): through dividing the traffic density by the time of one hour (the measurement interval of data traffic is one hour).

It is not our purpose to insist on the accuracy of the obtained traffic density for representing the spatial distribution of user traffic demand. Because of capacity limits for a BS, the actual traffic demand of users may exceed the measured figures of traffic volume. The area obtained from Voronoi cells does not exactly correspond with the actual coverage of BSs. However, the traffic density is a simple approximation for representing the intensity of user traffic demand in a unit area as well as the spatial difference of traffic demand across different cells of diverse sizes.

It is not our purpose to insist on the accuracy of the obtained traffic density for representing the spatial distribution of user traffic demand. Because of capacity limits for a BS, the actual traffic demand of users may exceed the measured figures of traffic volume.

The Spatial Distribution of the Cell Traffic

Figure 1a shows the downlink (DL) cell traffic distribution of 5763 cells in the target area at 9 pm of a week day when the use of data traffic is exhibited relatively high throughout the day. Since the histogram of the empirical data shows a highly right-skewed distribution: the measured values of skewness and kurtosis are 2.99 and 18.99 respectively, the distribution of log-transformed data is depicted. We found that the empirical distribution can be accurately fitted by Gaussian mixtures (i.e., the cell traffic distribution can be modeled by log-normal mixtures) where their means and variances are obtained by the Expectation Maximization (EM) algorithm with either 2 or 3 components in most cases, which is consistent with the results in [7].

In addition, the results of parametric fitting using maximum likelihood parameter estimates of the log-normal, gamma, and Weibull distributions, which are commonly used distributions to describe skewed empirical data, are also shown in Fig. 1a. An "eyeball" comparison indicates that the gamma and Weibull distributions show a close fit of the distribution of empirical data, but the log-normal distribution does not. The Kolmogorov-Smirnov test (K-S test) is used to check for goodness-of-fit of empirical data to test distributions. We test the distribution fitting of the cell traffic every hour for 12 days at the 5 percent significance level and found that the gamma and Weibull distributions are accepted by 14.3 percent and 34.7 percent of the total

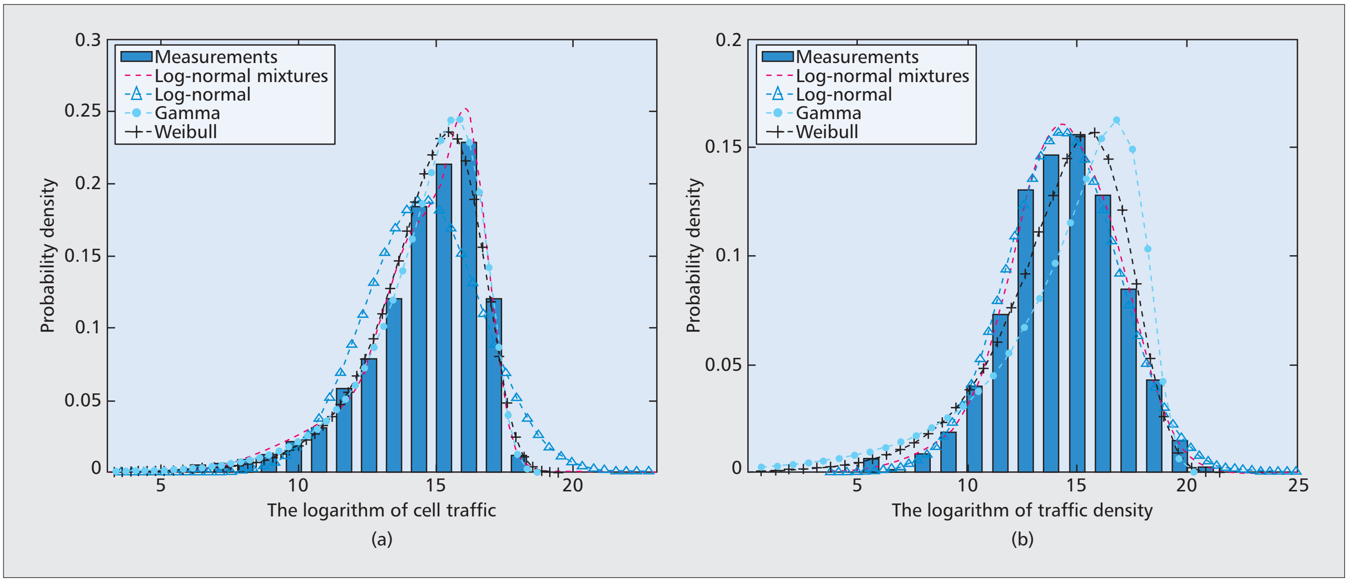

Figure 1. *The probability density function (PDF) of the cell traffic and the traffic density with distribution fittings: a) 5773 cells of whole target area; b) 2727 cells of a metropolitan area.*

measurement time respectively, while the distribution of log-normal mixtures is accepted all the times. Therefore, the findings from above measurements show that the spatial distribution of the cell traffic can be approximated by the gamma or Weibull distributions and accurately modeled by the mixture distribution such as log-normal mixtures in this article.

The Spatial Distribution of the Traffic Density

The traffic density is analyzed based on a grid basis. First, the cell traffic of each cell is divided by the corresponding Voronoi cell area to get the traffic density. Then the target area is divided into a square grid and each pixel (square) in the grid is assigned to the traffic density of its nearest BS (cell), and thus a matrix of traffic density can be obtained. All pixels within the same Voronoi cell have the same value, that is, we assume traffic demand within a cell is uniformly distributed in the measurement. Hence the accuracy of this approximation is high when the cell size is small.

A metropolitan area of 40 × 40 km, which includes a large city (population of more than 8 million) as well as surrounding suburban and rural areas, is selected for measuring the traffic density. The distribution of the traffic density is strongly positively skewed, so we also try to fit the empirical data with log-normal mixtures, log-normal, gamma, and Weibull distributions. The measured values of skewness and kurtosis of the empirical data are 19.95 and 712.52, respectively. Figure 1b shows the traffic density distribution of all the cells in the metropolitan area at 9 pm. It is shown that the distribution can be approximated by either log-normal mixtures or the log-normal distribution, but fitting with the gamma or the Weibull distribution shows poor performance.

The results of the K-S test show that all the distributions are rejected. This is because empirical data of the traffic density in this article are approximations because of the previously mentioned assumptions (i.e., the uniformly distributed traffic density within a cell and the disagreement of Voronoi cell areas and actual coverage areas), so empirical data does not exhibit complete statistics. However, the K-S statistics (the maximum distance between the cumulative distribution function (CDF) of empirical data and the reference distribution) exhibits low values, where the K-S statistics for log-normal mixtures and the log-normal distribution are 0.0193 and 0.0382, respectively.

We also checked traffic densities of some specific urban and rural areas extracted from the whole target area. Here we take two different dense urban areas, "urban area1" and "urban area 2," where cell density (the number of cells per square kilometer) is 28.6 and 42 respectively, as examples. The snapshot of the traffic density of "urban area1" at 9 pm is visualized in Fig. 2a. The side length of the square pixel for urban areas in Fig. 2a is set to 3/300 km. The distribution of a specific area also exhibits highly skewed distribution as the measured values of skewness and kurtosis of the empirical data are 6.18 and 70.84, respectively.

Figure 2b shows the CDFs of the traffic density of the two ares. Log-normal mixtures, the log-normal, Weibull, and gamma distributions are examined by the K-S test, but it is found that all of them are rejected in the same manner like the metropolitan area case. In terms of the K-S statistic, the distribution of log-normal mixtures shows the most accurate fit for the empirical distribution of both areas. The gamma distribution shows relatively bad fits of the empirical data in comparison with the others for almost all the place and time. Except log-normal mixtures, the Weibull distribution shows better approximation performance in case of "urban area 1," while the log-normal distribution is better in case of "urban area 2."

In fact, both the log-normal and Weibull dis-

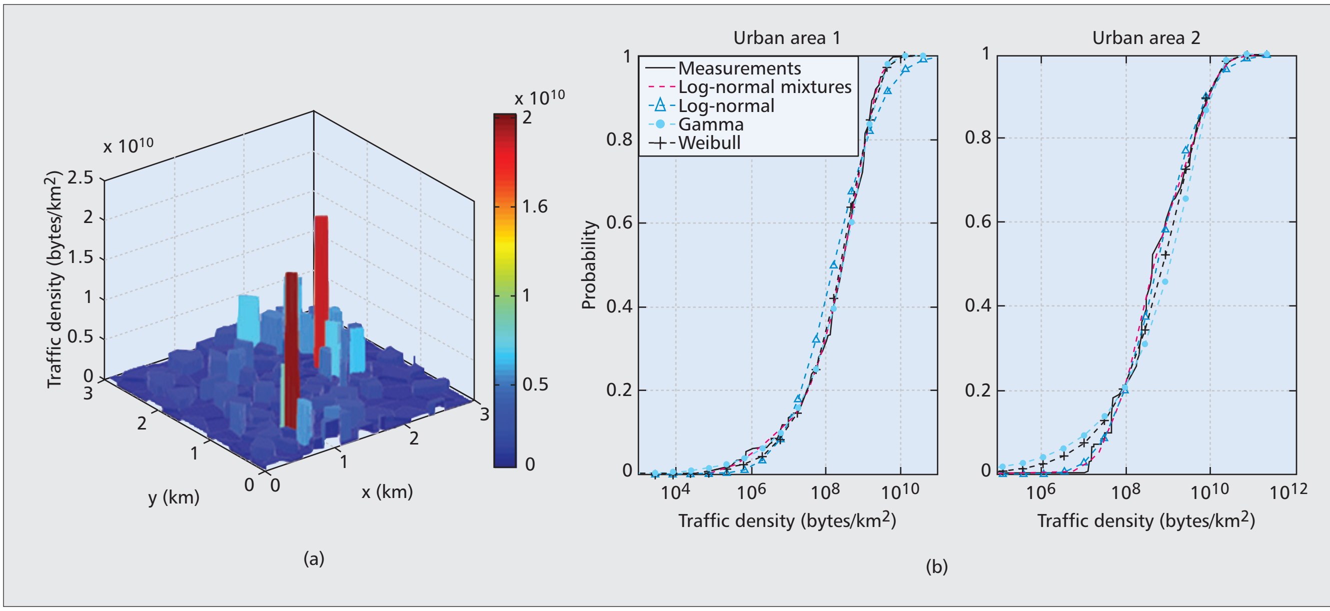

Figure 2. *Three-dimensional view of the DL traffic density and its distribution: (a) the traffic density of "urban area 1"; b) the cdf of traffic density measurements and distribution fitting of "urban area 1" (left) and "urban area 2" (right).*

tributions have been used together in many fields to describe the skewed empirical data. For example, in reliability engineering, they are alternatively used for measuring and estimating the product life [10]. From the results in Fig. 1b and Fig. 2b, we found that the log-normal and Weibull distributions can be selectively used to approximate the spatial distribution for a certain region at a certain time. Since the distribution of the traffic density varies across both space and time, one distribution can only provide an approximation of a small area at a specific time, where its distribution can be regarded as stationary. Through a large number of measurements from different areas and times, we conclude that *the spatial distribution of the traffic density in cellular networks can be approximated by the log-normal or the Weibull distribution, which is subject to change with time and place, and accurately described by the mixture distribution such as log-normal mixtures.*

In this article, the log-normal distribution is selected for spatial modeling of the traffic density which will be described later. It is necessary to obtain statistics of the traffic density for parameter settings of modeling. If Z is a random variable with a standard normal distribution, then $X = \exp(\sigma Z + \mu)$ has a log-normal distribution with the mean $M = \exp(\mu+\sigma^2/2)$ and variance $\$V = (\exp(\sigma^2) - 1)\exp(2\mu+\sigma^2)$, where the location parameter μ and scale parameter σ is the mean and variance of the variable's logarithm. Parameters of the distributions show diversity depending on time and area, but the distributions from almost all regions can be approximated by the log-normal or Weibull distribution. Hence, we just select the parameters of a typical urban and rural area, which are summarized in Table 1. They are obtained by parametric fitting with maximum likelihood estimates of log-normal mixtures and the log-normal distribution. The urban area in the table is the "urban area 1" which is the core of the city including a rail way station, high buildings, apartments, and etc. The rural area in the table is a typical agricultural area of 20 × 20 km where the number of cells per square kilometer is 0.19.

Modeling Spatial Correlations of Traffic Density

The most fundamental feature for characterizing the spatial distribution of the traffic density is the highly skewed distribution such as the log-normal distribution with the parameters specified in Table 1. Additionally, another kind of measure is required to describe spatial patterns, like smoothness or spatial fluctuations, of the traffic density. Hence, we newly define the measure of the coherence distance which can evaluate the correlation between the traffic density of adjoining regions. The coherence distance is defined as the distance that the two-dimensional autocorrelation function (ACF) of traffic density drops to the half of its peak value.

The two-dimensional sample ACF of the traffic density of Fig. 2a is depicted in Fig. 3a, and its enlarged figure of the cross-section is shown in Fig. 3b. The value of autocorrelation is normalized by its maximum value. Since the shape of the two-dimensional sample ACF is not symmetrical, we obtained the coherence distance by averaging the distances between the origin and the points that have a autocorrelation value of $0.5\pm\varepsilon$, where ε is set to 0.01. In Fig. 3b, the measured average coherence distance of the urban area is 71.4m, which means that traffic demands of two points separated by more than 71.4m have the correlation lower than 0.5. The coherence distance of the selected rural area specified in Table 1 is much larger than that of the urban area and spatial patterns of rural areas normally show slower fluctuations.

Since coherence distance is affected by target area, grid size, and cell area, the provided values

If one only has a purpose of research on the algorithm of network planning, the procedure of traffic estimation is not necessary. In that case, the spatial traffic model which can generate virtual spatial traffic patterns must be the fundamental requirement for examining their algorithm in simulations.

Statistics and parameters			Rural		Urban	
			UL	DL	UL	DL
A lognormal distribution	Location(μ)		11.573	12.572	17.7956	18.93
	Scale(σ)		2.3055	2.7985	2.1188	2.3991
Lognormal mixtures	Location	μ_1	12.2822	5.2920	17.6345	20.9990
		μ_2	9.2172	11.3766	19.4508	15.3133
		μ_3	13.9874	14.0221	15.1012	19.3297
	Scale	σ_1	0.2368	0.0283	1.2168	0.5081
		σ_2	4.3690	4.4479	0.7933	5.0516
		σ_3	0.3911	1.6622	3.2990	1.9255
	Mixture proportions	p_1	0.4081	0.0688	0.3287	0.2212
		p_2	0.3601	0.3210	0.4280	0.1915
		p_3	0.2318	0.6102	0.2433	0.5874
Spatial correlation	Coherence distance (m)		1075.7	1075.4	80	71.4
	The maximum spatial spread (ω_{max})		0.001202	0.001163	0.012673	0.011592

Table 1. *Evaluation of fitted log-normal distributions and spatial correlation.*

may not be consistent if settings of the measurement are different. Therefore, we do not claim the accuracy of the values of coherence distance but the necessity to reflect a measure of the spatial correlation for spatial traffic modeling, which was also stressed in [11]. In the following section, we present our spatial modeling scheme of the traffic density with the coherence distance and a log-normal distribution with the estimated mean and variance.

Spatial Modeling of the Traffic Density

From our measurements, we found that the spatial distribution of the traffic density is spatially correlated and can be approximated by the log-normal distribution. In our previous work [8], we proposed spatial modeling of SSLT which is able to generate large-scale spatial variations in the traffic density. It is a sum-of-sinusoids statistical model with introducing randomness into the variables in the model. A random traffic density map can be generated by the model and the statistics of the map can be adjusted for various scenarios by controlling the parameters.

Our model is built on a grid-based plane like the same manner as the measurement of traffic density in the previous sections. The target region A is divided into $M \times N$ square pixels and the pixel size is set to be the same as the one in the measurement. A pixel $g_{m,n}$, where $m = 1, \ldots, M$ and $n = 1, \ldots, N$, contains the traffic density $\rho_{m,n}$ (bytes/km^2). Let $\rho = (\rho_{m,n})_{m=1,\ldots,M;n=1,\ldots,N}$ denote the traffic density matrix which represents the random traffic density map.

Let $x_{m,n}$ and $y_{m,n}$ be the two-dimensional Cartesian coordinates of the center of the pixel $g_{m,n}$ in unit of meters. To generate a log-normally distributed traffic density map, a Gaussian random field, $\rho^{(G)} = (\rho^{(G)}_{m,n})_{m=1,\ldots,M;n=1,\ldots,N}$, is first generated by

$$\rho^{(G)}_{m,n} = \frac{2}{\sqrt{L}} \sum_{l=1}^{L} \cos(i_l x_{m,n} + \phi_l) \cos(j_l y_{m,n} + \psi_l) \quad (1)$$

where angular frequencies i_l and j_l are uniform random variables between 0 and ω_{max} and phases φ_l and ψ_l are uniform random variables between 0 and 2π. For a large enough L, $\rho^{(G)}_{m,n}$ can be approximated as standard Gaussian random variables according to the central limit theorem. We found that the value $L = 10$ should be large enough. We define ω_{max} as the maximum spatial spread which decides the rate of fluctuations of the random field.

Finally, by taking the exponential function of $\rho^{(G)}_{m,n}$ with the location parameter μ and the scaling parameter σ, we obtain the traffic density matrix whose elements are log-normally distributed as follows:

$$\rho_{m,n} = \exp(\sigma \rho^{(G)}_{m,n} + \mu), \quad (2)$$

where $m = 1, \ldots, M$ and $m = 1, \ldots, N$. By controlling μ and σ, the log-normal distributed random values ρ are scaled to fit the statistics of

[6] *Each BS may be equipped with multiple transmit antennas and each transmit antenna can be recognized as a transmit point.*

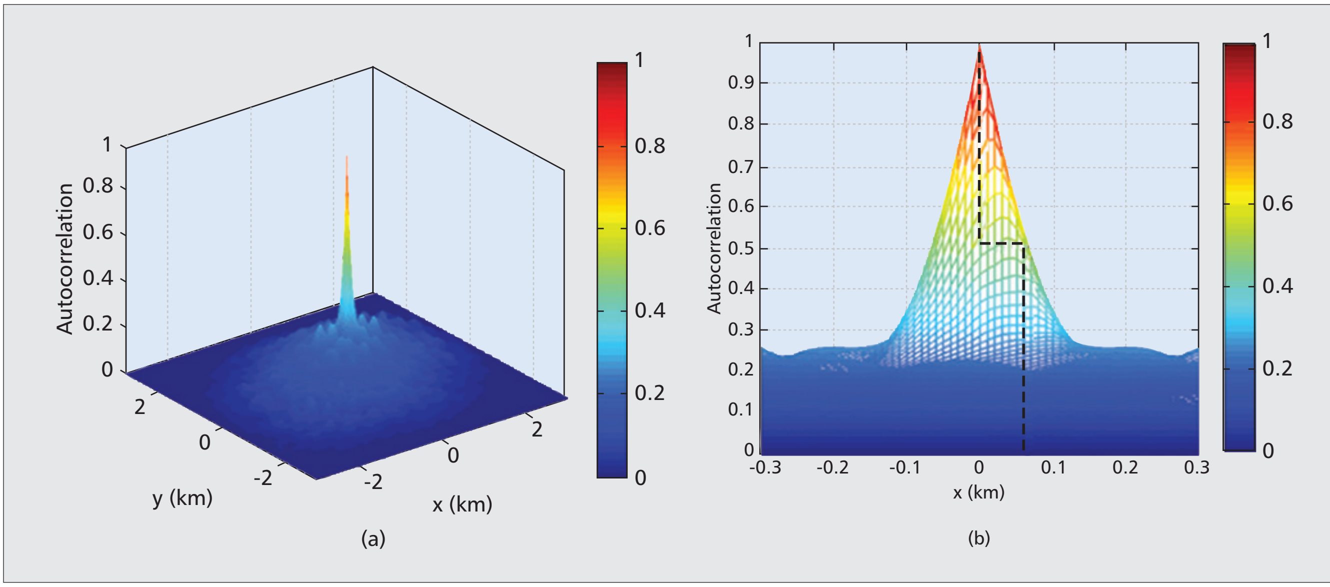

Figure 3. *The normalized two-dimensional sample autocorrelation function (ACF) of the traffic density shown in Fig. 2a: a) two dimensional visualization; b) the cross section of (a) along y axis.*

traffic density for specific regions, e.g., an urban or a rural area. The maximum spatial spread ω_{max} affects the two-dimensional ACF of Eq. 2 and the resulting coherence distance. The correlations between traffic densities of adjacent regions (or coherence distance) get smaller as ω_{max} gets larger, which means that the generated random traffic density map will be more fluctuated.

The sample of the traffic density of downlink generated by our model for an urban area is presented in Fig. 4a. Default parameter values of log-normal distributions in Table 1 are used for the generation. The maximum spatial spread ω_{max} in Table 1 are numerically obtained matching the coherence distance of a sample generated by Eq. 2 and that of measurements in Table 1. The cross section of its sample ACF is depicted in Fig. 4b. The measured coherence distance is 81.9m, which is similar to the measured value in Table 1. Note that the coherence distance of a traffic density map generated by Eq. 2 should be checked whether it is out of an error range, which can be set by one's requirements (30m for the urban area and 100m for the rural area are used in our simulations), comparing with the coherence distance in Table 1. That is because the model generates a random map of the traffic density every time and may occasionally produce a much different coherence distance.

The proposed model shapes the traffic density with only a log-normal distribution with scaling parameters μ and σ given in Table 1. However, as shown in Fig. 2b, the traffic density for a specific region can be accurately approximated by lognormal mixtures. For modeling by log-normal mixtures with three components, scaling parameters μ_i and σ_i (instead of assigning same values of μ and σ for all the pixels) can be stochastically selected for each pixel among three components according to the proportions p_i (the probabilities of three components) in Table 1, where $i = 1, 2, 3$. However, in that case, as uncorrelated values can be assigned to adjacent grids, spatial correlations may not be easily handled to fit the coherence distance. This issue is left for the future study.

Applications of Spatial Traffic Model

Our proposed model lays a foundation for the analysis and simulations of cellular networks. We provide the application of the proposed model on network planning, resource management, and performance analysis.

Traffic Demand Generation for Network Planning

Network planning requires the real spatial distribution of traffic demand in a target area. In [3], authors used geographical and demographical characteristics of the service area to estimate traffic demand, and generated traffic intensity matrix with demand nodes for network planning. The demand node represents a certain amount of traffic demand per unit area, which has been widely applied to generate traffic demand as the input of network planning algorithms. However, if one only has a purpose of research on the algorithm of network planning, the procedure of traffic estimation is not necessary. In that case, the spatial traffic model which can generate virtual spatial traffic patterns must be the fundamental requirement for examining their algorithm in simulations.

The demand nodes can be generated by our model. In Fig. 5, we give an example of the generation of traffic demand nodes by using our model with rural DL parameters in Table 1. The pixel color indicates the magnitude of the traffic density: the red color represents high density and the blue color represents low density. The basic generation method can be summarized as follows. Once we get the traffic density matrix ρ from our model, demand nodes can be randomly dropped in each pixel. Let a (km^2) denote the area of the pixel, where $a = (3/300)^2$ and $a =$

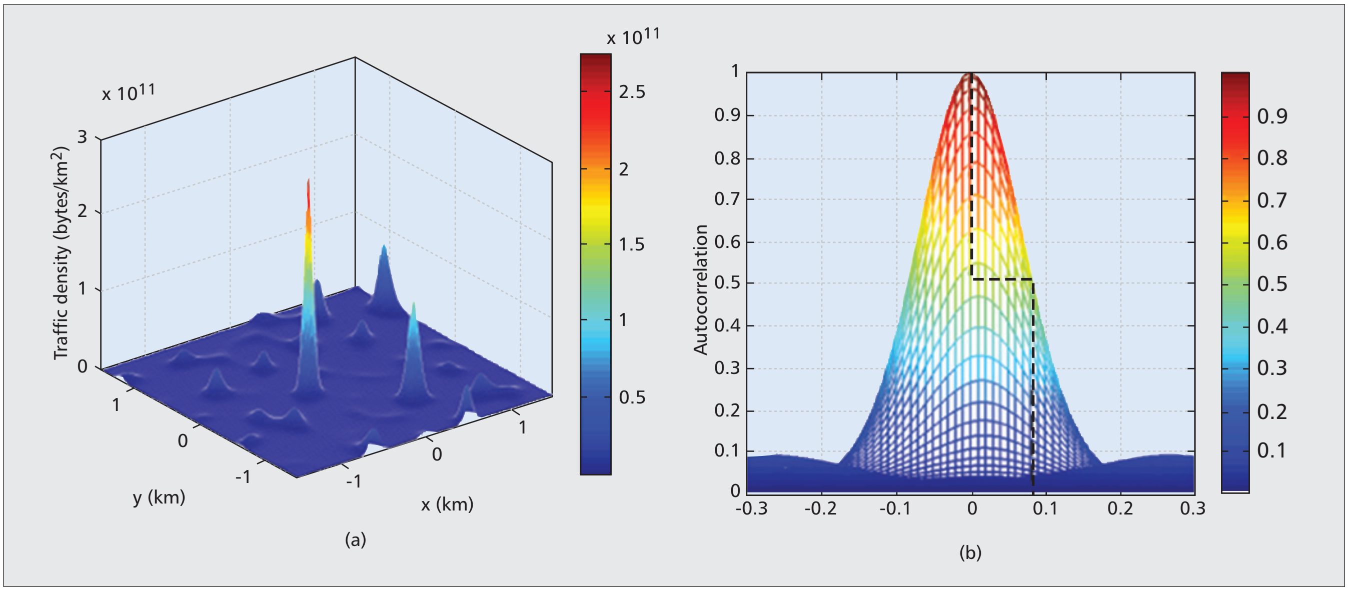

Figure 4. *Modeling examples: a) a sample map of the traffic density for an urban area; b) the cross section of the normalized sample ACF of (a).*

$(20/250)^2$ for an urban and a rural area in this article respectively. All the nodes are assumed to have the same amount of data rate requirement r (bytes/second) smaller than the minimum requirement $\rho_{min}a/3600$, where ρ_{min} (bytes/km^2) is the smallest element in ρ. Then the number of demand nodes in the pixel $g_{m,n}$ is thus $\lceil \rho_{m,n}a/(3600r) \rfloor$.

The general network planning problem is to deploy BSs minimizing the fixed and operational cost, while every demand node must be covered by at least one BS. Detailed explanations about the usage of the model in network planning can be found in [8]. Base stations positioned by the network planning algorithm in [8] are also depicted in Fig. 5. In analysis and simulation, it is important to investigate the impact of the spatial traffic distribution and spatial correlation on network planning optimization and the network performance. For example, we evaluate energy efficiency of the network with respect to the inhomogeneity of the spatial traffic distribution in [8]. Likewise, one can optimize the locations of cells and their BSs' transmission power levels under various distributions and coherence distances which can be generated by our spatial model.

Network Management

Since the spatial imbalance of traffic loads causes QoS degradations (e.g., under-utilization of resources, increase of outage probability in congestion cells), many research related to network management such as power control, channel allocation, fractional frequency reuse (FFR), and load balancing have been studied. For example, when a cell shows a much higher traffic load than neighboring cells, BSs of neighboring cells will lower their transmission power to avoid interference for the QoS of the cell. Especially in heterogeneous wireless networks, interference management is essential due to different types of coexisting BSs which will induce severer interference. Configuring system model with heterogeneous wireless networks, developing the algorithms optimized with spatial traffic distribution, and examining the algorithms with simulations all require the spatial traffic model.

We take load balancing as an example for the application of our model. One can choose a load balancing strategy to distribute the concentrated traffic load to neighboring cells by using cell size adjustment, bandwidth allocation, or user association algorithms [12]. Thus, in the same manner like network planning, research on load balancing naturally requires the non-uniform spatial traffic model, not only to optimize, but to also evaluate the algorithms.

In the simulation work in [12], the authors generate temporal and spatial variations of the user arrival rate for each cell by assuming the log-normal distribution for the spatial distribution of traffic load across cells. However, spatial correlations are not considered. Our proposed model can replace independent and identically distributed (i.i.d.) random variables (RVs) for the generation of user arrival rates of neighboring cells by spatially correlated RVs generated by Eq. 1.

Performance Analysis of Cellular Network

Stochastic geometry theory provides mathematical models and statistical methods to analyze geometrical structure of BSs and MSs so that the performance of wireless networks, such as coverage and throughput, can be analyzed [11]. For example, authors in [13] evaluated energy efficiency of Poisson-Voronoi tesselation (PVT) cellular networks considering a non-uniform spatial traffic distribution which is modeled by the Pareto distribution for the traffic intensity of MSs and the Poisson point process (PPP) [11] for the location of MSs. Instead of the Pareto distribution, the log-normal or Weibull distribution can be assumed for the non-uniform distribution of the spatial traffic intensity in order to evaluate the performance of cellular networks.

Considerations and Future Works

Spatial Randomness Measure

Some spatial randomness measures are required to objectively rate the spatial distribution of traffic demand and node configurations. In wireless networks, geometrical configurations of nodes affect overall system performance because of variations in distances between transmitters and receivers. Authors in [11] claimed that spatial randomness of node configurations is also an object that needs to be addressed and overcome like the wireless channel. To that end, the essential prerequisite is to establish both the spatial traffic model and the spatial randomness measure, which should be utilized as components of network traffic simulations.

The measure can be specified with the spatial traffic model in the assumptions of simulations like the Rayleigh-fading multipath channel model with the delay spread and the Doppler spread. The coherence distance proposed in this article can also be a type of spatial randomness measure which can evaluate the rate of spatial fluctuations. Actually, the relationship between the coherence distance and the maximum spatial spread ω_{max} in this article is similar to the relationship between the coherence time and the Doppler spread of the channel.

Our previous work [8] introduced the measure of inhomogeneity proposed in [14], which can rate the degree of inhomogeneity of the spatial distribution of nodes. If each node is assumed to have the same amount of traffic demand and the nodes are uniformly distributed, then the inhomogeneity has a value of zero. If there is only one hot spot in a target area and all the nodes are concentrated on a small area, it is close to one. However, it is ambiguous when there exist many small hot spots scattered in the target area, because the measure may have a value of near zero even if the spatial traffic pattern is non-uniform. Therefore, further research is needed to associate the inhomogeneity value with the size and number of hot spots or to define some other measure.

The Issue of Spatio-Temporal Traffic Modeling

Random spatial and temporal models of network traffic will play an important role in managing the increasing and changing mobile traffic. With respect to the enhancement of energy efficiency, our previous work [2] suggested BS sleep operation according to traffic dynamics. Switching off some under-utilized BSs in off-peak hour should consider spatial traffic distribution as well as its temporal variations. It means that switching off a BS should consider the state of neighboring cells' traffic loads during a certain time interval because neighboring BSs will need to accommodate traffic from the sleeping cell. The strategy actually exploits spatial and temporal traffic variations in order to improve system performance.

Existing works on cellular networks normally assumed Poisson arrivals and uniformly distributed MSs for every cell, but it would be more practical to use the spatio-temporal traffic model which can produce spatial inhomogeniety, temporal variations, and their spatio-temporal correlations. Complete modeling of traffic dynamics both in time and space is highly difficult because traffic dynamic is intrinsically stochastic and the combinatorial representation in space and time domain is complex.

Although this article only covers spatial correlations, research on adding a temporal correlation in each area and correlations between daily traffic variations of neighboring areas to the random spatio-temporal traffic model are needed to consider. On the other hand, in our measurements of temporal traffic dynamics, it is found that daily traffic patterns of cells are very diverse. For instance, the distribution of peak hours is diverse according to regional groups. Therefore, the spatio-temporal traffic model would reflect spatio-temporal correlations as well as geographically diverse patterns of temporal traffic variations.

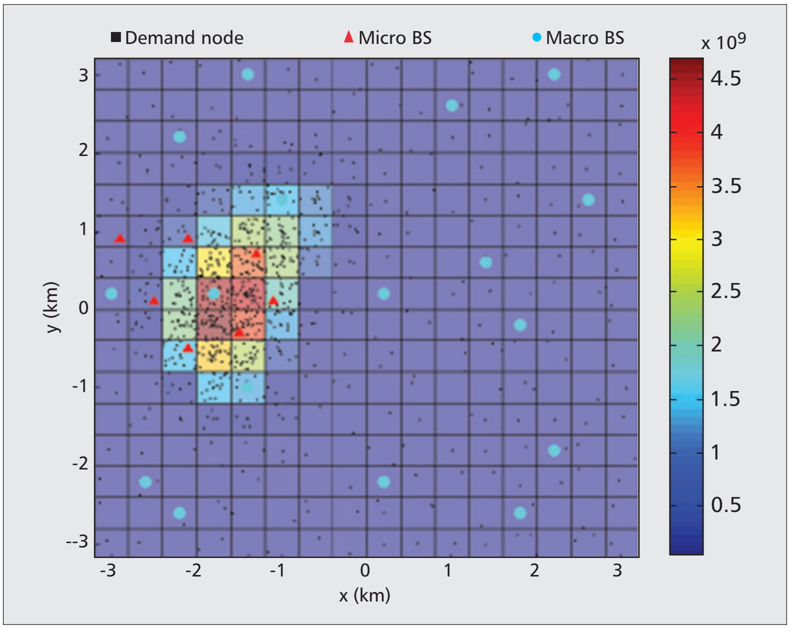

Figure 5. *Network planning based on demand nodes generated by the proposed spatial model of the traffic density.*

Applicability to the Next Generation Network

The spatial distribution of the traffic density is strongly related to human behaviors (e.g., smartphone usage, mobility) and the distribution of population over space. These factors are reflected in the distribution of the traffic density, because the spatial model in this article is based on the traffic volume that a BS actually serviced in a certain time interval; in other words, the traffic volume that MSs actually requested. Whether 3G or 4G networks are deployed, it is conjectured that the spatial traffic distribution will exhibit a highly skewed distribution which may be described by the log-normal and Weibull distribution families. While traffic analysis performed herein investigates GPRS/EDGE networks, much work is needed on confirmation of

We provide a realistic spatial model of the traffic density considering the log-normal distribution and spatial correlations. The proposed model is expected to have wide applications in the field of cellular network simulation.

the applicability of the spatial model in other 3G and 4G cellular networks.

CONCLUSION

Our measurement on commercial cellular networks found that the cell traffic can be approximated by the Weibull or gamma distribution; the traffic density, which is regarded as real traffic demand of people, can be approximated by the log-normal and Weibull distribution. The mixture distribution such as log-normal mixtures is required to accurately describe them because the distributions are non-stationary over space and time. We also analyze spatial traffic correlations by introducing the new measure of the coherence distance. The findings suggest that cellular networks should be designed considering such a highly skewed and spatially correlated distribution of cellular traffic so that resources are more efficiently utilized. Generally, ordinary researchers cannot access raw traffic data of cellular networks, and therefore research about cellular networks have been often hindered from lack of practical spatial traffic models. We provide a realistic spatial model of the traffic density considering the log-normal distribution and spatial correlations. The proposed model is expected to have wide applications in the field of cellular network simulation.

ACKNOWLEDGMENT

This work is sponsored in part by the National Basic Research Program of China (2012CB316000), the Nature Science Foundation of China (61201191, 60925002, 61021001), and Hitachi Ltd.

REFERENCES

[1] Cisco, "Cisco Visual Networking Index: Global Mobile Data Traffic Forecast Update, 2012–2017," White Paper, Feb. 2013.

[2] Z. Niu, "TANGO: Traffic-Aware Network Planning and Green Operation," *IEEE Wireless Commun.*, vol. 18, Oct. 2011, pp. 25–29.

[3] K. Tutschku and P. Tran-Gia, "Spatial Traffic Estimation and Characterization for Mobile Communication Network Design," *IEEE JSAC*, vol. 16, no. 5, June 1998, pp. 804–11.

[4] U. Paul *et al.*, "Understanding Traffic Dynamics in Cellular Data Networks," *Proc. IEEE INFOCOM*, Apr. 2011.

[5] S. V. Hanly, "An Algorithm for Combined Cell-Site Selection and Power Control to Maximize Cellular Spread Spectrum Capacity," *IEEE JSAC*, vol. 13, no. 7, Sept. 1995, pp. 1332–40.

[6] U. Gotzner, A. Gamst, and R. Rathgeber, "Spatial Traffic Distribution in Cellular Networks," *Proc. IEEE VTC*, May 1998.

[7] M. Michalopoulou, J. Riihijarvi, and P. Mahonen, "Towards Characterizing Primary Usage in Cellular Networks: A Traffic-Based Study," *Proc. IEEE DySPAN*, May 2011.

[8] D. Lee, S. Zhou, and Z. Niu, "Spatial Modeling of Scalable Spatially-Correlated Log-Normal Distributed Traffic Inhomogeneity and Energy-Efficient Network Planning," *Proc. IEEE WCNC*, Apr. 2013.

[9] J. Illian *et al.*, *Statistical Analysis and Modelling of Spatial Point Patterns*, John Wiley & Sons, 2008.

[10] J. F. Lawless, *Statistical Models and Methods for Lifetime Data*, New York: John Wiley & Sons, 1982.

[11] J. G. Andrews *et al.*, "A Primer on Spatial Modeling and Analysis in Wireless Networks," *IEEE Commun. Mag.*, Nov. 2010, pp. 156–63.

[12] Z. Yang and Z. Niu, "Load Balancing by Dynamic Base Station Relay Station Associations in Cellular Netwroks," *IEEE Wireless Commun. Lett.*, vol. 2, no. 2, Apr. 2013, pp. 155–58.

[13] L. Xiang *et al.*, "Energy Efficiency Evaluation of Cellular Networks Based on Spatial Distributions of Traffic Load and Power Consumption," *IEEE Trans. Wireless Commun.*, vol. 12, no. 3, Mar. 2013, pp. 961–73.

[14] U. Schilcher *et al.*, "Measuring Inhomogeneity in Spatial Distributions," *Proc. IEEE VTC*, May 2008.

BIOGRAPHIES

DONGHEON LEE (lidx10@mails.tsinghua.edu.cn) received his B.Sc. and M.Sc. in Electrical Engineering from Pusan National University, Busan, Korea, in 2007 and 2009, respectively. He is currently a Ph.D. student in the department of Electronic Engineering, Tsinghua University, Beijing, China. His research interests include performance evaluation of wireless cellular networks, teletraffic engineering, and green wireless communications.

SHENG ZHOU (sheng.zhou@tsinghua.edu.cn) received his B.S. and Ph.D. degrees in Electronic Engineering from Tsinghua University, China, in 2005 and 2011, respectively. He is currently an assistant professor of Electronic Engineering Department, Tsinghua University. From January to June 2010, he was a visiting student at Wireless System Lab, Electrical Engineering Department, Stanford University, CA, USA. His research interests include cross-layer design for multiple antenna systems, cooperative transmission in cellular systems, and green wireless cellular communications.

XUAN ZHOU (zhouxuan@zju.edu.cn) received his master degree from Beijing University of Post and Telecommunication, and he is now a Ph.D. candidate of Zhejiang University, Hangzhou, China. Meanwhile, he works as a senior system engineer in China Mobile. His research efforts focused on network analysis and management, human behavior modeling and network data mining.

XIAOFENG ZHONG (zhongxf@tsinghua.edu.cn) received his Ph.D. degree in Information and Communication System from Tsinghua University, Beijing, China, in 2005. He has been an Assistant Professor in the Department of Electronic Engineering of Tsinghua University. He researches in the field of mobile network, including the users' behaviors and traffic model analysis, MAC and Network protocol design, and resource management optimization. He has published more than 30 papers and owns 7 patents.

HONGGANG ZHANG (honggangzhang@zju.edu.cn) is an International Chair Professor, Université Européenne de Bretagne (UEB) & Supélec, France, a Full Professor of Zhejiang University, China, and an Honorary Visiting Professor of the University of York, UK. He received the Ph.D. degree in Electrical Engineering from Kagoshima University, Japan, in March 1999. He has served as the Chair of the IEEE ComSoc TCCN during 2011–`2012. He was the Co-Chairs of IEEE Globecom 2008 Symposium and IEEE ICC 2013 Symposium. He has been the Lead Guest Editor of the IEEE Communications Magazine feature topic issues on "Green Communications." He is the co-editor/co-author of two books with the titles of "Cognitive Communications" (Wiley) and "Green Communications: Theoretical Fundamentals, Algorithms and Applications" (CRC), respectively.

ZHISHENG NIU [F] (niuzhs@tsinghua.edu.cn) graduated from Beijing Jiaotong University, China, in 1985, and got his M.E. and D.E. degrees from Toyohashi University of Technology, Japan, in 1989 and 1992, respectively. After spending two years at Fujitsu Laboratories Ltd., Kawasaki, Japan, he joined Tsinghua University, Beijing, China, in 1994, where he is now a professor in the Department of Electronic Engineering and deputy dean of the School of Information Science and Technology. His current research interests include teletraffic theory, mobile Internet, radio resource management of wireless networks, and green communications.