

Temporal-spatial distribution nature of traffic and base stations in cellular networks

ISSN 1751-8628
 Received on 25th March 2017
 Revised 17th July 2017
 Accepted on 10th August 2017
 doi: 10.1049/iet-com.2017.0330
 www.ietdl.org

Zhifeng Zhao¹, Meng Li¹, Rongpeng Li¹ ✉, Yifan Zhou¹

¹College of Information Science and Electronic Engineering, Zhejiang University, Zheda Road 38, Hangzhou, People's Republic of China
 ✉ E-mail: lirongpeng@zju.edu.cn

Abstract: Recent years have witnessed the unprecedented surge of mobile traffic and base stations (BSs) deployment, which poses severe requirement for future communications systems. Understanding the distribution dynamics of traffic and BSs in time-space domain is of vital importance for better network design and resource management in cellular networks. In this study, a study on the statistical characteristics of cellular traffic series is carried out and α -stable distribution is verified to be valid for modelling the traffic series of each BS. On the other hand, inspired by the fact that BSs traffic series are spatially correlated, the authors study the statistical relationship between the correlation coefficient and the distance between BSs. Moreover, α -stable model is also suitable to describe the BSs deployment, thus conducing to prove the existence of self-similarity. In addition, both the traffic time series and the BSs spatial distribution are deeply associated with heterogeneity, so they come up with the density-based and distance-based methods to quantify their heterogeneous degree.

1 Introduction

The explosive growth of traffic [1] and numerous base stations (BSs) [2] bring severe challenges to current and future cellular networks. Thus it is imperative to fully understand the spatial-temporal characteristics of traffic and the spatial distributed patterns of BSs.

Prior research works have already presented that a good knowledge of the temporal-spatial traffic dynamics in cellular networks can conduce to analyse the potential effect that network resource usage pattern and subscriber behaviour have on the network protocol design, spectrum allocation and energy savings [3–5]. In this respect, understanding how cellular traffic is distributed temporally and spatially, could be an extremely beneficial step.

In the past several years, there have been a number of works that focus on traffic self-similarity nature in Ethernet networks [6] and *ad-hoc* networks [7]. Specifically, some traffic models based on the self-similar stochastic processes such as linear fractional stable noise process [8], heavy-tailed ON/OFF process [7] and so on, provided reasonable descriptions of the traffic scenarios. However, these traffic models seldom take human behaviour patterns into consideration [9]. Equally important as in wired core networks, statistical traffic model should also be studied in cellular networks. From the spatial domain, the authors of [10, 11] introduced the geospatial dynamics of application usage and the spatial characteristics of network resource usage, respectively, whereas the authors of [12, 13], with a statistical significance, studied the spatial distribution of cellular traffic by using log-normal distribution. However, temporal-spatial natures of traffic in cellular networks seldom are researched systematically, and detailed analysis on the spatial correlation of traffic usage at BSs is still missing.

For BSs spatial distribution, Poisson point process (PPP) is widely employed in characterising the stochastic deployment patterns of BSs [14]. Afterwards, Geyer saturation process [15], Poisson cluster process [16], and two-tier PPP [17] have been proposed to reflect the clustering property of BSs which keeps pace with the social clustering behaviours. Although the aforementioned models could provide mathematical tractability in the networking performance evaluation, they failed to reveal the intrinsic heavy-tailed feature of BSs deployment under the influence of traffic spatial distribution [3]. Moreover, the ever-increasing deployment of dense small cells and the multi-tier networking heterogeneity

cause the network topology much more complicated than before [18]. Therefore, it is necessary to find suitable mathematical approach to characterise the non-uniformity of the deployed BSs.

In view of the above problems, we conduct comprehensive analysis on the temporal-spatial distribution characteristics of traffic and BSs. Concretely, our work has made contributions in the following three aspects:

- Firstly, we study typical traffic characteristics such as periodicity, burstiness and self-similarity and propose to take advantage of α -stable distribution to model the traffic distribution in time domain. We also examine the spatial correlation of cellular network traffic between BSs.
- Secondly, we reconfirm that heavy-tailed phenomenon does exist in the BSs deployment, and BSs deployment can be characterised by α -stable model, which leads to the self-similarity of BSs spatial distribution.
- Thirdly, we put forward a density-based measure and a distance-based measure to evaluate the inhomogeneous degree of BSs spatial distribution and traffic temporal distribution.

The remaining of this paper is organised as following. In Section 2, we present a brief description on the dataset and the relevant mathematical background. Section 3 focuses on the temporal-spatial analysis of traffic in cellular networks. Then detailed spatial study of BSs deployment is made in Section 4. In Section 5, the heterogeneity of BSs deployment and traffic temporal distribution is investigated. In the end, we summarise the paper in Section 6.

2 Background

2.1 Dataset description

The measurement data used in this paper are obtained from cellular networks of an eastern China city, including traffic and commercial BSs information. In particular, our traffic datasets are based on a significant number of practical traffic records from China Mobile via the Gb interface of 2G/3G cellular networks or S1 interface of 4G cellular networks, while our geographical datasets are collected from the original engineering database of the operator. The downlink traffic is measured in the unit of bytes that each BS transmits to the covered users in one-hour interval for continuous 7 days, from 5573 cellular BSs. Therefore, the traffic record of each

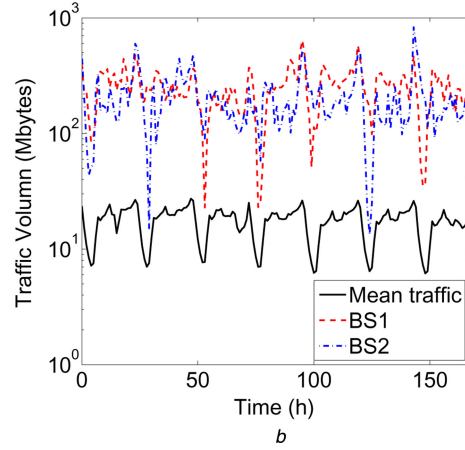
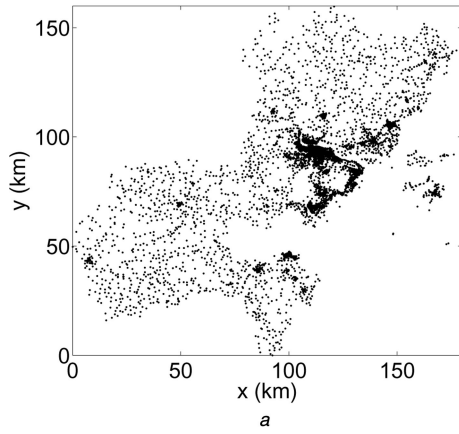


Fig. 1 The illustration of BSs and traffic

(a) All BS locations (black dots) of the eastern China city, (b) Traffic time series of all BSs, BS1 and BS2 during 7 days

Table 1 Dataset information

Traffic information		BSs information	
Traffic resolution	Duration	No. of BSs	Location information (longitude and latitude)
1 h	1 week	5573	yes

BS can be regarded as a time series. In addition, the geographic location (longitude and latitude) of each BS is available. Table 1 lists the corresponding details. For the purpose of confidentiality, we convert the longitude–latitude information of each BS to X , Y coordinates in this paper, which does not change the relative spatial positions of BSs. Fig. 1a draws the X – Y locations of all BSs on the two-dimensional (2D) plane.

2.2 Mathematical background

α -Stable model with the property of burstiness, long-range dependence (LRD) and heavy-tailed distribution, manifests itself in the capability to characterise the distribution of the normalised sums of a relatively large number of independent identically distributed random variables [19]. Since the probability density function (PDF) is not available in closed-form for most stable distributions, α -stable distribution is generally specified by its characteristic function.

Definition 1: A random variable X is said to follow α -stable distribution if there are parameters $0 < \alpha \leq 2$, $\sigma \geq 0$, $-1 \leq \beta \leq 1$, and $\mu \in \mathcal{R}$ such that its characteristic function is of the following form:

$$\begin{aligned} \phi(\omega) &= E(\exp(j\omega X)) \\ &= \exp\{-\sigma^\alpha |\omega|^\alpha (1 - j\beta(\operatorname{sgn}(\omega))\Phi) + j\mu\omega\}, \end{aligned} \quad (1)$$

with Φ being given by

$$\Phi = \begin{cases} \tan \frac{\pi\alpha}{2}, & \alpha \neq 1; \\ -\frac{2}{\pi} \ln|\omega|, & \alpha = 1. \end{cases} \quad (2)$$

Here, the function $E(\cdot)$ represents the expectation operation with respect to a random variable. α is called the characteristic exponent and indicates the index of stability, while β is identified as the skewness parameter. α and β together determine the shape of the distribution. Moreover, σ and μ are called scale and shift parameters, respectively. Specifically, if $\alpha = 2$, α -stable model reduces to Gaussian distribution.

Usually, it is challenging to prove whether a dataset follows a specific statistical distribution, especially for α -stable model without a closed-form expression for the PDF. Therefore, when a dataset is said to satisfy α -stable model, it usually means the

dataset is consistent with the hypothetical distribution and the corresponding properties. In other words, the validation needs to firstly estimate the parameters of α -stable model based on the given dataset, and then compare the real distribution of the dataset with the estimated α -stable model [20].

3 Temporal-spatial analysis of cellular traffic

3.1 Temporal dynamics of cellular traffic and its modelling

Due to subscribers' daily social behaviours, cellular traffic shows temporal-spatial variations [3]. Fig. 1b plots the three empirical traffic curves of all BSs and the representative BSs (BS1 and BS2). Obviously, the aggregated behaviour of all BSs expresses periodic feature. For individual BS, traffic volume is different, which implies the imbalanced and heterogeneous features of cellular traffic at each BS. Furthermore, the traffic record of each BS possesses significant volatility, thus indicating the existence of burstiness of cellular traffic.

In order to analyse the burstiness of cellular traffic at different time scales, we employ m -aggregated series as follows:

Definition 2: Given a discrete time series $X = (X_1, X_2, \dots)$, m -aggregated series $X^{(m)}$ ($X^{(m)} = X_n^{(m)}$; $n = 1, 2, 3, \dots$) is the average of the original series X over non-overlapping blocks of size m [6]

$$X_n^{(m)} = \frac{1}{m} \sum_{i=nm-(m-1)}^{nm} X_i. \quad (3)$$

As depicted in Figs. 2a and b, the three-aggregated traffic series becomes smoother (less bursty) to some degree. Meanwhile, the variance of the aggregated traffic series at different time scales is computed to weigh the degree of burstiness. The slowly decaying variance displayed in Figs. 2c and d reveal that the bursty of cellular traffic remains significant as time scale increases. What's more, the three-aggregated traffic series resembles the original one on shapes. In this regard, self-similarity comes naturally which reflects the LRD feature of network traffic [9].

Definition 3: Given a zero-mean, stationary time series $X = (X_1, X_2, \dots)$, we say that X has H -self-similarity, if for all positive $m \in \mathcal{N}$, the sum of the original series X over non-overlapping blocks of size m has the same distribution as X is rescaled by m^H [7]. That is

$$X_n \triangleq \frac{1}{m^H} \sum_{i=nm-(m-1)}^{nm} X_i = \frac{1}{m^H} m X_n^{(m)}. \quad (4)$$

Here H is called as Hurst parameter with the value ranging from 0.5 to 1, and can be used to measure the degree of self-similarity. The larger H is, the stronger self-similarity becomes. Specifically, if $H=1/2$, the sampled series can be said to be short-range dependent (SRD) and lack self-similarity [9].

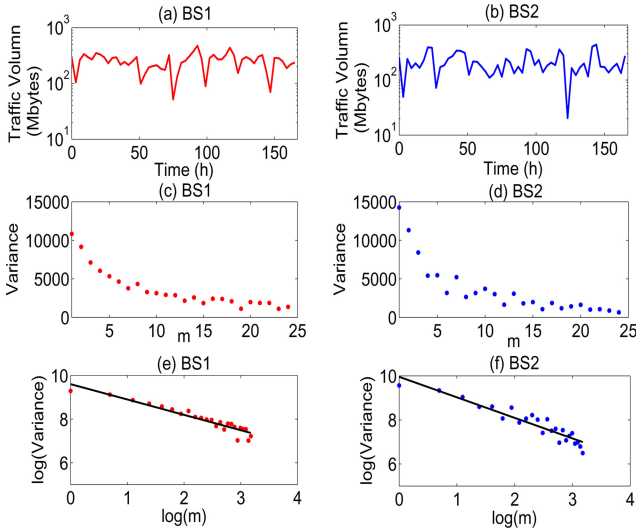


Fig. 2 Cellular traffic and its modelling (a), (b) Three-aggregated traffic series of BS1 and BS2, (c), (d) Variances of the aggregated traffic series and, (e), (f) Variance-time plot and the fitting result

Variance-time plot is adopted in this paper to measure self-similarity and estimate Hurst parameter [6], relying on the fact that the variance of the sample traffic series decreases more slowly than the reciprocal of the sample size for a self-similar series. The variance of $X^{(m)}$ is plotted against m on a log-log plot; a simple least squares line with slope $(-\beta)$ greater than -1 suggests self-similarity, and H is given by $H = 1 - \beta/2$ [7]. Figs. 2e and f illustrate the variance-time log-log plots of BS1 and BS2, where appropriate fitting results (H is 0.711 and 0.648 for the traffic series of BS1 and BS2, respectively) prove the existence of self-similarity of the sampled traffic series. Furthermore, we apply variance-time method into the whole dataset and give the CDF of Hurst parameter in Fig. 4a. Here, R -square is used to examine the accuracy of the linear trend of variance-time log-log plot. About 94% of R -square is larger than 0.8, which indicates a good linear fit.

The infinity of variance caused by LRD of network traffic calls for α -stable marginal distributions [8]. Therefore, α -stable distribution with the characteristics of burstiness and self-similarity is used here to model the traffic patterns. The traffic series of BS1 and BS2 are taken as representative examples, and Fig. 3 presents the corresponding comparison between the fitting results and the real ones in term of CDF (cumulative distribution function). The Kolmogorov–Smirnov test (K-S test) [21] is employed to check for goodness-of-fit of the empirical data. The K-S test results show that α -stable distribution cannot be rejected at the 5% significance level. Meanwhile, the measured K-S test statistic, which is the maximum distance between the CDF of empirical data and the reference distribution, are 0.0434 and 0.0671 for BS1 and BS2, respectively.

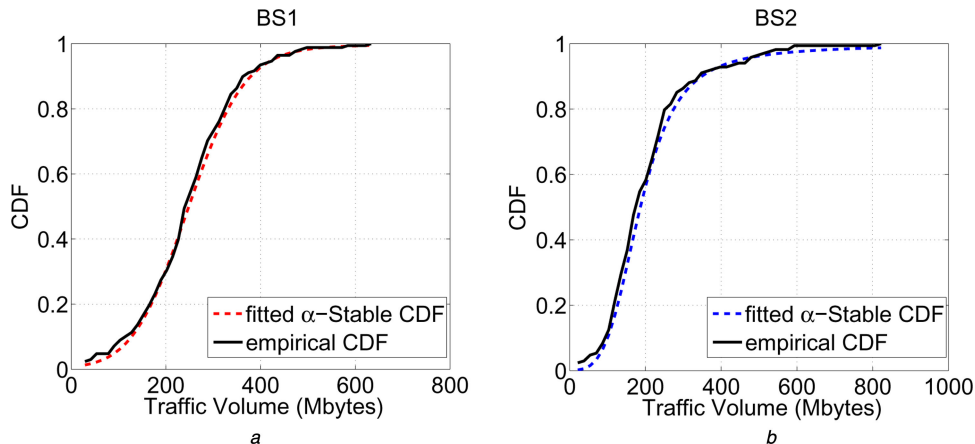


Fig. 3 For the traffic series of BS1 and BS2, α -stable distribution fitting results versus the empirical ones

Remark 1: Traffic expresses significant burstiness and self-similarity in cellular networks. Meanwhile, based on K-S test result, α -stable distribution is suitable to model the temporal pattern of cellular traffic.

3.2 Spatial dynamics of cellular traffic

Traffic usage over the spatial domain (geographic area), is closely associated with network performance [10]. Therefore, to help operators design network more efficiently and reasonably, we mainly analyse the spatial correlation of BSs traffic series from the global view (all BSs) as well as the local view (individual BS).

3.2.1 Global spatial correlation.: We calculate the Pearson correlation coefficient [22] between pairs of BSs by using the time series of traffic record. Fig. 4b shows the CDF of Pearson correlation coefficient between all pairs of BSs (black solid line) as well as the pairs of BSs within different mutual distances (dashed line). For all pairs of BSs, the median of the Pearson correlation coefficient is about 0.32, which manifests the traffic at BSs are indeed spatially correlated. For pairs of BSs in the category of different distances, all of the medians are larger than 0.32, i.e. 0.41 (≤ 0.5 km), 0.39 (≤ 1 km), 0.36 (≤ 5 km) and 0.36 (≤ 20 km). Clearly, closer BSs show higher Pearson correlation coefficient. Furthermore, it is noteworthy that CDFs of the Pearson correlation coefficient for BSs within 5 and 20 km are almost conforming.

3.2.2 Local spatial correlation.: From a local perspective, we calculate the correlation coefficient between a targeted BS and other surrounding BSs. Here, coefficient-distance curves of three typical BSs are depicted in Fig. 5. Fig. 5 mainly conveys two key meanings. Firstly, the traffic series of BSs are spatially correlated. Secondly, as the geographical distance between BSs increases, the correlation coefficient would not decrease accordingly. In other words, the correlation coefficients between pairs of BSs look like the wave trends that first fall and then rise and again fall. We may call it spatial long-range dependence. Taking advantage of this finding, we could understand why the CDFs of BSs within 5 and 20 km are almost conforming in the previous part.

Remark 2: Traffic usage at different BSs exhibits spatial correlation, even spatial long-range dependence to some extent.

4 BSs spatial deployed pattern

4.1 Modelling for BSs deployment

Paul *et al.* [3] pointed out that less than 10% of the subscribers generate 90% of the traffic load while 10% of the BSs carry 50–60% of the traffic load, which implies that traffic spatial dynamics exhibit heavy-tailed statistical feature as well. Humans with similar social behaviours tend to live together, which leads to various traffic hotspots and causes BSs to be deployed densely as clusters

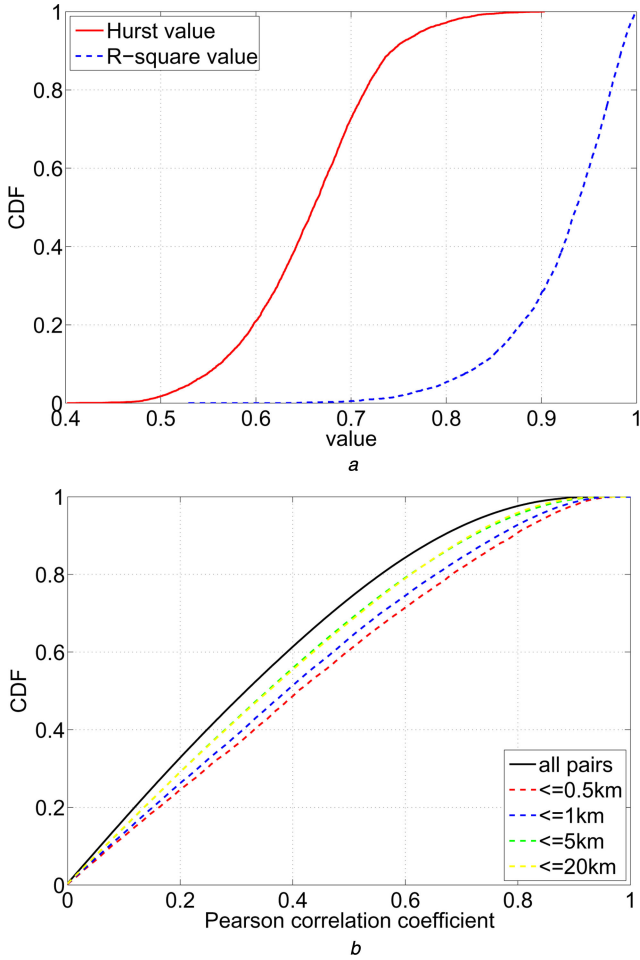


Fig. 4 Variance-time method into the whole dataset
 (a) CDF of Hurst parameter and R -square, (b) CDF of Pearson correlation coefficient between pairs of BSs (all pairs and pairs within different distances)

in certain areas. Intuitively, the BS spatial distribution would be heavy-tailed just like the spatial traffic dynamics. For example, a real region with size $45 \times 90 \text{ km}^2$ is selected from Fig. 1a and is depicted in Fig. 6a. Obviously, it is different from the traditional distribution pattern simulated by PPP like Fig. 6b. Furthermore, besides Poisson distribution, we also choose several representative heavy-tailed distributions in Table 2 as candidates to characterize the realistic BS deployment.

Specifically, we first sample the targeted region with a fixed sliding window randomly. Then we compute the BSs density for different 10,000 sample windows and obtain the empirical BS density distribution. Next, we estimated the unknown parameters in each candidate distributions. Finally, we compared the induced PDF with the exact (empirical) one.

In terms of the targeted region, we compute the PDF of BSs density with the sliding window $4 \times 4 \text{ km}^2$. After fitting the empirical PDF to candidate distributions (the estimated parameters

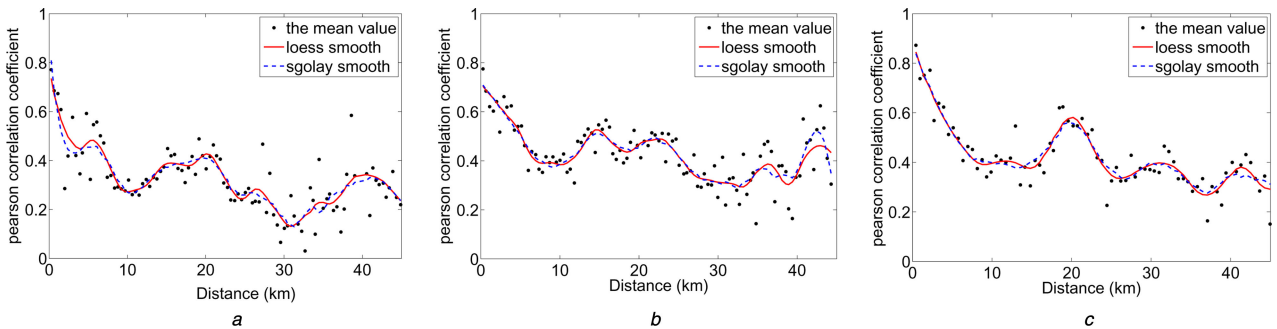


Fig. 5 Variation of Pearson correlation coefficient with the increasing of the distance between BSs
 (a) Example BS 1; (b) Example BS 2; (c) Example BS 3

in Table 2), we provide the comparison between the empirical BSs density distribution with the candidate ones in Fig. 7. As depicted in Fig. 7, the statistical pattern of BSs obviously exhibits heavy-tailed characteristics. Besides, among all candidate distributions, α -stable distribution most precisely matches the empirical PDF in terms of root mean square error (RMSE). The RMSE value of α -stable, Poisson, log-normal, Weibull is 0.0198, 0.2780, 0.0546 and 0.0942, respectively. These results indicate that the real distributed pattern of BSs is far away from complete randomness, which can be verified in Figs. 6a and b intuitively.

4.2 Self-similarity in BSs deployment

The applicability of α -stable distribution to BSs spatial deployment motivate us to research spatial self-similarity of BSs distribution. The region ($45 \text{ km} \times 90 \text{ km}$) is divided into 4050 small windows and BS number in each window is counted. Then we compute the Hurst coefficient of BSs density by the variance-time plot method. Figs. 8a and b illustrate the variances at different time scales and the variance-time log-log plot, respectively. The higher Hurst coefficient (0.78) proves the existence of self-similarity of BSs density in spatial domain.

Remark 3: Spatial pattern of the deployed BSs exhibit strong heavy-tailed and self-similarity characteristics, and α -stable distribution manifests itself as the most suitable one to model the BSs deployment in terms of RMSE.

5 Heterogeneity of BSs deployment and traffic distribution

Based on the statement in Sections 3 and 4, the inhomogeneous phenomenon does exist in BSs spatial distribution and traffic temporal distribution, which makes substantial effect on the network design and resource management. In this section, we focus on discussing how to quantise the heterogeneity degree for BSs and cellular traffic.

5.1 Spatial heterogeneity of BSs

In this part, we introduce two methods to study the spatial heterogeneity of BSs.

5.1.1 Density-based measure: The first method is the density-based measure that divides the whole region into smaller window and count the number of pattern points in each window. The detailed experimental procedure is listed as follows:

- Firstly, we subdivide the targeted region ($X \times Y$) into N same regions and obtain the number of BSs in each small region, namely, $(u_i, i = 1, 2, \dots, n)$.
- Then, according to u_i , we sort all the small regions in ascending order as $u_{(i)}$: $u_{(1)} \leq u_{(2)} \leq \dots \leq u_{(n)}$.
- Finally, the target curve is obtained as the ratio of the cumulative amount of $u_{(i)}$ to the total amount, which can be expressed as follows:

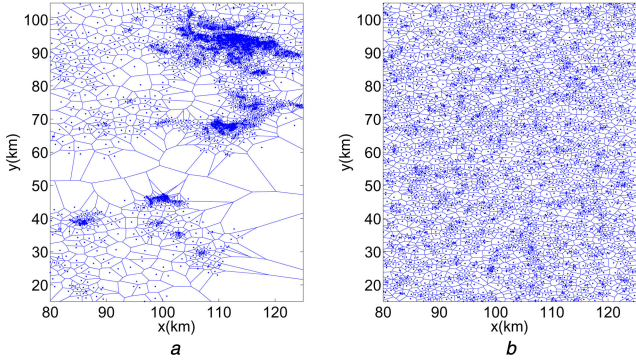


Fig. 6 BS spatial distribution
(a) Real BS locations (blue dots) and corresponding Voronoi cell area, (b) Traditional BS locations generated by PPP (3097 blue dots) and corresponding Voronoi cell area

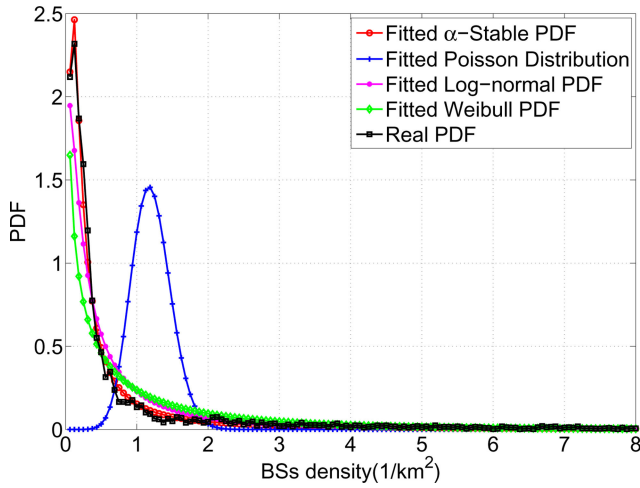


Fig. 7 Results after fitting BSs density to the candidate distributions, when the sliding window size is $4 \times 4 \text{ km}^2$

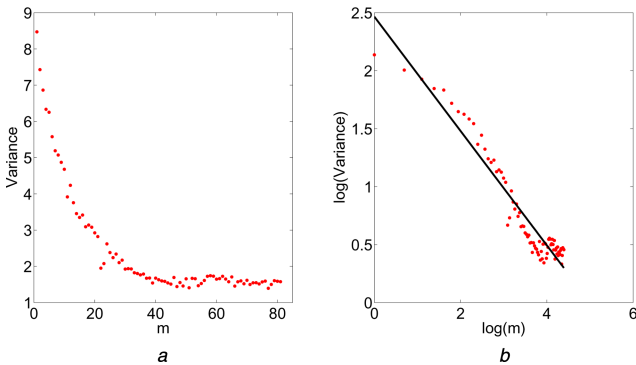


Fig. 8 Variances at different time scales and the variance-time log-log plot
(a) Variances at different time scales, (b) Variance-time log-log plot and fitting result

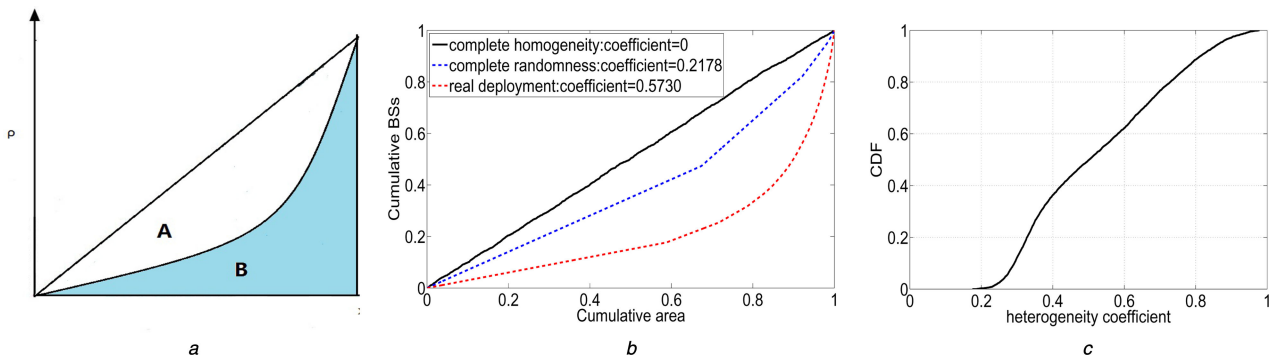


Fig. 9 Coefficient h of BSs heterogeneity
(a) Curve for $\rho(x)$, (b) Heterogeneity coefficient in different BSs distributed scenarios, (c) CDF of heterogeneity coefficient in terms of traffic series in time domain

Table 2 PDF and estimated parameters of candidate distributions

Distribution	PDF	Estimated parameters
Weibull	$pqx^{q-1}e^{-px^q}$	$p = 1.1668$ $q = 0.6519$
Log-normal	$\frac{1}{\sqrt{2\pi}nx}e^{-(\ln x - m)^2/(2n^2)}$	$m = -0.9658$ $n = 1.3646$
α -Stable	closed form not always exists	$\alpha = 0.6764$ $\beta = 0.9992$ $\sigma = 0.1288$ $\mu = -0.0519$
Poisson	$\frac{\lambda^k}{k!}e^{-\lambda}$	$\lambda = 19.2511$

$$\rho(x) = \frac{\sum_{i=1}^{nx} u(i)}{\sum_{i=1}^n u(i)}, \quad 0 \leq x \leq 1. \quad (5)$$

In order to describe the heterogeneity degree of BSs deployment, we present its coefficient h based on the well-known definition of Gini coefficient [23]. The coefficient h of BSs heterogeneity is computed as the ratio by area A and area B (shown in Fig. 9a)

$$h = 1 - \frac{A}{A+B} = 1 - 2 \int_0^1 \rho(x) dx. \quad (6)$$

The coefficient h ($0 \leq h \leq 1$) can be used as a metric of BSs heterogeneity degree. A larger h shows more BSs concentrated in the spatial domain, while a smaller h shows more equally and randomly distributed BSs. In other words, if the distributed BSs are more homogeneity-oriented, the curve is close to the 45° line. However, when BSs are more centralised, the curve will be closer to the lower right corner in Fig. 9a.

Basing the two aforementioned situations, shown in Figs. 6a and b, we carry out the above experimental steps with small window $1 \times 1 \text{ km}^2$ and get the simulation results in Fig. 9b. Obviously, the coefficient h of the real BSs deployment is much larger than the contrast one (random situation). These observations imply that the real deployed BSs have strong heterogeneity.

5.1.2 Distance-based measure. For density-based measure, it is challenging to find a suitable small window size. Therefore, in order to avoid this kind of problem, the distance-based measure, considering the distance between every two neighbouring points, is proposed. For 1D domain, the natural ordering of points are determined and the distance between every two neighbouring points can be captured easily. For 2D or multi-dimension, however, there is no natural ordering for points, thus the definition of neighbouring points is a problem.

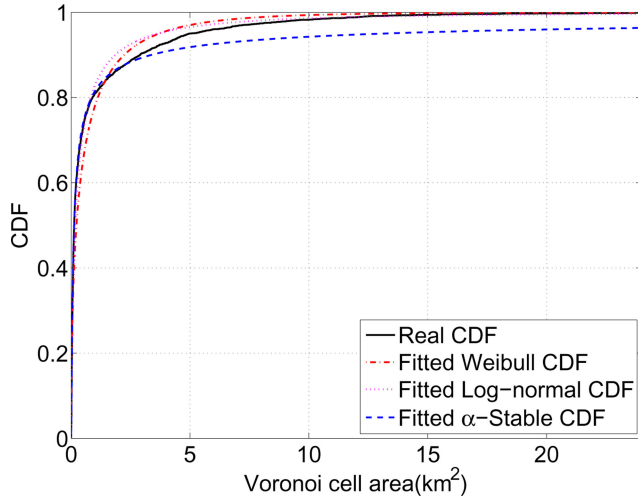


Fig. 10 CDF of Voronoi cell area among various distributions

Based on the above consideration, we adopt Voronoi cell area (V) as our choice in the spatial domain, instead of the distance popularly used in 1D domain.

Here, the coefficient of variation (CoV), which is the standard deviation normalised by mean, of Voronoi cell area (V), is used to measure the deviations of the distributed BSs from homogeneity. It can be defined as follows:

$$C = \frac{\sigma_V}{\mu_V}. \quad (7)$$

The larger is the value of C ($C \geq 0$), the higher is the BSs heterogeneity degree. If the BSs are perfectly homogeneous, the value of C is 0. According to the computed results, the values of C are 2.4727 and 0.8864 for the real BSs deployment and the random BSs deployment, respectively. It is consistent with the observations in the previous part.

In addition, for the purpose of deep understanding on the statistical characteristic of Voronoi cell area (V), we analyse its distribution fitting with the aforementioned representative distributions. Fig. 10 shows us the final fitting results. The K-S test results show that all candidate distributions would be rejected at the 5% significance level. However, the measured K-S test statistics are 0.0424, 0.0693 and 0.1144 for α -stable distribution, log-normal distribution and Weibull distribution, respectively. Therefore, in terms of the K-S statistics, α -stable distribution shows better fitting performance.

5.2 Temporal heterogeneity of cellular traffic

Inspired by the density-based method that measures the spatial heterogeneity of BSs distribution, we also adopt this method to estimate the temporal heterogeneity degree of cellular traffic. The experimental procedure is similar to the procedure in Section 5.1.1 except the following items:

- The target discussed now is time domain rather than space domain, which continues for 7 days.
- The small sliding window size is 1 h.
- In each small window, we compute the traffic volume instead of the number of BSs.

We take all traffic time series (5573) into computation, and present the CDF of heterogeneity coefficient in Fig. 9c. Obviously, the median of heterogeneity coefficient is about 0.5. In particular, when h is 0, it means that traffic load is totally equal at any time point. When h is 1, it means that all traffic loads are converged at certain time point.

Remark 4: Both the density-based measure metric and distance-based measure metric, evaluating the deviations of distributed BSs from homogeneity, are put forward. They both prove that the real

deployment pattern of BSs in cellular networks demonstrates strong heterogeneity. Meanwhile, traffic temporal heterogeneity degree is also quantised by density-based method.

6 Discussion and conclusion

6.1 Discussion

Mobile users, BSs and traffic demand are three elementary parts in the service chain of cellular networks, especially in the radio access networks. In this paper, we reveal the heterogeneity of BSs deployment and traffic demand on both temporal and spatial dimensions, based on realistic data records from operators. Although the information of individual mobile users are not available here to confirm the trinity-like connections between these entities, the linear dependence between traffic and BSs still enables an insight into the deployment of cellular networks.

For example, as depicted in this paper, the spatial density of both traffic and BSs exhibit heavy-tailed property, and the density-based or distance-based measure can be adopted to characterise the degree of heterogeneity. In real deployment, the operators do not expect the traffic to be too imbalanced, that's to say, the degree of heterogeneity cannot exceed some threshold. In this case, the operators need to deploy new BS to address this overflowing problem. However, it's difficult to simulate the resulting traffic distribution considering the new BS deployment even if the location is determined in advance. On the other hand, it's easier to calculate the heterogeneity degree of BSs given the location of to-be-deployed BSs in addition to the spatial distribution of all BSs in the cellular networks. Therefore, the heterogeneity degree of traffic can be derived according to the spatial relationship between traffic and BSs, thus forecasting the resulting effect of new deployment of BSs, which can be shown as an application of the results in this paper. Besides the spatial part, the temporal characterisation of traffic demand can also be used to minimise the spatial density of BSs or the corresponding on-off policies of BSs, where the cooperation between BSs need to be investigated.

6.2 Conclusion

This paper focuses on the comprehensive study about the statistical distribution characteristics of traffic and BSs in cellular network.

For traffic, we verify the periodicity of the aggregated behaviour of all BSs as well as the unbalanced feature for single BS. Then burstiness of cellular traffic at different time scales and the induced self-similarity are also confirmed. In consideration of the traffic nature, α -stable distribution is preferred here to extract the traffic characteristics, and K-S test results prove its applicability. Moreover, we also find that positive correlation does exist between the traffic series at BSs.

For BSs, we verify the existence of heavy-tailed characteristic in realistic BS deployment and α -stable distribution manifests its higher accuracy in modelling the BSs locations. Afterwards, the self-similarity of network topology is also proved by the Hurst coefficient. At the same time, considering the clustering effect of BSs deployment in real network topology, we present the density-based metric and the distance-based metric to evaluate the spatial heterogeneity of the deployed BSs. Both the density-base metric and the distance-based metric, prove that the real spatial pattern of BSs shows strong heterogeneity.

In summary, all findings based on the practical measurement data from mobile operator, provide helpful guidance for future network design and optimisation.

7 Acknowledgments

This work was supported in part by the Program for Zhejiang Leading Team of Science and Technology Innovation (No. 2013TD20), National Natural Science Foundation of China (No. 61731002, 61701439), the National Postdoctoral Program for Innovative Talents of China (No. BX201600133), and the Project funded by China Postdoctoral Science Foundation (No. 2017M610369).

8 References

- [1] Cisco. Cisco Visual Networking Index: Global Mobile data Traffic Forecast Update, 2014–2019, 2015, http://www.cisco.com/c/en/us/solutions/collateral/service-provider/visual-networking-index-vni/white_paper_c11-520862.html
- [2] Hasan, Z., Boostanimehr, H., Bhargava, V.K.: ‘Green cellular networks: a survey, some research issues and challenges’, *IEEE Commun. Surv. Tutor.*, 2011, **13**, (4), pp. 524–540
- [3] Paul, U., Subramanian, A.P., Buddhikot, M.M., *et al.*: ‘Understanding traffic dynamics in cellular data networks’. Proc. IEEE INFOCOM, Shanghai, China, 2011
- [4] Zhong, Y., Quek, T.Q., Ge, X.: ‘Heterogeneous cellular networks with spatio-temporal traffic: delay analysis and scheduling’, *IEEE J. Sel. Area. Commun.*, 2017, **35**, (6), pp. 1373–1386
- [5] Li, R., Zhao, Z., Zhou, X., *et al.*: ‘Intelligent 5G: when cellular networks meet artificial intelligence’, *IEEE Wirel. Commun.*, 2017, accepted and available at <http://ieeexplore.ieee.org/abstract/document/7886994/>
- [6] Leland, W.E., Taqqu, M.S., Willinger, W., *et al.*: ‘On the self-similar nature of ethernet traffic’, *ACM SIGCOMM Comput. Commun. Rev.*, 1993, **23**, (4), pp. 183–193
- [7] Yin, S., Lin, X.: ‘Traffic self-similarity in mobile ad hoc networks’, Proc. IFIP WOCN 2005, Dubai, United Arab Emirates, 2005
- [8] Karasarisidis, A., Hatzinakos, D.: ‘Network heavy traffic modeling using α -stable self-similar processes’, *IEEE Trans. Commun.*, 2001, **49**, (7), pp. 1203–1214
- [9] Cappe, O., Moulines, E., Pesquet, J.-C., *et al.*: ‘Long-range dependence and heavy-tail modeling for teletraffic data’, *IEEE Signal Process.*, 2002, **19**, (3), pp. 14–27
- [10] Shafiq, M.Z., Ji, L., Liu, A.X., *et al.*: ‘Characterizing geospatial dynamics of application usage in a 3G cellular data network’. Proc. IEEE INFOCOM, Orlando, USA, 2012
- [11] Paul, U., Subramanian, A.P., Buddhikot, M.M., *et al.*: ‘Understanding spatial relationships in resource usage in cellular data networks’. Proc. INFOCOM, Orlando, USA, 2012
- [12] Rathgeber, R.: ‘Spatial traffic distribution in cellular networks’. Proc. IEEE VTC, Ottawa, Ont, 1998
- [13] Michalopoulou, M., Riihijärvi, J., Mähönen, P.: ‘Towards characterizing primary usage in cellular networks: a traffic-based study’. Proc. IEEE DySPAN, Aachen, Germany, 2011
- [14] Haenggi, M., Andrews, J.G., Baccelli, F., *et al.*: ‘Stochastic geometry and random graphs for the analysis and design of wireless networks’, *IEEE J. Sel. Area Commun.*, 2009, **27**, (7), pp. 1029–1046
- [15] Taylor, D.B., Dhillon, H.S., Novlan, T.D., *et al.*: ‘Pairwise interaction processes for modeling cellular network topology’. Proc. IEEE GLOBECOM, Anaheim, CA, 2012
- [16] Lee, C.-H., Shih, C.-Y., Chen, Y.-S.: ‘Stochastic geometry based models for modeling cellular networks in urban areas’, *Wirel. Netw.*, 2013, **19**, (6), pp. 1063–1072
- [17] Zhang, J., Wang, W., Zhang, X., *et al.*: ‘Base stations from current mobile cellular networks: measurement, spatial modeling and analysis’. Proc. IEEE WCNC, Shanghai, China, 2013
- [18] Ge, X., Tu, S., Mao, G., *et al.*: ‘5G ultra-dense cellular networks’, *IEEE Wirel. Commun.*, 2016, **23**, (1), pp. 72–79
- [19] Samorodnitsky, G.: ‘Stable non-Gaussian random processes: stochastic models with infinite variance’ (Chapman and Hall/CRC, New York, 1994), <http://www.amazon.com/Stable-Non-Gaussian-Random-Processes-Stochastic/dp/0412051710>
- [20] Ge, X., Zhu, G., Zhu, Y.: ‘On the testing for α -stable distributions of network traffic’, *Comput. Commun.*, 2004, **27**, (5), pp. 447–457
- [21] Kolmogorov-Smirnov test, https://en.wikipedia.org/wiki/Kolmogorov-Smirnov_test
- [22] Lawrence, I., Lin, K.: ‘A concordance correlation coefficient to evaluate reproducibility’, *Biometrics*, 1989, **45**, (1), pp. 255–268
- [23] Gini-Coefficient, <https://en.wikipedia.org/wiki/Gini-coefficient>