Human Mobility Patterns in Cellular Networks

Xuan Zhou, Zhifeng Zhao, *Member, IEEE*, Rongpeng Li, Yifan Zhou, Jacques Palicot, and Honggang Zhang, *Senior Member, IEEE*

***Abstract*—This letter investigates inter-arrival time, dwell time distributions and other mobility patterns in mobile cellular networks. It has been generally assumed and widely accepted that both inter-arrival time and dwell time distributions can be well approximated by exponential distribution. However, based on real cellular data measurements, we evaluate the fitness of various typical statistical distributions such as power-law, exponential, Weibull, lognormal and Rayleigh distributions, and find that a power-law distribution fits both inter-arrival time and dwell time more precisely. Besides, mobility patterns in daytime, night, rural and urban areas provide further demonstrations of the power-law model. Moreover, new models on the distributions of inter-departure time and the number of arrived subscribers are also proposed to characterize other mobility patterns, and the corresponding simulation results are well consistent with the empirical ones.**

***Index Terms*—Human mobility, inter-arrival time, dwell time, inter-departure time, the number of arrived subscribers, power-law, cellular network.**

I. Introduction

HUMAN mobility patterns play an important role in protocol design and performance analysis of cellular networks. During the past ten years, many researchers have designed and analyzed various cellular network management schemes, taking for granted that inter-arrival time and dwell time are either exponentially distributed [1] or other similar distributions [2]. On the other hand, several power-law (Pareto) distribution phenomena have been discovered in various areas of human behaviors lately, such as time intervals between consecutive emails sent by users obey power-law [3], dwell time and sign-on inter-arrival time in WLAN adhere to power-law as well [4]. Other non-exponential distributions have also been observed for the dwell time in WLAN scenario [5]. [6] reports that travel length and pause time of human can be modeled as truncated Pareto distributions by GPS traces. [7] also concludes that inter-contact time can be described by the power-law distribution. The findings above inspire us to speculate that the inter-arrival time and dwell time in cellular networks might follow power-law distribution rather than exponential distribution.

Manuscript received April 22, 2013. The associate editor coordinating the review of this letter and approving it for publication was B. Bellalta.

X. Zhou, Z. Zhao, R. Li, Y. Zhou, and H. Zhang are with the Dept. of Information Science and Electronic Engineering, Zhejiang University, Hangzhou 310027, China (e-mail: {zhouxuan, zhaozf, lirongpeng, zhouyftt, honggangzhang}@zju.edu.cn). H. Zhang is also with the Université Européenne de Bretagne & Supélec (e-mail: honggang.zhang@supelec.fr).

J. Palicot is with Supélec, Avenue de la Boulaie CS 47601, F-35576 Cesson-Sévigné Cedex, France (e-mail: jacques.palicot@supelec.fr).

This letter is partially supported by the National Basic Research Program of China (973Green, No. 2012CB316000) and the National Natural Science Foundation of China (NSFC) under grant number 61071130.

Digital Object Identifier 10.1109/LCOMM.2013.090213.130924

To validate the above assumptions and models, the most effective way is to carry out measurements in real cellular networks, and examine the fitness of various distributions to the observations. In this letter, we make an intensive study on temporal and spatial human mobility patterns through the collected practical cellular networks' observations in one month. Our measurement results validate that the inter-arrival time and dwell time of mobile subscribers can be more accurately modeled using power-law distribution, compared to exponential distribution. Moreover, although the inter-arrival and dwell time have been previously discussed by a number of papers [8], however to our best knowledge, no paper has ever investigated inter-departure time of mobile subscribers within cellular networks. We model the distribution of inter-departure time according to the patterns generated from the power-law distributed inter-arrival time and dwell time, and verify the newly proposed model through the comparison with the measurements of actual traces of mobile users. Meanwhile, to describe the spatial imbalance features of mobility patterns in a large area, we also analyze the distribution of the numbers of arrived subscribers in unit square area during a period of time, and find that power-law distribution is suitable to fit as well.

II. Dataset Description

Our dataset is based on 1 month anonymous traffic records collected from 7 million subscribers, originated from about 15000 GSM or UMTS base stations (BSs) of China Mobile within a region of 3000 km^2. Each record includes timestamp, anonymous subscriber ID, location area code (LAC) and cell ID. Besides that, each traffic record also contains the subscriber service activities such as calls and short messages, as well as network activities like location updates, handovers and pagings. The abundant information in daily traffic records amounts up to 30 Gigabytes and provides a sufficient dataset foundation for our analysis and evaluation. Furthermore, the dataset is sorted and classified by temporal and spatial properties, like daytime/night and rural/urban area. Here, we regard one daytime subset as the group of records with timestamps from 9 AM to 5 PM, while the timestamps in a night subset range from 10 PM to 5 AM. Similarly, the rural areas consider the places like countrysides and villages while the urban areas merely include cities. Suburb areas are included in neither rural nor urban subsets, so as to make distinct decisions.

III. Inter-Arrival Time and Dwell Time

In this section, inter-arrival time and dwell time of subscribers in a BS are discussed. The time between two consecutive arrivals of subscribers can be obtained based on the arrival

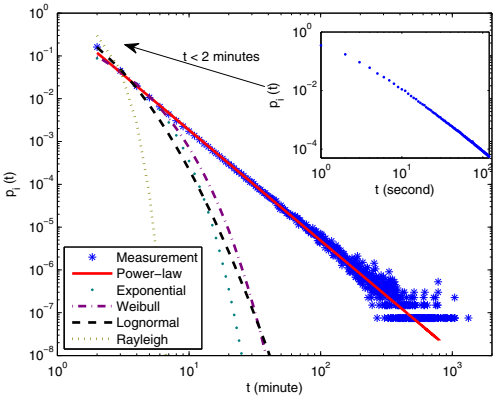

Fig. 1. The fitting results of inter-arrival time by candidate distributions.

TABLE I
INTER-ARRIVAL TIME: AIC TEST OF CANDIDATE DISTRIBUTIONS.

Distribution	Probability Density Function	AIC	Weight
Power-law (PL)	at^{-b}	13833	1.000
Exponential (EXP)	ae^{-bt}	20395	0.000
Weibull (WB)	$abt^{b-1}e^{-at^b}$	14137	7.4×10^{-67}
Lognormal (LN)	$\frac{1}{bt\sqrt{2\pi}} \exp[\frac{(\ln(t)-a)^2}{-2b^2}]$	19455	0.000
Reyleigh (RL)	$\frac{t}{a^2} \exp[-\frac{t^2}{2a^2}]$	30311	0.000

timestamps of subscribers in the traffic records. Therefore, the probability density function (PDF) of inter-arrival time, denoted as $p_i(t)$, can be calculated from the recorded dataset, where t is sampled in the order of minute. Afterwards, the distribution of inter-arrival time is fitted using the common heavy-tail distributions listed in Table I. Fig. 1 depicts the relevant fitting results with respect to the empirical data, where the unknown parameters (e.g., a, b) in the distributions are found by maximum likelihood estimation (MLE) method. Together with the MLE, an Akaike information criterion (AIC)[1] [9] is applied so as to quantitatively find the best distribution model. Table I summarizes the AIC results and indicates that a power distribution $p_i(t) = 0.706t^{-2.58}$ mostly approximates the empirical inter-arrival time.

To examine the influence of the temporal and spatial factors on inter-arrival time, the dataset is classified into several subsets in either temporal or spatial dimension. Fig. 2(a) and (b) show the impact of temporal and spatial properties on $p_i(t)$, respectively. Besides, Fig. 2(c) illustrates the distribution of inter-arrival time during one day, where the sample unit is zoomed from minute to second. As Table I shows that Weibull

[1]AIC values are calculated by the following formula [9]:

$$AIC = -2\log(\mathcal{L}(\cdot|\text{data})) + 2K \quad (1)$$

where $\mathcal{L}(\cdot)$ is the likelihood function and K is the number of estimable parameters in the approximating model. Let AIC_{min} be the minimum of AIC values of M candidate distributions, the Akaike weight of the ith model, which is considered as the normalization of the model likelihoods, is defined as [9]:

$$w_i = \frac{\exp[(AIC_{min} - AIC_i)/2]}{\sum_{i=1}^{M} \exp[(AIC_{min} - AIC_i)/2]} \quad (2)$$

The Akaike weights can be interpreted as the relative preciseness when all the models try to fit the empirical observations on the basis of minimal (estimated) information loss. Therefore, the model with largest weight is the most promising one to describe the data distribution in the candidate models.

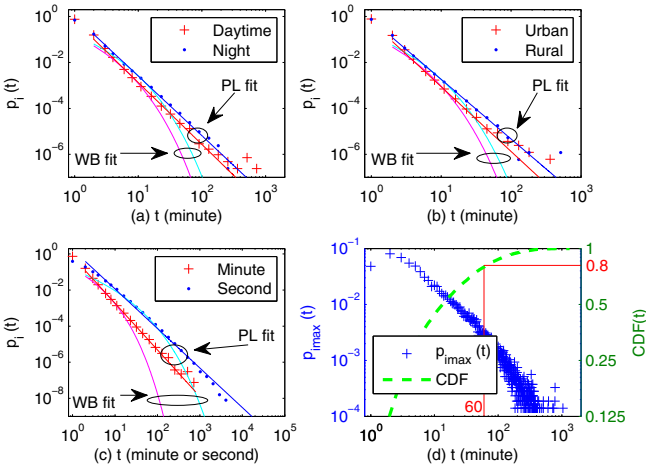

Fig. 2. The distributions of inter-arrival time under different conditions.

TABLE II
PARAMETERS OF POWER-LAW DISTRIBUTION AND COMPARISON OF AIC WEIGHTS IN FIG. 2(A)(B)(C).

Parameters	Daytime	Night	Rural	Urban	Second	Minute
a	0.817	0.661	0.576	0.722	2.867	0.706
b	2.598	2.585	2.518	2.666	2.388	2.582
w_{PL}	1.000	1.000	1.000	1.000	1.000	1.000
w_{WB}	7.5×10^{-25}	0.000	0.000	0.000	1.4×10^{-15}	0.000

distribution is the second best model for inter-arrival time, the fit goodness of the power-law distribution is only compared with Weibull distribution for simplicity. Table II lists the fitted parameters of the power-law distribution (i.e., a and b in at^{-b}) and AIC weights in each case. It can be found that in all situations, the slopes (i.e., b in at^{-b}) are almost identical. Additionally, the slopes of the daytime and urban subset are a little steeper than those of the night and rural subset, because the mobility is more frequent in the former scenarios.

Fig. 1 also indicates that most of the inter-arrival time takes small values. Therefore, a few large values will be submerged in daytime. Consequently, it's still worthwhile to investigate the distribution of maximal inter-arrival time in each cell. Fig. 2(d) presents the measured result and cumulative distribution function (CDF) of maximal inter-arrival time. In Fig. 2(d), the solid lines indicate that 20% of cells have a maximal inter-arrival time larger than 1 hour. After giving a deeper investigation on the timestamps and cells with maximal inter-arrival time larger than one hour, we also observe that 90.4% of the timestamps appear at night while the corresponding ratio is 3.2% in daytime. Moreover, the phenomenon emerges in 48.9% of the cells locating in rural areas and 21.1% in urban areas.

Furthermore, after categorizing the incoming subscribers of a cell according to their next hops (adjacent cells), we discover that the slopes of $p_i(t)$ are irrelevant with the subscribers' destinations (adjacent cells) and remain basically the same (Mean=2.53, Standard Deviation=0.12). This conclusion will be used in the next section to model the arrival-departure process.

Similar to $p_i(t)$, the PDF of dwell time denoted as $p_d(t)$

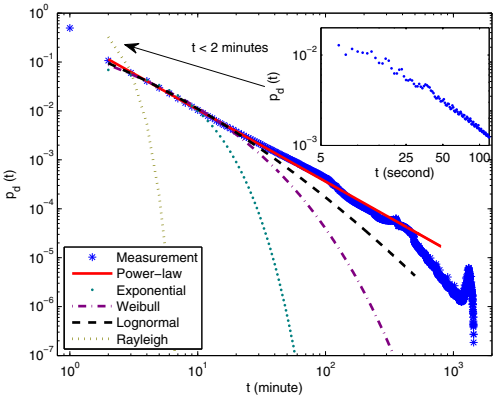

Fig. 3. The fitting results of dwell time by candidate distributions.

TABLE III
DWELL TIME: AIC TEST OF CANDIDATE DISTRIBUTIONS.

AIC test	PL	EXP	WB	LN	RL
Values	8834	13557	12506	9128	30981
Weights	1.000	0.000	0.000	1.63×10^{-64}	0.000

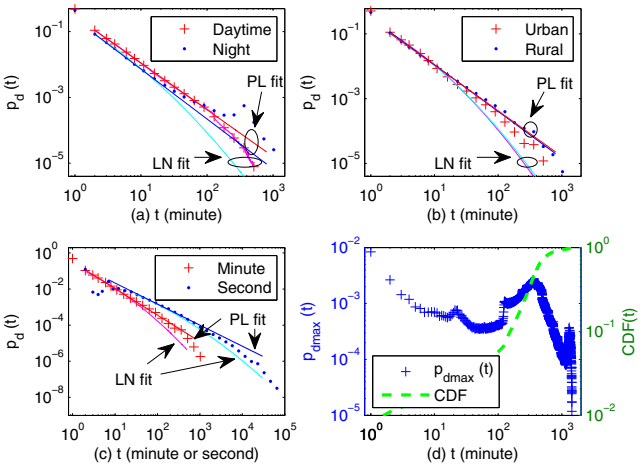

Fig. 4. The distributions of dwell time under different conditions.

TABLE IV
PARAMETERS OF POWER-LAW DISTRIBUTION AND COMPARISON OF AIC WEIGHTS IN FIG. 4(A)(B)(C).

Parameters	Daytime	Night	Rural	Urban	Second	Minute
a	0.295	0.239	0.270	0.356	0.980	0.314
b	1.422	1.517	1.407	1.442	1.359	1.47
w_{PL}	1.000	1.000	1.000	0.862	0.896	1.000
w_{LN}	0.000	3.4×10^{-48}	0.000	0.138	0.104	0.000

can be calculated, where t is also rounded to minute. Fig. 3 compares the fitting results using heavy-tail distributions with MLE and Table III summarizes the goodness of fit by AIC test. It can be easily observed that the power-law distribution fits more precisely to the measured $p_d(t)$ using $0.3137t^{-1.47}$. Thus, the power-law distribution $p_d(t) = at^{-b}$ is adopted for further discussions. Here, it is worthy to note that lognormal distribution becomes the second best model, hence we merely compare power-law with lognormal in the study of the dwell time distribution.

The PDFs of dwell time with respect to temporal and spatial views are plotted in Fig. 4(a)(b), as well as in different time granularities in Fig. 4(c). Besides, the fitting parameters in power-law distribution and corresponding AIC weights are listed in Table IV. It comes to the conclusion that cells with different temporal and spatial properties have almost identical distributions which could be well fitted by power-law functions, where the only distinction between them is pointed to the heavy-tails. The distribution of the maximum dwell time shown in Fig. 4(d) is quite different from that of the maximum inter-arrival time in Fig. 2(d); here, several peak values appear at 1 minute, 2 hours and 6 hours. These values might be attributed to the following reasons: 1 minute dwell time originates from the subscribers passing through the cells, 6 hours dwell time represents the schedule of normal office workers, and periodic location updates of terminals (T3212 in 3GPP TS 24.008) exert influence on 2 hours dwell time.

Subsequently, based on the power-law distribution at^{-b}, we look into the conditional probability of inter-arrival and dwell time deduced as follows:

$$\Pr(T > s + t | T > s) = p(t|s) = \frac{(b-1)(1+t/s)^{-b}}{s} \quad (3)$$

where t denotes the excess waiting time given that the previous subscriber has arrived since s minutes ago. (3) implies that the power-law distribution exhibits a memory property, differing from the exponential distribution. Hence we can predict the following arrival time and the excess dwell time by (3) whose distributions are depicted in Fig. 5(a)(b), respectively. Interestingly, a is irrelevant to $p(t|s)$, and $p(t|s)$ performs like a power-law distribution when $t > s$.

IV. INTER-DEPARTURE TIME AND THE NUMBER OF ARRIVED SUBSCRIBERS

Suppose that a cell has N adjacent cells (i.e., C_{adj1}, $C_{adj2}, \cdots, C_{adjN}$). Therefore, the inter-arrival time for subscribers, whose next hop is C_{adji}, still adheres to power-law distribution $p_i(t)$, whereas the corresponding dwell time for a subscriber obeys power-law distribution $p_d(t)$. Recall the conclusion in Section III that the dwell time of a subscriber is independent of both inter-arrival time and the corresponding next hop. Hence, the arrival-departure relationship can be modeled as N memory birth-death (MBD) processes (Fig. 6(a)), whose PDFs of inter-birth time and life time are $p_i(t)$ and $p_d(t)$, respectively. As a result, the inter-death time of N MBD processes can explicitly describe the inter-departure time of the cell (Fig. 6(b)). In Fig. 6(d), the simulation and measurement distributions are plotted whereas Fig. 6(c) depicts the fitting result using MLE and shows the inter-departure time is still power-law distributed. Unlike traditional birth-death (BD) process whose birth rate and death-rate are memoryless [10], MBD process with power-law distributed inter-birth time has a complex form of birth rate [11].

The last proposed model concerns the number of arrived subscribers into a pixel region within a time interval T. We define a $50m \times 50m$ region as a pixel and T as 24 hours. Usually, each pixel belongs to a certain BS, whereas one BS could cover several pixels. Without loss of generality, the

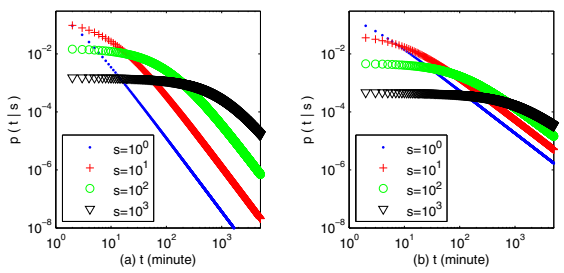

Fig. 5. (a) The conditional PDF of inter-arrival time. (b) The conditional PDF of dwell time.

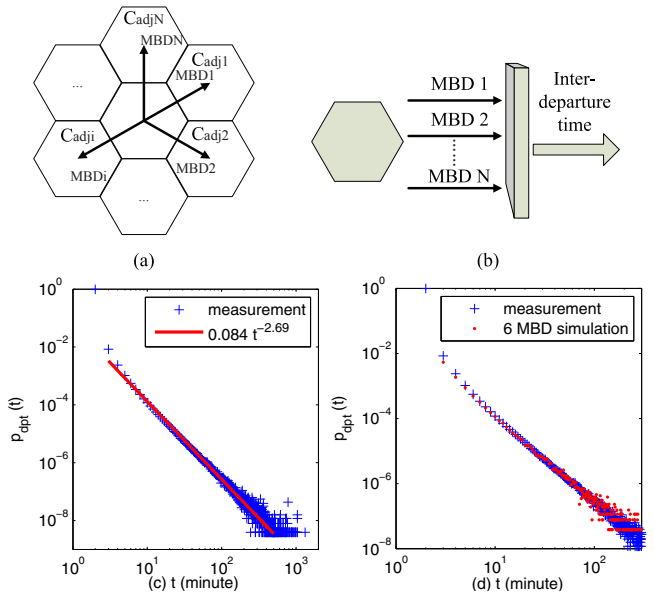

Fig. 6. (a) An illustration of N MBD processes model of arrival-departure relationship. (b) An illustration of the inter-departure model with N MBD processes. (c) The fitting function and measurements of inter-departure time. (d) The simulation result and measurements of inter-departure time.

proportions of pixels with k arrived subscribers in the interval T are denoted as $p_a(k)$. In Fig. 7(a), the measured PDF $p_a(k)$ and CDF are plotted. It suggests an imbalance pattern that in one day about 9% of pixels have less than one subscriber arrived, whereas 50% of pixels have less than 70 arrivals. More importantly, the PDF $p_a(k)$ can be well fitted using a power-law distribution with MLE, as shown by the AIC test in Table V. Yet, Fig. 7(b)(c) and Table VI illustrate that the slope of $p_a(k)$ at night is a little steeper than that in daytime.

V. Conclusion

In this letter, we study the human mobility models of inter-arrival time, dwell time, inter-departure time and number of arrived subscribers in unit square area, based on the dataset from on-operating cellular networks. In particular, we show that the human mobility models exhibit considerable power-law characteristics. These findings also indicate that the human activities have memory property, which make the activity prediction reasonable. Moreover, the imbalance property of number of the arrived subscribers establishes a macroscopical view of human clustering and aggregation attributes.

TABLE V
NUMBER OF ARRIVALS: AIC TEST OF CANDIDATE DISTRIBUTIONS.

AIC Test	PL	EXP	WB	LN	RL
Values	7291	9563	7405	8393	18720
Weights	1.000	0.000	1.51×10^{-25}	0.000	0.000

TABLE VI
FITTING RESULTS OF NUMBER OF ARRIVED SUBSCRIBERS IN FIG. 7

Parameters	Daytime	Night	Rural	Urban	Total
a	0.117	0.225	0.167	0.127	0.087
b	0.981	1.195	1.013	0.968	0.927

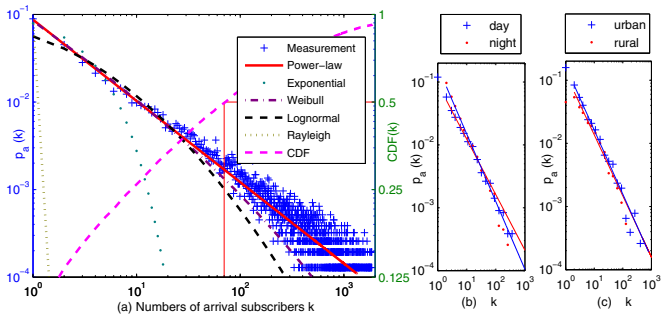

Fig. 7. (a) The PDF and CDF of the number of arrived subscribers. (b) The number of arrived subscribers in day and night. (c) The number of arrived subscribers in urban and rural.

References

[1] Y. Du, J. Fan, and J. Chen, "Experimental analysis of user mobility pattern in mobile social networks," in *2011 IEEE Wireless Communications and Networking Conference*.
[2] E. Jugl and H. Boche, "Dwell time models for wireless communication systems," in *1999 IEEE Vehicular Technology Conference – Fall*, vol. 5.
[3] A.-L. Barabási, "The origin of bursts and heavy tails in human dynamics," *Nature*, vol. 435, pp. 207–211, May 2005.
[4] S. Thajchayapong and J. M. Peha, "Mobility patterns in microcellular wireless networks," *IEEE Trans. Mob. Comput.*, vol. 5, pp. 52–63, Jan. 2006.
[5] E. Zola and F. Barcelo-Arroyo, "Impact of mobility models on the cell residence time in WLAN networks," in *2009 IEEE Sarnoff Symposium*.
[6] I. Rhee, M. Shin, S. Hong, K. Lee, S. J. Kim, and S. Chong, "On the Levy-walk nature of human mobility," *IEEE/ACM Trans. Netw.*, vol. 19, no. 3, pp. 630–643, Jun. 2011.
[7] T. Karagiannis, J.-Y. Le Boudec, and M. Vojnovic, "Power law and exponential decay of intercontact times between mobile devices," *IEEE Trans. Mob. Comput.*, vol. 9, pp. 183–194, Oct. 2010.
[8] N. Aschenbruck, A. Munjal, and T. Camp, "Trace-based mobility modeling for multi-hop wireless networks," *Computer Commun.*, vol. 34, no. 6, pp. 704–714, May 2011.
[9] K. P. Burnham and D. R. Anderson, "Multimodel inference: understanding AIC and BIC in model selection," *Sociological Methods and Research*, vol. 33, no. 2, pp. 261–304, Nov. 2004.
[10] J. Virtamo, "Queueing theory," *Lecture Notes*, Fall, 2005.
[11] D. A.Benson, R. Schumer, and M. M. Meerschaert, "Recurrence of extreme events with power-law interarrival times," *Geophysical Research Lett.*, vol. 34, no. 16, Aug. 2007.