Scale-free user-network approach to telephone network traffic analysis

Yongxiang Xia,* Chi K. Tse,† Wai M. Tam,‡ Francis C. M. Lau, and Michael Small
Department of Electronic and Information Engineering, Hong Kong Polytechnic University, Hong Kong, China
(Received 4 April 2005; published 16 August 2005)

The effect of the user network on the telephone network traffic is studied in this paper. Unlike classical traffic analysis, where users are assumed to be connected uniformly, our proposed method employs a scale-free network to model the behavior of telephone users. Each user has a fixed set of acquaintances with whom the user may communicate, and the number of acquaintances follows a power-law distribution. We show that compared to conventional analysis based upon a fully connected user network, the network traffic is significantly different when the user network assumes a scale-free property. Specifically, network blocking (call failure) is generally more severe in the case of a scale-free user network. It is also shown that the carried traffic is practically limited by the scale-free property of the user network, rather than by the network capacity.

DOI: 10.1103/PhysRevE.72.026116 PACS number(s): 89.75.Hc, 89.75.Da, 89.75.Fb

I. INTRODUCTION

A recent study of the scale-free property of so-called *complex networks* has motivated research in the modeling of practical networks based upon certain specific network topologies that possess properties closely resembling those of realistic physical networks [1,2]. In general terms, a *complex network* may be characterized by a large number of nodes and a set of complex relationships between them. Numerous examples of complex networks are found in social, information, technological, and biological systems [3]. The scale-free property provides a new perspective to analyze these systems.

The telephone networks, including the conventional telephone network and the cellular mobile network, are kinds of complex networks. They have undergone rapid developments in the past few decades. With a growing number of end users and increasing demand for a greater variety of services, operators are facing challenges in providing a variety of communication services and at the same time maintaining an adequate level of the *quality of service* (QoS) [4,5]. In order to facilitate a better network planning, traffic analysis that can reflect the true network behavior is indispensable.

The objective of traffic modeling is to construct models that can capture the salient statistical properties of the traffic. Several traffic models for communication networks have been proposed [6]. They are derived by fitting the existing traffic data under particular sets of conditions. Since the underlying mechanisms of the complex network behavior are unknown or simply not taken into account in the modeling process, such models fall short of a clear connection with the actual physical processes that are responsible for the behavior observed in the traffic data.

From the viewpoint of complex networks, the user network underlying any communication network is a complex network exhibiting the scale-free property [2]. Up until now, complex network behavior in telephone networks has been rarely considered. Aiello *et al.* [15] studied the scale-free property in the daily traffic of long-distance calls in a telephone network. However, to the best of our knowledge, traffic analysis based upon a scale-free user network is completely unavailable.

Some other traffic networks, such as the Internet [7–12] and airline networks [13,14], have already been studied from the viewpoint of complex networks. The Internet can be considered as a hugh traffic network in which data sources and sinks are interconnected by a network of routers. Moreover, the Internet is a packet-switching system, meaning that the data sources need to first break down the information into smaller units called *packets* that then transverse independently through the network of routers before arriving at the data sinks. The role of the routers is thus to accept data packets from the input and retransmit them at the appropriate output based on the destination address. During the transmission process, no dedicated connection is set up between the source and destination, and, consequently, packets from different sources may share the communication link between the routers. Compared to the Internet, the airline traffic network has a very similar configuration in that passengers and goods are like packets and the airports can be treated as routers. Passengers and goods arriving at the airports will be transported to other airports or their destinations by planes. In this sense, air traffic systems can also be considered as a kind of packet-switching systems. In contrast, telephone networks are circuit-switching systems, in which a dedicated connection is established between the caller and the receiver for each telephone call. Moreover, this connection cannot be used by other users during the call conversation. Hence, although the essence of the Internet, the airline traffic networks, and the telephone networks is the same—the traffic load, the network theories applying to the study of the Internet or airline traffic networks are not applicable to the telephone networks.

In this paper we attempt to incorporate a *user network model* in the analysis of telephone network traffic. Our purpose is twofold. First, we aim to provide a clear connection between the user network behavior and the network traffic. Second, we aim to illustrate how network traffic data can be more realistically simulated with the inclusion of a proper

*Electronic address: enyxxia@eie.polyu.edu.hk
†Electronic address: cktse@eie.polyu.edu.hk
‡Electronic address: tamwm@eie.polyu.edu.hk

user network behaviorial model. This study clears up several misconceptions. Telephone traffic (including mobile network traffic) cannot be considered without taking into account the way in which human users are connected in the real world. The fact that human networks possess the scale-free property can change the way network resources have to be planned. For instance, we will show that limited network capacity is not always the cause of call failures, while the scale-free property of the user network is the real evil. Thus, increasing network capacity can be useless or irrelevant to enhancing the traffic in a telephone network.

II. USER NETWORK CONFIGURATION

Formally, we may describe a user network in terms of nodes and connections. A node is a user, and a connection between two nodes indicates a possibility that these two users may call each other, i.e., a connection connects a pair of acquaintances.

In the classical traffic analysis, each user can call any other user with equal probability. Thus, the user network is a fully connected network. In such a user network, the effect of each user is assumed to be identical.

However, in reality, some users make more calls than do others. A relatively small group of users are usually responsible for most of the calls and hence have a comparatively greater impact on the traffic. Our basic assumption of the user network is that it is not uniform, i.e., a user does not call every user in the network with equal probability. In fact, users usually call only their own acquaintances, such as family members, colleagues, and friends. If a user has more acquaintances, the probability of him making or receiving a call at any time is higher. Thus, in the real user network, user i only has n_i connections that connect him to his n_i acquaintances.

It has been found that many human networks are scale-free networks, with n_i typically following a power-law distribution [2]:

$$p(n_i) \sim n_i^{-\gamma}, \qquad (1)$$

where $\gamma > 0$ is the characteristic exponent.

In our study, the following two-step method is used to construct the scale-free user network. First, the number of acquaintances n_i for user i is determined by a power-law distributed random number. In other words, the size of the acquaintance list for each user is fixed in this step. Next, the acquaintance list of user i is filled by selecting users in the network randomly. The relationship of acquaintance is bidirected, i.e., if user i is selected as an acquaintance of user j, then user j is automatically added into user i's acquaintance list.

When a user is going to make a call, he randomly chooses a receiver from his acquaintance list. The user network configuration is shown in Fig. 1, which is a typical scale-free configuration [1]. Specifically, each node represents a user, and a link between two nodes indicates that these two users are acquaintances. The degree (i.e., number of links) of node i is equal to n_i. Figure 2 shows a power-law distribution of n_i in a scale-free user network. We clearly see that a relatively small number of users have a large number of acquaintances. In a study of long distance call traffic by Aiello *et al.* [15], the incoming and outgoing connections were found to follow a power-law distribution, similar to (1), and the exponents $\gamma_{\text{in}} = \gamma_{\text{out}}$ was about 2.1. This clearly suggests that users do not contribute equally to the network traffic. In the following sections, we will study this effect in detail.

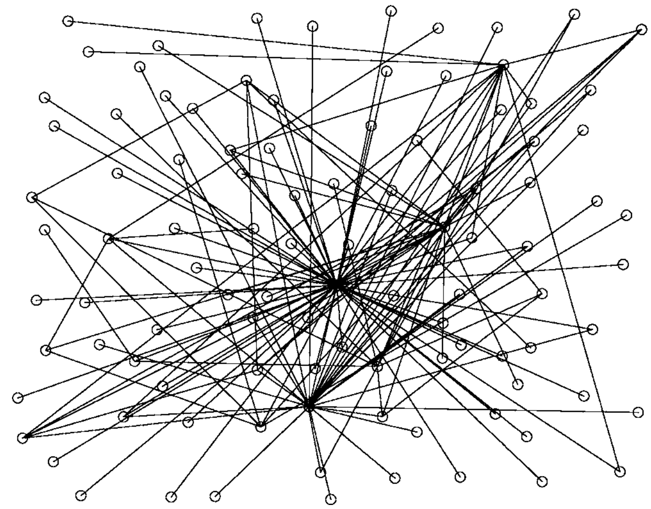

FIG. 1. User network configuration.

III. TRAFFIC ANALYSIS

In a telephone network, "traffic" refers to the accumulated number of communication channels occupied by all users. Different from the user network, the telephone network is a directed complex network, in which each edge has a direction from the caller to the receiver. For each user, the call arrivals can be divided into two categories: *incoming calls* and *outgoing calls*. (The term *call arrival* in a network has been customarily used to refer to both receiving and initiating calls.) Here, incoming calls refer to those being received

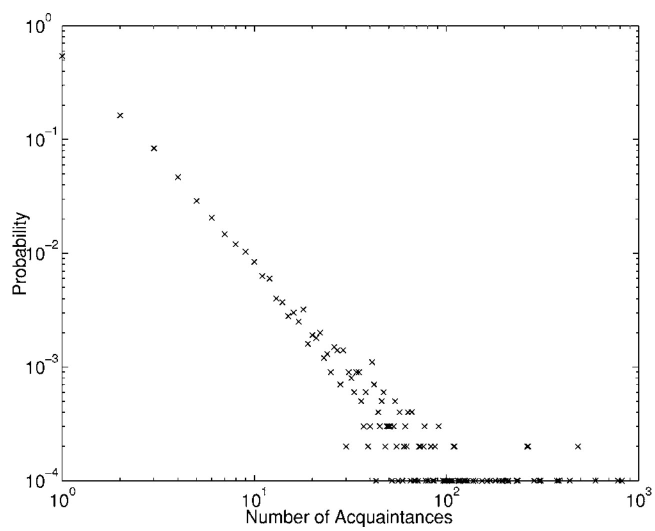

FIG. 2. Power-law distribution of the number of acquaintances showing scale-free property. The mean n_i is 5.

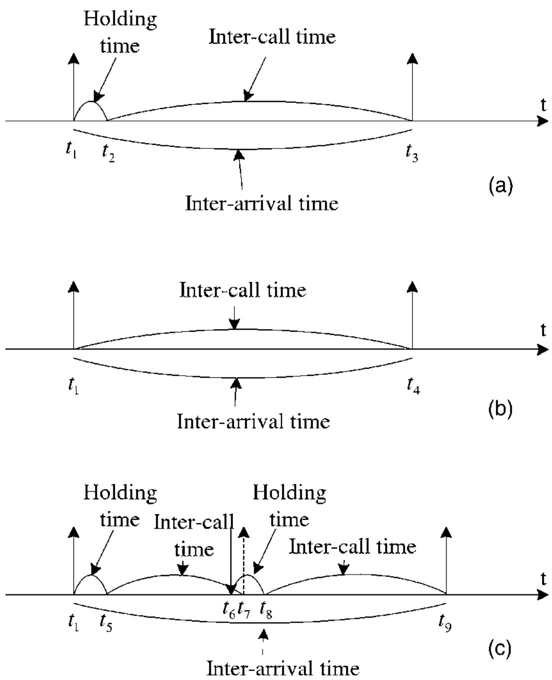

FIG. 3. Three typical calling processes.

by a user, and outgoing calls refer to those being initiated by that user. Since every incoming call for one user must be an outgoing call from another user, we only need to consider outgoing calls from each user when we analyze the network traffic.

Outgoing calls are initiated randomly. If a call arrives and the conversation is successfully established, both the caller and the receiver will be engaged for a certain duration commonly known as *holding time*. The length of the holding time is also a random variable. Thus, the traffic load depends on the rate of call arrivals and the holding time for each call.

Figure 3 shows three typical cases of the calling process.

Case I: When an outgoing call arrives at time t_1, a receiver is randomly selected. If this receiver is idle at that time, a call is successfully established, and the caller and receiver will be engaged for a duration of holding time $(t_2 - t_1)$. The call ends at time t_2. The *intercall time* $(t_3 - t_2)$ is the duration between the end of this call and the beginning of the next outgoing call arrival. Also, the interarrival time is equal to the sum of the holding time and the intercall time. This is the normal calling process, which is depicted in Fig. 3(a).

Case II: For an outgoing call arriving at time t_1, it may be blocked because the receiver is engaged with another call at time t_1 or all channels are occupied at that time. Under such circumstances, a *call blocking* is said to occur. The telephone network is usually considered as a "lossy" system, in which the blocked call simply disappears from the network. In this case the interarrival time is equal to the intercall time (i.e., $t_4 - t_1$, where t_4 is the arrival time of the next outgoing call), as shown in Fig. 3(b).

Case III: In this case, an outgoing call is supposed to take place at time t_7. However, if an incoming call has arrived ahead of it and the conversation is still going on at time t_7, the outgoing call attempt will be cancelled. (Since this call attempt has not been initiated, it is counted as neither a call arrival nor a call blocking.) When the conversation ends at time t_8, another intercall time is assumed before the next outgoing call arrives at time t_9. In this case, the interarrival time is $(t_9 - t_1)$, as illustrated in Fig. 3(c). Of course, at time t_9 it is also possible that the user is engaged with another call. Then, the call arrival at time t_9 will be cancelled, and the interarrival time will become longer accordingly.

In our subsequent analysis, the above three cases of the call arrival process will be considered. Here, we note that in some previous study, simplifying assumptions are made about this process leading to a drastic simplification of the analysis [16,17]. However, we prefer to study the traffic without making any simplifying assumptions on the call arrival process to avoid obscuring the effects of the scale-free user networks.

The holding time and the intercall time are usually modeled by some random variables with exponential distribution. The probability density function (PDF) of the holding time is given by

$$f_1(t) = \frac{1}{t_m} e^{-t/t_m}, \tag{2}$$

where t_m is the average holding time. The PDF of the intercall time is given by

$$f_2(t) = \mu_i e^{-\mu_i t}, \tag{3}$$

where $1/\mu_i$ is the average intercall time. The holding times are distributed identically for all users, but the mean values of the intercall times for different users may be different.

As shown in Fig. 3, the interarrival times for the three cases are different. However, if we examine the traffic over a sufficiently long period of time (e.g., 60 min), we can obtain the average call arrival rate λ_i, which is the average number of call arrivals per unit time, for user i. Thus, the average arrival rate for the whole network is

$$\lambda = \sum_{i=1}^{N} \lambda_i \tag{4}$$

where N is the total number of users in the network.

The volume of traffic carried over a period of time can be found as the sum of the holding times of all call conversations during that time period. A more useful measure of traffic is the *traffic intensity* [6], which is defined by

$$A = \lambda t_m. \tag{5}$$

Thus, traffic intensity A represents the average activity in a period of time. Although A is dimensionless, it is customarily expressed in units of erlangs, after the Danish pioneer traffic theorist Erlang. The maximum traffic intensity of a single channel is 1 erlang, which means that the channel is always busy. Similarly, the maximum traffic intensity in erlangs of a group of channels is equal to the number of channels.

In a telephone network, there are two distinct kinds of traffic: *offered traffic* and *carried traffic*. The offered traffic is the total traffic that is being requested by users. The carried traffic is the actual traffic that is being carried by the network. In practice, due to limited network capacity and some user behavior, the carried traffic is smaller than the offered traffic, and a certain percentage of the offered traffic experiences network blocking.

The telephone network is typically measured in terms of the average activity during the busiest hour of a day [6]. During the busiest hour, each user contributes to a traffic load that is between 0.025 and 0.05 erlang. For an average holding time of 3 to 4 min, there will be one or two calls for a typical user during the busiest hour. To be consistent with the conventional definition of traffic load, two channels will be used for each successfully established call because both users stay in the same network.

IV. TRAFFIC ANALYSIS WITH DIFFERENT USER NETWORKS

We consider a telephone network of N users. Users are located in M subnetworks, each supporting N/M users. (In a fixed telephone network, the subnetworks are the central offices; in a cellular mobile network, the subnetworks are referred to as cells.) Here, for simplicity, we assume that users remain in their subnetworks for the entire period of simulation. (In the case of mobile networks, the traffic behavior may be further complicated by the dynamics of users moving from one subnetwork to another at different times.) Two user network configurations, namely, the fully connected network and scale-free network, are considered.

In a fully connected user network, the effect of each user is assumed to be identical. Thus, each user has the same average arrival rate, i.e., $\mu_i = \bar{\mu}$ and $\lambda_i = \bar{\lambda}$ for all i. In this way, the classical traffic analysis ignores the effect of user network behavior on the traffic.

In a scale-free user network, as mentioned before, the probability that a user with more acquaintances makes/receives a call is higher. Then, the mean value of his intercall time is smaller. In order to show this inequality, we assume

$$\mu_i = p_0 n_i, \tag{6}$$

where p_0 is a constant of proportionality.

The simulated call arrivals and traffic intensities are shown in Figs. 4 and 5 for the fully connected user network and the scale-free user network. The parameters are set as follows:

$$N = 10\,000, \quad M = 4, \quad \bar{n} = \text{average } n_i = 5,$$

$$p_0 = 1/500 \text{ call/(minute acquaintance)},$$

$$\text{average } t_m = 4 \text{ min},$$

$$\bar{\mu} = p_0 \bar{n} = 0.01 \text{ call/min}.$$

It should be noted that we have assumed an infinite network capacity in our simulations. Thus, the call blockings are not consequences of limited network capacity.

Referring to Fig. 3, the minimum value of the interarrival time is equal to the intercall time (case II). Thus, the upper bound of the average arrival rate of a user is given by

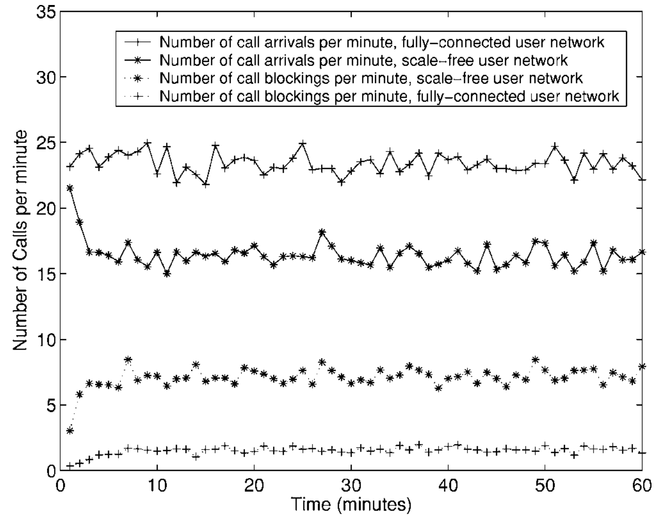

FIG. 4. Call arrivals and call blockings per minute.

$$\lambda_i < \mu_i = p_0 n_i. \tag{7}$$

Figure 6 shows the simulated result for λ_i. From the figure, we see that as n_i increases, the actual arrival rate has more clearance from its upper bound. In general, the upper bound of the average arrival rate of a subnetwork can be derived as

$$\lambda = \sum_i \lambda_i < \sum_i \mu_i = p_0 \sum_i n_i = N p_0 \bar{n}/M = 25 \text{ calls/min}. \tag{8}$$

This upper bound is reached only when each call process assumes that of case II. In practice, such a situation is unlikely. Furthermore, for the scale-free user network, the difference between the upper bound and the actual value of λ_i is larger than that for the fully connected user network. Therefore, the simulated call arrival rate of the scale-free user network is lower than that of the fully connected user network, as shown in Fig. 4.

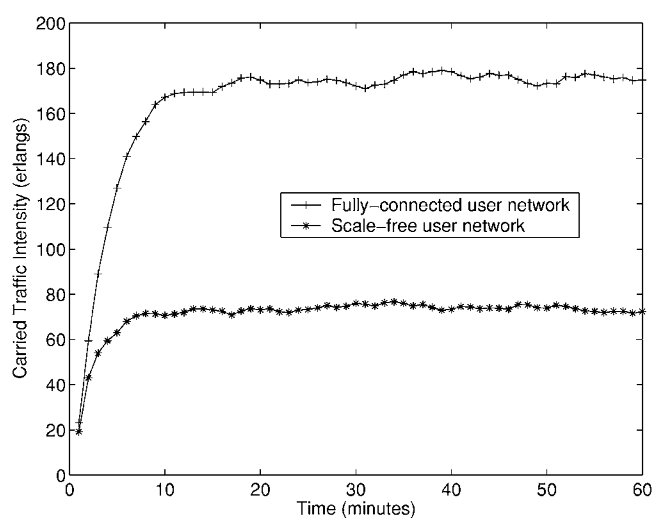

FIG. 5. Carried traffic intensity.

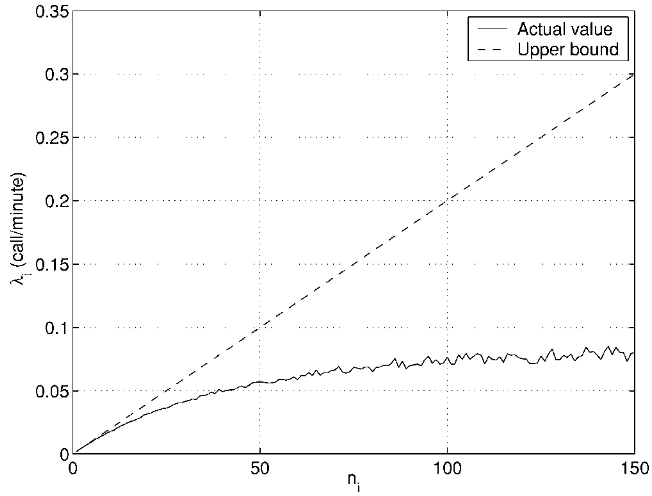

FIG. 6. Actual average arrival rate.

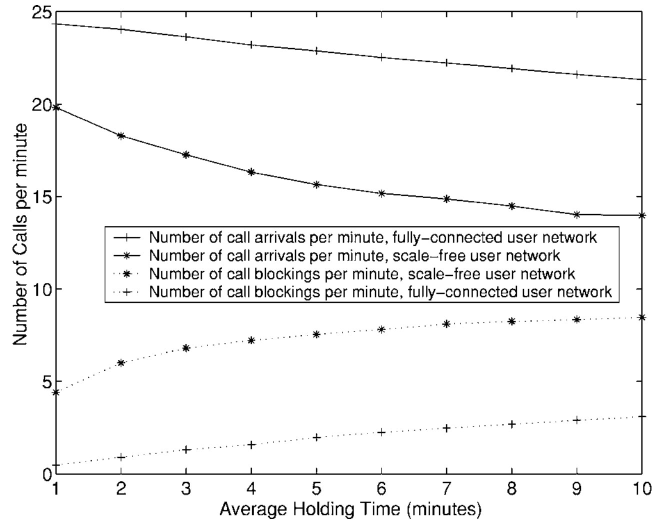

FIG. 7. Call arrivals vs average holding time t_m.

From the simulated traffic data, two important observations can be made. First, the carried traffic intensity of the scale-free user network shows some noticeable differences. Because of the power-law characteristic of the acquaintance distribution, the call arrival rates for different users are not identical. The calls tend to concentrate on a small number of users who have a relatively large number of acquaintances. At the same time, the majority of users, who only have a few acquaintances, are contributing much less to the traffic load. Clearly, network blocking is more severe in the case of the scale-free user network because of the presence of the few very heavy users and each user being able to make at most one call at a time. Thus, in the scale-free user network, the number of call blockings is significantly higher. Second, we note that the scale-free property of the user network has a great impact on the extent of network blocking. We may conclude that increasing the network capacity beyond a threshold value does not help reduce blocking.

V. ROLES OF NETWORK PARAMETERS

The network traffic is determined by three factors, i.e., t_m, p_0, and the acquaintance distribution of the user network. In this section, we investigate the effects of the choice of parameters on the network traffic. For the fully connected network, each user can call any other user. For a fair comparison between the fully connected user network and the scale-free user network, the same set of average intercall time $(1/\bar{\mu})$ and average holding time (t_m) will be used in both user networks.

Figure 7 shows the call arrivals versus the average holding time t_m. In both user network configurations, we observe that by increasing t_m, the number of call arrivals decreases. This can be reasoned as follows. For the usual case I, since the interarrival time is the sum of the holding time and the intercall time, the interarrival time increases as t_m increases and hence the call arrival rate decreases. Also, as t_m increases, case III of the calling process occurs with a higher probability, meaning that more call attempts are cancelled without being counted as a call arrival. Furthermore, with a larger t_m, each call lasts longer. Hence, when an incoming call arrives, the probability that it will be blocked is higher. Thus, the number of blocking increases with t_m.

The carried traffic intensity versus t_m can also be estimated easily using the following equation, and hence simulated results are omitted here:

$$\begin{aligned} A &\approx 2 \times \text{average number of successful new calls} \times t_m \\ &= 2 \times (\text{average number of call arrivals} \\ &\quad - \text{average number of call blockings}) \times t_m. \end{aligned} \tag{9}$$

As an illustration, suppose t_m=4 min. As shown in Fig. 7, for the fully connected user network, the average number of call arrivals is 23.2 calls/min, and the average number of call blockings is 1.5 calls/min. Then the carried traffic intensity can be estimated as $2\times(23.2-1.5)\times 4=173.6$ erlangs, which is consistent with Fig. 5. For the scale-free user network, the average number of call arrivals is 16.3 calls/min and the average number of call blockings is 7.2 calls/min.

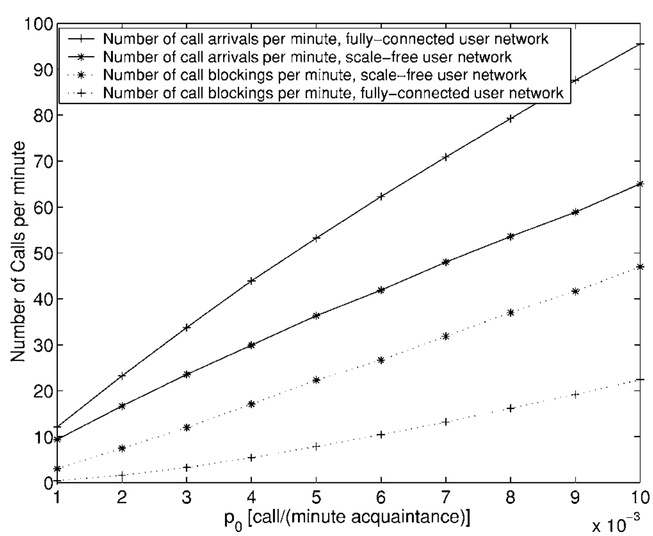

FIG. 8. Call arrivals vs p_0.

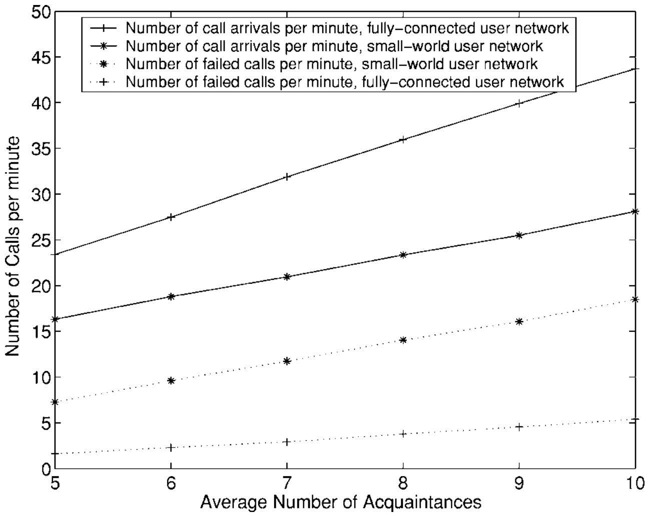

FIG. 9. Call arrivals vs the average number of acquaintances.

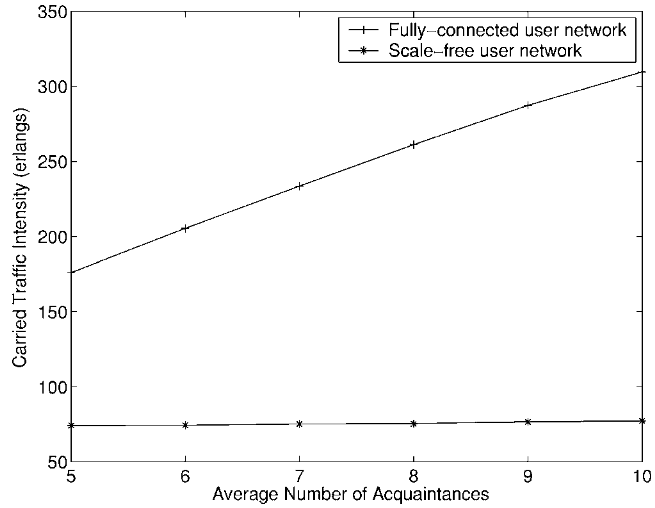

FIG. 10. Carried traffic intensity vs average number of acquaintances.

The carried traffic intensity is $2\times(16.3-7.2)\times 4=72.8$ erlangs, also consistent with Fig. 5.

In the second set of simulations, we vary the proportionality constant p_0 in the scale-free user network. Correspondingly, the value of $\bar{\mu}$ in the fully connected network changes according to $\bar{\mu}=5p_0$ to maintain the same average intercall time $(1/\bar{\mu})$ in both networks. Figure 8 shows the effect of changing p_0 (and hence $\bar{\mu}$ in the case of the fully connected network). For a larger p_0, the probability of initiating a call from any user is higher. The number of call arrivals thus increases. At the same time, the number of call blockings increases for the same reason, although the extent of the increase is smaller than the number of call arrivals. Thus, the difference between them becomes larger as p_0 increases. Therefore, the carried traffic intensity increases with p_0. Comparing the fully connected user network and the scale-free user network, the carried traffic intensity for the scale-free user network grows much slower than that for the fully connected user network, as p_0 increases. In other words, the scale-free user network is less sensitive to the variation of p_0.

Finally, the effect of varying the average number of acquaintances is shown in Fig. 9. For the scale-free user network, we change this parameter by changing γ of the power law distribution. As shown in (1), a smaller γ corresponds to a gentler slope in the power-law distribution, which means that more users have a large number of acquaintances. Hence, the average number of acquaintances increases as γ decreases. Correspondingly, in the fully connected user network, we change the value of $\bar{\mu}$ according to $\bar{\mu}=\bar{n}/500$ to maintain the same average intercall time $(1/\bar{\mu})$ in both user networks. Figure 9 shows how the numbers of call arrivals and call blockings grow with the average number of acquaintances. For the fully connected user network, it is clear that the carried traffic intensity also increases. However, for the scale-free user network, the increase in the number of call arrivals is nearly the same as the increase in the number of call blockings, and the average number of successful new calls is almost fixed. As shown in Fig. 10, the carried traffic intensity of the scale-free user network is not sensitive to the average number of acquaintances.

VI. CONCLUSIONS

In this paper we study the telephone network traffic from a scale-free user network perspective. The simulation results show that the network traffic assuming a scale-free user network is quite different from the traffic assuming a conventional fully connected user network. For the scale-free user network, the traffic load arises mainly from a small number of users who have a relatively large number of acquaintances. This concentration causes a higher blocking probability. At the same time, the majority of users, who have a few acquaintances, contribute much less to the traffic load. In this paper we have studied the effects of different network parameters on the calling process. A possible extension is therefore to characterize a particular society by a set of appropriate network parameters. A city with a bias to particular kinds of activities may have a more or less “uniform” user network (larger or smaller γ), for instance, and this may affect the way its telephone network should be planned.

Our final conclusions are that telephone network traffic is greatly influenced by user behavior, and that network blockings are not likely to be reduced by increasing network capacity (adding extra resources or intensifying investments), which would have been the usual expectation. Thus, a clear, though obvious, lesson to be learned from this traffic analysis is that any strategy for altering the traffic in any manner must take into account the scale-free property of user networks. For instance, network providers may make use of some pricing plans to alter the network traffic, e.g., by penalizing the highly connected users.

ACKNOWLEDGMENT

This work was supported by a research grant (Project No. G-YE38) provided by Hong Kong Polytechnic University.

[1] D. J. Watts and S. H. Strogatz, Nature (London) **393**, 440 (1998).
[2] R. Albert and A. L. Barabási, Rev. Mod. Phys. **74**, 47 (2002).
[3] M. E. J. Newman, SIAM Rev. **45**, 167 (2003).
[4] A. Leon-Garcia and I. Widjaja, *Communication Networks: Fundamental Concepts and Key Architectures* (McGraw-Hill, New York, 2000).
[5] K. Anurag, D. Manjunath, and J. Kuri, *Communication Networking: An Analytical Approach* (Elsevier/Morgan Kaufmann, Amsterdam, 2004).
[6] J. C. Bellamy, *Digital Telephony* 3rd Edition, (Wiley, New York, 2000).
[7] T. Ohira and R. Sawatari, Phys. Rev. E **58**, 193 (1998).
[8] R. V. Solé and S. Valverde, Physica A **289**, 595 (2001).
[9] R. Guimerà, A. Arenas, A. Díaz-Guilera, and F. Giralt, Phys. Rev. E **66**, 026704 (2002).
[10] L. Zhao, Y.-C. Lai, K. Park, and N. Ye, Phys. Rev. E **71**, 026125 (2005).
[11] R. Albert, A.-L. Barabási, H. Jeong, and G. Bianconi, Science **287**, 2115a (2000).
[12] S.-H. Yook, H. Jeong, and A.-L. Barabási, Proc. Natl. Acad. Sci. U.S.A. **99**, 13382 (2002).
[13] L. A. N. Amaral, A. Scala, M. Barthélémy, and H. E. Stanley, Proc. Natl. Acad. Sci. U.S.A. **97**, 11149 (2000).
[14] R. Guimerà and L. A. N. Amaral, Eur. Phys. J. B **38**, 381 (2004).
[15] W. Aiello, F. Chung, and L. Lu, in *Proceedings of 32nd Annual ACM Symposium on Theory of Computing*, ACM, New York, 2000, p. 171.
[16] Siemens Aktiengesellschaft, *Telephone Traffic Theory: Tables and Charts* (Siemens, Berlin, 1970).
[17] R. A. Thompson, *Telephone Switching Systems* (Artech House, Boston, 2000).