Wireless Networks (2021) 27:3613–3619
https://doi.org/10.1007/s11276-019-02233-x

User behavior and user experience analysis for social network services

Rong Bao[1] · Lei Chen[1] · Ping Cui[1]

Published online: 1 January 2020

Abstract

The user behavior characteristics of mobile social network services are of guiding significance to the evaluation of user experience, and test cases and test scenarios should be designed according to user behavior characteristics. Current studies have heavily addressed the action sequence and the frequency distribution of user behavior. However, there is little research on the amount of user action triggered by the communication angle and the fluctuation of the user's action communication performance under different scenarios. This paper analyzes the distribution of data concerning different user actions and tests the waiting time and success rate of different user actions in different scenes. The results suggest that the complex scenarios can consist of some typical user behaviors.

Keywords Quantify of user experience · Social network service · User behavior · Communication scenarios · Mobile internet

1 Introduction

With the development of 4G/5G communication technology, a large number of Social Network Services (SNSs) have emerged. Mobile operators and equipment providers pay more attention to the Quality of User Experience (QoE) of such services, and improve the user experience through more intelligent scheduling strategies [1]. Some scheduling strategies adopt Deep Packet Inspection (DPI) technology to identify the specific services to which packets belong, and then use different scheduling strategies according to different services. This reflects that the communications industry has recognized that different services impact the user experience in different ways.

Therefore, based on the service recognition and efficient traffic analysis method [2–5], an internet service provider can employ many new techniques to improve the user experience and save energy. For example, more complex scheduling strategies are employed to select better access networks [6, 7], and some resource allocation schemes are developed to save energy [8–12] and to more effectively route traffic [13] in complex networks. Some approaches are designed to deal with security problems [14] and measure network performance [15]. To the capacity of a wireless channel in a mobile network, the effective capacity concept is proposed [16] and is employed to measure the probability of a QoS outage [17–19]. Many emerging technologies are employed in SDN networks [20, 21], optical networks [22, 23], sensor networks [24, 25], and networks migration [26]. To evaluate these new techniques, we need to build a large number of test scenarios. However, the traditional evaluation system mainly measures the service quality of a large-scale business using indistinguishable general indicators. However, many test systems are still based on the data flow model [27, 28], and cannot effectively trigger intelligent scheduling strategies; therefore, it is impossible to evaluate the effectiveness of these intelligent scheduling strategies. Many communication enterprises have to use the manual dial-up method to verify a new scheduling strategy, but this method is unable to simulate large-scale scenarios.

A possible approach is to replay the real data packets that are captured from networks in a simulation system

✉ Lei Chen
chenlei@xzit.edu.cn

Rong Bao
11719609@qq.com

Ping Cui
xzcuiping@xzit.edu.cn

[1] Jiangsu Province Key Laboratory of Intelligent Industry Control Technology, Information and Electrical Engineering College, Xuzhou University of Technology, Xuzhou 221018, Jiangsu, China

[15]. These packets might be generated by user actions, such as logins, sending messages, comments, and so on. A complex test scenario should consist of these actions according to user behavior. Therefore, the new problems are how the user behavior affects the QoE, and how to reconstruct user behavior using the captured data packets.

The remainder of paper is divided into five sections. We first introduce the related works about user behavior distribution analysis. Next, we give an approach to build the test scenarios based on the user behavior central distribution. Thirdly, we explore the SNS user behaviors to demonstrate the user behavior distribution such as the operation frequency, the size of the posted message, and the total number of operations. Then, we test how the user behavior affects the end user experience indicators. We conclude in the fifth section.

2 Related works

There are different user actions for different mobile SNSs. It is difficult to analyze the behavioral characteristics of such a large number of user actions and design test cases when building communication scenarios. However, in recent years, several mainstream services have formed a large proportion in the communication market, and thus the research focus can be narrowed.

As shown in Fig. 1, in the four mainstream mobile business markets, a few brands have occupied large market shares [29]. This trend of branding greatly reduces the business scope for user experience evaluations. In addition, the phenomenon of crowd gathering in similar businesses also brings opportunities for user behavior analysis.

Through studying the communication records between operators' terminal nodes, Y.Jin found that users have a strong degree of aggregation [30]. As shown in Fig. 2, the PEARSON correlation can be used to describe the correlation among users, and the correlation is scaled between [− 10, 10]. It is obvious that besides the low degree of user

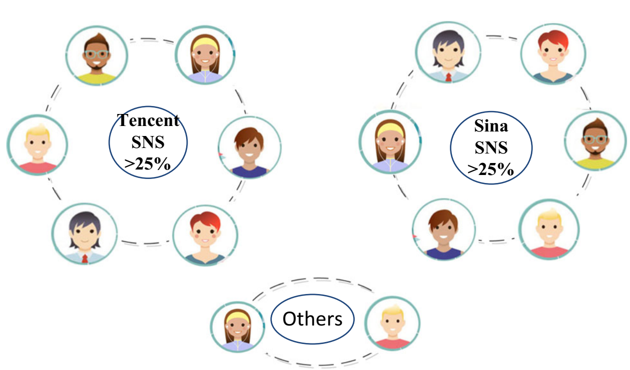


Fig. 1 Market penetration of the main SNSs in China

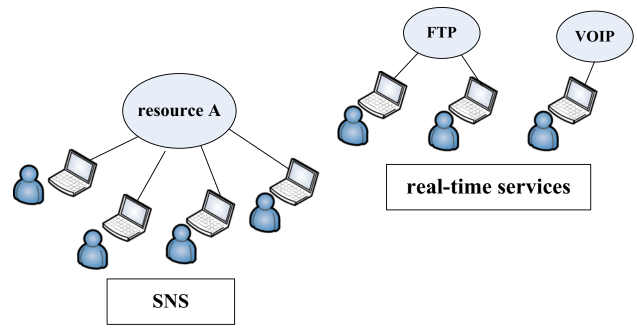


Fig. 2 Level of aggregation of user behaviors

aggregation of real-time services such as FTP large file transfers and VoIP, SNSs show a very high degree of user aggregation. This phenomenon implies that users' behavioral habits are likely to influence each other in the same type of user groups. This aggregation phenomenon will result in the centralized distribution of user behavior, thus forming some typical user behavior. Typical behaviors will have a greater impact on the characteristics of the scenario than other atypical behaviors, and this will make it possible to build complex scenarios using a few typical user behaviors.

Through the analysis of the user behavior characteristics of the main SNSs, some researchers found concentrated distributions of the common actions, action frequency and information length of users.

Facebook is a popular social platform across the world, and thus, it provides a wealth of data on user behaviors. F. Schneider studied the behavioral characteristics of Facebook users [31]. As shown in Fig. 3, there are only five frequent user actions. In addition, the action sequence also has certain rules to follow, as shown in Fig. 4.

Some studies suggest that the traffic model of a traditional communication network can be used as a reference to analyze the communication behaviors of computer

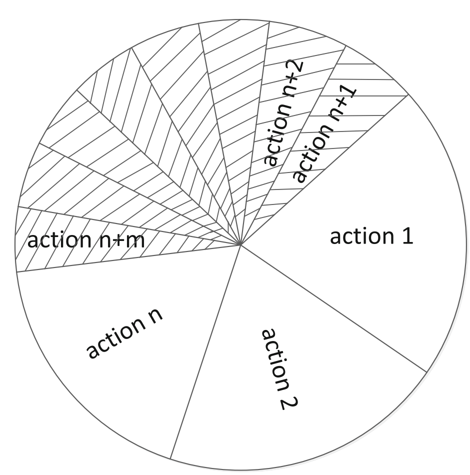


Fig. 3 Frequency actions on Facebook

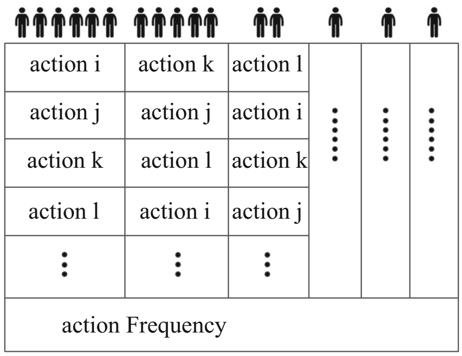


Fig. 4 Action sequences on SNS

network users [32]. This model will provide support for simulating the network communication scenarios with respect to user behaviors and testing the combined influence of user behaviors and scheduling strategies on user experience. In the model, the network user behaviors can be described by the Poisson distribution of parameter λ:

$$P_u(i,t) = \frac{(\lambda t)^i}{i!} e^{-\lambda t} \tag{1}$$

where i is the number of services during t. The amount and length of the data flow of each service follow the geometric distributions of E_f and E_l, respectively:

$$P_f(n) = \frac{1}{E_f}\left(1 - \frac{1}{E_f}\right)^{n-1} \tag{2}$$

$$P_l(n) = \frac{1}{E_l}\left(1 - \frac{1}{E_l}\right)^{k-1} \tag{3}$$

where n and k are the number and length of the data flows, respectively. The network traffic $n_{ij}(T)$ can thus be obtained based on the number of users (T), the number of data flows $f_i(T)$ and the length of data flow $n_{ij}(T)$ during $(0, T)$:

$$N(T) = \sum_{i=1}^{u(T)} \sum_{j=1}^{f_i(T)} n_{ij}(T). \tag{4}$$

Then, the network traffic during $(t, t+\tau)$ can be written as follows:

$$N_t(\tau) = N(t+\tau) - N(t). \tag{5}$$

In this traffic model, the parameter λ depends on the user density and usage habits of the users in the scene, E_f depends on the operation sequences and action frequencies of the users, and E_l is determined by the distribution of user actions. Due to the additivity of the Poisson flow, if there is a centralized distribution of the user behaviors, a few typical action sequences can be used to simulate the entire communication scenario and form impact scenarios that are similar to the real scenarios. In addition, since user habits are somewhat fixed, typical user behaviors have a higher degree of reusability. The model in Fig. 5 can be adopted to collect real user communication data, and form a simple script describing the user behaviors to build a complex communication scenario. Through setting the ratio among the typical users, we build the scenario and illustrate the process in Fig. 6. Then, we can evaluate the QoE in this simulation scenario.

The key point of using the above model to evaluate user experience is whether the user action frequency and the amount of data that is triggered by the action obey a centralized distribution, and whether the user behavior impacts the QoE indicators.

3 User behavior analysis of an SNS

Using the above model, it is possible to reconstruct a real scene with a small number of typical user behaviors if the data length and frequency of user behaviors obey a central distribution. Related studies have found centralized distributions of user behaviors in SNSs. In this paper, we analyze the online times and operation frequencies of SNS users. To observe the impacts of user behaviors on QoE, we also test the appreciable indicators of the end users in different scenarios with different user densities and behaviors.

3.1 Central distribution of the online frequency

To analyze the distribution of the online frequency for an SNS, we survey 35 college students on their use counts per day. As shown in Fig. 7, the results show that this special user group has similar usage habits.

Figure 8 shows the distribution of users' usage times for Weibo on campus. This implies a centralized distribution of usage habits in this special communication scenario.

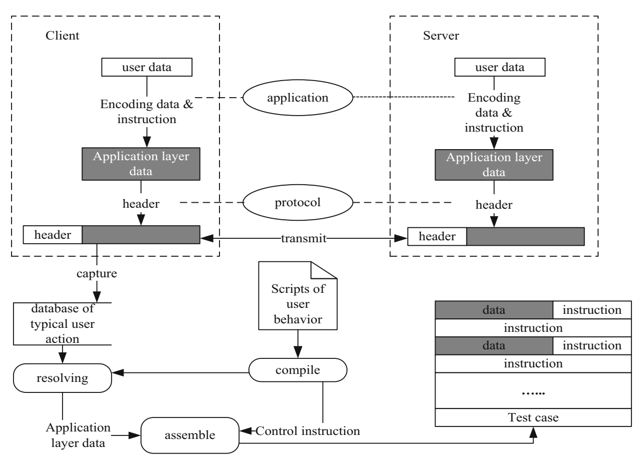


Fig. 5 Building a complex scenario

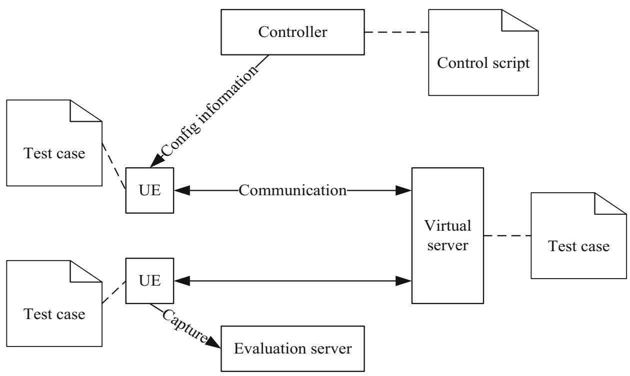


Fig. 6 Testing process

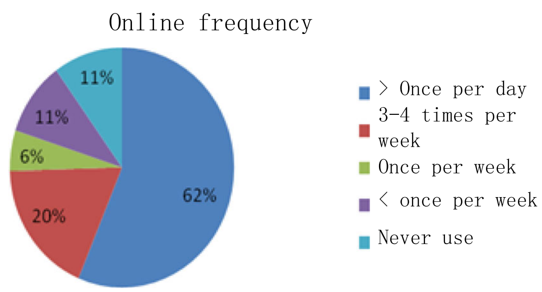


Fig. 7 Online frequency of Weibo

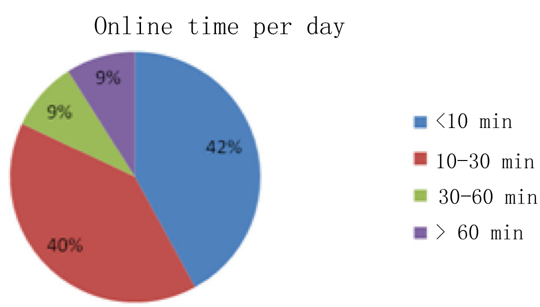


Fig. 8 Online time of Weibo

3.2 Central distribution of user data

We collect the Weibo information of 60,000 users. We compute the average number of posts that these users published per day. The number of tweets per person per day was calculated. As shown in Fig. 9, for most users, they make less than 10 daily posts, and the distribution has a long tail.

In addition, the length of the post is also restricted by the usage habits. In this paper, the length of the 2622 Weibo posts without links is analyzed in Fig. 10, where the vertical coordinate is the number of blog posts and the horizontal coordinate is the length of the posts.

As shown in Fig. 10, the lengths of most posts are between 10 and 50 bytes. The oscillation in Fig. 10 is caused by the double-byte representation of Chinese characters in the computer. When Fig. 11 retains only the double-byte data, the oscillation disappears. This result shows a centralized, heavy-tailed distribution.

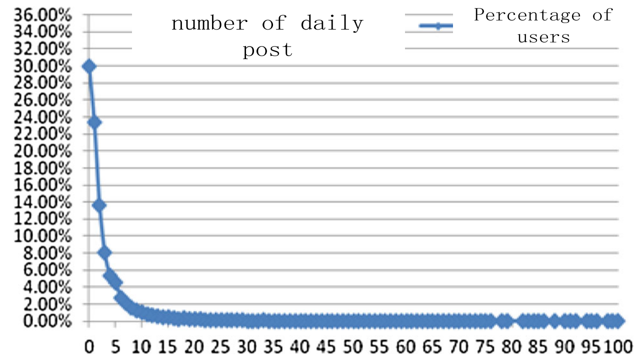


Fig. 9 Number of daily posts

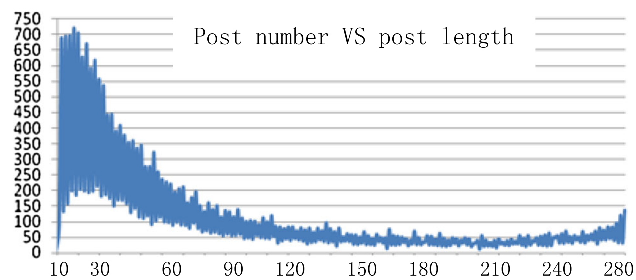


Fig. 10 Distribution of post length

4 Impact of the scenario on QoE indicators

We survey the degree to which the communication scenario affects the user experience. In this paper, the appreciable indicators of the SNS are evaluated in an urban area. The cellular network access is provided by the same communication operators. Therefore, the infrastructure conditions of the tests are similar, and the performance fluctuations should be caused by the behavioral features of the different user groups, including the user density, operation frequency, and other operation habits.

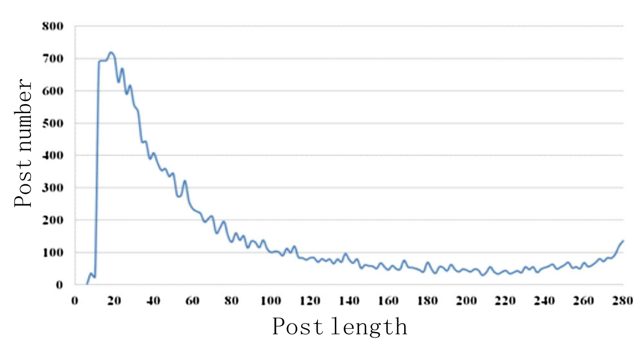


Fig. 11 distribution of post lengths (only even bytes)

To quickly test and analyze various user perception indicators, we define the user perception indicators of SNSs, such as QQ, MSN and Weibo. The algorithm for calculating these indicators is shown in Fig. 12. Firstly, the key attributes are extracted from the network packets. Then, the key messages are identified according to the combination of key attributes. Finally, important attributes such as the time stamps and serial numbers of key messages are used to calculate user perception indicators (e.g., post delay, login success rate, etc.).

By using the above algorithm, the user perception indicator is measured in real scenes, and the significant variances in user groups among scenarios are helpful for observing the impacts of user behavior on the QoE indicators.

Figures 13–16 show the delay test results of QQ messages that are sent in different scenarios. The test results show that the distribution ranges of the delays are approximately the same. However, the number of delays with a significant deviation from the main distribution range is obviously different. In particular, Fig. 16 shows that the delay difference between the different time periods in the same place is also obvious. The test results indicate that user behavior is key factor of communication scenarios, which has a certain impact on the QoE indicators.

In the same area, the communication capacities of the communication operators are more similar. In different scenarios, the group behaviors of users are different, and the usage habits are thus different. Therefore, the degree to which the user behavior affects QoE indicators can be observed. In this paper, we chose to test the delay of posting Weibo comments through a cellular network under different scenarios in a large research and development base. The scenarios include an office area, experiment area, restaurant, lounge, and so on. Except that the restaurant is significantly more densely populated than other areas, all other areas have similar user densities. In addition, the communication operators are known to identify data packets and adopt special scheduling strategies for SNSs.

The test results in Fig. 17 also indicate that in the case of the same user density and a similar communication capacity, user behavior causes delayed fluctuations. Because a delay jitter bigger than 0.1 s is appreciable, the delay fluctuation in Fig. 17 is obvious for users.

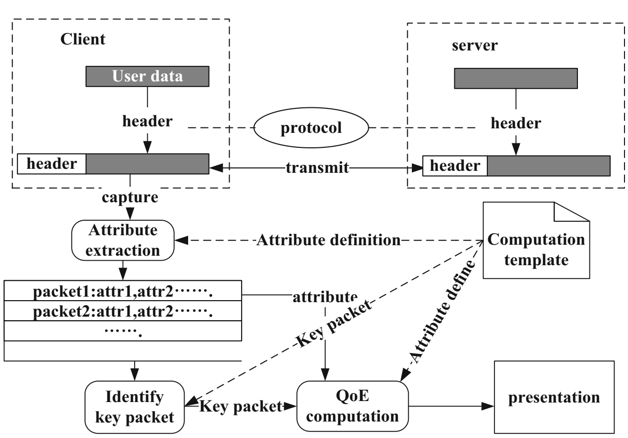


Fig. 12 Framework of the automatic evaluation method

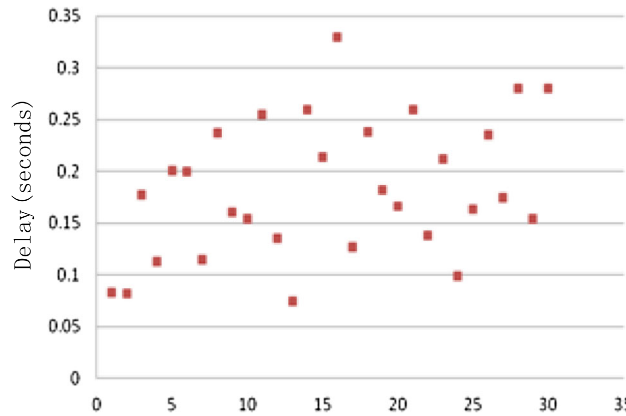


Fig. 13 QQ message delays in a shopping mall

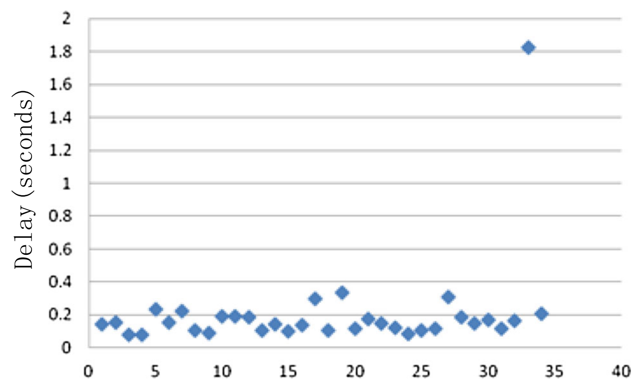


Fig. 14 QQ message delays in an exhibition center

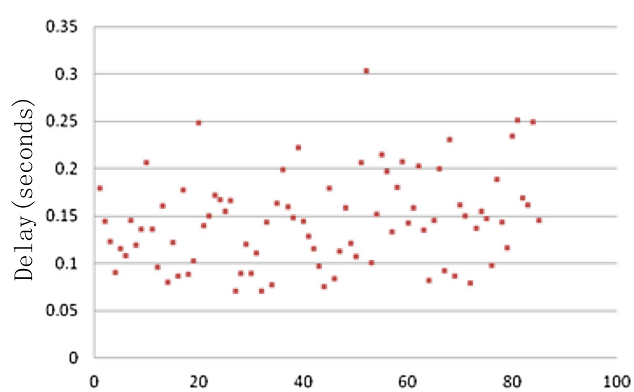


Fig. 15 QQ message delays in the International Finance Centre (indoors)

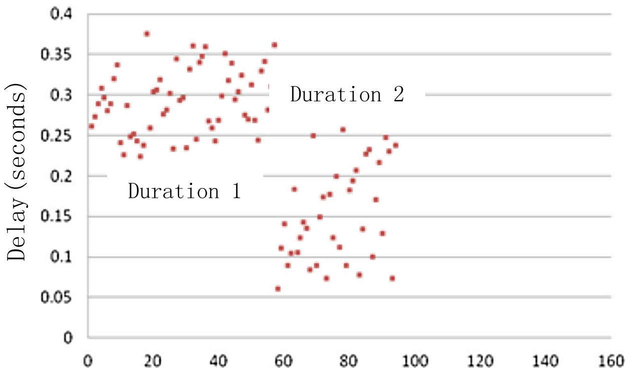


Fig. 16 QQ message delays in the International Finance Centre (outdoors)

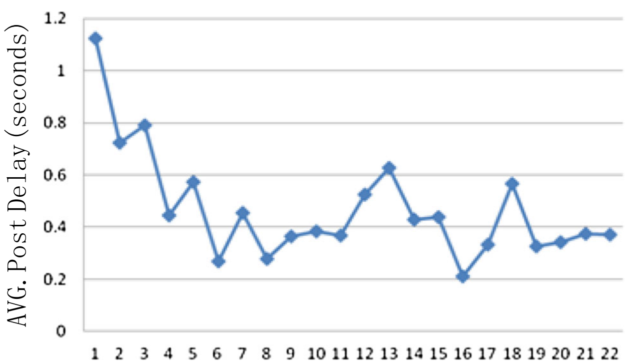


Fig. 17 Average post delays at 22 test points (5 tests for each point)

5 Conclusions

In this paper, we mainly analyze the distribution of user behavior for an SNS, and the degree of impact of user behavior on QoE indicators. The investigation results show the centralized distribution of user behavior. The test results in real scenarios imply that the user behavior actually affects the user experience. Therefore, we propose that a QoE test scenario can be built using the typical user behaviors. The relationship between user behavior and the network load should be further studied to guide the test case design.

Acknowledgements This work is partly supported by the Natural Science Foundation of Jiangsu Province of China (No. BK20161165); the applied fundamental research Foundation of Xuzhou of China (No. KC17072); and the Open Fund of the Jiangsu Province Key Laboratory of Intelligent Industry Control Technology, Xuzhou University of Technology.

References

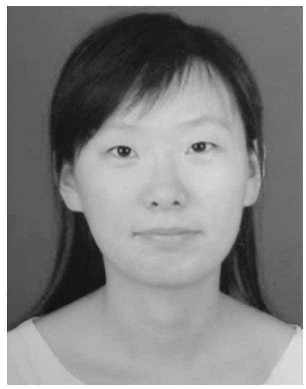
1. Wang, Y., Li, P. L., Jiao, L., et al. (2017). A data-driven architecture for personalized QoE management in 5G wireless networks. *IEEE Wireless Communications*, *24*(1), 102–110.
2. Jiang, D., Wang, W., Shi, L., & Song, H. (2018). A compressive sensing-based approach to end-to-end network traffic reconstruction. *IEEE Transactions on Network Science and Engineering*.. https://doi.org/10.1109/tnse.2018.2877597.
3. Jiang, D., Huo, L., & Song, H. (2018). Rethinking behaviors and activities of base stations in mobile cellular networks based on big data analysis. *IEEE Transactions on Network Science and Engineering*, *1*(2), 1–12.
4. Zhu, J., Song, Y., Jiang, D., & Song, H. (2018). A new deep-Q-learning-based transmission scheduling mechanism for the cognitive Internet of Things. *IEEE Internet of Things Journal*, *5*(4), 2375–2385.
5. Cao, X., Yoshikane, N., Popescu, I., et al. (2017). Software-defined optical networks and network abstraction with functional service design. *IEEE/OSA Journal of Optical Communications and Networking*, *9*(4), C65–C75.
6. Jiang, D., Huo, L., Lv, Z., Song, H., & Qin, W. (2018). A joint multi-criteria utility-based network selection approach for vehicle-to-infrastructure networking. *IEEE Transactions on Intelligent Transportation Systems*, *19*(10), 3305–3319.
7. Wunder, G., Jung, P., Kasparick, M., et al. (2014). 5GNOW: Non-orthogonal, asynchronous waveforms for future mobile applications. *IEEE Communications Magazine*, *52*(2), 97–105.
8. Cui, Q., Gu, Y., Ni, W., & Liu, R. P. (2017). Effective capacity of licensed-assisted access in unlicensed spectrum for 5G: From theory to application. *IEEE Journal on Selected Areas in Communications*, *35*(8), 1754–1767.
9. Jiang, D., Li, W., & Lv, H. (2017). An energy-efficient cooperative multicast routing in multi-hop wireless networks for smart medical applications. *Neurocomputing*, *2017*(220), 160–169.
10. Huo, L., & Jiang, D. (2019). Stackelberg game-based energy-efficient resource allocation for 5G cellular networks. *Telecommunication Systems*, *23*(4), 1–11.
11. Huo, L., Jiang, D., & Lv, Z. (2018). Soft frequency reuse-based optimization algorithm for energy efficiency of multi-cell networks. *Computers & Electrical Engineering*, *66*(2), 316–331.
12. Jiang, D., Zhang, P., Lv, Z., et al. (2016). Energy-efficient multi-constraint routing algorithm with load balancing for smart city applications. *IEEE Internet of Things Journal*, *3*(6), 1437–1447.
13. Wang, F., Jiang, D., & Qi, S. (2019). An adaptive routing algorithm for integrated information networks. *China Communications*, *7*(1), 196–207.
14. Lv, Z., Kong, W., Zhang, X., Jiang, D., Lv, H., & Lu, X. (2019). Intelligent security planning for regional distributed energy internet. *IEEE Transactions on Industrial Informatics*. https://doi.org/10.1109/tii.2019.2914339.
15. Chen, L., Jiang, D., Song, H., et al. (2018). A lightweight end-side user experience data collection system for quality evaluation of multimedia communications. *IEEE Access*, *6*, 15408–15419.
16. Wu, D., & Negi, R. (2003). Effective capacity: A wireless link model for support of quality of service. *IEEE Transactions on Wireless Communications*, *2*(4), 630–643.
17. Shehab, M., Alves, H., & Latva-aho, M. (2019). Effective capacity and power allocation for machine-type communication. *IEEE Transactions on Vehicular Technology*, *68*(4), 4098–4102.
18. Guo, C., Liang, L., & Li, G. Y. (2019). Resource allocation for low-latency vehicular communications: An effective capacity perspective. *IEEE Journal on Selected Areas in Communications*, *37*(4), 905–917.
19. Xiao, C., Zeng, J., Ni, W., Liu, R. P., Su, X., & Wang, J. (2019). Delay guarantee and effective capacity of downlink NOMA fading channels. *IEEE Journal of Selected Topics in Signal Processing*, *13*(3), 508–523.
20. Jiang, D., Huo, L., & Li, Y. (2018). Fine-granularity inference and estimations to network traffic for SDN. *PLoS ONE*, *13*(5), 1–23.

21. Xiong, Y., Li, Y., Zhou, B., et al. (2018). SDN enabled restoration with triggered precomputation in elastic optical inter-datacenter networks. *IEEE/OSA Journal of Optical Communications and Networking*, *10*(1), 24–34. https://doi.org/10.1364/jocn.10.000024.
22. Filer, M., Gaudette, J., Ghobadi, M., et al. (2016). Elastic optical networking in the microsoft cloud. *Journal of Optical Communications and Networking*, *8*(7), A45–A54.
23. Tanaka, T., Inui, T., Kadohata, A., et al. (2016). Multiperiod IP-over-elastic network reconfiguration with adaptive bandwidth resizing and modulation. *Journal of Optical Communications and Networking*, *8*(7), A180–A190.
24. Sun, M., Jiang, D., Song, H., et al. (2017). Statistical resolution limit analysis of two closely-spaced signal sources using Rao test. *IEEE Access*, *5*, 22013–22022.
25. Jiang, M., Jiang, L., Jiang, D., Li, F., & Song, H. (2018). A sensor dynamic measurement errors prediction model of sensors based on NAPSO-SVM. *Sensors*, *18*(233), 1–14.
26. Patri, S. K., Grigoreva, E., Kellerer, W., et al. (2019). Rational agent-based decision algorithm for strategic converged network migration planning. *IEEE/OSA Journal of Optical Communications and Networking*, *11*(7), 371–382.
27. Zhang, H., Wang, Y., Qiu, X., et al. (2015). Network operation simulation platform for network virtualization environment. In *17th Asia-Pacific network operations and management symposium (APNOMS)*. https://doi.org/10.1109/apnoms.2015.7275351.
28. Jiang, D., Xu, Z., & Xu, H. (2015). A novel hybrid prediction algorithm to network traffic. *Annals of Telecommunications*, *70*(9), 427–439.
29. China social networking application user behavior research report. China Internet network information center (2014) **(in Chinese)**.
30. Jin, Y., Duffield, N., Haffner, P., et al. (2011). Can't see forest through the trees. In *Proceedings of 9th workshop on mining and learning with graphs*. San Diego, USA.
31. Schneider, F., Feldmann, A., Krishnamurthy, B., et al. (2009). Understanding online social network usage from a network perspective. In *Proceedings of the 9th ACM SIGCOMM conference on Internet measurement conference*. Chicago, USA.
32. Zhang, S., Zhao, Z., Guan, H., & Yang, H. (2017). A modified poisson distribution for smartphone background traffic in cellular networks. *International Journal of Communication Systems*, *30*(6), e3117.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Rong Bao was born Xuzhou, China, in 1968. She received the Ph.D. degree in control theory and control engineering from the China University of Mining and Technology, China. She is currently a Professor with the School of Information and Electrical Engineering, Xuzhou University of Technology, China. Her research interests include net-work measurement, computer networks, and data mining.

Lei Chen was born Xuzhou, China, in 1980. He received the Ph.D. degree in control theory and control engineering from Tongji University. He is currently an Associate Professor with the School of Information and Electrical Engineering, Xuzhou University of Technology, China. His research interests include Internet traffic model, QoS evaluation, and QoE evaluation.



Ping Cui was born Xuzhou, China, in 1980. She received the Master degree in software engineering from Sichuan University. She is currently an Experimenter with the Xuzhou University of Technology.