Received December 1, 2017, accepted January 4, 2018, date of publication January 16, 2018, date of current version April 4, 2018.

Digital Object Identifier 10.1109/ACCESS.2018.2794354

A Lightweight End-Side User Experience Data Collection System for Quality Evaluation of Multimedia Communications

LEI CHEN[1], DINGDE JIANG[2], (Member, IEEE), HOUBING SONG[3], (Senior Member, IEEE), PING WANG[4], RONG BAO[1], KAILIANG ZHANG[1], AND YI LI[5]

[1]Jiangsu Province Key Laboratory of Intelligent Industry Control Technology, Xuzhou University of Technology, Xuzhou 221000, China
[2]School of Astronautics and Aeronautics, University of Electronic Science and Technology of China, Chengdu 611731, China
[3]Department of Electrical, Computer, Software, and Systems Engineering, Embry-Riddle Aeronautical University, Daytona Beach, FL 32114, USA
[4]The College of Electronics and Information Engineering, Tongji University, Shanghai 201804, China
[5]Xuzhou Construction Machinery Group Research Institute, Xuzhou 221000, China

Corresponding author: Dingde Jiang (jiangdd@uestc.edu.cn)

This work was supported in part by the Nature Science Research Fund of Jiangsu Province of China under Grant BK20161165, in part by the Open Fund of the Jiangsu Province Key Laboratory of Intelligent Industry Control Technology, Xuzhou University of Technology, in part by the European Commission within the Seventh Framework Programme under Grant PIRSES-GA-2013-612652, in part by the National Natural Science Foundation of China under Grant 61571104 and Grant 61071124, in part by the General Project of Scientific Research of the Education Department of Liaoning Province under Grant L20150174, in part by the Program for New Century Excellent Talents in University under Grant NCET-11-0075, in part by the Fundamental Research Funds for the Central Universities under Grant ZYGX2017KYQD170, Grant N150402003, Grant N120804004, and Grant N130504003, and in part by the State Scholarship Fund under Grant 201208210013.

ABSTRACT A lightweight evaluation system, which can be deployed on common user equipment in commercial mobile networks, is proposed for measuring the user experience of multimedia services. The user experience evaluation system can measure the key quality indicators of traditional and emerging services in different scenarios. In contrast to traffic models, this system models typical user behavior to construct complex scenarios of communication networks. In commercial network experiments, the proposed evaluation system achieves stable and efficient performance in complex scenarios, which consist of different types of services and typical user behaviors.

INDEX TERMS Quality of experience, quality of service, mobile communications, multimedia service, end-to-end evaluation.

I. INTRODUCTION AND MOTIVATION

With the rapid technological improvement and infrastructure deployment in wireless communication technologies, including 4G (and later 5G) and high-speed Wi-Fi, there has been dramatic growth in the number of traditional mobile multimedia applications (i.e., mobile video, mobile social network, and VoIP) and emerging mobile applications (i.e., 3D video stream and Tactile Internet) [1]. These diverse content-rich multimedia applications lead to highly complex traffic patterns and face high user requirements on Quality of Experience (QoE). However, the inherent features of wireless communications, such as scarce bandwidth, interference, fading, error-prone channels, diverse access technologies and mobility, lead to a high level of dynamics of available communication resources that can severely deteriorate the quality of mobile multimedia applications with QoE constraints. There are bottlenecks in applying existing wireless techniques for ensuring wireless multimedia QoE. The mismatch between the multimedia quality requirements and the service that is offered by the underlying communication infrastructure makes it greatly challenging to develop mobile multimedia applications over wireless networks. Although network operators and service providers make large investments in improving the system availability, security and performance, mobile multimedia users still suffer from poor QoE frequently. Thus, new and efficient technologies are needed to improve the QoE for wireless multimedia applications. However, it is difficult to evaluate these new technologies from the viewpoint of end user experience.

In this paper, we propose a QoE evaluation system, which is capable to measure the QoE of different services in the commercial communication networks and the laboratory.

Differing from the traditional evaluation systems that employ the traffic model of network side to build communication scenarios, our proposed system simulate the user behaviors of end user side to recreate the real scenarios.

We first briefly introduce some analytical methods for QoE evaluation of multimedia service in this section. A few more-detailed structures of this lightweight system will be discussed in the next section and section three. In section four, the experimental results will be presented, which show that this lightweight system is easy to deploy in a common portable computer. The conclusions of the paper are presented in section five.

A. ANALYTICAL METHODS OF QoE

The concept of QoE (also called quality of user experience) in the field of communications is a product of the so-called experience economy era, in which ''experience'' is viewed as a sort of commodity, or at least determines a commodity value. In the European Qualinet community, QoE was defined as ''the degree of delight or annoyance of the user of an application or service. It results from the fulfillment of his or her expectations with respect to the utility and/or enjoyment of the application or service in the light of the user's personality and current state [2].'' In the communication business, the strategies of Internet Service Providers (ISPs) are undergoing a shift from conventional Customer Relationship Management (CRM) to Customer Experience Management (CEM). CEM requires the service providers to pay more attention to the subjective experiences of their customers or end users, rather than simply the objective technological performance indicators of their communication services. Different from Quality of Service (QoS), which has unified objective evaluation standards, QoE is a subjective evaluation index, which varies from person to person and from time to time. Thus, evaluating QoE requires multidisciplinary investigations since not only engineering but also psychology, cognitive science, and economics affect the results. QoE depends on many technical and non-technical factors. To provide competitive multimedia services to retain existing customers and attract new customers, it is important for service providers to correctly evaluate QoE. Generally, there are two types of methods for assessing and analyzing QoE: subjective methods and objective methods.

Benefiting from existing research efforts, QoE can be partly quantified, and its value can be predicted at some level by using objective methodology and subjective projection models. Subject methods (see e.g., [3], [4]) rely on user-related data and can be carried out by analyzing the responses from actual users toward a certain service, which are collected via polls or tests. In subjective assessment, testers specify services in a controlled environment and assign quality scores based on their own experiences. The results can be represented by metrics such as Mean Opinion Score (MOS). In general, the results more accurately reflect the actual user satisfaction level. Nevertheless, this assessment methodology is time-consuming and excessively costly, which makes it difficult to employ in a large-scale assessment. To reduce cost and promote objectivity, psychophysiology-based QoE measurements, such as electroencephalography and gaze tracking, provide valuable insight into internal physiological processes. A substantial body of research has emerged over the years, with earlier works focusing on auditory and visual quality perception. However, these methods can only be generalized to a limited number of applications [5].

For large-scale commercial networks, objective assessment has been widely adopted by researchers and engineers from both academia and industry. Many such methods have been presented in the literature and target various multimedia services. The rationale behind objective methods is that it is possible to establish mathematical models that project subjective indicators to objective indicators. A structural connection between subjective and objective indicators, which respectively reflect QoS and QoE, is proposed in [6]. Various standardization organizations, including ISO, ITU and ETSI, have also published a series of recommendations that describe the objective QoE indicators [7], [8]. In these objective assessments, QoE is modeled with objectively measurable factors such as input variables, and the value of QoE is the corresponding output. The hybrid assessment uses subjective methods to obtain accurate QoE measurements, establishes the model with respect to objective factors in a laboratory environment, and employs objective assessment methods to calculate QoE outside the laboratory [9]. To narrow the gap between the laboratory environment and the real world, crowd testing is developed to collect subjective experiences and objective indicators in commercial networks at the same time. In addition, a quality analysis of testers is proposed for filtering the feedback of users in large-scale commercial networks [10].

To study the QoE modeling method in large-scale commercial networks, a lightweight QoE evaluation system must be designed to collect the real-time key quality indicators (KQIs) within different complex scenarios of commercial networks.

B. MAIN CHALLENGES

Because of the large perceptual differences among the many mobile services that are running in commercial networks, it is impossible to measure the user experience via a single method for all services. Among various mobile Internet applications, the multimedia services are especially important. For instance, based on the forecast by CISCO, by the year 2019, as one of the most important types of multimedia traffic, video (including TV, video on demand (VoD), Internet and Peer to Peer networks) will account for as much as 80-90% of global consumer traffic. Moreover, by then, traffic from wireless devices will be much higher than that from wired devices [11]. The range that is covered by multimedia services is quite broad since multimedia data basically refer to information in coded form, which is ready for transmission via multiple media rather than a single medium,

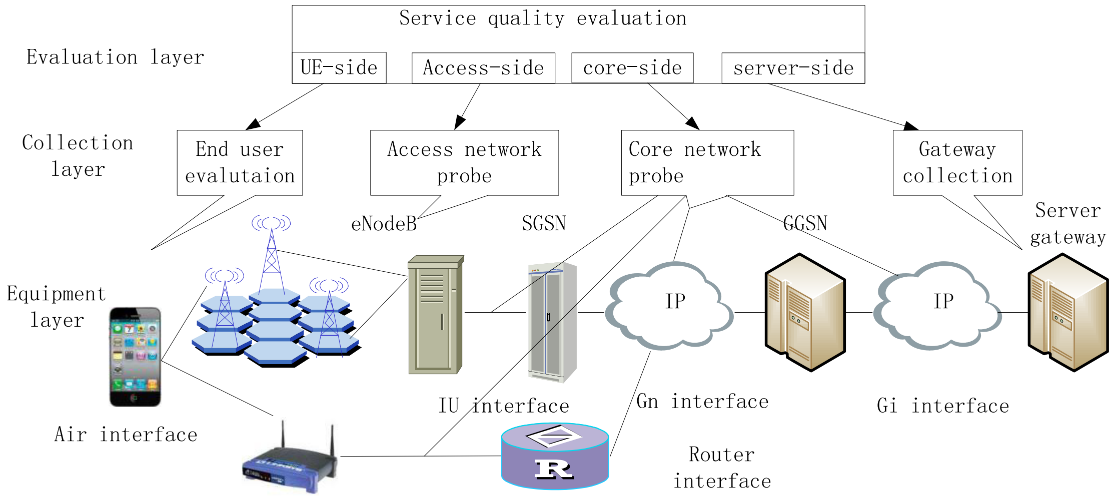

FIGURE 1. Framework of the validation platform for a QoE-driven commercial network.

including characters/text, graphics, still pictures, audio, video, and interactive data [12]. Typical multimedia services include web surfing, video streaming, voice over Internet protocol (VoIP), audio/video broadcasting, teleconferencing, and mobile online gaming.

In general, different multimedia services have various service requirements for different domains, such as the contextual domain, the technological domain and the business domain [13]. However, they can be generally categorized based on their demands on service latency and degrees of data symmetry. Depending on whether a strict latency requirement is imposed (i.e., whether it is time-sensitive), a service can be classified as either a real-time or a non-real-time service. Services can also be classified according to the method of user perception. A comprehensive evaluation system should be able to activate different main mobile services and measure their QoE performances via the corresponding methods.

II. FRAMEWORK OF QoE EVALUATION SYSTEM

For the validation of a QoE-driven network design and improvement, the QoE evaluation and analysis platform should be able to collect information at each key interface in the network. A framework for such a platform is illustrated in Fig. 1. The information that is collected on both the network side and the end-user side will facilitate the identification by researchers of the connection between QoE and network performance, and the location of potential problems.

Currently, there are some well-designed evaluation systems on the core-side and the access-side of communications networks, which only need to evaluate the common key performance indicators (KPIs) of the underlying network. However, these indicators affect the user experience on the user side in very various ways. Thus, it is difficult to assess the QoE indicators using only the KPIs. The subjective experience is affected by a variety of factors from different fields. It has been widely accepted that QoE is a multi-disciplinary metric. Therefore, various information, such as network status, service type, and user type, should be collected in the end device. Then, these objective indicators in a data-driven method could be inferred to describe the user's subjective experience [9].

Our proposed end-to-end evaluation system is implemented on the user side and the service side, which can automatically activate services according to the designed use case script and output their QoE performance based on the data that were collected from the user side. The structure of the evaluation system is illustrated in Fig. 2. The client software is set up in the end user device, the controller send the scripts describing typical user behavior to the clients and server. According these scripts, the clients communicate with the server. At the same time the communication data are recorded in the clients. These data will be sent to the controller for QoE analysis. In this way, the proposed system can generate the network traffic in the end device to describe the characteristics of and user behaviors in communication networks. The network traffic in a large-scale network exhibits periodic change [14]. Thus, we can simply repeat some periodic user behaviors, which can be described by case scripts, to construct typical scenarios of communication networks.

In this system, the designed case scripts that describe the user behaviors are stored in the controller. The case scripts provide the original transmission data, the action frequency, the number of users and the order of execution for simulating real users in real scenarios. The controller activates the data flow between the user equipment and the virtual servers according to the case scripts about the typical habits of Internet users. By using the Application Programming Interfaces (APIs) of mobile services and the collected network packets from the user side, the controller computes the user perceptual indicators and transmits the analysis results to the servers.

A consistent user experience is expected in all types of scenarios, including in office towers, dense residential areas,

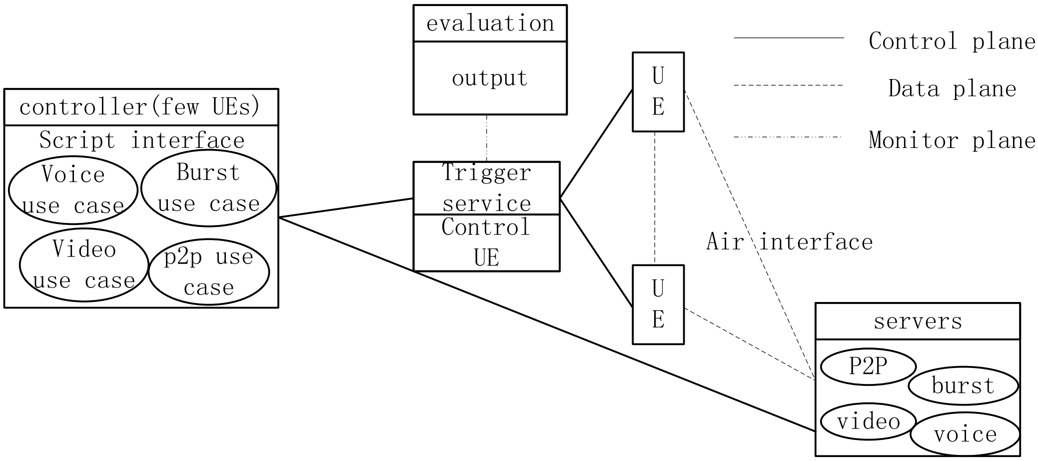

FIGURE 2. Structure of the evaluation system.

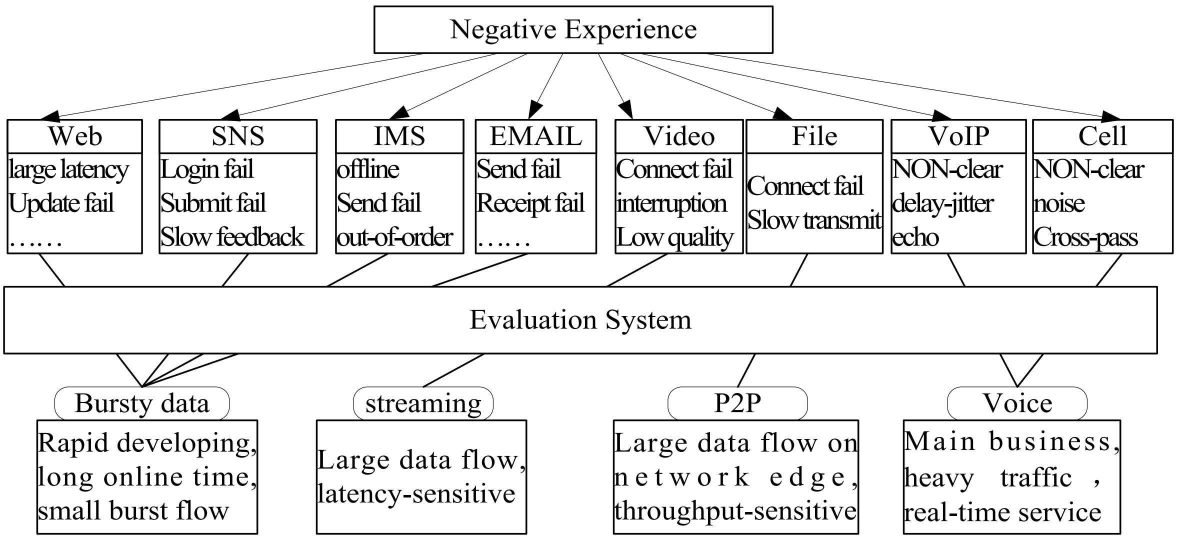

FIGURE 3. Categories of mobile services according to characteristics of user experience.

stadiums, subways, and on highways. Since the key factors that affect QoE may vary over a wide range of scenarios, and users may have different expectations for their service experience in different scenarios, the system must be able to construct all kinds of complex scenarios and evaluate different services. Thus, the evaluation system consists of four subsystems, which can run in parallel on the end device.

III. STRUCTURE OF SUBSYSTEMS

According to the collected customer complaints and the perceptual characteristics of each service, the main mobile services are divided into four categories in the proposed evaluation system, which is illustrated in Fig. 3. Four subsystems are designed, with one corresponding to each category, which can run in parallel to simulate complex real scenarios. The remainder of this section will describe the subsystems.

A. VOICE SUBSYSTEM

Speech transmission is probably the most traditional mobile communication service. The conventional circuit switch (CS) and packet switch (PS), which respectively originate from Public Switched Telephone Network (PSTN) and General Packet Radio Service (GPRS) technologies, are two main voice transportation methodologies that are implemented in 2G and 3G cellular networks. In 4G/LTE cellular networks, the overall mobile voice service is transported over the IP network. Furthermore, the Voice over LTE (VoLTE), which is based on the IP multimedia subsystem (IMS), will deliver both service and media planes as data flows within the LTE data bearer. E-model is a popular approach for estimating the experience quality of a VoIP user using KPIs. It maps the R-Factor to MOS. The expression for the R-Factor is:

$$\mathrm{R} = \mathrm{R}_0 - \mathrm{I}_s - \mathrm{I}_d - \mathrm{I}_{e,\mathrm{eff}} + \mathrm{A}. \quad (1)$$

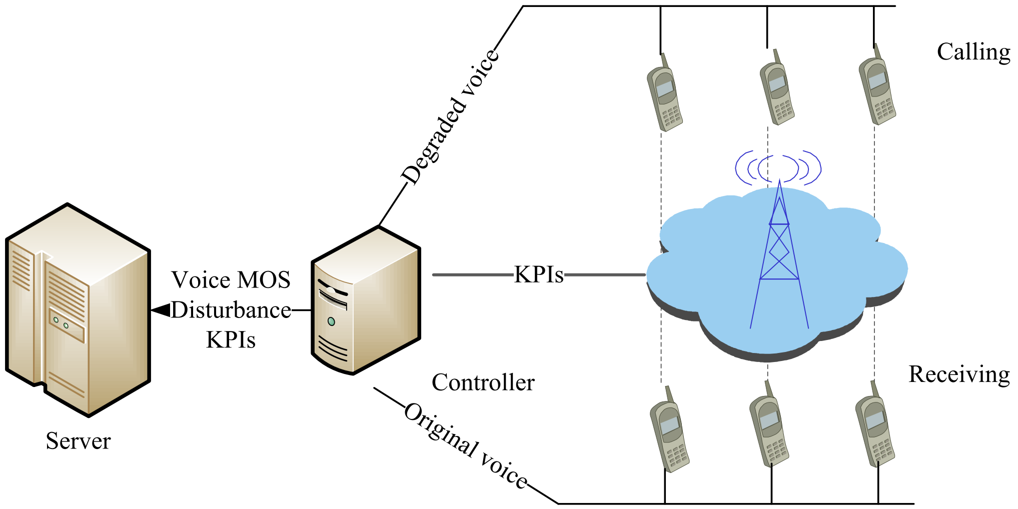

FIGURE 4. Structure of voice subsystem.

where R_0 is the signal-to-interference ratio, I_s is a combination of different impairment factors, and I_d is an impairment factor due to talk and listener echoes and delay contributions. $I_{e,eff}$ denotes the equipment factor, which represents the impairment that is caused by a low-bit-rate CODEC and packet loss. *A* denotes the expectation or advantage factor, and it assumes different values, from 0 to 20 [15].

Based on subjective assessment, the clarity of speech is still the major concern in mobile voice services. The most widely adopted subjective evaluation methods are absolute category rating (ACR), degraded category rating (DCR), and compared category rating (CCR). In these methods, the objective user perception is quantified as the mean opinion score (MOS), which is based on the mean rating for quality evaluation of encoded speech. MOS describes the difficulty levels that end-users have in hearing speech. Based on psychoacoustics, some objective reference methods, such as PESQ [16] and PAMS [17], have been designed for predicting the MOS. To make the deployment more flexible, a non-reference evaluation method is recommended by ITU [18]. Because of the concern that PS cannot guarantee the delay requirement, the delay jitter becomes an important monitoring indicator for QoE in PS-based voice quality evaluation; e-model, which builds the relationship between the network performance indicators and user perception [19]. In addition to clarity, the disturbances of voice, such as noise and single pass, also significantly deteriorate the user experience. An objective evaluation method for such disturbances is proposed in [19]. The evaluation system uses (2) to check the noise:

$$\varsigma < \frac{\left(\sum\limits_{i\in\Lambda_k}\frac{\rho_i\cdot 10^7}{\rho_i+10^5}\right)\cdot\left(\sum_{j\in\Lambda_k}p_j\right)}{\max\left(\sum\limits_{q\in\mathrm{pre}_k}p_q,\sum_{h\in\mathrm{pos}_k}p_h\right)}, \tag{2}$$

where the ρ_i is the asymmetrical factor of the ith frame, which is given by the P.862 algorithm; p_j is the pitch of the jth frame; Λ_k, pre_k and pos_k are the potential noise frame, the previous temporal masking frame and the post-temporal masking frame, respectively; and ς is a threshold for determining whether the noise in area Λ_k is audible. For checking cross passing, we calculate the cosine distance between the original voice and degraded voice as:

$$\boldsymbol{J} = \left\{ j_i \,\middle|\, j_i = \frac{C_k\cdot C'_k}{\|C_k\|\cdot\|C'_k\|},\ \|C_k\| > 10^{11}\wedge\|C'_k\| > 10^{11}\right\}, \tag{3}$$

where the C_k and C'_k are the values of the frequency-domain sampling of the kth frame in the original and degraded voice samples, respectively. In (4), ∂ is the threshold for checking the cross-passing disturbance.

$$\frac{\sum_{j_i\in J} j_i}{|J|} > \partial. \tag{4}$$

The single-passing disturbance is checked through voice activity detection.

In our proposed platform, the structure of the voice subsystem is illustrated in Fig. 4. The controllers send original voice signal to the calling phones and obtain degraded voice data from the receiving phones. The case script will provide the information about calling duration, number of users and execution time. Firstly, the ''Voice Activity Detection'' algorithm is employed to check the single passing. If no voice activity area is detected, the single passing is reported. P.563 is used to decide whether the received signal is background noise or voice. If the received signal is only background noise, the single passing is also reported. P.862 and the algorithm for checking the disturbance are employed to evaluate the user experience. By computing the similarity of frequency spectrum between degraded signal and original signal, the inequality (4) is used to detect the cross passing and echo. The potential noise area is detected by asymmetrical factor of P.862. And then the time realignment algorithm and the inequality (2) are employed to detect the noise. At the same time the MOS of P.862 is also recorded. The controller collects these voice experience information for evaluating the performance of the communication system.

B. VIDEO SUBSYSTEM

Real-Time Streaming Protocol/Real-Time Transport Protocol (RTSP/RTP) [20], HTTP progressive download, and HTTP live streaming are three main video streaming technologies. RTSP/RTP live streaming and HTTP live streaming support real-time video services and adaptive bit-rates. HTTP progressive download only supports non-real-time transport. Because of its lower bit-rate requirement and deployment cost, HTTP progressive download has been adopted by the most web-based video service providers, as a share of the global video service providers.

For mobile video services, the service type, video content and user expectation are important subjective factors that affect the user experience. Depending on different objective factors, the subjective indices influence the user experience in very diverse ways. Moreover, different services have different objective performance indices. For real-time video services, video quality (such as resolution, color rendition, motion portrayal, overall quality, and sharpness), voice quality and audio-video synchronization significantly affect user perceptions of the services. For non-real-time video services, such as HTTP progressive download, the main objective indices that are closely related to user experience include only the code rate, buffering time and interruption frequency because video quality is fixed in the transmission process. For both types of video services, the delay, jitter and loss rate are important network performance indices. In general, the QoE of video service can be formulated as an exponential function of some QoS indicators [21]. The exponential model can be formulated as:

$$\mathrm{QoE} = \mathrm{a}e^{-\beta \mathrm{QoS}} + \gamma, \tag{5}$$

where a, β and γ are coefficients and QoS is a function of a KPI parameter, such as jitter or delay. To include more QoS parameters, the model can be rewritten as:

$$\mathrm{QoE} = \mathrm{e}^{\mathrm{a}_0}\mathrm{e}^{\mathrm{a}_1\mathrm{QoS}_1+\mathrm{a}_2\mathrm{QoS}_2+\ldots+\mathrm{a}_n\mathrm{QoS}_n}. \tag{6}$$

For non-real-time video services, the key to evaluating the impacts of buffering and interruption time on user experience is to convert actual time to psychological time in different typical scenarios. Some projection models have been proposed that achieve over 90% correlation between actual waiting time and user-perceived time [22], [23]. For the evaluation of video quality, ITU has published several recommendations. For example, ITU published ITU-R BT.500 [24] for the subjective assessment of the quality of television pictures. This recommendation has been revised several times and is still widely used. In 2007, ITU published ITU-R BT.1788 [25] for the subjective assessment of video quality in multimedia applications. There are also some objective assessment models for video services. However, there is no rigorous visual system model for accurately describing user perception. Therefore, the peak-signal-to-noise ratio (PSNR) is still widely used to evaluate video quality. Based on PSNR statistics, some machine learning approaches were proposed for evaluating the QoE of HTTP adaptive streaming services [26].

To enhance the QoE, HTTP adaptive streaming has appealing advantages over conventional streaming technologies [27]. Currently, the HTTP adaptive streaming service has held most of the mobile video market share. The structure of the video subsystem is illustrated in Fig. 5. The controller accesses the video service via the mobile equipment. A video player is implemented in the controller, which can play the video that is specified by the case scripts and recode the code rate, the buffering time and the interruption time. In this subsystem, the case script provides the execution order, maximum number of users, code rate and video duration, for simulating different real scenarios. The controller evaluates the quality of user experience by using the method that was proposed in [23].

The formula for the video MOS that is used in this subsystem was proposed in [23]:

$$MOS = = 5 \cdot \frac{5 \cdot e^{-5.7\alpha}}{5} \cdot \frac{5 \cdot e^{-1.607\beta}}{5} \cdot \frac{5 \cdot e^{-0.0416\gamma}}{5} \cdot \frac{5 \cdot \rho^{8.94} e^{-8.94(\rho-1)}}{5}, \tag{7}$$

where α, β, γ, and ρ are the underflow time ratio, loss rate, initial playout delay (seconds), and normalized playout rate, respectively.

C. BURSTY DATA SUBSYSTEM

The bursty data services usually have a long online time, but its actual active time is relatively short compared to the online time. Users can use services of this type to browse web news, communicate with other users, and upload and download pictures and short videos/voices. Typical services include social network services (SNSs), instant messaging (IM) services, mobile office automatic (OA) services, and mobile finance services.

Most bursty data services adopt round-trip communications. Therefore, the response time is an important factor that affects the user experience. However, the response time that users perceive is the duration of the whole interactive process, which usually includes one additional round trip in the transmission layer according to various proprietary application-layer protocols. Depending on different user behaviors, the tolerance levels for response latency are very different. Some latencies that users cannot directly perceive also influence the user experience. Since services of this type usually demand a constant online status for a long duration, the dropping rate should also be used as a performance indicator, even for the non-active duration.

Since different services have different characteristics, it is difficult to establish a unified QoE evaluation standard for bursty data services. Many research efforts focus on some

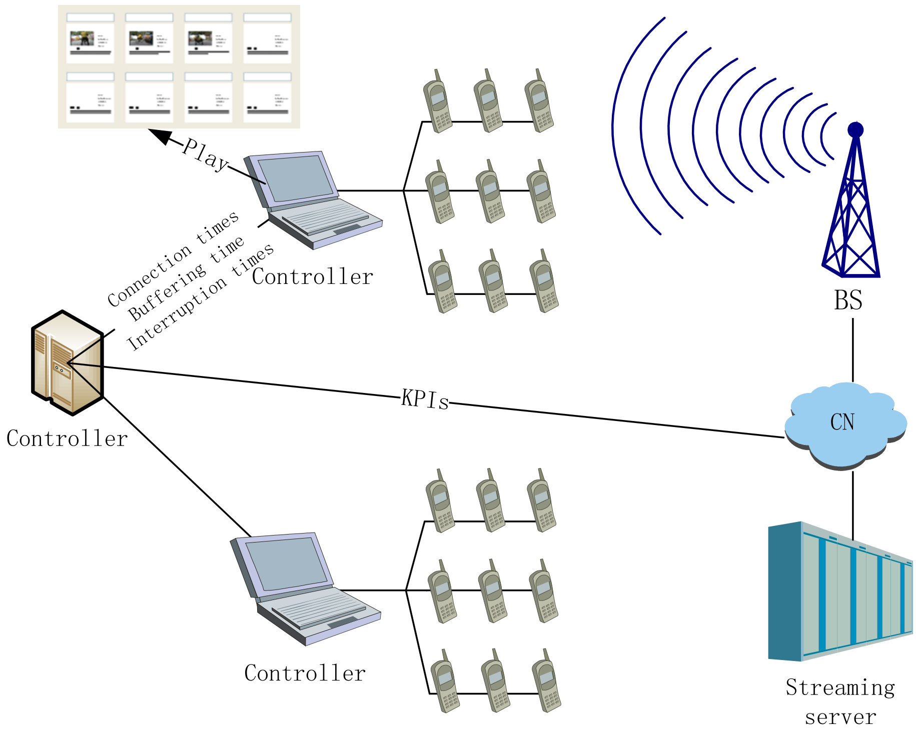

FIGURE 5. Structure of the video subsystem.

common web-based services, and some analytical models have been proposed for projecting the actual waiting time to user experience [28], [29]. Therefore, for other types of services (of which there are many), it is still necessary to explore the impact of waiting time on the psychologically expected and acceptable time.

The latency and success rate of human-perceived round-trip processes in the services are closely related to user experience. We split the round-trip processes of bursty data services into two classes: (1) processes that can be perceived by users. The delay and success probabilities of some actions can be perceived by users, e.g., the delay of publishing a short video on a website and that of submitting a comment on a blog; (2) processes that cannot be perceived by users. Although delays of some actions cannot be directly perceived by users, delays that are too long will cause confusion among users. For instance, in the chat window of an IM service, users cannot perceive the latency of sending messages. However, long delays might cause the messages to arrive at the receivers in the wrong order, thereby leading to a terrible logical mess for the users. Note that since a high dropping rate might lead to a negative experience, the round-trip communications for maintaining a constant presence online also might influence the user experience. There are two key principles for identifying these human-perceived round-trip processes: (1) The whole human-perceived round-trip process should be identified as a basic element in QoE-driven scheduling. This process usually includes several round-trip communication processes in the application layer and network layer, which cannot be perceived by users. (2) Some processes deteriorate QoE with long response times, and some cause negative experiences through failed communications. There are different concerns for different types of processes. Moreover, different types of actions affect user behavior in different ways. For instance, some actions are delay-sensitive and some actions can tolerate relatively long delays. Therefore, the projection models from actual time to psychological time should be very different. Because most bursty data services have their own proprietary application-layer protocols between server and client, understanding those proprietary protocols is another key problem in computing the perceptual indicators. To identify the beginning packet and the end packet of a whole perceptual communication process and compute the user experience indicators, this subsystem requires a script that was proposed in [30].

As illustrated in Fig. 6, the server sends case scripts and commands to both a virtual server and virtual users, which simulate the real server and real users on the other side. The virtual users and virtual server simulate the communication behaviors of real users according to the case scripts. In this subsystem, the case scripts must give the order of the user actions, the action frequencies and the real data packets that result from the user actions. When a virtual user and the virtual server replay the user behavior according to the case script, the controller captures the data packets of the underlying-level networks on the user side and computes the perceptual indicators through analyzing the public and private protocols. In this subsystem, the case scripts are composed

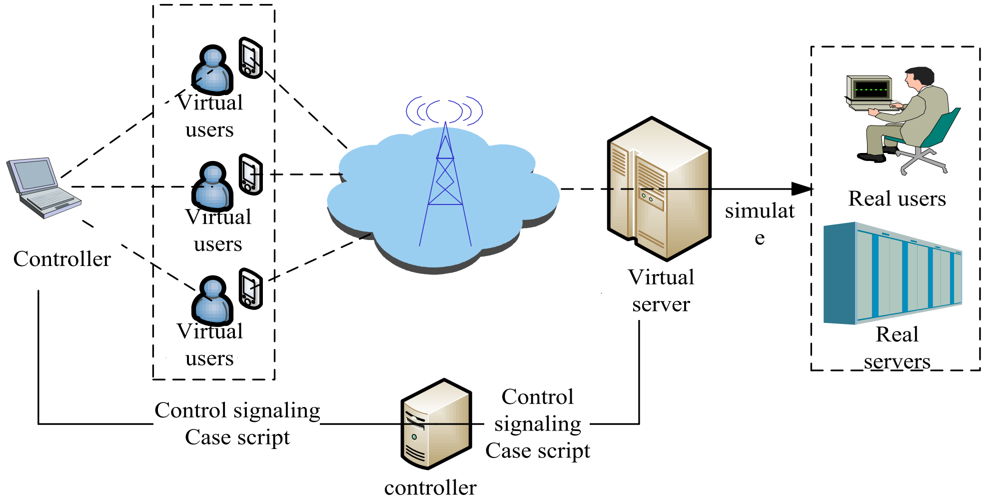

FIGURE 6. Structure of the bursty data subsystem.

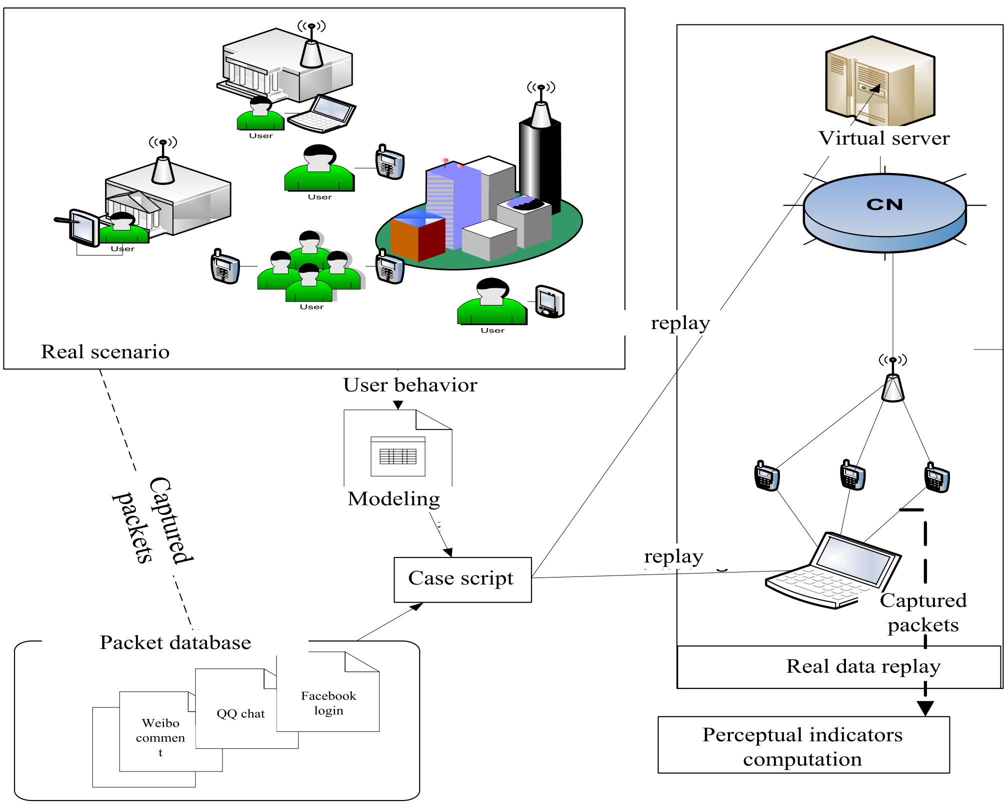

FIGURE 7. Steps of the bursty data service test.

of real data packets, which are collected from real scenarios, and the action script, which describes the real user behaviors. The whole process of simulating real bursty data services is illustrated in Fig. 7.

Most of the bursty services are latency-sensitive applications. According to the Weber-Fechner law [28], the relationship between user experience and waiting time can be expressed by a general formula:

$$MOS = \omega \cdot \ln(\tau \cdot T + \lambda) + \eta, \quad (8)$$

where T is the waiting time, η is the upper limit of the value of MOS and λ is the minimum perceptual delay. In our proposed evaluation system, ω and τ are used to adjust the time sensitivity for different services. For a simple web browsing

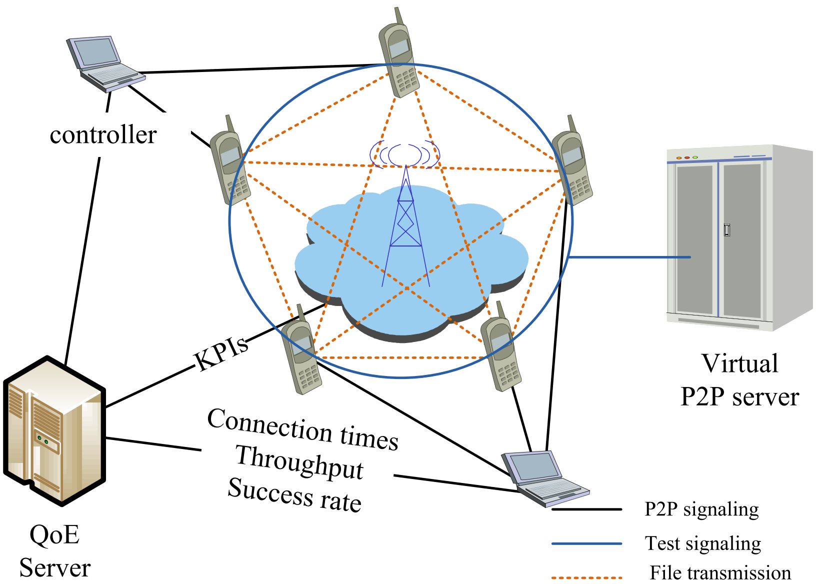

FIGURE 8. Structure of the P2P file transmission subsystem.

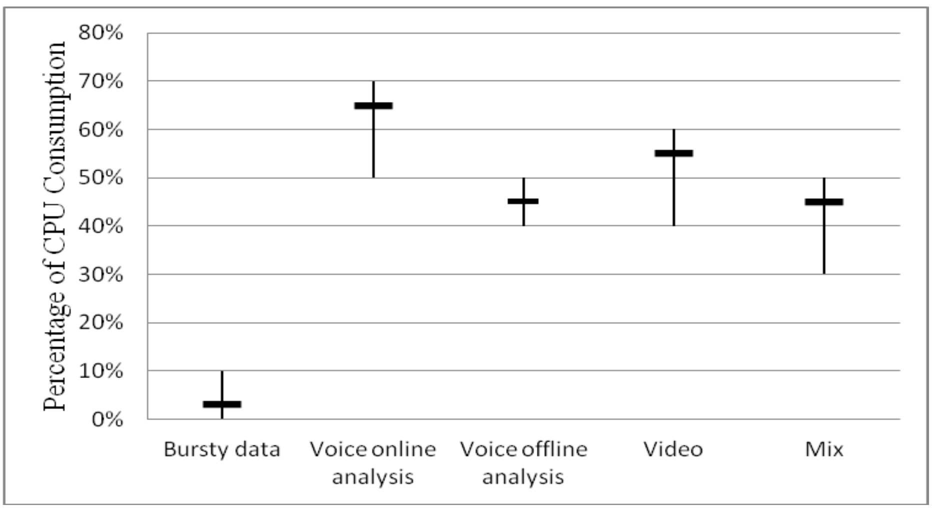

FIGURE 9. CPU consumptions of the terminal system in commercial networks.

service, the MOS is calculated by [29]:

$$MOS = \frac{4\cdot (\ln (ST) - \ln (0.003 \cdot Max + 0.12))}{\ln ((0.003 \cdot Max + 0.12) / Max)} + 5, \quad (9)$$

where ST is the waiting time and Max is the maximum waiting time that the user can tolerate. A simpler formula is proposed in [31]:

$$\mathrm{MOS} = 5 - \frac{578}{1 + (11.77 + 22.61/\mathrm{ST})^2}. \quad (10)$$

D. FILE TRANSMISSION SUBSYSTEM

File transmission does not directly affect the user experience; however, it consumes some bandwidth at the edge of the network. To simulate real scenarios, a P2P subsystem is deployed in the user equipment. The controllers activate the P2P services in the smart terminals and the virtual P2P server according to the case scripts. The connection times, throughput and download success rate are collected in the end side and sent to the service to analyze the user experience. The structure of the P2P subsystem is illustrated in Fig. 8.

For file transmission services, the MOS can be calculated by [32]:

$$MOS_{FTP} = \left\{ \begin{array}{cc} 1 & u < u^- \\ b_1 \cdot \log_{10} (b_2 \cdot u) & u^- < u < u^+ \\ 5 & u^+ < u \end{array} \right\}, \quad (11)$$

where u represents the data rate of the correctly received data, and b_1 and b_2 are obtained from the upper (u^-) and lower (u^+) rate expectations for the FTP services.

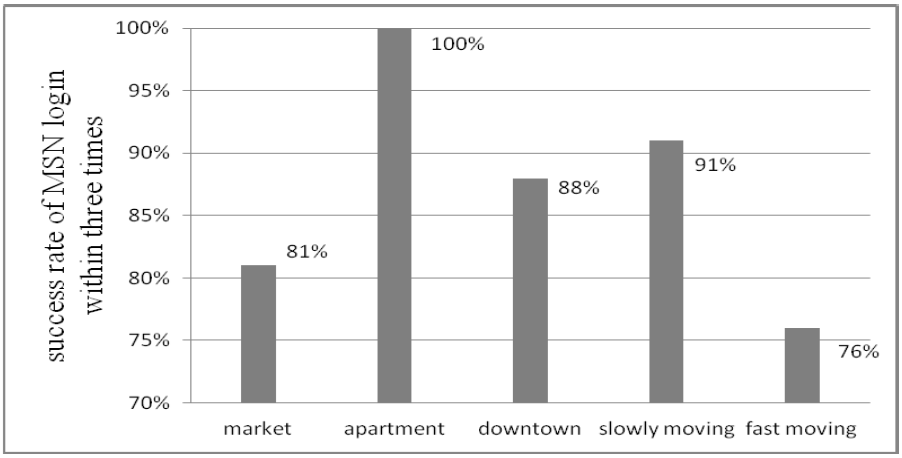

FIGURE 10. Success rates of MSN login in the various environments.

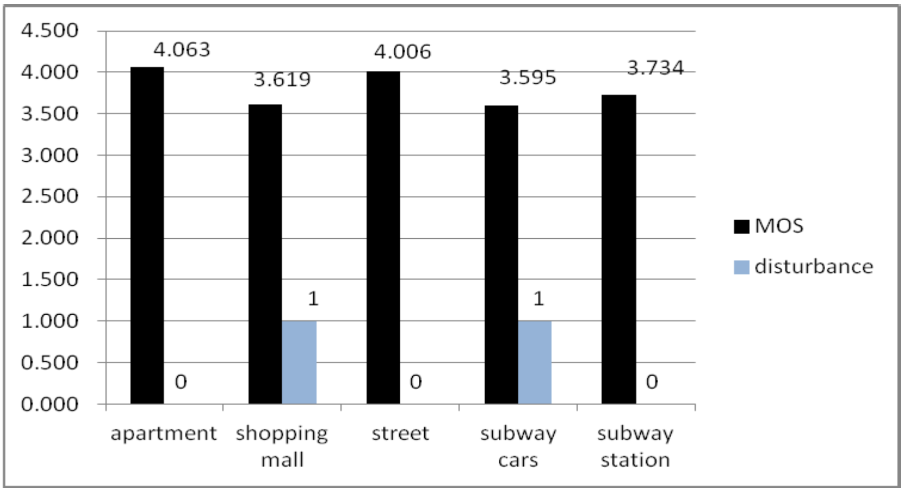

FIGURE 11. MOSs and numbers of voice disturbances in the various environments.

IV. EXPERIMENTS AND ANALYSIS

For commercial network testing, the system must be deployed in the portable computer. The system needs to activate the service, capture the packet and analyze the QoE indicators. Thus, it is essential to determine how many virtual users the portable computer can simulate. In our experiments, we use a PC with 1.6 G Intel dual E2140 CPU and 2 G DDR memory, whose computing power is less than that of a common portable computer. The affordable maximum number of users for each service is listed in Table 1. Fig. 9 shows the maximum, minimum and average percentages of CPU consumption for these test cases.

The CPU consumption of the portable computer is illustrated in Fig. 9. The bursty data services and voice with offline QoE analysis result in low CPU consumption, even with many virtual users. The voice and video services occupy more CPU time. The experiment results of the Mix test case show that a common computer can simulate at least two users who are running all four of these services. Thus, this lightweight evaluation system is easy to deploy in commercial networks.

TABLE 1. Experimental parameters.

Test Name	Number of Users	Test Case	Running Time
Bursty data	120	QQ frequent chat and MSN frequent chat	30 minutes
Voice online analysis	4	play 10 seconds of voice (English, German and Chinese) with online QoE analysis	20 hours
Voice offline analysis	8	play 10 seconds of voice (English, German and Chinese) with offline QoE analysis	30 minutes
Video	20	250 kbps video	2 hours
Mix	2	1 QQ chat, 1 MSN chat, 1 voice service, 1 video service and 1 P2P file download	1 hour

The system is verified in different environments for completeness. We observe the performances of the different

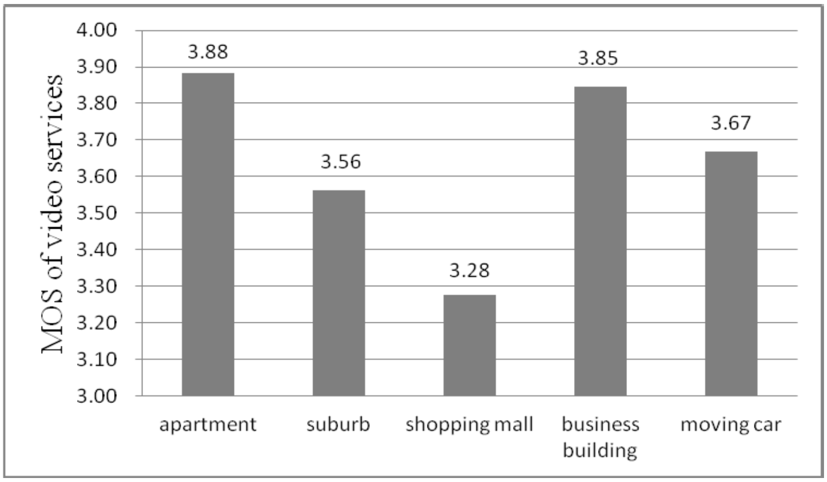

FIGURE 12. MOSs of video services in various environments.

services when using the "Mix" case script. For the bursty data service, we measure the success rate of MSN login for less than two additional connections in five environments. As expected, crowded and fast-moving environments deteriorate the performance. The results are illustrated in Fig. 10. Fig. 11 shows the average MOS value and the number of disturbances of voice service. The test in the running subway cars yields the lowest MOS value, and audible noise occurs once in the subway cars.

The experiment results for video service show the same trend as in Fig. 12. The fewer people that share the bandwidth, the more likely that the communication service obtains a higher score. The test results of this evaluation system are mostly consistent with our daily experience.

V. CONCLUSIONS

In this paper, the main mobile services of commercial networks are divided into four categories according to the perceptual indicators. An evaluation system is implemented to measure the user experiences for these four categories of services. Through a well-designed case script, the terminal system can replay real user behaviors to simulate real scenarios in commercial networks. Then, the terminal system analyzes the network performance in terms of user experience. The server collects the analysis results from the mobile test terminals, which are distributed across various geographies. A test of computational complexity is executed in commercial networks. The test results show that this system can run on a low-specification portable computer. Thus, it is easy to deploy in commercial networks.

ACKNOWLEDGMENT

The authors wish to thank the reviewers for their helpful comments.

REFERENCES

[1] G. Wunder *et al.*, "5GNOW: Non-orthogonal, asynchronous waveforms for future mobile applications," *IEEE Commun. Mag.*, vol. 52, no. 2, pp. 97–105, Feb. 2014.

[2] K. Brunnström *et al.*, "Qualinet white paper on definitions of quality of experience," Qualinet, Novi Sad, Serbia, White Paper 6, Mar. 2013.

[3] *Interactive Test Methods for Audiovisual Communications*, document Recom. P.920, ITU-T, 2000.

[4] M.-D. Cano and F. Cerdan, "Subjective QoE analysis of VoIP applications in a wireless campus environment," *Springer Telecommun. Syst.*, vol. 49, no. 1, pp. 5–15, Jan. 2012.

[5] U. Engelke *et al.*, "Psychophysiology-based QoE assessment: A survey," *IEEE J. Sel. Topics Signal Process.*, vol. 11, no. 1, pp. 6–21, Sep. 2016.

[6] M. Volk, J. Sterle, U. Sedlar, and A. Kos, "An approach to modeling and control of QoE in next generation networks," *IEEE Commun. Mag.*, vol. 48, no. 8, pp. 126–135, Aug. 2010.

[7] *Definitions of Terms Related to Quality of Service*, document Recom. E.800, ITU, 2008.

[8] *Speech and Multimedia Transmission Quality; QoS Aspects for Popular Services in Mobile Networks; Part 2: Definition of Quality of Service Parameters and Their Computation*, document ETSI.TS 102 250-2, 2011.

[9] Y. Wang *et al.*, "A data-driven architecture for personalized QoE management in 5G wireless networks," *IEEE Wireless Commun.*, vol. 24, no. 1, pp. 102–110, Feb. 2017.

[10] R. K. P. Mok, R. K. C. Chang, and W. Li, "Detecting low-quality workers in QoE crowdtesting: A worker behavior-based approach," *IEEE Trans. Multimedia*, vol. 19, no. 3, pp. 530–543, Mar. 2017.

[11] *Cisco Visual Networking Index: Forecast and Methodology (2014-2019)*, Cisco Syst., San Jose, CA, USA, 2015.

[12] *Framework Recommendation for Multimedia Services*, document Recom. F.700, ITU-T, 2000.

[13] K. U. R. Laghari, N. Crespi, and K. Connelly, "Toward total quality of experience: A QoE model in a communication ecosystem," *IEEE Commun. Mag.*, vol. 50, no. 4, pp. 58–65, Apr. 2012.

[14] D. Jiang, Z. Xu, and H. Xu, "A novel hybrid prediction algorithm to network traffic," *Ann. Telecommun.*, vol. 70, no. 9, pp. 427–439, 2015.

[15] *The E-Model: A Computational Model for Use in Transmission Planning*, document Recom. G.107, ITU-T, 2011.

[16] *Perceptual Evaluation of Speech Quality (PESQ): An Objective Method for End-to-End Speech Quality Assessment of Narrow-Band Telephone Networks and Speech Codecs*, document Recom. P.862, ITU-T, 2001.

[17] A. W. Rix, J. G. Beerends, D.-S. Kim, P. Kroon, and O. Ghitza, "Objective assessment of speech and audio quality; technology and applications," *IEEE Trans. Audio, Speech, Language Process.*, vol. 14, no. 6, pp. 1890–1901, Nov. 2006.

[18] *Single-Ended Method for Objective Speech Quality Assessment in Narrow-Band Telephony Applications*, document Recom. P.563, ITU-T, 2004.

[19] L. Chen *et al.*, "A method for detecting noise, single pass, cross pass and echo of mobile phone," in *Proc. Int. Conf. Meas., Instrum. Autom.*, Guilin, China, 2012, pp. 415–421.

[20] *RTP: A Transport Protocol for Real—Time Applications*, document RFC 3550, 2003.

[21] M. Vaser and S. Forconi, "QoS KPI and QoE KQI relationship for LTE video streaming and VoLTE services," in *Proc. 9th Int. Conf. Next Generat. Mobile Appl., Services Technol.*, Cambridge, U.K., 2015, pp. 318–323.

[22] T. Hossfeld, S. Egger, R. Schatz, M. Fiedler, K. Masuch, and C. Lorentzen, "Initial delay vs. interruptions: Between the devil and the deep blue sea," in *Proc. 4th Int. Workshop Quality Multimedia Exper.*, Yarra Valley, VIC, Australia, 2012, pp. 1–6.

[23] M. Li, "QoE-based performance evaluation for adaptive media playout systems," *Adv. Multimedia*, vol. 2013, Jan. 2013, Art. no. 152359.

[24] *Methodology for the Subjective Assessment of the Quality of Television Pictures*, document Recom. BT 500-11, ITU-R, 2002.

[25] *Methodology for the Subjective Assessment of Video Quality in Multimedia Applications*, document Recom. BT.1788, ITU-R, 2007.

[26] F. Wang, Z. Fei, J. Wang, Y. Liu, and Z. Wu, "HAS QoE prediction based on dynamic video features with data mining in LTE network," *Sci. China Inf. Sci.*, vol. 60, no. 4, pp. 042404:1–042404:14, Apr. 2017.

[27] M. Seufert, S. Egger, M. Slanina, T. Zinner, T. Hoßfeld, and P. Tran-Gia, "A survey on quality of experience of HTTP adaptive streaming," *IEEE Commun. Surveys Tuts.*, vol. 17, no. 1, pp. 469–492, 1st Quart., 2015

[28] S. Egger, T. Hossfeld, R. Schatz, and M. Fiedler, "Waiting times in quality of experience for Web based services," in *Proc. 4th Int. Workshop Quality Multimedia Exper.*, Yarra Valley, VIC, Australia, 2012, pp. 86–96.

[29] *Estimating End-to-End Performance in IP Networks for Data Applications*, document Recom. G.1030, ITU-T, 2005.

[30] L. Chen *et al.*, "A replay approach for remote testing user experience of mobile bursty data application," *Int. J. Online Eng.*, vol. 9, no. 7, pp. 18–23, Nov. 2013.

[31] P. Ameigeiras, J. J. Ramos-Munoz, J. Navarro-Ortiz, P. Mogensen, and J. M. Lopez-Soler, "QoE oriented cross-layer design of a resource allocation algorithm in beyond 3G systems," *Comput. Commun.*, vol. 33, no. 5, pp. 571–582, Mar. 2010.

[32] D. Tsolkas, E. Liotou, N. Passas, and L. Merakos, "A survey on parametric QoE estimation for popular services," *J. Netw. Comput. Appl.*, vol. 77, pp. 1–17, Jan. 2017.

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 a Lecturer 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.

DINGDE JIANG (S'08–M'09) received the Ph.D. degree in communication and information systems from the University of Electronic Science and Technology of China, Chengdu, China, in 2009. From 2013 to 2014, he was a Visiting Scholar with the Department of Computer Science and Engineering, University of Minnesota, Minneapolis, MN, USA. He was a Professor with the School of Computer Science and Engineering, College of Information Science and Engineering, Northeastern University, Shenyang, China, in 2017. He is currently a Professor with the School of Astronautics and Aeronautics, University of Electronic Science and Technology of China. His research interests include network measurement, modeling and optimization, performance analysis, network management, network security in communication networks, particularly in software-defined networks, information-centric networking, energy-efficient networks, and cognitive networks.

His research was supported by the National Science Foundation of China, the Program for New Century Excellent Talents in University of Ministry of Education of China, and so on. He has received the best paper awards at several international conferences. He served as a technical program committee member for several international conferences. He has been serving as an Editor for one international journal.

HOUBING SONG (M'12–SM'14) received the Ph.D. degree in electrical engineering from the University of Virginia, Charlottesville, VA, USA, 2012. He served on the Faculty of West Virginia University from 2012 to 2017. In 2007, he was an Engineering Research Associate with the Texas A&M Transportation Institute. In 2017, he joined the Department of Electrical, Computer, Software, and Systems Engineering, Embry-Riddle Aeronautical University, Daytona Beach, FL, USA, where he is currently an Assistant Professor and the Director of the Security and Optimization for Networked Globe Laboratory. He is an author of over 100 articles. His research interests include cyber-physical systems, cybersecurity and privacy, Internet of Things, edge computing, big data analytics, unmanned aircraft systems, connected vehicle, smart and connected health, and wireless communications and networking. He is an editor of four books the *Smart Cities: Foundations, Principles and Applications* (Hoboken, NJ, USA: Wiley, 2017), the *Security and Privacy in Cyber-Physical Systems: Foundations, Principles and Applications* (Chichester, U.K.: Wiley–IEEE Press, 2017), the *Cyber-Physical Systems: Foundations, Principles and Applications* (Boston, MA, USA: Academic Press, 2016), and the *Industrial Internet of Things: Cybermanufacturing Systems* (Cham, Switzerland: Springer, 2016). He serves as an Associate Technical Editor for the *IEEE Communications Magazine*.

Dr. Song is a Senior Member of the ACM. He was a very first recipient of the Golden Bear Scholar Award, the highest campus-wide recognition for Research Excellence from the West Virginia University Institute of Technology in 2016.

PING WANG received the Ph.D. degree from the Department of Computer Science and Engineering, Shanghai Jiao Tong University, in 2007. He is currently an Associate Professor with the Department of Information and Communication Engineering, Tongji University. His main research interests include routing algorithms and resource allocation of wireless networks.

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 network measurement, computer networks, and data mining.

KAILIANG ZHANG is currently an Experimenter with the Xuzhou University of Technology.

YI LI is currently a Senior Engineer with the Xuzhou Construction Machinery Group Research Institute.