A DATA-DRIVEN ARCHITECTURE FOR PERSONALIZED QOE MANAGEMENT IN 5G WIRELESS NETWORKS

YING WANG, PEILONG LI, LEI JIAO, ZHOU SU, NAN CHENG, XUEMIN (SHERMAN) SHEN, AND PING ZHANG

ABSTRACT

With the emergence of a variety of new wireless network types, business types, and QoS in a more autonomic, diverse, and interactive manner, it is envisioned that a new era of personalized services has arrived, which emphasizes users' requirements and service experiences. As a result, users' QoE will become one of the key features in 5G/future networks. In this article, we first review the state of the art of QoE research from several perspectives, including definition, influencing factors, assessment methods, QoE models, and control methods. Then a data-driven architecture for enhancing personalized QoE is proposed for 5G networks. Under this architecture, we specifically propose a two-step QoE modeling approach to capture the strength of the relationship between users and services. Thereafter, the preferences of a user is introduced to model the user's subjectivity toward a specific service. With the comprehension of users' preferences, radio resources can be distributed more precisely. Simulation results show that overall QoE can be enhanced by 20 percent, while 96 percent of users have an improved QoE, which validates the efficiency of the proposed architecture.

INTRODUCTION

Triggered by the exponential growth of mobile data and the proliferation of smart devices, the investigation of the new fifth-generation (5G) cellular networks has become a focus of both academia and industry. 5G networks are expected to provide a wide range of mobile applications and service with quality of service (QoS) guarantees. Video-based services are one of the most critical services supported by 5G networks. It is estimated that in 2019, 80 percent of Internet traffic will be video traffic. In order to improve video quality, many efforts have been made to optimize the QoS of the video transmission system, such as video compression optimization and network resource allocation [1, 2]. In addition to monitoring and controlling QoS, it is more crucial to assess video quality from users' perspective, which is known as quality of experience (QoE). With the development of video transmission systems and consumer video technologies, a better understanding of the user experience is urgent. There is a consensus that the user QoE is becoming one of the primary focuses, and will be the key to the competition in 5G wireless networks. Therefore, it is critical to investigate the needs of users, measurements of the user's subjective feelings, and methods to improve the user QoE.

QoE is highly related to user acceptability. In general, QoE is based on the quality of interactions between users and applications, while QoS depends on the quality of interactions between applications and networks. Therefore, the technologies based on QoE can satisfy the requirements of end users in a better manner than those based on QoS. In order to deal with new issues caused by enriched services, QoE parameters are designed to represent video quality, which includes video structural similarity (VSSIM), video quality metric (VQM), and moving picture quality metric (MPQM). However, those parameters are still not adequate, especially for the service context and human subjective factors. Consequently, one of the main research targets of 5G is to identify more efficient parameters for the users. This tendency makes QoS less capable of measuring the service quality perceived by users for the emerging communication services. Instead, QoE, which is a direct measurement of human perception of the quality of communication services, is more promising and beneficial for future communication systems.

In 5G/future networks, QoE management will be a challenging task, as QoE is expected to be well and autonomously managed for each user and service [3]. As illustrated in Fig. 1, a soccer game video is played to Bob, who is a soccer fan, and Lin, who is not. Obviously, the fan's expectation for the video quality is higher. Under the limitation of network resources, it should be advantageous to allocate more resources to Bob. Furthermore, the estimation and resource allocation should be completed quickly. As a result, user/service specific modeling and sufficient computing capacity are required for such "personalized QoE." Fortunately, with the development of cloud computing, online large-scale modeling

Ying Wang , Peilong Li , and Ping Zhang are with Beijing University of Posts and Telecommunications.

Lei Jiao is with the University of Agder.

Zhou Su is with Shanghai University.

Nan Cheng and Xuemin (Sherman) Shen are with the University of Waterloo.

Digital Object Identifier: 10.1109/MWC.2016.1500184WC

and real-time computations are available, such as Google advertising and Netflix movie recommendations. Moreover, with the permission of users, the historical usage and preference information of each user can be preserved and analyzed on the cloud. Therefore, personalized QoE is achievable within the architecture of 5G networks.

In this article, we summarize the state of the art of QoE and the corresponding challenges in 5G networks. Then a new architecture for QoE management is proposed. The remainder of this article is organized as follows. In the following section, the state of the art of QoE is presented. The challenges of QoE in 5G are then discussed. Following that a data-driven architecture for personalized QoE in 5G networks is presented and examined, followed by the conclusion.

State of the Art

This section elaborates on the state of the art of QoE from a bottom-up perspective. First, the definition of QoE is introduced. Second, we explore the factors that may influence QoE. Then, we discuss QoE assessment methods. Finally, QoE models and control schemes are summarized.

Definition of QoE

Currently, there is no strict definition of QoE. At the International Telecommunication Union (ITU-T), QoE is defined as "*the overall acceptability of an application or service, as perceived subjectively by the end user*." Recently, 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 (w.r.t.) the utility and/or enjoyment of the application or service in the light of the user's personality and current state*." [4]. However, the general understanding of QoE is the same to a large extent: QoE is a new measurement approach for communication services, and it is determined by the interactions between users and services.

Influencing Factors

In [5], possible influencing factors related to QoE are categorized into four domains: human, technology, business, and context. In more detail, factors such as age, gender, memory, attention, satisfaction, and human roles are within the scope of the human domain. The business domain includes advertising, pricing, churn rate, and so on. In the context domain, there are factors such as global position service (GPS) data, temporal information, and social factors. The technological domain covers all technological aspects of the service life cycle from service design to delivery.

It should be emphasized that not all factors mentioned above were studied sufficiently. First, some factors are too sophisticated to be quantified, especially subjective factors such as user interests, user expectations, service ease of use, and so on. Therefore, analyzing these factors in QoE is difficult. Second, some factors are too complicated to obtain in real-life environments, such as brain waves, user memories, and some contextual factors (e.g., alone or with people, quiet or noisy). Third, the relation between those factors and QoE is too complicated to be measured and modeled accurately, and more research effort is required. We summarize some of the studied factors in Table 1.

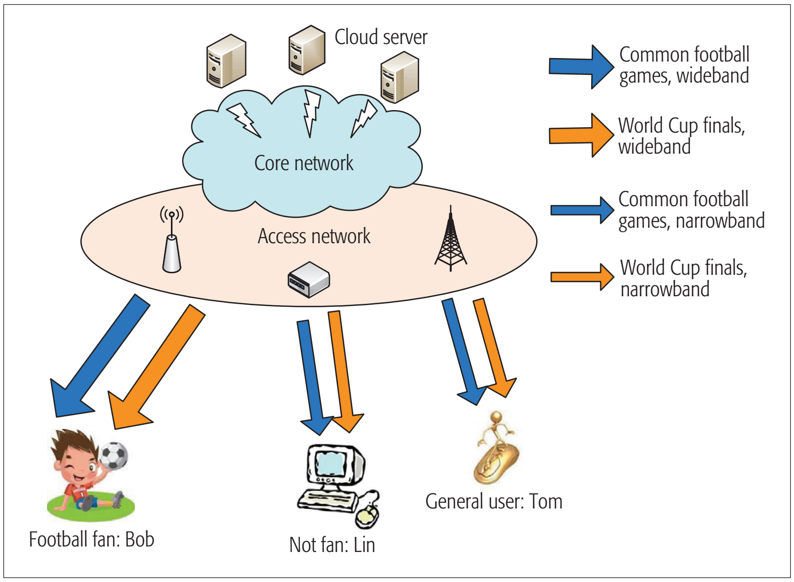

Figure 1. An example of personalized QoE management.

Assessment Methods

QoE assessment can be categorized into subject assessment, object assessment, and hybrid assessment.

Subjective assessment is based on psychoacoustic/visual experiments, in which participant testers specify services in a controllable environment and give quality scores based on their own experiences. Subjective assessment is the most direct way to assess users' QoE since the results are given by humans directly. Nevertheless, the assessment methodology is excessively costly, which makes it difficult to be employed in a large-scale assessment.

In objective assessment, QoE is mathematically modeled with some objectively measurable factors as input variables, and the value of QoE is the corresponding output. The advantage of objective methods is the convenience and tractability. However, the evaluated QoE from the objective approach is an approximation, rather than the exact measurement for each user.

The hybrid assessment uses subjective methods to obtain the accurate QoE, establishes the model with respect to objective factors in a laboratory environment, and then employs objective assessment methods to calculate QoE outside the laboratory. A classic hybrid assessment called pseudo subjective quality assessment (PSQA) has been proposed in order to provide accurate QoE assessment as perceived by humans widely.

QoE Models

It is obvious that any description of QoE can be regarded as a type of QoE model. In this article, however, the definition of the QoE model is narrowed down to the mathematical relation between QoE and the influencing factors. We categorize QoE models into mathematic models and machine learning models.

Mathematic Models: Mathematical models are traditional methods to study the relation between the influencing factors and the users' QoE. Gener-

Based on experiences researching QoE, it is realized that in many situations, the relation between influencing factors and QoE cannot be expressed by mathematical formulas explicitly. Furthermore, even if a mathematical model can be established, system parameters may be hidden in the beginning.

Author	Factor	Domain	Application	QoE assessment or model	Optimization method
Khan *et al.* [6]	Code rate Delay Packet loss Bandwidth PSNR Packet error probability	Technology	Voice File downloading Streaming video	Logarithmic IQX	Greedy algorithm on symbol rate
Hossfeld *et al*. [7]	Page loading time Memory effect Type of score	Technology Human	Video Web	SVM	Not discussed
Thakolsri *et al*. [8]	VSSIM	Technology	Video	Linear model	Greedy algorithm on data rate
Hosfeld *et al*. [9]	Freeze Block	Technology	Video	Crowdsourcing SOS Model	Not mentioned
Ghareeb *et al*. [10]	I-frame losses P-frame losses B-frame losses Group of picture sizes	Technology	Video	PSQA	Streaming path selecting
Sackl *et al*. [11]	Expectation Access method	Human Context	Not specified	Subjective assessment	Not discussed

TABLE 1. Summary of well-studied factors.

ally, related parameters are collected in laboratory environments, along with the users' QoE. Statistic analysis is conducted to find the specific QoE formulation with respect to the parameters. Some models are introduced below.

The E-model recommended by ITU-T SG12 is a classic linear model that provides predictions for the overall quality in a voice conversation at the network planning stage. The E-model is a mature model, but it is limited to voice service over telecommunication networks.

Reichl *et al*., argue that QoE has a logarithmic nature based on the Weber-Fechner Law (WFL) [12]. The basic idea of the WFL is that the human sensory system can be traced back to the percipience of "just noticeable differences", and the differential perception is directly proportional to the relative change of physical stimulus. By taking network QoS as stimulus and QoE as the perception, we can obtain a logarithmic relation mapping from QoS to QoE. The logarithmic model is very convincing because it has a strong link with psychological theory. However, it also has limitations in that QoS should be interpreted as physical stimulus while many QoS factors cannot.

Another mathematical QoE model that is based on the "IQX hypothesis" is discussed in [5]. The basic idea of this hypothesis is that a change of QoE depends on both the identical QoS changes and the actual level of QoE. After the integration of the formulation of QoE and QoS, we can obtain a negative exponential function of QoE w.r.t. QoS impairment factors such as packet loss. It should be noted that the IQX hypothesis correlates QoE to QoS impairments such as packet loss or network delay, while the logarithmic model correlates QoE to the perceivable QoS resources such as bandwidth or bit rate.

Machine Learning Model: Based on experiences researching QoE, it is realized that in many situations, the relation between influencing factors and QoE cannot be expressed by mathematical formulas explicitly. Furthermore, even if a mathematical model can be established, system parameters may be hidden in the beginning. In order to explore the implicit relation and the system parameters, machine learning methods are widely applied.

A classical machine learning approach is the recurrent neural network (RNN) model used in the PSQA assessment method. RNN is composed of a set of neurons that exchange signals in the form of spikes. Each neuron's state is a non-negative integer called potential to be changed by the arriving signals. After the training process, the state of each neuron can be determined, and then the assessment can be carried out by the RNN based on these potentials.

The support vector machines (SVM) model is adopted by Hossfeld in [6]. The input factors are the current page load time and the past values of user's rating. A hyperplane is calculated through the SVM approach based on the training set. Once the hyperplane is settled down, users' current QoE can be assessed easily.

There are other machine learning mechanisms such as decision tree and Bayesian network. With the development of the machine learning technologies, the use of these technologies in QoE modeling needs to be studied further.

QoE Control

Given the QoE model, communication resources should be controlled intelligently to further improve the overall QoE of users. There are different QoE controlling strategies based on the type of communication resources. We will introduce several typical QoE control schemes.

In [5], a cross-layer model is established for multimedia traffic in mobile communication systems. The symbol rate is mapped to QoE, which is a type of network layer parameter. The maximized QoE is achieved by adjusting the symbol rate for each user with a greedy algorithm.

Ghareeb *et al*., divide a video stream into sub-streams using the multiple description coding scheme [9]. In that scheme, QoE is assessed by PSQA when two sub-streams arrive at a client. The server then makes a decision to redirect the two sub-streams over new paths if the QoE cannot meet users' requirement.

Challenges of QoE in 5G

Although 5G networks are still at conceptual stage, there is some consensus about the notable trends of 5G, such as explosive growth of data traffic, massive increase in the number of interconnected devices, continuous emergence of new services and application scenarios, and so on. These trends will bring new challenges to QoE management, which are discussed as follows.

Challenges from Various Communication Scenarios

In 5G, typical scenarios include communications when people are at work, when people are at home or at entertainment events, and when people are on the move [13]. A consistent user experience is expected in all types of scenarios, including in office towers, dense residential areas, stadiums, subways, highways, and so on. These scenarios, which are characterized by ultra-high traffic volume density, ultra-high connection density, or ultra-high mobility, are challenging for QoE management in 5G. For example, in high speed scenarios, maintaining a satisfactory level of service continuity is of great importance. Similarly, in the ultra-dense small cells scenario, designing an appropriate mobility management strategy in order to, for example, reduce the number of handovers, is a key issue to improve QoE. Therefore, maintaining a consistent and optimum user experience in such complicated network scenarios requires close coordination of the various QoE control strategies listed above.

Since the key factors affecting QoE may vary in a wide range of scenarios, and users may have different expectations for their service experience in distinct scenarios, QoE becomes a comprehensive metric, and providing a personalized user experience through more refined resource management in such complicated 5G scenarios is an important topic.

Challenges Due to Emerging Applications

It can be foreseen that the applications for 5G mobile communication will be much richer. In 2G, the main applications are phone calls and SMS. In 3G and 4G, all kinds of smart phone applications spring out, and the applications extend to mobile games, mobile music, mobile video, and so on. As a result, cellular technology has changed our society and the way we communicate [14]. With further development in 5G, applications such as virtual reality, 3D videos, and interactive games may emerge. These emerging applications could bring two challenges in the field of QoE.

One challenge in 5G resulting from emerging applications is the requirement of new QoE models. With the rapid development of computer hardware and software as well as the standardization of Internet technology, intelligent mobile terminals and their corresponding services become diverse [15]. Different from traditional terminals, smart wearable terminals have various forms, such as smart watches, smart glasses, sports wristbands, smart jewelry, and so on. Meanwhile, mobile services, which are based on intelligent terminals, are no longer limited to telephone, text messaging, and video services. They are more involved in medical monitoring, interactive games, information exchange, and other emerging businesses.

Since most existing QoE models focus on VoIP, video, and HTTP services, the new service characters and user demands should be studied for the above-mentioned recently developed applications in order to set up a proper QoE model.

Imagine a medical monitoring service running on wearable equipment, which monitors blood pressure and pulse and even identifies when the user has an irregular heartbeat. In this case, the QoE model should combine the individual user's information to make accurate decisions to achieve a high level of QoE. For example, for a user with hypertension according to their health history, when symptoms appear, the system is required to ensure real-time transmission and a certain level of reliability to activate the emergency alarm and other corresponding medical-related services. From this example, we can see that new personalized services require a combination of various factors in order to guarantee QoE.

With new attractive applications, users may use their cell phones more frequently. Consequently, the mobile phone should have a long enough battery life to ensure reliable service even with more extensive usage. Otherwise, QoE may drop significantly. Therefore, in terms of QoE, energy efficiency is also an important metric in 5G.

With new attractive applications, users may use their cell phones more frequently. Consequently, the mobile phone should have a long enough battery life to ensure reliable service even with more extensive usage. Otherwise, QoE may drop significantly. Therefore, in terms of QoE, energy efficiency is also an important metric in 5G.

Challenges Related to Big Data

In the 5G era, with the exponential growth in network data traffic and richer applications, it can be easily predicted that huge amounts of data will be generated in mobile communications. This feature will also result in several challenges in the fields of QoE research.

One research challenge related to big data is the subjective factors affecting QoE. It is mainly embodied in two aspects. One aspect is the research of different kinds of subjective factors. By subjective factors we primarily mean a user's specific preferences. As a matter of fact, however, there are many other kinds of subjective factors that also matter. These factors include the user's mood, attention, expectations, etc. [16]. Consequently, a series of problems are involved with respect to these factors:

- How to quantize these factors?
- How to normalize these factors among users after quantization?
- How to capture them in a real time application?

It can be observed that the solutions to these problems are related not only to technical fields, but also to social psychology and cognitive science. Therefore, the research is still in its infancy.

The other aspect is the research of the analytical relation between QoE and the influencing factors. It has been widely accepted that QoE is a multi-disciplinary metric. In other words, QoE is affected by a variety of factors from different fields. However, if all the factors are taken into consideration, tackling their influence explicitly might be very difficult. Therefore, it is concluded that this problem requires further study.

Another valuable challenge is the evaluation of QoE from the perspective of big data. Until now, almost all research on QoE is based on

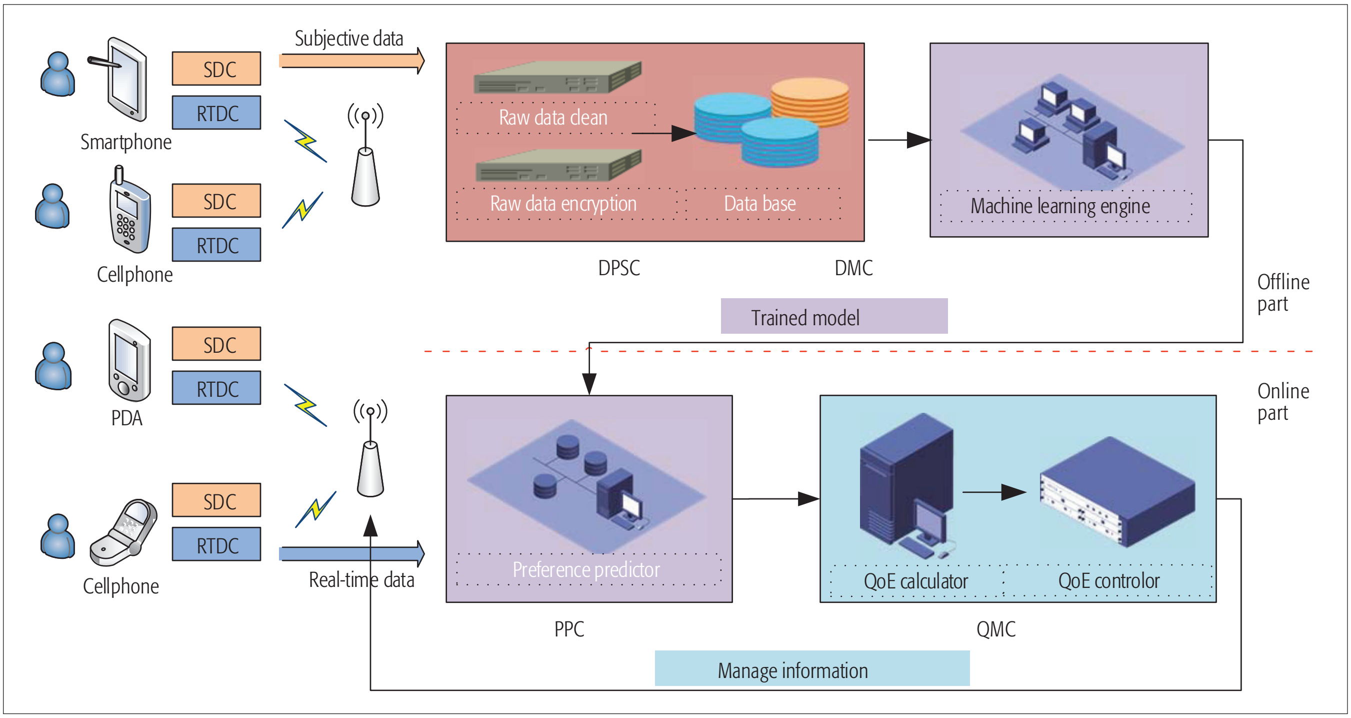

FIGURE 2. A data-driven architecture for personalized QoE.

direct feedback from users in order to obtain a subjective evaluation. However, it is expensive to collect useful feedback information, and data collection is limited by the controlled environment. In the era of 5G and big data, in addition to the traditional method that employs data collection through user feedback, some objective indicators in a data-driven method could be inferred to describe the user's subjective experience. For example, for online video services, the viewing time, the total number of clicks on a certain video, the number of daily accesses, etc., could be measured and analyzed to determine a user's QoE. For web services, QoE attributes could be extracted from online reviews that reflect user experience feedback on the web services. This process faces challenges such as natural language processing and unstructured data processing. Therefore, extracting features and deriving/analyzing the user's experience from a large amount of data should be a focus of research in the future.

It is also challenging to handle the security and privacy related to intelligent terminals, which comes along with the trend that intelligent terminals play a key role in 5G, especially in the big data era.

For those who use intelligent terminals, intentionally or unintentionally, their own personal details, such as their personal information, contacts, download history, application usage records, system logs, etc., are kept in the terminal devices. Intelligent terminals can infer a user's own personality traits and preferences, consumption habits, and even values, in order to provide personalized QoE. Since personalized service involves the collection and analysis of a user's personal information, balancing the protection of a user's privacy and personalized QoE management is also a major challenge.

A Data-Driven Architecture for Personalized QoE in 5G Networks

As mentioned before, the essential bottleneck for assessing QoE from a user's perspective is to involve the subjective characteristics of the users, which should be captured by the personalized QoE management architecture. For this purpose, a data-driven architecture for personalized QoE in 5G networks is proposed. First, several requirements the proposed architecture fulfills are explained in detail below:

- The architecture should contain a monitor to obtain real time information about the application that the users are using and the QoS status.
- The architecture should maintain a data mining scheme that can predict a user's preference/expectations toward the applications in use.
- The architecture should manage the communication resources based on the QoS status and the predicted preference/expectation to maintain a satisfactory QoE.

Figure 2 illustrates the two parts (the offline part and the online part) of the proposed architecture.

Offline Part

The offline part is responsible for training and tuning the user preference prediction model. The granularity of the model is per-user per-service, that is, the inputs are a specific user and a specific service in use, and the output is this user's expectation toward this service. To train this model, some basic components are necessary.

Subjective Data Collector: All data are originally collected by the subjective data collector (SDC). Since the data is used for model training, it is expected to have included different dimensions and levels such that the user and the type of service can be recognized with adequate detail.

Almost all research works concerning QoE evaluations have a scheme to collect related data. Considering the specific characteristics of this architecture, the distributed data collecting agent is suitable because of the open environment in which the system is deployed and the multi-dimension of data that the system collects.

The central part of this data collecting component is a mobile agent installed on the end-user devices. Three different entities are logically partitioned in this agent: the QoS monitoring entity, the contextual monitoring entity, and the experience monitoring entity. In addition, a local data repository is necessary for the storage of measured data and to synchronize with the data processing and storage component. The structure of the mobile agent is shown in Fig. 3, and the modules within this structure are detailed as follows:
- The QoS monitoring entity is responsible for measuring the technical parameters, including the device information, such as operating system and screen size; the network information, such as the access type, throughput, delay, and jitter; and the application information, such as the application type.
- The contextual monitoring entity is in charge of collecting context information of the application, which contains the location, the mobility, and some other available data from sensors on the device or around the body. This entity is extensible for all kinds of newly developed sensors.
- The experience monitoring entity interacts with users by gathering explicit feedback in the form of questionnaires

The collection method follows the experience sampling method (ESM) to minimize possible disruptions. Both closed and open-ended question formats are supported. In addition, the experience probe can be extended with new modules and parameters. With further study of implicit experience measures, e.g., brain activity, heart activity, and eye movement, this entity can be easily extended.

Data Processing and Storing Component: All the data from the distributed SDC is synchronized to the data processing and storing component (DPSC). The main function of this component is to pre-process and then store the data into databases, preparing for data mining afterward.

Once new data is received, an initial data cleaning process is carried out to remove damaged data. Then, anonymization is employed by adopting pseudonyms and reducing location accuracy. The cleaned data is further encrypted for security reasons and a set of features is extracted from the data. Both extracted features and encrypted data are stored within distributed databases. Only the extracted features can be accessed by other components in the proposed architecture.

Note that a user may use various end devices or login as different accounts for distinct services. To associate data from individual sources, a unique user ID is required. The concept known as OpenID can serve this purpose, among other uniquely identifiable ID schemes. A unique user ID can be used as authenticated information that is linked to a user's cell phone number or verifiable email address. With a unique user ID, the DPSC can merge data for each user every time after obtaining data from SDC.

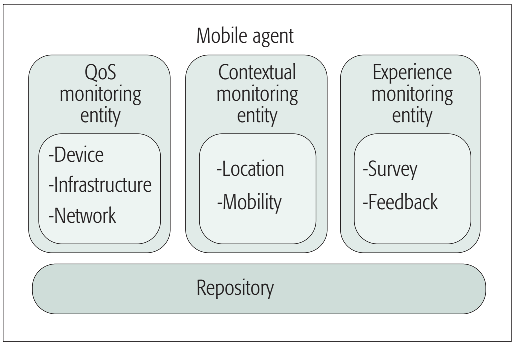

FIGURE 3. The structure of mobile agent.

Data Mining Component: The data mining component (DMC) is responsible for training the data mining model to understand user preferences. This is the core component of the entire system. Most traditional QoE modeling approaches focus on the mapping from network layer parameters to QoE parameters. However, a specific user has very different preferences for distinct services. Thus, in the same network layer condition, users could have different satisfaction levels for distinct services due to different preferences. Consequently, one or more data mining models with respect to user preferences are built into this component, and they are trained by the data obtained from DPSC. After being validated, the trained models can be employed in the online part.

ONLINE PART

The online part obtains information about the users and the corresponding services in real time. With the help of the trained data mining models in the offline part, user preferences based on services can be predicted. The online part manages all of the users' QoE according to their preferences and the network status. The online part is formed by three components: the real-time data collector, the preference predictor, and the QoE controller.

Real-Time Data Collector: The real-time data collector (RTDC) is different from the SDC. The main purpose of the RTDC is to capture the current information of the user, the services the user is using, and the network resources, while the SDC collects users' subjective information. Three types of data are gathered by the RTDC: the user ID, the service in use, and the network status. Similar to the SDC, an agent is included in the end device, which sends the user-service data to the preference prediction component (PPC), and the network status information to the QoE controller.

Preference Prediction Component: The core of the preference prediction component (PPC) is a preference predictor. The task of the predictor is to process information obtained by the PPC, by selecting the right data mining model that was trained in the offline part, to further predict preferences for this specific user. Note that different models may be trained for various types of services, and the current model is automatically selected by the corresponding predictor.

QoE Management Component: User preferences, derived from the PPC, is sent to the QoE management component (QMC). The QMC maintains a QoE calculator and a QoE controller. The controller is aware of the status of the objec-

The offline part is responsible for training and tuning the user preference prediction model. The granularity of the model is per-user per-service, that is, the inputs are a specific user and a specific service in use, and the output is this user's expectation toward this service.

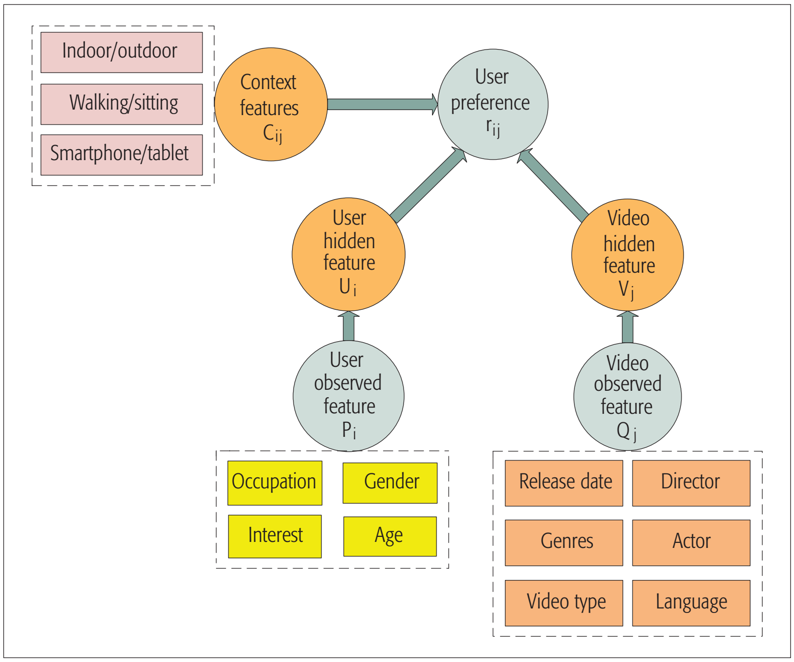

FIGURE 4. User-service preference model.

tive network and the subjective preferences of users in the system. To manage QoE, a key issue is to model the subjective factors and objective factors that would affect QoE quantitatively. A reasonable model can be selected from the models summarized above. With a QoE optimization target such as maximizing the total QoE in the system or improving the QoE fairness for all users, an optimization scheme can be designed based on the current system's status and user preferences.

There are many parameters that can be tuned in a communication system. In access networks, factors such as bandwidth, transmission channels, and the resource block (RB) can be scheduled. In core networks, different routing schemes can be selected for different types of streaming. Furthermore, there are also various optimization approaches for streaming coding. In practice, to optimize system performance according to the users' QoE, the QMC can maintain a strategy pool from which different strategies can be selected and implemented for various parts of the system. As a result, the optimization can be employed at the system level. In more detail, we can adopt a suitable optimization strategy on various system parts based on the analysis of service characteristics and the current system status.

USE CASE STUDY

In order to validate the proposed architecture, we study a preliminary use case. In this case, we collected user data from volunteers from the university campus. These data include gender, age, occupation, and preferred video type. The service in this case is configured as online video, and the videos are categorized into 16 classes. The release dates of the videos are also recorded. The context features we considered include whether the user is indoors or outdoors, whether the user is walking or sitting, and whether the user watches video using a smartphone or a tablet.

In this case, a two-step QoE modeling method is presented, which depends not only on network-layer parameters, but also on user-service preferences. In the first step, the user-service preference is modeled and the preference value is predicted as the feature of QoE modeling. In the second step, QoE is modeled according to both network layer parameters and user preferences.

The regression-based latent factor model (RLFM) is adopted for user preference prediction, which is shown in Fig. 4, where subscripts i and j stand for the user and service, respectively. As can be seen from the figure, the model contains three layers. In the top layer, r_{ij} represents user x_i's preference for video y_j, and C_{ij} denotes specific context features related to r_{ij}. In the bottom layer, P_i and Q_j represent the observed user and the video feature vectors, respectively. The U_i and V_j in the middle layer represent the hidden (latent) feature vectors for the user and the video, respectively. These three layers construct a Bayesian graphic, in which r_{ij} is subject to the distribution determined by U_i and V_j, and Ui and V_j, in turn, are subject to the distribution determined by P_i and Q_j. The Monto Carlo expectation-maximization (EM) algorithm is adopted in this case to train the model.

Since the feature data considered in our architecture is of large amount, and the RLFM can accurately predict user responses for large-scale data in the presence of the feature, this model is suitable for preference prediction in our application. In addition, the performance of RLFM has been verified on certain benchmark datasets. The learning procedure of the PPC is illustrated in Fig. 5a, in which the root mean square error (RMSE) of user preference can be observed for both the training set and the validation set. It can be seen from the figure that after some iterations, the RMSE of the predicted preference of the validation set stays at a low level, and more importantly, it becomes close to that of the training set. This is strong evidence showing that the PPC can efficiently predict user preferences. In addition, even better performance is expected with enhanced learning technology and models.

Then, the QoE model is developed using the popular sigmoid function with respect to user preference r_{ij} and QoS denoted by S, that is,

$$QoE = \frac{\theta}{1 + e^{(-\alpha S + \beta r_{ij} + \gamma)}},$$

where α, β, γ, and θ are the parameters constraining the quantization of QoE, and for simplicity, the bit rate is utilized to represent the QoS a user can obtain.

We assume that the users are randomly distributed in the coverage areas of five APs, and each AP has an upper bound of the total bit rate. An optimization problem is formulated to describe the maximization of the total QoE. By solving the problem, a properly tuned bit rate could be achieved for the optimization of QoE.

We compared the QoE distribution in the proposed architecture with that of the traditional water-filling algorithm. In the traditional water-filling algorithm, an iterative bit rate allocation solution is obtained as a result of total QoE maximization without considering user expectations/preferences, which results in more bit rate allocated to the user with lower QoE. Simulation

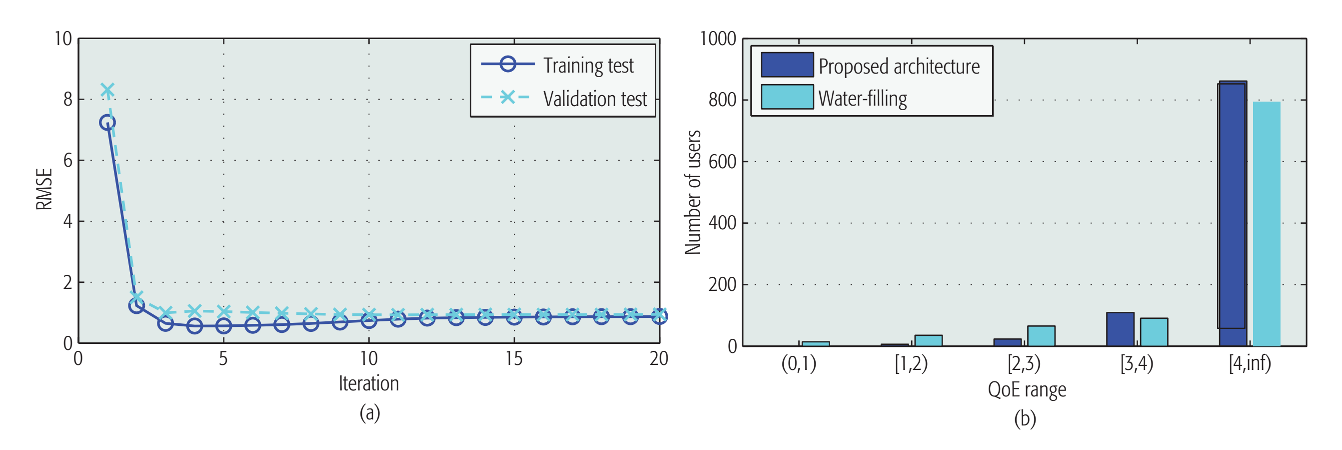

FIGURE 5. Use case figures: a) Learning procedure of PPC; b) QoE comparison.

results show improved performance of the system compared with the traditional resource allocation method, with an increase of the total QoE by 20 percent, and 96 percent of the users obtaining a higher QoE, as shown in Fig. 5b.

CONCLUSIONS

In this article, we summarize in detail the state-of-the-art techniques for QoE and discuss the challenges to employ QoE in 5G wireless communication networks. Based on the observations of the existing QoE studies and the understanding of wireless networks, we propose a systematic architecture for QoE optimization in 5G wireless networks that consists of an online layer and an offline layer. To evaluate the proposed architecture, we deploy a preliminary use case. Simulation results illustrate that the performance of the proposed architecture increases total QoE by 20 percent, while 96 percent of users obtain a higher QoE compared with the traditional water-filling method.

ACKNOWLEDGMENT

This work is supported by the National Nature Science Foundation of China (61372113, 61421061), the National 863 Project (2014AA01A701), and the Beijing Natural Science Foundation (4132050).

REFERENCES

[1] J. Qiao *et al.*, "Video Quality Provisioning for Millimeter Wave 5G Cellular Networks with Link Outage," *IEEE Trans. Wireless Commun.*, vol. 14, no. 10, 2015, p. 5692–5703.

[2] M. Ismail and W. Zhuang, "Statistical QoS Guarantee for Wireless Multi-Homing Video Transmission," *Proc. IEEE Global Communications Conference (GLOBECOM)*, USA, Dec. 2013.

[3] Y. Cheng, A. Leon-Garcia, and I. Foster, "Toward an Autonomic Service Management Framework: A Holistic Vision of SOA, AON, and Autonomic Computing," *IEEE Commun. Mag.*, vol. 46, no. 5, 2008, pp. 138–46.

[4] K. Brunnström *et al.*, "Qualinet White Paper on Definitions of Quality of Experience," *European Network on Quality of Experience in Multimedia Systems and Services (COST Action IC 1003)*, Lausanne, Switzerland, Version 1.2, Mar. 2013.

[5] K. Laghari and K. Connelly, "Toward Total Quality of Experience: A QoE Model in a Communication Ecosystem," *IEEE Commun. Mag.*, vol. 50, no. 4, Apr. 2012, pp. 58–65.

[6] S. Khan *et al.*, "MOS-based Multiuser Multiapplication Cross-Layer Optimization for Mobile Multimedia Communication," *Advances in MultiMedia*, vol. 2007, no. 1, Jan. 2007, pp. 6–6.

[7] T. Hosfeld *et al.*, "The Memory Effect and its Implications on Web QoE Modeling," *Proc. 23rd Int'l. Teletraffic Congress (ITC)*, USA, Sept. 2011.

[8] S. Thakolsri, W. Kellerer, and E. Steinbach, "QoE-Based Cross-Layer Optimization of Wireless Video with Unperceivable Temporal Video Quality Fluctuation," *Proc. IEEE Int'l. Conf. Commun. (ICC)*, Japan, Jun. 2011.

[9] T. Hossfeld *et al.*, "Quantification of Youtube QoE via Crowdsourcing," *Proc. IEEE Int'l. Symposium Multimedia (ISM)*, USA, Dec. 2011.

[10] M. Ghareeb and C. Viho, "Hybrid QoE Assessment is Well-Suited for Multiple Description Coding Video Streaming in Overlay Networks," *Proc. Eighth Annual Communication Networks and Services Research Conf. (CNSR)*, USA, May 2010.

[11] A. Sackl *et al.*, "Wireless vs. Wireline Shootout: How User Expectations Influence Quality of Experience," *Proc. Fourth Int'l. Workshop Quality Multimedia Experience (QoMEX)*, Australia, Jul. 2012.

[12] P. Reichl *et al.*, "The Logarithmic Nature of QoE and the Role of the Weber-Fechner Law in QoE Assessment," *Proc. IEEE Int'l. Conf. Communications*, South Africa, May 2010.

[13] A. Osseiran *et al.*, "Scenarios for 5G Mobile and Wireless Communications: The Vision of the METIS Project," *IEEE Commun. Mag.*, vol. 52, no. 5, May 2014, pp. 26–35.

[14] G. Fettweis and S. Alamouti, "5G: Personal Mobile Internet Beyond What Cellular did to Telephony," *IEEE Commun. Mag.*, vol. 52, no. 2, Feb. 2014, pp. 140–45.

[15] B. Bangerter *et al.*, "Networks and Devices for the 5G Era," *IEEE Commun. Mag.*, vol. 52, no. 2, Feb. 2014, pp. 90–96.

[16] M. Seufert *et al.*, "A survey on Quality of Experience of HTTP Adaptive Streaming," *IEEE Commun. Surveys & Tutorials*, vol. 17, no. 1, Sep. 2014, pp. 469–92.

BIOGRAPHIES

YING WANG [M'04] received her Ph.D. in circuits and systems from Beijing University of Posts and Telecommunications (BUPT) in 2003. She is a professor at BUPT and the director of the Radio Resource Management Lab, Wireless Technology Innovation Institute, BUPT. Her research interests are in the area of cooperative and cognitive systems, radio resource management, and mobility management in 5G systems. She won first prize of the Scientific and Technological Progress Award from the China Institute of Communications in 2006 and 2009, and second prize of the National Scientific and Technological Progress Award in 2008. She has published more than 100 papers in international journals and conference proceedings.

PEILONG LI received the B.S. degree and M.S. degree in information and communications engineering from the Beijing University of Posts and Telecommunications in 2012 and 2015, respectively. He currently works at Baidu.com Inc. His research interests are user preferences and QoE management.

LEI JIAO [S'08, M'12] received the B.E. degree in telecommunications engineering from Hunan University, Changsha, China, in 2005, the M.E. degree in communication and information systems from Shandong University, Jinan, China, in 2008, and the Ph.D. degree in information and communication technology from the University of Agder (UiA), Norway, in 2012. He is now working in the Department of Information and Communication Technology, UiA, as an associate professor. His research interests include cognitive radio networks, wireless sensor networks, and artificial intelligence.

ZHOU SU [S'03, M'06] received the B.E and M.E degrees from Xi'an Jiaotong University, Xi'an, China, in 1997 and 2000, respectively, and the Ph.D. degree from Waseda University, Tokyo, Japan, in 2003. His research interests include multimedia communication, web performance, and network traffic. He received the best paper award at the International Conference ChinaCom2008, and the Funai Information Technology Award for Young Researchers in 2009.

NAN CHENG [S'13] received his B.S. and M.S. degrees from Tongji University, China in 2009 and 2012, respectively. He is currently pursuing a Ph.D. degree in the Department of Electrical and Computer Engineering, the University of Waterloo, Waterloo, ON, Canada. His research interests include vehicular communication networks, cognitive radio networks, and resource allocation in smart grids.

XUEMIN (SHERMAN) SHEN [M'97, SM'02, F'09] is a professor and University Research Chair, Department of Electrical and Computer Engineering, University of Waterloo, Canada. His research focuses on resource management in interconnected networks, wireless network security, smart grid, and vehicular ad hoc and sensor networks. He serves/has served as Technical Program Committee chair/co-chair for IEEE GLOBECOM 2007 and 2016, IEEE INFOCOM '14, IEEE VTC-Fall '10, General Chair for ACM Mobihoc'15, Symposia Chair for IEEE ICC '10, and Tutorial Chair for IEEE VTC-Spring '11 and IEEE ICC '08. He also serves or has served as Editor-in-Chief for *IEEE Network*, *Peer-to-Peer Networking and Applications*, and *IET Communications*. He is a registered Professional Engineer of Ontario, Canada, an Engineering Institute of Canada Fellow, a Canadian Academy of Engineering Fellow, a Royal Society of Canada Fellow, and a Distinguished Lecturer of the IEEE Vehicular Technology Society and the IEEE Communications Society.

PING ZHANG received his Ph.D. degree from BUPT in 1990. He is currently a professor at BUPT and vice director of the Ubiquitous Networking Task Commission of the Chinese Communication Standardization Association. He is the chief scientist of the National Program on Key Basic Research Project (973 Program), and also serves as a member of the China 3G Group China 863 FUTURE project. His research interests mainly focus on the fields of mobile communications, ubiquitous networking, and service provisioning.