A Network and Device Aware QoS Approach

for Cloud-Based Mobile Streaming

Chin-Feng Lai*, Member, IEEE*, Honggang Wang*, Senior Member, IEEE*, Han-Chieh Chao*, Senior Member, IEEE*, and Guofang Nan*, Member, IEEE*

***Abstract—*Cloud multimedia services provide an efficient, flex- ible, and scalable data processing method and offer a solution for the user demands of high quality and diversified multimedia. As intelligent mobile phones and wireless networks become more and more popular, network services for users are no longer limited to the home. Multimedia information can be obtained easily using mobile devices, allowing users to enjoy ubiquitous network services. Considering the limited bandwidth available for mobile streaming and different device requirements, this study presented a network and device-aware Quality of Service (QoS) approach that provides multimedia data suitable for a terminal unit en- vironment via interactive mobile streaming services, further considering the overall network environment and adjusting the interactive transmission frequency and the dynamic multimedia transcoding, to avoid the waste of bandwidth and terminal power. Finally, this study realized a prototype of this architecture to validate the feasibility of the proposed method. According to the experiment, this method could provide efficient self-adaptive mul- timedia streaming services for varying bandwidth environments.**

***Index Terms—*Adaptive Qos, cloud multimedia, network and device-aware.**

I. INTRODUCTION

LOUD computing has become the development trend of the Internet [1]. Massive amounts of data are calculated simultaneously and user demands are met rapidly, based on the architecture of cloud resource virtualization. The basic tech- nique of cloud computing is derived from distributed computing and grid computing. In recent years, as mobile devices have de- veloped rapidly, users have been able to access network ser-

**C**

Manuscript received March 05, 2012; revised July 26, 2012; accepted November 13, 2012. Date of publication January 15, 2013; date of current version May 13, 2013. This work was supported by the National Science Council and Science Park Administration of the Republic of China, Taiwan under Contract NSC 101-2628-E-197-001-MY3, 101-2221-E-197-008-MY3,

101-2219-E-197-004 and 101MG07-2. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Chang Wen Chen.

C.-F. Lai is with the Department of Computer Science and Information Engineering, National Chung Cheng University, Chiayi, Taiwan (e-mail: [cinfon@ieee.or](mailto:cinfon@ieee.org)g).

H. Wang is with the Department of Electrical and Computer Engineering, University of Massachusetts Dartmouth, North Dartmouth, MA 02747 USA (e-mail: [hwang1@umassd.ed](mailto:hwang1@umassd.edu)u).

H.-C. Chao is with the Institute of Computer Science and Information Engi- neering, National ILan University, I-Lan, Taiwan (e-mail: [hcc@niu.edu](mailto:hcc@niu.edu.tw).tw).

G. Nan is with the Institute of Systems Engineering, Tianjin University, Tianjin, China, and also with the Department of Electronics, Polytechnic University of Turin, Turin, Italy (e-mail: [gfnan@tju.edu](mailto:gfnan@tju.edu.cn).cn).

Digital Object Identifier 10.1109/TMM.2013.2240270

vices anywhere and at anytime. Especially with the develop- ment of 3G and 4G networks, multimedia services have become universal application services. The media cloud is an extended technology developed to meet the fast-changing information in- dustry and user’s demand for higher multimedia quality and var- ious terminal units. It realizes multimedia computing, storage space configuration, and sharing services based on the powerful arithmetic capability of cloud computing [2]–[5]. As intelligent mobile devices and multimedia technology have begun to pop- ularize, the public has started to use mobile devices such as in- telligent mobile phones or tablets to view multimedia videos by means of streaming [6]–[8]. Generally speaking, accessing mul- timedia video services through networks is no longer a problem. The major video platforms, such as Youtube and Amazon, have good management styles and provide users to share multimedia videos easily with diversified services.

No matter what the service is, users will always expect powerful, sound and stable functions. For multimedia videos, stability is of the greatest importance. Users expect to watch videos smoothly and at a certain level of quality, no matter what changes occur in the network environment. However, the existing video platforms often provide inconsistent playback, resulting from the fluctuation of network on-line quality, es- pecially with mobile devices, which have limited bandwidth and terminal unit hardware resources. As the number of net- work users is rapidly increasing, bandwidth insufficiency will occur and then network multimedia services will be affected significantly. Differing from general services that have a high acceptance rate for packet loss, multimedia packets emphasize the correctness, sequence order and real-time nature of packets. When a multimedia video service is applied, the service quality declines greatly while trying to meet the demands of video transmission. Users often view live videos that freeze have intermittent sound, or even failure to operate. Therefore, how to execute smooth playback with limited bandwidth and the different hardware specifications of mobile streaming is an interesting challenge.

H.264/SVC is an extended coding and decoding architecture

based on H.264/AVC. The benefit of H.264/SVC is that it can adjust the image quality dynamically, according to the band- width of the receiving end. The draft was proposed in April of

2004 and was decided on in November of 2007. SVC puts for- ward a brand-new layered architecture. This hierarchical struc- ture can realize the scalability of temporal, spatial and quality dimensions. The spirit of SVC is that the receiving end is sure to receive image packets of the lowest quality for decoding. The image layer with the lowest quality is called the base layer. The

1520-9210/$31.00 © 2013 IEEE

base layer of SVC is completely compatible with H.264/AVC, and when there is enough bandwidth to receive image packets with higher quality, the decoder will carry out reference de- coding according to the received packets, which is to say, high quality image packets cannot resolve images independently; the image packet of the base layer must be consulted for decoding. In terms of the scalability of the three dimensions, SVC uses the hierarchical B-picture method to realize temporal scalability, down/up sampling filters and inter-layer prediction to realize spatial scalability, and Signal-to-Noise Ratio (SNR) scalability and an Metal Gear Solid (MGS) Codec to realize quality scal- ability. The scalability of hierarchies in the video can be deter- mined during the course of coding.

In addition, interactive mobile multimedia services commu- nicate and coordinate the mobile device with the server-side to select the multimedia file applicable to the device environment (bandwidth, resolution and arithmetic capability) [9], [10], so as to realize an optimal multimedia streaming service. This pre- vious study proposed an interactive mobile multimedia service over cloud computing, which was introduced in [23], [24]. In the previous service, the mobile device side exchanges informa- tion with the cloud environment, so as to determine an optimum multimedia video. Scholars have done numerous researches to- ward conventional platform (CDN) to store different movie for- mats in a multimedia server, to choose the right video stream according to the current network situation or the hardware cal- culation capabilities. To solve this problem, many researches have attempted dynamic encoding to transfer media content, but still cannot offer the best video quality. This is due to the time consuming fact that traditional encoding requires re-coding of the entire multimedia content. This research targets the char- acteristic of streaming protocols to record the current stream video content and the bandwidth state of the user while also analyzing the past bandwidth fluctuations to evaluate and pre- dict the possible bandwidth changes in the future while using map and reduce algorithm in cloud computing to immediately transfer the video encoding to quickly transfer the most suitable video format for the user. It performed well, both in power con- sumption and streaming quality. However, a number of issues were found in the experimental results, which are described in the following.

1) Frequent commutations cause extra bandwidth and elec-

trical power consumption. The mechanism of feedback at the fixed frequency adopted in the previous mechanism set- ting can be a good dynamic adjustment setting in a drasti- cally changing bandwidth environment; however, in a rela- tively stable bandwidth environment, too-frequent conver- sations will cause extra electrical power consumption in the terminal units.

2) It is a challenge to provide multimedia content man- agement in a dynamic environment. Previous studies have shown there are two major modes for providing multimedia data: the multi-source content mode, and the transcoding mode. The multi-source content mode pro- vides multimedia files of different resolutions and bitrates for the terminal units to choose from. Although this mode can provide a simple multimedia streaming service that can effectively reduce delay, it is a difficult challenge

to provide appropriate multimedia files for diversified terminal units. In the transcoding mode, multimedia files are transcoded dynamically, in order to be applicable to the device side according to the terminal environment. This mode needs to consider the real-time problem, es- pecially for H.264/SVC coding, as the time required for transcoding causes difficulties for real-time streaming. Al- though SVC is applicable to varying bandwidth networks due to its multilayer architecture, how to provide a multi- media hierarchy that is suitable for dynamic environment variations according to the terminal unit is an interesting research project.

Therefore, the network and device-aware QoS approach pro- posed in this study aimed to solve the aforesaid problems. It followed the previously proposed overall architecture, and fur- ther discussion was presented on how to dynamically adjust the communication mechanism and manage multimedia files to pro- vide self-adaptive multimedia streaming services according to the environmental parameters of various devices in a cloud en- vironment. This study made a number of major contributions.

1) Different from mobile streaming, this work introduces cloud-based real time transcoding for adaptive mobile streaming. Based on cloud computing this researches offers a more efficient method that uses map-reduce to separate the video content into different clips. Also performs distributed encoding to solve the immediate multimedia transcoding problem and further introduce a dynamic communicative and predictive bandwidth for dynamic mobile environments. This can provide a more stable streaming service for an unstable network. Com- pared to a stable network, the losses of bandwidth and power in the terminal units caused by excessive packet transmission can be reduced.

2) For the device parameters, this study used a Bayesian net- work environment and a multi-variable linear regression model to predict whether the multimedia files in different formats would conform to the required real-time needs for the streaming mechanism and to consider whether the overall electric consumption of the device could provide a complete multimedia file playback service.

3) This study further determined the optimum image adjust-

ment layers according to the environmental parameters (arithmetic capability, electric quantity of the battery, electric consumption rate and bandwidth) fed back from the device for SVC multimedia files, so as to obtain a better dynamic adjustment mechanism and to avoid the provision of a fixed quality which would reduce the image quality.

4) This study used a virtual environment to simulate the overall cloud multimedia environment for the proposed mechanism, and then used stream-limiting software to realize a prototype module of the overall architecture. The feasibility of the proposed method was validated and the related experiment was analyzed by constructing the prototype environment.

This paper is organized as follows. Section II introduces studies related to cloud media streaming. Section III describes the overall situation architecture and introduces definitions of

the parameters of various components as well as the internal module. Section IV realizes the overall prototype environment of the theoretical mechanism and conducts an experiment for analysis and comparison. The final section provides the conclusion and direction for future development.

II. RELATED WORK

Multimedia cloud computing can be divided into two major parts. One is the media cloud, and the other is cloud-aware mul- timedia (cloud media) [3]. Studies related to these two parts are introduced as follows.

*A. Review of Multimedia-Aware Cloud*

The multimedia-aware cloud mainly aims at multimedia files in the cloud environment. Meeting the requirement of QoS, according to the arithmetic capability of the cloud [11], [12], involves the balanced sharing of the load, real-time dynamic coding and heterogeneous work conversions. Fan *et al.* [13] used job scheduling for various clusters in the cloud environ- ment to attain load balance. Zheng *et al.* [14] proposed a collab- orative quality ranking framework and used the greedy mode to calculate the speeds of various components in the cloud, in order to efficiently improve the overall applications. Diaz-Sanchez *et al.* [15] used the Digital Living Network Alliance (DLNA) and Universal Plug and Play (UPnP) as an open multimedia plat- form for domestic multimedia sharing, and then used either the http or UDP transmission protocol to obtain open cloud com- puting middleware. Fengy *et al.* [16] proposed a streaming-ori- ented management service architecture according to different multimedia data in the cloud environment.

*B. Review of Cloud-Aware Multimedia*

Differing from the multimedia-aware cloud, cloud-aware multimedia inclines to the streaming service of front-end mul- timedia data. Multimedia video files are usually quite large. In order to provide the multimedia streaming service of real-time videos, files are divided into packets, so that the client-side can view multimedia information instantly as received instead through downloading. Multimedia streaming usually contains several major elements, such as the encoder, the decoder, the streaming server and the player. Streaming technology can be classified into two types: the web server and the streaming server. The web server is a general web server, using Hyper- Text Transfer Protocol (HTTP) as the communication medium. Streaming in this mode is called HTTP streaming. It is con- venient because no additional streaming servers are required; therefore it is also known as serverless streaming. The HTTP protocol uses Transmission Control Protocol (TCP). A hand- shake must be done before transmission, and retransmission is required if a data packet is lost, which causes severe delay.

Therefore, a set of HTTP live streaming protocols was

proposed for Apple Inc., which provided streaming services that allowed packet interruption. Garcia [17] used the http live streaming transmission mode to transcode multimedia files at the mobile device side for live streaming. Jin [18] allocated an applicable bandwidth environment to each mobile device prop- erly in a P2P streaming environment, used game formulation to

model the collaborative streaming problem, and predicted the transmission probability of each peer to calculate the overall bandwidth allocation. Trajkovska *et al.* [19] proposed a QoS function for the heterogeneous architecture of a cloud envi- ronment. They calculated the QoS parameter environment of each client side and executed QoS services according to the hybrid environment. The interactive streaming service aims to provide preferable mutual information between the server and client-side. Liu *et al.* [20] proposed a recommendation service that transmits recommended multimedia browse file to mobile devices according to the supported resolution fed back from the mobile devices. Yin *et al.* [21] proposed the present contemporary ubiquitous media services and recommended multimedia files according to direct resizing, attention-based adaptation, and user guided semantic adaptation models. The authorsprevious study was published in [22]–[24]. It considered the arithmetic capability and resolution of the terminal unit environment and the power factor at the mobile device side for the selection of multimedia streaming.

*C. Review of Dynamitic SVC Multimedia Service*

SVC contains spatial, temporal or quality three-dimensional parameters that can be adjusted dynamically and are provided in multilayer transmission, so as to provide a preferable streaming service for network instability, which has been mostly studied in mobile multimedia services. The methods proposed by Sohn *et al.* [25] and Choi *et al.* [26] assures the spatial, temporal or quality layer number preferentially according to the user’s preference setting, and gives the largest number of SVC layers within the capacity of the available bandwidth to match the variance in the network environment and to increase the band- width utilization rate. Huang *et al.* [27] used a cloud-based SVC proxy for the SVC format conversion and transmission of mul- timedia files to provide better streaming services. The Priority based Layer Switching (PLS) algorithm proposed by Zhang *et al.* [28] determines the receiving bit rate according to RTCP and changes the SVC layer quantity according to the receiving bit rate, in order to match variances in the network environment. The method proposed by Chen *et al.* [29] predicts the available bandwidth according to the history and network monitoring, and then abstracts the image data containing partial SVC layers for transmission, allowing the streaming the bit rate to approach to the predicted available bandwidth. Differing from only consid- ering the network quality, Ramos *et al.* [30] believed that the de- vice capability is an important factor that limits video quality. They proposed Device and Network-Aware Scaling (DeNAS) for multiple mobile devices, determining the appropriate SVC layer number according to the device capability, the network ca- pacity and the user’s request, so as to meet the video quality re- quirement of the device and to increase the efficiency of the net- work bandwidth service. The results showed that even in mul- tiple shared environments with limited bandwidth, the frame- work can provide SVC streaming with preferable image quality. Lee *et al.* [31] proposed the Adaptive Spatial Resolution Con- trol (ASRC) algorithm in order to reduce the power consump- tion of mobile devices. This algorithm considers picture quality, available bandwidth and energy consumption. The energy for decoding is calculated before the video data is transmitted, and

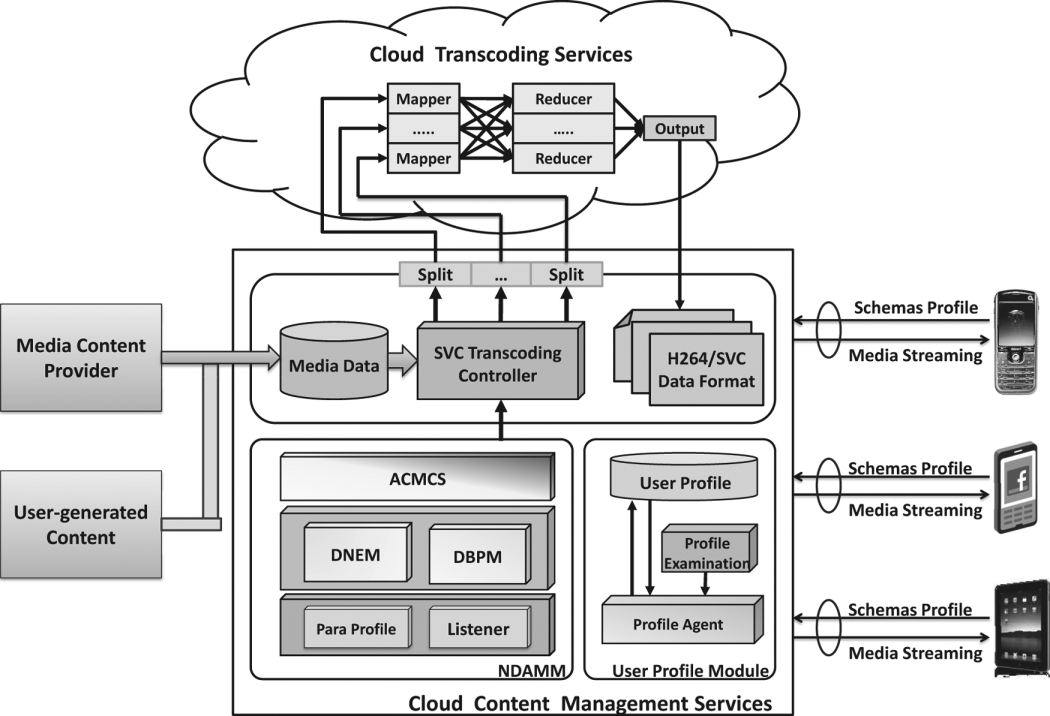


Fig. 1. Proposed System Structure of Cloud-based Mobile Streaming.

the energy consumption and picture quality are evaluated before the SVC layer number to be downloaded is determined.

III. PROPOSED SYSTEM STRUCTURE AND

MODULE PARAMETER DEFINITION

The proposed system provided an efficient interactive streaming service for diversified mobile devices and dynamic network environments. The overall system structure is shown in the following Fig. 1. When a mobile device requests a multimedia streaming service, it transmits its hardware and network environment parameters to the profile agent in the cloud environment, which records the mobile device codes and determines the required parameters, and then transmits them to the Network and Device-Aware Multi-layer Management (NDAMM). The NDAMM determines the most suitable SVC code for the device according to the parameters, and then the SVC Transcoding Controller (STC) hands over the transcoding work via map-reduce to the cloud, in order to increase the transcoding rate. Finally, the multimedia video file is trans- mitted to the mobile device through the service.

*A. User Profile Module*

The profile agent is used to receive the mobile hardware envi- ronment parameters and create a user profile. The mobile device transmits its hardware specifications in XML-schema format to the profile agent in the cloud server, as shown in the following Table I. The XML-schema is metadata, which is mainly se- mantic and assists in describing the data format of the file. The metadata enables non-owner users to see information about the files, and its structure is extensible. However, any mobile device that is using this cloud service for the first time will be unable to provide such a profile, so there shall be an additional profile ex- amination to provide the test performance of the mobile device and sample relevant information. Through this function, the mo- bile device can generate an XML-schema profile and transmit it to the profile agent. The profile agent determines the required parameters for the XML-schema and creates a user profile, and then transmits the profile to the DAMM for identification.

TABLE I

DATA STRUCTURE OF FEATURE PARAMETERS

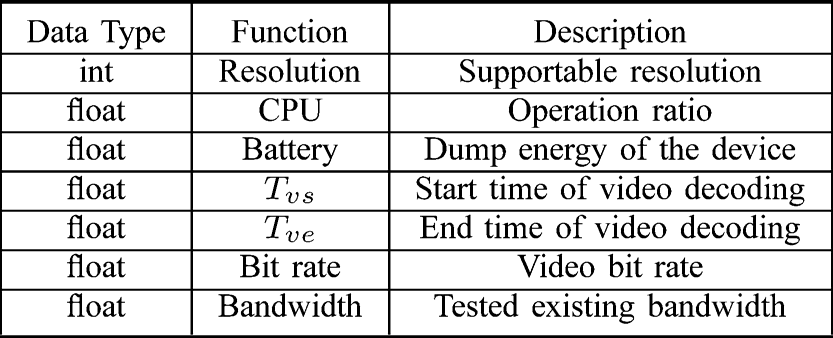
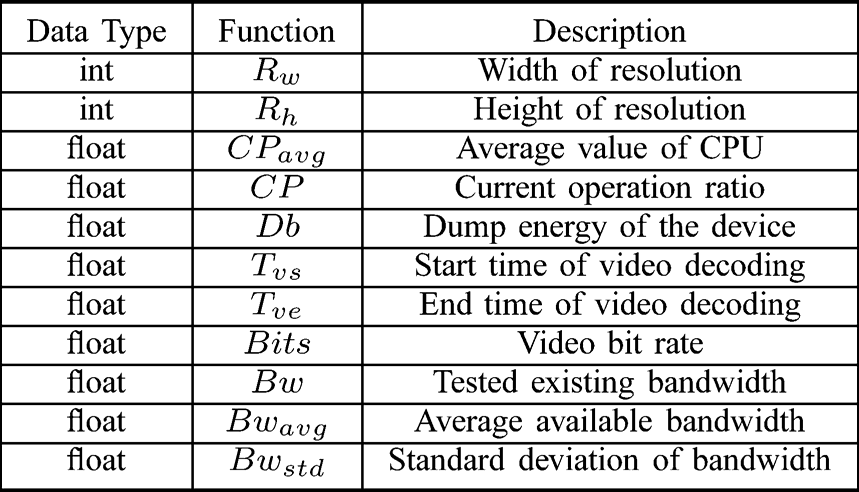


TABLE II

DATA STRUCTURE OF PREDICTION PARAMETERS



*B. Network and Device Aware Multi-Layer Management*

*(NDAMM)*

The NDAMM aims to determine the interactive communi- cation frequency and the SVC multimedia file coding parame- ters according to the parameters of the mobile device. It hands these over to the STC for transcoding control, so as to reduce the communication bandwidth requirements and meet the mo- bile device user’s demand for multimedia streaming. It consists of a listen module, a parameter profile module, a network es- timation module, a device-aware Bayesian prediction module, and adaptive multi-layer selection. The interactive multimedia streaming service must receive the user profile of the mobile de- vice instantly through the listen module. The parameter profile module records the user profile and determines the parameter list required in this study, as shown in Table II. This is provided to both the network estimation module and the device-aware Bayesian prediction module to predict the required numerical values. and represent the width and height of the sup- portable resolution for the device, and and represent the present and average CPU operating speed. and represent the existing energy of the mobile device and energy consumption rate, and , , represent the ex- isting, average and standard deviation values of the bandwidth. When this parameter form is maintained, the parameters can be transmitted to the network estimation module and the de- vice-aware Bayesian prediction module for relevant prediction.

*C. Dynamitic Network Estimation Module (DNEM)*

The DNEM is mainly based on the measurement-based pre- diction concept [32]; however, it further develops the Exponen- tially Weighted Moving Average (EWMA). The EWMA uses the weights of the historical data and the current observed value to calculate gentle and flexible network bandwidth data for the

dynamic adjustment of weights. In order to determine the pre- cise network bandwidth value, the EWMA filter estimates the network bandwidth value, shown as:

(1)

in which is the estimated bandwidth of the No. t time interval, is the bandwidth of the No. time in- terval, and is the estimation difference. For different mobile network estimations, this study considered the error correction of estimation and the overall standard difference and estimated the different bandwidths by adjusting the weights:

(2)

among which, is the moving average weight and is the stan- dard deviation weight. However, in a mobile environment, the modifier formula for the two weights is shown as:

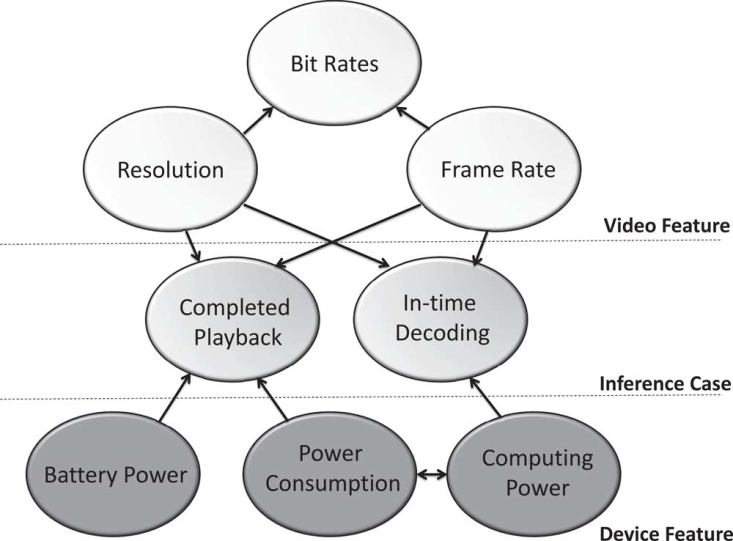


Fig. 2. Network and Device Aware Bayesian Network.

however, the question remains of how to choose an appropriate video format according to the available resources of various devices. Hereby, in order to conform to the real-time require-

if

if

if

|  |  |  |  |
| --- | --- | --- | --- |
|  |  |  |  |
|  |  | | |

if

|  |  |  |  |
| --- | --- | --- | --- |
|  |  |  |  |
|  |  | | |

(3)

(4)

ments of mobile multimedia, this study adopted Bayesian theory to infer whether the video features conformed to the decoding action. The inference module was based on the following two conditions:

• The LCD brightness does not always change This hypoth- esis aims at a hardware energy evaluation. The literature

among which, and are the corrected parameters of the weights, and is the error boundary value (set as

in this study). When the prediction error is greater than , the system shall reduce the weight modification of the predicted difference; relatively, when the prediction error is less than , the system shall strengthen the weight modification of the pre- dicted difference. When the changed bandwidth of the system is greater than the standard difference, the predicted weight will increase as the corrected value of the standard deviation is reduced. The predictor formula for the overall mobile network quality uses the standard normal state value range concept of plus-minus three standard deviations of statistics, referring to [33] to identify the stable or unstable state of the current mobile network. If the present mobile network is in a stable state, it shall conform to the following equation:

(5)

viation. The value is almost 1.128 [34]. If the network band- width value of this time cycle is within plus-minus three standard deviations of the standard value, the present mobile network will be in a stable state; otherwise it will be in a fluc- tuating state.

*D. Network and Device-Aware Bayesian Prediction Module*

*(NDBPM)*

The SVC hierarchical structure provides scalability of the temporal, spatial and quality dimensions. It adjusts along with the FPS, resolution and video variations of a streaming bitrate;

[34] states that TFT LCD energy consumption accounts for about 20%–45% of the total power consumption for different terminal hardware environments. Although the overall power can be reduced effectively by adjusting the LCD, with multimedia services, users are sensitive to brightness; they dislike video brightness that repeatedly changes. As changing the LCD brightness will influence the energy consumption evaluation value, the LCD bright- ness of the mobile device is assumed to not able to change at will during multimedia service.

• The energy of the mobile device shall be sufficient for playing a full multimedia video Full multimedia service must be able to last until the user is satisfied. This assumed condition is also the next main decision rule.

As for the three video parameters of FPS, resolution and bit

rate, the bit rate depends on the frame rate and resolution, so the Bayesian network shown Fig. 2 adopts the frame rate and resolution as the video input features and uses the bit rate as parameter considered next, in which the , and

If , , in which is the aggregate of the resolutions supported by the mobile device, and is the aggregate of videos at different frame rates. Event is

set as the event that can be decoded smoothly, and is the event

of video completely decoded by the device, defined as:

It is necessary to determine at what resolution and frame rate an entire multimedia file can be decoded completely and

smoothly. Considering the above , , , and are inde- pendent events:

(6)

speed is greater than 1, the multimedia video cannot be played smoothly, and the expectation can be set as 0. As the CP has a decreasing video resolution, this study used the multi-variable linear regression model and the Gaussian probability distribu- tion to obtain the most-possible solution, as shown in (11)–(14). The procedure is described below.

1) The equation, with k unknowns, is:

(11)

(7)

among which, , , and can be determined by the Gaussian probability distribution.

and depend on the resolution and frame rates. As for the influence of power on video decoding time, according to the power consumption of the processor in the CMOS process

2) The minimum residual value of the regression model is obtained from the given resolution and the data. If there are given data, let

. . . . . . .

technology:

and let the residual value be .

, The sum of the

among which, is the effective switched capacitance,

(8)

residual values is:

(12)

represents the operating voltage, and is the operating fre-

quency. The relationship between frequency and voltage is ex- pressed as the following equation:

(9)

3) The partial differential of vector B can be carried out by using the above equation to obtain the minimum residual value. Vector B can be expressed as:

among which, is a constant, represents the threshold voltage, and is the electron coefficient [35]. The

defined as . can be calculated by the following equation:

(13)

4) This study used k = 2 to construct the linear regression model to obtain the estimated value.

(14)

(10)

among which, represents the cycle number for the workload. Equation (10) indicates that although accelerating the operating frequency can reduce the workload effectively, it relatively ac- celerates the consumption of power. The probability distribution of can be corrected by the following condition.

if elseif else

among which, is the cycle number required for the entire multimedia video, and is the required cycle number for a one-second segment. If the energy of the mobile device is in- sufficient for playing the full video at the computing speed, the expected value will be 0. If the one-second segment decoding

5) The average variance corresponding to and the given resolution is used as the parameter

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  |  |  |  |  |
|  | | | | |

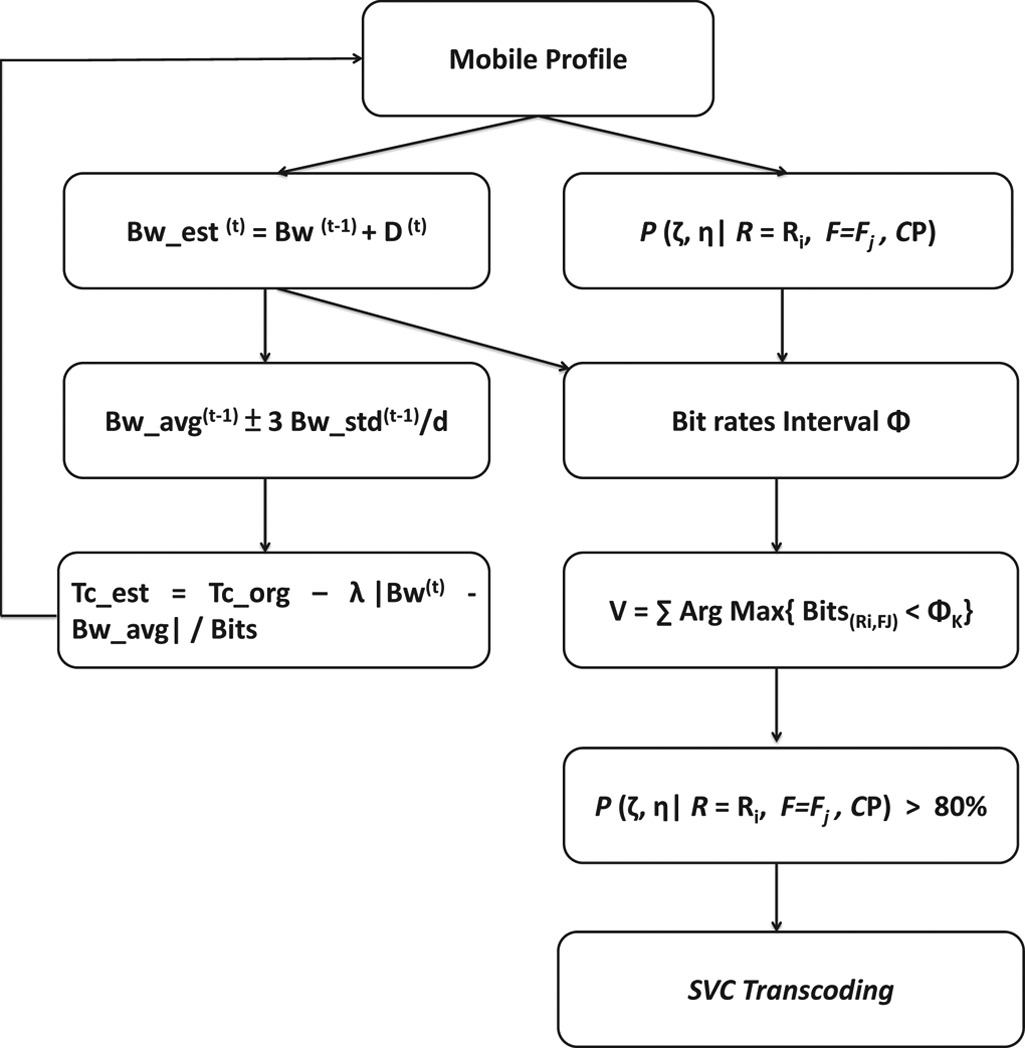
of the Gaussian probability distribution, and a plus-minus one standard deviation is taken as the probable .

*E. Proposed Adaptive Communication and Multi-Layer*

*Content Selection (ACMCS)*

When the predicted bandwidth state and the Bayesian predic- tive network are determined, the cloud system will further de- termine the communication and the required multimedia video files according to the information.

*1) Communication Decision:* A good dynamic communi- cation mechanism can reduce the bandwidth needs and the power consumption of the device resulting from excessive packet transmission, and the transmission frequency can be determined according to the bandwidth and its fluctuation ratio based on such dynamic decision-making. The transmit mode is engaged until the device finds a variation of the transmitted variables that exceeds a threshold. Although the threshold can reduce the communication frequency effectively and precisely, in this mode the mobile device must start up additional threads

for continuous monitoring; thus, the load on the device side is increased. The communication time is shown as:

(15)

among which, represents the corresponding factor of the band- width difference to the stream flow, and is the set max- imum communication time, considering the effect of (5) on net- work quality, for the correction of .

if (16)

if

When the network bandwidth difference exceeds a triple stan- dard deviation, this indicates the present network is unstable. The overall communication frequency shall incline to frequency to avoid errors; however, when the network bandwidth differ- ence is less than a triple standard deviation, the current network is still in a stable state, and the influence on bandwidth differ- ence can be corrected gradually.

*2) SVC Multi-Layer Content Decision:* SVC is an improve- ment over traditional H.264/MPEG-4 AVC coding, as it has higher coding flexibility. It is characterized by temporal scal- ability, spatial scalability and SNR scalability, allowing video transmissions to be more adaptable to heterogeneous network bandwidth. This study investigated how to determine an appro- priate multimedia video streaming service according to these three major characteristics. First, the appropriate bandwidth in- terval was determined, in which the average bandwidth

was used as the standard value and each standard deviation was the bandwidth interval segment. Let be the bandwidth aggre- gate, then:

(17) A quadruple standard difference is assumed to be the

|  |  |  |  |
| --- | --- | --- | --- |
|  |  |  |  |
|  |  | | |

boundary value. As the communication and prediction mech- anisms are constructed, the system will correct the overall threshold according to the bandwidth variation gradually, in order to avoid the bandwidth boundary exceeding the prac- tical situation. When the bandwidth interval is completed, it becomes the criterion of the video streaming bit rate. The appropriate resolution and frame rate can then be determined as the streaming data. Among these, the multimedia file shall conform to the following conditions:

(18) (19)

As mentioned above, the system will determine the ultimate resolution and a frame rate that are lower than each frequency range interval as the multimedia space and time features. The file must enable the device to decode smoothly and finish the complete decoding. The overall work flow is shown in the fol- lowing Fig. 3. When the mobile device transmits the current

Fig. 3. Flowchart of SVC Content Decision.

network and hardware features to the cloud environment, the NDAMM will predict the bandwidth at the next time point ac- cording to the bandwidth and standard deviation and will iden- tify whether the bandwidth state is stable or not. The DBPM infers whether the multimedia video, at different resolutions and frame rates, can complete smooth decoding and whether the hardware can provide complete video playback services, according to the profile examination and subsequent hardware features. When the Bayesian inference table is completed, the next communication time can be determined, and the SVC mul- timedia coding applicable for the mobile device can be provided according to the predicted and inferred network and hardware features.

IV. PROTOTYPE IMPLEMENTATION AND EXPERIMENTAL ANALYSIS

In order to validate the feasibility of the architecture, this study realized a prototype of the overall architecture and car- ried out a simulation to test its transmission quality in various networks.

*A. Prototype Implementation*

The realized prototype is shown in the following Fig. 4. This study used three sets of IBM BladeCenter HX5 servers for a cluster series connection, and used VMware vSphere Hyper- visor software in each server to simulate six computers.

A total of 16 computers were used for SVC transcoding, and

one computer was in charge of receiving and managing services. For SVC coding, Joint Scalable Video Model (JSVM) software was used for coding and decoding. The mobile device was a Samsung Galaxy Nexus, and the wireless network environment was Wi-Fi (IEEE 802.11g) or 3G for transmission service. The

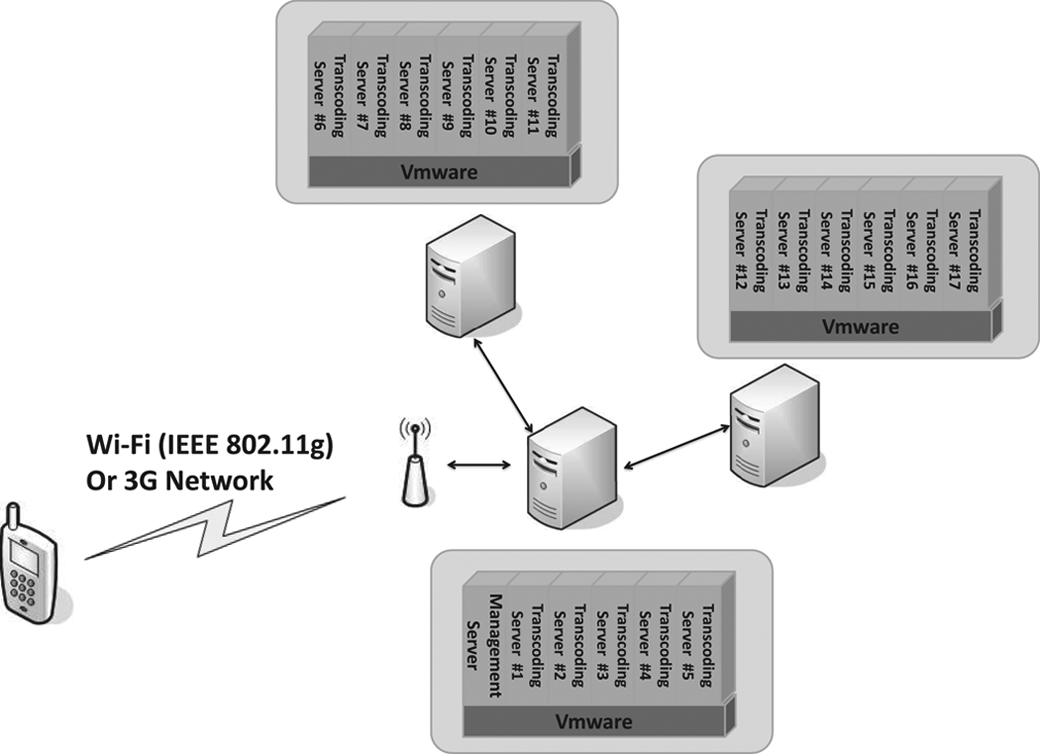
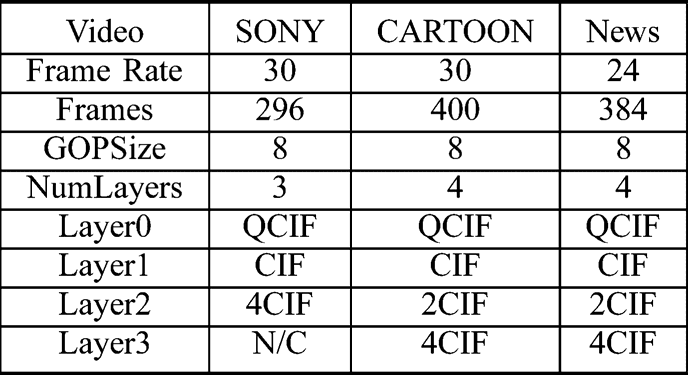


Fig. 4. Prototype of Cloud-based Mobile Media Streaming.

TABLE III

TESTED VIDEO PARAMETERS



test video and coding parameters are shown in the following

Table III.

The observed encoding speed is shown in the following Fig. 5. It was observed that the SVC coding still had diffi- culties in real-time. The coding speed in one computer was about one sheet per second, but this could be accelerated by using multiple computers in order to decentralize the coding action. However, for the SVC coding action, the decentralized calculation still depended on the key picture data. According to the test results, if more than 10 computers were used for decentralizing the calculation, the increase in speed became minimal. More detailed coding action allocation or dependent segmentation planning could be required to accelerate the demand for real-time services effectively.

*B. Video Quality Experiment*

According to the experimental result in the previous sec- tion, real-time coding was difficult; therefore, the video was transcoded off-line and the network and hardware features were turned into profiles to be handed over to the system. After coding, the wireless base station limited stream for the simulation experiment. This study first considered the video quality in a static network. The result is shown in the following Figs. 6 and 7.

Wireless multimedia streaming was carried out at a fixed

bandwidth. It was observed that in the case of stream limiting caused by a fixed bandwidth, the mechanism proposed in this study used the network mean value as the intermediate unit of the video quality. This had a better effect in comparison to the

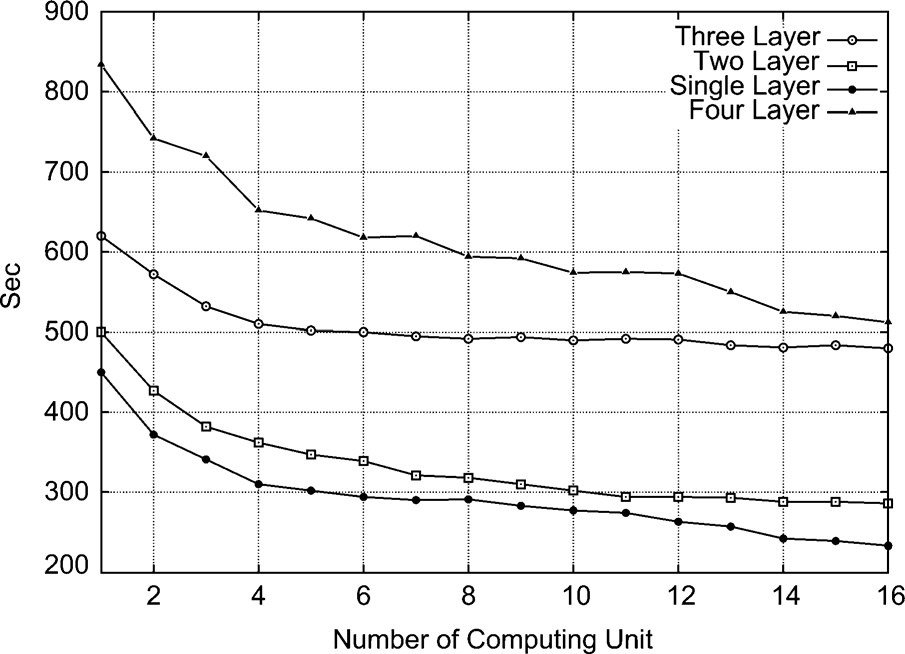


Fig. 5. Encoding Time with Multi-computing for SVC Transcoding.

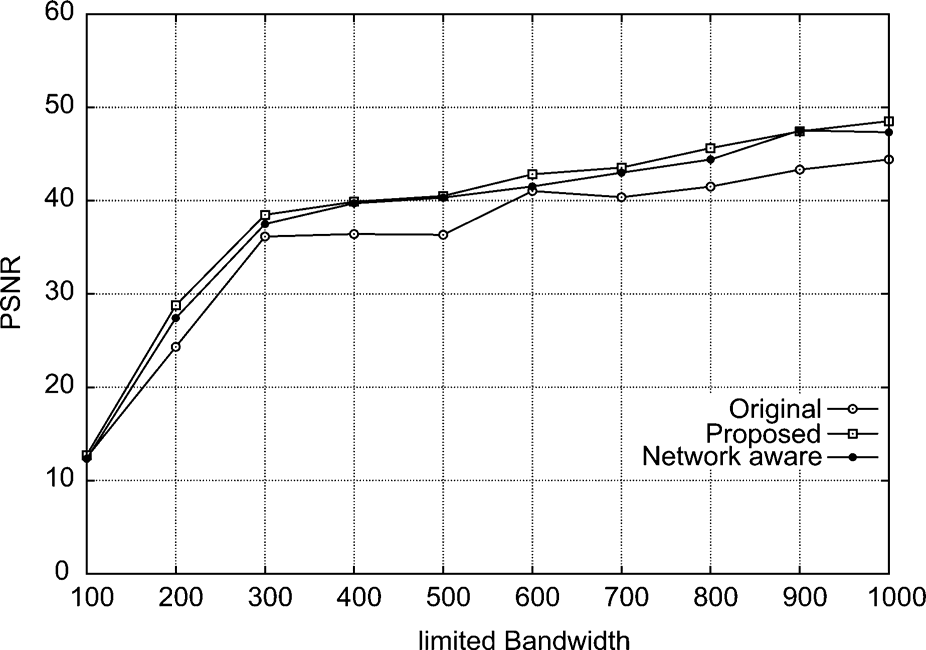


Fig. 6. Video Quality in Static Network.

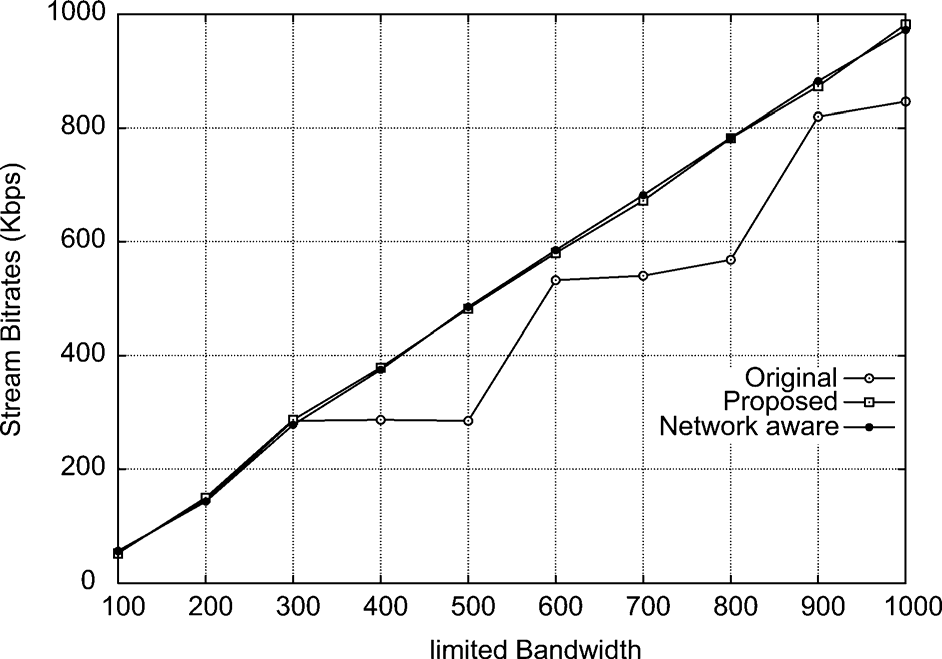


Fig. 7. Bit rates in Static Network.

other two fixed quality codings, but the overall improvement effect was not obvious. For a dynamic network, this study used bandwidth-recording software to record at intervals of

3 seconds. The user was allowed to be in three situations:

walking, driving, and taking train. Each situation was recorded for 10 minutes as a dynamic network. The network bandwidth record is shown in the following Fig. 8.

The wireless streaming experiment was carried out according to the three kinds of network quality, and the result is shown in the following Fig. 9. The network in the walking state was

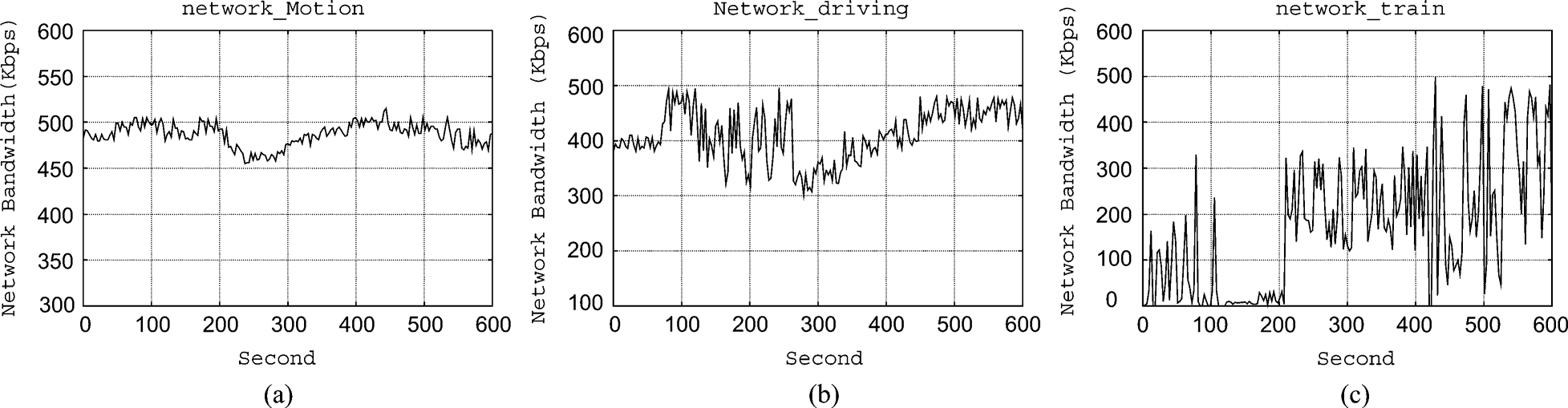


Fig. 8. Dynamic Network in Different Environments. (a) Motion. (b) Driving. (c) Train.

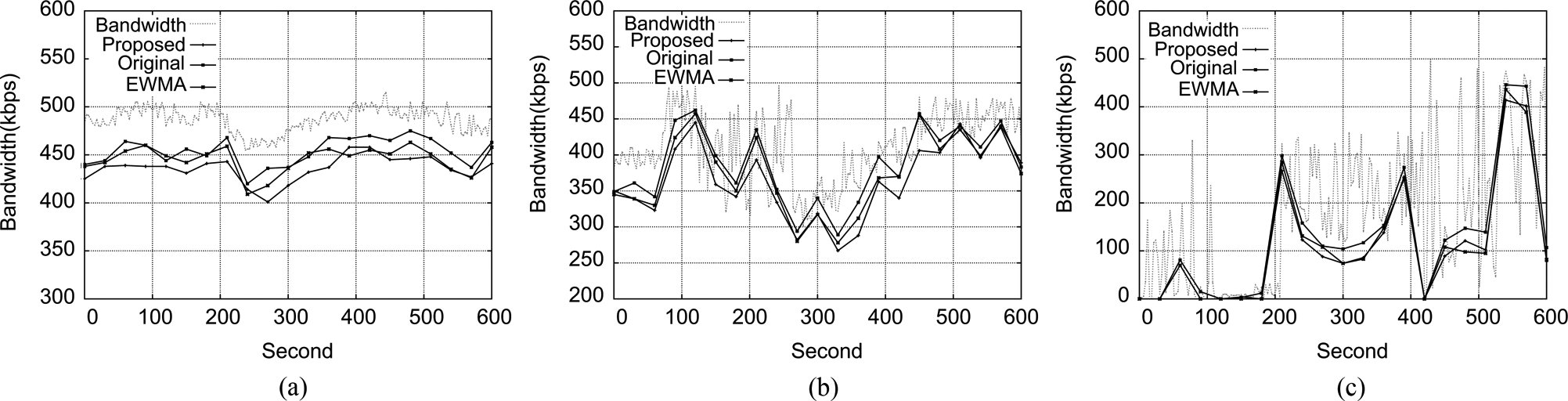


Fig. 9. Streaming Bit Rates in Different Environments. (a) Motion. (b) Driving. (c) Train.

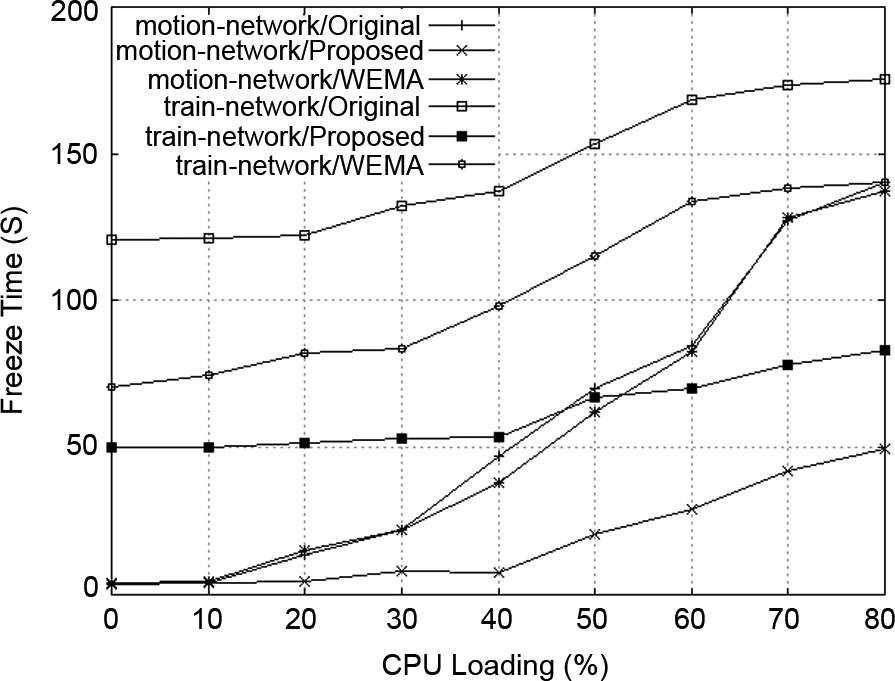


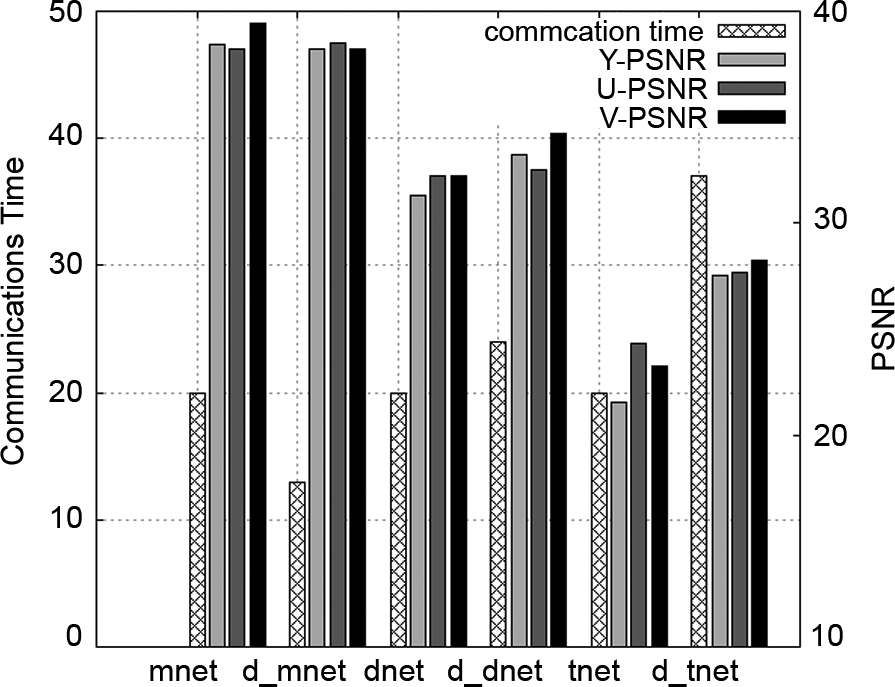
Fig. 10. Freeze Time with Different Hardware and Network.

similar to a state of fixed network bandwidth. However, in the network states of driving and taking mass rapid transit, as the quality change of the dynamic network was large, the original SVC coding reduced the video quality in comparison to the single-layer mode. This was because the changed bandwidth did not conform to the SVC layer number planning, which caused extra bandwidth loss. The proposed SVC coding consid- ered averaging the network and bandwidth standard deviation during interval planning, in order to better meet mobile network changes and improve the video quality.

*C. Hardware Feature Experiment*

For mobile multimedia services, the characteristics of the mo- bile devices had to be considered. Different battery energy were set and the number of applications was increased to change the

Fig. 11. Video Quality Analysis of Communication Mechanism.

CPU load, and the instances of video freeze in the three net- work states were compared. The result is shown in the Fig. 10. It was obvious that decoding was likely to fail at a high CPU utilization rate. In terms of battery energy, although it reduced the multimedia resolution and frame rate, complete multimedia playback service could be assured. According to the test result, the effects of the CPU were mostly seen in the frame rate, and the battery energy parameters could be improved effectively by changing the multimedia resolution.

*D. Communication Mechanism Analysis*

In order to validate the dynamic communication mecha- nism, the communication time was identified according to the aforesaid dynamic network record, and the result is shown as Fig. 11 which show the static communication mechanism

on different network condition (mnet means network\_motion condition, dnet means network\_driving condition, and tnet means network\_teaining condition) and dynamic communica- tion mechanism symboled as d\_ . As the network state for mass rapid transit was dynamic, the conversation frequency was accelerated. This study used a timing conversation mechanism to determine the influence of the communication mechanism on the quality. The result showed that in a stable network, the dynamic communication mechanism maintained the general maximum permissible time, whereas in an dynamic state, although the communication frequency was accelerated, the multimedia quality could still be maintained effectively.

V. CONCLUSION

For mobile multimedia streaming services, how to provide appropriate multimedia files according to the network and hard- ware devices is an interesting subject. In this study, a set of adap- tive networks and a device aware QoS approach for interactive mobile streaming was proposed. The DNEM and DBPM were used for the prediction of network and hardware features, and the communication frequency and SVC multimedia streaming files most suitable for the device environment were determined according to these two modules. In the experiment, the overall prototype architecture was realized and an experimental anal- ysis was carried out. The experimental data proved that the method could maintain a certain level of multimedia service quality for dynamic network environments and ensure smooth and complete multimedia streaming services. Cloud services may accelerate research on SVC coding in the future.

REFERENCES

[1] M. Armbrust, A. Fox, R. Griffith, A. D. Joseph, R. Katz, A. Konwinski, G. Lee, D. Patterson, A. Rabkin, I. Stoica, and M. Zaharia, “A view of cloud computing,” *Commun. ACM*, vol. 53, p. 508, Apr. 2010.

[2] M. F. Tan and X. Su, “Media cloud: When media revolution meets rise of cloud computing,” in *Proc. IEEE 6th Int. Symp. Service Oriented Syst. Eng.*, 2011, pp. 251–261.

[3] W. Zhu, C. Luo, J. F. Wang, and S. P. Li, “Multimedia cloud com- puting,” *IEEE Signal Process. Mag.*, vol. 28, no. 3, pp. 59–69, 2011.

[4] G. Q. Hu, W. P. Tay, and Y. G. Wen, “Cloud robotics: Architecture, challenges and applications,” *IEEE Network, Special Issue on Machine and Robotic Networking*, vol. 26, no. 3, pp. 21–28, May-Jun. 2012.

[5] S. Ferretti, V. Ghini, F. Panzieri, and E. Turrini, “Seamless support of multimedia distributed applications through a cloud,” in *Proc. IEEE*

*3rd Int. Conf. Cloud Comput. (CLOUD)*, 2010, pp. 548–549.

[6] G. Zhang, Y. G. Wen, J. Zhu, and Q. H. Chen, “On delay minimiza- tion for file-based content uploading to media cloud via collaborative wireless networks (invited paper),” in *Proc. 2011 Int. Conf. Wireless Commun. and Signal Process. (WCSP1)*, Nanjing, China, Nov. 9–11,

2011.

[7] A. Khan and K. K. Ahirwar, “Mobile cloud computing as a future of

mobile multimedia database,” *Int. J. Comput. Sci. Commun.*, vol. 2, no.

1, pp. 219–221, 2011.

[8] L. Zhou, X. Wang, W. Tu, G. Mutean, and B. Geller, “Distributed scheduling scheme for video streaming over multi-channel multi-radio multi-hop wireless networks,” *IEEE J. Select. Areas Commun.*, vol. 28,

no. 3, pp. 409–419, Apr. 2010.

[9] K. E. Psannis, Y. Ishibashi, and M. G. Hadjinicolaou, “QoS for wireless interactive multimedia streaming,” in *Proc. ACM Workshop QoS and Security for Wireless and Mobile Networks*, 2007, pp. 168–171.

[10] S. M. Saranya and M. Vijayalakshmi, “Interactive mobile live video learning system in cloud environment,” in *Proc. Int. Conf. Recent*

*Trends in Inf. Technol. (ICRTIT)*, 2011, pp. 673–677.

[11] W. Song, J. Hai, C. Jie, and F. Kaiqin, “A novel cache scheme for cluster-based streaming proxy server,” in *Proc. IEEE Int. Conf. Dis- tributed Computing Syst. Workshops*, 2005, pp. 727–733.

[12] Y. Hang, C. C. J. Kuo, and I. Ahmad, “Energy efficiency in data cen- ters and cloud-based multimedia services: An overview and future di- rections,” in *Proc. Int. Green Computing Conf.*, 2010, pp. 375–382.

[13] P. Fan, J. Wang, Z. Zheng, and M. R. Lyu, “Toward optimal deploy-

ment of communication-intensive cloud applications,” in *Proc. IEEE Int. Conf. Cloud Computing*, 2011, pp. 460–467.

[14] Z. Zheng, Y. Zhang, and M. R. Lyu, “Cloudrank: A QoS-driven com- ponent ranking framework for cloud computing,” in *Proc. IEEE Symp. Reliable Distributed Syst.*, 2010, p. 18493.

[15] D. Diaz-Sanchez, F. Almenarez, A. Marin, D. Proserpio, and P. A.

Cabarcos, “Media cloud: An open cloud computing middleware for content management,” *IEEE Trans. Consum. Electron.*, vol. 57, pp.

970–978, 2011.

[16] J. Feng, P. Wen, J. Liu, and H. Li, “Elastic stream cloud (ESC): A

stream-oriented cloud computing platform for rich internet applica- tion,” in *Proc. Int. Conf. High Performance Computing and Simulation*

*(HPCS)*, 2010, pp. 203–208.

[17] A. Garcia and H. Kalva, “Cloud transcoding for mobile video content delivery,” in *Proc. 2011 IEEE Int. Conf. Consumer Electronics (ICCE)*,

2011, pp. 379–380.

[18] X. Jin and Y. K. Kwok, “Cloud assisted p2p media streaming for band- width constrained mobile subscribers,” in *Proc. IEEE Int. Conf. Par-*

*allel and Distributed Syst.*, 2010, pp. 800–805.

[19] I. Trajkovska, J. Salvachúa, and A. M. Velasco, “A Novel P2P and Cloud Computing Hybrid Architecture for Multimedia Streaming with QoS Cost Functions,” *Proc. ACM Multimedia*, p. 1227230, 2010.

[20] Y. Liu, Z. Yang, X. Deng, J. Bu, and C. Chen, “Media browsing for mo- bile devices based on resolution adaptive recommendation,” in *Proc.*

*Int. Conf. Commun. and Mobile Computing*, 2009, pp. 285–290.

[21] W. Yin, X. Zhu, and C. W. Chen, “Contemporary ubiquitous media services: Content recommendation and adaptation,” in *Proc. IEEE PerCom Workshop Pervasive Communities and Service Clouds*, 2011, p. 12934.

[22] C. F. Lai and M. Chen, “Playback-rate based streaming services for

maximum network capacity in IP multimedia subsystem,” *IEEE Syst. J.*, vol. 5, no. 4, pp. 555–563, 2010.

[23] C. F. Lai and A. V. Vasilakos, “Mobile multimedia services over cloud computing,” *IEEE COMSOC MMTC E-Lett.*, vol. 5, no. 6, 2010.

[24] S. Y. Chang, C. F. Lai, and Y. M. Huang, “Dynamic adjustable multi- media streaming service architecture over cloud computing,” *Comput.*

*Commun.*, vol. 35, no. 15, pp. 1798–1808, Sep. 2012.

[25] H. Sohn, H. Yoo, Y. B. Lee, C. S. Kim, W. D. Neve, and Y. M. Ro, “MPEG-21-Based scalable bitstream adaptation using medium grain scalability,” in *Proc. IEEE Region 10 Conf.*, 2008, pp. 1–5.

[26] H. Choi, J. W. Kang, and J. G. Kim, “Dynamic and interoperable adap- tation of SVC for QoS-enabled streaming,” *IEEE Trans. Consum. Elec-*

*tron.*, vol. 53, no. 2, pp. 384–385, 2007.

[27] Z. Huang, C. Mei, L. E. Li, and T. Woo, “CloudStream: Delivering High-quality Streaming Videos through a Cloud-based SVC Proxy,” in *Proc. IEEE INFOCOM*, , 2011, pp. 201–205.

[28] L. N. Zhang, C. Yuan, and Y. Z. Zhong, “A novel SVC VoD system with rate adaptation and error concealment over GPRS/EDGE net-

work,” in *Proc. Congr. Image and Signal Process.*, , vol. 1, pp.

349–354.

[29] P. Chen, X. Jeongyeon, L. B. Lee, M. Kim, S. Hahm, B. Kim, and K.

L. Park, “A network-adaptive SVC streaming architecture,” in *Proc. Int. Conf. Advanced Commun. Technol.*, 2007, vol. 2, pp. 955–960.

[30] N. Ramos and S. Dey, “A device and network-aware scaling frame-

work for efficient delivery of scalable video over wireless networks,”

in *Proc. IEEE Int. Symp. Personal, Indoor and Mobile Radio Commun.*,

2007, pp. 1–5.

[31] H. Lee, Y. Lee, J. Lee, D. Lee, and H. Shin, “Design of a mobile video streaming system using adaptive spatial resolution control,” *IEEE*

*Trans. Consum. Electron.*, vol. 55, no. 3, pp. 1682–1689, 2009.

[32] H. Yoon and J. Kim, “Measurement-based achievable throughput esti- mation in IEEE 802.11a WLANs,” *IEEE Commun. Lett.*, vol. 11, no.

9, pp. 714–716, Sep. 2007.

[33] Y. Lin, S. Cheng, W. Wang, and Y. Jin, “Measurement-based TFRC: Improving TFRC in heterogeneous mobile networks,” *IEEE Trans.*

*Wireless Commun.*, vol. 5, no. 8, pp. 1971–1975, Aug. 2006.

[34] M. Kim and B. Nobe, “Mobile network estimation,” in *Proc. ACM Mobicom*, , Jul. 2001, pp. 298–309.

[35] Y. K. Lai, Y. F. Lai, and P. Y. Chen, “Content-based LCD backlight power reduction with image contrast enhancement using histogram analysis,” *J. Display Technol.*, vol. 7, no. 10, pp. 550–555, 2011.

[36] S. Yaldiz, A. Demir, and S. Tasiran, “Stochastic modeling and opti-

mization for energy management in multicore systems: A video de- coding case study,” *IEEE Trans. CAD Integr. Circuits Syst.*, vol. 27, p.

1264277, Jul. 2008.

**Chin-Feng Lai** (M’07) was born in Taiwan in 1980, and is now an assistant professor at Institute of Computer Science and Information Engineering, National Ilan University since 2011. He received the Ph.D. degree in Department of Engineering Science from the National Cheng Kung University, Taiwan, in 2008. His research interests include multimedia communications, sensor-based healthcare and em- bedded systems. After receiving the Ph.D. degree, he has authored/co-authored over 60 refereed papers in journals, conference, and workshop proceedings

about his research areas within two years. He was selected as an honorary member of the Phi Tau Phi Scholastic Honor Society of the Republic of China in 2009 and TC member of Multimedia Systems & Applications Technical Committee (MSATC), IEEE Circuits and Systems Society, 2009. He is also a project leader of several international, industrially-funded multimedia projects. Now he is making efforts to publish his latest research in the IEEE TRANSACTIONS ON MULTIMEDIA and the IEEE TRANSACTIONS ON CIRCUIT AND SYSTEMS IN VIDEO TECHNOLOGY. He is also a member of the IEEE Circuits and Systems Society and IEEE Communication Society.

**Honggang Wang** (SM’06) received his Ph.D. in Computer Engineering at University of Ne- braska-Lincoln in 2009. He is an assistant professor at University of Massachusetts Dartmouth. His re- search interests include wireless communication and networking, sensor networks, multimedia communi- cations, social networks and wireless healthcare. He has published more than 60 papers in his research areas. He is the winner of the Best Paper Award of the 2008 IEEE Wireless Communications and Networking Conference (WCNC). He serves as an

Associate Editor of Wiley’s Security and Communication Networks (SCN) Journal. He is the TPC member for IEEE INFOCOM 2013, IEEE ICC 2011, IEEE Globecom 2010–2012, and IEEE WiMob 2008, 2009. He currently serves as a Board member of IEEE MMTC (Technical Committee on Multimedia Communications) Services and Publicity.

**Han-Chieh Chao** (SM’92) is a joint appointed Full Professor of the Department of Electronic Engineering and Institute of Computer Science & Information Engineering where he also serves as the president of National Ilan University, I-Lan, Taiwan. He has been appointed as the Director of the Computer Center for Ministry of Education starting from September 2008 to July 2010. His research interests include High Speed Networks, Wireless Networks, IPv6 based Networks, Digital Creative Arts and Digital Divide. He received his M.S. and

Ph.D. degrees in Electrical Engineering from Purdue University in 1989 and

1993 respectively. He has authored or co-authored 4 books and has published about 280 refereed professional research papers. He has completed 100 MSEE thesis students and 3 Ph.D. students. Dr. Chao has received many research awards, including Purdue University SRC awards, and NSC research awards (National Science Council of Taiwan). He also received many funded research grants from NSC, Ministry of Education (MOE), RDEC, Industrial Tech- nology of Research Institute, Institute of Information Industry and FarEasTone Telecommunications Lab. Dr. Chao has been invited frequently to give talks at national and international conferences and research organizations. Dr. Chao is the Editor-in-Chief for IET Communications, Journal of Internet Technology, International Journal of Internet Protocol Technology and International Journal of Ad Hoc and Ubiquitous Computing. Dr. Chao has served as the guest editors for Mobile Networking and Applications (ACM MONET), IEEE JSAC, IEEE Communications Magazine, Computer Communications, IEE Proceedings Communications, the Computer Journal, Telecommunication Systems, Wireless Personal Communications, and Wireless Communications

& Mobile Computing. Dr. Chao is a Fellow of IET (IEE). He is a Chartered

Fellow of British Computer Society.

**Guofang Nan** (M’04) received the Bachelor degree from Beijing Institute of Technology in 1998, and his M.Sc. degree and Ph.D. degree in the Institute of Systems Engineering from Tianjin University in 2002 and 2004, respectively. He has been with the Tianjin University in China since 2006 as an associate professor. His research interests include optimization and control of complex network system, information systems and intelligent computing. He worked as a visiting scholar at Department of Computing in Imperial College in 2005, and worked

as postdoctoral research associate in Polytechnic Institute of Turin from 2008 to 2009. He was selected in the Program for New Century Excellent Talents in Universities in 2008. He has published more than 30 journal and conference papers in recent years.