Game-theoretic Scalable Offloading for Video Streaming Services over LTE and WiFi Networks

Donghyeok Ho, Gi Seok Park, and Hwangjun Song

Abstract— This paper presents a game-theoretic scalable offloading system that provides seamless video streaming services by effectively offloading parts of video traffic in all video streaming services to a WiFi network to alleviate cellular network congestion. The system also consolidates multiple physical paths in a cost-effective manner. In the proposed system, the fountain encoding symbols of compressed video data are transmitted through long term evolution (LTE) and WiFi networks concurrently to flexibly control the amount of video traffic through the WiFi network as well as mitigate video quality degradation caused by wireless channel errors. Furthermore, the progressive second price auction mechanism is employed to allocate the limited LTE resources to multiple user equipment in order to maximize social welfare while converging to the ε-Nash equilibrium. Specifically, we design an application-centric resource valuation that explicitly considers both the realistic wireless network conditions and characteristics of video streaming services. In addition, the scalability and convergence properties of the proposed system are verified both theoretically and experimentally. The proposed system is implemented using network simulator 3. Simulation results are provided to demonstrate the performance improvement of the proposed system.

Index Terms— Fountain code, Game theoretic resource allocation, Progressive second price auction, Scalable traffic offloading, Video streaming

1 Introduction

THE growth in global mobile data traffic has been tremendous. According to the Cisco visual network index [1], the overall mobile data traffic is expected to grow at a compound annual growth rate (CAGR) of 47% from 2016 to 2021, reaching 49 exabytes per month by 2021, which represents a nearly sevenfold increase over 2016. The main catalyst of this exponential rise in mobile data traffic is the explosive increase in smartphones, tablets, laptops, and advanced multimedia applications such as YouTube and Netflix. In [1], Cisco estimates that mobile video traffic will attain at a CAGR of 54% per year for the period 2016–2021 and account for more than 78% of mobile data traffic in 2021. Unfortunately, this massive surge in mobile video traffic has caused an unprecedented degree of pressure on the limited capacity of cellular networks, and ultimately degrades the user-perceived video quality [2]. The provision of support for the explosive traffic growth in mobile networks is quite challenging. The most straightforward solution is to increase the cellular network capacity by adding base stations, or by upgrading the cellular networks to next generation advanced networks such as long term evolution (LTE) [3], LTE-Advanced (LTE-A) [4], and worldwide interoperability for microwave access (WiMAX) release 2 (IEEE 802.16 m) [5]. However, simply increasing cellular network capacity may not always be economical. Even in fourth-generation (4G) network, bandwidth remains a limited resource because of the rapid growth of user demand for advanced multimedia applications. Moreover, these approaches require both large capital (CAPEX) as well as operational (OPEX) expenditures.

Mobile data offloading [6] refers to the use of complementary networks such as WiFi and femtocell to deliver mobile data traffic originally planned for transmission over cellular networks. In [1], Cisco estimates that 63% of global mobile data traffic will be offloaded to WiFi or small-cell networks by 2021. In recent years, global network operators such as AT&T, T-Mobile, Orange, and China Mobile have deployed carrier-grade access points (APs) in high density user locations like malls, markets or cafes to increase the volume of offloaded traffic [7], [8]. WiFi offloading is emerging as a cost-effective solution for operators to accommodate the tremendous growth of mobile data traffic, because it is much cheaper to install new WiFi APs than to upgrade cellular network equipment. There are practical issues faced by operators when offloading mobile data traffic to a WiFi network, such as the quality of the WiFi experience, limitations of the WiFi planning tools, deployment of WiFi hotspots, manual offloading, and pricing [9]. The foremost challenge associated with mobile data offloading is user experience [6]. Service providers must ensure consistent user experience and service continuity. However, it is not easy to main-

- *Donghyeok Ho is with the Division of IT Convergence Engineering, Pohang University of Science and Technology (POSTECH), San 31, Hyoja-dong, Nam-gu, Pohang-si, Republic of Korea 790-784. E-mail: positiveho@postech.ac.kr.*
- *Gi Seok Park is with the Division of IT Convergence Engineering, Pohang University of Science and Technology (POSTECH), San 31, Hyoja-dong, Nam-gu, Pohang-si, Republic of Korea 790-784. E-mail: kiseok@postech.ac.kr.*
- *Hwangjun Song is with the Department of Computer Science and Engineering, Pohang University of Science and Technology (POSTECH), San 31, Hyoja-dong, Nam-gu, Pohang-si, Republic of Korea 790-784. E-mail: hwangjun@postech.ac.kr.*

xxxx-xxxx/0x/$xx.00 © 200x IEEE Published by the IEEE Computer Society

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TMC.2017.2748592, IEEE Transactions on Mobile Computing

tain the quality-of-experience (QoE) of a video streaming service when WiFi offloading is performed. This is due to performance limitations in terms of data rate and service range needed to deliver high-quality video streaming services in a single wireless network environment [10]. Recently, a cooperation among heterogeneous wireless networks has attracted a huge amount of interest.

Auction theory, which is a subfield of game theory, provides useful mathematical tools to analyze the complex interactions among interdependent rational players (i.e., buyers and sellers). Auction has been widely exploited to provide mobile data offloading solutions since it is suitable for a large-scale decentralized and competitive environment. The fundamental design issues when applying auction theory in various wireless communications and networking domains is presented in [11]. In [12], a reverse auction-based incentive mechanism is proposed to motivate user equipment (UEs) to assist mobile data offloading, where the cellular operator offers an incentive to UEs in exchange for delay during which the data are not received via the cellular network. In the iterative double auction mechanism [13], the buyers (cellular operators) and sellers (AP owners) submit their bids to the broker to maximize their payoffs, and the broker then returns the auction outcome to maximize the social welfare (i.e., the sum of players' resource valuations). In [14], auction-based WiFi offloading in vehicular environments is proposed to improve the utility of vehicle users and increase the revenue of cellular network operator. In [15], the opportunistic decision-making algorithm is designed to maximize the vehicle user's satisfaction, which is modeled as the downloading cost and downloading delay. In [16], a seamless WiFi-based Internet access scheme is proposed to overcome high packet loss rate and frequent connection disruptions. By configuring all APs with the same medisum access control (MAC) and internet protocol (IP) addresses, the vehicle users get a graceful illusion that only one AP exists. In [17], a combinational reverse auction is proposed to implement a marketplace for mobile data offloading. AP owners lease the unused network capacity and the cellular operator rents the available AP bandwidth. When performing mobile data offloading, efficient resource allocation is in particular important for multimedia applications, since the necessary bandwidth for these applications is huge and varies continuously based on the content. However, most of the previous studies allocate resources without considering the actual benefit to users in terms of video quality. Thus, these studies are not appropriate for the multimedia resource allocation scenario over time-varying wireless networks.

In this paper, we propose a game-theoretic scalable offloading system for video streaming services over heterogeneous wireless networks. The main contributions are summarized as follows.

- We propose scalable offloading, which finely controls the amount of traffic transmitted through LTE and WiFi networks, to provide seamless video streaming services and alleviate cellular network congestion. When using multiple physical paths with diverse delays for proposed scalable offloading, the packets may arrive out-of-order to the receiver [18]. We adopt the fountain code [19], [20] as a forward error correction (FEC) scheme to solve the packet reordering and packet loss problems in error-prone wireless networks while adjusting the amount of offloading traffic in a flexible scale.
- We frame the scalable offloading problem as a progressive second price (PSP) auction game [21] in which LTE resources are shared among multiple UEs. More importantly, we define an application-specific resource valuation function based on the rate-distortion models of the encoded video sequences.
- We design a feedback-based UE bid decision algorithm with low communication and computational overheads to improve the convergence speed to an equilibrium point. In addition, the scalability of the proposed algorithm is verified both theoretically and experimentally.
- For performance verification under the large-scale and decentralized wireless network environment, we implement the proposed system using network simulator 3 (NS-3) open source software [22].

The remainder of this paper is organized as follows. The background and related work are reviewed in section 2. The details of the proposed game-theoretic scalable offloading are described in section 3 and experimental results are given in section 4. Concluding remarks are provided in section 5.

2 Background and Related Work

In this section, we briefly review fountain code and the PSP auction mechanism in sections 2.1 and 2.2, respectively. We then introduce the related works about mobile data offloading in section 2.3.

2.1 Fountain Code

Fountain codes such as Luby Transform (LT) [20], Online [23], and Raptor [24], [25] are block-based FEC schemes that provide high coding efficiency, flexibility, and low encoding and decoding processing times. Fig. 1 provides an example of an LT code deployment scheme for a video streaming service. For fountain encoding, the data stream should be divided into source blocks. Each source block is partitioned into source symbols of a predefined size. As

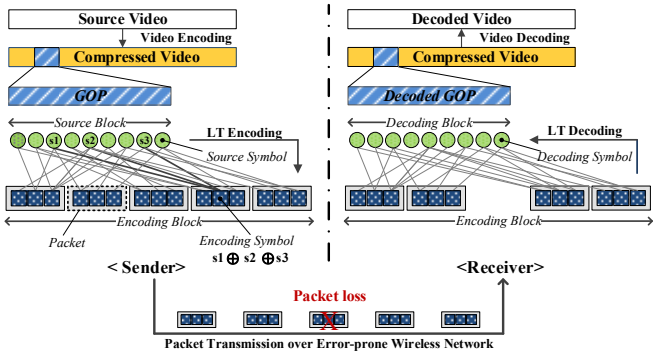

Fig. 1. Example of LT code deployment scheme for a video streaming service over error-prone wireless networks.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TMC.2017.2748592, IEEE Transactions on Mobile Computing

shown in Fig. 1, an encoding symbol is mapped to certain source symbols based on the degree distribution probability. Each encoding symbol is generated by performing bitwise XOR operations on selected source symbols. This process is repeated until the last encoding symbol is created.

In general, the number of encoding symbols (k') required for successful fountain decoding is calculated by

$$k' = (1+\delta)\cdot k,$$

where k is the number of source symbols and δ is the symbol overhead with a very small real number. The aforementioned equation means that the number of received encoding symbols must be slightly larger than k to reconstruct source symbols successfully. Fountain code can generate encoding symbols infinitely from a given finite number of source symbols (this is called rateless code). That is,

$$c = k/n,$$

where c is the code rate and n is the number of encoding symbols. Thus, if a sufficient number of encoded symbols are available, the receiver can reconstruct all the source symbols even if some packets are lost or out-of-order. The decoding process is equivalent to solving linear equations. In general, the simple decoding algorithm through message passing [19] is widely used because of its low complexity. The decoding complexity of LT code is known to be $O(k\log(k))$. Raptor code [24], [25], which is included in the 3rd generation partnership project (3GPP) [26] and digital video broadcasting-handheld (DVB-H) system [27], achieves linear time encoding and decoding complexity through a pre-coding stage of the source symbols [28]. In recent years, a more advanced Raptor code with greater flexibility and improved reception overhead, called RaptorQ, has been introduced into the IETF [29].

2.2 Progressive Second Price Auction Mechanism

A PSP auction [21], which is a generalized Vickrey auction [30] (also called second-price sealed-bid auction), is a decentralized mechanism to share variable-sized resources among multiple users in a network framework. It is assumed that the total amount of network resource is R, and there are M players participating in the auction, indexed by $i \in \mathbf{I} = \{1,2,...,M\}$. Players submit a bid sequentially and modify their bids as a reply to those submitted by others with a bid change fee ε. The i_{th} player's bid is defined by $b_i = (q_i, p_i) \in \boldsymbol{B}_i$ $\left(=[0,R]\times[0,\infty)\right)$, where q_i is the desired quantity of resources, p_i is the offered price per unit, and $\boldsymbol{B}_i$ is the set of possible bid actions. The bid profile is $\boldsymbol{b} = \{b_1,...,b_M\} = (b_i, \boldsymbol{b}_{-i}) \in \boldsymbol{B}$, where $\boldsymbol{b}_{-i} = \{b_1,...,b_{i-1},b_{i+1},...b_M\} \in \boldsymbol{B}_{-i}$ denotes the bid profile of the i_{th} player's opponents, and $\boldsymbol{B} = \prod_{i\in\mathbf{I}} \boldsymbol{B}_i$. After collecting bids, the auctioneer executes two functions: resource allocation and payment computation. The allocation rule of the i_{th} player is defined by $s_i(\boldsymbol{b}) = (a_i(\boldsymbol{b}), \tau_i(\boldsymbol{b})) \in \boldsymbol{B}_i$, where $a_i(\boldsymbol{b})$ is the resources allocated to the i_{th} player, and $\tau_i(\boldsymbol{b})$ represents the payment for $a_i(\boldsymbol{b})$. Note that p_i is the price per unit and $\tau_i(\boldsymbol{b})$ is the total cost. An allocation rule is deemed feasible if it satisfies the following.

$$\begin{aligned} &\sum_{i\in\mathbf{I}} a_i(\boldsymbol{b}) \le R, \text{ for } \forall \boldsymbol{b} \in \boldsymbol{B}, \\ &0 \le a_i(\boldsymbol{b}) \le q_i, \text{ for } \forall i \in \mathbf{I}, \\ &0 \le \tau_i(\boldsymbol{b}) \le q_i \cdot p_i, \text{ for } \forall i \in \mathbf{I}. \end{aligned} \tag{1}$$

The allocation rule satisfying (1) is depicted in Fig. 2. When the i_{th} player participates in the auction with the bidding price p_i, the amount of available resource (r_i^{avail}) that can be assigned to the i_{th} player is obtained by $r_i^{avail}(p_i, \boldsymbol{b}_{-i}) = \left[R - \sum_{k\in\mathbf{I}_{-i}, p_k \ge p_i} q_k\right]^+$, where $[x]^+ = \max\{x, 0\}$. Then, $a_i(\boldsymbol{b})$ is determined by $a_i(\boldsymbol{b}) = \min\left\{q_i, r_i^{avail}(p_i, \boldsymbol{b}_{-i})\right\}$. The payment $\tau_i(\boldsymbol{b})$ is calculated based on the externality that the player imposes on others through his or her participation using $\tau_i(\boldsymbol{b}) = \sum_{k\in\mathbf{I}_{-i}} p_k \cdot \left[a_k(0, \boldsymbol{b}_{-i}) - a_k(b_i, \boldsymbol{b}_{-i})\right]$.

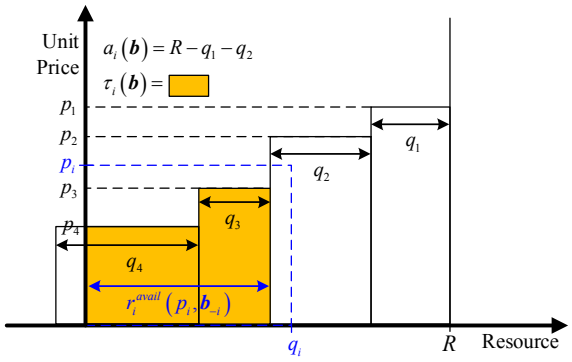

Fig. 2. Example of the PSP auction-based resource allocation mechanism for the i_{th} player with given opponent profiles.

In game theory, a Nash equilibrium is a configuration of player strategies such that nobody can improve his or her utility by unilaterally changing strategy. In PSP, the ε-Nash equilibrium is formally defined as follows.

ε-Nash Equilibrium: For given opponent profiles, define the set of ε-best replies by

$$\boldsymbol{B}_i^{\varepsilon}(\boldsymbol{b}_{-i}^*) = \left\{b_i^* \in \boldsymbol{B}_i : u_i(b_i^*, \boldsymbol{b}_{-i}^*) \ge u_i(b_i, \boldsymbol{b}_{-i}^*) - \varepsilon, \text{ for } \forall b_i \in \boldsymbol{B}_i \text{ and } \forall i \in \mathbf{I}\right\},$$

where $u_i(\cdot)$ represents the utility of the i_{th} player, b_i^* is the ε-best reply of the i_{th} player with regard to $\boldsymbol{b}_{-i}^*$, and ε $(\varepsilon > 0)$ can be interpreted as a bid fee each time players submit a bid. An ε-Nash equilibrium is a fixed point of $\boldsymbol{B}^{\varepsilon}(\boldsymbol{b}) = \prod_{i\in\mathbf{I}} \boldsymbol{B}_i^{\varepsilon}(\boldsymbol{b}_{-i}^*)$. Under elastic demand, the PSP auction is incentive compatible and stable, in that it has a truthful ε-Nash equilibrium in which all players bid at prices equal to their marginal valuation of the resource. In addition, PSP is economically efficient because the equilibrium allocation maximizes the social welfare.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TMC.2017.2748592, IEEE Transactions on Mobile Computing

2.3 Related Work

Thus far, a considerable amount of research effort has been devoted to alleviating cellular network congestion by means of mobile data offloading. A comprehensive survey of mobile data offloading, including technical and business perspectives, was offered in [6]. Dimatteo et al. [31] evaluated the potential costs and gains for WiFi offloading in a metropolitan area by means of using large scale high mobility traces with the opportunistic networking paradigm. Lee et al. [32] showed the promising results of WiFi offloading by means of using empirical pedestrian traces. The AP deployment algorithm based on realistic mobility characteristics was proposed in [33], and more granular deployment algorithms in a city-wide scenario was studied in [34]. In addition, the feasibility of WiFi offloading in high-speed vehicular environments was demonstrated in [35], [36], [37]. Motivated by these potential economic benefits, thus far efficient offloading mechanisms have been proposed. Wiffler [38] was developed to augment third-generation (3G) network capacity for delay-tolerant applications over vehicular networks. MultiNets [39] is a seamless offloading system that allows a mobile device to automatically switch to the most suitable network interface based on user-defined policies such as energy saving and throughput. Wiffler and MultiNets make offloading decisions based on historical bandwidth records and signal strength that do not reflect the current traffic load. Thus, UEs cannot reasonably respond to real-time wireless network states. MADNet [40] is a collaborative mobile data offloading architecture, which is designed mainly for extending mobile battery life. However, MADNet has focused on the vertical handover. In addition to added implementation cost and effort, these approaches require sophisticated architecture and mutual agreement between stakeholders. Although interoperability is supported, an abrupt bandwidth change and substantial handover latency are unavoidable [41]. These negative effects lead to bursty packet loss of video data and error propagation in the successive frames until the next I-frame is received successfully. For these reasons, the vertical handover may seriously degrade the QoE of a video streaming service [42], [43].

Path diversity means that a mobile user can establish multiple physical paths over heterogeneous wireless access networks. Path diversity has some advantages [44]: it can eliminate problematic vertical handover delay and increase the transmission rate and reliability, because the physical paths are independent of each other. In recent years, some efficient multipath transport protocols [18], [45], [46] and end-to-end virtual path construction systems [44], [47] have been proposed to overcome the limitations of single path transmission by harmonizing the multiple physical paths simultaneously. In addition, an access network discovery and selection function (ANDSF), which is an optional network element in the 3GPP evolved packet core (EPC), has been specified in 3GPP standards [48], [49] for efficient interworking between 3GPP and non-3GPP access networks. ANDSF assists UEs in discovering non-3GPP access networks such as WiFi and WiMAX, and provides UEs with operator-defined policies to support network selection and traffic routing decisions. However, several limitations of the ANDSF have been expounded in recent literature [50]. The current ANDSF is not combined with the actual usage of the access networks. If ANDSF treats the traffic information of several access networks, traffic offloading techniques can efficiently resolve the congestion of 3GPP access networks. In addition to the network-side information, a considerable amount of intelligence is available locally in UEs, such as battery, mobility, and user preference (e.g., service type, quality, and cost). This information can be used to improve the performance of network selection and decisions. 3GPP also offers some alternative offloading mechanisms that take advantage of the hybrid architecture of the EPC such as local IP access (LIPA), selected IP traffic offload (SIPTO) [51], and IP flow mobility (IFOM) [52].

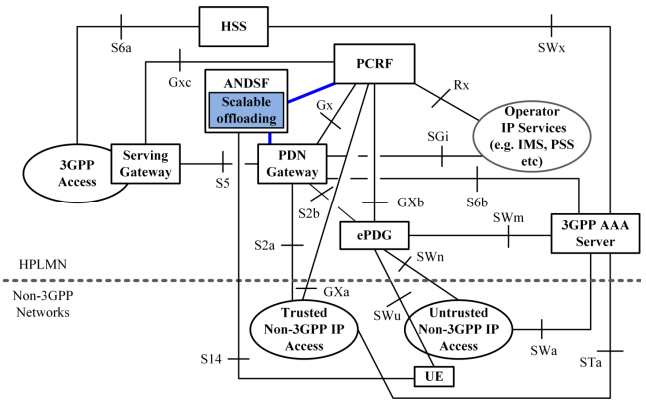

Fig. 3. Proposed interworking architecture between 3GPP and non-3GPP access networks based on the 3GPP standard architecture.

3 Proposed Game-Theoretic Scalable Offloading System

The goal of the proposed system is to alleviate LTE network congestion by offloading parts of video traffic through a WiFi network in a flexible scale while improving video quality of all UEs at low monetary cost by efficiently allocating limited LTE resources. Fig. 3 represents the proposed interworking architecture between 3GPP and non-3GPP access networks, which is based on the 3GPP standard architecture [48], [49]. As shown in Fig. 3, the proposed scalable offloading module is implemented at the ANDSF server for compatibility with existing architecture. In addition, ANDSF has additional interfaces to the policy and charging rule function (PCRF) and packet data network gateway (P-GW) to apply real-time network condition-aware offloading policies. In this paper, WiFi APs are assumed to be deployed by the cellular network operator. Cellular network operator can effectively offload the mobile traffic by turning on their APs when the current LTE network capacity experiences difficulty supporting UEs.

3.1 System Architecture

The proposed system architecture under consideration is given in Fig. 4. The proposed system includes three main

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TMC.2017.2748592, IEEE Transactions on Mobile Computing

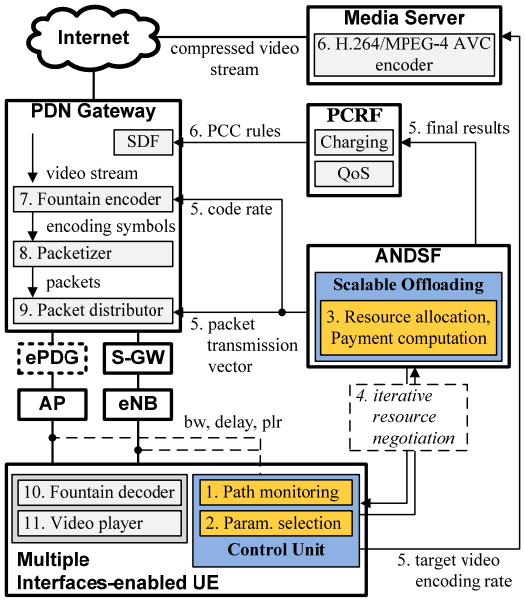

Fig. 4. Proposed system architecture under consideration.

entities: ANDSF, P-GW, and UE. In addition, it is tightly coupled with a media server to accommodate the various mobile users. Specifically, the media server performs real-time transcoding to seamlessly provide a video stream compatible with the end-user's device, and to meet the bandwidth constraints imposed by the user's resource allocation. Mobile devices are assumed to be equipped with LTE and WiFi network interfaces. In order for path diversity to be utilized efficiently, the characteristics of each physical path must be recognized. In this paper, the pathChirp algorithm [53] is employed to estimate the available bandwidth of each physical path. The propagation delay and packet loss rate (PLR) are obtained by using the time stamp and index number in each packet header, respectively. Based on the path monitoring information, the control unit of UE determines parameters such as the fountain code rate, the packet transmission vector (i.e., the number of packets transmitted through LTE and WiFi networks), and the target video encoding rate while satisfying delay and fountain decoding failure rate constraints to support seamless video streaming. After collecting the aforementioned parameters from UEs, the proposed offloading module in ANDSF computes the number of LTE packets that can be allocated to each UE as well as the corresponding payment for allocated resources, and then delivers interim results to all participating UEs. This resource negotiation process is repeated until convergence is achieved. Finally, the optimal control parameters are forwarded to each associated entity.

Now, the H.264/MPEG-4 AVC encoder compresses a video sequence at the target video encoding rate. At the P-GW, the fountain encoder generates encoding symbols at the fountain code rate, the packetizer allocates the encoding symbols to packets, and then the packet distributor assigns packets to selected physical paths according to the packet transmission vector. It is shown in [44], [54] that the packetization of encoding symbols is closely associated with the fountain decoding probability. In this paper, an encoding pattern-aware packetization algorithm [44] is adopted as a component.

Meanwhile, PCRF creates policy and charging control (PCC) rules based on the final negotiation results, which are forwarded by ANDSF. New rules are then delivered to policy and charging enforcement function (PCEF) (i.e., P-GW). P-GW enforces the PCC rules to the user service data flow (SDF). When IP packets arrive, P-GW detects SDFs to which each packet belongs, and applies the PCC rule to the packet according to its SDF.

3.2 Problem Description

In this section, we formulate the proposed scalable offloading problems for the video streaming services regarding UEs and the cellular network operator. Before presenting a detailed description, some symbolic descriptions are given in the following. $\mathbf{I}=\{1,2,...,M\}$ represents a set of UEs, where M is the number of UEs. The control unit of UE determines its own packet transmission vector for an encoding block, represented by $\vec{q}_i=\left(q_i^{LTE}, q_i^{WiFi}\right)$ for $1\le i\le M$, where q_i^{LTE} and q_i^{WiFi} denote the number of requested packets through LTE and WiFi networks, respectively. Considering the delay and complexity, a source block during the fountain encoding process is made up of a group of pictures (GOP) in a compressed video stream. The target video encoding rate, $r_{gop}\left(\vec{q}_i, c_i\right)$, for a GOP is calculated by

$$r_{gop}\left(\vec{q}_i,c_i\right)=\frac{\left(q_i^{LTE}+q_i^{WiFi}\right)\cdot S_{pkt}\cdot c_i}{\left(N_{frm}/FR\right)}, \tag{2}$$

where S_{pkt} is the size of a packet payload, N_{frm} is the number of frames in a GOP, and FR is the video frame rate. We can now formulate the optimal problems as follows.

Problem Formulation for the i_{th} UE: Determine the packet transmission vector $\vec{q}_i$ and the fountain code rate c_i for $\forall i\in\mathbf{I}$ to minimize the following function with the minimum monetary cost

$$d_{gop}\left(r_{gop}\left(\vec{q}_i, c_i\right)\right), \tag{3}$$

$$\text{s.t. } t_{blk}\le T_{blk}^{\max}, \tag{4}$$

$$\pi_{blk}\le \Pi_{blk}^{\max}, \tag{5}$$

where $d_{gop}(\cdot)$ denotes the average distortion of compressed video in a GOP, t_{blk} is the end-to-end delay for an encoding block, $T_{blk}^{\max}$ denotes the tolerable maximum end-to-end delay to avoid the receiver buffer underflow for an encoding block, π_{blk} is the fountain decoding failure rate for a source block, and $\Pi_{blk}^{\max}$ denotes the tolerable maximum fountain decoding failure rate for a source block. The rate-distortion modeling techniques are essential to developing a real-time video encoding algorithm with low computational complexity over a time-varying wireless network. In this paper, we employ an empirical data-based GOP-level rate-distortion model [55] to esti-

mate the corresponding average distortion (d_{gop}) of compressed video in a GOP without repeated video compression processes, that is,

$$\tilde{d}_{gop}\left(r_{gop}\left(\vec{q}_i, c_i\right)\right) = \alpha \cdot r_{gop}\left(\vec{q}_i, c_i\right)^{\beta}, \quad (6)$$

where α $(\alpha > 0)$ and β $(\beta < 0)$ are the model parameters depending on the video sequence.

The cellular network operator then allocates LTE resources to UEs to maintain congestion at a tolerable level. The LTE packet allocation vector is represented by $\vec{a}_{LTE} = \left(a_1^{LTE}, a_2^{LTE} ..., a_M^{LTE}\right)$, where a_i^{LTE} denotes the number of LTE packets allocated to the i_{th} UE. Now, it can be formulated as follows.

Problem Formulation for Cellular Network Operator: Determine the LTE packet allocation vector $\vec{a}_{LTE}$ to maximize

$$\sum_{i \in \mathbf{I}} a_i^{LTE}, \quad (7)$$

$$\text{s.t.} \quad \sum_{i \in \mathbf{I}} a_i^{LTE} \le R, \quad (8)$$

$$0 \le a_i^{LTE} \le q_i^{LTE}, \text{ for } \forall i \in \mathbf{I}, \quad (9)$$

where R denotes the available LTE network resource for video streaming services, which is shared among UEs. The cellular network operator can determine R by establishing the specific policies of the LTE radio resource management (RRM) module, such as dynamic resource allocation (DRA) and packet scheduling for video streaming flows [56].

The aforementioned UE problem can be simplified to obtain an effective solution with low computational complexity using the following properties.

Property 1. To satisfy (4), the maximum number of packets through each physical path is calculated based on its available bandwidth and observed delay. The transmission delay of a packet through LTE and WiFi networks is calculated by $t_{LTE,i}^{trans} = S_{pkt} / bw_{LTE,i}$ and $t_{WiFi,i}^{trans} = S_{pkt} / bw_{WiFi,i}$, where $bw_{LTE,i}$ and $bw_{WiFi,i}$ are the available bandwidths of the LTE and WiFi networks, respectively. The maximum number of packets that can be transmitted through LTE and WiFi networks within $T_{blk}^{\max}$ for an encoding block is then obtained by $Q_i^{LTE} = \left\lfloor \frac{T_{blk}^{\max} - t_{LTE,i}^{prop}}{t_{LTE,i}^{trans}} \right\rfloor$ and $Q_i^{WiFi} = \left\lfloor \frac{T_{blk}^{\max} - t_{WiFi,i}^{prop}}{t_{WiFi,i}^{trans}} \right\rfloor$, where $t_{LTE,i}^{prop}$ and $t_{WiFi,i}^{prop}$ denote the propagation delay of the LTE and WiFi networks, respectively, and $\lfloor x \rfloor$ refers to the largest integer less than x. Now, the feasible ranges of the number of packets transmittable through LTE and WiFi networks are determined by $0 \le a_i^{LTE} \le Q_i^{LTE}$ and $0 \le a_i^{WiFi} \le Q_i^{WiFi}$.

Property 2. The optimal fountain code rate (c_i^{opt}) that satisfies the fountain decoding failure rate constraint can be represented as a function of $\vec{q}_i$ under the assumption that each channel condition is known during the subsequent block transmission [47]. That is,

$$c_i^{opt}\left(\vec{q}_i\right) = \arg\min_{c_i}\left(g\left(\vec{q}_i, c_i\right)\right) \text{ for } 0 < c_i \le 1, \quad (10)$$

$$g\left(\vec{q}_i, c_i\right) = \begin{cases} \Pi_{blk}^{\max} - \pi_{blk}\left(\vec{q}_i, c_i\right) & \text{if } \pi_{blk}(\cdot) \le \Pi_{blk}^{\max} \\ \infty & \text{otherwise,} \end{cases} \quad (11)$$

where $\pi_{blk}\left(\vec{q}_i, c_i\right)$ denotes the fountain decoding failure rate calculated based on the packet loss rates of each physical path [47]. Based on the aforementioned properties, the previous problem formulation for a UE is simplified as follows under the assumption that free WiFi resources are fully utilized to reduce monetary cost.

Simplified Problem Formulation for the i_{th} UE: Determine q_i^{LTE} for $\forall i \in \mathbf{I}$ with the given q_i^{WiFi} to minimize the following function with the minimum monetary cost

$$\tilde{d}_{gop}\left(r_{gop}\left(q_i^{LTE}\right)\right)\Big|_{q_i^{WiFi} = Q_i^{WiFi}}, \quad (12)$$

$$\text{s.t.} \quad q_i^{LTE} \le Q_i^{LTE}. \quad (13)$$

3.3 Proposed PSP Auction-based Resource Negotiation

In a large-scale decentralized and competitive environment, it is difficult for UEs and cellular network operator to solve the aforementioned optimization problems. In this paper, we employ the PSP auction mechanism [21] to obtain an effective solution with relatively low computational complexity. Since the PSP auction is designed to assign variable-sized portions of a divisible resource among multiple bidders, it is more suitable to the proposed system to allocate wireless resources to UEs. We first present the PSP auction-compatible system model and then verify the feasibility of resource valuation function, designed for video streaming services. Finally, the proposed bid decision algorithm and its convergence analysis are described in detail.

3.3.1 PSP Auction-compatible System Model

In this paper, we envision an offloading market where a set of UEs (the buyers) competes to obtain the LTE resources of the cellular network operator (the seller) for the video streaming service. It is assumed that there is only one seller, and this seller performs the role of an auctioneer who conducts the auction and clears the market. As soon as the current traffic demand exceeds the available LTE resource R, the cellular network operator announces the auction for sharing the LTE resource and collects bids from the UEs. After collecting bids from the UEs, the cellular network operator computes the optimal resource allocation according to the PSP rules, defined by $s_i(\boldsymbol{b}) = \left(a_i^{LTE}(\boldsymbol{b}), \tau_i(\boldsymbol{b})\right)$ for $\forall i \in \mathbf{I}$ and delivers the feedback to corresponding UE. The UE waits for an auction announcement, and then decides whether or not to bid. Although the LTE network is congested, some UEs may not want to participate in the auction to obtain LTE resources, because they perceive a good WiFi channel. In this case, the video streaming services are provided through a WiFi single network only. If the UE decides to

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TMC.2017.2748592, IEEE Transactions on Mobile Computing

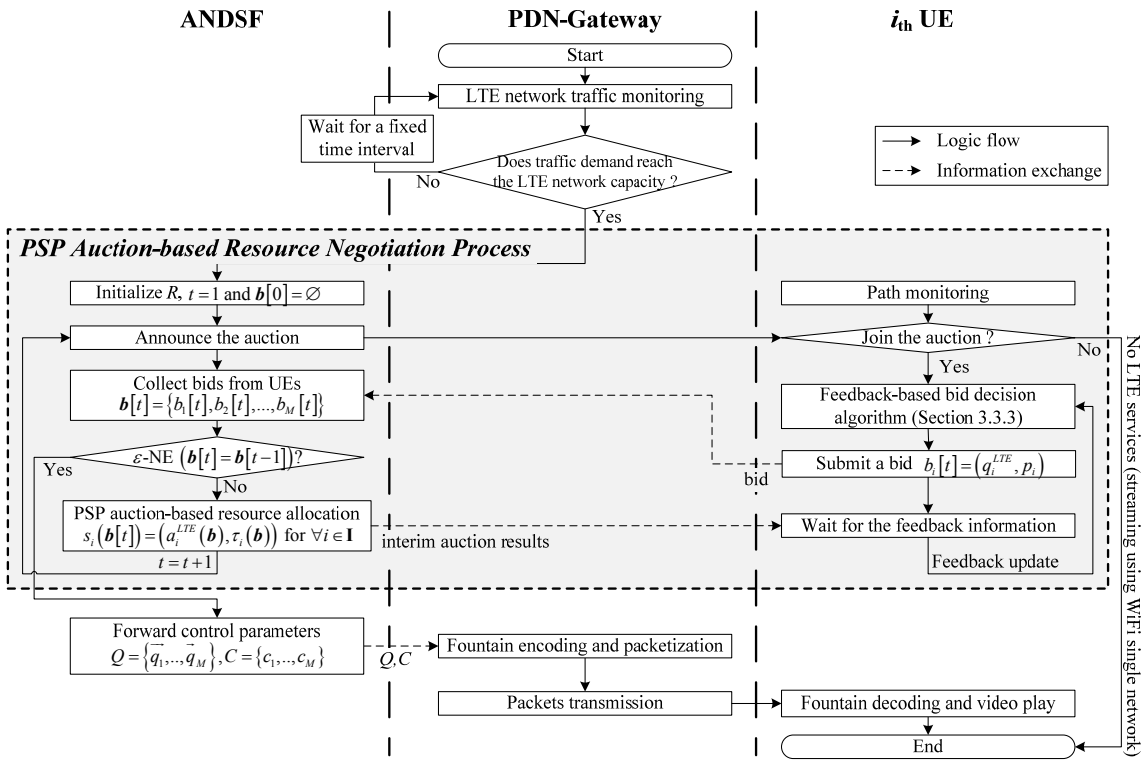

Fig. 5. Proposed PSP auction-based resource negotiation process.

join the auction, it submits the bid $b_i=(q_i^{LTE},p_i)$ and waits for the interim auction results (the details of the bid decision algorithm are described in section 3.3.3). When the interim auction results have been revealed, the UEs update their bid based on the feedback information to achieve better utility. This resource negotiation process is repeated until all UE bids are converged to the ε-Nash equilibrium to satisfy the objectives of both UEs and the cellular network operator simultaneously. The overall system procedure is presented in Fig. 5.

Now, the simplified UE problem formulation in section 3.2 is changed to be compatible with the PSP auction mechanism as follows.

Modified Problem Formulation for the i_{th} UE Compatible with PSP Auction: Determine a bid $b_i=(q_i^{LTE},p_i)$ for $\forall i\in\mathbf{I}$ with the given q_i^{WiFi} to maximize the following utility function

$$u_i(b_i,\boldsymbol{b}_{-i})\big|_{q_i^{WiFi}=Q_i^{WiFi}}=\theta_i\left(a_i^{LTE}(b_i,\boldsymbol{b}_{-i})\right)\big|_{q_i^{WiFi}=Q_i^{WiFi}}-\tau_i(b_i,\boldsymbol{b}_{-i}), \quad (14)$$

$$\text{s.t. } q_i^{LTE}\le Q_i^{LTE}, \quad (15)$$

where $\theta_i(\cdot)$ denotes the resource valuation of the i_{th} UE, which is defined by

$$\theta_i\left(q_i^{LTE}\right)\big|_{q_i^{WiFi}=Q_i^{WiFi}}=PSNR\left(q_i^{LTE}\right)\big|_{q_i^{WiFi}=Q_i^{WiFi}}-PSNR(0)\big|_{q_i^{WiFi}=Q_i^{WiFi}}, \quad (16)$$

where $PSNR\left(q_i^{LTE}\right)\big|_{q_i^{WiFi}=Q_i^{WiFi}}$ represents the peak-signal-to-noise ratio (PSNR), which is defined by $10\cdot\log\left(255^2/\tilde{d}_{gop}\left(r_{gop}\left(q_i^{LTE}\right)\right)\big|_{q_i^{WiFi}=Q_i^{WiFi}}\right)$, and $\tau_i(b_i,\boldsymbol{b}_{-i})$ is a payment for allocated resources to the i_{th} UE (for details, refer to section 2.2).

3.3.2 Verification of Utility and Resource Valuation Functions

In this section, we verify the feasibility of our utility and resource valuation functions in (14) by confirming the requirements of the PSP auction. Fig. 6 depicts the i_{th} UE's utility and payment according to the amount of allocated resource (a_i^{LTE}) under the assumption that the opponent profiles are given. The diagonal region (red) and rectangles (yellow) represent the utility and payment values of the i_{th} UE, respectively. When the utility function in (14) is concave, the PSP auction-based resource allocation mechanism guarantees that the social welfare, $\sum_{i\in\mathbf{I}}\theta_i\left(a_i^{LTE}\right)$, is maximized at the ε-Nash equilibrium [21] while satisfying the given constraints in (1). As shown in

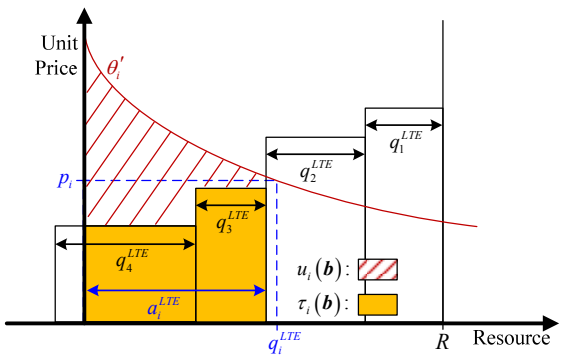

Fig. 6. Example of the i_{th} UE's utility and payment with given opponent profiles.

Fig. 6, the second term of (14), namely, $\tau_i(\boldsymbol{b})$, is a non-decreasing convex function with respect to a_i^{LTE} because the first derivation of $\tau_i(\boldsymbol{b})$ is represented as a staircase function. Thus, the resource valuation function must be monotonically non-decreasing and concave with respect to increased resources in order to guarantee that the ε-Nash equilibrium exists.

Proposition 1. *The resource valuation $\theta_i(\cdot)$ in (16) is differentiable, $\theta_i(0)=0$, monotonically non-decreasing with increased resources, and concave.*

Proof. By combining (2) and (16), $\theta_i\left(q_i^{LTE}\right)\Big|_{q_i^{WiFi}=Q_i^{WiFi}}$ can be represented as follows.

$$\theta_i\left(q_i^{LTE}\right)\Big|_{q_i^{WiFi}=Q_i^{WiFi}} = PSNR\left(q_i^{LTE}\right)\Big|_{q_i^{WiFi}=Q_i^{WiFi}} - PSNR(0)\Big|_{q_i^{WiFi}=Q_i^{WiFi}}$$
$$= 10\cdot\beta\cdot \left[\log\frac{Q_i^{WiFi}}{q_i^{LTE}+Q_i^{WiFi}} - \log c_i^{opt}\left(q_i^{LTE}\right)\Big|_{q_i^{WiFi}=Q_i^{WiFi}} + \log c_i^{opt}(0)\Big|_{q_i^{WiFi}=Q_i^{WiFi}}\right].$$

Since it is difficult to represent $c_i^{opt}(\cdot)$ as a closed form as shown in (10) and (11), we empirically derive the estimation model ($\tilde{c}_i^{opt}$) by using a curve fitting method. That is,

$$\tilde{c}_i^{opt}\left(q_i^{LTE}\right)\Big|_{q_i^{WiFi}=Q_i^{WiFi}} = \gamma\cdot\left(q_i^{LTE}+Q_i^{WiFi}\right)^{\kappa}, \quad (17)$$

where γ and κ are the model parameters. With values of Q_i^{WiFi} varying from 50 to 200, and average PLR from 0 to 0.05, we verify the accuracy of the proposed code rate model as shown in Fig. 7. The resulting R^2 values are larger than 0.98 in all cases, as shown in Table 1. Consequently, the proposed model fits the observed data well. Furthermore, the table clearly indicates that both γ and κ are always positive.

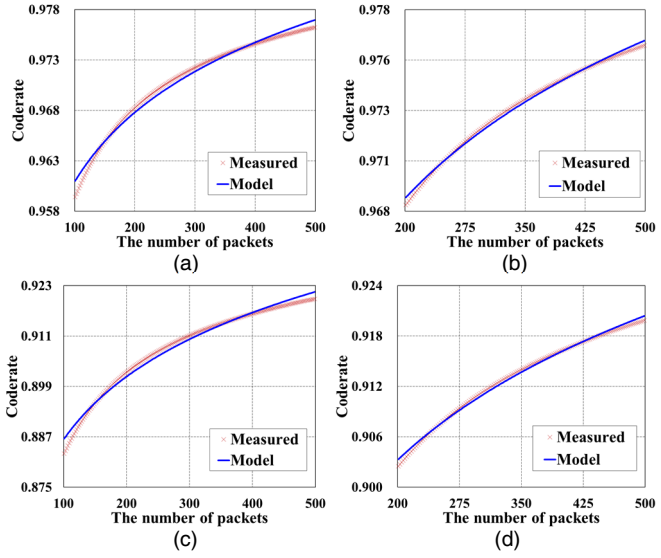

Fig. 7. Examples of the proposed code rate model: (a) code rate with Q_i^{WiFi} = 100 and PLR = 1%, (b) code rate with Q_i^{WiFi} = 200 and PLR = 1%, (c) code rate with Q_i^{WiFi} = 100 and PLR = 5%, and (d) code rate with Q_i^{WiFi} = 200 and PLR = 5%.

TABLE 1
CODE RATE MODEL PARAMETERS AND R-SQUARED VALUE

Model Index	γ	κ	R^2
Fig. 7 (a)	0.9163	0.0103	0.9885
Fig. 7 (b)	0.9245	0.0088	0.9964
Fig. 7 (c)	0.7929	0.0242	0.9878
Fig. 7 (d)	0.8099	0.0206	0.9962

Now, $\theta_i\left(q_i^{LTE}\right)\Big|_{q_i^{WiFi}=Q_i^{WiFi}}$ is approximately represented by

$$\theta_i\left(q_i^{LTE}\right)\Big|_{q_i^{WiFi}=Q_i^{WiFi}} \approx 10\cdot\beta\cdot\left[\log\frac{Q_i^{WiFi}}{q_i^{LTE}+Q_i^{WiFi}} - \log\left(\gamma\cdot\left(q_i^{LTE}+Q_i^{WiFi}\right)^{\kappa}\right) + \log\left(\gamma\cdot\left(Q_i^{WiFi}\right)^{\kappa}\right)\right] = 10\cdot\beta\cdot(1+\kappa)\cdot\left(\log\frac{Q_i^{WiFi}}{q_i^{LTE}+Q_i^{WiFi}}\right). \quad (18)$$

The first derivation of $\theta_i\left(q_i^{LTE}\right)\Big|_{q_i^{WiFi}=Q_i^{WiFi}}$ is then obtained by

$$\theta_i'\left(q_i^{LTE}\right)\Big|_{q_i^{WiFi}=Q_i^{WiFi}} = -\frac{10\cdot\beta\cdot(1+\kappa)}{\ln 10}\cdot\frac{1}{\left(q_i^{LTE}+Q_i^{WiFi}\right)}. \quad (19)$$

Because β is negative and κ is positive, $\theta_i'\left(q_i^{LTE}\right)>0$ in $q_i^{LTE}\geq 0$. The second derivation of $\theta_i\left(q_i^{LTE}\right)\Big|_{q_i^{WiFi}=Q_i^{WiFi}}$ is next obtained by

$$\theta_i''\left(q_i^{LTE}\right)\Big|_{q_i^{WiFi}=Q_i^{WiFi}} = \frac{10\cdot\beta\cdot(1+\kappa)}{\ln 10}\cdot\frac{1}{\left(q_i^{LTE}+Q_i^{WiFi}\right)^2}. \quad (20)$$

For the same reason, $\theta_i''\left(q_i^{LTE}\right)<0$ in $q_i^{LTE}\geq 0$. Thus, the resource valuation $\theta_i(\cdot)$ is differentiable, monotonically non-decreasing, and concave with respect to allocated resources. In addition, $\theta_i(0)=0$. □

3.3.3 Feedback-based Bid Decision Algorithm

During the PSP auction, the game converges to an ε-Nash equilibrium using the truthful ε-best reply strategy [21] under the assumption that a player determines a bid based on complete knowledge of its opponent profiles ($\boldsymbol{b}_{-i}$) in the previous round and that bids are always submitted asynchronously, with only one player submitting a new bid in each round. However, when a time-varying wireless network environment is considered, the aforementioned assumption is not reasonable [57]. Each mobile device has limited computational capabilities, so it may be infeasible for it to process all of its opponent profiles in order to determine the bid that will maximize its payoffs within a tolerable delay. In addition, considerable communication overhead is incurred as players repeatedly exchange bid information every auction period. Thus, it is important to design a practical bid decision algorithm that can be deployed successfully in a realistic environment. In this section, we design the feedback-based bid decision algorithm to improve the convergence speed to an equilibrium point while maximizing social welfare without complete knowledge of opponent information. It

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TMC.2017.2748592, IEEE Transactions on Mobile Computing

is assumed that the cellular network operator sends UEs the feedback information on the previously submitted bid. To determine the i_{th} UE's bid in round t, two forms of feedback information are required: 1) $q_i^{LTE}[t-1]$, which is the amount of requested resource in round $t-1$; and 2) $a_i^{LTE}(\boldsymbol{b}[t-1])$, which is the amount of allocated resource in round $t-1$. It is assumed that all participating players are rational and selfish to maximize their own utility according to the following.

Property 3. If the requested resource from the previous round is fully allocated (i.e., $a_i^{LTE}(\boldsymbol{b}) = q_i^{LTE}$), UE will not decrease the amount of requested resource in the next round.

Property 4. If the requested resource from the previous round is not fully allocated, (i.e., $a_i^{LTE}(\boldsymbol{b}) < q_i^{LTE}$), UE will not increase the amount of requested resource in the next round.

Property 5. The UE submits a truthful bid reflecting individual's true valuation (i.e., $p_i = \theta_i'(q_i^{LTE})$ for $\forall i \in \mathbf{I}$), which is a dominant strategy in PSP [21].

Properties 3 and 4 can be easily verified by contradiction. The proposed algorithm is designed based on the above properties, and the details are described as follows.

Step 1. Initialize system parameters such as $T_{blk}^{\max}$, $\Pi_{blk}^{\max}$, and ε, and $\boldsymbol{b}[0] = \varnothing$.

Step 2. In round $t=1$, each UE initializes the bid $b_i[1] = \left(q_i^{LTE}[1], \theta_i'(q_i^{LTE}[1])\right)$, where $q_i^{LTE}[1]$ is set to Q_i^{LTE}. According to Property 5, p_i is set to $\theta_i'(q_i^{LTE})$. Go to Step 4.

Step 3. In round $t \in \{2,3,...\}$, each UE updates its bid based on the feedback information. If the requested resource is fully allocated in round $t-1$ (i.e., $a_i^{LTE}(\boldsymbol{b}[t-1]) = q_i^{LTE}[t-1]$), then go to Step 3-1. Otherwise, go to Step 3-2.

Step 3-1. According to Property 3, the UE does not decrease the amount of requested resource. The bid $b_i[t] = \left(q_i^{LTE}[t], \theta_i'(q_i^{LTE}[t])\right)$ in round t is determined by $q_i^{LTE}[t] = q_i^{LTE}[t-1]$. This means that the UE continuously uses the previous round bid without incurring a bid submission fee ε [21]. Go to Step 4.

Step 3-2. According to Property 4, the UE does not increase the amount of requested resource. The bid $b_i[t] = \left(q_i^{LTE}[t], \theta_i'(q_i^{LTE}[t])\right)$ in round t is determined by

$$q_i^{LTE}[t] = \begin{cases} q_i^{new} & \text{if } \theta_i(q_i^{new}) - \theta_i(a_i^{LTE}(\boldsymbol{b}[t-1])) > \varepsilon \\ q_i^{LTE}[t-1] & \text{otherwise,} \end{cases} \quad (21)$$

where $q_i^{new} = a_i^{LTE}(\boldsymbol{b}[t-1]) + \omega_i \cdot (q_i^{LTE}[t-1] - a_i^{LTE}(\boldsymbol{b}[t-1]))$ and ω_i $(0 \le \omega_i \le 1)$ is the weighting factor. The UE will send a bid as long as it improves the current utility by ε.

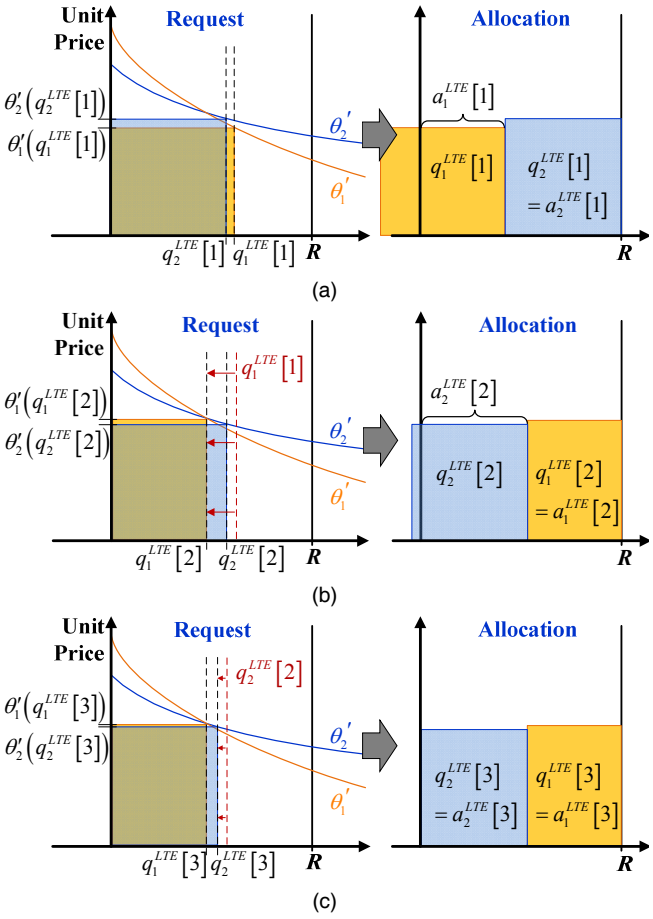

Fig. 8. Example of the proposed feedback-based bid decision algorithm with two UEs: (a) round 1 when $a_1^{LTE}(\boldsymbol{b}[1]) < q_1^{LTE}[1]$ and $a_2^{LTE}(\boldsymbol{b}[1]) = q_2^{LTE}[1]$, (b) round 2 when $a_1^{LTE}(\boldsymbol{b}[2]) = q_1^{LTE}[2]$ and $a_2^{LTE}(\boldsymbol{b}[2]) < q_2^{LTE}[2]$, and (c) round 3 when $a_1^{LTE}(\boldsymbol{b}[3]) = q_1^{LTE}[3]$ and $a_2^{LTE}(\boldsymbol{b}[3]) = q_2^{LTE}[3]$.

Step 4. If all of UE's current bids are the same as the previous bid profiles (i.e., $\boldsymbol{b}[t] = \boldsymbol{b}[t-1]$), then terminate the process under the assumption that the ε-Nash equilibrium is achieved. Otherwise, the ANDSF calculates $a_i^{LTE}(\boldsymbol{b}[t])$ which is delivered to UE. With the updated feedback information, go to Step 3 to proceed to the next round auction.

An example of the proposed bid decision algorithm with two UEs is depicted in Fig. 8. In round 1, each UE requests the LTE packets, initialized by $q_i^{LTE}[1] = Q_i^{LTE}$. It is observed in Fig. 8 (a) that UE 1 does not receive as many LTE packets as requested. In this case, UE 1 decreases the number of requested LTE packets in round 2 according to Property 4. On the other hand, UE 2 maintains the same bid in round 2 according to Property 3, because the amount of requested resource in round 1 is fully allocated. In round 2, UE 1 successfully obtains the requested LTE resource since UE 1's marginal valuation becomes larger

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TMC.2017.2748592, IEEE Transactions on Mobile Computing

than that of UE 2, as shown in Fig. 8 (b). Thus, UE 2 attempts to reduce the number of requested LTE packets in round 3 to improve its utility, as shown in Fig. 8 (c). Finally, the amount of LTE resource, requested by all UEs, is comprehensively assigned in round 3. The UEs no longer change their bids, since convergence is achieved, and thus the algorithm is terminated.

3.3.4 Convergence Analysis of the Proposed Feedback-based Bid Decision Algorithm

In this section, we theoretically analyze the convergence of the proposed algorithm, i.e., all participating UEs reach a mutual agreement. If the auction game ends at the disagreement point, some UEs receive an unfair resource allocation which may cause a serious QoE degradation. Thus, the convergence is one of the most important factors in the auction game. We now show that the bid profiles $\boldsymbol{b}$ converge in a finite number of negotiation rounds.

Proposition 2. *In the proposed feedback-based bid decision algorithm, the bid profiles* $\boldsymbol{b}$ *converge in* $O\left(\left(\sum_{i\in \mathrm{I}} Q_i^{LTE} - R\right) \Big/ \theta_i^{-1}(\varepsilon)\right)$ *rounds.*

Proof. In the auction game, the amount of requested resource can be classified into two cases: 1) $\sum_{\forall i\in \mathrm{I}} q_i^{LTE}[t-1] \le R$ and 2) $\sum_{\forall i\in \mathrm{I}} q_i^{LTE}[t-1] > R$. In the proposed algorithm, the UEs in each case determine their bid in round t as follows.

1) $\sum_{\forall i\in \mathrm{I}} q_i^{LTE}[t-1] \le R$: the number of packets requested from all UEs is fully allocated. In this case, $q_i^{LTE}[t] = q_i^{LTE}[t-1]$ for $\forall i \in \mathrm{I}$, and the algorithm is terminated.

2) $\sum_{\forall i\in \mathrm{I}} q_i^{LTE}[t-1] > R$: there exists at least one UE where the number of requested packets is partially allocated. In this case, $q_i^{LTE}[t]$ for $\forall i \in \mathrm{I}$ is determined as follows.

If $a_i^{LTE}(\boldsymbol{b}[t-1]) = q_i^{LTE}[t-1]$, then $q_i^{LTE}[t] = q_i^{LTE}[t-1]$.

If $a_i^{LTE}(\boldsymbol{b}[t-1]) < q_i^{LTE}[t-1]$, then $q_i^{LTE}[t] \le q_i^{LTE}[t-1]$.

Note that if $a_i^{LTE}(\boldsymbol{b}[t-1]) < q_i^{LTE}[t-1]$, the UE reduces $q_i^{LTE}[t]$, provided that it improves the current utility by ε. In every round, there must exist at least one UE $\forall i \in \mathrm{I}$ such that $q_i^{LTE}[t] \le q_i^{LTE}[t-1]$. Thus, the number of packets requested from all UEs in round t is always smaller than that in round $t-1$. Consequently, the sequence $\left\{\sum_{i\in \mathrm{I}} q_i^{LTE}[t]:\ t \ge 1\right\}$ gradually decreases and thus converges in $O\left(\left(\sum_{i\in \mathrm{I}} Q_i^{LTE} - R\right) \Big/ \theta_i^{-1}(\varepsilon)\right)$ rounds. □

4 Simulation Results

The proposed system is implemented using NS-3, which is a well-known discrete-event network simulator. The simulation environment is set up as follows, and details of the system parameters are summarized in Table 2.

(1) **Network:** We create a network topology consisting of 25 UEs, two eNBs, and four APs (802.11g/n/ac standards). The initial position and distance between the entities is shown in Fig. 9. The UEs are randomly located within 50 m of the AP, a greater number of UEs are distributed around AP 4. The UE mobility is arbitrarily set between 1 and 5 m/s. The Friis propagation loss model [58] and a trace-driven fading loss model specified by 3GPP [59] are employed for the LTE network. For the WiFi channel, yet another network simulator (YANS) model [60] is employed, which is most commonly used in NS-3. During the simulation, R is fixed to 12 Mbps for the eNB 1 and 20 Mbps for the eNB 2. The simulation is performed over 300 s.

(2) **Handover:** For X2 handover between eNBs, the reference signal received quality (RSRQ)-based handover algorithm is employed, in which the handover decision is primarily based on event A2 measurements (serving cell's RSRQ becomes worse than threshold) and event A4 measurements (neighbor cell's RSRQ becomes better than threshold) [61]. In contrast, the current IEEE 802.11 specification does not explicitly support the handover. Thus, we have implemented a signal-to-noise ratio (SNR)-based handover algorithm. When the signal strength of the AP weakens, a user scans available channels and then connects to the AP with the best SNR under the assumption that adjacent APs have identical identifiers.

(3) **Video:** We use two full high-definition (HD)-sized

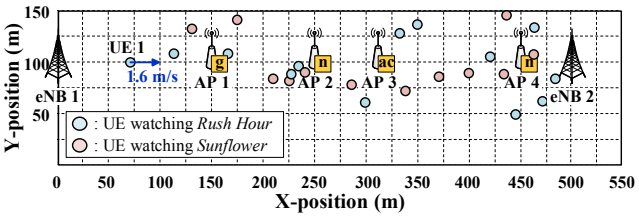

Fig. 9. Wireless network environment for performance evaluation.

TABLE 2
SYSTEM PARAMETERS

Parameter	Description	Value
M	Number of UEs	25
$T_{blk}^{\max}$	Tolerable maximum end-to-end delay	500 ms
ω_i	Weighting factor for bid update	0.3 ~ 0.9
S_{pkt}	Packet payload size	1024 bytes

TABLE 3
RATE-DISTORTION MODEL PARAMETERS

GOP Index	*Rush Hour*		*Sunflower*	
	α	β	α	β
1	3087.3	-1.049	1970.4	-1.011
2	8102.6	-1.233	1747.3	-0.997
3	4917.0	-1.168	3094.5	-1.098
4	4800.9	-1.158	2160.3	-1.028
5	4588.2	-1.154	2753.5	-1.055

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TMC.2017.2748592, IEEE Transactions on Mobile Computing

(1920 x 1080) video sequences, *Rush Hour* and *Sunflower* [62]. The video stream is encoded by H.264/MPEG-4 AVC codec with 25 frames per second, and a GOP consists of 25 frames (IPPP…PPP). The rate-distortion model parameters for each GOP are provided in Table 3 (parameters after the 5th GOP are omitted because of space limitations).

(4) **Fountain code:** LT code [20] is employed during the simulation. The symbol size is set to 32 bytes. Thus, one packet includes 32 encoding symbols. During the simulation, $\Pi_{blk}^{\max}$ is fixed to 0.01 and δ is empirically set to the average value for 1,000 receptions to maintain the successful fountain decoding probability at 0.99 [44]. Thus, in most cases, fountain decoding is successful, but sometimes fail when the wireless channel state is very unstable. In this case, if the receiver buffer contains enough video data (e.g., several GOPs) for playout, the receiver requests more encoding symbols and attempts to decode again. Otherwise, the receiver attempts decoding with an insufficient number of encoding symbols. Because we adopt message passing [19] as a decoding algorithm, the source block can be partially recovered even if sufficient encoding symbols are not available.

4.1 Performance Verification of the Proposed PSP Auction-based Resource Negotiation

In this section, we demonstrate the performance of the proposed PSP auction-based resource negotiation process. During the simulation, ε is set to 0.001. The simulation results are presented in Fig. 10. Ten UEs participate in the auction of eNB 1 (UEs except for 1, 2, 6, and 9 are omitted in the figures because of space constraints). Since each UE experiences different WiFi channel conditions, the number of transmitted WiFi packets is variously determined, as shown in Fig. 10 (a). With the given number of WiFi packets, the number of requested LTE packets and corresponding offered price are initialized (section 3.3.3 Step 2). As the round proceeds, each UE dynamically updates the bid to maximize his or her utility based on the feedback information from the previous round bid (section 3.3.3 Step 3). Fig. 10 (b), (c), and (d) present this negotiation process in detail. In round 13, all UE bidding prices are converged to the ε-Nash equilibrium that maximizes social welfare, as shown in Fig. 10 (c) and (e). Accordingly, the payment also reaches the stable state, as shown in Fig. 10 (f).

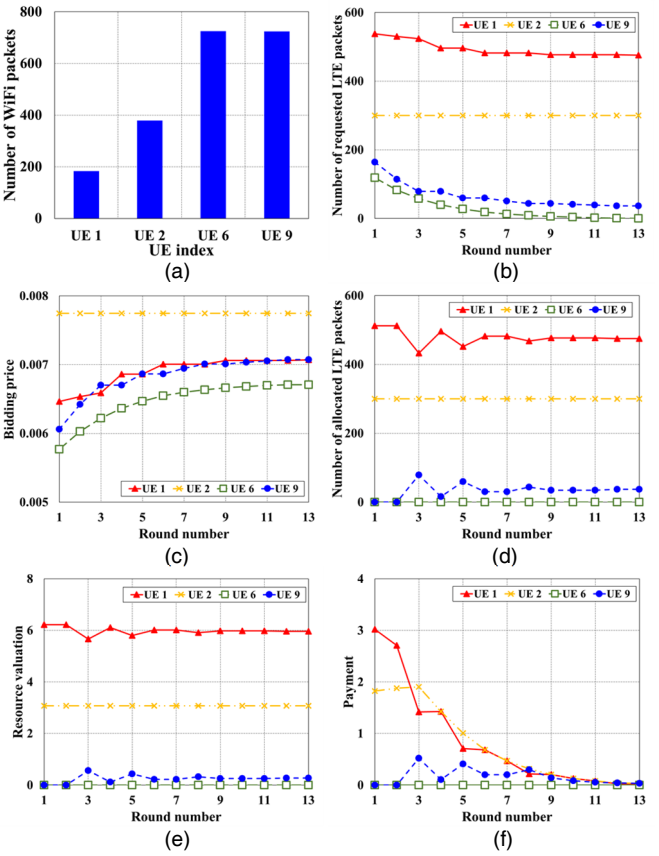

Fig. 10. Example of the PSP auction-based resource negotiation process with the proposed algorithm (ε = 0.001): (a) the number of WiFi packets, (b) the number of requested LTE packets, (c) offered price, (d) the number of allocated LTE packets, (e) resource valuation, and (f) payment.

To verify the performance of the proposed feedback-based bid decision algorithm, we implement three additional bid decision algorithms: bisection method, greedy algorithm [57], and ε-best reply [21]. The bisection method is a well-known technique for solving the optimization problem because of its simplicity and robustness. In a greedy algorithm for a PSP auction, a time-varying bid-quantity step size is selected such that each step results in an approximately equivalent PSNR drop or increase. During the simulation, the constant PSNR step size is set to 0.1 dB. In the ε-best reply (ε = 0.001), the UE submits the optimal bid with complete information of the opponent

TABLE 4
AVERAGE NUMBER OF ROUNDS, SOCIAL WELFARE, AND PAYMENT COMPARISON WITH EXISTING BID DECISION ALGORITHMS

	Bisection	Proposed feedback-based bid decision				Greedy	ε-best reply (ε = 0.001)
		ε = 1	ε = 0.1	ε = 0.01	ε = 0.001		
Avg. number of rounds	9.43	4.62	9.65	13.99	14.99	30.25	331
Max. number of rounds	10	8	14	18	19	46	524
Min. number of rounds	9	2	6	10	11	8	186
Avg. social welfare	22.25	22.72	22.76	22.77	22.77	22.75	22.78
Avg. payment	11.82	12.01	4.19	0.12	0.06	0	0
Required feedback information	q_i^{LTE}, a_i^{LTE}						opponent bid profiles ($\boldsymbol{b}_{-i}$)

profiles by a round-robin order. Table 4 shows the results of performance comparisons for 300 repeated auctions. The bisection method achieves the fast convergence. However, it is not efficient because the equilibrium is reached at the point at which low social welfare and high payment are generated. In the greedy, ε-best reply, and proposed algorithms, each UE produces a better bid through iterative interaction, thereby improving their utility. As a result, on average, the equilibrium point is obtained in round 30.25, producing 22.75 social welfare with the greedy algorithm; in round 331, producing 22.78 social welfare with the ε-best reply; and in round 14.99, producing 22.77, social welfare with the proposed algorithm. Although the optimal solution can be obtained by the ε-best reply, it takes a long time to converge. Thus, the ε-best reply algorithms are difficult to apply to the time-varying wireless network environment. In the proposed algorithm, an effective tradeoff between the number of rounds and social welfare can be attained by finely adjusting ε. When ε is set to 0.001, the proposed algorithm achieves near-optimal social welfare with relatively fast convergence speed compared to existing bid decision algorithms.

4.2 Performance Verification of the Proposed System with Dynamic Network Environment

In a dynamic and realistic wireless network environment, we verify the performance of the proposed PSP-based scalable offloading system. As shown in Fig. 11, the UE 1 moves at a speed of 1.6 m/s and is initially connected to the eNB 1. At about 25 s, the UE 1 detects an available WiFi AP and executes a scalable offloading by using LTE and WiFi networks, simultaneously. The performance in terms of PLR, code rate, bitrates, and PSNR observed by the UE 1 are provided in Fig. 12. Except for the cell boundary area of eNB (between 100 and 140 s), PLR over LTE networks is relatively stable compared to WiFi networks, as shown in Fig. 12 (a). For WiFi networks, PLR increases when a handover occurs between APs or a large number of UEs are concentrated, such as is the case for AP 4. Based on these wireless network states, the fountain code rate is adaptively adjusted to mitigate wireless channel errors, as shown in Fig. 12 (b). Fig. 12 (c) and (d) present the estimated and measured bitrates over WiFi and LTE networks, respectively. It is observed in Fig. 12 (c) that 802.11n and 802.11ac provide a generally higher bitrate as compared to 802.11g. However, even with 802.11n, the bitrate may vary depending on the density of the users in the service area of the AP. In the proposed system, the UE requests the LTE resource based on the

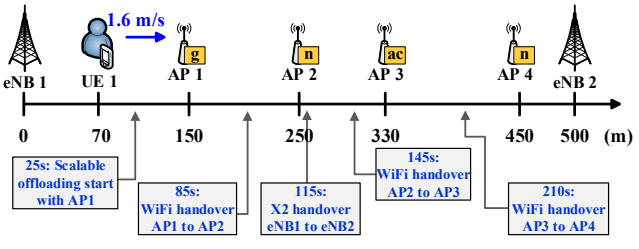

Fig. 11. WiFi and LTE network connection status according to the user mobility.

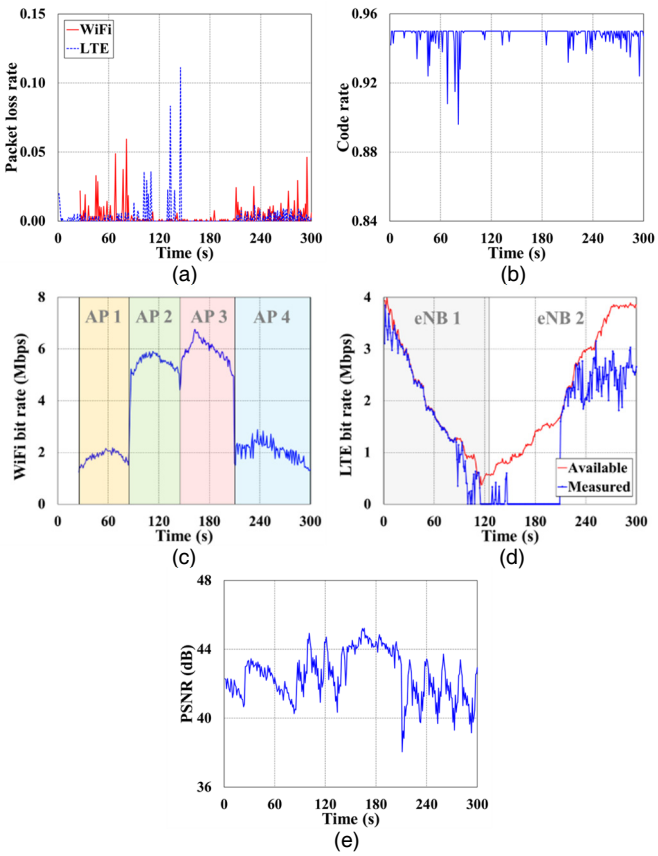

Fig. 12. Performance verification of UE 1: (a) PLR, (b) code rate, (c) WiFi bitrate, (d) LTE bitrate, and (e) PSNR.

estimated available bandwidth, as shown by the red line in Fig. 12 (d). In fact, the accuracy of the bandwidth estimation algorithm affects the performance of the proposed system. If the estimated available bandwidth is greater than the actual bandwidth, the PLR is significantly increased. Although fountain code is adopted to improve error robustness, the video quality may deteriorate when the UE cannot receive a sufficient number of encoding symbols to successfully reconstruct the source block. As shown by the blue line in Fig. 12 (d), the UE may not always obtain the requested resources because the LTE resource is allocated by considering user's WiFi network state. According to the PSP auction rules, UEs with poor WiFi network conditions are generally allocated more LTE resources to avoid video quality degradation, and vice versa. Between 145 s and 210 s, the UE 1 has sufficient WiFi resources available because it is connected to AP 3 with 802.11ac. Therefore, LTE resources are rarely allocated during that time, as shown in Fig. 12 (d). The corresponding video quality is shown in Fig. 12 (e). By efficiently utilizing the resources of WiFi and LTE networks, the quality of video streaming is almost guaranteed in the 40–45 dB PSNR range.

4.3 Performance Comparison with Existing System

In this section, we compare the performance of the proposed PSP-based scalable offloading with those of the two existing systems, namely, the traditional PSP [21] and on-the-spot offloading [32]. In the traditional PSP, UEs

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TMC.2017.2748592, IEEE Transactions on Mobile Computing

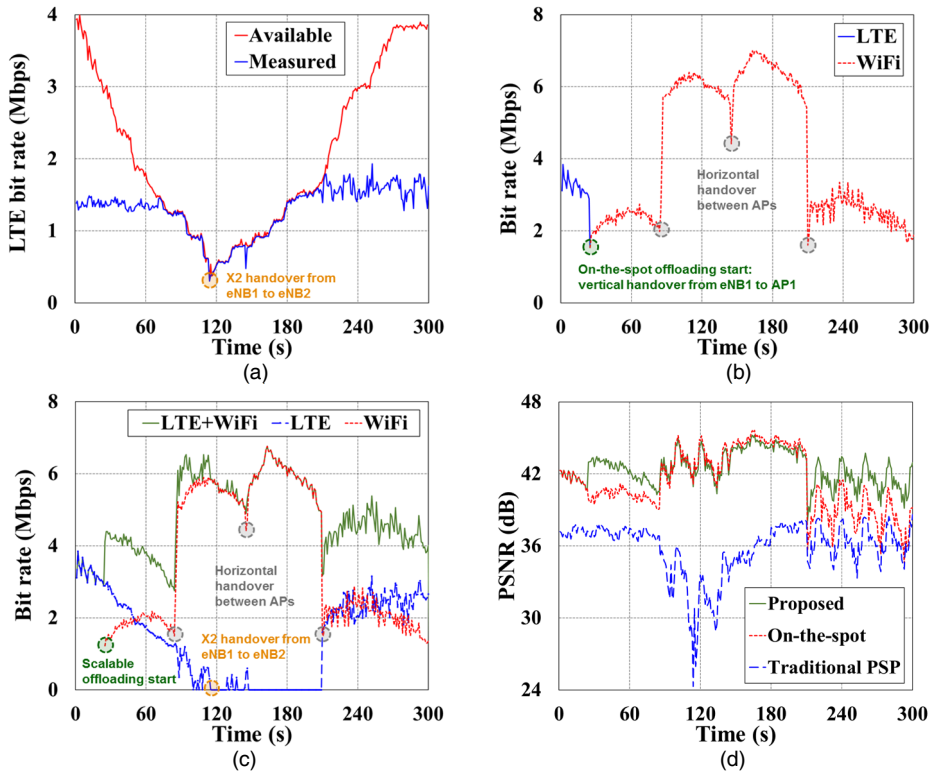

Fig. 13. Performance comparison with existing systems: (a) LTE bitrate with traditional PSP, (b) LTE and WiFi bitrate with on-the-spot offloading, (c) LTE and WiFi bitrate with proposed PSP-based scalable offloading, and (d) PSNR.

share the single LTE network resources according to the PSP auction mechanism without activating the WiFi network interface. On-the-spot offloading [32] is based on a spontaneous connectivity to WiFi, and traffic is only transmitted over the cellular network when WiFi is not available. During the on-the-spot offloading, users perform vertical handovers between heterogeneous wireless networks. When a vertical handover occurs, a substantial handover delay is unavoidable in the process of measurement, initialization, decision, and execution [41]. However, we try to compare the performance of the proposed scalable offloading with the more advanced on-the-spot offloading. In the ideal on-the-spot offloading, the user knows in advance that a vertical handover will occur. Accordingly, the packets are buffered in advance by the network until the next wireless station is prepared to accept the packets, and the video encoding rate is appropriately adjusted according to the significant bandwidth variation during a vertical handover to avoid discontinuity in the video playback.

The simulation results are provided in Fig. 13. Fig. 13 (a) presents a timeline plot of the LTE bitrate at UE 1 with the traditional PSP over single LTE network. As the UE 1 moves away from eNB 1, the bitrate gradually decreases, resulting in an X2 handover to eNB 2 at about 115 s. Although the requested LTE resources are almost fully allocated through the auction, it is difficult to ensure QoS for the video streaming because only a small amount of LTE resources are available due to poor LTE channel conditions during X2 handover. As a result, the video quality degrades significantly, as shown by the blue dotted line in Fig. 13 (d). In the on-the-spot offloading, the UE 1 is initially connected to the eNB 1. At about 25 s, the UE 1 detects an available WiFi AP and performs a vertical handover (from eNB 1 to AP 1 with 802.11g). Subsequently, the UE 1 performs a horizontal handover to other APs supporting various WiFi standards such as 802.11n/ac according to the mobility. In fact, the performance of on-the-spot offloading depends on several factors [63]: available wireless technology (e.g., 802.11a/b/g/n/ac), the characteristics of the APs (e.g., range and capacity), the density of the users in the service area of the AP, and the simultaneous data sending/receiving requests from its users. Thus, it is difficult to guarantee the quality of the WiFi AP selected for mobile data offloading [6]. It is evident from Fig. 13 (b) that the bitrate sometimes decreases when connecting to a low-performance AP (AP 1 with 802.11g between 25 s and 85 s) or sharing the same AP among dense users (AP 4 with 802.11n between 210 s and 300 s). Accordingly, the video quality may deteriorate, as shown by the red dotted line in Fig. 13 (d). To alleviate the phenomenon in which video quality is entirely dependent on the WiFi network states, users in the proposed system effectively consolidate multiple physical paths. By default, all UEs fully utilize the available WiFi resources, and LTE resources are allocated to maximize the social welfare according to the PSP auction rule. Fig. 13 (c) and (d) reveal that the proposed system achieves relatively high and consistent PSNR compared to existing systems. To provide a subjective video quality comparison, the captured 91th frame of Rush Hour of the proposed and the existing systems are presented in Fig. 14.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TMC.2017.2748592, IEEE Transactions on Mobile Computing

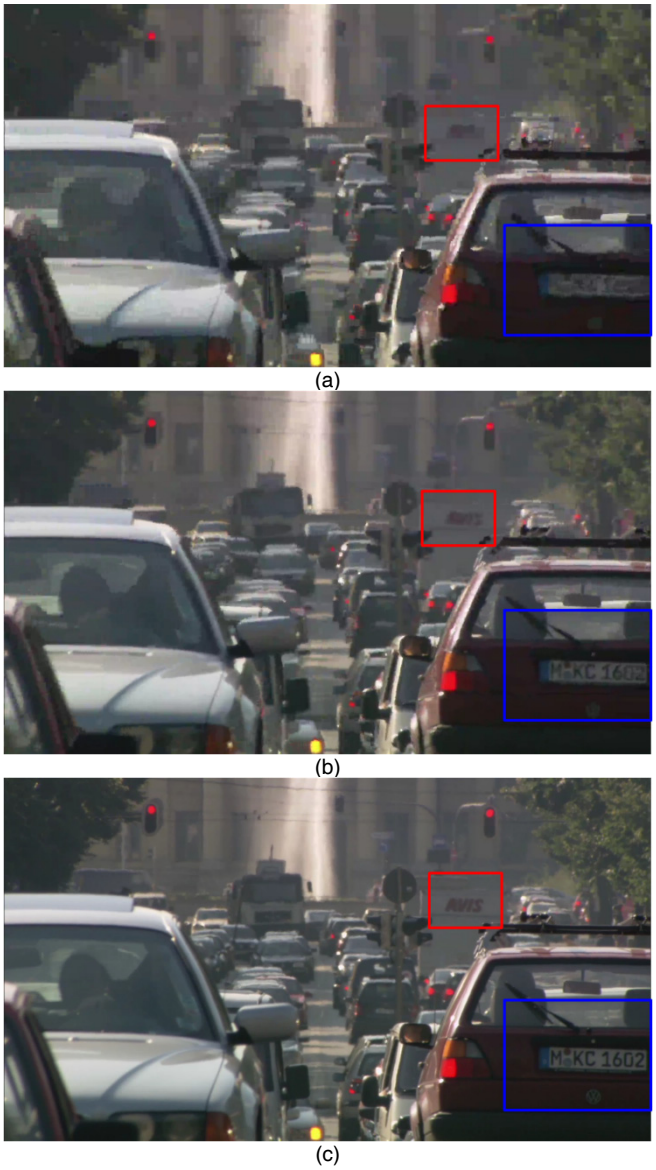

Fig. 14. Subjective video quality comparison of the 91th frame: (a) traditional PSP over single LTE network, (b) on-the-spot offloading, and (c) proposed PSP-based scalable offloading.

TABLE 5
AVERAGE BITRATE AND PSNR COMPARISON WITH EXISTING SYSTEMS

	LTE bitrate (Mbps)	WiFi bitrate (Mbps)	LTE+WiFi bitrate (Mbps)	PSNR (dB)	PSNR std. dev.
Proposed	1.44	3.62	5.06	42.61	1.38
Traditional PSP	1.45	0	1.45	35.81	2.25
On-the-spot	0.26	4.01	4.27	41.16	2.65

As can be seen in the figures, the subjective video quality of the proposed system is much better than that of the existing systems. The average bitrate and PSNR values of all UEs are summarized in Table 5.

4.4 Convergence and Scalability Evaluation of the Proposed System

In this section, we evaluate the performance of the proposed system in terms of the scalability and convergence speed, i.e., how fast the resource negotiation process converges with an increasing number of UEs and the bid submission fee. During the simulation, the available LTE network resource R is fixed at 30 Mbps and the bid submission fee ε at 0.001. For each number of UEs, we perform 100 times auctions, and obtain the average number of rounds until convergence. The simulation results are provided in Fig. 15. It is observed that the average number of resource negotiation is less than 20 rounds, regardless of the number of UEs. In fact, the number of UEs that are not assigned the resources, i.e., $a_i^{LTE} < q_i^{LTE}$, increases as more UEs participate in the auction. As a result, the total number of packets requested from all UEs rapidly approaches R, and thus, the negotiation process can be terminated in a short round. Fig. 16 shows an example of the social welfare trend as the round proceeds when ε is set to 0.001. It is apparent that that the social welfare value is gradually increased and maximized in the final

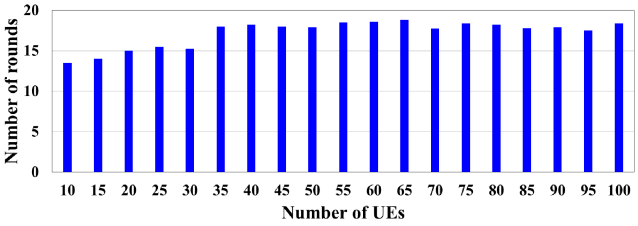

Fig. 15. Average number of rounds until the convergence according to the number of UEs ($\varepsilon = 0.001$).

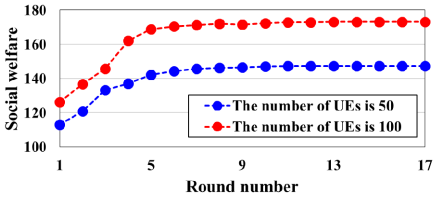

Fig. 16. Social welfare trend as the round proceeds ($\varepsilon = 0.001$).

TABLE 6
AUCTION RESULT SUMMARY ACCORDING TO THE NUMBER OF UEs ($\varepsilon = 0.001$)

	M=20	M=40	M=60	M=80	M=100
Avg. round	15	18.2	18.6	18.2	18.4
Max. round	19	21	20	22	21
Min. round	9	11	12	13	12
Avg. social welfare	126.8	145.44	162.5	171.1	182.5

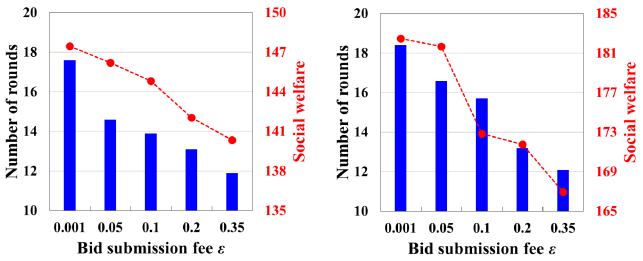

Fig. 17. Relation between the convergence speed and social welfare according to bid submission fee: (a) when the number of UEs is 50 and (b) when the number of UEs is 100.

round. The simulation results are summarized in Table 6.

We now discuss the relation between the convergence speed and social welfare according to ε. As mentioned in Proposition 2, the convergence speed strongly influences the bid submission fee ε. The smaller ε, the closer to the value-optimal allocation can be achieved; however, when the value of ε is smaller, the iteration takes longer to converge. Fig. 17 clearly reveals that the system more quickly reaches convergence as ε increases. In addition, the social welfare value in Fig. 17 (b) is larger than that in Fig. 17 (a). In fact, the social welfare value becomes larger as more UEs participate in the auction since the UE's resource valuation function in (16) is monotonically increasing and concave.

5 Conclusion

In this paper, we have presented a game-theoretic scalable offloading for video streaming services over LTE and WiFi networks. One of the important features of the proposed system is that scalable offloading of video streaming traffics through LTE and WiFi networks is realized without any noticeable video quality degradation. We have adopted the PSP auction mechanism to efficiently allocate the limited LTE resources to UEs. In addition, fountain code was deployed to simultaneously deliver the video traffic through LTE and WiFi networks with diverse characteristics since it can adjust the amount of offloading traffic in a flexible manner and solve packet reordering and packet loss problems in error-prone heterogeneous wireless networks. The proposed feedback-based bid decision algorithm maximizes social welfare while achieving the truthful ε-Nash equilibrium with fast convergence compared to existing bid decision algorithms. For performance verification, we have built the test-bed using NS-3. The simulation results have demonstrated that the proposed system efficiently alleviates the LTE traffic congestion by scalable offloading through a WiFi network, and provides a higher quality of video streaming services than existing system.

There are still problems to be addressed in the proposed system. One of the challenges is the energy consumption of mobile devices, which have limited battery capacity. In general, more power is spent to enable multiple network interfaces concurrently. In addition, power consumption depends on service type and data patterns because the on/off process of network interfaces uses more power and incurs longer delays. Large-scale deployment of the proposed system is another challenge because the current cellular network architecture is difficult to expand, difficult to manage, inflexible, and expensive [64]. Recently, a 5G cellular network architecture based on software-defined networking (SDN) [64], [65] has been proposed to develop a programmable mobile cellular network that allows greater flexibility in management and configuration, and thus overcomes the shortcomings of traditional wireless networks. By deploying the emerging SDN technology, the majority of the LTE EPC functional entities can be migrated to the application plane of the controller. Furthermore, new applications or advanced functions will be easily implemented in software without a specific hardware support. Under the SDN-enabled wireless environment, we will frame the mobile data offloading problem for video streaming to optimize network utilization and application performance, and to enhance the user experience.

References

[1] Cisco White Paper, "Cisco Visual Networking Index – Global Mobile Data Traffic Forecast, Update, 2016–2021," 2017.

[2] Y. Choi, H. Ji, and J. Park, "A 3W Network Strategy for Mobile Data Traffic Offloading," *IEEE Communications Magazine*, pp. 118–123, Oct. 2011.

[3] D. Astely, E. Dahlman, A. Furuskär, Y. Jading, M. Lindström, and S. Parkvall, "LTE: The Evolution of Mobile Broadband," *IEEE Communications Magazine*, vol. 47, no. 4, pp. 44–51, 2009.

[4] 3GPP TR 36.913 v12.0.0, "Technical Specification Group Radio Access Network; Requirements for Further Advancements for Evolved Universal Terrestrial Radio Access (E-UTRA) (LTE-Advanced) (Release 12)," Sep. 2014.

[5] IEEE 802.16 Broadband Wireless Access Working Group, "IEEE Standard for Local and Metropolitan Area Networks, Part 16: Air Interface for Broadband Wireless Access Systems, Amendment 3: Advanced Air Interface," *IEEE 802.16mTM*, May 2011.

[6] A. Aijaz, H. Aghvami, and M. Amani, "A Survey on Mobile Data Offloading: Technical and Business Perspectives," *IEEE Wireless Communications*, vol. 20, no. 2, pp. 104–112, April 2013.

[7] Tefficient industry analysis, "Using Public Wi-Fi as Customer Magnet," 2016.

[8] R. Trestian, O. Ormond, and G.-M. Muntean, "Game Theory-based Network Selection: Solutions and Challenges," *IEEE Communications Surveys and Tutorials*, vol. 14, no. 4, pp. 1212–1231, 2012.

[9] A. Aijaz, N. Uddin, O. Holland, and A. H. Aghvami, "On Practical Aspects of Mobile Data Offloading to Wi-Fi Networks," *Ithaca*, New York, 2015.

[10] Y. Go, O. C. Kwon, and H. Song, "An Energy-Efficient HTTP Adaptive Video Streaming with Networking Cost Constraint over Heterogeneous Wireless Networks," *IEEE Trans. on Multimedia*, vol. 17, no. 9, pp. 1646–1657, 2015.

[11] Y. Zhang, C. Lee, D. Niyato, and P. Wang, "Auction Approaches for Resource Allocation in Wireless Systems: A Survey," *IEEE Communications Surveys and Tutorials*, vol. 15, no. 3, pp. 1020–1041, 2013.

[12] X. Zhuo, W. Gao, G. Cao, and S. Hua, "An Incentive Framework for Cellular Traffic Offloading," *IEEE Trans. on Mobile Computing*, vol. 13, no. 3, pp. 541–555, Mar. 2014.

[13] G. Iosifidis, L. Gao, J. Huang, and L. Tassiulas, "A Double-Auction Mechanism for Mobile Data-Offloading Markets," *IEEE/ACM Trans. on Networking*, vol. 23, no. 5, pp. 1634–1647, Oct. 2015.

[14] N. Cheng, N. Lu, N. Zhang, X. Zhang, X. S. Shen, and J. W. Mark, "Opportunistic WiFi Offloading in Vehicular Environment: A Game-Theory Approach," *IEEE Trans. on Intelligent Transportation Systems*, vol. 17, no. 7, pp. 1944–1955, 2016.

[15] N. Wang and J. Wu, "Opportunistic WiFi Offloading in a Vehicular Environment: Waiting or Downloading Now?," *IEEE International Conference on Computer Communications*, 2016.

[16] P. LV, X. Wang, X. Xue, and M. Xu, "SWIMMING: Seamless and Efficient WiFi-Based Internet Access from Moving Vehicles," *IEEE Trans. on Mobile Computing*, vol. 14, no. 5, 2015.

[17] S. Paris, F. Martignon, I. Filippini, and L. Chen, "An Efficient Auction-based Mechanism for Mobile Data Offloading," *IEEE Trans. on Mobile Computing*, vol. 14, no. 8, pp. 1573–1586, Aug. 2015.

[18] O. C. Kwon, Y. Go, Y. Park, and H. Song, "MPMTP: Multipath Multimedia Transport Protocol using Systematic Raptor Codes over Wireless Networks," *IEEE Trans. on Mobile Computing*, vol. 14, no. 9, pp. 1903–1916, Sep. 2015.

[19] D. MacKay, "Fountain Codes," *IEEE Communications*, vol. 152, no. 6, pp. 1062–1068, 2005.

[20] M. Luby, "LT Codes," *IEEE Symp. on Foundations of Computer Science*, pp. 271–280, 2002.

[21] A. Lazar and N. Semret, "Design and Analysis of the Progressive Second Price Auction for Network Bandwidth Sharing," *Telecommunications Systems*, 1999.

[22] The network simulator 3 (NS-3), Available: http://www.nsnam.org.

[23] P. Maymounkov, "Online Codes," Research Report TR2002-833, New York University, 2002.

[24] A. Shokrollahi, "Raptor Codes," *IEEE Trans. on Information Theory*, vol. 52, no. 6, pp. 2551–2567, 2006.

[25] M. Luby, "Raptor Forward Error Correction Scheme for Object Delivery," RFC 5053, Oct. 2007.

[26] 3GPP TS 26.346 v11.2.0, "Technical Specification Group Services and System Aspects; Multimedia Broadcast/Multicast Service; Protocols and Codecs," 2012.

[27] ETSI EN 302 304 v1.1.1, "Digital Video Broadcasting (DVB); Transmission System for Handheld Terminals (DVB-H)," 2004.

[28] T. Stockhammer, A. Shokrollahi, M. Watson, M. Luby and T. Gasiba, "Application Layer Forward Error Correction for Mobile Multimedia Broadcasting," in Handbook of Mobile Broadcasting: DVB-H DMB ISDB-T and Media FLO, pp. 239–280, 2008, CRC Press.

[29] M. Luby, A. Shokrollahi, M. Watson, T. Stockhammer, L. Minder, "RaptorQ Forward Error Correction Scheme for Object Delivery," IETF RFC 6330, 2011.

[30] V. Krishna, Auction Theory, Academic Press, 2002.

[31] S. Dimatteo, P. Hui, B. Han, and V. Li, "Cellular Traffic Offloading through WiFi Networks," *Proc. of IEEE MASS*, pp. 192–201, 2011.

[32] K. Lee, J. Lee, Y. Yi, I. Rhee, and S. Chong, "Mobile Data Offloading: How Much Can WiFi Deliver?" *IEEE/ACM Trans. on Networking*, vol. 21, no. 2, pp. 536–550, Apr. 2013.

[33] T. Wang, G. Xing, M. Li, and W. Jia, "Efficient WiFi Deployment Algorithms based on Realistic Mobility Characteristics," *IEEE MASS*, 2010.

[34] E. Bulut, and B. Szymanski, "WiFi Access Point Deployment for Efficient Mobile Data Offloading," Proc. of ACM PINGEN, 2012.

[35] N. Cheng, N. Lu, N. Zhang, X. S. Shen, and J. W. Mark, "Vehicular WiFi Offloading: Challenges and Solutions," *Vehicular Communications*, 2014.

[36] V. F. Mota, D. F. Macedo, Y. Ghamri-Doudane, and J. M. S. Nogueira, "On the Feasibility of WiFi Offloading in Urban Areas: The Paris Case Study," *IEEE Wireless Days*, 2013.

[37] V. Bychkovsky, B. Hull, A. Miu, H. Balakrishnan, and S. Madden, "A Measurement Study of Vehicular Internet Access using in Situ Wi-Fi Networks," *Proc. of ACM MobiCom*, 2006.

[38] A. Balasubramanian, R. Mahajan, and A. Venkataramani, "Augmenting Mobile 3G using WiFi," *Proc. of ACM MobiSys*, pp. 209–222, Jun. 2010.

[39] S. Nirjon, A. Nicoara, C.-H. Hsu, J. Singh, and J. Stankovic, "MultiNets: A System for Real-Time Switching between Multiple Network Interfaces on Mobile Devices," *ACM Trans. on Embedded Computing Systems*, vol. 13, no. 4, 2014.

[40] A. Y. Ding, Y. X. B. Han, P. Hui, A. Srinivasank, M. Kojo, and S. Tarkoma, "Enabling Energy-Aware Collaborative Mobile Data Offloading for Smartphones," *Proc. of IEEE SECON*, pp. 1–9, New Orleans, LA, Jun. 2013.

[41] S. Ko and K. Chung, "A Handover-aware Seamless Video Streaming Scheme in Heterogeneous Wireless Networks," *Annals of Telecommunications*, pp. 1–12, 2013.

[42] J.-Y. Pyun, "Context-aware Streaming Video System for Vertical Handover over Wireless Overlay Network," *IEEE Trans. on Consumer Electronics*, vol. 54, no. 1, pp. 71–79, Feb. 2008.

[43] A. Calvagna and G. D. Modica, "A User-centric Analysis of Vertical Handovers," *Proc. of ACM WMASH*, 2004.

[44] S. C. Han, H. Joo, D. Lee, and H. Song, "An End-to-End Virtual Path Construction System for Stable Video Streaming over Heterogeneous Wireless Networks," *IEEE J. on Selected Area in Communications*, vol. 29, no. 5, May 2011.

[45] A. Ford, C. Raiciu, M. Handley, S. Barre, and J. Iyengar, "Architectural Guidelines for Multipath TCP Development," IETF RFC 6182, Mar. 2011.

[46] V. Sharma, K. Kar, K.K. Ramakrishnan, and S. Kalyanaraman, "A Transport Protocol to Exploit Multipath Diversity in Wireless Networks," *IEEE/ACM Trans. on Networking*, vol. 20, no. 4, pp. 1024–1039, Aug. 2012.

[47] D. Ho, Y. Park, and H. Song, "QoS-Supporting Video Streaming System with Minimum Data Service Cost over Heterogeneous Wireless Networks," *J. of Visual Communication and Image Representation*, vol. 24, no. 18, pp. 1293–1302, Nov. 2013.

[48] 3GPP TS 23.402 v13.2.0, "Technical Specification Group Services and System Aspects; Architecture Enhancements for Non-3GPP Accesses (Release 13)," Jun. 2015.

[49] 3GPP TS 24.312 v12.9.0, "Technical Specification Group Core Network and Terminals; Access Network Discovery and Selection Function (ANDSF) Management Object (MO) (Release 12)," Jun. 2015.

[50] 4G America White Paper, "Integration of Cellular and Wi-Fi Networks," Sep. 2013.

[51] 3GPP TR 23.829 v10.0.1, "Technical Specification Group Services and System Aspects; Local IP Access and Selected IP Traffic Offload (LIPA-SIPTO) (Release 10)," Oct. 2011.

[52] 3GPP TS 23.261 v12.0.0, "Technical Specification Group Services and System Aspects; IP Flow Mobility and Seamless Wireless Local Area Network (WLAN) Offload (Release 12)," Sep. 2014.

[53] V. Ribeiro, R. Riedi, R. Baraniuk, J. Navratil, and L. Cottrell, "PathChirp: Efficient Available Bandwidth Estimation for Network Paths," *Passive and Active Measurement Workshop*, 2003.

[54] D. Lee and H. Song, "A Robust Luby Transform Encoding Pattern-Aware Symbol Packetization Algorithm for Video Streaming over Wireless Network," *IEEE Trans. on Multimedia*, vol. 13, no. 4, pp. 788–796, Aug. 2011.

[55] D. Jurca and P. Frossard, "Media Flow Rate Allocation in Multipath Networks," *IEEE Trans. on Multimedia*, vol. 9, no.6, pp. 1227–1240, Oct. 2007.

[56] TATA White Paper, "Radio Resource Management – Radio Admission Control and Bearer Control," 2013.

[57] N. H. Mastronarde and M. van der Schaar, "Automated Bidding for Media Services at the Edge of A Content Delivery Network," *IEEE Trans. on Multimedia*, vol. 11, no. 3, 2009.

[58] H. T. Friis, "A Note on a Simple Transmission Formula," *Proc. of the IRE*, vol. 34, no. 5, pp. 254–256, May 1946.

[59] 3GPP TS 36.104 v13.0.0, "Technical Specification Group Radio Access Network; Evolved Universal Terrestrial Radio Access (E-UTRA); Base Station (BS) Radio Transmission and Reception (Release 13)," Jul. 2015.

[60] M. Lacage and T.R. Henderson, "Yet Another Network Simulator," *WNS2'06*, 2006.

[61] 3GPP TS 36.331 v14.2.0, "Technical Specification Group Radio Access Network; Evolved Universal Terrestrial Radio Access (E-UTRA); Radio Resource Control (RRC); Protocol specification (Release 14)," Mar. 2017.

[62] Video test sequence, Available: http://cs-nsl-wiki.cs.surrey.sfu.ca/wiki/ Video_Library_and_Tools.

[63] E. Bulut and B. K. Szymanski, "Rethinking Offloading WiFi Access Point Deployment from User Perspective," *Proc. of IEEE WiMob*, 2016.

[64] Open Networking Foundation, "Software-Defined Networking: The New Norm for Networks," White Paper, Apr. 2012.

[65] A. Bradai, K. Singh, T. Ahmed, and T. Rasheed, "Cellular Software Defined Networking: A Framework," *IEEE Communications Magazine*, vol. 53, no. 6, pp. 36–43, 2015.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TMC.2017.2748592, IEEE Transactions on Mobile Computing

Donghyeok Ho received the B.S. degree in Dept. of Media from Soongsil University, Korea in 2010. He received his M.S. degree from Dept. of Computer Science and Engineering, Pohang University of Science and Technology (POSTECH), Korea in Feb. 2012. Currently, he is a Ph.D. student in the Division of IT Convergence and Engineering, POSTECH, Korea. His research interests include adaptive video streaming, game theoretic resource allocation, traffic control, and software defined-networking.

Gi Seok Park received the B.S. degree (*summa cum laude*) in Electrical Engineering from Dongguk University, Seoul, Korea, in 2010. He received his M.S. degree in IT Convergence Engineering from Pohang University of Science and Technology (POSTECH), Korea in 2013. Currently, he is a Ph.D. student in IT Convergence and Engineering, POSTECH, Korea. His research interests are in broad area of network optimization and future internet.

Hwangjun Song received the B.S. and M.S. degrees from Dept. of Control and Instrumentation (EE), Seoul National University, Korea in 1990 and 1992, respectively, and Ph.D. degree in Electrical Engineering-Systems, University of Southern California, Los Angeles, CA, USA in 1999. From 1995 to 1999, he was a research assistant in SIPI (Signal and Image Processing Institute) and IMSC (Integrated Media Systems Center), Univ. of Southern California. From 2000 to 2005, he was an assistant professor/vice dean of admission affairs at Hongik University, Seoul, Korea. Since Feb. 2005, he has been with Dept. of Computer Science and Engineering, POSTECH (Pohang University of Science and Technology), Korea. He received Haedong Paper Award from Korean Institute of Communication Science in 2005. He was a courtesy associate professor at Univ. of Florida in 2011-2012. He served as a vice president of Korean Institute of Information Scientists and Engineers in 2016-2017. He is an APSIPA distinguished lecturer for 2017-2018. He is an editorial board member of Journal of Visual Communication and Image Representation and an associate editor of Journal of Communications and Networks, and served as an editorial board member of International Journal of Vehicular Technology and a guest editor of Special Issue on "Network technologies for emerging broadband multimedia services" in the Journal of Visual Communication and Image Representation and Special Issue on "Wireless & Mobile networks" in the International Journal of Ad Hoc and Ubiquitous Computing. His research interests include multimedia signal processing and communication, image/video compression, digital signal processing, network protocols necessary to implement functional image/video applications, and control system.