

# SERVICE PROVISIONING AND PROFIT MAXIMIZATION IN NETWORK-ASSISTED ADAPTIVE HTTP STREAMING

Zhisheng Yan<sup>1</sup>, Cedric Westphal<sup>2,3</sup>, Xin Wang<sup>2</sup>, and Chang Wen Chen<sup>1</sup>

<sup>1</sup>State University of New York at Buffalo, <sup>2</sup>Huawei US R&D Center, <sup>3</sup>UCSC

## ABSTRACT

Mobile adaptive HTTP streaming with centralized consideration of multiple streams has gained increasing interest. It poses a special challenge that the interests of both content provider and network operator need to be deliberately balanced. More importantly, the adaptation is required to be flexible enough to be ported to various systems that work under different network environments, QoE levels, and economic objectives. To address these challenges, we propose a Markov Decision Process (MDP) based network-assisted adaptation framework, wherein cost of buffering, significant playback variation, bandwidth management and income of playback are jointly investigated. We then demonstrate its promising service provisioning and maximal profit for a mobile network in which fair or differentiated service is required.

## 1. INTRODUCTION

Thanks to strong scalability and versatility, HTTP adaptive streaming has established itself as the dominant technique for future Internet video delivery. The most challenging research task for HTTP adaptive streaming is the rate selection process, that is, which quality version of a segment should be streamed. Many works [1, 2] have been proposed towards the optimal adaptation conducted by the client itself using local measurements and estimation. Although these rate adaptation algorithms showed promising performance in single-user networks, they cannot be directly applied to multi-client mobile networks with shared bandwidth bottleneck. Playback instability and unfairness have been commonly identified as among the weaknesses for client-side adaptation in practical experiments, where two or more clients compete for the same bottleneck [3, 4]. In fact, the fundamental problem behind the issues is that client-side adaptation naively assumes the wireless network is utilized by the client itself and has no knowledge of other competing streams in the shared bottleneck.

Therefore, it is natural for the network operator to arbitrate such competition and globally adapt the clients' decisions by considering multiple players' status and the overall mobile network. MPEG and 3GPP have included such ideas into its working draft [5] or technical report [6] in order to further enhance the quality of experience (QoE) of Internet

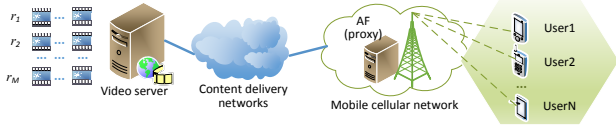
streaming. In addition to performance improvement, there are also economic incentives for a network operator to join the adaptive streaming ecosystem. For instance, content provider *Netflix* has recently entered a peering agreement on smooth streaming with network operator *Comcast* [7]. One of the most critical challenge, however, is how to balance the interests of a content provider and a network operator. Indeed, a content provider always wants its users to experience satisfactory QoE whereas a mobile operator sometimes might be more concerned about its bandwidth cost under congestion. Hence, the adaptation strategy shall be able to guarantee the users' QoE, as well as maximizing the operator's profit. On top of this important challenge, it is also crucial to design the joint adaptation framework in a flexible and systematical way such that the framework can be easily ported to various applications and network environments. For example, there are diverse types of service provision, such as fair experience versus differentiated QoE, high-quality playback versus highly-smooth playback, and different cost models of network operators, such as bandwidth-centric versus user-number-centric.

Only a few research works have been focused on Internet streaming with competing clients. Some exploited traffic shaping in the server side to orchestrate the experience of two competing users [3, 4]. Others aimed at optimizing various kinds of QoE utility, such as rate-resolution utility [8], QoE continuum [9], rate-buffering utility [10], and loss-jitter-delay mapped utility [11], by developing individual optimal or heuristic algorithm. Finally, there are some interesting works using admission control of new clients to maximize the bandwidth utilization [12]. Nonetheless, all these works concentrate on a specific QoE objective only addressing the interests of content providers, which limits their applicability.

The major contribution of this research is in the development of a *network-assisted* adaptive HTTP streaming framework for wireless bottleneck link from the perspective of both content providers and network operators. The scheme can flexibly achieve different system-level service requirements by tuning certain framework parameters and can maximize the economic profit of the operator at the same time. Specifically, we propose a Markov Decision Process (MDP) based framework to carry out the adaptation, where we jointly consider the clients' bandwidth and service requirements, and thoroughly investigates the operator's economic benefits. In

---

This work was done during an internship at Huawei US R&D Center.



**Fig. 1.** Architecture of the proposed mobile streaming system.

the following, we focus on mobile cellular networks to introduce the framework and we describe the system from the point of view of mobile operators. Then we demonstrate two case studies where service fairness and service differentiation (for premium users and regular users) is required by a content provider within an operator-controlled mobile network. These have been identified as major use cases of network-assisted streaming by MPEG [5]. However, it should be noted that the applicability of the framework is general and can be properly adapted to other service requirements.

## 2. MDP-BASED NETWORK-ASSISTED ADAPTATION FRAMEWORK

We consider a cellular Internet streaming system as shown in Fig. 1. We target one cell, where  $\mathcal{U}$  is the set of users and each user is indexed by  $i = 1, 2, \dots, N$ . There are a set of videos  $\mathcal{V}$  with different quality, characterized by bit-rate  $r_j$  where  $j = 1, 2, \dots, M$ . Each quality version is split into  $T_{seg}$ -second segments. We shift the intelligence from clients to a logically centralized controller within mobile operator's network, i.e., the Application Function (AF) within 3GPP networks [6], for the network-assisted adaptation.

The proposed system works as follows. The video server initially sends out a description of video versions. At each switching point that occurs every  $T_{seg}$  seconds, a user requests a segment based on local throughput measurements that need very low complexity. Such request only serves as a *preliminary suggestion* and therefore no corresponding local optimization in the client is carried out. In other words, there would be no conflict if this request is modified later. Unlike conventional client-side adaptation where the mobile operator simply forwards the client requests to the video server, the mobile operator here will re-write the request based on MDP-based cell-wide adaptation done by the AF. The client feedback for adaptation, such as available video representations and bandwidth measurement, can be feasibly communicated to the AF. For instance, it can be achieved by standardized quality metrics reporting process [13] or new techniques like URL parameter insertion [14]. The final adaptation decision then will be delivered to the cellular scheduler and the video server. We adopt a rate proportional scheduling wherein the downlink resources are allocated based on the upper layer video rate. That way, the video version decided from system's perspective can be effectively streamed to the users.

### 2.1. Wireless Bandwidth Model

We model the last-hop cellular link for each client as a Markov process, which has been widely studied and proved

to be effective [15]. Each link has  $K$  states of available bandwidth, namely  $\mathcal{B} = \{bw_k | k = 1, 2, \dots, K\}$ .

The clients measure their own downlink bandwidth and periodically report it to AF, which serves as the input to the adaptation framework. Note that the measured bandwidth may not be exactly the value of  $bw_k$ . We thereby divide the available bandwidth domain into  $K$  regions and map the real bandwidth to the corresponding region.

### 2.2. Mobile Operator Profit Model

**Normal Playback Income.** At switching point  $t$ , the mobile operator is expected to obtain income  $I_{play}$  from the content provider if it can guarantee a normal playback with rate  $R_{t+1}$  during the upcoming period. Formally, this implies  $R_{t+1} \leq BW_{t+1}$ , where  $R_{t+1} \in \mathcal{V}$  and  $BW_{t+1} \in \mathcal{B}$  is the estimated bandwidth of link. Considering the logarithmic relationship in rate distortion theory, we model the income from client  $i$  as

$$I_{i,play} = \Gamma_{play} \alpha \frac{\log \frac{R_{i,t+1}}{R_{min}}}{\eta_{play}} \quad (1)$$

where  $\alpha$  ( $0 \leq \alpha \leq 1$ ) is the sharing weight that normal playback places on QoE and  $R_{min} = \min\{r_j | r_j \in \mathcal{V}\}$ . We denote the normalization factor as  $\eta_{play} = \log \frac{R_{max}}{R_{min}}$ , where  $R_{max} = \max\{r_j | r_j \in \mathcal{V}\}$ .  $\Gamma_{play}$  is an indicator function for the normal playback, namely  $\Gamma_{play} = 1$  if  $R_{t+1} \leq BW_{t+1}$  and  $\Gamma_{play} = 0$  otherwise.

**Playback Buffering Cost.** If the operator-adapted bit-rate exceeds the available bandwidth, playback buffering would occur, which should not be encouraged in the service contract. Hence, we impose certain penalty cost  $C_{buf}$  on player  $i$ .

$$C_{i,buf} = (1 - \Gamma_{play}) \beta \frac{\log \frac{R_{i,t+1} - BW_{i,t+1}}{\delta_{min}}}{\eta_{buf}} \quad (2)$$

where  $\beta$  ( $0 \leq \beta \leq 1$ ) is the re-buffering weight, and  $\delta_{min} = \min\{r_j - bw_k | r_j \in \mathcal{V}, bw_k \in \mathcal{B}\}$ . The normalization factor of playback buffering is  $\eta_{buf} = \log \frac{R_{max} - BW_{min}}{\delta_{min}}$ , where  $BW_{min} = \min\{bw_k | bw_k \in \mathcal{B}\}$ .

**Playback Smoothness.** If the rate variation in two consecutive periods is too significant, the content provider would penalize the mobile operator some money  $C_{var}$ .

$$C_{i,var} = \Gamma_{var} \gamma \frac{\log \frac{|R_{i,t} - R_{i,t+1}|}{\Delta}}{\eta_{var}} \quad (3)$$

where  $\gamma$  ( $0 \leq \gamma \leq 1$ ) is the smoothness weight and  $\alpha + \beta + \gamma = 1$ ,  $\Delta$  is the rate variation threshold below which no penalty would occur, and  $\eta_{var} = \log \frac{R_{max} - R_{min}}{\Delta}$  is the normalization factor. We can thus express the indicator function  $\Gamma_{var}$  as,  $\Gamma_{var} = 1$  if  $R_t - R_{t+1} \geq \Delta$  and  $\Gamma_{var} = 0$  otherwise.

**Bottleneck Bandwidth Cost.** Finally, the mobile cellular network may experience radio congestion, during which the mobile operator would have to spend proportionally more bandwidth cost to manage the network:

$$C_{bw} = \Gamma_{bw} \theta \left( \sum_{i \in \mathcal{U}} R_{i,t+1} - R_{th} \right) \quad (4)$$

where  $R_{th}$  is the total service rate constraint above which the mobile operator has to spend extra money,  $\theta$  is the cost per unit exceeded rate,  $\Gamma_{bw} = 1$  when  $\sum_{i \in \mathcal{U}} R_{i,t+1} > R_{th}$ , and otherwise  $\Gamma_{bw} = 0$ .

### 2.3. Markov Decision Process Formulation

MDP is a 4-tuple reinforcement learning task [16] that can intelligently interact with the uncertain bandwidth estimation and make the adaptation decision accordingly.

**System States.** The system state at time period  $t$  can be defined as  $s_t = (\mathbf{R}_t, \mathbf{BW}_t)$ , where the vector represents the bandwidth and quality version for all clients in  $\mathcal{U}$ . Besides, the system state set  $\mathcal{S} = \{s_1, s_2, \dots, s_T\}$  essentially keeps track of the evolution of the entire streaming system, where  $T$  is session duration. Since the system state depends only on its most recent (previous) state, the Markov property holds.

**System Actions.** The action set  $\mathcal{A} = \{a_1, a_2, \dots, a_T\}$  defines the operator-selected quality version of video for all the users. In particular,  $a_t = \mathbf{R}_{t+1}$ , denotes the quality version to be streamed in the upcoming adaptation period.

**State Transitions.** The state transition from  $s_t$  to  $s_{t+1}$  is determined by the decided quality version and available bandwidth at time period  $t$ . Given that the mobility-decided bandwidth variation is independent of bit-rate decisions, the state transition probability  $\mathcal{P}_{a_t}(s_t, s_{t+1})$  can be derived as,

$$\begin{aligned} \mathcal{P}_{a_t}(s_t, s_{t+1}) &= \Pr\{s_{t+1}|s_t\} \\ &= \Pr\{(\mathbf{R}_{t+1}, \mathbf{BW}_{t+1})|(\mathbf{R}_t, \mathbf{BW}_t)\} \\ &= \Pr\{\mathbf{R}_{t+1}|\mathbf{R}_t, \mathbf{R}_{t+1} = a_t\} \Pr\{\mathbf{BW}_{t+1}|\mathbf{BW}_t\} \end{aligned} \quad (5)$$

where  $\Pr\{\mathbf{BW}_{t+1}|\mathbf{BW}_t\}$  can be obtained from the transition matrix of the channel model and  $\Pr\{\mathbf{R}_{t+1}|\mathbf{R}_t, \mathbf{R}_{t+1} = a_t\}$  is decided by the action policy.

**Profit Function and Optimization Objective.** The profit function  $\mathcal{R}_{a_t}(s_t, s_{t+1})$  is the overall monetary profit of the mobile operator from  $s_t$  to  $s_{t+1}$ , i.e.,

$$\mathcal{R}_{a_t}(s_t, s_{t+1}) = \sum_{i \in \mathcal{U}} \lambda_i (I_{i,play} - C_{i,buf} - C_{i,var}) - C_{bw} \quad (6)$$

where  $\lambda_i$  ( $0 \leq \lambda_i \leq 1$ ) is the service priority coefficient for user  $i$  and  $\sum_{i \in \mathcal{U}} \lambda_i = 1$ . Hence, the optimization objective of the network-assisted streaming is that we seek to take the optimal policy  $\pi$  such that the profit of mobile operator during the  $T$ -second session can be maximized. Thus the objective is  $\mathcal{OBJ} = \max_{\pi} \sum_{t=0}^{t=T} \mathcal{R}_{a_t}(s_t, s_{t+1})$ .

**Framework Flexibility.** All the above models can be customized for the mobile operator or content providers. Importantly, these modified frameworks can still be optimally solved as introduced next because they only lead to a different profit function  $\mathcal{R}_{a_t}(s_t, s_{t+1})$ .

### 2.4. MDP Solution

We propose a dynamic programming based algorithm to solve the profit-maximized rate adaptation problem. We first let

$v(s_t)$  be the maximal expected profit from  $s_t$  to the end state  $s_T$ . Based on Bellman value iteration [16], we have

$$v(s_t) = \max_{a_t} \left\{ \sum_{s_{t+1} \in \mathcal{S}} \mathcal{P}_{a_t}(s_t, s_{t+1}) (\mathcal{R}_{a_t}(s_t, s_{t+1}) + v(s_{t+1})) \right\} \quad (7)$$

where  $\mathcal{S} = \mathcal{V}^N \times \mathcal{B}^N$  is the state space. This essentially means, for a given current state  $s_t \in \mathcal{S}$  and a given action  $a_t$ , we have multiple possible next states  $s_{t+1} \in \mathcal{S}$  and need to compute the expected rewards for all the transitions. Then we can obtain the optimal rewards by selecting the action achieving the highest expected rewards. By substituting  $v(s_T) = 0$ , the iteration will have a initial value. Iteratively, we can compute the optimal  $v(s_0)$  (equal to the objective  $\mathcal{OBJ}$ ) for all possible states  $s_0 \in \mathcal{S}$  and accordingly the optimal policy.

$$\pi = \arg \max_a \left\{ \sum_{s_1 \in \mathcal{S}} \mathcal{P}_a(s_0, s_1) (\mathcal{R}_a(s_0, s_1) + v(s_1)) \right\} \quad (8)$$

---

#### Algorithm 1 Dynamic Programming Adaptation Algorithm

---

- 1: **Initialization:**  $t \leftarrow T - 1$ ,  $v(s_{t+1}) \leftarrow 0$
  - 2: **while**  $t \geq 0$  **do**
  - 3:   **for** all possible  $s_t \in \mathcal{S}$  **do**
  - 4:      $v(s_t) \leftarrow \max_{a_t} \{ \sum_{s_{t+1} \in \mathcal{S}} \mathcal{P}_{a_t}(s_t, s_{t+1}) (\mathcal{R}_{a_t}(s_t, s_{t+1}) + v(s_{t+1})) \}$
  - 5:      $+ v(s_{t+1})) \}$
  - 6:      $v(s_{t+1}) \leftarrow v(s_t)$
  - 7:    $t \leftarrow t - 1$
  - 8: **Output** the optimal deterministic policy  $\pi$ .
- 

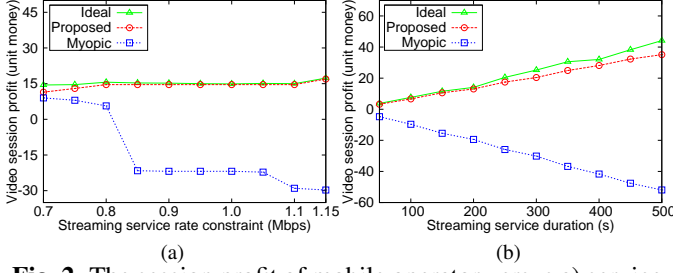
An table that maps each state to the optimal quality version would be generated. Then the mobile operator can make the decision by looking up the table at each switching point. We summarize the iterative algorithm in Algorithm 1.

### 3. PERFORMANCE EVALUATIONS

We evaluate the proposed systems using two case studies, i.e., fair services and differentiated services, with two competing users using *ns-2*. We use H.264/AVC encoded sequence, “Big Buck Bunny”. The bit-rate of 5 quality levels is 95.11, 183.53, 364.63, 493.02, and 798.09 Kbps, respectively.

We use a four-state Markov last-hop channel. Each state lasts for 1 second. We also select segment length  $T_{seg}$  to be 1 second to correspond to the channel coherence time. Since we are interested in the service provisioning by the mobile operator, we adopt the same Markov model for all users to eliminate the impacts of channel differentiation. We adopt a Markov channel matrix used in [2] to simulate the cellular link, i.e.,  $\mathbf{P}_t = [.5, .5, 0, 0; .2, .6, .2, 0; 0, .1, .7, .2; 0, 0, .2, .8]$ . The bandwidth region boundaries for the four state are 256, 512, and 896 Kbps. We use the core network settings in [10].

Regarding the profit model, we assume the overall service rate constraint  $R_{th}$  is 850 Kbps. The congestion penalty  $\theta$  is set to  $\infty$ , which indicates that aggressive bit-rate selection is completely forbidden. We set the sharing weight  $\alpha$ ,  $\beta$  and  $\gamma$  as 0.3, 0.5, and 0.2, respectively, for higher emphasis of buffering. We have run simulations with different sets of



**Fig. 2.** The session profit of mobile operator versus a) service rate constraint  $R_{th}$  during congestion; b) session duration  $T$ .

weights, and found them insensitive to service provisioning. However, the actual profit of operator would be varied, which indicates mobile operators' potential tradeoff between profit and service agreement. The threshold for rate consistency  $\Delta$  is set to be 350 Kbps. The service priority coefficient  $\lambda$  is set to be 0.5/0.5 and 0.7/0.3 for the two users in the fair and differentiated services case, respectively. The initial quality are the lowest version and the player's buffer is set to be 80 frames. The video session runs 200 seconds for 15 times.

### 3.1. Fair Services Provisioning

In this section, we evaluate the performance of different adaptation algorithms when applied in the network-assisted system, where fair service provisioning is required:

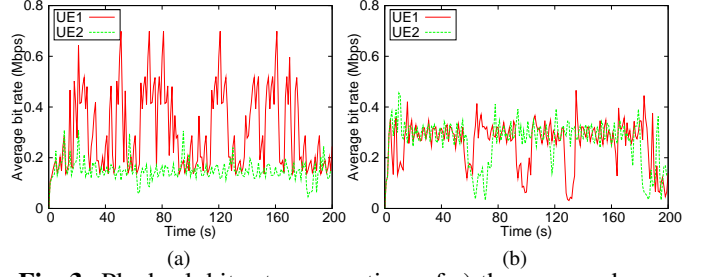
- i) **Upper bound algorithm** (referred as *Ideal*): Same as Algorithm 1, except that future channel condition is perfectly known by the mobile operator such that the MDP transition probability becomes 1.0 when considering an action. Thus we can compute the actual rather than the expected rewards.
- ii) **Myopic algorithm** (referred as *Myopic*): The network-assisted system directly employs the local throughput-based adaptation decision from clients.

We show the total profit of mobile operator versus service rate constraint in Fig. 2a. We also demonstrate the profit versus session durations in Fig. 2b. As shown in the figures, the proposed algorithm achieves a similar performance as *Ideal*. This actually demonstrates the insensitivity of the proposed algorithm to channel model accuracy, which generalizes its applicability. Additionally, the proposed algorithm shows a superior performance over *Myopic* because *Myopic*, without studying the profit factors, may decide a high bit-rate that causes client re-buffering or too significant rate variations.

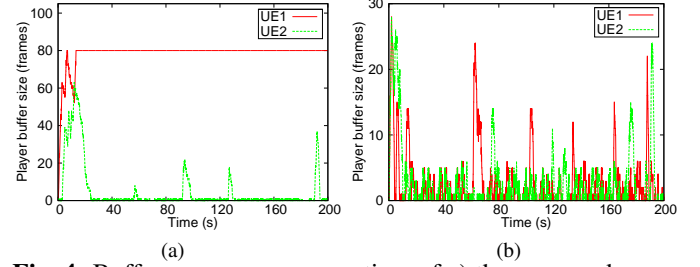
### 3.2. Differentiated Services Provisioning

We now evaluate the performance of proposed system under the requirement of service differentiation. Since we obtained a similar profit comparison curve as Fig. 2 for the differentiated services case, we will focus on evaluating other playback metrics, compared with conventional client-centric system.

Fig. 3 shows playback bit-rate in one session. The high-priority user (UE1) achieves a significant higher rate in the network-assisted system. This is because the mobile operator consider the differentiated service by tuning the priority coef-



**Fig. 3.** Playback bit-rate versus time of a) the proposed system; b) conventional client-centric system.



**Fig. 4.** Buffer occupancy versus time of a) the proposed system; b) conventional client-centric system.

**Table 1.** Playback Quality Metrics

UE1/UE2	Average bit-rate	Buffering ratio	Occurrence of buffering
Proposed	303.95 / 150.18 (Kbps)	0 / 17.71%	0 / 4.43 (per second)
Client-centric	260.64 / 266.06 (Kbps)	18.69% / 18.79%	4.67 / 4.70 (per second)

ficient, accordingly making service-aware adaptation. However, users enjoy a similar playback rate in client-centric system since they essentially compete for the shared bandwidth. We also found that significant rate variations ( $>350$  Kbps) for the proposed system takes more than 5 segments and there is no variations within one segment exceeding the threshold.

For buffer occupancy as shown in Fig. 4, we observe that the high priority user never suffers buffer underflow in the network-assisted system even though congestion occurs. This is at the expense of performance degradation of the low priority user, which is acceptable based on service differentiation agreement. In the client-centric system, nonetheless, the competing players is frequently re-buffered.

We also show industry-standard performance metrics [17] in Table 1. We observe that the proposed system generally outperforms the client-centric system. It can also satisfactorily meet the service differentiation requirement.

## 4. CONCLUSION

We proposed a generalized MDP-based adaptation framework for network-assisted mobile streaming by considering the playback, bandwidth, and economic factors in the mobile operator's view. With two case studies, the proposed framework is shown to outperform conventional client-side adaptation in terms of service provisioning. It also achieves near-ideal profit for the operator. Despite of potential complexity of MDP, the burden can be significantly released when implemented in the operator's proxy clouds. Future work can be focused on extending the framework to a larger user scale.

## 5. REFERENCES

- [1] Z. Li, X. Zhu, J. Gahm, R. Pan, H. Hu, A. Begen, and D. Oran, "Probe and adapt: rate adaptation for HTTP video streaming at scale," *IEEE J. Sel. Areas Commun.*, vol. 32, no. 4, pp. 719–733, Apr. 2014.
- [2] M. Xing, S. Xiang, and L. Cai, "Rate adaptation strategy for video streaming over multiple wireless access networks," in *IEEE Global Communications Conference (GLOBECOM)*, Anaheim, USA, Dec. 2012, pp. 5745–5750.
- [3] R. Houdaille and S. Gouache, "Shaping HTTP adaptive streams for a better user experience," in *ACM conference on Multimedia systems (MMSys)*, Chapel Hill, USA, Feb. 2012, pp. 1–9.
- [4] S. Akhshabi, L. Anantkrishnan, C. Dovrolis, and A. Begen, "Server-based traffic shaping for stabilizing oscillating adaptive streaming players," in *ACM Workshop on Network and Operating Systems Support for Digital Audio and Video (NOSSDAV)*, Oslo, Norway, Feb. 2013, pp. 19–24.
- [5] A. Giladi and T. Stockhammer, "Descriptions of core experiments on DASH amendment," MPEG N14619, Jul. 2014, Available: <http://goo.gl/NJjnAz>.
- [6] *Improved Support for Dynamic Adaptive Streaming over HTTP in 3GPP*, Std. 3GPP TR 26.938, 2014, Available: <http://www.3gpp.org/DynaReport/26938.htm>.
- [7] Netflix to Pay Comcast for Smoother Streaming, Available: <http://goo.gl/NPWxYm>.
- [8] D. De Vleeschauwer, H. Viswanathan, A. Beck, S. Benno, G. Li, and R. Miller, "Optimization of HTTP adaptive streaming over mobile cellular networks," in *IEEE INFOCOM*, Turin, Italy, Apr. 2013, pp. 898–997.
- [9] Z. Yan, J. Xue, and C. W. Chen, "QoE continuum driven HTTP adaptive streaming over multi-client wireless networks," in *IEEE International Conference on Multimedia and Expo (ICME)*, Chengdu, China, Jul. 2014, pp. 1–6.
- [10] W. Pu, Z. Zou, and C. W. Chen, "Video adaptation proxy for wireless dynamic adaptive streaming over HTTP," in *IEEE International Packet Video Workshop (PV)*, Munich, Germany, May 2012, pp. 167–172.
- [11] W.-H. Hsu and C.-H. Lo, "QoS/QoE mapping and adjustment model in the cloud-based multimedia infrastructure," *IEEE Syst. J.*, vol. 8, no. 1, pp. 247–255, Mar. 2014.
- [12] Z. Yan, C. W. Chen, and B. Liu, "Admission control for wireless adaptive HTTP streaming: An evidence theory based approach," in *ACM International Conference on Multimedia*, Orlando, USA, Nov. 2014, pp. 893–896.
- [13] *Progressive download and dynamic adaptive streaming over http*, Std. 3GPP TS 26.247 V12.1.0, 2013, Available: <http://goo.gl/4EJbvd>.
- [14] X. Wang and S. Zhang, "Contribution to DASH core experiment on parameter insertion in url," MPEG m28174, Jan. 2013.
- [15] H. S. Wang and N. Moayeri, "Finite-state markov channel-a useful model for radio communication channels," *IEEE Trans. Veh. Technol.*, vol. 44, no. 1, pp. 163–583, 1995.
- [16] R. S. Sutton and A. G. Barto, *Reinforcement Learning: An Introduction*, 1st ed. MIT Press, 1998.
- [17] A. Balachandran, V. Sekar, A. Akella, S. Seshan, I. Stoica, and H. Zhang, "Developing a predictive model of quality of experience for internet video," in *ACM SIGCOMM*, Aug. 2013, pp. 339–350.