Context Aware Energy Efficient Optimization for Video On-demand Service over Wireless Networks

Changyang She and Chenyang Yang
School of Electronics and Information Engineering, Beihang University, Beijing, China
Email: {cyshe,cyyang}@buaa.edu.cn

***Abstract*—Video on-demand (VOD) service is widely requested, and is becoming a dominant application in wireless networks, where energy efficiency (EE) is a major design goal. Reducing the energy consumption of VOD service with given quality of service (QoS) requirement can improve the EE of wireless systems. Recent finding shows that user mobility is highly predictable, and hence future location information is possible to know. In this paper, we study EE-optimal transmit power and bandwidth allocation for VOD service in orthogonal frequency division multiplexing (OFDM) system by exploiting context information. We consider two kinds of context information, i.e., the predictive average channel gains and the QoS of VOD service. The optimal resource allocation policy that minimizes the average energy consumption for VOD service is proposed. At the beginning of the service, the average transmit power and number of used subcarriers are allocated with future average channel gains. During the procedure of the service, the instantaneous transmit power is allocated to different subcarriers after the instantaneous channel gains becomes available. Simulation results show that the energy consumed for VOD service with the proposed policy is around half of that with an existing policy.**

I. Introduction

Global mobile data traffic has grown dramatically over the past years, where a large portion of the data is generated by mobile video. As shown in [1], 55 % of the overall data traffic is mobile video in 2014, and the percentage is predicted to become 72 % by 2019. Moreover, the video on-demand (VOD) service is widely provided by content providers like Netflix and YouTube. This indicates that VOD service has a significant contribution to the increasing mobile data traffic.

On the other hand, improving energy efficiency (EE) is an important design goal for fifth generation (5G) cellular networks [2]. However, the requirements of the VOD service have not been well addressed in the EE design.

Inspired by the recent finding in [3] that user mobility is highly predictable, improving EE by exploiting context information has drawn significant attention as the smart phone popularizes. Context information can be classified into application (e.g., quality of service (QoS)), network (e.g., congestion status), user (e.g., user location), and device levels [4]. In prior works with the predicted user location [5, 6], the data rate is adjusted by adapting the service time. In orthogonal frequency division multiplexing (OFDM) systems that are prevalent in existing and future cellular networks, it is possible to adjust the number of active subcarriers and transmit power on different subcarriers to further improve the EE.

In this paper, we study EE-optimal resource allocation for OFDM systems by exploiting user- and application-level context information, i.e., the future average channel gains and QoS requirement of VOD. Different from [6] that is with an implicit assumption of perfectly known future instantaneously channel gains, we assume that only the future average channel gains are available. We optimize the resource allocation for VOD service that maximizes the EE of a multi-cell OFDM network, which includes a pre-resource allocation and a dynamic power allocation. At the beginning of the VOD service, the *pre-resource allocation* finds the average transmit power and the number of used subcarriers with predictive average channel gains. During the transmission procedure for the VoD service, the *dynamic power allocation* allocates the transmit power among subcarriers based on the instantaneous channel gains. Simulation results show that with the optimal resource allocation policy the energy consumed by the VOD service is around half of that when the policy in [6] is applied.

II. System Model

We consider the scenario that one mobile user with VOD service travels across multi-cells, and the user is served by the nearest BS. Suppose that the video file is encoded into multiple layers with scalable video coding (SVC), and is transmitted via HTTP-based protocols [7]. The SVC-encoded HTTP adaptive video is divided into several segments, and each segment is encoded into one base layer and several enhance layers. If the data in the base layer of each segment is conveyed to the user before deplaying, then no playback interruption will occur. Data loss in enhance layers only leads to the video quality deterioration. Since the transportation in core network is not a bottleneck, we assume that the requested data is available at each BS when the user is accessed to the BS.

A. Transmission and Channel Models

Since there may exist some real time services (e.g. voice) in the network, the BSs can only employ the remaining resources to serve the user of VOD. Consider an OFDM system, where $P^{\max}$ and $K^{\max}$ denote the transmit power and number of subcarriers available for the VOD service, respectively.

Suppose that the video file is divided into N_v segments with equal playback time $L_v\Delta T$, where N_v and L_v are integers. The overall video playback time can be divided into N_L frames each with duration ΔT, and each frame is divided into N_S time slots each with duration $\tau \ll \Delta T$. Consider frequency-selective block Rayleigh fading channel. Assume that the average channel gain is constant within duration ΔT and changes from one frame to another, and the instantaneous channel gain is constant within duration τ and is independently and identically distributed (i.i.d.) among time slots and among subcarriers.

Denote $\alpha_i g_{ij}^k$ as the composite channel gain on the kth subcarrier in the jth time slot of the ith frame, where α_i and g_{ij}^k are the average and instantaneous channel gains, respectively. For the considered Rayleigh fading channel, g_{ij}^k is exponentially distributed with mean of one.

With the predicted user location [3] and radio map, the average channel gains $\alpha_i, i = 1, ..., N_L$ are predictable. We assume that they are perfectly known by the 1st BS the user accessed to before the video transmission starts. The instantaneous channel gain g_{ij}^k is assumed to be perfectly known at the user and the BS it is accessed to at the beginning of the jth time slot of the ith frame. With the capacity achieving coding, the instantaneous transmission/service rate in the jth time slot of the ith frame can be expressed as,

$$s_{ij} = B \sum_{k=1}^{K_i} \log_2 \left(1 + \frac{\alpha_i}{\sigma_0^2} p_{ij}^k g_{ij}^k \right) \quad \text{bits/s}, \tag{1}$$

where B is the subcarrier spacing, σ_0^2 is the variance of Gaussian noise, K_i is the number of used subcarriers in the ith frame, and p_{ij}^k is the transmit power on the kth subcarrier in the jth time slot of the ith frame.

B. Queueing Model for VOD Service

Denote the amount of data transmitted to the buffer at the user and displayed in the ith frame as S_i and R_i, respectively. Then, $S_i = \tau \sum_{j=1}^{N_S} s_{ij}$. When a certain quality level of the video is chosen by the user, R_i, $i = 1, ..., N_L$, are given. Assume that the buffer size is larger than the size of the video, which is reasonable for smart phones since storage devices are cheap nowadays. Hence, we do not need to consider the buffer overflow probability.

To guarantee the requested video quality, the video segments need to be transmitted to the buffer at the user before they are displayed. Then, the QoS requirement of the VoD service can be expressed as the following constraints [6],

$$Q_0 + \sum_{n=0}^{l-1} \sum_{i=nL_v+1}^{(n+1)L_v} S_i \geq \sum_{n=0}^{l} \sum_{i=nL_v+1}^{(n+1)L_v} R_i, l = 1, ..., N_v - 1,$$

where $Q_0 = \sum_{i=1}^{L_v} R_i$ is the initial queue length, which reflects the fact that the video playback starts after the user has received the first video segment.

For notational simplicity but without loss of generality, we set $L_v = 1$ in the following. Then, $N_v = N_L$.

In wireless channels, such QoS constraints are hard to satisfy with probability one due to channel fading. To overcome this difficulty, alternative constraints that are less stringent are considered in the sequel. Denote the average service rate in the ith frame as $\bar{s}_i$, which can be obtained from (1) as,

$$\bar{s}_i = B \sum_{k=1}^{K_i} \mathbb{E} \left[\log_2 \left(1 + \frac{\alpha_i}{\sigma_0^2} p_{ij}^k g_{ij}^k \right) \right] \quad \text{bits/s}, \tag{2}$$

Since $S_i = \tau \sum_{j=1}^{N_S} s_{ij}$, we have $\frac{1}{\Delta T} S_i = \frac{1}{N_S} \sum_{j=1}^{N_S} s_{ij}$. When $N_S \to \infty$, $\frac{1}{N_S} \sum_{j=1}^{N_S} s_{ij} \to \bar{s}_i$. then the QoS constraints are equivalent to

$$\sum_{i=1}^{l} \bar{s}_i \geq \frac{1}{\Delta T} \sum_{i=2}^{l+1} R_i, l = 1, ..., N_L - 1. \tag{3}$$

In practice, the number of time slots in each frame, N_S, is large but finite. Then, some data in each segment may not be conveyed by a transmit policy subject to constraint (3) before displaying. Nonetheless, with the SVC-based HTTP adaptive video, the loss of a small amount of data in the enhance layers does not lead to severe deterioration of user experience.

C. Power Model and EE definition

The total energy consumed by the BSs during the ith frame can be modeled as [8]

$$E_i = \frac{1}{\rho} \sum_{j=1}^{N_S} \sum_{k=1}^{K_i} \tau p_{ij}^k + \Delta T P_c K_i + \Delta T P^{rt}, \tag{4}$$

where $\rho \in (0, 1]$ is the power amplifier efficiency, P_c is the circuit power consumed for baseband processing on each subcarrier, P^{rt} is the power consumption for serving other real time services.

With the average channel gains available, the EE of the network can be defined as the ratio of the average data transmitted to the average total energy consumed at the BSs during the overall video playback duration, i.e.,

$$EE \triangleq \frac{\mathbb{E}_g \left(\sum_{i=1}^{N_L-1} \tau \sum_{j=1}^{N_S} s_{ij} + S^{rt} \right)}{\mathbb{E}_g \left(\frac{1}{\rho} \sum_{i=1}^{N_L-1} \sum_{j=1}^{N_S} \sum_{k=1}^{K_i} \tau p_{ij}^k + \Delta T P_c \sum_{i=1}^{N_L-1} K_i + T P^{rt} \right)},$$

where S^{rt} and P^{rt} are the amount of data transmitted and the transmit power for the real time services, respectively. For VOD service, the amount of transmitted data equals to the amount of data that is required to transmit. Thus, $\mathbb{E}_g \left(\sum_{i=1}^{N_L-1} \tau \sum_{j=1}^{N_S} s_{ij} \right) = \sum_{i=1}^{N_L-1} \Delta T \bar{s}_i = \sum_{i=2}^{N_L} R_i$, which is given when the video is chosen by the user at a certain quality level. Since we are only interested in resource allocation for the VOD service, maximizing the EE is equivalent to minimizing the average energy consumed for the VOD service, i.e.,

$$\mathbb{E}_g \left(\frac{1}{\rho} \sum_{i=1}^{N_L-1} \sum_{j=1}^{N_S} \sum_{k=1}^{K_i} \tau p_{ij}^k + \Delta T P_c \sum_{i=1}^{N_L-1} K_i \right) \tag{5}$$

III. Energy Efficient Resource Allocation Policy

A. Problem Formulation

Since g_{ij}^k is not predictable, we cannot find p_{ij}^k to minimize (5) at the beginning of the VOD service. However, if the distribution of g_{ij}^k is known (e.g., Rayleigh fading as we assumed), then (5) can be minimized by optimizing the average transmit power and bandwidth allocated to the VOD service in the N_L-1 frames, i.e., $\bar{P}_i \triangleq \mathbb{E}_g\left(\sum_{k=1}^{K_i} p_{ij}^k\right)$ and K_i, $i=1,...,N_L-1$. With the assumption that g_{ij}^k is i.i.d. among slots, we have $\mathbb{E}_g\left(\tau\sum_{j=1}^{N_S}\sum_{k=1}^{K_i} p_{ij}^k\right) = \Delta T\bar{P}_i$. Then, (5) can be rewritten as $\Delta T\sum_{i=1}^{N_L-1}\left(\frac{1}{\rho}\bar{P}_i + P_c K_i\right)$.

With $\alpha_i, i=1,...,N_L-1$, at the beginning of the VOD service, the first BS accessed by the user can find $\bar{P}_i$ and K_i to minimize the average energy consumption. Then, the BS sends the policy to the subsequent BSs. We refer to $\{\bar{P}_i, K_i, i=1,...,N_L-1\}$ as the *pre-resource allocation policy*.

After g_{ij}^k, $k=1,...,K_i$ are available at a BS accessed by the user in the jth time slot of the ith frame, the BS allocates p_{ij}^k with given $\bar{P}_i$ and K_i to satisfy the QoS constraints. We refer to $p_{ij}^k = p(\bar{P}_i, K_i, g_{ij}^k), k=1,...K_i, j=1,...,N_S$ as the *dynamic power allocation policy* in the ith frame.

The optimal average transmit power and number of used subcarriers in different frames that minimize the average energy consumption under the QoS constraints in (3) can be found from the following problem,

$$\min_{K_i,\bar{P}_i} E_{\text{ave}} \triangleq \Delta T\sum_{i=1}^{N_L-1}\left(\frac{1}{\rho}\bar{P}_i + P_c K_i\right) \tag{6}$$

$$\text{s.t.} \sum_{i=1}^{l}\bar{s}_i \geq \frac{1}{\Delta T}\sum_{i=2}^{l+1} R_i, \tag{6a}$$

$$\bar{P}_i \leq P^{\max} \text{ and } K_i \leq K^{\max}, \tag{6b}$$

where $l, i = 1,...,N_L-1$.

For Rayleigh fading channel we conidered, the average service rate in (2) can be expressed as $\bar{s}_i = K_i\int_0^\infty B\log_2\left[1+\frac{\alpha_i}{\sigma_0^2}p\left(\bar{P}_i, K_i, g\right)g\right]e^{-g}dg$, which depends on the form of the function $p(\bar{P}_i, K_i, g)$, where g is the value of the instantaneous channel gain on a subcarrier in a slot. This indicates that dynamic power allocation policy affects the constraints in (6a), and hence affects the optimal solution of problem (6). Therefore, the optimal value of the objective function in (6) is a function of $p(\bar{P}_i, K_i, g)$. We denote it as $E^*_{\text{ave}}\left[p(\bar{P}_i, K_i, g)\right]$.

Then, the optimal dynamic power allocation policy can be obtained by minimizing $E^*_{\text{ave}}\left[p(\bar{P}_i, K_i, g)\right]$, denoted as $p^*(\bar{P}_i, K_i, g)$, with which we can find the optimal pre-resource allocation policy, $\{\bar{P}_i^*, K_i^*, i=1,...,N_L-1\}$.

B. Optimal Dynamic Power Allocation Policy

It is very hard to find the optimal dynamic power allocation policy that minimizes $E^*_{\text{ave}}\left[p(\bar{P}_i, K_i, g)\right]$, because the expression of $E^*_{\text{ave}}\left[p(\bar{P}_i, K_i, g)\right]$ can not be obtained. In what follows, we use an alternative approach to find $p^*(\bar{P}_i, K_i, g)$.

Inspired by the fact that a policy that maximizes $\bar{s}_i$ with given $\bar{P}_i$ and K_i can minimize the $\bar{P}_i$ with given $\bar{s}_i$ and K_i, we first find the power allocation policy that maximizes $\bar{s}_i$ with given $\bar{P}_i$ and K_i. As shown in [9], such a policy is water-filling,

$$p^w\left(\frac{\bar{P}_i}{K_i}, g\right) = \begin{cases} \frac{\sigma_0^2}{\alpha_i}\left(\frac{1}{g_i^{\text{th}}} - \frac{1}{g}\right), g \geq g_i^{\text{th}}, \\ 0, \quad\quad g < g_i^{\text{th}}. \end{cases} \tag{7}$$

In Rayleigh fading channels, the water level g_i^{th} can be obtained from

$$\int_{g_i^{\text{th}}}^{\infty}\frac{\sigma_0^2}{\alpha_i}\left(\frac{1}{g_i^{\text{th}}} - \frac{1}{g}\right)e^{-g}dg = \frac{\bar{P}_i}{K_i}. \tag{8}$$

Then, when g_{ij}^k, $k=1,...K_i$, are known at the BS in the jth time slot of the ith frame, $p_{ij}^k = p^w(\frac{\bar{P}_i}{K_i}, g_{ij}^k)$.

Now we show that $p^*(\bar{P}_i, K_i, g) = p^w\left(\frac{\bar{P}_i}{K_i}, g\right)$, i.e., (7) is the optimal dynamic power allocation policy.

Proposition 1. Consider an arbitrary dynamic power allocation policy $p'(\bar{P}_i, K_i, g)$ that is different from $p^w\left(\frac{\bar{P}_i}{K_i}, g\right)$. The optimal solutions of problem (6) with policies $p^w\left(\frac{\bar{P}_i}{K_i}, g\right)$ and $p'(\bar{P}_i, K_i, g)$ are denoted as $\{\bar{P}_i^w, K_i^w, i=1,...,N_L-1\}$ and $\{\bar{P}'_i, K'_i, i=1,...,N_L-1\}$, respectively. Then,

$$E^*_{\text{ave}}\left[p(\bar{P}_i^w, K_i^w, g)\right] \leq E^*_{\text{ave}}\left[p(\bar{P}'_i, K'_i, g)\right]. \tag{9}$$

Proof: See Appendix A. ■

C. Optimal Pre-resource Allocation Policy

In the sequel, we propose a two-step method to find the optimal solution of problem (6).

In the first step, we derive the minimum average power consumption to support the average service rate $\bar{s}_i$, and denote it as $P_i^{vod*}(\bar{s}_i)$, which is referred to as *optimal power-rate relation*. The optimal power-rate relation in each frame, say the ith frame, can be obtained from the following of problem,

$$\min_{\bar{P}_i, K_i} P_i^{vod}(\bar{s}_i) = \frac{1}{\rho}\bar{P}_i + P_c K_i \tag{10}$$

$$\text{s.t. } K_i\int_0^\infty B\log_2\left[1+\frac{\alpha_i}{\sigma_0^2}p^w\left(\frac{\bar{P}_i}{K_i}, g\right)g\right]e^{-g}dg = \bar{s}_i, \tag{10a}$$

$$\bar{P}_i \leq P^{\max} \text{ and } K_i \leq K^{\max}. \tag{10b}$$

The optimal solution is denoted as $\{\bar{P}_i^{opt}(\bar{s}_i),\ K_i^{opt}(\bar{s}_i)\}$. Then, $P_i^{vod*}(\bar{s}_i) = \frac{1}{\rho}\bar{P}_i^{opt}(\bar{s}_i) + P_c K_i^{opt}(\bar{s}_i)$

In the second step, we find the optimal average service rate in all frames by solving the following problem,

$$\min_{\bar{s}_i, i=1,...,N_L-1}\sum_{i=1}^{N_L-1} P_i^{vod*}(\bar{s}_i) \tag{11}$$

$$\text{s.t. } \sum_{i=1}^{l}\bar{s}_i \geq \frac{1}{\Delta T}\sum_{i=2}^{l+1} R_i, l=1,...,N_L-1, \tag{11a}$$

$$0 \le \bar{s}_i \le s_i^{\max}, \tag{11b}$$

where $s_i^{\max}$ is the maximum average service rate the system can support in the ith frame, which can be expressed as

$$s_i^{\max} = K^{\max}\mathbb{E}_g\left\{B\log_2\left[1+\frac{\alpha_i}{\sigma_0^2}p^w\left(\frac{P^{\max}}{K^{\max}},g\right)g\right]\right\}. \tag{12}$$

Denote $\{\bar{s}_i^*, i = 1, ..., N_L - 1\}$ as the optimal solution of problem (11). The following proposition indicates that $\{\bar{P}_i^{opt}(\bar{s}_i^*), K_i^{opt}(\bar{s}_i^*), i = 1, ..., N_L-1\}$ is the optimal solution of problem (6).

Proposition 2. Denote $\{\bar{P}_i'(\bar{s}_i'), K_i'(\bar{s}_i'), i = 1, ..., N_L - 1\}$ as an arbitrary feasible solution of problem (6), where $\bar{s}_i', i = 1, ..., N_L - 1$ are the related average service rates in different frames. Then,

$$\sum_{i=1}^{N_L-1} P_i^{vod*}(\bar{s}_i^*) \le \sum_{i=1}^{N_L-1}\left[\frac{1}{\rho}\bar{P}_i'(\bar{s}_i') + P_c K_i'(\bar{s}_i')\right].$$

Proof: The proof of this proposition is similar to that of Proposition 1, and is omitted due to the lack of space. ■

In what follows, we respectively find the solutions of problems (10) and (11).

1) Optimal Power-Rate Relation: Under constraints (10a), minimizing the objective function in (10) is equivalent to maximizing the ratio of (10a) to (10), i.e.,

$$\frac{\int_0^\infty B\log_2\left[1+\frac{\alpha_i}{\sigma_0^2}p^w\left(\frac{\bar{P}_i}{K_i},g\right)g\right]e^{-g}dg}{\frac{1}{\rho}\frac{\bar{P}_i}{K_i}+P_c} \triangleq \frac{F_i(\bar{P}_i^s)}{\frac{1}{\rho}\bar{P}_i^s+P_c}, \tag{13}$$

which is the ratio of average service rate per subcarrier to the average power consumption per subcarrier, where $\bar{P}_i^s \triangleq \frac{\bar{P}_i}{K_i}$ is the average transmit power per subcarrier.

Proposition 3. $F_i(\bar{P}_i^s)$ is strictly concave in $\bar{P}_i^s$.

Proof: See Appendix B. ■

Proposition (3) indicates that (13) is strictly quasiconcave [10]. Therefore, the maximum of (13) is unique. We denote the optimal average transmit power per subcarrier as $\bar{P}_i^{s^*}$.

Further considering the maximal resource constraints in (10b), the optimal power-rate relation $P_i^{vod*}(\bar{s}_i)$ has following two properties. The proofs are omitted due to the lack of space.

Property 1. If the required value of $\bar{s}_i$ is low such that

$$\bar{s}_i < s_i^{\text{th}} \triangleq F_i\left(\bar{P}_i^{s^*}\right)\min\left\{\frac{P^{\max}}{\bar{P}_i^{s^*}}, K^{\max}\right\}, \tag{14}$$

then $P_i^{vod*}(\bar{s}_i)$ is linear.

Property 2. If $\bar{s}_i \in [s_i^{\text{th}}, s_i^{\max}]$, then $P_i^{vod*}(\bar{s}_i)$ is strictly convex.

If $\bar{s}_i < s_i^{\text{th}}$, then the average service rate $\bar{s}_i$ can be achieved when the maximal resource constraints in (10b) are inactive, i.e., $\bar{P}_i < P^{\max}$ and $K_i < K^{\max}$. If $\bar{s}_i \in [s_i^{\text{th}}, s_i^{\max}]$, then at least one of the maximal resource constraints in (10b) is active to achieve the required average service rate $\bar{s}_i$.

The properties suggest that $P_i^{vod*}(\bar{s}_i)$ first linearly increases with $\bar{s}_i$, and then becomes strictly convex in $\bar{s}_i$. Therefore, $P_i^{vod*}(\bar{s}_i)$ is convex in $\bar{s}_i, \forall \bar{s}_i \in [0, s_i^{\max}]$.

2) Optimal Average Service Rate Allocation: Since $P_i^{vod*}(\bar{s}_i)$ is convex in $\bar{s}_i$, the objective function in (11) is convex in $\bar{s}_i$, $i = 1, ..., N_L - 1$. With linear constraints, problem (11) is a convex programming, which can be solved numerically by techniques such as the interior-point method.

IV. Numerical and Simulation Results

In this section, we evaluate the proposed policy by comparing with an existing policy proposed in [6], which also considers VOD and exploits the context information.

A. The Policy for Comparison

The policy in [6] does not include dynamic power allocation, since the channel was assumed as flat fading and constant within the overall video playback time, i.e., $g_{ij}^k = 1, \forall i, j, k$. Moreover, the maximal bandwidth and transmit power were applied to serve the user of VOD, and the service rate is adjusted by adapting the service time in each frame. After regular manipulations, we can show that the power-rate relation achieved by this policy is always linear, i.e.,

$$\tilde{P}_i^{\text{lin}}(\bar{s}_i) = \frac{P^{\max}/\rho + P_c K^{\max}}{\tilde{s}_i^{\max}}\bar{s}_i, \forall \bar{s}_i \in [0, \tilde{s}_i^{\max}], \tag{15}$$

where $\tilde{s}_i^{\max} = K^{\max}B\log_2\left(1+\frac{\alpha_i P^{\max}}{\sigma_0^2 K^{\max}}\right)$.

In the scenario where real-time service exists in the network and hence the BSs cannot turn into idle mode, the rate allocation policy in [6] in fact minimizes $\sum_{i=1}^{N_L-1}\tilde{P}_i^{\text{lin}}(\bar{s}_i)$ under constraints (11a) and $0 \le \bar{s}_i \le \tilde{s}_i^{\max}$.

To compare with our policy fairly over the frequency-selective block fading channel, we extend the policy in [6] by incorporating the water-filling power allocation with the rate allocation in [6]. To be specific, if the user is served in a certain time slot, the BS will allocate transmit power according to (7) over $K^{\max}$ subcarriers given the maximal transmit power $P_T^{\max}$. Then, the achieved power-rate relation becomes

$$P_i^{\text{lin}}(\bar{s}_i) = \frac{P^{\max}/\rho + P_c K^{\max}}{s_i^{\max}}\bar{s}_i, \forall \bar{s}_i \in [0, s_i^{\max}], \tag{16}$$

where $s_i^{\max}$ is shown in (12). Then, the extended policy in fact minimizes $\sum_{i=1}^{N_L-1}P_i^{\text{lin}}(\bar{s}_i)$ under constraints (11a) and (11b), which is denoted as "Existing policy" in the legend.

B. Performance Evaluation

We use the first two minutes of a six-layer video in [11] to evaluate the performance of different policies. The sizes of video segments of different layers can be obtained from "Sony_G16B15_CIF_DQP6_5EL_48_42_36_30_24_18/" in [12].

The considered scenario is shown in Fig. 1. The user travels from the start point of $(0, 0)$ m to the end point of $(2400, 0)$ m

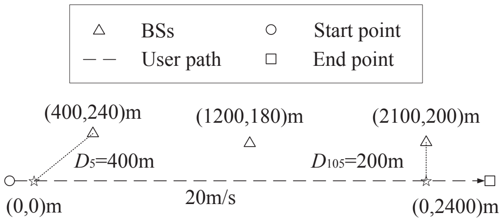

Fig. 1. Scenario for numerical and simulation results.

with velocity 72 km/h (i.e., 20 m/s). Assume that the video playback starts when the user is located at $(0,0)$ m, where the first video segment has been conveyed to the buffer of the user. During the overall video playback time of 120 s, the user is respectively served by each of the three BSs, and the user always accesses to the nearest BS. The path loss model is $35.3 + 37.6\log_{10} D_i$ dB, where D_i is the BS-user distance in the ith frame. We take $D_5 = 200$ m and $D_{105} = 400$ m as examples to show the power-rate relations with different distances. Other parameters are listed in Table II, where the transmit power and number of subcarriers available for the VOD service are set as $\frac{1}{8}$ of those for a macro BS [13].

TABLE I
LIST OF PARAMETERS [8, 13]

Available transmit power $P^{\max}$	5.0 W
Available number of subcarriers $K^{\max}$	128
Subcarrier space B	15 kHz
Power amplifier efficiency ρ	38.8 %
Circuit power consumption for one subcarrier P_c	0.02 W
Channel gain-to-noise ratio when $D_i = 500$ m	20 dB
Duration of each frame ΔT	1 s
Duration of each time slot τ	5 ms

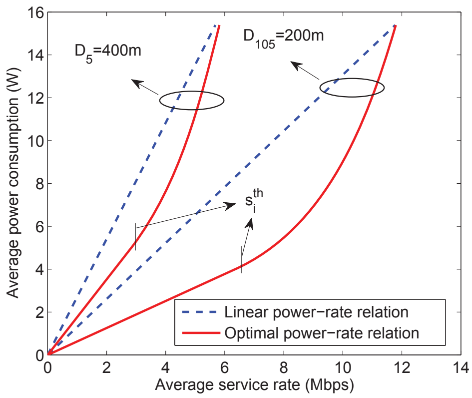

Fig. 2. Power-rate relations with different user-BS distances.

Figure 2 illustrates the optimal power-rate relation $P_i^{vod*}(\bar{s}_i)$ and the linear power-rate relation in (16). The results show that the minimal average power consumption required by our policy to support a given average service rate is lower than that required by the extended policy in [6].

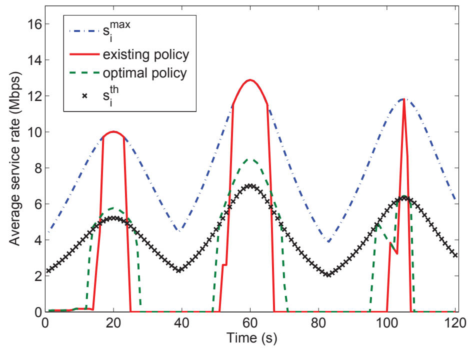

Fig. 3. Average service rate allocated by different policies.

Figure 3 shows the impact of power-rate relation on the rate allocation policy. The average service rate of "Existing policy" and our policy (with legend "Optimal policy") are respectively obtained with the linear power-rate relation in (16) and the optimal power-rate relation $P_i^{vod*}(\bar{s}_i)$. With "Existing policy", the BSs serve the user when it is close to one of the BSs (i.e., when the maximal service rate $s_i^{\max}$ in (12) is high). Besides, the service rates equal to $s_i^{\max}$ in most of the active frames with $\bar{s}_i > 0$. By contrast, with "Optimal policy", the BSs do not serve the user with the maximal service rate $s_i^{\max}$.

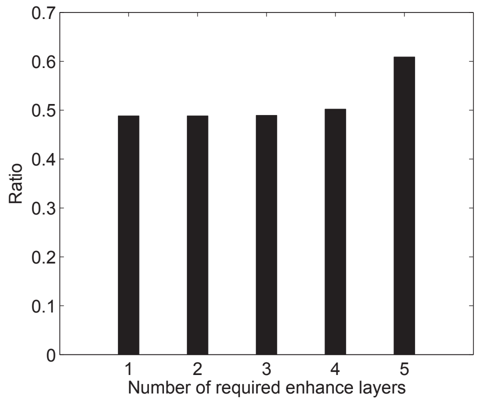

Fig. 4. Ratio of energy consumption of the proposed policy to that of the extended policy in [6].

Figure 4 shows the energy saving ratio of the proposed policy with respective to the extended existing policy with different video qualities. The results show that the energy consumed by the proposed policy is nearly halved compared to the energy consumed by the existing policy.

V. CONCLUSION

In this paper, we optimized resource allocation to improve the EE of an OFDM system serving VOD user by exploiting user-level and application-level context information. The optimal resource allocation policy that jointly allocates transmit power and the number of subcarrier was proposed, which consists of a pre-resource allocation and a dynamic power allocation. At the beginning of the service, the predictive average channel gains are employed to find the average transmit power and the number of subcarriers in each frame. During the video transmission, the transmit power is allocated to different subcarriers after the instantaneous channel gains become available. Simulation results showed that about half of the energy consumed for the VOD service can be saved by the proposed policy compared to the existing policy.

Appendix A
Proof of Proposition 1

Proof: Denote the average service rates achieved by the dynamic power allocation policy $p'(\bar{P}_i, K_i, g)$ with pre-resource allocation policy $\{\bar{P}_i', K_i', i = 1, ..., N_L - 1\}$ as $\bar{s}_i'$, $i = 1, ..., N_L - 1$.

To prove the proposition, we need the following property: the water-filling policy can minimize $\bar{P}_i$ with given K_i and average service rate $\bar{s}_i$ [9]. According to the property, given K_i' and $\bar{s}_i'$, $i = 1, ..., N_L - 1$, the average transmit power is minimized with $p^w\left(\frac{\bar{P}_i}{K_i}, g\right)$. Denote the related minimal average transmit power in the ith frame as $\bar{P}_i^{\min}$. Then, $\bar{P}_i^{\min} \leq \bar{P}_i'$, $i = 1, ..., N_L - 1$. Hence

$$\sum_{i=1}^{N_L-1}\left(\frac{1}{\rho}\bar{P}_i^{\min} + P_c K_i'\right) \leq \sum_{i=1}^{N_L-1}\left(\frac{1}{\rho}\bar{P}_i' + P_c K_i'\right). \quad \text{(A.1)}$$

Moreover, with $p^w\left(\frac{\bar{P}_i}{K_i}, g\right)$, the optimal pre-resource allocation is $\{\bar{P}_i^w, K_i^w, i = 1, ..., N_L\}$. Thus,

$$\sum_{i=1}^{N_L-1}\left(\frac{1}{\rho}\bar{P}_i^{w} + P_c K_i^w\right) \leq \sum_{i=1}^{N_L-1}\left(\frac{1}{\rho}\bar{P}_i^{\min} + P_c K_i'\right). \quad \text{(A.2)}$$

From (A.1) and (A.2), we have (9). The proof is complete. ∎

Appendix B
Proof of Proposition 3

Proof: Substituting (7) into $F_i\left(\bar{P}_i^s\right)$ defined in (13), we can obtain $F_i(\bar{P}_i^s)$ as follows,

$$F_i\left(\bar{P}_i^s\right) = \frac{B}{\ln 2}\left[\int_{g_i^{\mathrm{th}}}^{\infty} \ln(g)\, e^{-g} dg - \ln\left(g_i^{\mathrm{th}}\right) e^{-g_i^{\mathrm{th}}}\right], \quad \text{(B.1)}$$

where the relation between g_i^{th} and $\bar{P}_i^s$ can be obtained from (8). Then, $F_i\left(\bar{P}_i^s\right)$ is a composition function, $F_i\left[g_i^{\mathrm{th}}\left(\bar{P}_i^s\right)\right]$, whose second order derivative is

$$\frac{d^2 F_i\left[g_i^{\mathrm{th}}\left(\bar{P}_i^s\right)\right]}{d(\bar{P}_i^s)^2} = \frac{d^2 F_i}{d\left(g_i^{\mathrm{th}}\right)^2}\left(\frac{d g_i^{\mathrm{th}}}{d\bar{P}_i^s}\right)^2 + \frac{dF_i}{dg_i^{\mathrm{th}}}\frac{d^2 g_i^{\mathrm{th}}}{d(\bar{P}_i^s)^2}. \quad \text{(B.2)}$$

To prove that $F_i\left[g_i^{\mathrm{th}}\left(\bar{P}_i^s\right)\right]$ is concave in $\bar{P}_i^s$, we only need to prove that (B.2) is negative. From (B.1) we can derive that

$$\frac{dF_i}{dg_i^{th}} = -\frac{B}{\ln 2}\frac{1}{g_i^{th}}e^{-g_i^{th}}, \quad \text{(B.3)}$$

$$\frac{d^2 F_i}{d\left(g_i^{th}\right)^2} = \frac{B}{\ln 2}\left(\frac{1}{g_i^{th}}\right)^2 e^{-g_i^{th}} + \frac{B}{\ln 2}\frac{1}{g_i^{th}}e^{-g_i^{th}}. \quad \text{(B.4)}$$

From (8), we can find the relationship between $\bar{P}_i^s$ and g_i^{th} as $\bar{P}_i^s = \frac{\sigma_0^2}{\alpha_i}\left(\frac{1}{g_i^{\mathrm{th}}}e^{-g_i^{\mathrm{th}}} - \int_{g_i^{\mathrm{th}}}^{\infty}\frac{1}{g}e^{-g}dg\right)$. Then, we can derive that $\frac{d\bar{P}_i^s}{dg_i^{\mathrm{th}}} = -\frac{\sigma_0^2}{\alpha_i}\left(\frac{1}{g_i^{\mathrm{th}}}\right)^2 e^{-g_i^{\mathrm{th}}}$. According to the property of inverse function, i.e., $\frac{d\bar{P}_i^s}{dg_i^{\mathrm{th}}}\frac{dg_i^{\mathrm{th}}}{d\bar{P}_i^s} = 1$ at any point $(g_i^{\mathrm{th}}, \bar{P}_i^s)$, we have

$$\frac{dg_i^{\mathrm{th}}}{d\bar{P}_i^s} = -\frac{\alpha_i}{\sigma_0^2}\left(g_i^{\mathrm{th}}\right)^2 e^{g_i^{\mathrm{th}}}. \quad \text{(B.5)}$$

Further considering that $\frac{d^2 g_i^{\mathrm{th}}}{d(\bar{P}_i^s)^2} = \frac{d\left(\frac{dg_i^{\mathrm{th}}}{d\bar{P}_i^s}\right)}{dg_i^{\mathrm{th}}}\frac{dg_i^{\mathrm{th}}}{d\bar{P}_i^s}$, we have

$$\frac{d^2 g_i^{\mathrm{th}}}{d\left(\bar{P}_i^s\right)^2} = \left[2\left(g_i^{\mathrm{th}}\right)^3 + \left(g_i^{\mathrm{th}}\right)^4\right]\left(\frac{\alpha_i}{\sigma_0^2}\right)^2 e^{2g_i^{\mathrm{th}}}. \quad \text{(B.6)}$$

Substituting (B.3), (B.4), (B.5) and (B.6) into (B.2), we can finally derive that

$$\frac{d^2 F_i\left[g_i^{\mathrm{th}}\left(\bar{P}_i^s\right)\right]}{d\left(\bar{P}_i^s\right)^2} = -\frac{B}{\ln 2}\left(\frac{\alpha_i}{\sigma_0^2}\right)^2\left(g_i^{\mathrm{th}}\right)^2 e^{g_i^{\mathrm{th}}} < 0.$$

This completes the proof. ∎

References

[1] Cisco, "Cisco visual networking index: Global mobile data traffic forecast update, 2014-2019," *Cisco white paper*, 2015.

[2] G. Wu, C. Yang, S. Li, and G. Li, "Recent advance in energy-efficient networks and its application in 5G systems," *IEEE Wireless Commun. Mag.*, vol. 22, no. 2, pp. 145 – 151, Apr. 2015.

[3] C. Song, Z. Qu, N. Blumm, and A.-L. Barabasi, "Limits of predictability in human mobility," *Science*, vol. 327, no. 5968, pp. 1018 – 1021, Feb. 2010.

[4] C. Park, Y. Seo, K. Park, and Y. Lee, "The concept and realization of context-based content delivery of NGSON," *IEEE Commun. Mag.*, vol. 50, no. 1, pp. 74–81, Jan. 2012.

[5] M. Draxler, P. Dreimann, and H. Karl, "Anticipatory power cycling of mobile network equipment for high demand multimedia traffic," in *Proc. IEEE Online GreenComm*, Nov. 2014.

[6] H. Abou-zeid, H. S. Hassanein, and S. Valentin, "Energy-efficient adaptive video transmission: Exploiting rate predictions in wireless networks," *IEEE Trans. Veh. Technol.*, vol. 63, no. 5, pp. 2013 – 2026, Jun. 2014.

[7] Michael Seufert, *et al.*, "A survey on quality of experience of HTTP adaptive streaming," *IEEE Commun. Surveys Tuts.*, vol. 17, no. 1, pp. 469 – 492, 2015.

[8] Claude Desset, *et al.*, "Flexible power modeling of LTE base stations," in *Proc. IEEE WCNC*, Apr. 2012.

[9] A. Goldsmith, *Wireless Communications*. Cambridge University Press, 2005.

[10] S. Boyd and L. Vandanberghe, *Convex Optimization*. Cambridge Univ. Press, 2004.

[11] P. Seeling and M. Reisslein, "Video transport evaluation with H.264 video traces," *IEEE Commun. Surveys Tuts.*, vol. 14, no. 4, pp. 1142 – 1165, 2012.

[12] Video Trace Library. [Online]. Available: http://trace.eas.asu.edu/videotraces2/cgs/cif/

[13] G. Auer, O. Blume, V. Giannini, I. Gódor, *et al.*, "D 2.3: Energy efficiency analysis of the reference systems, areas of improvements and target breakdown," *EARTH*, Jan. 2012. [Online]. Available: https://www.ict-earth.eu/publications/deliverables/deliverables.html