Energy-Budget-Compliant Cloud Video Delivery to Mobile Devices

Mohammad Hosseini  
Department of Computer Science  
University of Illinois at Urbana-Champaign  
Email: shossen2@illinois.edu

Anduo Wang  
Department of Computer Science  
University of Illinois at Urbana-Champaign  
Email: adw@illinois.edu

Seyed Rasoul Etesami  
Electrical and Computer Engineering  
University of Illinois at Urbana-Champaign  
Email: etesami1@illinois.edu

Abstract—Advances in computing hardware and novel multimedia applications have urged the development of handheld mobile devices such as smartphones and PDAs. Video content is increasingly consumed daily in mobile devices as well as in first response emergency services. With this significant increase of mobile video, one of the challenges is how to efficiently transmit the bulky videos to resource-constrained mobile devices. Despite the attractive features of the widely-used approaches on the bitrate adaptation such as DASH standard, video delivery impose significant demands on the limited battery capacity of mobile devices. Thus the development of efficient approaches to decrease the amount of streamed data while maintaining maximum possible quality with the aim of increasing the battery lifetime has become a key research topic.

In this initial study, we run an offline study towards design of energy-efficient delivery of video segments to mobile devices with limited energy budget over wireless networks, with the aim of integrating into DASH standard. Our results show that using our proposed adaptation significantly improves the quality of information per unit of energy consumed to stream videos to mobile devices.

I. INTRODUCTION

The growth of multimedia services, including streaming and conversational services, is one of the key drivers of the evolution to new mobile technologies and standards. Aside from the extensive use of mobile devices in daily life, handheld mobile devices also have been significantly considered in tactical edge \(^1\), and is reaching first responders, disaster-relief workers, and soldiers in the field to aid in different missions. An underlying assumption in these scenarios is connectivity to the cloud, which is not always available or reliable at the tactical edge. Thus the users are considered to be connected to a server as cloudlet, a private cloud by using WiFi located in a stable and secure environment, far from physical threats \([1], [2]\).

Unfortunately WiFi-enabled mobile devices need a high power to generate a stronger signal compared to other radio-based interface cards such as Bluetooth and 3G \([3], [4]\). A recent study precisely measured energy consumption for different parts of a mobile phone mainly for wireless communications. The results showed that the WiFi IEEE 802.11 Network Interface Controller uses as much as 24 times more power while downloading data compared to the idle mode \([5]\).

Rahmati, et al \([6]\) and Balasubramanian, et al \([7]\) investigated a simple linear-cost energy model for wireless data transfers, assuming constant network conditions throughout a single transfer. They modeled the energy cost for establishing a connection and transferring \(n\) megabytes of data as

\[
E = E_c + n \cdot E_t \tag{1}
\]

where \(E_c\) is the energy cost for connection establishment and \(E_t\) the energy per MByte of data transfer. Figure 1 shows the average energy consumed for downloading data of different sizes against varying inter-transfer times in WiFi.

Cameras on mobile devices also have given rise to significant sharing of videos in the context of tactical edge. It enables scenarios where there is an immediate need for video information. However, wireless delivery of huge data is power intensive, and query results can be large as videos are bulky. So it is necessary not only to consider that mobile devices offer budget-based battery power, but also power-intensive tasks such as video delivery consume large amounts of battery power, and that bandwidth in tactical edges, such as those experienced by first responders and soldiers in the battlefield, is limited.

Unfortunately due to these resource limitations, there must

---

\(^1\)First responders operating in crisis situations such as earthquakes, tsunamis, and other similar emergencies can also be considered to be operating at the tactical edge.
be a balance between the requirements of data and the consumption of limited resources, both at the server side and mobile client side. One of the challenges to achieving this balance is to meet the demands to achieve maximum Quality of Information (QoI) despite the limited battery power of the mobile device with minimum negative impact. While there has been considerable research to enhance computation, delivery time, and battery life, most of these work assume reliable connectivity and do not provide any energy-budget-compliant approaches; an invalid assumption in tactical edge and hostile environments.

In this paper, we run an offline study to efficiently deliver data to mobile devices with limited energy budget over wireless networks, targeting video data. Given the video which should be delivered to the user, our goal is to maximize the total QoI of the transmitted video data so to satisfy an energy budget specified by the user’s mobile device. In this pilot study, our approach is to estimate a download budget that achieves the energy budget based on the relationship between energy consumption and the amount of data downloaded, and to selectively choose the proper resolutions of the video chunks, so that the total amount of video data transmitted to a mobile device satisfies the available budget.

The paper is organized as follows: in Section 2, we discuss relevant work on adaptive multimedia delivery. Section 3 explains our methodology for energy-budget-compliant adaptive video delivery. Section 4 presents our evaluation and experimental results, and Section 5 contains our conclusions and a discussion of future work.

II. RELATED WORK

Work that systematically explored energy-aware adaptation in general, was done in the late 90s and early 2000s [8], [9] and studied laptops, whose energy profile is very different than that of modern smartphones. On the other hand, recently there have been interesting researches on multimedia optimization and HTTP streaming for mobile clouds focusing on the network coding, rate determinations, and quality of experience [10], [11], [12], [13]. Dynamic Adaptive Streaming over HTTP (DASH) in specific, also known as MPEG-DASH standard, is the first adaptive bitrate streaming solution which provides adaptive bitrate streaming whereby the resolution of a video is altered while it is being sent from server to client [14]. The multimedia content is stored on an HTTP server, and consists of a Media Presentation Description (MPD), which describes a manifest of the available segments, their various alternatives, their URL addresses, and other characteristics. Unfortunately there is no “energy-aware” implementations for MPEG-DASH.

While there are numerous work proposing methods to reduce energy for video delivery, to the best of our knowledge, there is no application-level work that takes into account the available energy budget for adaptive delivery of cloud video to battery-operated devices, given the energy characteristics of Wireless Network Interface Controllers (WNIC). X. Liu in their famous study [15] made a case for a video control plane combined with video bitrate selection/adaptation using large-scale measurements gathered from over 50 million users. In their work it is assumed that a video content provider such as YouTube or Hulu runs such a control plane to monitor and improve the video experience for its customers. The control plane can use a global optimization of client and network conditions to dynamically optimize the video bitrate. While interesting, this work does not provide a “local optimization” module for each user. In our work, we further enhance the optimization module on the control plane and modify it by embedding a “local” optimization module integrated with the global optimization module, and try to specifically address local optimization. Figure 2 shows an overview of the control plane, while the highlighted module depicts our modification. In our previous work [16], we studied how to efficiently transmit bulky 3D graphics information to bandwidth- and power-limited mobile devices targeting mobile 3D Games. In this study, we borrow concepts from [14], [15], and [16] and take a step towards design of an adaptive approach for efficient delivery of videos to mobile devices with a limited energy budget over wireless networks.

III. PROPOSED FRAMEWORK

In this section, we explain our methodology regarding the implementation of different parts of the adaptive video delivery system. Our design allows for efficient delivery of video segments or chunks already available via cloudlet to mobile devices while satisfying a download budget which is estimated based on an energy budget as explained in Section I, with the aim of decreasing power consumption. Figure 3 shows a detailed overview of different processes in our framework. As can be seen in the figure, the system consists of two parts: client-side and server (or cloudlet) -side. Prior to the video data receipt, the client device sends the current available energy budget to the server (we do not distinguish the control plane and server from the cloudlet), which is then used by the server to optimize the segments that will be delivered given their priority as an output of the priority system, which specifies how important each segment is for
the receiver to the context of the session or mission. To provide a more accurate energy-budget translation to a download budget, we can embed a complementary module which uses the feedbacks from devices’ power traces for further fine-grained refinements to our energy model as defined in equation 1. To achieve this, the system uses a classification list to prioritize the currently required segments. Then, based on the budget constraint received from the user and the prioritized list, proper resolutions of the video segments are selectively chosen, serialized and streamed to the target mobile device. It is assumed different resolution of video segments will be available on the cloudlet. For communication, DASH is used as a complementary technique to send the segments over the wireless network. Finally, the client receives the segments and in parallel, as a part of the client-side DASH, the newly received segments are buffered and video playout will continue. Receiving an optimum and efficient size of video segments reduces the network bandwidth and thus the energy consumption of the handheld device that receives the data as well as other limited resources (such as memory) for less-prioritized segments.

\[ \text{Maximize } \sum_{\tau_i \in T} Q_{\tau_i} \text{s.t. } \sum_{\tau_i \in T} s_{\tau_i} \leq W. \]  

This selection scheme is the well-known 0-1 Knapsack optimization problem. The 0-1 Knapsack problem is NP-hard but there are good, efficient, approximation algorithms (fully polynomial approximation schemes), so this approach is computationally feasible. However, by using this method, only a subset of the segments would be selected and transmitted to the client, which can not be acceptable since the user intends to receive all the segments.

To overcome the shortcomings of the 0-1 Knapsack approach, we propose heuristic algorithms that send all segments, but with different resolutions according to their priorities. This is the multiple-choice knapsack problem, in which the items (in this terminology, segments) are subdivided into \( k \) different groups, and one segment from each group should be chosen, so that finally all segments can be delivered within available budget. We define three heuristic algorithms to choose which resolution to choose for each segment, which are described in subsection III-B.

The idea of multiple-choice knapsack problem has been applied to certain contexts. Y. Song et al in their paper [17] investigated the multiple-choice knapsack problem and its applications in cognitive radio networks. In their paper, a centralized spectrum allocation in cognitive radio networks has been formulated as a multiple-choice knapsack problem. Lamani et al [18] also proposed an end-to-end quality of service to tackle the problem of setting end-to-end connections across heterogeneous domains modeling the complexity as a multiple choice knapsack problem.

A. Problem Definition

For segments resolution selection and streaming, the most significant factor in both battery and bandwidth usage is the amount of data downloaded by the mobile device’s Wireless Network Interface Controller (WNIC). As discussed in Section I, we suppose that the mobile client specifies an energy budget, and a download budget that achieves the energy budget is estimated. If the total size of the video segments does not exceed the download budget, then all of them can be delivered. If the download budget \( W \) is insufficient to deliver all of the segments, then the total size must be reduced. Every segment has a specific size given the MPD, and we must decide how to deliver them within the budget \( W \).

One approach to reducing the total size of the segments that are delivered is to transmit a subset of them. Let \( T = \{ \tau_1, \tau_2, \tau_3, \ldots \} \) be the set of segments. Each segment \( \tau_i \) has a size \( s_{\tau_i} \) and an associated value \( v_{\tau_i} \), which is based on the priority system. We just use the term value to represent the importance value. We also define the QoI contribution of each segment to be \( Q_{\tau_i} \), which is a function of its size and value. The goal is to stream a subset of segment \( T' \subseteq T \) that maximizes the total QoI of the delivered segments without exceeding the budget \( W \). In other words

\[ \text{Maximize } \sum_{\tau_i \in T'} Q_{\tau_i} \text{s.t. } \sum_{\tau_i \in T'} s_{\tau_i} \leq W. \]
that, we can tag the segments into two different importance classes: Less-Important ($C_1$) and Important ($C_2$).

First, a list is created containing all segments classified by their importance class. Each $\tau_{ij}$ in the list represents a single segment where $i$ signifies the importance class, and $j$ is the index of the segment within that specific class. Given different resolutions represented as various levels, we denote $R(i)$ to represent resolution of level $i$. For every level $i$ and $j$, the sizes of segments in resolutions $R(i)$ and $R(j)$ are available.

Now we describe how our main heuristic algorithm works.

Our problem is to transmit segments to a mobile device in a way that maximizes the total QoI of the transmitted segments within an energy budget that is specified by the mobile client. We cannot know the energy consumption associated with segments in advance, but as argued in Section I, it is closely related to the sizes of the segments needed to be streamed to the client. So we estimate a download limit $W$ based on the energy budget. We embed a complementary module to achieve more accurate estimation based on fine-grained feedbacks collected from the mobile device.

Reducing the resolution also leads to reduced QoI. Thus we set a user-defined minimum acceptable resolution. Let $R_{(min)}$ be the minimum resolution that is acceptable to the mobile user. The two priority classes, $C_1$ and $C_2$, has an associated relative value ($v_1$ and $v_2$). The assignments to classes and the associated values are decided by the users. We normalize the values by setting $v_1 = 1$. Let $S_1$ be the total original size of segments in $C_1$ before adaptation, $S_2$ the total original size of segments in $C_2$ before adaptation, and $S = S_1 + S_2$.

For simplicity, we assume that the QoI of a single transmitted segment is a function of its size and its relative value. We use the simplest of these functions - the product of size and value. For example, a segment $\tau_i$ of original size $s_{\tau_i}$ with reduction factor $r_{\tau_i}$ and value $v_{\tau_i}$ has QoI $\sum_{j} r_{\tau_i}^{x} v_{\tau_i}^{x}$. Our goal is to select appropriate segment resolutions in a way that maximizes the total QoI subject to the constraints of the download budget $W$ and minimum resolution $R_{(min)}$. We have proposed three different heuristic algorithms which we describe in the following.

**Algorithms**

Calculate $S_1$ (total size of all segments in $C_1$), $S_2$ (total size of all segments in $C_2$), $S = S_1 + S_2$, $\tau_{R_{(min)}}$ (the size of segment $\tau_i$ when in minimum resolution), and $W$ (available budget).

- If $S \leq W$, then no adaptation is needed.
- If $\sum_{\tau_i} \tau_{R_{(min)}} > W$ then the optimization problem cannot be solved within the constraints $W$ and $R_{(min)}$ to stream all the segments. However, in practice we find a maximum cut-off $l$ and stream the corresponding subset $T_l \subseteq \mathcal{T}$ so that $\sum_{\tau_i \in T_l} \tau_{R_{(min)}} \leq W$.
- Otherwise, we solve one of the following subproblems.

**Subproblem 1:**

If $S_2 + \sum_{\tau_i \in C_1} \tau_{R_{(min)}} \leq W$ then all segments in $C_2$ can be sent as is and the problem is to solve the problem for $C_1$ in a way that maximizes the QoI of the adapted segments in $C_1$ within the download budget $W_1 = W - S_2$.

**Subproblem 2:**

If $S_2 + \sum_{\tau_i \in C_1} \tau_{R_{(min)}} > W$ then we transmit all segments in $C_1$ by $R_{(min)}$ and the problem is to adapt the segments in $C_2$ in a way that maximizes the QoI of the adapted segments in $C_2$ within the download budget $W_2 = W - \sum_{\tau_i \in C_1} \tau_{R_{(min)}}$.

**Algorithms for subproblem 1 (Algorithm 1):**

Assuming segment $\tau_i$ with original resolution $R(i)$ is represented by $\tau_{R(i)}$, then calculate minimum $j_1$ such that:

$$\sum_{\tau_i \in C_1} \tau_{R(i - j_1)} \leq W_1 < \sum_{\tau_i \in C_1} \tau_{R(i - j_1 + 1)} \quad (3)$$

In other words, find the minimum $j_1$ such that all segments in $C_1$ can be transmitted within the budget $W_1$ if their resolution is reduced by $j_1$ levels.

The particular goal here is to maximize the total size of the segments sent within the budget $W_1$. To achieve this, we reduce the resolution of the first segment in $C_1$ by $j_1$ resolution levels. Suppose that after reduction, it has size $x$. This leaves a budget of $W_1 - x$ for the remaining segments. We then calculate a new $j_2$ for the remaining segments using budget $W_1 - x$ and reduce the resolution of the second segment in $C_1$ by $j_2$ levels. This is repeated until all segments have been adapted.

The other two heuristic algorithms for subproblem 1 (namely aggressive mode and round robin) are represented in Algorithm 2 and Algorithm 3 respectively. However, for the rest of the paper, we suppose Algorithm 1 is our main adaptation approach unless it is directly stated.

**Algorithm for subproblem 2:**

The algorithm is the same as for subproblem 1, except that we
are adapting segments in $C_2$ with an available budget $W_2 = W - \sum_{\tau_i \in C_1} \tau_i R(\text{min})$.

IV. EVALUATION

We evaluated our work based on a video dataset consisting of 47 videos, with segments of different resolutions. The priority system is designed in a way that allows the users an option to tag the video segments by manually setting the importance value. Based on the segments context, the important and less-important segments account for 41.1% and 58.9% of total size, respectively.

In our experiments we used HTTP as a communication protocol for streaming of videos from the server to the mobile device. To evaluate our proposed energy-efficient adaptation algorithms, the client device was an HTC 3D EVO smartphone with a Snapdragon S3 chipset and a dual core 1.2 GHz processor. During our experiments, the distance of the client device with the 802.11g WiFi router was 5 meters, receiving a signal strength of -60 dBMW. To calculate the amount of consumed energy, we used PowerTutor [19] profiler which measures the power consumption of various hardware components using device’s built-in battery voltage sensors.

To evaluate our proposed adaptation algorithms, we ran our experiments with the available budget $W$ set to be different percentages of $S$ (total size of all segments). In particular, we set $W$ to $0.2S$, $0.3S$, $\cdots$, $1.0S$ corresponding to 20%, 30%, $\cdots$, 100% of the total size of segments. We chose $R(\text{min})$ (i.e. the minimum acceptable resolution) to be $320 \times 240$ that resulted in some videos in $C_2$ being adapted for most values of $W$. In practice, a user will choose $R(\text{min})$ based on the perceived quality of the videos. We did all the experiments with the segments in each importance class sorted by both decreasing and increasing size, and also unsorted based on random ordering. Each trial of our experiment was run until segments are fully received, and we repeated each test several times to ensure that the standard deviation of the measurements was within acceptable limits. Using PowerTutor measurements, Figure 4 shows the average energy consumption of the WiFi 802.11g WNIC in Joules with a precision of 0.1J, and showed how it changes as we apply our adaptation approach for different values of $W$ (as a percentage of $S$). We measured the total energy to stream the sorted (increasing) segment list. As can be seen, the increase in total energy consumption grows roughly linearly with the amount of data being received at the client side in agreement with Equation (1) in Section I and Figure 1. The dashed line shows the linear trend-line for the energy consumption of the segment list.

Figure 5 shows the results for the average size reduction in terms of average adaptation ratio, measured for all segments in $C_1$ and $C_2$, for sorted (increasing) segment list. We also measured the approximation error relative to the optimal solution as a factor of how well the three heuristic algorithms work, for three different ordering of segments, as shown in Figure 6. It should be noted that the approximation ratio of all three heuristic algorithms are more than 99% relative to the optimal solution, resulting in an approximation error of less than 1%. Also as can be seen, Algorithm 2 (i.e. round robin) is not performing as well as the other two algorithms for small amount of available budgets. Since the difference of the three algorithms is trivial, we considered the first proposed algorithm to be the main adaptation approach.

As discussed previously, we assume that the QoI of a streamed segment is a function of its size and a relative value. If $S_1'$ and $S_2'$ are the total sizes of the streamed segments in classes $C_1$ and $C_2$, respectively, that are received by the client, and $v_1$ and $v_2$ are the associated relative values, then we define the total QoI to be

$$\text{Total QoI} = \frac{v_1 \cdot S_1' + v_2 \cdot S_2'}{S \cdot (v_1 + v_2)}$$ (4)

where $S$ is the total size of the segments before adaptations. We normalized the relative values by setting $v_1=1$. The total QoI is a measure of the effectiveness of an approach to maximizing the total amount of data based on prioritizations, with larger values being more effective.

Figure 7 (top) shows our experimental results for total QoI per available budget for sorted (increasing order) segment list. We used three different pairs ($v_1=1$, $v_2$) of relative values, (1,1), (1,2) and (1,3), to differentiate the priorities of segments in $C_1$ and $C_2$. As can be seen in the figure, the total QoI increases as the ratio $\frac{v_2}{v_1}$ increases, confirming that our proposed
Fig. 6: Relative approximation error measured for three different ordering of segments: increasing (top), random (middle), and decreasing (bottom) order, each for our three heuristic algorithms.

Fig. 7: Comparison of the normalized total QoI (top) and QoI per unit of energy (bottom) measured for three relative value pairs $(v_1=1,v_2)$.

tagged segments are adapted. This specific point is considered as the peak of QoI. As we go ahead with adaptation of the segments in $C_2$, the gain in QoI brought by our approach is being decreased; a fact which is confirmed by tracking the segments being adapted in $C_1$ and $C_2$ along with the achieved QoI.

However, our method does provide quality degradation in general, but considering the power savings achieved using our method, it is reasonable to believe that users in situations with severe resource constraints would make this sacrifice in quality in exchange for respecting their available energy budgets.

V. C

In this paper we introduced an adaptive prioritized approach to manage delivery of bulky videos to mobile devices with energy constraints, especially those seen in the context of crisis situations. Our approach is to selectively choose different segments resolutions so that the overall amount of data transferred to a mobile device does not exceed an available budget, with the aim of decreasing the amount of energy needed to download the segments. The evaluation results show that our energy-aware adaptations improves the total gained QoI of videos by making best use of the limited energy of mobile devices.

We are currently extending our approach for multiple levels of priorities as opposed to just two. We also plan to propose online adaptation algorithms, as well as distributing the system to collect measurements that will help us to quantify both user satisfaction, and network dynamics.
REFERENCES


