

Video-Streaming for Fast Moving Users in 3G Mobile Networks

Anna Kyriakidou
Deprt. of Informatics &
Telecommunications
Univ. of Athens, Greece
a.kyriakidou@di.uoa.gr

Nikos Karelos
Deprt. of Informatics &
Telecommunications
Univ. of Athens, Greece
n.karelos@di.uoa.gr

Alex Delis
Deprt. of Informatics &
Telecommunications
Univ. of Athens, Greece
ad@di.uoa.gr

ABSTRACT

The emergence of third-generation (3G) mobile networks offers new opportunities for the effective delivery of data with rich content including multimedia messaging and video-streaming. Provided that streaming services have proved highly successful over stationary networks in the past, we anticipate that the same trend will soon take place in 3G networks. Although mobile operators currently make available pertinent services, the available resources of the underlying networks for the delivery of rich data remain inherently constrained. At this stage and in light of large numbers of users moving fast across cells, 3G networks may not be able to warrant the needed quality-of-service requirements. The support for streaming services necessitates the presence of content or media servers properly placed over the 3G network; such servers essentially become the source for streaming applications. Evidently, a centralized approach in organizing streaming content might lead to highly congested media-nodes which in presence of moving users will certainly yield increased response times and jitter to user requests. In this paper, we propose a workaround that enables 3G networks to offer uninterrupted video-streaming services in the presence of a large number of users moving in high-speed. At the same time, we offer a distributed organization for the network's media-servers to better handle over-utilization.

Categories and Subject Descriptors

H.3.4 [Systems and Software]: Information networks; H.3.5 [Online Information Services]: Data sharing; C.2.4 [Distributed Systems]: Distributed applications

General Terms

Algorithms, Design, Performance

Keywords

Streaming for moving users, mobile multimedia services, rate adaptation, real-time streaming

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

MobiDE'05, June 12, 2005, Baltimore, Maryland, USA.
Copyright 2005 ACM 1-59593-088-4/05/0006 ...\$5.00.

1. INTRODUCTION

The third generation (3G) mobile phone system *UMTS* enables better quality and allows for more convenient use of multimedia messaging and video-services by offering higher bandwidth and lower latency than its GSM and GPRS predecessors [15, 1]. *UMTS* furnishes upto 2Mbps rates for indoor and upto 384Kbps for outdoor environments. Clearly, much improvement in terms of allocated resources has been made for the handling of "rich" data including multimedia messages and video-services. Nevertheless, the available resources still present significant limitations for the scale up of services when very large number of clients are present in a cell of a network. Perhaps, the most daunting challenge comes from moving users who access multimedia data and video feeds with the help of their mobile phones and PDAs while traveling on either private vehicles or mass transportation means such as commuter trains and public busses. Evidently, a large number of concurrent connections soliciting data resources in a cell and being handled in real-time pose significant capacity problems for the underlying 3G network. The situation becomes even more challenging when users attempt to follow streaming sources while on the move. We consider streaming services to be of key importance as they will ultimately offer information on demand for the moving user in any geographic position and at any time. In order to facilitate video streaming over *UMTS* networks a number of issues have to be addressed so that users do not experience delays and discontinuities during playback. The two core aspects that require attention are the variations of the available bandwidth as users enter and leave cells as well as the effective management of handovers as roaming users attach to different base-stations along their trajectory. The problem of graceful transition when moving between base-stations becomes more critical when the users are on high-speed motorways. In this case, handovers become more frequent and traffic load at successive base-stations may vary. In this paper, we outline the above emerging problem and propose a scheme that allows for not only improved network resource sharing but also for enhanced management of streaming sources to the mobile user. It is expected that base-stations are to transmit in different bitrates throughout the journey of an individual as cells will undoubtedly present diverse levels of congestion and availability of connections.

When considering vehicular users in general, one can exploit the fact that the user's trajectory can be predicted in a satisfactory manner. An early method to attain this goal is to keep an aggregate history of observations made regarding the movement of users within each cell [9]. Based on this information, probability density functions for the prediction for the next-cell-to-move can be derived and used. Traffic authorities imposed speed limits and road signals can

also assist in the more accurate estimation of a user's average speed. In addition, the user's direction can be predicted reasonably well by keeping track of her trajectory thus far. Although a model for precise prediction is beyond the scope of this paper, we can assume that there are already techniques that can offer a good estimation of a moving user path. For instance, an individual's geographic location could be tracked with the assistance of *UMTS Location Service (LCS)* [2] that can identify the cell that a user presently appears in. If a user is moving along a highway, one could easily estimate not only the direction of his movement but also his average speed along a given trajectory. Finally, the soon anticipated incorporation of *Global Positioning System (GPS)* receivers into mobile phones through *A-GPS* features [30] will help in the very accurate user positioning and extraction of their movement characteristics. It is our conjecture that at this stage, simply knowing the overall direction of a user's trajectory in conjunction with the highway that she travels on can ensure timely video-streaming and playout continuity for users.

In our streaming environment, there exist three distinct types of synergistic computing systems: media-servers, base-stations, and user equipment. These systems are organized in three functional layers as Figure 1 depicts. The role of the media-servers is to predominantly manage content in a highly distributed fashion where fast content retrieval and data resilience can be guaranteed. Base-stations handle all user initiated connections to the 3G network and through their channels offer users requested data. The last tier of Figure 1 consists of cellular phones and PDAs equipped with appropriate video-players and featuring minimal buffer capabilities to support streaming.

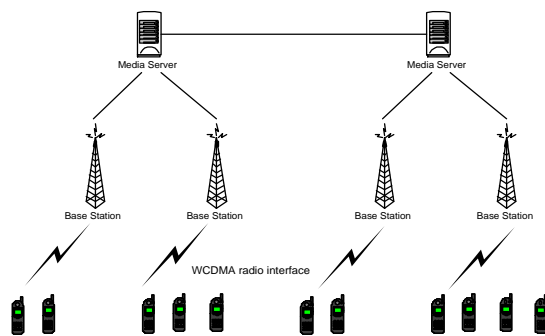


Figure 1: Three-Tier Organization for Streaming Provision.

Media and base-stations are internetworked via high-speed wired links while *UMTS* offer wireless connections between user equipment and base-stations. This distributed media-server architecture provides for dividing of large video streams into multiple segments [34, 14]. A media-server initially retrieves a solicited video object either from its storage units or from another remote media-server. In this paper, we take the approach that instead of transmitting the entire object to a single base-station, we first segment it and then forward its segments to the base-stations along the user's path. Our rationale is that an individual base-station handles only a section of the video file; the size of the section in discussion is commensurate to the duration of a user's trip inside the cell. Clearly, video-object segmentation reduces both network transmission costs between media and base-stations and start-up latencies that users experience upon cell entrance. On the other hand, segmentation might get more complex if a user remains longer in a cell than her estimated time and so she may face delays in the reception of

frames. We address this issue by continual monitoring of both user speed and position and by doing so giving the base-station the option to receive additional video increments and sufficiently feed a user at all times.

Our work requires minimal buffer capabilities for mobile stations so that a sufficient number of frames can be accommodated. Buffer presence assures that the playout does not stop if the base-station emits at lower bitrate due to 3G network congestion. We propose a rate adaptation scheme which allows a base-station to adjust its transmitting bitrate at any time according to base-station load and the states of the client-side buffer. The *UMTS* streaming class defines that the bitrate assigned to a moving user is guaranteed even though it might be less than the maximum video bitrate [15]. The base-station's decision of accepting a video streaming session has an impact on all subsequent base-stations that follow up on the video delivery process. While a streaming session is in progress, the load of base-stations-to-serve may dynamically change and potentially lead to session drops. Such drops are highly undesirable and we adopt a policy to address this issue. Our proposed scheme gives a base-station the opportunity to appropriately alter the transmission bitrate by taking into account the current base-station load and simultaneously ensuring that the client buffer does not starve.

The rest of the paper is organized as follows: Section 2 presents the overall system architecture and examines the interaction between media-servers and base-stations. Section 3 proposes our bitrate adaptation scheme and Section 4 discusses the results of our preliminary experimentation. Finally, related work and conclusions can be found in Sections 5 and 6 respectively.

2. MEDIA-SERVERS/BASE-STATIONS INTERACTION

The fast movement of users via different cells of the 3G network imposes a set of new requirements for the entire video delivery system. As the user relocates rapidly, she faces a high number of handovers during her journey and, as a consequence, a large video stream has to be fetched from different base-stations in order to warrant continuous playback. As suggested earlier, we assume the deployment of dedicated media-servers which undertake both the storage and distribution of video-streams to underline base-stations. It is imperative that media-servers, base-stations, and user-equipment involved in a streaming session must cooperate in order to guarantee *QoS* for the video reception of the moving individual. In this section, we outline our overall architecture, discuss the content delivery process that media-servers carry out, and present specific algorithms for video segmentation and content distribution.

2.1 Architecture

The three distinct types of cooperative computing systems (shown in Figure 1) organized in a multi-tier architecture constitute our proposed streaming environment. We assume that base-stations communicate with the mobile stations through the *WCDMA* radio interface [15]. Each cell of the *UMTS* network is served by a different base-station whose responsibility is to deliver the video streams to its constituent mobile users. A streaming service necessitates the use of media servers that will handle the storage and delivery of video files [3]. Although we could adopt a centralized approach to accommodate the media in delivery, high contention and resource over-utilization would impact user request response times greatly. Clearly a distributed approach that webs media-servers together is required. High-speed wired communication means link these servers and all share required meta-data structures.

Due to incurred costs and the fact that users move at high-speeds, having a dedicated media server for each base-station would be a poor decision. If the mobile user is traveling at a speed of 100 km/h and the cell radius is 0.7 km, then he will pass through the given cell in 50.4 seconds at maximum, assuming an hexagonal shape. This implies that the number of frequent handovers taking place increase the interaction among media-servers that will have to be involved throughout the streaming session. Furthermore, in order to avoid under-utilization of media-servers and strike a balance in the aggregate use for facilitating streaming, we group 3G cells into groups as Figure 2 depicts. This assignment is expected to happen in a static manner although it could be modified to reflect emerging new realities in the core network. In this regard, Figure 2 shows a network layout in which sets of sixteen cells are configured to function as a group. In this example, the mobile user is currently in a cell of group A and is heading towards group F. The media-servers that can be involved in the delivery of video objects are A, D, E and F. The server A is expected to interact with server D, D with E, and E with F. In this chained-fashion, we anticipate that the media servers notify each other about the streaming session of the oncoming mobile user. In addition, the media-servers send and receive in pipelined fashion the video object under transmission.

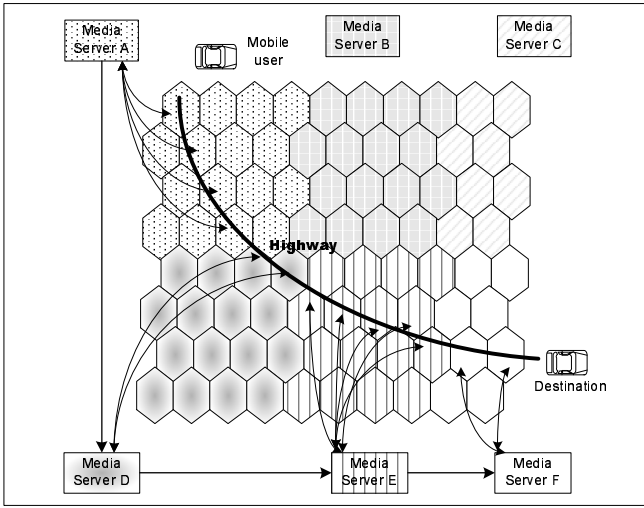


Figure 2: Grouping of Cells

There may be other interactions as well. For instance, there must be cooperation between media server A and F if the requested video object is initially located at F. A media server accepts video requests from base-stations residing in its group; if it does not currently have the object it is responsible for locating—using the meta-data structure—and fetching it. Due to the location awareness of our approach, we assume that each server predominantly stores streams specific to its own geographic area. For instance, if the route drawn in Figure 2 crosses a county, video clips with showing traffic conditions ahead in specific points may be requested. Similarly in a city setting, such requests may entail multimedia location-based virtual presentations.

2.2 Content Distribution

Provided that a media server has a video clip in place, a straightforward approach would be to transmit the object into all base-stations operating in cells located on or near-by the user's trajectory. This is not only wasteful in both network and base-station

resources but also increases the user-perceived playout latency. Therefore, we resort to using video segmentation [14], [21] in order to reduce network transmission costs between the video-holding media server and its subordinate base-stations. Segmentation also decreases the start-up latency that users experience upon cell entrance.

The length of a video-segment, denoted S_t , sent to each base-station is proportional to the average time that the user is expected to stay in the specific cell. The process of segmenting the video streams into chunks of frames of specific duration assumes that the media servers are aware of the underlying cell configuration. In particular, a media-server has to be aware of the precise manner with which a motorway cuts across its subordinate cells, the direction as well as the speed of moving users. With this data available, the media-server in discussion can approximate the time that a user spends in a cell. For example, if a motorist moving at 100km/h has just entered a cell and departs after traversing a 1km route portion, the server can compute the duration of the user's stay to be at approximately 36 seconds. The media-server can capitalize on this very information to appropriately segment the streamed video; it only dispatches enough frames for a playout period of 36 seconds.

The duration of a user's presence within a cell may vary according to speed changes with clearly lower speeds leading to elongated stays in the cell and vice versa. Should the speed be decreasing, the base-station will ultimately require more frames from the media server than the number predicted once the user appeared in the cell. Such a request constitutes a "cache miss" which will not be noted by the user if detected on time and acted upon by the coordinating base-station. Imposing a minimum threshold in the number of frames always available for delivery at a base-station may help overcome such "cache-misses". Therefore, when the number of frames awaiting transmission on a base-station falls below the abovementioned threshold, the base-station signals its need for additional frames to its overseeing media-server; should the latter act upon this request, additional frames arrive on time at the base-station for delivery. On the other hand, as soon as a user increases speed, she will depart the cell extend earlier than initially anticipated. The drawback here is that the media-server has provided the base-station with more frames than those eventually needed. During the handover process, the coordinating media-server has to generate a video-fragment which in its opening contains frames that have already been transmitted to the previous base-station but not-yet-seen by the user.

Our approach allows a base-station to dynamically alter the transmission bitrate according to its current load. Under light load, a base-station may opt to increase the transmission rate for a specific video-stream thus leading to potential frame shortage. To easily avoid such shortage, we use the minimum allowed vehicle speed to compute the size of the video-segment S_t to be transported to base-stations:

$$S_t = \frac{Distance}{Minimum.Speed} \quad (1)$$

In most freeways there are authority-posted limits for minimum allowed speed in each road segment. As media-servers are aware of the geographic area that they serve, such minimum speed rates are statically designated for each cell in their jurisdiction. Evidently, the video-segment size that we potentially use as safety against frame shortage is:

$$S'_t = \frac{Distance}{Minimum.Speed} - \frac{Distance}{Average.Speed} \quad (2)$$

Algorithms 1 and 2 depict the video segmentation and distribution that we follow in our media distribution. Upon a new video-streaming request, we assume that the media-server can efficiently retrieve the corresponding video-file either from local storage options or remote servers via its low-latency/high-bandwidth wired networking infrastructure. The identification of the user's current location, the precompiled knowledge of the traveled distance within a cell, in conjunction with the minimum allowed speed on pertinent highway segments, permit for the estimation of the maximum user stay S_t in a specific cell. Subsequently, the media-server can create the first segment of frames needed for transmission via the base-station to the requesting client. The size of a video-segment is given by $V = \sum_{i=1}^N F_i$, where F_i is the size of the i -th frame and N is the number of frames in the segment; we can easily compute N by multiplying the frame-rate (frames/second) with the duration of stay in a cell S_t .

Algorithm 1 Video Segmentation at Media-Server

```

1: Get Minimum User Speed  $MinSpeed$ 
2: Get  $PathLength$  in cell range
3:  $p \leftarrow$  last frame transmitted
4: if (New Session) then
5:    $S_t = \frac{PathLength}{MinSpeed}$ 
6: else
7:   // Shortage of Frames
8:    $S_t = \delta * \frac{PathLength}{MinSpeed}$ 
9:   // with  $\delta < 1$ 
10: end if
11:  $V \leftarrow i \sum_{i=p+1}^{p+N} F_i$ ,
    where  $F_i$  is size of  $i$ -th frame,  $N=FrameRate \times S_t$ 

```

In light of frame shortage, our video-segmentation algorithm dispatches into the base-station with need an increment of frames. This is defined as a fixed fraction of the length of S in the current cell (line 7 of Algorithm 1). Requests of such increments may occur multiple times before a motorist leaves a cell due to congestion.

A handover might find a moving user either serviced by a base-station in the realm of the current media-server or under the authority of a completely new media-server along the motorist's path. In the first case, the media-server initiates the delivery of the next video-fragment to the next-base-station encountered. The just departed base-station can help in determining the appropriate stream position p from which the segmentation will have to resume. The duration of the video-segment is computed anew using the same algorithm that now takes into consideration the data points from the new cell. Clearly, the length of the route as well as designated minimum speed limits may be different from those encountered in the previous cell.

In the second scenario, a handover may force a user to operate in an entirely new group of cells supported by a new media-server. In general, the portion of the "not-yet-seen" stream has to be forwarded from the previous to the new media-server unless the latter already maintains its own copy. If we are not interested in reducing the transmission costs, we can transport the entire video object to the new media-server using the assumed high-speed link. The media-server now in charge takes over the session identifier of the moving user and along with user state data from its previous location can help coordinate the delivery of the video in the new cell. To enhance coordination among media-servers in the highest-level of Figure 1, prefetching could be used [7, 34]. We could de-

Algorithm 2 Video Distribution from Media-Server to Base-Station(s)

```

1: while (OutstandingRequests) do
2:   if New Session then
3:     Start session (user's location, video stream, cell ID)
4:     if Video Stream Not in Storage then
5:       Get Video Stream from corresponding Media-Server
6:     end if
7:   end if
8:   if (not(Handover)) then
9:     Apply Video Segmentation Algorithm
10:    Send Video-Segment  $V$  to Base-Station
11:  else
12:    if (new Base-Station within Media-Server realm) then
13:      Apply Video Segmentation Algorithm
14:      Send Segment  $V$  to Base-Station
15:      Send playback position  $p$  to new Base-Station
16:    else
17:      Send Video Stream to next Media-Server
18:      Send playback position  $p$  to next Media-Server
19:    end if
20:  end if
21: end while

```

ploy prefetching of video-segments to media-servers and/or base-stations to facilitate playout continuity and minimize the start-up latencies.

Users moving with similar speed and nearby to the streaming user can benefit from the already segmented video stream and start the playout immediately. Caching efficiency is limited by the fact that only users with similar traveling behavior may use the video segments. We can overcome this limitation if the starting point of the video segment at the next cell corresponds to users traveling at the maximum speed within the current cell. At the same time, the total size of the segment caters for users that travel at minimum speed within the next cell, thus remaining longer in the cell's range. This ensures that successive base stations hold a sufficient amount of frames to serve users traveling at different speeds.

3. RATE ADAPTATION

In this section, we propose a rate adaptation scheme whose objective is to better serve the overall needs of fast-roaming users. More specifically, we present a mechanism used by base-stations to control the rate at which they transmit video to each user when the cell becomes overloaded and the transmission bitrate eventually needs to be decreased. In light of this reduction, we seek ways to avoid discontinuities in user playback and cell bandwidth overutilization lowering so the probability for a session drop.

While focusing on bitrate management between the second and last tier of Figure 1, we assume that pertinent video-segment data is available at a base-station. Upon session initiation, the size Q of the mobile device buffer becomes known to the managing base-station. In general, we assume that a video-object is divided into frames of constant duration. Frames that belong to the same file vary in size depending on the encoding rate and the scene content. A time domain perspective allows us to control the transmission rate examining the time interval between transmission of successive frames rather than their respective sizes. If the inter-departure time corresponds to the rate instructed by the file's frame rate¹,

¹Typical frame rate values are 25 frames/sec for the PAL color sys-

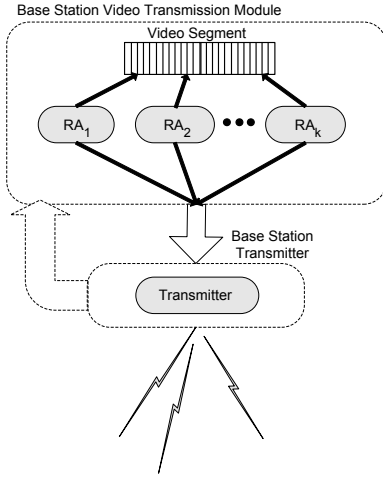


Figure 3: Rate Adaptation Modules within a base-station

the transmission bitrate corresponds to the file's encoding bitrate. Alterations in the inter-departure times result in the inversely proportional changes in transmission bitrate.

Let $\{X_k^i\}_{k \geq 1}$ denote the departure process of the video frames from the base-station, for the i -th user. If τ_k^i is the departure time for frame k , $X_k^i = \tau_{k+1}^i - \tau_k^i$ is the inter-departure interval for the k -th frame. In the absence of buffering capabilities on the mobile device, the smoothness of $\{X_k^i\}$ is critical for the smoothness of playback at the user's end. Thus, the following should hold:

$$P\{X_k^i = T\} \rightarrow 1 \quad (3)$$

where T is the inter-departure interval specified by the video-object frame rate. The buffer support that we assume available at the user-end enables the modification of the $\{X_k^i\}$ process reflecting modifications to the actual transmitting bitrate of the base-station.

Video Streaming modules are integral parts of the base-station configuration and each such module handles the transmission of a video-stream. Hence, a segment of a video-stream is assigned to an instance of a *Video Streaming* module for final delivery to the user's equipment. A *Rate Adaptation (RA)* element is assigned to each user session for the specified video stream. An *RA* is aware of the user's buffer size and is responsible for the forwarding of video frames to the actual *Transmitter* of the base-station. As information about the station's load is fed-back by the *Transmitter*, the *Rate Adaptation* element regulates the inter-departure process of video frames from the station to a user, in a way that preserves playback continuity. Figure 3 depicts the interaction among these two elements and the *Transmitter* at a base-station that serves k concurrent sessions for the same video-object.

The operation of the *Rate Adaptation* element is governed by periodic time intervals of constant duration, termed *Control Cycles*. Operating in the time domain, the module is aware of the exact number of frames the media player at the user-end will need over a specific period of time to ensure smooth playback. Let A denote the duration of the control cycle. Also, let $Q_A > 0$ be the occupancy (i.e., number of frames) of the buffer at the beginning of the control cycle and N_A the number of frames that will be reproduced at the user-end during the control cycle. Since N_A frames are requested

from the media-player and Q_A frames are accumulated the rate adapter needs only transmit $N_A - Q_A$ frames at minimum over the control cycle.

The video frame rate instructs that a frame be transmitted every $T = \frac{A}{N_A}$ *mseconds*. Each one of the $N_A - Q_A$ frames transmitted at minimum during the control cycle will depart the base-station at longer intervals equal to $T = \frac{A}{N_A - Q_A}$. The initial inter-departure time has been spread by a tolerance factor α ($\alpha \geq 0$) where:

$$\frac{A}{N_A - Q_A} = (1 + \alpha) \frac{A}{N_A} \Rightarrow \alpha = \frac{Q_A}{N_A - Q_A} \quad (4)$$

The *tolerance factor*, α , is a parameter of the control cycle; α may turn negative only when $Q_A > N_A$. Thus, a more specific definition of α would be

$$\alpha = \frac{Q_A}{N_A - Q_A} * 1_{\{Q_A \leq N_A\}} + 1_{\{Q_A > N_A\}} \quad (5)$$

If the *RA* element forwards frames at the rate instructed by the tolerance factor, the transmission bitrate over the control cycle will be equal to $B/(1+\alpha)$, where B is the encoding bitrate of the video-stream. A control cycle during which the base-station transmits at the minimum bitrate instructed by α is called a *degraded* cycle. A degraded cycle will lead to zero buffer occupancy at the end of the control cycle and the tolerance factor for the next control cycle will be equal to zero. Therefore, no two successive degraded cycles may occur.

Non-zero buffer occupancy at the beginning of a control cycle will be present only if the overall transmission rate over the previous control cycles exceeded B . This can be achieved if the *RA* element forwards frames at a higher rate when the station is underutilized. Let β denote the *speed-up factor*, the factor by which the bitrate increases in this case. An expression for the speed-up factor may be obtained if we consider that the maximum transmission bitrate will lead to a full user buffer at the end of the control cycle. If Q_A is the buffer occupancy at the beginning of the cycle, then the station may transmit at maximum $Q - Q_A$ frames over the control cycle. In this case, each frame will be transmitted every $\frac{A}{Q - Q_A}$ *mseconds*, with inter-departure interval having been decreased by β :

$$\frac{A}{Q - Q_A} = (1 - \beta) \frac{A}{N_A} \Rightarrow \beta = 1 - \frac{N_A}{Q - Q_A} \quad (6)$$

An upgraded cycle will transmit at a rate of $B/(1 - \beta)$. The speed-up factor may turn negative only when the free buffer space is less than the frames that will be played back during the control cycle. In this case, the cycle is *forced* to operate in degraded mode, so that we can avoid buffer overflow.

It is clear that the n -th control cycle may forward frames at a rate in the range of:

$$\frac{B}{\max\{(1 - \beta), (1 + \alpha)\}} \leq B_n \leq \frac{B}{(1 - \beta)} \quad (7)$$

The respective inter-departure process, $\{X_n\}$ will be in the range of:

$$(1 - \beta)T \leq \{X_{n,k}\} \leq \max\{(1 - \beta)T, (1 + \alpha)T\} \quad (8)$$

The *RA* element knows at any time the exact number of frames that have been transmitted to the user, and it also knows the time that has passed since the session initiation, which corresponds to the number of frames the playback process has consumed. The difference between the two values denotes the user buffer occupancy,

so no feedback mechanism is required as far as the user buffer occupancy is concerned. The algorithm followed by each *Rate Adaptation* element in the *Video Streaming* module is outlined in Algorithm 3.

Algorithm 3 *Rate Adaptation* element operation

```

1: // Executed at the beginning of every control cycle
2: // for user  $i$ 
3:  $Q_A^i = FramesTransmit_i - FramesPlayed_i$ 
4:  $\alpha = Q_A^i / (FramesCycle - Q_A^i)$ 
5:  $\beta = 1 - FramesCycle / (Q - Q_A^i)$ 
6:  $MinInterval_i = (1 - \beta) \times T$ 
7: if  $\beta < 0$  then
8:    $MaxInterval_i = MinInterval_i$ 
9: else
10:   $MaxInterval_i = (1 + \alpha) \times T$ 
11: end if
12:  $Interval_i = MinInterval_i +$ 
     $(CellLoadPerc/100) \times (MaxInterval_i - MinInterval_i)$ 

```

Since the duration of the control cycle is constant, multiple control cycles may occur during a user’s presence in the range of a single cell, depending on the size of the cell and the user’s speed. We assume that the each cell handover always initiates a new control cycle.

Algorithm 3 allows for alteration in the transmission bitrate by providing upper and lower bounds (i.e., $MinInterval_i$ and $MaxInterval_i$) to ensure the smoothness of the playout process. The choice of the actual bitrate within the specified range, at which the base-station transmits during a control cycle, is ultimately a function of the station’s load at the time. This load is continually estimated with the help of the *Transmitter* module. This feedback enables each *Rate Adaptation* element to cater for buffer occupancy increase, taking advantage of low system load periods. At the same time, by detecting high system load, the *Rate Adaptation* element lowers the transmission bandwidth, allowing for more sessions to be accommodated, while at the same time the playback process is not distorted.

4. EVALUATION RESULTS

In order to reproduce and experiment with the behavior of our proposed architecture and bitrate adaptation scheme, we have set-up a simulation testbed. We have assumed a user trajectory with duration of *200 seconds*. The user traverses numerous cells of different sizes. Each base-station is equipped with the *Video Streaming* module as described earlier. A *Control Cycle* of *5 seconds* is adopted by all elements. The buffer size at the user-equipment is assumed to be large enough to store *10 seconds* which is readily met by modern cellular phones and/or PDAs. We designate ten levels of base-station load with load changing at random times. The maximum duration of each load state is *30 seconds*. The *PAL* color system is assumed for the video being transmitted, so the default inter-departure time for each frame is set at *40 mseconds*.

At the beginning of each control cycle, the *Rate Adaptation* element applies the proposed Algorithm 3. The testbed initially computes the tolerance and speed-up factors thus generating the acceptable inter-departure times range. The actual inter-departure time for the control cycle is proportional to the station’s load at the time. If the station is lightly loaded, the minimum interdeparture time (maximum bandwidth) is applied. Conversely when the base-station is heavily loaded, the frames are forwarded to the transmit-

ter at the minimum rate instructed by the maximum interdeparture time. For intermediate load levels an appropriate value from the inter-departure times range is selected according to Algorithm 3 in a uniform fashion.

Figure 4 shows the evolution of the base stations’ load during the user’s trajectory, over all cells that the individual travels in. Having defined load of value 5 as “normal” load, the base-station load remains relatively high through the user trajectory with a few very short periods of low load.

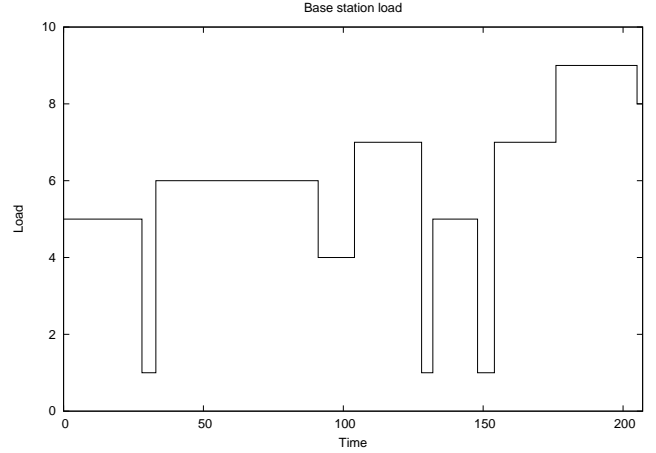


Figure 4: Base-station load during user trajectory

Figure 5 shows the applied transmission bitrate, along with the minimum and maximum bitrates allowed for every control cycle. The y -axis represents the percentage of the actual transmission bitrate to the video encoding bitrate. The bitrate is inversely proportional to the inter-departure times which are illustrated in figure 6. If we compare the curves of Figures 5 and 4, we can easily establish that the actual bitrate applied is a function of the base-station load and the calculated allowed bitrate range. At times of high load, the applied bitrate is closer to the minimum acceptable bitrate and conversely at times of light load, the applied bitrate is closer to the maximum acceptable bitrate. Figure 6 shows that the inter-departure times are indeed proportional to the base-station load.

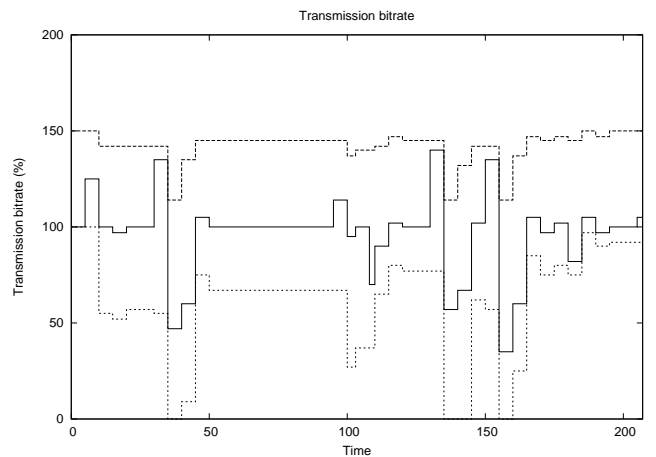


Figure 5: Allowed transmission bitrate limits and applied transmission bitrate.

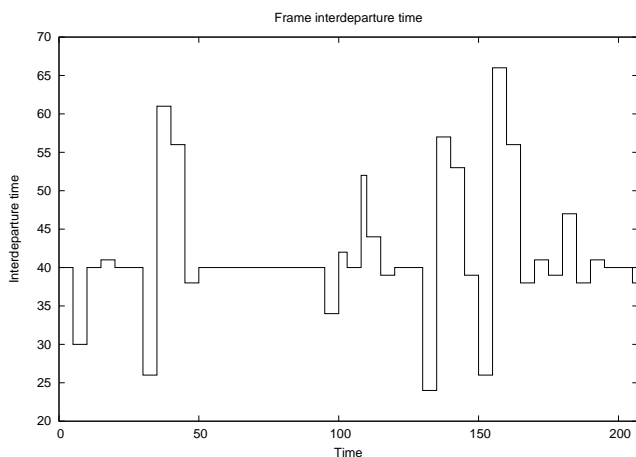


Figure 6: Applied interdeparture interval.

We show the user buffer occupancy throughout the trajectory in Figure 7. The buffer does not starve at any time suggesting the usefulness of our proposed scheme. The occupancy increases at times where the base station load falls below the normal load.

Under overall higher-than-normal base-station load settings, our tests show that the playback continuity is preserved by only taking advantage of relatively small periods of station underutilization to increase the transmission bitrate. The range of the acceptable transmission rate includes the bandwidth already guaranteed by the network upon session acceptance at all times. Therefore, the network was never *forced* to transmit at a higher-than-agreed bitrate. At times of station over-utilization, the decreased transmission bitrate allows for more calls to be accommodated significantly decreasing the probability of session drop.

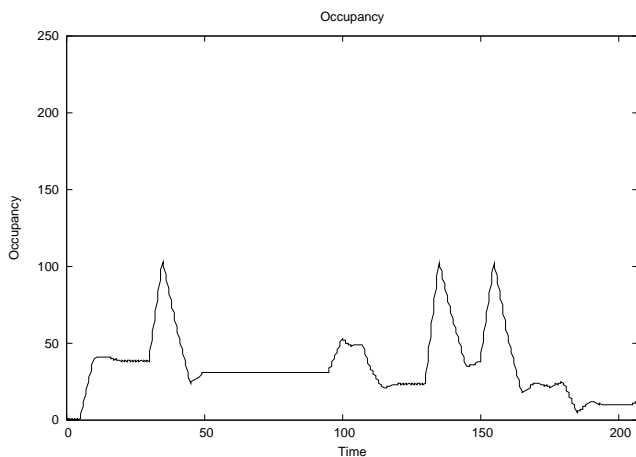


Figure 7: User buffer occupancy.

A system that adopts no rate adaptation scheme would constantly require a transmission bitrate equal to the 100% of the video encoding bitrate throughout the session. Although the bitrate of real-time video streaming sessions is guaranteed by the *UMTS* specifications, at near-capacity situations, the network would have to either drop a session or be forced to transmit with lower bitrate. The former case is clearly undesirable and the latter generates jitter effects for the end-user. This would happen even if the station

gets overloaded only for a period of time equal to the proposed scheme's control cycle duration. Discontinuities in the playback process may be observed at a system adopting the proposed rate adaptation scheme as well, however only in the case when the base-station is constantly load saturated, thus not allowing for any up-graded cycles to take place.

5. RELATED WORK

There has been a large amount of reported work in related areas that include caching for video systems, management of moving objects, and rate adaptation for streaming systems on wired networks. Data caching and mirroring has been proposed as a way to help the scalability of video delivery systems [32]. By placing content closer to the consuming clients not only network costs can be curtailed but also the load of streaming servers can be reduced. Various aspects of the use of proxy servers for video objects has been examined in [14, 26, 23, 10]. In [14], the segmentation and caching of streaming objects is proposed and the merging of requests that are temporally related is investigated. This merging idea has been used in [8, 18, 11] to save bandwidth in light of requests for the same video object that arrive closely in time.

The partial caching of two successive intervals of a video stream is proposed in [10] as a way to speed-up the serving of follow-up requests for the the same video object. In [26], the caching of initial frames of a video object is used in a proxy-setting to reduce startup latency. In the same direction, the storage of the bursty parts of a video-stream in a proxy is advocated in [34]; the remaining parts of the video are directly retrieved from the source helping significantly reduce peak bandwidth requirements in the backbone. A caching mechanism for layered encoded multimedia streams is suggested in [23]; the objective of the technique is to selectively deliver stream quality by differentiating on the client network connection. Stream quality differentiation is also addressed in [24], in conjunction with a seamless handoff mechanism for mobile users.

A formal spatiotemporal framework for modeling moving objects and a query language is discussed in [27]. Efficient techniques for indexing moving objects in one and two dimensions are proposed in [12] while in [5] the trade-offs for indexing schemes to answer interval, rectangle, approximate nearest-neighbor, approximate farthest-neighbor and convex-hull queries are examined. The indexing of current and anticipated positions of moving objects in the context of location based services is examined in [25]. Much work has been also reported in data broadcast and dissemination over wireless networks during the last decade [4, 22, 33, 19, 6].

Rate adaptation for wired streaming systems has been extensively studied in [28, 20, 31, 29, 16, 17]. These studies assume an adaptable video encoding system that changes the encoding parameters on the fly based on feedback information about the channel state. The notion of cycle-based operation is used in [13] with cycles being successively alternating between good and bad cycles. Our work differs in that it does not require a prefetching period so that an initial occupancy is built up in the buffer before the playback begins. Also, our algorithm functions in a graceful manner when the base-station load does not allow for aggressive use of channel resources.

6. CONCLUSIONS

In this paper, we address the problem of efficient video delivery in real-time to high-speed users roaming a 3G network. We propose a network of media servers handling the content distribution on top of the mobile environment that closely cooperates with the

base-stations and user-equipment for the provision of continuous video playout. We segment video streams into variable-sized parts according to the user's speed and traversal path through different cells. In this manner, we minimize the transmission costs between media-servers and base-stations as well as the start-up latency experienced by users during handover. We adopt the use of *Video Streaming* modules along with their *Rate Adaptation* elements with the infrastructure of base-stations to ensure smoothness of the playout process. Preliminary experimentation results through simulation show that the proposed scheme rapidly adapts to changes in load conditions at base-stations, thus minimizing the probability of buffer starvation or even session drops. The low complexity of the proposed mechanism makes it suitable for real-time applications.

7. REFERENCES

- [1] 3rd Generation Partnership Project. Universal Mobile Telecommunication System/IMT-2000 Spectrum. Technical Report 6, *UMTS Forum*, 1998.
- [2] 3rd Generation Partnership Project. Stage 2 Functional Specification of Location Services in UTRAN. Technical Report (3G TR 25.923 version 1.4.0), *UMTS Forum*, 1999. Technical Specification Group(TSG) RAN, Working Group 2 (WG2).
- [3] 3rd Generation Partnership Project. Transparent End-to-End Packet Switched Streaming Service (PSS) General Description (Release 4). Technical Report 3GPP-TS-26.233-V4.0.0, *UMTS Forum*, 2000. Technical Specification Group Services and System Aspects.
- [4] S. Acharya, M.J. Franklin, and S. B. Zdonik. Balancing Push and Pull for Data Broadcast. In *Proceedings of SIGMOD 1997*, Tucson, AZ, May 1997.
- [5] P.K. Agarwal, L. Arge, J. Erickson, and H. Yu. Efficient Tradeoff Schemes in Data Structures for Querying Moving Objects. In *12th Annual European Symposium on Algorithms (ESA)*, pages 4–15, Bergen, Norway, September 2004.
- [6] D. Aksoy, M. Altinel, R. Bose, U. Cetintemel, M. Franklin, J. Wang, and S. Zdonik. Research in Data Broadcast and Dissemination. In *Proceedings of International Conf. on Advanced Multimedia Content Processing (AMCP)*, Osaka, Japan, November 1998.
- [7] P. Cao, E. W. Felten, A. R. Karlin, and K. Li. A Study of Integrated Prefetching and Caching Strategies. In *Proceedings of ACM SIGMETRICS Conf.*, pages 188–197, Ottawa, Canada, May 1995.
- [8] S. Chan and F. Tobagi. Caching schemes for distributed video services. In *Proceedings of the 1999 IEEE International Conference on Communications (ICC'99)*, Vancouver, Canada, June 1999.
- [9] S. Choi and K. G. Shin. Predictive and Adaptive Bandwidth Reservation for Hand-Offs in QoS-Sensitive Cellular Networks. In *Proceedings of ACM-SIGCOMM*, pages 155–166, 1998.
- [10] A. Dan and D. Sitaram. A Generalized Interval Caching Policy for Mixed Interactive and Long Video Environments. In *Proceedings of IST/SPIE Multimedia Computing and Networking Conference*, San Jose, CA, January 1996.
- [11] A. Dan, D. Sitaram, and P. Shahabuddin. Dynamic Batching Policies for an On-demand Video Server. *Multimedia Systems*, 4(3):112–121, 1996.
- [12] G. Kollios and D. Gunopulos and V.J. Tsotras. On Indexing Mobile Objects. In *Proceedings of 1999 ACM SIGACT-SIGMOD-SIGART Symposium on Principles of Database Systems (PODS)*, Philadelphia, PA, 1999.
- [13] M. Hassan, L. Atzori, and M. Krunz. Video Transport over Wireless Channels: A Cycle-based Approach for Rate Control. In *Proceedings of the ACM Multimedia 2004 Conference*. ACM Press, October 2004.
- [14] M. Hofmann, E. Ng, K. Guo, S. Paul, and H. Zhang. Caching Techniques for Streaming Multimedia over the Internet. Technical report, Bell Laboratories, April 1999. BL011345-990409-04TM.
- [15] H. Holma and A. Toskala. *WCDMA for UMTS Radio Access for Third Generation Mobile Communications*. John Wiley & Sons Inc., New York, NY, 2nd edition, 2002.
- [16] C.-Y. Hsu, A. Ortega, and A.R. Reibman. Joint Selection of Source and Channel Rate for VBR Video Transmission Under ATM Policing Constraints. *IEEE Journal of Selected Areas in Communications*, 15(6):1016–1028, 1997.
- [17] P.-C. Hu, Z.-L. Zhang, and M. Kaveh. Channel Condition ARQ Rate Control for Real-time Wireless Video Under Buffer Constraints. In *Proceedings of the IEEE International Conf. on Image Processing*, Vancouver BC, Canada, September 2000.
- [18] K.A. Hua, Y. Cai, and S. Sheu. Patching: a Multicast Technique for True Video-on-demand Services. In *Proceedings of the 6th ACM International Conference on Multimedia*, pages 191–200. ACM Press, 1998.
- [19] T. Imielinski, S. Viswanathan, and B. R. Badrinath. Data on Air: Organization and Access. *IEEE Transactions on Knowledge and Data Engineering*, (3):353–372, 1997.
- [20] H.-J. Lee, T. Chiang, and Y.-Q. Zhang. Scalable Rate Control for MPEG-4 Video. *IEEE Transactions On Circuits and Systems for Video Technology*, 10(9):878–894, September 2000.
- [21] S.-J. Lee, W.-Y. Ma, and B. Shen. An Interactive Video Delivery and Caching System Using Video Summarization. *Computer Communications*, 25(4):424–435, March 2002.
- [22] E. Pitoura and P.K. Chrysanthis. Multiversion Data Broadcast. *IEEE Transactions on Computers*, 51(10):1224–1230, 2002.
- [23] R. Rejaie, H. Yu, M. Handley, and D. Estrin. Multimedia Proxy Caching Mechanism for Quality Adaptive Streaming Applications in the Internet. In *Proceedings of INFOCOM(2)*, pages 980–989, 2000.
- [24] S. Roy and B. Shen and V. Sundaram. Application Level Handoff Support for Mobile Media Transcoding Sessions. In *12th International Workshop on Network and Operating System Support for Digital Audio and Video*, Miami, FL, 2002.
- [25] S. Saltinis and C.S. Jensen. Indexing of Moving Objects for Location-Based Services. In *Proceedings of the IEEE International Conference on Data Engineering (ICDE)*, pages 463–472, 2002.
- [26] S. Sen, J. Rexford, and D. F. Towsley. Proxy Prefix Caching for Multimedia Streams. In *Proceedings of INFOCOM(3)*, pages 1310–1319, New York, NY, 1999.
- [27] A.P. Sistla, O. Wolfson, S. Chamberlain, and S. Dao. Modeling and Querying Moving Objects. In *Proceedings of the 13th IEEE International Conf. on Data Engineering*, Birmingham, UK, April 1997.
- [28] H. Song and C.-C.-J. Kuo. Rate control for Low Bit Rate Video via Variable Encoding Frame Rates. *IEEE Transactions On Circuits and Systems for Video Technology*, 11(4):512–521, April 2001.
- [29] W. Tawbi, F. Horn, E. Horlait, and J.-B. Stefani. Video Compression Standards and Quality of Service. *The Computer Journal*, 36(1):43–54, 1993.
- [30] Texas Instruments. Mobile Connectivity: Assisted-GPS. <http://focus.ti.com/general/docs/wtbu>, 2004.
- [31] T. Wiegand, M. Lightstone, T. Campbell, D. Mukherjee, and S. Mitra. Rate-Distortion Optimized Mode Selection for Very Low Bit Rate Video Coding and the Emerging H.263 Standard. *URL: citeseer.ist.psu.edu/wiegand95ratedistortion.html*, 1999.
- [32] D. Wu, Y.T. Hou, W. Zhu, Y.-Q. Zhang, and J.M. Peha. Streaming Video over the Internet: Approaches and Directions. *IEEE Transactions on Circuits and Systems for video Technology*, 11(1):1–20, February 2001.
- [33] X. Yang and A. Bouguettaya. Broadcast-Based Data Access in Wireless Environments. In *Proceedings of the EDBT International Conference*, Prague, Czech Republic, 2002.
- [34] Z.-L. Zhang, Y. Wang, D.H.C. Du, and D. Shu. Video staging: a proxy-server-based approach to end-to-end video delivery over wide-area networks. *IEEE/ACM Transactions on Networking*, 8(4):429–442, 2000.