A Mobile Proxy Architecture for Video Services Over High-Speed Rail Environments in LTE-A Networks

Ming-Chin Chuang and Meng Chang Chen

***Abstract*—With the rapid developments in wireless communication technology, people access Internet applications anytime and anywhere. Recently, academics and industry researchers have been paying increasing attention to the video streaming services over high-speed rail environments. For providing the quality of service of video streaming, we present a novel mobile proxy architecture for improving the system performance over high-speed trains in Long Term Evolution–Advanced networks. The mobile proxy not only saves the wireless bandwidth consumption between the base station and the high-speed train but also reduces the starting delay time of the video. Moreover, we propose a score-based replacement scheme in the memory management of the proxy server based on the feature of the video. Our simulation results demonstrate that the proposed scheme provides a better performance than other existing cache management schemes.**

***Index Terms*—High-speed rail, mobile proxy, score-based replacement (SBR), video services.**

I. INTRODUCTION

RECENTLY, the high-speed rail system has been becoming increasingly popular, because it is convenient and comfortable to passengers. Many people on the train access the Internet during their journey. According to Cisco's white papers [1], the mobile video (e.g., live TV, video on demand, and video conferencing) will occupy over 70% of mobile data traffic by 2016. Therefore, it becomes a critical demand that the high-speed rail system provides the video services for the passengers.

Nowadays, the 3GPP Long Term Evolution–Advanced (LTE-A) [2] is an innovative wireless access technique. The LTE physical layer uses modulation and coding scheme (MCS), multiple-input–multiple-output (MIMO), coordinated multipoint transmission/reception (CoMP), and carry aggregation (CA) schemes to support high data rate at speeds up to 350 km/h and even 500 km/h in rural areas. As a result, it is a feasible solution in practice that the Internet service provider (ISP) uses the LTE-A technique to support the video streaming services (e.g., video on demand) for the passengers. Fig. 1 shows the scenario that the passengers (PAs) access the video services over high-speed train in LTE-A networks.

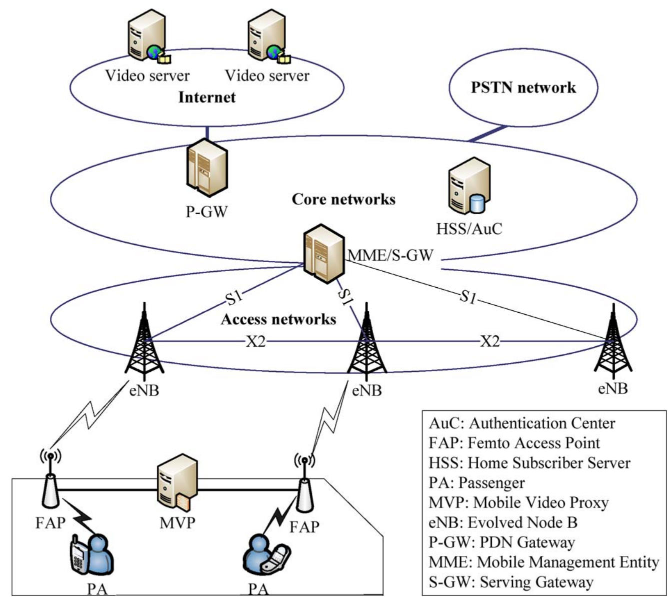


Fig. 1. Passengers access the video services over high-speed train in LTE-A networks.

Unfortunately, the wireless bandwidth of the LTE-A is still not able to support a lot of passengers to access the video services at the same time because the video belongs to the huge data that need more bandwidth to transmit it. Therefore, it calls for a good mechanism for supporting better video quality and admitting more requests of passengers. In this paper, we present a novel mobile proxy architecture for improving the system performance over high-speed trains in LTE-A networks. The mobile proxy not only saves the wireless bandwidth consumption between the base station and the high-speed train but also reduces the start-up waiting time of the video. In addition, we propose a score-based replacement (SBR) scheme in the memory management based on the feature of the video. Moreover, we show that the SBR scheme has the higher byte hit rate and the lower starting delay time.

The rest of this paper is organized as follows. Section II describes some backgrounds and related work. Section III shows the system architecture, and Section IV presents the proxy operations in detail. In Section V, we show the performance of the proposed scheme. Finally, Section VI summarizes this paper.

Manuscript received November 4, 2013; revised February 26, 2014 and July 20, 2014; accepted August 28, 2014. Date of publication September 17, 2014; date of current version November 20, 2015.

The authors are with the Institute of Information Science, Academia Sinica, Taipei 115, Taiwan (e-mail: speedboy@gmail.com; mcc@iis.sinica.edu.tw).

Digital Object Identifier 10.1109/JSYST.2014.2354435

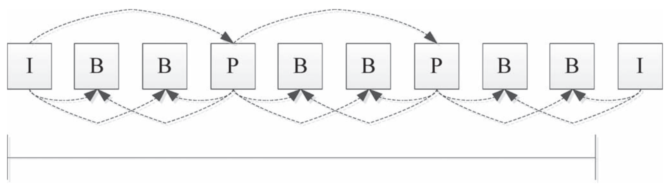


Fig. 2. Typical MPEG-4 GoPs.

II. Background

A. MPEG-4 Overview

MPEG-4 is a mature technique for compressing the audio and visual digital data, such as streaming media, videophone, and broadcast television applications. It was introduced in late 1998 and designated a standard for a group of audio and video coding formats. In MPEG-4 video coding [8], the video would be divided into a series of consecutive group of pictures (GoPs) when it is delivered in networks. The GoPs include three kinds of frames, namely, intracoded (I), bipredictive (B), and predicted (P) frames. The I frame is the main frame in the video and coded independently of any of the other frames in the GoP, such as a starting point for the next sequence of frames. The quality of the video will decrease significantly if the I frame is lost. This frame is also used to resynchronize the entire scene at the receiver. The B frame provides the highest level of compression, and it is predicted by using both the forward and backward directional changes in motion. The P frame is compressed and used as a reference point for B frames. The compression of the P frame is based on the motion-compensated prediction. Fig. 2 shows the sequence numbers of video frames in typical MPEG-4 GoPs. Recently, studies [27], [28] have aimed at the mobile video (e.g., MPEG-4 and H.264/AVC) to design the enhanced resource allocation schemes. Tamimi *et al.* [27] proposed a simple seasonal ARIMA model (SAM) for mobile video, and then, a video frame generator has been developed based on the SAM model for the resource allocation in Worldwide Interoperability for Microwave Access (WiMAX) networks. Sheu *et al.* [28] proposed a multicast resource allocation scheme based on scalable video coding technique to maximize the network throughput in WiMAX relay networks. Wu *et al.* proposed a joint source channel coding scheme [30] for video transmission over bandwidth-limited and error-prone heterogeneous wireless networks. Later, Wu *et al.* proposed a loss-tolerant bandwidth aggregation approach [31] to improve the video quality.

B. Replacement Mechanisms

The proxy plays an essential role in modern network systems to smooth the performance gap between the application server and the user. The replacement policy had a great effect upon the performance of the proxy. Many replacement mechanisms had been proposed [9]–[16], such as least recently used (LRU), least frequently used (LFU), LRU-K, and least recently frequently used (LRFU). The studies [17], [18] collate these replacement mechanisms and show the comparisons of the performance. However, these mechanisms were designed for web access, and they do not take the features of the video into account. Therefore, these replacement algorithms cannot directly be applied to the cache management of the video proxy server. Recently, some researchers have focused on multimedia streaming services, and they also use the proxy server to improve the video quality. Studies [19], [20] use the dynamic segment-based cache replacement schemes for video streaming objects. However, these schemes lack of the prefetching and call admission control mechanisms, resulting in the limited performance improvement.

C. LTE-A Systems for High-Speed Rail

Many studies [21]–[26], [32]–[37] had been implanted the testbed and analyzed the capacity of LTE-A under high-speed rail environment. Tian *et al.* [21], Karimi *et al.* [22], and Pan *et al.* [36] used multiple antennas and relay technique to achieve seamless handover. Moreover, their simulation results show that the high-speed train owns enough bandwidth to support the multimedia services. Song *et al.* [37] investigated the handover problem that the railway communication system based on C/U plane split heterogeneous networks, whose macrocell is based on LTE networks. Studies [23]–[26], [32], [33], [35] investigate the channel model of the high-speed rail under varied environments. These study results can prove that the LTE-A technology supports high data rate transmission in high-speed rail environment even if the received signaling of the train changes fast. Luo *et al.* [34] proposed a position assisted coordinate hybrid automatic repeat request scheme to reduce the Voice-over-Internet Protocol packet error rate in LTE systems for high-speed railway.

III. System Architecture

Fig. 3 shows the proposed mobile video proxy architecture. There are two components (i.e., the call admission control function and the cache management) participating in the video caching process. According to the available resource (i.e., wireless bandwidth, available memory, and system loading), the call admission control limits the number of video requests in order to prevent the network congestion and call dropping. The proxy server uses the cache mechanism to provide the smooth media streams [3], [4], saves the bandwidth consumption, and reduces the request response time.

The proxy server intercepts the PA's request, initiates the request resolution procedure, which is responsible for locating the requested object, and obtains the PA's service class (e.g., free or charge member) and requirements (e.g., high-quality or low-quality video). There are four scenarios that the proxy server delivers the video block, as shown in Fig. 3. In scenario 1 (i.e., Step 1-1 and Step 1-2), the proxy delivers the video block to the PA directly if the video is cached in the memory (i.e., a hit). In scenario 2 (i.e., Step 2-1 to Step 2-3), when the object is not stored in the memory (i.e., a hit miss), it is loaded from the secondary storage and then sent to the PA. In scenario 3 (i.e., Step 3-1 to Step 3-4), the proxy does not have the requested object and thus forwards the request to the remote content server, downloads the object in the secondary storage, loads the object in the memory, and transmits the block to the PA. As a

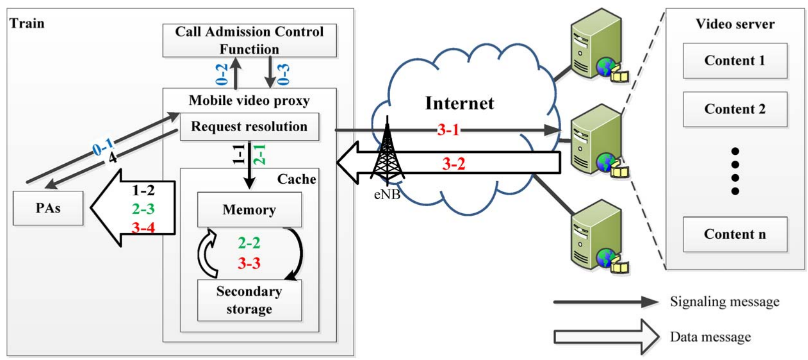


Fig. 3. Diagram of the mobile video proxy architecture.

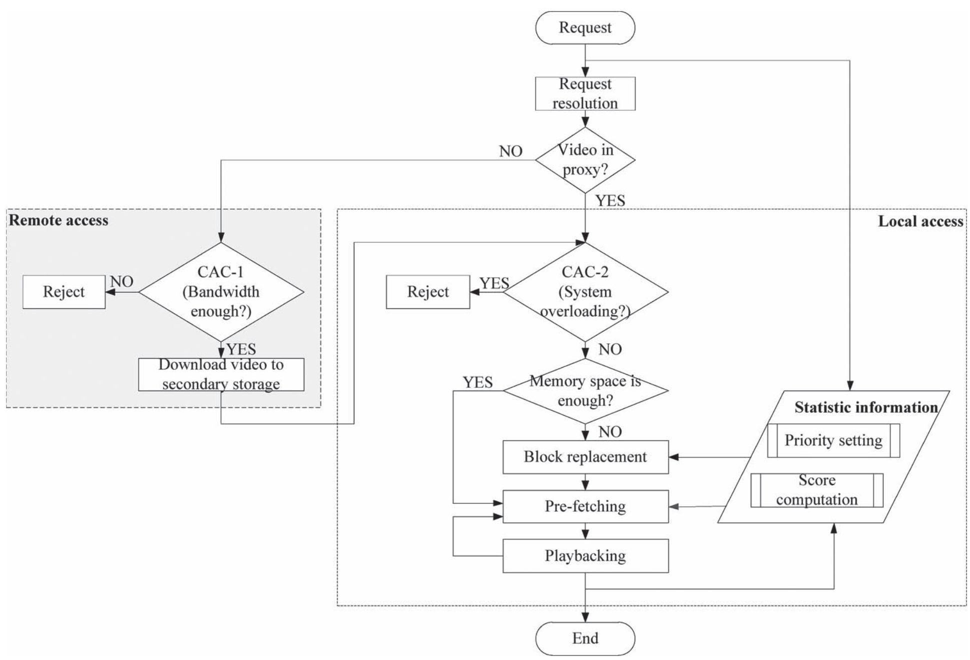


Fig. 4. Flowchart of the mobile proxy operations.

result, the access cost is scenario 1 < scenario 2 < scenario 3. To ensure quality of service (QoS), a call admission control function is applied [5]. Scenario 4 is that the proxy rejects the PA's request because of insufficient resource. As the secondary storage is generally large enough (e.g., a 2-TB storage can keep about 1100 1-h-long 4-Mb/s video objects) to keep the downloaded popular video clips such as primetime news and sport event broadcasts of the day, the proposed secondary storage management is simple but effective that it keeps all the downloaded video objects and uses LRU policy for replacement.

IV. Mobile Proxy Operations

In order to reduce the consumption of the network bandwidth, the proxy architecture is introduced. Moreover, many cache replacement algorithms (e.g., LRU, LFU, and LRFU) in the proxy have been proposed for web browsing to save the bandwidth consumption and reduce the response time. However, these proxy cache management schemes focusing on web traffic are not appropriate for video access on a high-speed train due to the unique characteristics of mobile video access, e.g., large data size, sequential playing, and unreliable transmission. Here, we describe three important observations of the features of video objects and propose a memory management scheme to take advantage from the observations.

Observation 1: As generally a video object is too large to be completely loaded in the memory, we partition a video object into several blocks and consider a block as the basic loading unit.

Observation 2: As generally user starts from the beginning and views a video object continuously, if a block is playbacked, then all other following blocks are likely to be playbacked, and the closer the blocks are, the more probable they are to be playbacked.

Observation 3: Each video frame has a different weight. In MPEG-4 video streaming, the importance of the frame is shown as follows: I frame > P frame > B frame. This means that the I frame has the greatest effect upon the video quality.

This section describes the mobile proxy operations in detail, particularly the memory replacement mechanism. Fig. 4 shows the flowchart of the mobile proxy operations, which include request resolution, call admission control, memory block allocation, prefetching, and SBR phases.

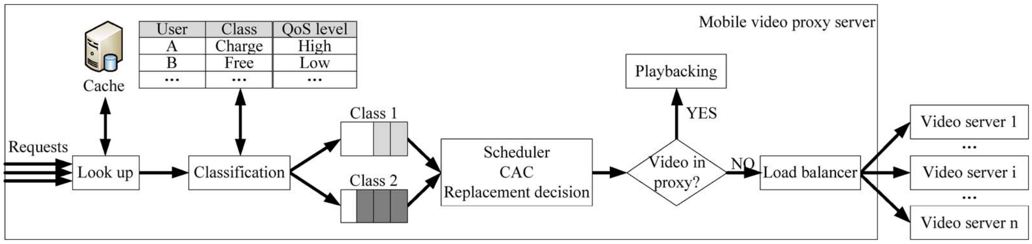


Fig. 5. Request resolution operations of the video proxy server.

A. Request Resolution

The video proxy server intercepts the PA's request when the passengers connect to the Internet and transmit the requests through the femto access point (FAP). Then, the mobile video proxy server initiates a request resolution procedure, which is responsible for locating the requested object. Moreover, the proxy obtains the PA's parameters (i.e., service classes and requirements) from the request message, builds a user priority table, classifies the categorization of passengers (e.g., charge or free member), schedules the requests, and transmits these parameters for the call admission control and the replacement decision. Finally, the video proxy server forwards the requests to the remote video servers if the requested objects are not stored in the cache. The goal of the load balancer is to distribute the workload of the remote video servers. On the other hand, the mobile video proxy server will play back the video stream for PAs if the requested objects are stored in the cache. Fig. 5 shows the request resolution operations of the video proxy server.

B. CAC

The call admission control (CAC) procedure is performed after the request resolution procedure. Generally, the call admission control function has two considerations: the remainder wireless bandwidth and the QoS provided to the user. For interactive video (e.g., video conferencing), the proxy rejects the request if the video transmission bit rate from the video server to the proxy (note that the bottleneck link is generally the wireless link from the ground to the train) is lower than the playbacking bit rate. For noninteractive video (e.g., video on demand), the proxy rejects the request if it cannot meet the predefined QoS requirement [6]. In our architecture, CAC-1 and CAC-2 are performed at the eNB and the mobile video proxy server, respectively. We describe these two CAC mechanisms as follows.

1) Requested Video Is Stored in the Remote Video Server: In this situation, the eNB first executes the CAC-1 procedure, and then, the mobile video proxy server executes the CAC-2 procedure if the CAC-1 procedure is passed. Usually, the bottleneck will happen between the eNB and the train during the transmission process since the wireless bandwidth is not enough. Therefore, the CAC-1 procedure in the eNB focuses on the admission control of the bandwidth. Because each video frame has different importance according to Observation 3, the eNB first transmits the high-priority frame.

2) Requested Video Is Stored in the Mobile Video Proxy Server: The CAC-2 procedure is performed at the mobile video proxy server if the requested video is stored or downloaded in the cache. If the proxy server is overloading (i.e., no free memory), the proxy denies the request. In addition, the CAC function notifies the proxy to retrieve the allocated blocks of the low-level user if a new high-level request arrives and the system resource is not enough.

C. Memory Block Allocation

As aforementioned, the video object is segmented into several blocks, and only a partial block is loaded in the memory for playbacking. For a video object access request, the proxy system provides the corresponding QoS and assigns a certain number of memory blocks with different block sizes according to the PA's service classes and video characteristics [3], [4], [6]. A high-level class service (e.g., charge member) receives high-quality video service, which generally means having more memory blocks of large block size. Moreover, the video is playbacked smoother [3] if a video request acquires more number of blocks.

For example, in Youtube, for a video with MPEG-4 coding, a low-quality (i.e., 640×360 pixels per frame) video needs a rate of 1 Mb/s, and a high-quality (i.e., 1920×1080 pixels per frame) video needs a rate of 6 Mb/s. If a low-level class request is viewing a low-quality Youtube video, it is assigned 1 Mb/s with two 1.25-MB memory blocks, each for 10-s playback; in contrast, if a high-level class request is viewing a high-quality Youtube video, it is assigned 6 Mb/s with three 7.5-MB memory blocks that can endure up to further packet delay of 10 s.

D. Prefetching

In order to save bandwidth and reduce service latency, the proxy needs efficient prefetching algorithm to maintain high byte hit rate. When the proxy admits a request, it allocates at least two blocks of memory for the request. One block is for storing the current playbacking video object, and the other block buffers the video for later playback (i.e., for prefetch). According to Observation 2, the proxy also schedules the block prefetching requests based on the early deadline first (EDF) scheduling algorithm [7] to meet the QoS requirements (e.g., playback delay and jitter) and fully utilize the available bandwidth and the system resource. In addition, the I frame is prior to B and P frames according to Observation 3.

E. SBR Mechanism

When the free space in cache memory is insufficient to accommodate a new non-video-conferencing request, a replacement policy determines which blocks should be replaced. Moreover, the replacement mechanism is the most important part in the proxy system because it affects the overall system performance. The traditional cache replacement mechanisms are not suitable for the video proxy server because they do not take the properties of the video caching into account. Here, we propose an SBR mechanism to improve the performance of the proxy, such as the byte hit rate and the start-up waiting time. Our replacement scheme not only considers the frequency and recency but also takes the popularity trend of the video into account. We assume that the users' access behaviors are following the mentioned observations. The priority setting and the popularity score computation of video are obtained from the periodical statistic information.

1) Priority Setting Phase: The video is classified into two kinds in the memory: high priority and low priority. The video is considered high priority if the video is being playbacked by at least one user, which means the video is likely to be playbacked repeatedly during the journey and deserves being allocated more blocks. On the other hand, the video is deemed low priority if there is no request for this video. Hence, low-priority video is replaced prior to high-priority video.

2) Popularity Score Computation of Video Phase: After the priority classification, the proxy compares the video's popularity score and then decides the candidate video to be reclaimed. The popularity trend of video indicates the likelihood of revisit. The proxy periodically records the number and time of video playback to compute a score of popularity of each video. Before describing the score computation, we first define the relevant notations. Let t_i be the observation time and c_i be the total number of hits between the t_0 and t_i, where $i = 0, \ldots, n$. The proxy then computes the score of each video. The replacement order is determined according to the score of the video if the videos have the same priority. The score of the video (i.e., VS_i) exploits the exponential moving average, expressed as follows:

$$\mathrm{VS}_i = (1-\alpha)^* f_{\mathrm{AVG}}(c,t) + \alpha^* f_i(\Delta c, \Delta t), 0 < \alpha < 1 \quad (1)$$

where α is a weight factor, $f_{\mathrm{AVG}}(c,t)$ is the total number of hits from c_0 to c_i divided by the observation time from t_0 to t_i (i.e., $(c_i - c_0)/(t_i - t_0)$), and $f_i(\Delta c, \Delta t)$ is the total number of hits from c_{i-1} to c_i divided by the observation time from t_{i-1} to t_i (i.e., $(c_i - c_{i-1})/(t_i - t_{i-1})$).

We further observe the trend of the video popularity, which can be described in three situations.

1) $f_{\mathrm{AVG}}(c,t) > f_i(\Delta c, \Delta t)$: This means that the popularity of the video is decreasing.
2) $f_{\mathrm{AVG}}(c,t) = f_i(\Delta c, \Delta t)$: This means that the popularity of the video is unchanged.
3) $f_{\mathrm{AVG}}(c,t) < f_i(\Delta c, \Delta t)$: This means that the popularity of the video is increasing.

In [29], the popularity distributions within videos accessed at an entertainment video-on-demand server were found to obey a k-transformed Zipf-like distribution. Therefore, the proxy can dynamically adjust the value of α according to the suitable Zipf-like function. When the score is below a threshold value, the memory blocks are the candidate to be reclaimed by the caching system. Moreover, the video is allocated more blocks in the initiation of the prefetching phase if it has higher VS_i (i.e., this is a popular video).

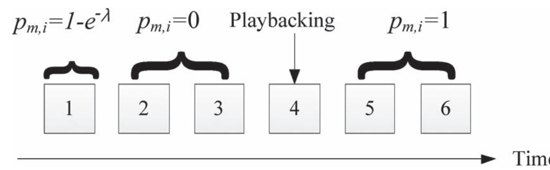


Fig. 6. Probability of future use of the block.

3) Block Replacement Phase: The block replacement phase is performed when a new request arrives and the available memory blocks cannot satisfy the request. The proxy starts searching replacement candidate block from the video with lowest popularity scores to high scores until it finds sufficient replacement blocks. If the proxy fails to find one, the new video request will be rejected.

In lowest score video, the proxy needs to compute the score of each block (i.e., $\mathrm{BS}_{m,i}$) and then reclaims the lower score blocks. If many memory blocks have the same score, the proxy reclaims the block from the biggest serial number block to the smallest one. The score of the block considers the service class (or fee), the block size, and the probability of future use. Note that the higher user service class needs to pay more fees. The score of the block is computed as follows:

$$\mathrm{BS}_{m,i}(N) = \frac{p_{m,i} \times \sum_{j=1}^{N} f_{m,i}}{s_{m,i}} \quad (2)$$

where $\mathrm{BS}_{m,i}(N)$ is the score function of block i of movie m, $p_{m,i}$ is the probability of future use of this block i of movie m, $s_{m,i}$ is the size of block i of movie m, N is the clicked number for this block i of movie m at the moment, and $f_{m,i}$ is the fee of the block i of movie m.

About the $p_{m,i}$, we use an example to explain, as shown in Fig. 6. Block 4 is playbacking, and the $p_{m,i}$ of blocks 5 and 6 are equal to 1 according to Observation 2. The $p_{m,i}$ of blocks 2 and 3 are 0 since we assume that the video starts from the beginning. Because of the user's random visit, it is assumed that request arrivals follow a Poisson process with mean rate λ. Therefore, the $p_{m,i}$ of the first block is $1 - e^{-\lambda}$.

Finally, we propose a multistage greedy method to acquire the maximum memory utilization (i.e., the sum of block score is maximal in the memory). The priority of the video is the first-stage filter. The low-priority video is first replaced because there is no request for this video. Then, the comparison of the video score (i.e., VS_i) is the second-stage filter. The video score represents the popularity of a video. In order to maximize the memory utilization, the system needs to select the video with lower score to reclaim its unused memory blocks. We can guarantee the playing and popular videos in the memory after the proxy performs the previous two stages. Next, the memory of the proxy keeps these blocks that have the higher block score. As a result, the proxy can obtain the maximum sum of block score (i.e., $\mathrm{MAX}\Sigma BS_{m,i}(N)$). The proposed algorithm is shown as follows. Note that the block score function and the prefetching mechanisms need to be redesigned if the behaviors of audiences do not follow Observation 2.

Algorithm 1: Maximum Memory Utilization Algorithm

Input: a new request arrives
Output: $MAX\Sigma BS_{m,i}(N)$ //the maximum sum of the block score in the memory
While (1) //Proxy waits to receive the new request
 If (each video only allocates two blocks and videos are playing)
  The system is overloading;
  Reject this request;
 Else
  Continue the procedure;
 End if
Stage 1: Priority comparisons
 If (low-priority video in memory)
  Replace the low-priority video;
 Else
  Execute Stage 2;
 End if
Stage 2: Video score comparisons
 Compute the score of each video;
 Sort the video score;
 If (score of each video is different)
  Select the lowest score video to reclaim its unused blocks;
 Else
  Select the video with decreasing popularity to reclaim their unused blocks; //$f_{AVG}(c,t) > f_i(\Delta c, \Delta t)$
 End if
Stage 3: Block score comparisons
 Compute the score of each block;
 Sort the block score;
 If (score of each block is different)
  Select the smallest block score to be a candidate;
 Else
  Select the block from the biggest serial number block to the smallest one;
 End if
 If (new $MAX\Sigma BS_{m,i}(N)$ > original $MAX\Sigma BS_{m,i}(N)$)
  Execute the replacement;
 Else
  Reject;
 End if
 Update $MAX\Sigma BS_{m,i}(N)$;
End while

V. Performance Evaluations

Here, we present evaluations of the cache management scheme of the proxy. We implemented popular replacement algorithms, which are the first-in–first-out (FIFO), LFU, LRU, and segment-based schemes, for the performance comparison with the proposed SBR scheme. LFU scheme counts how often an item is needed, and it discards the item with the lowest reference frequency. LRU scheme discards the least recently used items first.

TABLE I
Simulation Parameters

Parameter	Value
Active users	100
Transmission range of cell	10 km
Velocity of train	100~400 km/hr
Transmission data rate	100 Mbps
Number of videos	100
Number of requests	500
Handover failure rate	1%
The total number of requests	500
Memory size	2~16 GB

A. Performance Metrics

We define three metrics to evaluate the performance: 1) byte hit ratio; 2) start-up waiting time; and 3) average peak signal-to-noise ratio (PSNR).

1) Byte hit ratio: A byte hit ratio is the ratio of the volume of data that are encountered in the memory to the entire volume of data requested. Note that the unit of data is in bytes.
2) Start-up waiting time: This duration time means that the PA sends the server a request until his device plays back the video. Generally, the PA feels higher user satisfaction when the start-up waiting time is lower.
3) Average PSNR: PSNR is calculated by comparing the difference in quality between original videos and received videos. Higher PSNR means higher video quality. The PA obtains a low-quality video if several frames (i.e., I, B, and P frames) are lost. PSNR is defined via the mean square error (MSE). Given a noise-free $m \times n$ monochrome image I and its noisy approximation K, MSE is defined as

$$\text{MSE} = \frac{1}{mn}\sum_{i=0}^{m-1}\sum_{j=0}^{n-1}[I(i,j) - K(i,j)]^2.$$

Then, the PSNR is computed as

$$\text{PSNR} = 10\log_{10}\left(\frac{\text{MAX}_I^2}{\text{MSE}}\right) = 20\log_{10}(\text{MAX}_I) - 10\log_{10}(\text{MSE})$$

where MAX_I is the maximum possible pixel value of the image. Therefore, the average PSNR is computed as the sum of the PSNR of received videos divided by the sum of the PSNR of original videos.

B. Simulation Setting

We use C++ to implement the SBR scheme based on the code of Karimi *et al.* [22]. Their code is publicly available at http://www.sfu.ca/~oba2/dls/. In addition, the access pattern of the video follows the proposed observations. There were 100 constant bit rate (CBR) video objects in the simulated environment, which includes high- and low-quality videos, and

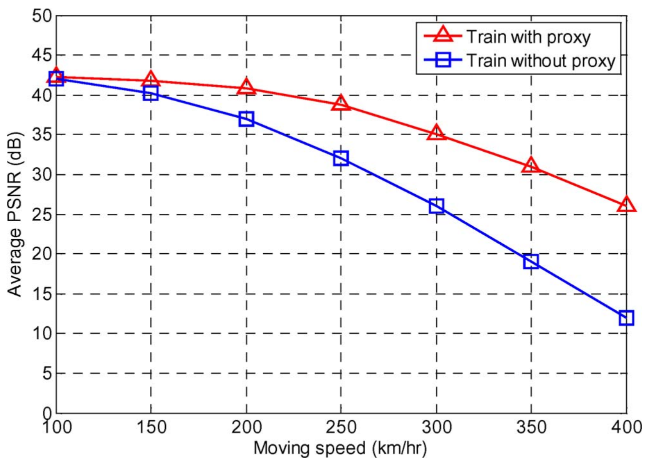


Fig. 7. Variation of PSNR under different moving speeds.

each video is segmented into 100 blocks. The total number of requests is 500 with arrival rate of 1 request per second. Table I shows the simulation parameters.

C. Simulation Results

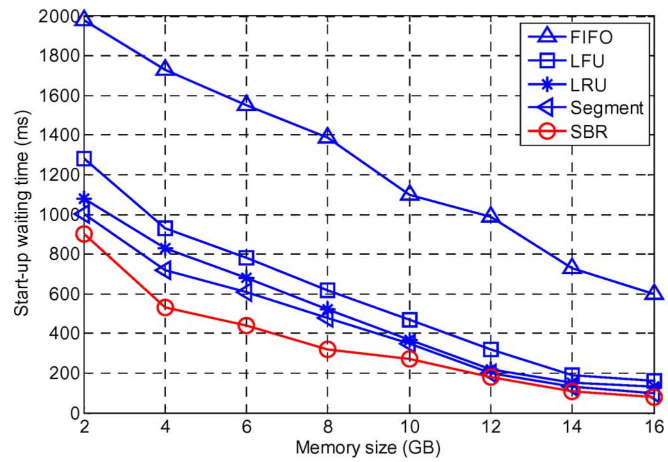
Fig. 7 shows the average PSNR of the video under different train speed environments. We can see that the Doppler effects are very severe, which cause the degradations of PSNR of the video when the speed of the train increases. However, the receiver has the better video quality when the train equips the mobile video proxy server, because the proxy server over the train performs the prefetching mechanism, CAC procedure (i.e., the frame of the highest priority will be transmitted first), stores the downloaded videos, and plays the buffer role to adapt the playback of the video. Therefore, the PSNR of the video degrades slowly when the wireless link between the train and the eNB is unstable. Oppositely, the PSNR of the video degrades quickly if the train does not equip the mobile video proxy server.

Fig. 8(a) and (b) depicts the start-up waiting time for the user. We can see that the larger memory size has higher byte bit ratio to decrease the start-up waiting time, and more requests occur higher probability to download the video from the remote video server, resulting in the higher start-up waiting time. Because the FIFO scheme does not consider the access habit of PA and the character of the video, this scheme has the worst result.

The other schemes, which consist of LFU, LRU, Segment, and SBR, improve the performance of the start-up waiting time. As a result, our replacement scheme has the best performance because SBR uses the prefetching scheme to further shorten the start-up waiting time.

Fig. 9 shows the performance of the byte hit rate with different mixed quality videos. We can see that SBR has the highest byte hit rate in all cases.

This resulted because our scheme not only considers the replacement procedure, which simultaneously considers the frequency, recency, and popularity trend of the video, but also takes the prefetching procedure based on the EDF algorithm into account. On the other hand, LRU does not consider the frequency and viewing sequence properties of blocks, and LFU does not consider the popularity trend of the video, resulting in a cache pollution problem.



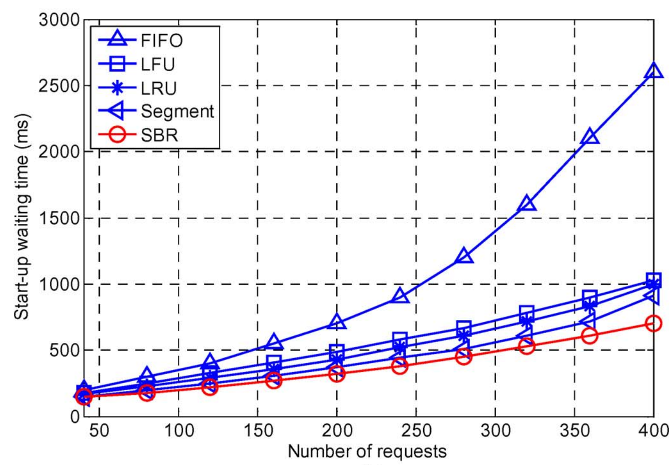


Fig. 8. Start-up waiting time of different replacement schemes. (a) Memory of the proxy server is from 2 to 16 GB. (b) Number of requests is from 40 to 400.

Table II shows the performance of the average PSNR when the moving speed of the train is 300 km/h. The receiver cannot successfully decode the video, resulting in the quality of video decreasing if he/she only receives the B or P frames. Once again, our scheme has the best result because the most important frame (i.e., I frame) is always transmitted and recovered first. On the other hand, the other schemes do not take the character of the video (i.e., Observation 3) into account. Therefore, they obtain the lower average PSNR when the packets are lost.

VI. Conclusion

In this paper, we have presented a mobile proxy architecture for video services over high-speed trains in LTE-A networks. According to the video features, we propose an SBR mechanism for the proxy system. Moreover, we use the prefetching scheme to improve the performance of the byte hit rate. The SBR scheme not only considers the frequency and recency but also takes the popularity trend of the video into account. The simulation results show that our SBR outperforms other replacement schemes in terms of the byte hit rate, the start-up waiting time, and the average PSNR.

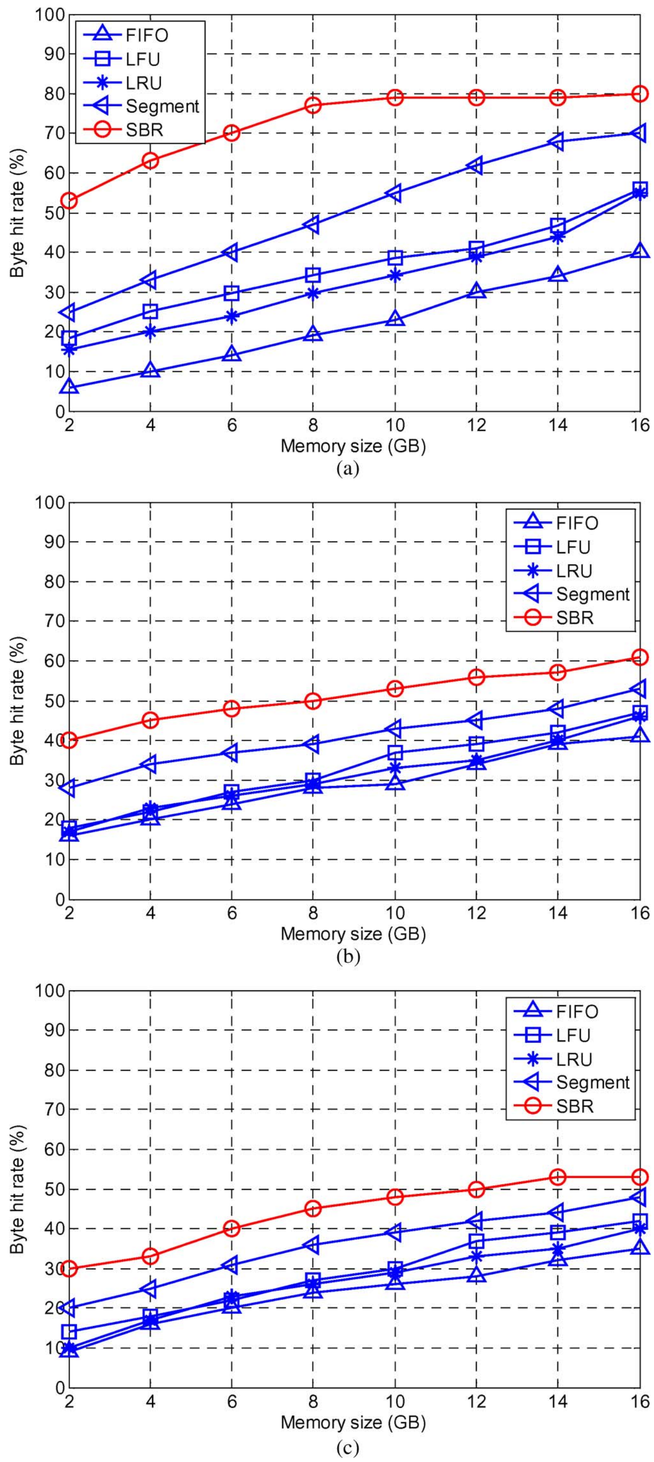


Fig. 9. Byte hit rate of different replacement schemes. (a) All videos are low quality. (b) 50% high-quality videos and 50% low-quality videos. (c) All videos are high quality.

TABLE II
PERFORMANCE OF THE AVERAGE PSNR

Methods	Average PSNR (dB)
FIFO	22.4
LFU	28
LRU	28.7
Segment	32.1
SBR	35.5

TABLE III
ACRONYM TABLE

Abbreviation	Full name
QoS	Quality-of-service
LTE-A	Long term evolution–advanced
MCS	Modulation and coding scheme
MIMO	Multiple input multiple output
CoMP	Coordinated multi-point transmission/reception
CA	Carry aggregation
ISP	Internet services provider
GoPs	Group of pictures
LRU	Least recently used
LFU	Least frequently used
FIFO	First in first out
LRFU	Least recently frequently used
CAC	Call admission control
EDF	Early deadline first
PSNR	Peak signal-to-noise ratio
MSE	Mean squared error
CBR	Constant bit rate
NEMO	Network mobility

In the future, we will consider the mobility management issues of the high-speed train.

1) The mass handover problem occurs when all PAs perform handover from one eNB to another eNB simultaneously. This creates a signaling message storm and generates a large number of processing demands on the wireless link. Therefore, we will integrate the proposed architecture with the network mobility (NEMO) protocol to solve the mass mobility problem.
2) The seamless handover procedure will be proposed for reducing the handover latency and packet loss.

APPENDIX

The acronyms used in this paper are presented in Table III.

REFERENCES

[1] Cisco company, "Cisco visual networking index: Global mobile data traffic forecast update, 2013–2018," Feb. 2014. [Online]. Available: http://www.cisco.com/c/en/us/solutions/collateral/service-provider/visual-networking-index-vni/white_paper_c11-520862.pdf

[2] "Requirements for Further Advancements for E-UTRA (LTE-Advanced)," Sophia-Antipolis Cedex, France, TR 36.913 V8.0.0, Sep. 2012, Release 11.

[3] R.-I. Chang, M. C. Chen, J.-M. Ho, and M.-T. Ko, "An effective and efficient traffic smoothing scheme for delivery of online VBR media streams," in *Proc. IEEE INFOCOM*, Mar. 1999, pp. 447–454.

[4] J. D. Salehi, Z.-L. Zhang, J. Kurose, and D. Towsley, "Supporting stored video: Reducing rate variability and end-to-end resource requirements through optimal smoothing," *IEEE/ACM Trans. Netw.*, vol. 6, no. 4, pp. 397–410, Aug. 1998.

[5] B. Rong, Y. Qian, K. Lu, H.-H. Chen, and M. Guizani, "Call admission control optimization in WiMAX networks," *IEEE Trans. Veh. Technol*, vol. 57, no. 4, pp. 2509–2522, Jul. 2008.

[6] S.-H. Lee, K.-Y. Whang, Y.-S. Moon, W.-S. Han, and I.-Y. Song, "Dynamic buffer allocation in video-on-demand systems," *IEEE Trans. Knowl. Data Eng.*, vol. 15, no. 6, pp. 1535–1551, Nov./Dec. 2003.

[7] C. L. Liu and J. W. Layland, "Scheduling algorithms for multiprogramming in a hard real-time environment," *J. ACM*, vol. 20, no. 1, pp. 46–61, Jan. 1973.

[8] I. E. Richardson, *The H.264 Advanced Video Compression Standard*. Hoboken, NJ, USA: Wiley, Aug. 2010.

[9] A. Dan and D. Towsley, "An approximate analysis of the LRU and FIFO buffer replacement schemes," in *Proc. ACM SIGMETRICS Conf. Meas. Modeling Comput. Syst.*, May 1990, pp. 143–152.

[10] J. T. Robinson and M. V. Devarakonda, "Data cache management using frequency-based replacement," in *Proc. ACM SIGMETRICS Conf. Meas. Modeling Comput. Syst.*, May 1990, pp. 134–142.
[11] E. J. O'Neil, P. E. O'Neil, and G. Weikum, "An optimality proof of the LRU-K page replacement algorithm," *J. ACM*, vol. 46, no. 1, pp. 92–112, Jan. 1999.
[12] D. Lee *et al.*, "LRFU: A spectrum of policies that subsumes the least recently used and least frequently used policies," *IEEE Trans. Comput.*, vol. 50, no. 12, pp. 1352–1361, Dec. 2001.
[13] T. Johnson and D. Shasha, "2Q: A low overhead high performance buffer management replacement algorithm," in *Proc. Int. Conf. VLDB*, Sep. 1994, pp. 439–450, ACM.
[14] N. Megiddo and D. Modha, "ARC: A self-tuning, low overhead replacement cache," in *Proc. 2nd USENIX Conf. File Storage Technol.*, Mar. 2003, pp. 115–130.
[15] S. Jiang and X. Zhang, "Making LRU friendly to weak locality workloads: A novel replacement algorithm to improve buffer cache performance," *IEEE Trans. Comput.*, vol. 54, no. 8, pp. 939–952, Aug. 2005.
[16] S. Jiang, X. Ding, F. Chen, E. Tan, and X. Zhang, "DULO: An effective buffer cache management scheme to exploit both temporal and spatial localities," in *Proc. 4th USENIX Conf. File Storage Technol.*, Dec. 2005, pp. 101–114.
[17] N. Megiddo and D. Modha, "Outperforming LRU with an adaptive replacement cache algorithm," *IEEE Comput.*, vol. 37, no. 4, pp. 58–65, Apr. 2004.
[18] Y. J. Kim and J. Kim, "ARC-H: Adaptive replacement cache management for heterogeneous storage devices," *J. Syst. Archit.*, vol. 58, pp. 86–97, Jan. 2012.
[19] K. L. Wu, P. S. Yu, and J. L. Wolf, "Segment-based proxy caching of multimedia streams," in *Proc. Int. Conf. WWW*, May 2001, pp. 36–44, ACM.
[20] T. H. Hsu and Y. H. Li, "A weighted segment-based caching algorithm for video streaming objects over heterogeneous networking environments," *Exp. Syst. Appl.*, vol. 38, no. 4, pp. 3467–3476, Apr. 2011.
[21] L. Tian, J. Li, Y. Huang, J. Shi, and J. Zhou, "Seamless dual-link handover scheme in broadband wireless communication systems for high-speed rail," *IEEE JSAC*, vol. 30, no. 4, pp. 708–718, May 2012.
[22] O. Baghban Karimi, J. Liu, and C. Wang, "Seamless wireless connectivity for multimedia services in high speed trains," *IEEE JSAC*, vol. 30, no. 4, pp. 729–739, May 2012.
[23] L. Liu *et al.*, "Position-based modeling for wireless channel on high-speed railway under a viaduct at 2.35 GHz," *IEEE JSAC*, vol. 30, no. 4, pp. 834–845, May 2012.
[24] K. Guan, Z. Zhong, and B. Ai, "Assessment of LTE-R using high speed railway channel model," in *Proc. IEEE Int. Conf. CMC*, Apr. 2011, pp. 461–464.
[25] B. Dusza, C. Ide, P.-B. Bok, and C. Wietfeld, "Optimized cross-layer protocol choices for LTE in high-speed vehicular environments," in *Proc. IEEE IWCMC*, Jul. 2013, pp. 1046–1051.
[26] F. J. Martin-Vega *et al.*, "LTE performance over high speed railway channel," in *Proc. IEEE VTC Fall*, Sep. 2013, pp. 1–5.
[27] A. K. AI Tamimi, C. So-In, and R. Jain, "Modeling and resource allocation for mobile video over WiMAX broadband wireless networks," *IEEE JSAC*, vol. 28, no. 3, pp. 354–365, Apr. 2010.
[28] J.-P. Sheu, C.-C. Kao, S.-R. Yang, and L.-F. Chang, "A resource allocation scheme for scalable video multicast in WiMAX relay networks," *IEEE Trans. Mobile Comput.*, vol. 12, no. 1, pp. 90–104, Jan. 2013.
[29] J. Yu, C. Chou, X. Du, and T. Wang, "Internal popularity of streaming video and its implication on caching," in *Proc. IEEE Int. Conf. AINA*, Apr. 2006, pp. 1–6.
[30] J. Wu *et al.*, "Joint source-channel coding and optimization for mobile video streaming in heterogeneous wireless networks," in *Proc. EURASIP J. Wireless Commun. Netw.*, 2013, pp. 1–16.
[31] J. Wu, Y. Shang, B. Cheng, B. Wu, and J. Chen, "Loss tolerant bandwidth aggregation for multihomed video streaming over heterogeneous wireless networks," *Wireless Pers. Commun.*, vol. 75, no. 2, pp. 1265–1282, Mar. 2014.
[32] Z. Min *et al.*, "Analysis and modeling of the LTE broadband channel for train-ground communications on high-speed railway," in *Proc. IEEE VTC Fall*, Sep. 2013, pp. 1–5.
[33] Y. Zhou, Z. Hou, Z. Pan, J. Shi, and J. Wang, "Dynamic Doppler tracking for LTE-based broadband communications on high speed rails," in *Proc. IEEE ChinaSIP*, Jul. 2013, pp. 389–393.
[34] W. Luo, X. Fang, S. Li, and Y. Xia, "Position assisted coordinate HARQ in LTE systems for high speed railway," in *IEEE VTC Spring*, Jun. 2013, pp. 1–5.
[35] N. Imtiaz Bin Hamid, N. Salele, M. T. Harouna, and R. Muhammad, "Analysis of LTE radio parameters in different environments and transmission modes," in *Proc. IEEE Int. Conf. EICT*, Feb. 2014, pp. 1–6.
[36] M. Pan, T. Lin, and W. Chen, "An enhanced handover scheme for mobile relays in LTE-a high-speed rail networks," *IEEE Trans. Veh. Technol.*, May 2014, DOI: 10.1109/TVT.2014.2322374.
[37] H. Song, X. Fang, and L. Yan, "Handover scheme for 5G C/U plane split heterogeneous network in high-speed railway," *IEEE Trans. Veh. Technol.*, Apr. 2014, DOI: 10.1109/TVT.2014.2315231.

Ming-Chin Chuang received the B.S. degree in computer and information science from Aletheia University, Taipei, Taiwan, in 2003, the M.S. degree in computer science and information engineering from Chaoyang University of Technology, Taichung, Taiwan, in 2005, and the Ph.D. degree from National Chung Cheng University, Minxiong, Taiwan, in 2012.

He is currently a Postdoctoral Fellow with the Institute of Information Science, Academia Sinica, Taipei. His research interests include mobility management, network security, cloud computing, and vehicular ad hoc network.

Meng Chang Chen received the Ph.D. degree in computer science from the University of California, Los Angeles, CA, USA, in 1989.

From 1989 to 1992, he was with AT&T Bell Labs, Murray Hill, NJ, USA. He is currently a Research Fellow with the Institute of Information Science, Academia Sinica, Taipei, Taiwan. His current research interests include wireless network, quality of service networking, information retrieval, and data and knowledge engineering.