CONSTRAINED CONSONANT BROADCASTING – A GENERALIZED PERIODIC BROADCASTING SCHEME FOR LARGE SCALE VIDEO STREAMING

W. C. Liu and Jack Y. B. Lee

Department of Information Engineering
The Chinese University of Hong Kong
Shatin, N.T., Hong Kong

ABSTRACT

Video-on-demand (VoD) systems that can serve hundreds to thousands of concurrent users are already widely available. However, to deploy metropolitan-scale VoD services for potentially tens of millions of users, current VoD systems are still limited in capacity, and expensive in cost. To tackle this challenge, this study proposes a new periodic broadcasting scheme, called Constrained Consonant Broadcasting (CCB), for large-scale video streaming. CCB outperforms all existing periodic broadcasting schemes while at the same time addresses two important constraints in practice, namely client access bandwidth and buffer requirements. For example, with a client access bandwidth constraint of twice the video bit-rate, a client buffer of 20% of the video size, and a total system bandwidth equal to six times the video bit-rate, the proposed CCB scheme can reduce the maximum startup latency by 72% and 70% compared to the current state-of-the-art Skyscraper Broadcasting and Greedy Disk-Conserving Broadcasting schemes respectively.

1. INTRODUCTION

Nowadays, video-on-demand (VoD) systems that can serve hundreds to thousands of concurrent users are already widely available. However, to deploy metropolitan-scale VoD services serving potentially tens of millions of users, current VoD systems are still limited in capacity, and expensive in cost. The primary reason is the use of unicast in streaming video data to the clients, making the system cost increases at least linearly with the system scale.

To tackle this challenge, researchers have recently proposed a number of new VoD architectures that make use of network multicast and client-side resources (i.e., bandwidth and buffer) to achieve significant resource savings in large-scale VoD systems [1-5]. Different designs result in different tradeoffs between the three key system resources, namely system bandwidth (i.e., server bandwidth and/or backbone network bandwidth, whichever is smaller), client access bandwidth, and client buffer. More importantly, the resource requirements and performance of these periodic broadcasting systems are independent of the system scale. This open-loop approach differs from close-loop multicast video streaming architectures such as the patching scheme proposed by Hua et. al. [6]. Therefore periodic broadcasting scheme can potentially serve an unlimited number of concurrent users, as long as the network infrastructure can accommodate them.

In this study, we propose a new Constrained Consonant Broadcasting (CCB) scheme that outperforms existing periodic broadcasting schemes. For example, with a client access bandwidth constraint of twice the video bit-rate, a client buffer of 20% of video size and a total system bandwidth equal to six times the video bit-rate, the proposed CCB scheme can reduce the maximum startup latency by 72% and 70% compared to the Skyscraper Broadcasting and Greedy Disk-Conserving Broadcasting schemes respectively, which are among the current state-of-the-art periodic broadcasting schemes.

The rest of paper is organized as follows: In Section 2, we review some previous works on periodic broadcasting. In Section 3, we present the details of the proposed Constrained Consonant Broadcasting scheme. In Section 4, we evaluate and compare the performance of the proposed CCB scheme with other periodic broadcasting schemes. Lastly, we conclude the paper in Section 5.

2. RELATED WORKS

In this section, we review some of the existing periodic broadcasting schemes [1-5]. To facilitate presentation, we summarize in Table 1 the notations used. We review three recent periodic broadcasting schemes in Section 2.1 to Section 2.3, and then present some fundamental theoretical results in Section 2.4.

TABLE I
SUMMARY OF NOTATIONS.

Symbol	Definition
L	The length of the video (sec)
L_i	The size of the i^{th} video segment (sec)
b	The playback rate of the video (Mbps)
N	The total number of video segments
B	The total system bandwidth (Mbps)
C	The client access bandwidth constraint (Mbps)
H	The client buffer constraint (% of video size)
T	The maximum startup latency (sec)

2.1. Skyscraper Broadcasting

Hua et al. proposed the Skyscraper Broadcasting scheme (SB) [2] in 1997 as an improvement to the Pyramid Broadcasting scheme proposed by Viswanathan and Imielinski [1]. Unlike the Pyramid Broadcasting scheme, where the video segments increase in size according to a geometric series, they partition the video according to a pre-defined function. To limit the client buffer requirement, they introduced a system parameter W to limit the size of the video segments to control the client buffer requirement, equal to $L_0b(W–1)$ [2]. The clients are required to download video data from two broadcasting channels simultaneously.

2.2. Greedy Disk-Conserving Broadcasting

Gao et al. proposed the Greedy Disk-Conserving Broadcasting scheme (GDB) [3] in 1998. It is a greedy algorithm that minimizes the number of server channels needed to guarantee a given maximum startup latency T and client I/O bandwidth requirement. Compared with the Skyscraper Broadcasting scheme, GDB has different video partition method and the clients are allowed to download video segments from n–1 broadcasting channels simultaneously, where n is defined as the order of this scheme (denoted as GDBn).

2.3. Poly-harmonic Broadcasting

Paris et. al. proposed the Poly-harmonic Broadcasting scheme (PHB) [5] in 1998 as an improved version of the original Harmonic Broadcasting scheme proposed by Juhn and Tseng in 1997 [4]. Unlike the original Harmonic Broadcasting scheme, the Poly-harmonic Broadcasting scheme guarantees continuous video playback and at the same time can achieve near-optimal performance. In the Poly-harmonic Broadcasting, it broadcasts fixed-size video segment L_i over i^{th} logical channel with $b/(m+i)$ Mbit/sec for i=0,1,…,N–1, where m is a configurable parameter to control the startup latency. Also, the clients must be able to receive video data from all broadcasting channels simultaneously.

2.4. Performance Bounds of Periodic Broadcasting

Common to all periodic broadcasting schemes, the key system parameters are startup latency, system bandwidth, client access bandwidth, and client buffer requirement. Different schemes can be considered as achieving different tradeoffs among these four parameters, and thus the natural question is whether there are any bounds on the system's performance.

This question has been investigated independently by Hu [7], Birk and Mondri [8], and the authors. Although the approaches and the derivations are different, they all arrive at the same results. Specifically, given a startup latency of T, it can be shown that the minimum system bandwidth needed for any periodic broadcasting scheme, is given by $B = b\cdot\ln(L/T+1)$, assuming there is no constraint on the client access bandwidth and client buffer requirement. Additionally, for any optimal periodic broadcasting scheme achieving the above performance bound, the client buffer requirement is bounded by 37% of the video size.

3. CONSTRAINED CONSONANT BROADCASTING

Among the periodic broadcasting schemes reviewed in Section 2, Poly-harmonic Broadcasting can achieve the lowest latency under any given system bandwidth constraint. However, Poly-harmonic Broadcasting scheme does have a shortcoming: client access bandwidth and client buffer requirement. In particular, Poly-harmonic Broadcasting requires a client to be able to receive all broadcasting channels simultaneously and has a buffer large enough to store up to 37% of the whole video. Given the often limited resources at a client, these requirements make Poly-harmonic Broadcasting difficult to implement in practice, despite its near-optimal performance.

Motivated by the above observation, we propose a new Constrained Consonant Broadcasting (CCB) scheme in this study to address the client access bandwidth and client buffer constraints. In CCB, a video title is divided into N equal size segments. The video segments are then broadcast periodically in separate broadcasting channels, with the i^{th} video segment (L_i) broadcast in the i^{th} channel, for i=0,1,…,N–1. Note that as video segments are of the same size and we assume the video is constant-bit-rate encoded, the playback duration for each video segment is the same, say U seconds.

To determine the bandwidth for the broadcasting channels, we need to first set a target latency T in multiples of video segment duration U and the number of segments N in the following equation:

$$T = \frac{m\cdot L}{N} \tag{1}$$

where m is a configurable parameter to tradeoff between system performance and system complexity. Increasing m can reduce the startup latency but more broadcasting channels will be needed. Next, we classify broadcasting channels into two types, namely Type-I and Type-II channels, and define their respective bandwidth partition schemes and client reception schedule in the following sections.

3.1. Type-I Channels

The set of Type-I channels starts with the first channel, with a bandwidth allocation of

$$B_0 = \frac{b}{m} \tag{2}$$

where m is the ratio of T and U, i.e., m=T/U. Subsequent channels are allocated with progressively less bandwidth as given by

$$B_i = \frac{b}{m+i},\quad i = 0,1,\dots,n_1-1 \tag{3}$$

for the i^{th} channel, where n_1 is the total number of Type-I channels. For the Type-I channels, the client is required to start receiving video segments upon entering the system and begin video playback in T seconds.

We can solve for n_1, such that the following constraints are satisfied:

$$\sum_{i=0}^{n_1-1} B_i \le C \quad\text{and}\quad \sum_{i=0}^{n_1} B_i > C \tag{4}$$

$$\max(H_i) \le H \cdot L \cdot b \qquad (5)$$

where $H_i = H_{i-1} - U \cdot b + \sum_{j=i}^{n_1-1} U \cdot B_j,\ i < n_1$ and $H_0 = \sum_{i=0}^{n_1-1} m \cdot U \cdot B_i$.

The constraint in (4) ensures that the total bandwidth required at the client is smaller than the available client access bandwidth *C*. Thus CCB will allocate as many channels as will fit within the available client access bandwidth to maximize its utilization. For the constraint in (5), H_0 computes the amount of video data received in the first *T* seconds while H_i computes the maximum amount of video data accumulated at the client buffer for the duration from $(m+i-1)\cdot U$ to $(m+i)\cdot U$ seconds after the client has entered the system. Thus (5) ensures that the client buffer requirement is not exceeded when receiving Type-I channels.

It is worth noting that if we remove both the client access bandwidth and client buffer constraints, the number of Type-I channels n_1 will be equal to *N*, i.e., all channels are Type-I. In this special case, CCB reduces to Poly-harmonic Broadcasting. This Poly-harmonic Broadcasting can be considered as a special case of CCB without client access bandwidth and client buffer constraints.

3.2. Type-II Channels

Type-II channels are divided into groups of consecutive channels. Channels within the same group have their bandwidth allocated according to (6) and subject to the client access bandwidth and client buffer constraints. The basic idea is that once a client completes receiving a video segment, the corresponding channel will be released. The client access bandwidth released then allows the client to begin receiving a new group of Type-II channels.

It may appear that it is simpler to reallocate all the available bandwidth to a single channel instead of a group of channels. However, doing so is in fact counter productive because there is more than enough time to transmit the latter video segments in the group. By transmitting a video segment in a *just-in-time* manner, we can further reduce the system bandwidth needed.

Let $n_{2,j}$ be the number of channels in group *j*, of which is created after channel *j* is released, where *j*=0,1,…, etc. Then the bandwidth allocation for channels in group *j* is given by

$$B_i = \frac{b}{i-j}, \text{ for } i \ge n_1 \qquad (6)$$

and the number of channels in group *j* can be determined from solving for $n_{2,j}$ in

$$\sum_{i=j+1}^{n_1+n_{2,0}+\ldots+n_{2,j}-1} B_i \le C \text{ and } \sum_{i=j+1}^{n_1+n_{2,0}+\ldots+n_{2,j}} B_i > C \qquad (7)$$

$$\max(H_i) \le H \cdot L \cdot b \qquad (8)$$

where $H_i = H_{i-1} - U \cdot b + \sum_{j=i}^{n_1+n_{2,0}+\ldots+n_{2,i-1}-1} U \cdot B_j$ and $H_0 = \sum_{i=0}^{n_1-1} m \cdot U \cdot B_i$. Similar to the case of Type-I channels, (7) and (8) represent the client access bandwidth and the client buffer constraints respectively.

4. PERFORMANCE EVALUATION

In this section, we evaluate the performance of CCB and compare it to Skyscraper Broadcasting, Greedy Disk-Conserving Broadcasting, and Poly-harmonic Broadcasting. In computing the numerical results, we use a video length of *L*=7200 seconds (2 hours). In computing the results, we applied the optimization procedure proposed by the original studies [2-3,5] to configure the parameters for each broadcasting schemes.

4.1. Startup Latency versus System Bandwidth

Startup latency is defined as the maximum time from a client entering the system to the time video playback starts. With a client access bandwidth of 2*b* and client buffer of 20% of video size, we plot in Fig. 1 the startup latency versus the system bandwidth ranging from 3*b* to 10*b*.

The results in Fig. 1 show that the Poly-harmonic Broadcasting scheme achieves the lowest startup latency, close to the theoretical lower bound (LB) when configured with large value of *m* (e.g. 16). However, unlike CCB, SB and GDB3, we did not apply the client access bandwidth and client buffer constraints in computing results for the Poly-harmonic Broadcasting scheme and thus the results are not directly comparable. Nevertheless, this shows the performance loss due to limited client resources.

Excluding the Poly-harmonic broadcasting scheme, it is clear from the results that the proposed CCB scheme achieves the lowest startup latency. This is true even for *m*=1, which generates the least number of broadcasting channels (and thus lowest system complexity) given the same system parameters. Increasing *m* further reduces the startup latency at the expense of higher system complexity. For a system bandwidth of 6*b*, CCB with *m*=1 achieves startup latency 72% and 70% lower than Skyscraper Broadcasting, and Greedy Disk-Conserving Broadcasting respectively.

4.2. Startup Latency versus Client Access Bandwidth

Fig. 2 plots the startup latency versus the client access bandwidth ranging from 2*b* to 6*b*, where *b* is the video bit-rate. The system bandwidth is equal to 6*b* and the client buffer is unlimited. There are three observations.

First, CCB clearly outperforms the other schemes, especially when the client access bandwidth is low. This is a desirable property as in practice the client access network will likely have much lower bandwidth than backbone networks. Second, the performance of Poly-harmonic Broadcasting, which has been shown to achieve near-optimal performance (without client resource constraints), degrades significantly when the client access bandwidth is limited. This is because Poly-harmonic Broadcasting requires a client access bandwidth that equals to the system bandwidth. Therefore in case the client access bandwidth becomes the bottleneck, the system bandwidth in fact cannot be fully utilized, thus leading to the poor performance.

Finally, we observe that the performance of Poly-harmonic Broadcasting and CCB converge when the client access bandwidth is increased to 6*b*, i.e., same as the system bandwidth. This verifies, as discussed in Section 3, that the Poly-harmonic

Broadcasting is a special case of CCB with the client access bandwidth and client buffer constraints removed.

4.3. Startup Latency versus Client Buffer Requirement

Fig. 3 plots the startup latency versus the client buffer requirement ranging from 10% to 50% of video size. The system bandwidth is equal to 8*b* and the client access bandwidth is equal to 2*b*.

Again, CCB outperforms the other two schemes, especially when the client buffer is small. Moreover, the Poly-harmonic Broadcasting scheme cannot work when the client buffer is reduced to smaller than 37% of the video size.

5. CONCLUSION

In this study, we proposed a new Constrained Consonant Broadcasting (CCB) scheme for large scale video streaming. CCB can be considered as a generalization of the Poly-harmonic Broadcasting scheme incorporating two important constraints, namely client access bandwidth and client buffer requirements. Our results showed that CCB outperforms current state-of-the-art periodic broadcasting schemes, especially when the client access bandwidth and client buffer is low, making it a potential candidate for building cost-effective, large-scale video streaming services.

6. ACKNOWLEDGE

This research is funded by a Direct Grant, and Earmarked Grant (CUHK 4328/02E) from the HKSAR Research Grant Council and the AoE-IT, a research grant from the HKSAR University Grants Council.

7. REFERENCES

[1] S. Viswanathan and T. Imielinski, “Metropolitan Area Video-on-Demand Service Using Pyramid Broadcasting,” *IEEE Multimedia Systems*, vol. 4, 1996, pp.197-208.

[2] K. A. Hua and S. Sheu, “Skyscraper Broadcasting: A New Broadcasting Scheme for Metropolitan Video-on-Demand Systems,” *Proceedings of the ACM SIG-COMM '97*, Cannes, France, Sep 1997, pp.89-100.

[3] L. Gao, J. Kurose, and D. Towsley, “Efficient Schemes for Broadcasting Popular Videos,” *Proceedings of the 8th International Workshop on Network and Operating Systems Support for Digital Audio and Video*, Cambridge, UK, Jul 1998.

[4] L. S. Juhn and L. M. Tseng, “Harmonic Broadcasting for Video-on-Demand Service,” *IEEE Transactions on Broadcasting*, vol.43(3), Sep 1997, pp.268-71.

[5] J. F. Paris, S. W. Carter, and D. D. E. Long, “A Low Bandwidth Broadcasting Protocol for Video on Demand,” *Proceedings of the 7th International Conference on Computer Communications and Networks*, Lafayette, LA, USA, Oct 1998, pp.690-7.

[6] K. A. Hua, Y. Cai, and S. Sheu, "Patching: A Multicast Technique For True Video-on-Demand Services," *Proceedings of the 6th International Conference on Multimedia*, Sep 1998, pp.191-200.

[7] A. Hu, “Video-on-Demand Broadcasting Protocols: A Comprehensive Study,” *Proceedings of the IEEE Infocom 2001*, Anchorage, AK, Apr 2001.

[8] Y. Birk and R. Mondri, “Tailored Transmissions for Efficient Near-Video-on-Demand Service,” *Proceedings of the IEEE International Conference on Multimedia Computing and Systems*, Florence, Italy, Jun 1999.

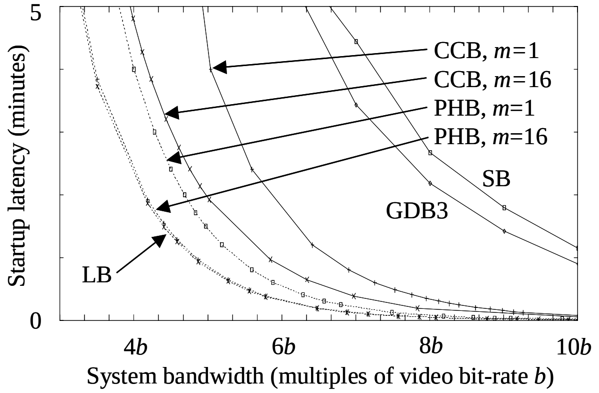

Figure 1. Startup latency versus system bandwidth (client access bandwidth = 2*b*, client buffer = 20% of video size)

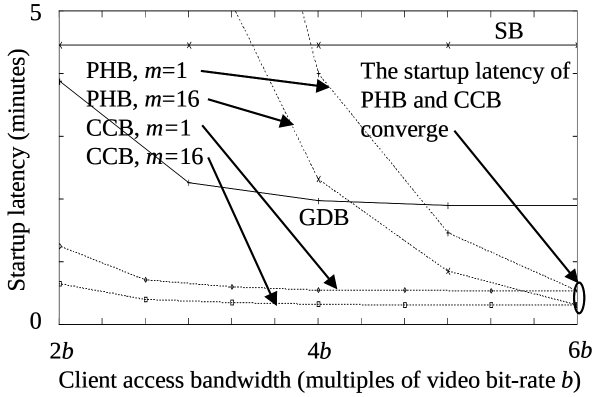

Figure 2. Startup latency versus client access bandwidth (system bandwidth = 6*b*, client buffer = unlimited)

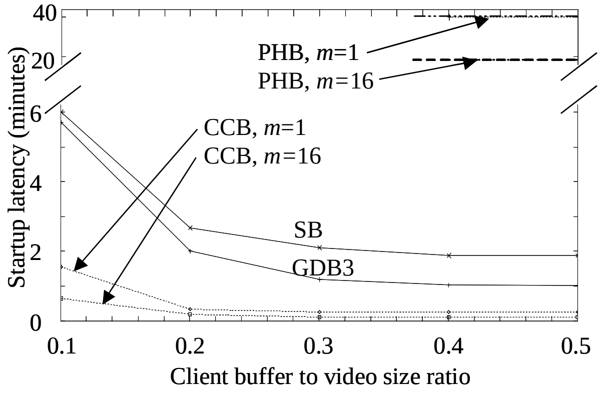

Figure 3. Startup latency versus client buffer requirement (system bandwidth = 8*b*, client access bandwidth = 2*b*)