Resource Allocation for Stored Video Delivery with Rate Availability Function

Ray-I Chang
Dept. of Inform. Management.
National Central University
Taiwan, ROC

Yu-Lin Chiu
Center of Inform. Management
CS Inst. of Science and Tech.
Taiwan, ROC

Jan-Ming Ho
Institute of Inform. Science
Academia Sinica
Taiwan, ROC

(The contact author is Ray-I Chang. Email: rchang@mgt.ncu.edu.tw, TEL: 03-4227151)

Abstract

Resource allocation for serving real-time video data with a significant multiple-time-scale bit-rate variability is complicated. In past years, different approaches were proposed to shape VBR (variable-bit-rate) video streams as smooth transmission schedules for efficient delivery. In [1], a linear-time method was proposed to construct transmission schedules with not only minimal client buffer and delay but also maximal bandwidth utilization. However, as many previous methods, it only considered the peak rate of schedule in resource allocation. In a real network, its available rate is time-varying and usually represented by a piecewise-linear function called *rate availability function (RAF)*. Previous methods those base on peak rate, instead of RAF, to construct transmission schedules would waste system resources in allocation. In this paper, we extend the algorithm proposed in [1] to consider RAF and explore how the available bandwidth can be used most effectively toward minimizing playback delay and client buffer requirements. The optimality of our approach is formally established.

Keywords[*]**:** resource allocation, transmission schedule, traffic shaping, real-time multimedia

[*] **This paper is partially supported by NSC under Grant 91-2416-H-008-008-**

1. Introduction

A broad range of applications (such as E-commerce, distance learning/collaboration, interactive virtual environments, and shopping/entertainment services) is enabled by the capacity to efficiently manage networked multimedia information [16]. However, multimedia data such as video and audio are VBR (variable-bit-rate) in nature [5]. For supporting jitter-free delivery, a multimedia system needs real-time transmission schedules to guarantee end-to-end QoS (quality-of-service) between servers and clients. Moreover, the requirement of resources such as network bandwidth, playback delay and client buffer should be minimized. In past years, different approaches [2] were proposed to shape stored VBR video streams as smooth traffic schedules for efficient delivery. However, they did not achieve the schedule result that optimizes playback delay, client buffer and bandwidth utilization at the same time. For example, as shown in a latest comprehensive survey [14] about this problem, the allocated playback delay and the bandwidth utilization were not minimized in previous methods (including the best-known MVBA, minimum variability bandwidth allocation, method introduced in [7]). In [1], a linear time method was proposed to resolve this drawback by constructing the transmission schedule that has not only minimal client buffer and delay but also maximal bandwidth utilization. However, as its problem definition is the same as previous methods, the method proposed in [1] only considered the peak rate of schedule in resource allocation. Due to the sharing of bandwidth for different users, the available rate of network is time-varying in nature. As shaped transmission schedules are usually piecewise linear in bandwidth consumption [14], the available rate can be represented by a piecewise-linear function called *rate availability function (RAF)*. In [17], RAF has been applied in admission control while a shaped transmission schedule was given. However, it wasn't considered in constructing transmission schedule directly before. Previous methods those base on peak rate, instead of RAF, to construct transmission schedules would waste system resources in allocation.

In this paper, we extend the algorithm proposed in [1] to apply RAF in the construction of transmission schedule. The proposed method can fully utilize available resources specified in RAF to serve as many users as possible. We have explored how the allocated bandwidth can be used most effectively toward minimizing playback delay and client buffer required. The optimality of our approach is formally established. The remainder of this paper is organized as follows. Basic concepts of video transmission and its resource allocation problem with RAF are introduced in Section 2. The proposed algorithm that constructs an optimal transmission schedule with minimal resource allocation is proposed in Section 3. Experiments are shown in Section 4. The last section shows the conclusion remarks.

2. Resource Allocation with Rate Availability Function

For supporting QoS delivery of a stored video, we need a good transmission schedule to deliver the retrieved data [15] from server to client. At the client site, a proper size of memory buffer is allocated. Incoming data are temporarily stored in the client buffer and consumed frame-by-frame periodically. Such a consumption schedule is also called *playback schedule* in this paper. Given a video stream V, it can be represented by a set of frames in the following.

$$V = \{ f_0, f_1, \ldots, f_{n-1} \} .$$

The number of frames in V is n and the ith frame is f_i (its frame size is represented as $|f_i| > 0$).

The ith accumulative fame size F_i can be computed as follows.

$$F_i = F_{i\text{-}1} + |f_i| .$$

The initial value is $F_k = 0$ for $k < 0$. The total frame size is just the video size $|V|$. We have $F_k = |V|$ for $k > n$-2. In this paper, we assume that a video is played at $t = 0$ and the time to play the ith frame is i (unit time). As the client plays the video frame-by-frame periodically, the playback schedule can be denoted by its accumulative playback function F(.).

$$F(t) = F_x \; ; \; \forall \; x \; \leq \; t < (x+1) .$$

Note that F(.) is a nonnegative stair function with jumps at time t for t = 0, 1, ..., n-1. The low corner and the up corner at time t are represented by $F(t)^- = F(t\text{-}1)$ and $F(t)^+ = F(t)$, respectively.

Based on the same idea, we define the transmission schedule G(.) as a function that cumulates the amount of data received at the client. As the playback of video is assumed to be started at $t = 0$, d is the playback delay if the start time of the transmission schedule is -d. Assume that video data are transmitted by rate $r(t)$ between time t-1 and time t. The transmission schedule can be represented by a integration function of transmission rate r(.) as follows.

$$G(t) = \sum_{i=-d}^{t} r(i)$$

The peak bandwidth of the network channel allocated for transmission is $r = \mathbf{max}\{ r(t) \mid \forall t \}$. According to the above formulation, $G(t)$ represents the amount of data sent by the server up to time t. Assume that there is no transmission error and the network delay is zero. $G(t)$ can also represent the amount of data received by the client up to time t. At a client, $G(t)$ and $F(t)$ represent the cumulated data received and consumed up to time t respectively. $b(t) = G(t) - F(t)$, called the buffer occupancy, would be the amount of transmitted data temporarily stored in the client buffer at time t. To avoid jitter in playback, a transmission schedule must be ahead of its playback schedule (such that $b(t) \geq |f_t|$ for any time t and the client buffer would not be underflow for playback). The minimal client buffer size required for supporting QoS delivery and playback is $b = \mathbf{max}\{ b(t) \mid \forall t \}$. Because $b(t) \geq |f_t|$ for any time t and $b \geq \mathbf{max}\{ |f_t| \}$, the required buffer size b is no smaller than the maximum frame size, and is not necessary to be larger than the size of video $|V|$.

Note that, given a limited buffer size, it results in loss of data if the transmission schedule sends too many data to the client buffer at the same time. Such an overflow condition should be avoided also in designing a *feasible* transmission schedule. In this paper, a transmission schedule is said to be feasible if it has no buffer overflow or underflow. Its upper bound $H(.)$ can be computed in the following.

$$H(t) = \mathbf{min}\{ |V|, F(t\text{-}1) + b \}.$$

For any time t, the value of $G(t)$ must be not smaller than its playback schedule $F(t)$. Moreover, $G(t)$ must be not larger than its upper bound $H(t)$. Given the playback schedule $F(.)$ of a stored video V, different approaches were proposed to construct its feasible transmission schedule $G(.)$ in past years [2]. A latest comprehensive survey about this problem can be found in [14]. In previous works, the performance of a transmission schedule is generally measured by its playback delay time, client buffer size, and bandwidth requirement. However, the bandwidth requirement they measured is the maximum network bandwidth required during transmission (called *peak rate*). A user request is admitted if its peak rate is smaller than the available bandwidth of the current network.

Notably, the network bandwidth is shared by different traffics. Therefore, the available rate of network is time-varying. It can be represented by a *rate availability function* (RAF) $z(.)$ where $z(t)$ is the available rate of network between time t-1 and time t. For supporting QoS delivery, the transmission schedule $G(.)$ must guarantee that its allocated rate $r(t)$ is not over the available rate $z(t)$ at any time t. In [17], RAF has been applied in admission control while a transmission schedule $G(.)$ with allocated rates $r(.)$ was given. Although the peak rate $r = \mathbf{max}\{ r(t) \mid \forall\ t \}$ may have been minimized, they did not consider available rates $z(t)$ for all time t in constructing transmission schedules. Therefore, the allocated rate $r(t)$ may be larger than its available rate $z(t)$ at time t. Previous methods those base on peak rate, instead of RAF, to construct transmission schedules would waste system resources in allocation. An example from the best-known algorithm [7] is shown in Fig. 1. The delay time is 2 and the rates allocated for the transmission schedule $G(t)$ are $r(-1)=1$, $r(0)=1$, $r(1)=1$, $r(2)=1$, $r(3)=1$, and $r(4)=2$. The peak rate allocated $r(4)=2$ is minimized and $G(t)$ is a feasible transmission schedule while RAF is not considered. Let the rates available in RAF be $r(-1)=3/4$, $r(0)=3/4$, $r(1)=3/4$, $r(2)=1$, and $r(x)=2$ for $x > 2$ as shown in Fig. 2. In view of the fact that $r(x) > z(x)$ for x = -1 to 1. The function $G(t)$ is not a feasible transmission schedule while RAF is considered. The system may require extending its delay time to provide $G(t)$ a guaranteed service. The extended delay time is 2 in our example (see Fig. 2).

3. Proposed Algorithms

A transmission schedule is said to be optimal in resource allocation if it requires the minimal network bandwidth, delay time and client buffer for supporting QoS delivery. In [1], a linear time method was proposed to construct an optimal transmission schedule that has not only minimal resource requirement but also maximal resource utilization. However, as it did not consider RAF in its procedure, the transmission schedule obtained may not be feasible. In this paper, we consider RAF in constructing transmission schedule directly to for fully utilizing available resources to serve as many users as possible. Given a stored video $V = \{ f_0, f_1, ..., f_{n-1} \}$ and RAF of network $z(.)$, the amounts of minimal resource required for guaranteeing jitter-free playback can be decided by ALGORITHM-1.

ALGORITHM-1:

// **INPUT**: the playback schedule $F(.)$ of a stored video $\boldsymbol{V} = \{ f_0, f_1, ..., f_{n-1} \}$,

// RAF of network $z(.)$

// **OUTPUT**: the transmission schedule $L(.)$ with transmission rates $r(.)$,

// the minimal client buffer size b and the minimum playback delay d

$L(n\text{-}1) = |V| = F(n\text{-}1);$

$t = n\text{-}1;$

while ($L(t) > 0$) **do** {

$t = t - 1;$

$L(t) = \mathbf{max}\{ F(t), L(t+1) - z(t+1) \};$

}

$d = -t;$

Initialize $b = 0$ and $r(i) = 0$ for all i;

for $t = -d+1$ **to** n-1 **do** {

$r(t) = L(t)-L(t-1)$;

$b =$ **max**{ $b, L(t)-F(t-1)$ };

}

An example to illustrate its computation is shown in Fig. 3. As the maximal available rate $z(t)$ is applied in each time t, the video data are transmitted and stored into the client buffer as late as possible. Therefore, the minimal buffer occupancy $L(t)-F(t-1)$ can be determined at any time t under guaranteed QoS. From the definitions shown in Section 1, it is not difficult to prove that the minimal client buffer size b and the minimum playback delay d can be achieved. Besides, given any transmission schedule $P(.)$ with RAF $z(.)$, we have $L(t) \leq P(t)$ for any time t. $L(.)$ is called the minimal $z(.)$-bounded transmission schedule.

Lemma-1: $L(.)$ is the minimal $z(.)$-bounded transmission schedule. It has the minimal buffer size and initial delay for all $z(.)$-bounded transmission schedules.

Proof:

(1) Suppose the contrary and let $P(.)$ be a $z(.)$-bounded transmission schedule, for which, there exists a time index x such that $L(x) > P(x)$ as shown in Fig. 4. Let y be the smallest time index that satisfies $x < y$ and $L(y) = F(y) = L(x) + \sum\{ z(i)$; for $i = x+1$ to y }. (The value y is existed. At least, we have the initial value $L(n-1) = F(n-1)$.) Follow the procedure steps of algorithm, $L(y) = F(y)$ implies $L(y+1) - z(y+1) \leq F(y)$. As $P(.)$ is $z(.)$-bounded, the relation $P(y) \leq P(x) + \sum\{ z(i)$; for $i = x+1$ to y } is true. We have $P(x) + \sum\{ z(i)$; for $i = x+1$ to y } $< L(x) + \sum\{ z(i)$; for $i = x+1$ to y }. That implies $P(y) < F(y)$. The underflow condition of the client buffer is occurred and

$P(.)$ is not a feasible transmission schedule. It is a contradiction and $L(.)$ is the minimal $z(.)$-bounded transmission schedule.

(2) Since $L(.)$ is the minimal $z(.)$-bounded transmission schedule, it sends the minimal amount of data to the client buffer for guaranteeing jitter-free playback. At any time t, we have $L(t) \leq Q(t)$ where $Q(.)$ is any other $z(.)$-bounded transmission schedule. As buffer occupancies have the relation $L(t)-F(t-1) \leq Q(t)-F(t-1)$. It implies that $L(.)$ has the minimal buffer size (the required buffer size $\mathbf{max}\{ L(t)-F(t-1) \mid \forall\, t \} \leq \mathbf{max}\{ Q(t)-F(t-1) \mid \forall\, t \}$).

(3) At time 0, we have $L(0) \leq Q(0)$. Given the available rate $z(0)$, we have $L(-1) = \mathbf{max}\{ 0, L(0) - z(0) \} \leq \mathbf{max}\{ 0, Q(0) - z(0) \} \leq Q(-1)$. Repeat the above step, $0 < L(-d+1) \leq Q(-d+1)$ and $0 = L(-d) \leq Q(-d)$. As any other $z(.)$-bounded transmission schedule $Q(.)$ has $0 < Q(-d+1)$ and $0 \leq Q(-d)$, the transmission schedule $L(.)$ has the minimal initial delay for all $z(.)$-bounded transmission schedules.

The lemma is proved. **Q.E.D.**

In our ALGORITHM-1, the minimum requirements in buffer size and playback delay are decided for the given video $V = \{ f_0, f_1, ..., f_{n-1} \}$ and RAF $z(.)$. While a user request is presented, we can compare the available buffer size B (the available playback delay D) and the minimum buffer size b (the minimum playback delay d) to make the admission. If the client buffer size $B \geq b$ and the playback delay $D \geq d$ are given, we can construct a simple transmission schedule by ALGORITHM-2.

ALGORITHM-2:

// **INPUT**: the playback schedule $F(.)$ of a stored video $\boldsymbol{V} = \{ f_0, f_1, ..., f_{n-1} \}$,

// RAF of network $z(.)$, the client buffer size B and the playback delay D

// **OUTPUT**: the transmission schedule $A(.)$ with transmission rates $r(.)$

$A(-D) = 0$;

$t = -D;$

while ($A(t) < |V|$) **do** {

$t = t + 1;$

$A(t) =$ **min**{ $H(t), A(t-1) + z(t)$ };

}

$t_e = t$; // the end point of the transmission schedule

Initialize $r(i) = 0$ for all i;

for $t = -D+1$ **to** t_e **do** $r(t) = A(t)-A(t-1)$;

An example to illustrate its computation is shown in Fig. 5. For any other transmission schedule $P(.)$ that has the same client buffer size $B \geq b$, the playback delay $D \geq d$ and RAF of network $z(.)$, we can prove $P(t) \leq A(t)$ for any time t. The obtained result $A(.)$ is called the maximal $z(.)$-bounded transmission schedule for the given client buffer size B ($B \geq b$) and playback delay D ($D \geq d$). As the video data have been transmitted to the client as early as possible, packet losses can be recovered by an error control scheme with data retransmission [19]. The transmission schedule obtained is robust against network errors.

Lemma-2: $A(.)$ is the maximal $z(.)$-bounded transmission schedule for the same client buffer size $B \geq b$ and playback delay $D \geq d$.

Proof:

Suppose the contrary. Let $P(.)$ be a $z(.)$-bounded transmission schedule, for which there exists a time index x such that $P(x) > A(x)$ as shown in Fig. 6. Let y be the largest time index that satisfies $y < x$ and $A(y) = H(y) = A(x) - \sum\{ z(i)$; for $i = x+1$ to y }. Follow the procedure steps of algorithm, $A(y) = H(y)$ implies $H(y) \leq A(y\text{-}1) + z(y)$. (The value y is existed. At least, we have $A(t_e) = H(t_e)$ for the end point t_e of the transmission schedule.) As $P(.)$ is $z(.)$-bounded, the relation $A(x)$ - $\sum\{ z(i)$; for $i = x+1$ to y } < $P(x)$ - $\sum\{ z(i)$; for $i = x+1$ to y } $\leq P(y)$ is true. That implies $H(y) < P(y)$. The overflow condition of the client buffer is occurred and $P(.)$ is not a feasible transmission schedule. It is a contradiction and the lemma is proved. **Q.E.D.**

4. Experiment Results

Our method is the first traffic shaping method that can consider RAF in constructing transmission schedules. It can decide the minimal client buffer and delay required under RAF constrains. As the optimality has been formally established, experiments on client buffer and delay required will not be addressed in this paper. Interested readers can find some basic results (without RAF) in our previous works [1]. In this paper, our experiments focus only on the number of users supported by the proposed method. Fig. 7(a) shows the cumulative playback function of test video *Princess Bride* encoded by Futuretel MPEG coder [9][10]. The length of *Princess Bride* is over 90-minutes long. As this video is encoded by hardware, its target rate is maintained with the same group-of-picture (GoP) size and frame rate. In our experiments, all users are assumed to request the same video stream *Princess Bride* with 200 KB memory buffer. Their start times are randomly and uniformly distributed with 200 ms playback delay. As the test video is nearly CBR (constant bit rate), its obtained result is supposed to have the best resource allocation for each single user. On an OC-3 link, the conventional method [7] with a fixed transmission schedule (with a fixed rate during transmission) can support nearly 130 users. However, in real networks, resources are sharing by different users. Allocating resources to different users without considering RAF may make incorrect decisions. A simple example is shown in Fig. 8(a) where the residue of (network bandwidth) / (allocated peak rate) would not be utilized. To resolve this drawback, our approach considers the current rate availability to allocates required rate. Notably, as shown in Fig. 8(b), the rate allocation is decided on not only the buffer availability but also the rate availability. Given the same OC-3 link, we can support over 155 users to watch *Princess Bride*. The improvement is over 15%. (Notably, the transmission schedule obtained by our method depends on current RAF specified.)

In modern packet networks, a video stream V is divided as a sequence of data packets $p_0p_1p_2\ldots p_m$ for delivery (see Fig. 9). Given an EDF (earliest deadline first) scheduler, the time constraints (ready time, deadline) = (s_x, e_x) (called schedulable region in [1]) for each packet p_x should be specified. Note that, at any time t, $A(t)$ represents the maximal amount of video data that could be received by client without buffer overflow. $L(t)$ is the minimal amount of video data that should be received before time t for supporting jitter-free playback. They show the upper bound and the lower bound for all transmission schedules with the same available rates $z(.)$, buffer size b, and playback delay d. When transmission schedules $A(.)$ and $L(.)$ are specified, we can decide when is the earliest time that p_x could start its transmission (the ready time s_x) and when is the latest time that p_x should be received (the deadline e_x). Therefore, a real-time transmission and error control mechanism [20] can be provided for the delivery of stored video across modern best-effort networks.

4. Conclusion

Due to the sharing of bandwidth for different users, the available rate of network is time-varying in nature. Usually, it is represented by a piecewise-linear function called *rate availability function (RAF)*. In [17], RAF has been applied in admission control while a shaped transmission schedule was given. However, it wasn't considered in constructing transmission schedule directly before. In this paper, we extend the algorithm proposed in [1] to consider RAF and explore how the available bandwidth can be used most effectively toward minimizing playback delay and client buffer requirements. The optimality of our approach is formally established. It is shown to be practical, efficient, and flexible in supporting continuous media transmission with RAF.

References

[1] Ray-I Chang, Meng-Chang Chen, Ming-Tat Ko and Jan-Ming Ho, "Schedulable Region for VBR Media Transmission with Optimal Resource Allocation and Utilization," *Information Sciences*, Vol. 141, Issue 1-2, pp. 61-79, 2002.

[2] W.-C. Feng and J. Rexford, "Performance evaluation of smoothing algorithms for transmitting prerecorded variable-bit-rate video," *IEEE Trans. Multimedia*, vol.1, no.3, pp. 302-313, 1999.

[3] Ray-I Chang, Meng-Chang Chen, Ming-Tat Ko and Jan-Ming Ho, "Bandwidth-Buffer Tradeoff for Delivery of Pre-Recorded Videos over Multiple Smoothing-Servers," *Journal of Applied Systems Studies*, Vol. 2, No. 3, 2001.

[4] Ray-I Chang, Meng-Chang Chen, Ming-Tat Ko and Jan-Ming Ho, "Online Traffic Smoothing for Delivery of VBR Media Streams," *Circuits, Systems, Signal Processing*, Vol. 20, No. 3, pp. 341-359, 2001.

[5] M. Krunz and H. Hughes, "A traffic model for mpeg-coded VBR streams," in proc. *ACM SIGMETRICS*, pp.47-55, 1995.

[6] J.M. McManus and K.W. Ross, "Video-on-demand over ATM: constant-rate transmission and transport," *IEEE Journal on Selected Areas in Communications*, Volume: 14 Issue: 6,

Page(s): 1087 –1098, Aug. 1996

[7] J. D. Salehi, Z. L. Zhang, J. F. Kurose, and D. Towsley, "Supporting stored video: reducing rate variability and end-to-end resource requirements through optimal smoothing," *IEEE/ACM Trans. Networking*, Vol.6, No.4, pp.397-410, 1998.

[8] K. Sohraby. On the theory of general on-off source with applications in high-speed networks. IEEE INFOCOM, pages 401-410, 1993.

[9] URL: ftp://thumper.bellcore.com/pub/vbr.video.trace

[10] Oliver Rose. "Statistical properties of MPEG video traffic and their impact on traffic modeling in ATM systems". Proceedings of the 20th Annual Conference on Local Computer Networks, Minneapolis, MN, 1995, pp. 397-406.

[11] Wu-chi Feng, "Video-on-Demand Services: Efficient Transportation and Decompression of Variable Bit Rate Video", Ph.D. Thesis, Univ. of Michigan, 1996.

[12] Y.C. Wang, S.L. Tsao, Ray-I Chang, M.C. Chen, J.M. Ho, and M.T. Ko. A fast data placement scheme for video server with zoned-disks. SPIE VVDC, pages 92-102, 1997.

[13] H. Zhang and E. W. Knightly. A new approach to support delay-sensitive VBR video in packet-switched networks. IEEE NOSSDAV, 1995.

[14] J. Zhang and J. Hui, "Traffic characteristics and smoothness criteria in VBR video traffic smoothing," *IEEE ICMCS*, pages 3-11, June 1997.

[15] Ray-I Chang, W.K. Shih and R.C. Chang, "Real-time disk scheduling for multimedia applications with deadline-modification-scan scheme," *Real-Time Systems*, vol.19, no.2, pages 149-168, 2000.

[16] Ray-I Chang, Meng-Chang Chen, Ming-Tat Ko and Jan-Ming Ho, "VBR Traffic Shaping for Streaming of Multimedia Transmission," *Multimedia Networking: Technology, Management and Applications*, Mahbubur Rahman Syed (Eds.), Idea-Group Pub., ISBN: 1-930708-14-9, 2002.

[17] V. Firoiu, J. Jurose and D. Towsley, "Efficient admission control of piecewise linear traffic envelopes at EDF schedulers," *IEEE/ACM Trans. Networking*, vol.6, no.5, pp.558-570, 1998.

[18] Jan-Ming Ho, Ray-I Chang, J.Y. Juang, Chia-Hui Wang, "Design and Implementation of a Web-Based Surveillance System using Internet Multicast Communications," *IEEE SoftCOM*, pp. 273-282, 2000.

[19] Chia-Hui Wang, Jan-Ming Ho, Ray-I Chang, Shun-Chin Hsu, "An Efficient Retransmission-Based Error Control Method in Real-Time Communication for Stored Multimedia," *submitted for publication*.

[20] Chia-Hui Wang, Jan-Ming Ho, Ray-I Chang, and Shun-Chin Hsu, "A Feedback-Controlled EDF Scheduling Algorithm for Real-time Multimedia Transmission," Technical Report TR-IIS-01-008, Academia Sinica, 2001.

[21] W. Feng, F. Jahanian, and S. Sechrest, "Optimal buffering for the delivery of compressed prerecorded video," IASTED International Conference on Networks, Jan 1996.

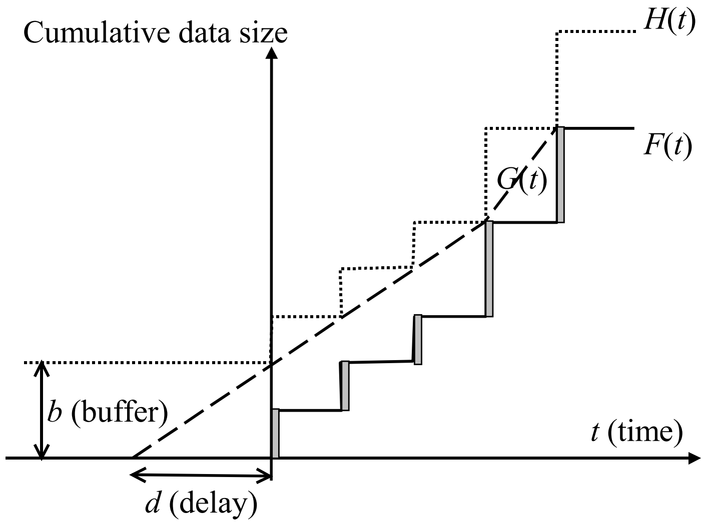


Fig. 1. The traffic-shaping result obtained by the best-known algorithm [7] is a feasible transmission schedule while RAF is not considered. The required buffer size and delay time are indicated.

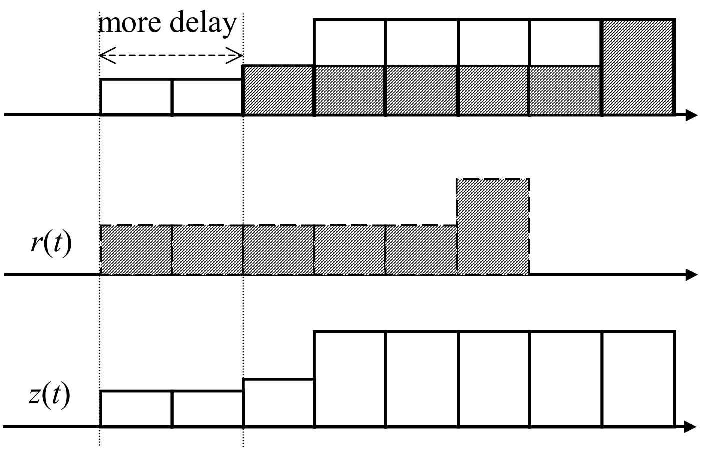


Fig. 2. As $r(x) > z(x)$ for x = -1 to 2, the function $G(t)$ is not still a feasible transmission schedule while RAF is considered. It may require extending the delay time of the transmission schedule to provide a guaranteed service.

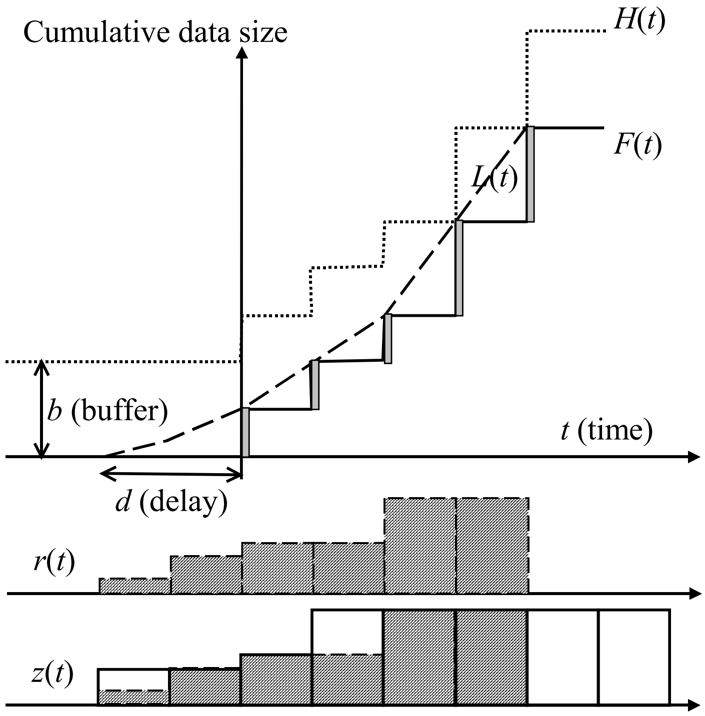


Fig. 3. The computation result of the proposed ALGORITHM-1 is illustrated. As video data are transmitted and stored into the client buffer as late as possible, the required buffer size and playback delay are minimal.

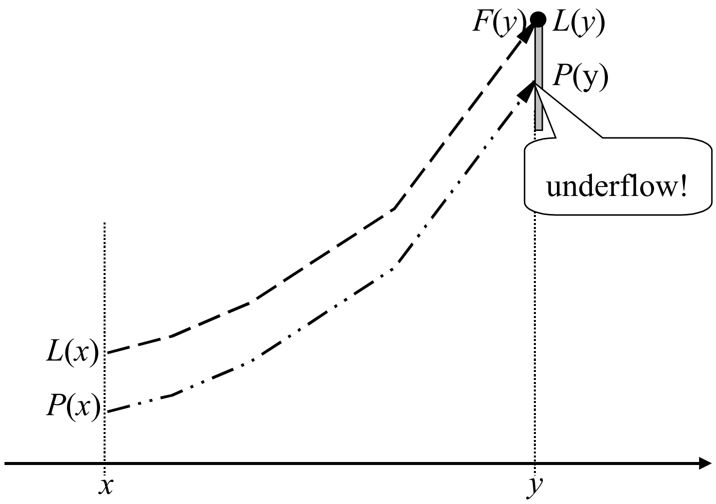


Fig. 4. A simple example is shown to prove that the obtained transmission schedule *L*(.) is the minimal *z*(.)-bounded transmission schedule.

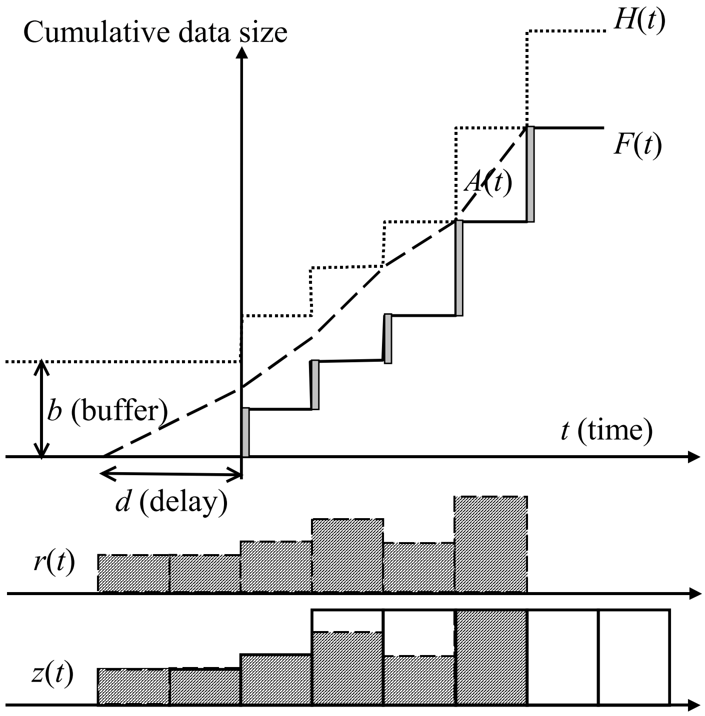


Fig. 5. The computation result of the proposed ALGORITHM-2 is illustrated. The obtained transmission schedule is robust against network errors because the video data have been transmitted to the client as early as possible.

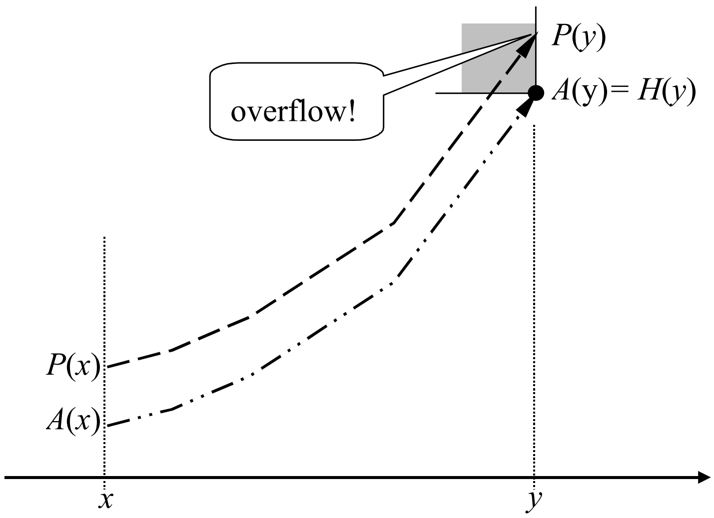


Fig. 6. A simple example is shown to prove that the obtained transmission schedule $A(.)$ is the maximal $z(.)$-bounded transmission schedule.

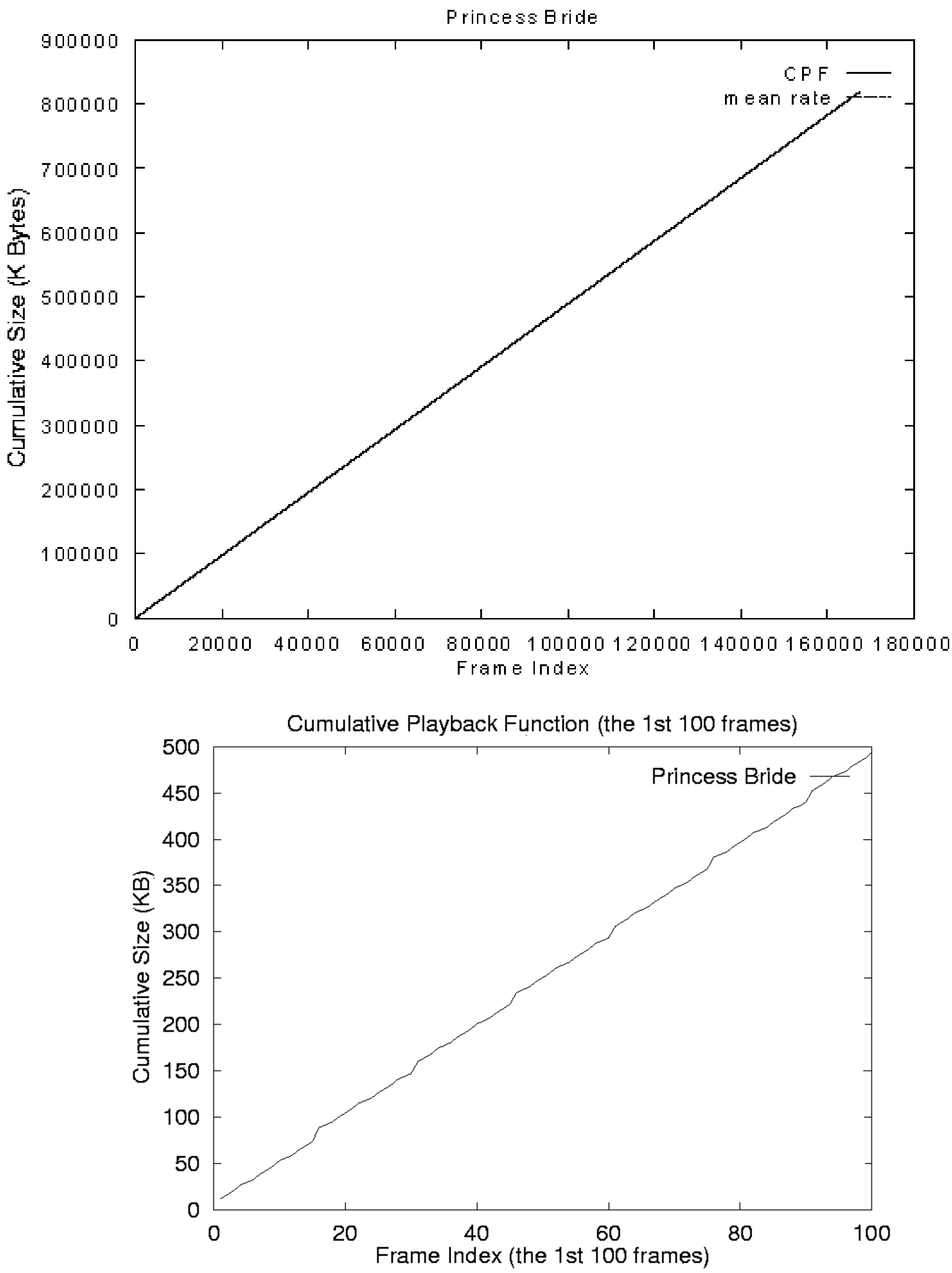
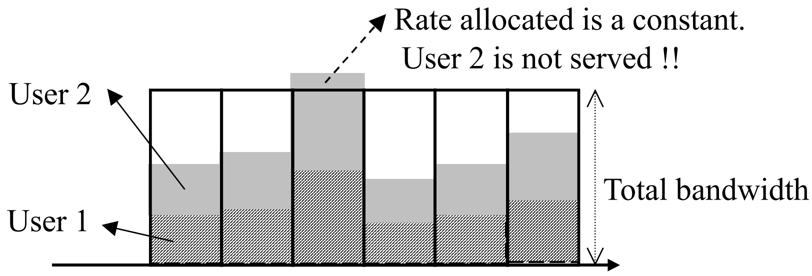
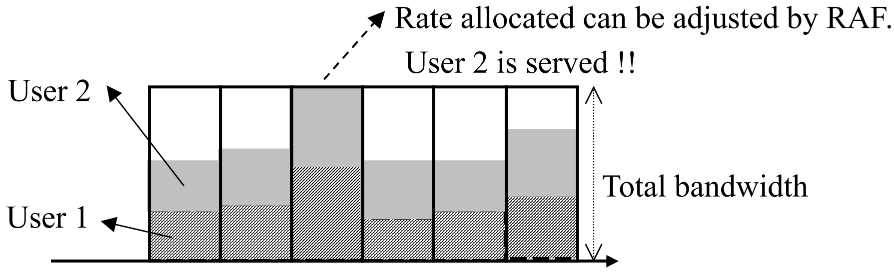


Fig. 7. The cumulative playback function of test video *Princess Bride* encoded by Futuretel MPEG coder. As the test video is nearly CBR, the transmission schedule obtained by the conventional method is supposed to have the best resource allocation for each single user.



(a)



(b)

Fig. 8. A simple example to show the difference between (a) the conventional method and (b) the proposed method.

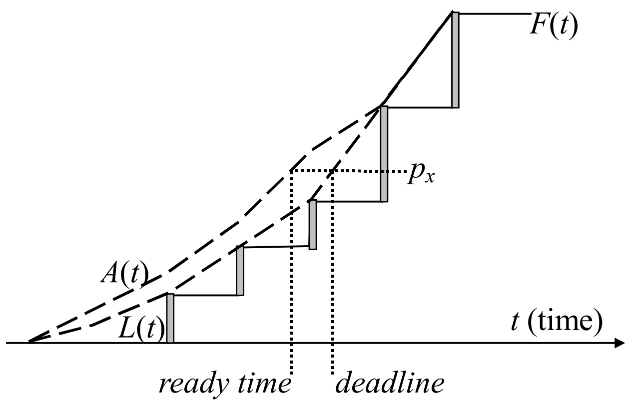


Fig. 9. A video stream V can be divided as a sequence of data packets $p_0p_1p_2\ldots p_m$ for delivery. We can utilize $A(.)$ and $L(.)$ to specify (ready time, deadline) for each p_x .