Computing and Informatics, Vol. 21, 2002, 1001–1018, No. 6, V 25. 3.

QOS-COMPETITIVE VIDEO BUFFERING*

Alexander Kesselman, Yishay Mansour

School of Computer Science
Tel-Aviv University
Tel-Aviv 69978, Israel
e-mail: {alx, mansour}@cs.tau.ac.il

Manuscript received 15 January 2002; revised 29 January 2003
Communicated by Burkhard Monien

Abstract. Many multimedia applications require transmission of streaming video from a server to a client across an internetwork. In many cases loss may be unavoidable due to congestion or heterogeneous nature of the network. We explore how discard policies can be used in order to maximize the quality of service (QoS) perceived by the client. In our model the QoS of a video stream is measured in terms of a cost function, which takes into account the discarded frames. In this paper we consider online policies for selective frame discard and analyze their performance by means of *competitive analysis*. In competitive analysis the performance of a given online policy is compared with that of an optimal offline policy. In this work we present competitive policies for a wide range of cost functions, describing the QoS of a video stream.

Keywords: Competitive analysis, QoS, video streaming, buffer policies, scheduling

1 INTRODUCTION

The emergence of high-speed networks facilitates many multimedia applications that rely on the efficient transfer of compressed video. Such applications include streaming video broadcasts, distance learning, shopping services, etc. However, compressed video, especially variable-bit-rate (VBR) video, typically exhibits significant burstiness on multiple time scales, owing to the encoding schemes and the content variation

* A preliminary version of this paper appeared at SIROCCO 2001.

between and within video scenes [6], [9]. This burstiness complicates the design of efficient transport mechanisms for such media. In a network where resources such as bandwidth and buffering capacity are constrained there is a need for an efficient video delivery system that can achieve high resource utilization and maximize the QoS perceived by the user.

In reality both network bandwidth and buffering capacity are likely to be limited. Under such circumstances there may be a situation in which a feasible transmission schedule is impossible, i.e., any transmission necessarily incurs loss of data. In this case instead of being denied service, clients may choose lower quality video streams with occasional frame loss. This is a natural situation in today's Internet where the network may not have sufficient bandwidth to support the peak rate of a video stream, or available bandwidth may fall below the requirement at the middle of a video transmission. A naive approach is to discard frames with no awareness of the video stream properties. As a result, the QoS perceived by the user may degrade drastically, even for small amounts of loss (e.g. tail drop of consecutive frames). In this paper we study intelligent selective frame discard policies, which take into account the application-specific properties in its decisions to discard particular frames, minimizing the QoS degradation.

Previous work. Recently a number of packet discarding schemes incorporating application-specific information have been proposed. In [12] appears a simple strategy called Frame-Induced Packet Discarding, in which upon detection of loss of a threshold number of packets belonging to a video frame, the network attempts to discard all the remaining packets of that frame. In [8] the problem of optimizing the quality of the transmitted video for a given cost function has been considered with leaky bucket constraints. Our work differs from theirs in that we are trying to optimize the QoS perceived by the user, rather than minimizing loss in general. In [24] *offline* algorithms for optimal selective frame discard have been considered. The notion of *selective frame discard* at the server has been introduced and the optimal selective frame discard problem using a QoS-based cost function has been defined. Unlike [24], we deal with *online* problem, when no information regarding the video stream is known a priori. Dropping schemes of MPEG frames are studied in [1, 23, 7].

Competitive analysis. We measure the performance of our algorithms using competitive analysis [18, 2]. In competitive analysis the performance of an online policy is compared with that of an optimal offline policy, which knows in advance the entire sequence of frame arrivals. Competitive analysis is a natural approach for Internet traffic, which is unpredictable and chaotic [19]. The advantage of competitive analysis is that a uniform performance guarantee is provided on all input sequences.

Our work. In this paper we study competitive online policies for intelligent selective frame discard at time of buffer overflow. We consider video streams without inter-frame dependencies and assume that frames have a fixed size. In order to measure the QoS perceived by the user we define a so called *well-behaved* cost function reflecting the playback discontinuity at the client. The cost of a discarded frame depends either on the distance to the closest discarded frame or on the position within a sequence of consecutive discarded frames it participates in. The QoS level function is the sum of the costs of the discarded frames. We consider various video stream settings, and our main result is that the competitive ratio of the Greedy Policy is bounded by a constant for all of these settings. At the end we show how our drop policy can deal with MPEG video streams having complex inter-frame dependencies.

Related work. In [11] the problem of smoothing real-time streams has been considered. Competitive analysis of jitter control online algorithms appears in [10]. Recently, several bandwidth smoothing techniques have been introduced to reduce the server and network resource requirements for transmitting pre-recorded video [4, 5, 15, 14]. These techniques are based on prior knowledge of the frame sizes of the entire video. In contrast, interactive video applications, such as video conferencing, typically have limited knowledge of the arrival sequence. Online smoothing techniques of live video streams have been presented in [17, 13].

The rest of this paper is organized as follows. Summary of results is presented in Section 2. Section 3 contains the model description. The Greedy Policy is defined in Section 4. Section 5 contains the analysis of the Greedy Policy. We show how our drop policy can be extended to handle MPEG video streams in Section 6. Section 7 contains concluding remarks.

2 SUMMARY OF RESULTS

In this section we give a brief overview of our main results while the formal definitions and proofs are deferred to the following sections. We analyze the natural Greedy Policy by means of global and local QoS metrics. The Greedy Policy always minimizes the loss incurred by a discarded frame, i.e., it always discards a frame that incurs the minimal cost with regards to the current state of the sequence. A formal definition of this policy is given later.

In our model each discarded frame has a non-zero cost given by a cost function. A cost function should provide measure of playback discontinuity, i.e., QoS degradation perceived by the user. We take two main aspects into consideration: the length of a sequence of consecutive discarded frames and the distance between two adjacent, but non-consecutive discarded frames. We define a family of *well-behaved* cost functions as follows. We require the cost of a single discarded frame to be bounded from below by L and from above by U (L and U are some positive constants). Similarly, the cost of a discarded frame at position l within a sequence of consecutive discarded frames is bounded by $L' \cdot l$ from below and by $U' \cdot l$ from above (again L' and U' are some positive constants). Intuitively a well-behaved cost function is a convex function monotonically decreasing on the distance between two non-consecutive discarded frames and increasing on the position of a discarded frame within a sequence.

We assume that the video stream is regulated by a leaky bucket. We consider moderate and large values of burst parameter σ (exact definition will be given later) defined with respect to the buffer capacity M (in frames). Notice that for $\sigma \leq M$ there is no loss since the maximal burst could be completely accommodated by the buffer. Moderate and large values of burst parameter correspond to $\sigma \leq \frac{5}{4}M$ and $\sigma > \frac{5}{4}M$, respectively. We choose this threshold because for $\sigma \leq \frac{5}{4}M$ the Greedy Policy does not discard consecutive frames.

The first set of our results deals with the competitive ratio of the Greedy Policy with regards to the global QoS metric, that is the sum of the costs of all the discarded frames under a well-behaved cost function. Our main result is presented in Table 1: for any sequence of frames, the competitive ratio of the Greedy Policy is $\min\left(L/U, \frac{LL'}{(12L+L')U'}\right)$.

Moderate σ	L/U
Large σ	$\frac{LL'}{(12L+L')U'}$

Table 1. Greedy competitive ratio (general cost)

Notice that in presence of large bursts the competitive ratio of the Greedy Policy *linearly* depends on the ratio between the constants. This is despite the fact that the cost of a sequence of discarded frames depends quadratically on its length.

The next set of results deals with local QoS metrics that may be viewed as "worst-case noise" metrics. We consider the main criteria affecting the playback discontinuity at the client: the minimal distance between two adjacent discarded frames and the maximal length of a sequence of consecutive discarded frames. We denote by d_{OPT} the minimal distance between two adjacent discarded frames produced by any *competitive online* policy (to be defined later) and by l_{OPT} the maximal length of a sequence of consecutive discarded frames created by an *optimal offline* policy. Table 2 compares the minimal distance between two adjacent discarded frames produced by the Greedy Policy with d_{OPT} for moderate values of the burst parameter and the minimal length of a sequence of consecutive discarded frames created by the Greedy Policy with l_{OPT} for large values of the burst parameter.

Minimal Distance (moderate σ)	$d_{OPT}/2 - 1$
Maximal Length (large σ)	$4l_{OPT} + 3$

Table 2. Worst-case noise metrics

3 MODEL DESCRIPTION

We consider a system with a single FIFO buffer delivering a video stream. We assume that the buffer can hold exactly M frames (we assume that all frames have the same size). Frames may arrive to the buffer at any time and send events are synchronized with time. We divide time into slots so that during each slot one video frame is sent. A buffer policy has to decide for each incoming frame whether it should be either rejected or accepted subject to the buffer capacity constraints. A frame in the buffer may be also *preempted* by the policy. However, all the frames must be sent in FIFO order.

Now we describe the traffic-shaping policy, i.e., the regulation of the rate at which a flow is allowed to inject packets into the QoS network. We consider leaky bucket traffic-shaping mechanism in which only a fixed amount of traffic is admitted to the network. Excess traffic is held in a queue until either it can be accommodated or must be discarded. A (σ, ρ) leaky bucket model with a burst parameter σ and a rate parameter ρ, during time interval of length t, has at most $\rho \cdot t + \sigma$ packet arrivals. Our results vary with the size of σ.

Definition 3.1. A (σ, ρ)-source is a source that generates stream that is shaped by a (σ, ρ) leaky bucket policer.

In our model each frame has the corresponding *cost*. The cost of a frame depends on the positions of previously discarded frames. The goal is to design a policy that minimizes the total cost incurred. Next we give a formal definition of the notation used. We start with a notation for a sequence of frames.

Definition 3.2. A sequence of frames is denoted by S. The start and the finish time of S are denoted by t^{start} and t^{final}, respectively.

Next we set the notation to indicate lost frames.

Definition 3.3. The loss indicator $X_f(t)$ corresponds to whether frame f was discarded at time t, i.e., $X_f(t)$ is 1 if f is discarded prior to time t and 0 otherwise.

Now we define the parameters which determine the cost of a discarded frame.

Definition 3.4. Let $d(f, t)$ be the distance between a frame f and the closest discarded frame preceding f at time t, or 0 if there is no such frame. Let $pos(f, t)$ be the position of a frame f within a sequence of consecutive discarded frames at time t, or 0 if f does not participate in such a sequence, i.e., if the previous and the following frames to f are not discarded at time t.

Once we have the necessary notation we can define the structure of the cost function.

Definition 3.5. The cost of a discarded frame f at time t is denoted by $c_{\phi_1,\phi_2}(f, t)$. If $d(f, t) > 0$ then $c_{\phi_1,\phi_2}(f, t) = \phi_1(d(f, t))$, otherwise $c_{\phi_1,\phi_2}(f, t) = \phi_2(pos(f, t))$.

We define the cost incurred by a policy over a sequence of frames as the sum of the costs of discarded frames.

Definition 3.6. The cost of a policy A while scheduling S is the sum of the costs of the discarded frames, i.e., $L_A(S) = \sum_f X_f(t^{final}) c_{\phi_1,\phi_2}(f, t^{final})$.

Now we introduce a class of well-behaved cost functions. In order to provide measure of playback discontinuity, a cost function should take two aspects into consideration: the length of a sequence of consecutive discarded frames and the spacing between two adjacent, but non-consecutive discarded frames (see [15]). The motivation behind the definition is as follows. For a single discarded frame the QoS degradation is minor while for sequences of consecutive discarded frames it is significant. Therefore, we would like the cost of a single discarded frame to be bounded by constant and the cost of a discarded frame within a sequence to be bounded by a constant factor of its position within a sequence of discarded frames (notice that the cost of a lost block would depend quadratically on its length). The cost of a frame within a sequence must be at least as high as the cost of a single discarded frame since it is always preferable, with respect to the video stream QoS, to drop a frame whose neighbors are not discarded. The QoS perceived by the user degrades as the distance between two discarded frames decreases or the length of a lost block increases. So we require a cost function to be monotonically increasing with: (1) decreasing distance between two non-consecutive discarded frames (2) increasing position of a discarded frame within a sequence of consecutive discarded frames. The best QoS is achieved when the loss is distributed as evenly as possible over the whole sequence. Therefore, the cost function should be minimal in this case. We require that the cost function is convex because the minimum of a convex function is established when all dropped frames are evenly spaced.

Definition 3.7. We say that a cost function c_{ϕ_1,ϕ_2} is well-behaved iff there exist positive constants L, U, L' and U', $0 < L \leq U \leq L' \leq U'$, satisfying the following constraints with respect to ϕ_1:

1. $\forall d: \ L \leq \phi_1(d) \leq U$ (bounded cost)
2. if $x < y$ then $\phi_1(x) > \phi_1(y)$ (ϕ_1 is anti-monotone)
3. $\forall x, y: \ \phi_1((x+y)/2) \leq (\phi_1(x) + \phi_1(y))/2$ (ϕ_1 is convex)

and with respect to ϕ_2:

1. $\forall l: \ L' \cdot l \leq \phi_2(l) \leq U' \cdot l$ (bounded average-cost)
2. if $x > y$ then $\phi_2(x) > \phi_2(y)$ (ϕ_2 is monotone)
3. $\forall x, y: \ \phi_2((x+y)/2) \leq (\phi_2(x) + \phi_2(y))/2$ (ϕ_2 is convex).

For instance, a well-behaved cost function c with $L = 1$, $U = 2$ and $L' = U' = 2$ might assign a cost to a discarded frame f as follows. If the frame is the l-th frame within a sequence of consecutive discarded frames then $c(f) = 2l$. Otherwise, the cost is defined based on its distance to the previous discarded frame d, and given by the formula $c(f) = 1 + 1/\sqrt{d}$. This cost function has been extensively studied in [20].

To analyze the performance of the Greedy Policy we introduce the following definitions.

Definition 3.8. Let the *lost set* of a policy at time t be the set of discarded frames that arrived during the time interval $[t - M + 1, t]$. Let the *frame set* of a policy at time t be the union of its lost set at time t and the frames in the buffer at time t. Let a *lost block* be a maximal sequence of consecutive discarded frames.

Now we turn to define competitive analysis. In competitive analysis (see [18]) the performance of an online policy is compared with that of the optimal policy (OPT), which knows in advance the entire sequence of frame arrivals. The competitive ratio is the minimum over all input sequences of the ratio between the cost incurred by OPT and the the cost incurred by the given online policy. More formally:

Definition 3.9. A policy A is c-competitive if for every sequence of frames S, $c \cdot L_A(S) \leq L_{OPT}(S)$. (Note that $0 \leq c \leq 1$). If $c > 0$ we also say that A is competitive.

Notice that a competitive online policy must not discard a frame from the input sequence S if OPT does not discard any frame from S.

4 GREEDY POLICY

We define a natural "Greedy Policy". The state of the policy depends on the previous decisions of the policy to discard particular frames.

Greedy Policy. Each time when the buffer is full and a frame arrives the policy discards a frame that minimizes the sum of the increase in cost of previously discarded frames plus the cost of the discarded frame itself.

4.1 Examples

Intuitively, when the Greedy Policy selects a frame to be discarded, the decision is optimized locally with respect to the current state of the system. On the other hand, OPT optimizes its decision globally with respect to the entire schedule. We study the performance of the Greedy Policy scheduling moderate and large bursts. Let us consider a system consisting of a buffer that is able to hold 3 frames. We define the *schedule* of a policy to be the sequence of arrive and send events (if any) for a stream of video frames.

Moderate bursts. For moderate bursts the Greedy Policy does not drop adjacent frames, as we prove later. Suppose that at time 0 the buffer is empty and a burst of 3 frames arrives. During the following t time units one frame is sent by the Greedy Policy and one frame arrives. Finally, at time t a burst of 4 frames arrives. In this case the Greedy Policy would drop 3 frames among the last six frames because its buffer is full. At the same time OPT would evenly distribute the loss over the whole sequence, which can be done since the buffer is not empty throughout this time interval. The resulting video streams appear on Figure 1 (dropped frames are marked by $\times$).

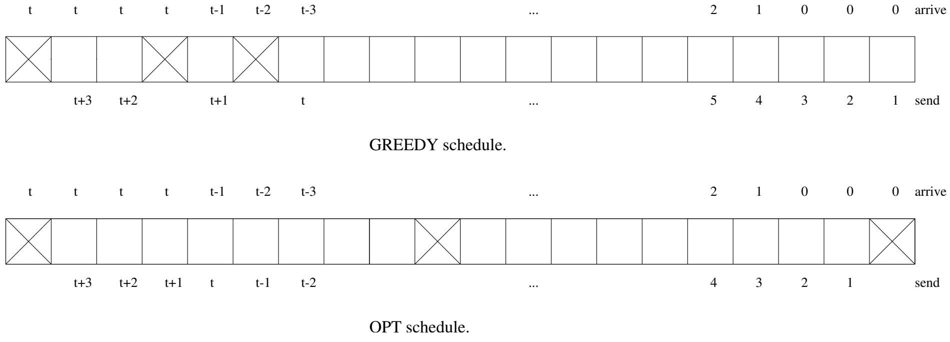

Fig. 1. Outcoming video streams of Greedy and OPT (moderate bursts)

Large bursts. In presence of massive bursts both OPT and the Greedy Policy are forced to drop consecutive frames. In this case the loss of OPT is uniformly divided between lost blocks while the Greedy Policy may create larger lost blocks. However, they are balanced later by future discards. Suppose that at time 0 the buffer is empty and a burst of 3 frames arrives. During the following t time units one frame is sent by the Greedy Policy and one frame arrives. Finally, at time t a burst of 8 frames arrives. Notice that the Greedy Policy would have its buffer full and OPT would have its buffer empty at this time. Now the Greedy Policy and OPT would discard 7 and 5 frames, respectively. The resulting video streams are shown in Figure 2.

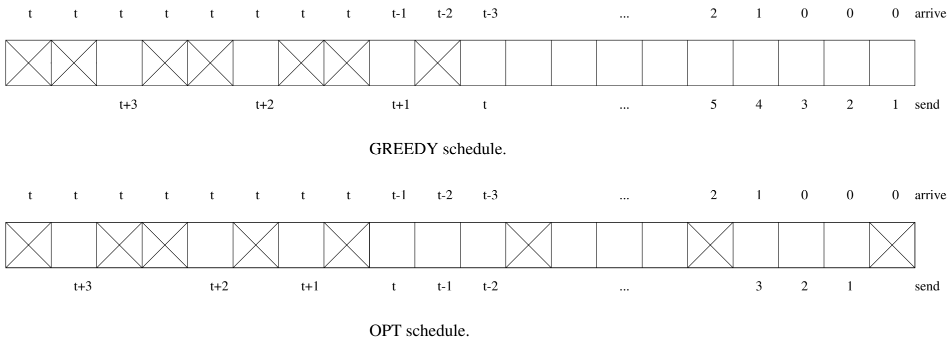

Fig. 2. Outcoming video streams of Greedy and OPT (large bursts)

5 ANALYSIS OF GREEDY POLICY

In this section we analyze the performance of the Greedy Policy. The following definition introduces the worst-case noise produced by an optimal offline policy for the case of moderate and large bursts.

Definition 5.1. Let the minimal distance between two neighboring discarded frames produced by any policy when k frames are lost during time interval of length s be denoted by $d_{OPT}(k,s)$. Let the maximal length of a lost block produced by any policy when k frames are lost during time interval of length s be denoted by $l_{OPT}(k,s)$[1].

5.1 Moderate Burst Size

We consider video streams generated by a (σ, ρ) source with $\sigma \leq \frac{5}{4}M$. We show that the competitive ratio of the Greedy Policy is at least L/U. To establish the competitive ratio we introduce an estimation of the optimal minimal distance between discarded frames. Then we derive the cost of the loss of the Greedy Policy using this estimation. In particular we demonstrate that the distance between two discarded frames under the Greedy Policy is at least half the estimation minus one, and show that it is almost optimal distance for a competitive online policy (up to a factor of 2). First we need some auxiliary lemmas.

Observation 5.1. When the number of frames lost in a time interval of s time units is k, the optimal minimal distance satisfies $d_{OPT}(k,s) \leq \lceil \frac{M+s}{k-1} \rceil$.

The observation holds since during s time units at most $M+s$ frames could be accepted. By the monotonicity and the convexity properties of ϕ_1, the cost function is minimized when the discarded frames are equally distributed, i.e., when the minimal distance between two discarded frames is maximized. A simple combinatorial argument shows that the maximum is achieved when discarded frames divide the $M+s$ accepted frames into $k-1$ balanced parts. The following lemma compares the performance of the Greedy Policy to d_{OPT}.

Lemma 5.2. When frames are scheduled according to the Greedy Policy and the size of the lost set is bounded by k, then the minimal distance between two adjacent discarded frames is at least $d_{OPT}(k,0)/2-1$.

Proof. Suppose by way of contradiction that the Greedy Policy discards a frame violating the condition of the lemma. Let the size of the lost set at this time be $m \leq k-1$. We assume that frames in the lost set are enumerated starting from the end of the sequence. Let d_0 denote the number of accepted frames in the buffer preceding the most recent discarded frame in the lost set, let d_i denote the distance between the i-th and $i+1$-th discarded frames in the lost set, and let d_m denote the number of accepted frames in the buffer succeeding the latest discarded frame in the lost set.

The Greedy Policy always tries to maximize the resulting minimal distance between two discarded frames by definition of well-behaved cost function. If the

[1] When there are no adjacent frames discarded in the time interval, then $l_{OPT}(k,s)=1$; when the maximum lost block in the time interval consists of two adjacent frames, then $l_{OPT}(k,s)=2$; etc.

condition of the lemma is violated, then the following holds. The distance between any pair of adjacent discarded frames is at most $d_{OPT}(k,0)-2$ and d_0 is not greater than $d_{OPT}(k,0)/2-1$. If it is not the case the Greedy Policy would have been able to discard a frame without having the minimal distance falling below $d_{OPT}(k,0)/2-1$.

Observe that the buffer is full and each of the frames currently in the buffer takes part in exactly one of the above distances. Thus, the sum of distances bounds M, and we have

$$M \le \sum_{i=0}^{m} d_i \le d_0 + (m-1)(d_{OPT}(k,0)-2) + d_m.$$

If there exists a discarded frame not in the lost set then $m \le k-2$ and $d_m \le d_{OPT}(k,0)-2$. In this case we obtain

$$\begin{aligned} & d_0 + (m-1)(d_{OPT}(k,0)-2) + d_m \\ \le\ & d_{OPT}(k,0)/2 - 1 + (k-3)(d_{OPT}(k,0)-2) + d_{OPT}(k,0) - 2 \\ \le\ & (k-1)(d_{OPT}(k,0)-2). \end{aligned}$$

If all the discarded frames belong to the lost set then $d_m \le d_{OPT}(m,0)/2-1$ and $m \le k-1$. In this case we get

$$\begin{aligned} & d_0 + (m-1)(d_{OPT}(k,0)-2) + d_m \\ \le\ & d_{OPT}(k,0)/2 - 1 + (k-2)(d_{OPT}(k,0)-2) + d_{OPT}(k,0)/2 - 1 \\ \le\ & (k-1)(d_{OPT}(k,0)-2). \end{aligned}$$

Consequently,

$$d_{OPT}(k,0) \ge \tfrac{M}{k-1} + 2,$$

which contradicts Observation 5.1. □

The following claim states that the Greedy Policy does not discard consecutive frames for the case of moderate bursts.

Claim 5.3. For any sequence generated by a (σ,ρ)-source with $\sigma \le \frac{5}{4}M$, the Greedy Policy does not discard consecutive frames.

Proof. Observation 5.1 and Lemma 5.2 imply that the minimal distance between two adjacent discarded frames is at least 1. □

In the next theorem we show that the Greedy Policy is L/U-competitive. Notice that the Greedy Policy schedules a maximum number of frames.

Theorem 5.4. For any sequence generated by a (σ,ρ)-source with $\sigma \le \frac{5}{4}M$, the competitive ratio of the Greedy Policy is at least L/U.

Proof. By Claim 5.3, the Greedy Policy incurs cost of at most U per frame whereas OPT discards the same number of frames with cost of at least L. □

Next we show that there exists a sequence of frames such that no online policy could improve the $d_{OPT}(M/4, 0)$ distance estimation while OPT is able to discard frames which are arbitrarily far apart. The intuition behind the proof is that no online policy can discard frames at time of underflow while being competitive. That means that a competitive online policy might receive a burst of maximum size when its buffer is full, while OPT could smoothly distribute the discarded frames throughout the entire sequence.

Theorem 5.5. There exists an input sequence generated by a (σ, ρ)-source with $\sigma \leq \frac{5}{4}M$ so that any online policy discards frames at distance of at most $d_{OPT}(M/4, 0)$ while in OPT the distance between any two discarded frames is arbitrarily large.

Proof. Suppose that frames are scheduled according to an online policy A. We construct a sequence of frames so that, on the one hand, OPT either does not discard any frame or the distance between two discarded frames is arbitrarily large and, on the other hand, A either discards at least one frame or discards $M/4$ frames that are very close.

Consider the following scenario. At time $t_s = 0$ the buffer is empty and M frames arrive. Then during sufficiently long scheduling period till time t_f one frame arrives every time unit. The online policy A cannot discard any of these frames without being 0-competitive.

A burst of $M/4$ frames arrives at the end of the period (at time t_f). Note that since A did not discard any frame its buffer is full. Henceforth, A necessarily discards $M/4$ frames when the burst arrives at time t_f. According to Observation 5.1 the optimal distance between two discarded frames is $d_{OPT}(M/4, 0)$. However, OPT could have evenly distributed the $M/4$ discarded frames until time t_f and accept all the burst of $M/4$ frames. Therefore, the distance between two discarded frames is $d_{OPT}(M/4, t_f - t_s)$, which can be made arbitrarily large. $\square$

The following is an immediate consequence of Theorem 5.5.

Corollary 5.6. The competitive ratio of any online policy is at most L/U.

5.2 Large Burst Size

We consider video streams generated by a (σ, ρ)-source with $\sigma > \frac{5}{4}M$. This is an interesting case for bursty VBR video streams. Notice that now the Greedy Policy may be forced to discard consecutive frames. In a similar spirit to the distance between discarded frames, we obtain bounds on the lengths of lost blocks. The problem here is that the dependence of a cost of such a sequence on its length is not linear but quadratic. This means that the cost of loss may increase drastically when σ increases.

Nevertheless, it turns out that in this case OPT also suffers large loss. We establish that the competitive ratio of the Greedy Policy is $\frac{LL'}{(12L+L')U'}$. In order to show this we introduce an estimation of the optimal maximal length of a lost block.

Then we derive the cost of the loss of the Greedy Policy and OPT in terms of this estimation. Moreover, we show that the length of a lost block produced by the Greedy Policy is at most larger by factor of 4 than the lower bound for OPT. Before we prove the main theorem we show a few lemmas. The following observation states the value of l_{OPT}.

Observation 5.7. When the number of frames lost throughout a time interval of s time units is k, then the maximal length of a lost block satisfies $l_{OPT}(k,s) = \lceil \frac{k}{M+s+1} \rceil$.

The observation holds since during s time units at most $M+s$ frames could be accepted. A simple combinatorial argument shows that any schedule will have a lost block of size at least $\left\lceil \frac{k}{M+s+1} \right\rceil$.

Lemma 5.8. When frames are scheduled according to the Greedy Policy and the size of the lost set is bounded by k, then the maximal length of a lost block is at most $2l_{OPT}(k,0)+1$.

Proof. Suppose by way of contradiction that the Greedy Policy discards a frame that arrived at time t and violates the condition of the lemma. The Greedy Policy always maintains the minimal possible length of the maximal lost block by definition of a well-behaved cost function. Therefore, if the condition of the lemma is violated then we have the following.

The cumulative size of the lost blocks in the lost set adjacent to an accepted frame in the buffer is at least $2l_{OPT}(k,0)+1$. If it is not the case the Greedy Policy would have been able to discard a frame without creating a lost block of length greater than $2l_{OPT}(k,0)+1$ (it would have discarded the frame between the two lost blocks instead of discarding the frame which violates the condition of the lemma). When we consider lost blocks adjacent to the accepted frames in the buffer, each lost block is counted at most twice. In addition we have a lost block of length at least $2l_{OPT}(k,0)+1$ that is adjacent to the most recent accepted frame solely. Thus, the sum of the lengths of all the lost blocks is at least $\left(\frac{M-1}{2}+1\right)(2l_{OPT}(k,0)+1) = \frac{M+1}{2}(2l_{OPT}(k,0)+1)$.

However, the size of the lost set is upper bounded by k; therefore

$$k \geq \frac{M+1}{2}(2l_{OPT}(k,0)+1),$$

implying that

$$l_{OPT}(k,0) \leq \frac{k}{M+1} - \frac{1}{2},$$

which contradicts Observation 5.7. □

For simplicity we assume in the sequel that $l_{OPT}(k,s) = \frac{k}{M+s+1}$.

Observation 5.9. The cost of k discarded frames divided into lost blocks of length l under a well-behaved function is bounded from below by $L' \cdot \frac{l+1}{2}k$ and from above by $U' \cdot \frac{l+1}{2}k$.

Proof. Consider a lost block of length l and let us denote the i-th frame within the block by f_i. By the definition of a well-behaved cost function, the cost of f_i is bounded from below and from above by $L' \cdot i$ and $U' \cdot i$, respectively. Thus, summing over all frames in the block we get

$$L' \cdot \frac{l(l+1)}{2} \leq \sum_{i=1}^{l} \phi_2(f_i) \leq U' \cdot \frac{l(l+1)}{2}.$$

Since there are k/l such blocks, the total cost is at least $L' \cdot \frac{l+1}{2}k$ and at most $U' \cdot \frac{l+1}{2}k$. $\square$

The next lemma shows that for large bursts OPT unavoidably pays the cost of producing sizable lost blocks.

Lemma 5.10. When frames are scheduled according to the Greedy Policy and the size of the lost set is $k > M$ then OPT discards at least $k - M$ frames from the lost set and pays cost of at least $L'(l_{OPT}(k-M, M)+1)(k-M)/2$.

Proof. The lost set contains frames that arrived during the last M time units. Observe that the number of frames accepted throughout this time interval by the Greedy Policy and by OPT may differ by at most M frames, i.e., OPT is able to accept M additional frames if its buffer is empty at the beginning of the interval. The lemma follows by Observation 5.7 and Observation 5.9. $\square$

The next proof uses a technique of "k-matching" between loss of the Greedy Policy and OPT. In k-matching the lost blocks of the online policy are divided into disjoint sets that are matched to sets of frames lost by OPT so that each frame lost by OPT appears in at most k such sets. The lost blocks of the Greedy Policy are matched either by 1-matching or by 3-matching. Evidently, the competitive ratio of the Greedy Policy is at least the minimum of the minimum ratio of 1-matched sets and one third of the minimum ratio among 3-matched sets.

Theorem 5.11. For any sequence generated by a (σ, ρ)-source with $\sigma > \frac{5}{4}M$, the competitive ratio of the Greedy Policy is at least $\frac{LL'}{(12L+L')U'}$.

Proof. We divide the schedule of the Greedy Policy into intervals of length M. Let us consider lost sets at the last time moment of every interval. Note that each discarded frame belongs to exactly one of these sets. First we identify and match the lost sets of the Greedy Policy that participate in 3-matching. Having finished with 3-matching, the remaining lost sets participating in 1-matching are arbitrarily matched to sets of the same cardinality formed from the remaining unmatched frames lost by OPT. Notice that both Greedy Policy and OPT lose the same number of frames. Thus, there always remains a sufficient number of remaining unmatched frames lost by OPT for 1-matching provided that 3-matched online lost sets have cardinality smaller than their offline counterparts.

We bound the cost of a lost set of the processed interval by determining the maximal length of a lost block to which the frames from the interval's lost set could belong. Suppose that we process the i-th interval $[Mi, M(i+1))$. Let k be the cardinality of the lost set of the interval (i.e. lost set at time $M(i+1)-1$) and let t_{max} be the time within the interval, $M(i-1) \le t_{max} < M(i+1)$, in which the cardinality k_{max} of the lost set is maximal. Clearly, frames from the lost set could belong to the maximal lost block whose length is given by the estimation of Lemma 5.8 at time t_{max}, that is $2l_{OPT}(k_{max}, 0) + 1$. We define a threshold value of $T = C \cdot M$, for a constant C. The relation between T and k_{max} determines the type of matching used. Finally, we choose the constant C so as to optimize the obtained competitive ratio. We consider two cases.

(1) If $k_{max} > C \cdot M$ then the lost set of the current interval takes part in 3-matching. We match the lost set to a set of $k_{max} - M$ frames that were necessarily lost by OPT during time interval $[t_{max} - M - 1, t_{max}]$. This is indeed a 3-matching since each of the offline lost frames could participate in at most three such sets, that is the matched set of the interval and the corresponding matched sets of the adjacent intervals.

Lemma 5.10 implies that the cost incurred by OPT is at least

$$L'(l_{OPT}(k_{max} - M, M) + 1)(k_{max} - M)/2.$$

At the same time by Observation 5.9 the cost incurred by the Greedy Policy is at most

$$U'(l_{OPT}(k_{max}, 0) + 1)k \le U'(l_{OPT}(k_{max}, 0) + 1)k_{max}.$$

Hence, the ratio between the cost of the offline and the corresponding online lost sets is

$$\begin{aligned} & \frac{L'(l_{OPT}(k_{max} - M, M) + 1)(k_{max} - M)/2}{U'(l_{OPT}(k_{max}, 0) + 1)k_{max}} \\ = \; & \frac{L'((k_{max} - M)/(2M+1) + 1)(k_{max} - M)/2}{U'(k_{max}/(M+1) + 1)k_{max}} \\ = \; & \frac{L' \frac{k_{max}+M+1}{2M+1}(k_{max} - M)/2}{U' \frac{k_{max}+M+1}{M+1} k_{max}} = \frac{L'(M+1)(k_{max} - M)}{2U'(2M+1)k_{max}} \\ > \; & \frac{L'(M+1)(k_{max} - M)}{2U'(2M+2)k_{max}} = \frac{L'(k_{max} - M)}{4U' k_{max}} \\ = \; & \frac{L'}{4U'}(1 - M/k_{max}) > \frac{L'}{4U'}(1 - 1/C) = \frac{(C-1)L'}{4CU'}. \end{aligned}$$

(2) If $k_{max} \leq C \cdot M$ then the lost set of the current interval takes part in 1-matching. By Observation 5.7 the optimal maximal length of a lost block is at most C. Applying Lemma 5.8 we obtain that the maximal length of a lost block of discarded frames for the Greedy Policy is at most $2C + 1$. Therefore, according to Observation 5.9 the cost of the lost set is upper-bounded by $(C + 1)U'k$. The ratio is kept above $\frac{L}{(C+1)U'}$ since each lost frame has cost at least L.

The competitive ratio of the Greedy Policy is the minimum of the ratio of 1-matched sets and one third the ratio of 3-matched sets. To derive the optimal C we equate these ratios:

$$\frac{(C-1)L'}{12CU'} = \frac{L}{(C+1)U'},$$

$$C^2 - \frac{12L}{L'}C - 1 = 0,$$

$$C = \frac{6L}{L'} + \sqrt{(\frac{6L}{L'})^2 + 1} \approx \frac{12L}{L'}.$$

Therefore, the optimized competitive ratio is $\frac{L}{(\frac{12L}{L'}+1)U'} = \frac{LL'}{(12L+L')U'}$.

□

Next we show that the length of a lost block created by the Greedy Policy is at most four times the optimal length of a lost block of OPT plus three.

Lemma 5.12. For any input sequence S the length of the maximal lost block satisfies $l_{GREEDY}(S) \leq 4 \cdot l_{OPT}(S) + 3$.

Proof. By Lemma 5.8 the maximal length of a lost block produced by the Greedy Policy is at most $2l_{OPT}(k, 0)+1$ when the size of the lost set is k. At the same time by Lemma 5.10 the optimal length of a lost block in this case is at least $l_{OPT}(k-M, M)$. The theorem follows since

$$2l_{OPT}(k, 0) + 1 \leq 4l_{OPT}(k - M, M) + 3.$$

□

We conclude with the following general theorem.

Theorem 5.13. For any sequence of frames, the competitive ratio of the Greedy Policy is at least $\min(L/U, \frac{LL'}{(12L+L')U'})$.

6 MPEG VIDEO STREAMS

In this section we discuss how to schedule MPEG video streams employing the Greedy Policy. In processing a video stream, the MPEG encoder produces three types of frames. The first type, I, are called intra coded frames. These are the simplest type of frame containing a coded representation of a still image to provide the decoder a starting point for decoding the next group of frames. P frames are the next type, called predicted frames. When decoding, they are created from information contained within the previous P or I frame. The last type of frame is the most common type, the B or bi-directional frame. B-frames are both forward and backward predicted and they are constructed from the previous and the next P or I frame. Finally, the encoder breaks up a sequence into Groups of Pictures (GOPs), with an I-frame at the beginning of each GOP, e.g. $IBBPBBPBBPBB$.

We can extend the cost of a frame to include the cost of all the dependent frames. This means that the cost of I frame will include the cost of dropping the whole GOP while the cost of P frame will subsume the cost of dropping the dependent B frames. In addition, the cost function could be modified to assign weights to frames with respect to their types. As a result, the Greedy Policy would first drop B frames. If that action is not sufficient, the Greedy Policy drops P frames as well, and in the worst case, it would drop all B and P frames as well as some I-frames. Observe that the Greedy Policy always tries to optimize the global QoS, while drop filters proposed earlier optimize only the local QoS within GOPs.

7 CONCLUSION

In this work we study competitive online buffering policies for video transmission across internetwork with leaky bucket constraints. First we consider video encoding schemes with independent frames. To measure the QoS we define a well-behaved cost function reflecting the playback discontinuity at the client. For moderate and large burst parameters we derived the competitive ratio of the Greedy Policy. In addition to the global QoS function we consider local metrics, such as the minimal distance between two discarded frames and maximal length of a sequence of consecutive discarded frames. The Greedy Policy is shown to be competitive with regard to these metrics as well. Then we demonstrate how our model can be extended to process MPEG video streams with complex inter-frame dependencies.

The proposed policy may be used for managing current Internet routers that wish to provide QoS. Some interesting future directions include studying more sophisticated cost functions and performing simulations in which performance of online dropping policies is estimated more properly using a global QoS function.

REFERENCES

[1] Bhattacharjee, S.—Calvert, K.—Zegura, E. W.: On Active Networking and Congestion. Technical Report GIT-CC-96-02, College of Computing, Georgia Tech.

[2] Borodin, A.—El-Yaniv, R.: Online Computation and Competitive Analysis. Cambridge University Press, 1998.

[3] Cruz, R. L.: A Calculus for Network Delay, Part 1. Network Elements in Isolation. IEEE Transactions on Information Theory, Vol. 37, 1991, No. 1, pp. 114–131.

[4] Feng, W.—Rexford, J.: Performance Evaluation of Smoothing Algorithms for Transmitting Prerecorded Variable-Bit-Rate Video. IEEE Transactions on Multimedia, September 1999, pp. 302–313.

[5] Feng, W.—Sechrest, S.: Smoothing and Buffering for Delivery of Prerecorded Compressed Video. Computer Communications, pp. 709–717, October 1995.

[6] Garrett, M.—Willinger, W.: Analysis, Modeling and Generation of Self-Similar VBR Video Traffic. Proceedings of ACM SIGCOMM, pp. 269–280, September 1994.

[7] Hemy, M.—Hengartner, U.—Steenkiste, P.—Gross, T.: MPEG System Streams in Best-Effort Networks. Proceedings of PacketVideo '99, April 1999.

[8] Hsu, C.-Y.—Ortega, A.—Reibman, A. R.: Joint Selection of Source and Channel Rate for VBR Video Transmission under ATM Policing Constraints. IEEE Journal on Sel. Areas in Communications, Vol. 15, 1997, pp. 1016–1028.

[9] Lakshman, T. V.—Ortega, A.—Reibman, A. R.: Variable Bit Rate (VBR) Video. Tradeoffs and potentials. Procedeends of the IEEE, Vol. 86, 1997, pp. 741–755.

[10] Mansour, Y.—Patt-Shamir, B.: Jitter Control in QoS Networks. 39th Annual Symposium on Foundations of Computer Science, 1998.

[11] Mansour, Y.—Patt-Shamir, B.—Lapid, O.: Optimal Smoothing Schedules for Real-Time Streams. PODC 2000.

[12] Ramanathan, S.—Rangan, P. V.—Vin, H. M.: Frame-Induced Packet Discarding. An Efficient QoS Management Strategy for Video Networking. Fourth International Workshop on Network and Operating System Support for Digital Audio and Video, Lancaster University, November 1993.

[13] Rexford, J.—Sen, S.—Dey, J.—Feng, W.—Kurose, J.—Stankovic, J.—Towsley, D.: Online Smoothing of Live, Variable-Bit-Rate Video. Proceedings of International Workshop on Network and Operating Systems Support for Digital Audio and Video, pp. 249–257, May 1997.

[14] Rexford, J.—Towsley, D.: Smoothing Variable-Bit-Rate Video in an Internetwork. IEEE/ACM Transactions on Networking, pp. 202–215, April 1999.

[15] Del Rosario, J. M.—Fox, G. C.: Constant Bit Rate Network Transmission of Variable Bit Rate Continuous Media in Video-On-Demand Servers. Multimedia Tools and Applications, Vol. 2, pp. 215–232, May 1996.

[16] Salehi, J. D.—Zhang, Z.-L.—Kurose, J. F.—Towsley, D.: Supporting Stored Video. Reducing rate variablity and end-to-end resource requirements through optimal smoothing. Proceedings of ACM SIGMETRICS, pp. 222–231, May 1996.

[17] Sen, S.—Rexford, J.—Dey, J.—urose, J.—Towsley, D.: Online Smoothing of Variable-Bit-Rate Streaming Video. IEEE Transactions on Multimedia, pp. 37–48, March 2000.

[18] Sleator, D.—Tarjan, R.: Amortized Efficiency of List Update and Paging Rules. CACM 28, pp. 202–208, 1985.

[19] Veres, A.—Boda, M.: The Chaotic Nature of TCP Congestion Control. Proceedings of IEEE INFOCOM 2000, pp. 1715–1723.

[20] Wijesekera, D.—Srivastava, J.: Quality of Service (QoS) Metrics for Continuous Media. Multimedia Tools and Applications, Vol. 2, pp. 127–166, September 1996.

[21] Wrege, D. E.—Knightly, E. W.—Zhang, H.—Liebeherr, J.: Deterministic Delay Bounds for VBR Video in Packet-Switching Networks. Fundamental Limits and Practical Tradeoffs. IEEE/ACM Transactions on Networking, Vol. 4, 1996, No. 3, pp. 352–362.

[22] Wrege, D. E.—Liebeherr, J.: Video Traffic Characterization for Multimedia Networks with a Deterministic Service. Proceedings of IEEE INFOCOM 96, pp. 537–544.

[23] Yeadon, N.—Garcia, F.—Hutchison, D.—Shepherd, D.: Filters. QoS Support Mechanisms for Multipeer Communications. IEEE Journal on Selected Areas in Computing (JSAC) special issue on Distributed Multimedia Systems and Technology, Vol. 14, 1996, No. 7, pp. 1245-1262.

[24] Zhang, Z. L.—Nelakuditi, S.—Aggrawal, R.—Tsang, R. P.: Efficient Selective Frame Discard Algorithms for Stored Video Delivery Across Resource Constrained Networks. Proceedings of IEEE INFOCOM 98.

Alex Kesselman completed his B.Sc. and M.Sc. degrees at the Ben Gurion University in 1995 and 1998, respectively. Following that he was a R & D staff member at Motorola Semiconductor Israel and CheckPoint Software Technologies. He currently studies towards Ph.D. degree in the Tel Aviv University. His research interests include communication networks, online algorithms and Quality of Service.

Yishay Mansour completed his BA and his MSc at the Technion in 1985 and 1987 and his PhD at MIT in 1990. Following that he was a postdoctoral fellow at Harvard University and a research staff member at IBM T. J. Watson Research Center. Since 1992 he is with the Computer Science Department, Tel-Aviv University, where he established (with Prof. Afek) the networking laboratory. He has written numerous papers in all areas of computer science with special emphasis on communication networks, machine learning, and theory of computation. He has supervised over a dozen of graduate students, served on numerous program committees and is currently an editor in three leading journals.