Optimal Packet Length Scheduling for Regulated Media Streaming

Premal Shah and Abhay Karandikar

Abstract— **We present here, for a prerecorded stream, an algorithm to determine a packet length schedule that is optimal with respect to appropriate measures of distortion and quality smoothness, when the traffic is required to adhere to specified Token Bucket Regulation constraints.**

Index Terms— **QoS, Packet Length Schedule, Token Bucket Traffic Regulation, Streaming**

I. INTRODUCTION

Networks that offer Quality of Service (QoS) guarantees, regulate the traffic sources using Traffic Regulators so that the resulting flows adhere to certain constraints. A simple Traffic Regulator that enforces a linearly bounded flow is the Token Bucket (or Leaky Bucket) Regulator which has two parameters - the token generation rate r and the size of the token bucket B [1], [2].

For Variable Bit Rate (VBR) traffic such as video streams, the selection of appropriate values for r and B can be both difficult as well as inefficient in terms of resource utilization [3], [4]. Further, the source would not always be able to procure for itself, its exact requirements from the network. Media streaming applications, though delay sensitive, are somewhat flexible in terms of their packet loss requirements and allow for a graceful degradation of quality with loss.

In this letter, we consider the problem where a source has been offered QoS guarantees subject to a particular Token Bucket Regulation. Further, by using flexible coding schemes, the source may schedule the size of its packets to adapt to the Token Bucket Regulation. For an offline (prerecorded) stream, we derive a packet length schedule that is optimal with respect to an appropriate measure of distortion and quality jitter as perceived by the receiver, while still honoring the imposed traffic constraints. We propose that MINMAX loss (in absolute or fractional terms) over sets of contiguous frames spanning disjoint time-intervals may be used as an effective combined measure of distortion and quality smoothness (consistency) of the stream.

II. THE MODEL

Consider a source that transmits for a finite duration N, at discrete times $n = 1, 2, ..., N$. The source is constrained by a generalized Token Bucket Regulator, henceforth denoted by $TBR(\boldsymbol{r}, \boldsymbol{B}, B_0)$. Here, we allow the token increments and the maximum number of residual tokens to vary with time n and denote them as vectors $\boldsymbol{r} = (r_1, r_2, ..., r_N)$ and $\boldsymbol{B} = (B_1, B_2, ..., B_N)$ respectively. B_0 is the initial token grant. The standard IETF Token Bucket Regulator is a special case realized by setting $r_n = r$, $B_n = B \quad \forall n : 1 \leq n \leq N$; r, B being constants. We denote this regulator as simply $TBR(r, B, B_0)$.

The source is characterized by a requirement schedule vector $\boldsymbol{y} = (y_1, y_2, ..., y_N)$. y_n denotes the number of bits required by the source to code the media content for interval n (defined to be from time $n-1$ to n)[1] to a desired level of quality. The transmission is termed 'offline' if each y_n is known apriori, e.g., streaming of a prerecorded video. Let $x_n : \boldsymbol{x} = (x_1, x_2, ..., x_N)$ denote the number of bits transmitted at time n. $TBR(\boldsymbol{r}, \boldsymbol{B}, B_0)$ constrains x_n so that

$$x_n \leq t_{n-1} + r_n \quad ; \forall n : 1 \leq n \leq N \tag{1}$$

Let $t_n : \boldsymbol{t} = (t_1, t_2, ..., t_N)$ be the number of tokens left in the bucket just after the n^{th} transmission. If all packet sizes x_n satisfy (1), i.e., the source is conforming, then t_n will evolve as

$$\begin{aligned} t_n &= \min(t_{n-1} + r_n - x_n, B_n) \; ; \\ t_0 &= B_0 \end{aligned} \tag{2}$$

We assume a zero effective playback buffer/delay scenario, i.e., the information pertaining to the interval n is useful only if transmitted by the source at n[2]. A conformant transmission is assumed to be error-free and at a constant delay.

We now consider the distortion associated with the transmission. For the zero-delay case, against a requirement of y_n bits only x_n are chosen to be transmitted. We define, for this case, the absolute loss α_n and fractional loss β_n in coding of the media content in the interval n

$$\alpha_n = \max(y_n - x_n, 0) \tag{3}$$

$$\beta_n = \max(\frac{y_n - x_n}{y_n}, 0) \tag{4}$$

We assume that the resulting perceptual distortion can be modeled as a function $d_\alpha(.)$ of α_n or a function $d_\beta(.)$ of β_n. The overall distortion D_α (or D_β) will be given by sum of the distortions in each of these non-overlapping time intervals, i.e., by

The authors are with the Department of Electrical Engineering, Indian Institute of Technology (IIT) Bombay, Mumbai 400 076, India (e-mail: {prem,karandi}@ee.iitb.ac.in).

[1] In general, we may consider the transmissions to take place at times $\boldsymbol{\tau} = (\tau_1, \tau_2, ..., \tau_N)$ where the time intervals are not constant. The model still applies, as the token increment r_n can vary with n, where n now denotes the time index.

[2] The analysis can easily be extended to the case of non-zero playback (absolute) delay bounds. We do not present this case here for the want of space.

(5) (or (6)).

$$D_\alpha = \sum_{n=1}^{n=N} d_\alpha(\alpha_n) \quad (5)$$

$$D_\beta = \sum_{n=1}^{n=N} d_\beta(\beta_n) \quad (6)$$

We do not assume any specific form for $d_\alpha(.)$ and $d_\beta(.)$ but only require them to be strictly increasing, convex and time-invariant[3]. The convexity of the distortion functions is a reasonable assumption, and follows from the nature of the Rate Distortion curves encountered in practice.

We now define a MINMAX type of measure on loss that will help model distortion even without the precise knowledge of $d_\beta(.)$ (or $d_\alpha(.)$). Consider the ordered loss vectors $\boldsymbol{\alpha'}, \boldsymbol{\beta'}$ - these are vectors of the loss at different times sorted in a decreasing order. Thus for fractional loss,

$$\beta'_1 \geq \beta'_2 \ldots \geq \ldots \beta'_N$$

in particular,

$$\beta'_1 = \max_{1\leq n\leq N} \beta_n \quad (7)$$

In our work, we seek a packet length schedule $\boldsymbol{x}^*$ that gives the minimum $\boldsymbol{\beta'}$ (or $\boldsymbol{\alpha'}$), comparisons made in lexicographic order, for a given requirement schedule $\boldsymbol{y}$ and $TBR(\boldsymbol{r}, \boldsymbol{B}, B_0)$ constraints.

This measure captures the idea of quality smoothness because a packet length schedule so optimized (i.e., for MINMAX loss) will also have optimal MINMAX distortion. Further, since the kind of $d_\beta(.)$ or $d_\alpha(.)$, that we are likely to encounter would show convexity of a high degree (typically superexponential), the schedule would also be close to optimal in terms of overall distortion of the stream. In fact, for the offline case, if the measure of distortion is based on absolute loss, i.e., D_α, the packet length schedule that optimizes $\boldsymbol{\alpha'}$ will also optimize D_α for any convex function $d_\alpha(.)$.

In the next section, we present an algorithm that gives the optimal $\boldsymbol{\beta'}$ for the offline case. An algorithm that would give the optimal $\boldsymbol{\alpha'}$ is similar with minor modifications. For mathematical convenience, it is assumed that $\boldsymbol{r}, \boldsymbol{B}, \boldsymbol{x}, \boldsymbol{y}, \boldsymbol{t}$ as well as $\boldsymbol{\alpha}, \boldsymbol{\beta} \in \{\mathcal{R}^+ \cup \{0\}\}^N$ rather than $\{\mathcal{Z}^+ \cup \{0\}\}^N$.

III. Optimal Packet Length Scheduling

Given the regulator $TBR(\boldsymbol{r}, \boldsymbol{B}, B_0)$, a packet length schedule $\boldsymbol{x}$ is conformant *iff* the packet lengths satisfy the set of constraints,

$$\sum_{n=i}^{n=j} x_n \leq B_{i-1} + \sum_{n=i}^{n=j} r_n \quad \forall (i,j) : 1 \leq i \leq j \leq N. \quad (8)$$

Also since the delay bound is zero, we consider only those packet length schedules $\boldsymbol{x}$ that satisfy

$$x_i \leq y_i \quad \forall i : 1 \leq i \leq N \quad (9)$$

[3]We do not consider any temporal data dependencies that may be present between frames belonging to different time intervals, as in video encoding, to keep the analysis simple.

The set of all allocation vectors $\boldsymbol{x}$ satisfying these constraints (8) and (9) are said to be admissible and the corresponding loss vectors $\boldsymbol{\beta}$ (or $\boldsymbol{\alpha}$) are said to be feasible. We seek to minimize $\boldsymbol{\beta'}$. Define

$$\gamma_{ij} = \max\{\frac{\sum_{n=i}^{n=j}[y_n - r_n] - B_{i-1}}{\sum_{n=i}^{n=j} y_n}, 0\} \quad (10)$$

$$\gamma^* = \max_{(i,j):1\leq i\leq j\leq N} \{\gamma_{ij}\} \quad (11)$$

$$(i^*, j^*) = \arg \max_{(i,j):1\leq i\leq j\leq N} \{\gamma_{ij}\} \quad (12)$$

Lemma 1: For an admissible allocation,

$$\beta'_1 \geq \gamma^* \quad (13)$$

Proof: We will prove this by contradiction. Assume that (13) does not hold then

$$\begin{aligned} \beta'_1 &< \gamma_{ij} \;\; for \;\; some \;\; (i,j) : 1 \leq i \leq j \leq N \\ as \;\; \beta_n &\leq \beta'_1 \;\; \forall n : 1 \leq n \leq N \\ \Rightarrow \sum_{n=i}^{n=j} \beta_n . y_n &< \sum_{n=i}^{n=j} [y_n - r_n] - B_{i-1} \\ \Rightarrow \sum_{n=i}^{n=j} x_n &> B_{i-1} + \sum_{n=i}^{n=j} r_n \end{aligned}$$

which would contradict (8). Hence the proof. □

Lemma 2: If $\boldsymbol{x}^*$ and $\boldsymbol{\beta'}^*$ denote the optimal packet length schedule and ordered fractional loss vector respectively then

$$\beta'^*_1 = \gamma^* \quad (14)$$

$$x^*_n = y_n(1 - \gamma^*) \quad \forall n : i^* \leq n \leq j^* \quad (15)$$

Proof: From Lemma 1, Lemma 2 would follow if (14) were feasible and (15) were a necessary condition for the same. The feasibility of (14) may be verified by considering a loss vector $\boldsymbol{\beta}$ where $\beta_i = \gamma^* \quad \forall i : 1 \leq i \leq N$ and using (12) and the admissibility constraints (8), (9). Claim (15) is equivalent to

$$\beta^*_n = \beta'^*_1 = \gamma^* \;\; \forall n : \; i^* \leq n \leq j^*$$

indeed, if this is not so,

$$\begin{aligned} \beta^*_n &< \beta'^*_1 = \gamma^* \;\; for \;\; some \;\; n : \; i^* \leq n \leq j^* \\ \Rightarrow \sum_{n=i^*}^{n=j^*} \beta^*_n . y_n &< \sum_{n=i^*}^{n=j^*} [y_n - r_n] - B_{i^*-1} \end{aligned}$$

Once again, this contradicts (8). Hence the proof. □

We present the recursive Algorithm 1 that gives the optimal packet length schedule to minimize $\boldsymbol{\beta'}$. The optimality follows from the lemmas proved above.

For illustration purposes, we have compared the performance of the above (Algorithm 1) with a naive algorithm. The naive allocation scheme would proceed by allocating as many of the required bits as possible at each n, i.e., $x_n = \min(t_{n-1} + r_n, y_n)$. The simulations have been performed on requirement vectors derived from high quality MPEG4 video traces obtained from [5]. The Token Bucket Regulator parameters were chosen to fall slightly short of the requirements. A typical distribution of

Algorithm 1 Algorithm for Offline Packet Length Scheduling

Inputs: vectors $\boldsymbol{y} := y_1^N, \boldsymbol{r} := r_1^N, \boldsymbol{B} := B_1^N; B_0 := B_0$
(a_p^q denotes the vector $(a_p, a_{p+1}, ..., a_q)$)
Output: $\boldsymbol{x}^*$

Compute γ^* and (i^*, j^*)
for all $n : i^* \leq n \leq j^*$ **do**
    $x_n^* := (1 - \gamma^*) y_n$
end for
if $\gamma^* = 0$ **then**
    for all $n : 1 \leq n \leq N$ **do**
        $x_n^* := y_n$
    end for
    return
end if
if $i^* > 1$ **then**
    $r_{i^*-1} := r_{i^*-1} - B_{i^*-1}$
end if
Apply Algorithm 1 with inputs $(y_1^{i^*-1}, r_1^{i^*-1}, B_1^{i^*-1}; B_0)$
Apply Algorithm 1 with inputs $(y_{j^*+1}^N, r_{j^*+1}^N, B_{j^*+1}^N; 0)$

fractional losses for the two allocations is shown in Figure 1. The fractional losses incurred by the optimal algorithm, as expected, are much more closely spaced than those by the naive scheme and the variance is less by a factor of about 14. We have verified through experiments that the scheme also leads to significant improvements in terms of overall distortion.

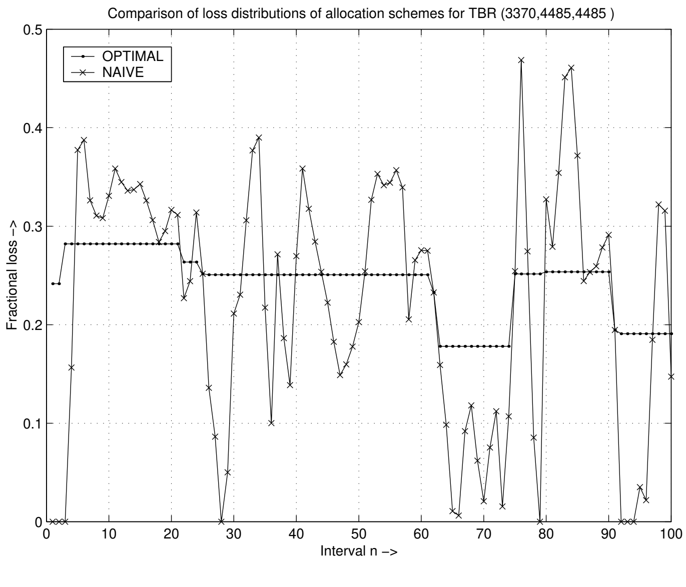

Fig. 1. β_n**s for different scheduling policies**

IV. Conclusions and Future Directions

In this letter, we have considered a problem where a media source that is regulated by traffic constraints, seeks to minimize the distortion and quality jitters in its transmission. The source may achieve this by sizing its packets appropriately. We have argued that MINMAX loss across coding intervals may be used as an objective function for this and presented an algorithm that gives the optimal packet length schedule based on the same. The proposed algorithm works for a generalized Token Bucket Regulator and is not necessarily optimal with respect to a time/space complexity measure. It may be possible to improve upon the complexity of this algorithm, especially when the regulator used is a standard IETF Token Bucket Regulator. The source may employ layered or even scalable coding to simplify implementation in terms of resizing packets. Although, we have considered a media stream for problem formulation, our analysis is applicable to any constrained traffic that exhibits the properties of loss tolerance.

It would be interesting to investigate similar packet length scheduling for the case of online transmission, where the decision about the length x_i must be based only on past and present inputs i.e., $y_j : j \leq i$ and apriori knowledge about the statistics of the input stream. In principle, this problem can be addressed within the Dynamic Programming framework but the solution would require a knowledge of the distortion function and would have a high time and/or space complexity. Determination of effective but implementable algorithms for the problem of online transmission is left as a future study.

References

[1] S Keshav, *An Engineering Approach to Computer Networking,* Addison Wesley, 2001

[2] Zheng Wang, *Internet QoS: Architectures and Mechanisms for Quality of Service,* Morgan Kaufmann Publishers, 2001

[3] P. Perry Tang and T. Tai, "Network Traffic Characterization using Token Bucket Model", *Proceedings of IEEE INFOCOM, March 1999*

[4] Constantinos Dovrolis, Maruthy Prasad Vedam and Parameswaran Ramanathan, "The Selection of the Token Bucket Parameters in the IETF Guaranteed Service Class", (1998, June) http://www.cc.gatech.edu/fac/Constantinos.Dovrolis/Papers/tokbuck.ps

[5] http://www-tkn.ee.tu-berlin.de/research/trace/ltvt.html