Optimization Bandwidth Sharing for Multimedia Transmission Supporting Scalable Video Coding

Mohammad S. Talebi
School of Computer Science, IPM
Tehran, Iran
Email: mstalebi@ipm.ir

Ahmad Khonsari
ECE Department, University of Tehran
School of Computer Science, IPM
Tehran, Iran
Email: ak@ipm.ir

Mohammad H. Hajiesmaili
ECE Department, University of Tehran
School of Computer Science, IPM
Tehran, Iran
Email: hajiesmaili@ipm.ir

***Abstract*—Wired and wireless data networks have witnessed a rapid proliferation of multimedia applications such as live-streaming applications, video conferencing, etc. A desirable key feature for multimedia transmission over multiuser environments with heterogeneous users is the ability of adapting rate and quality of video stream to different QoS conditions. The most efficient approach to address the scalability of multimedia applications is to encode video stream in compliance with Scalable Video Coding (SVC) standard, which is proposed as an extension to H.264/AVC standard. This paper addresses the utility-proportional optimization for multimedia applications that are relying on SVC-encoded video signals. We use the staircase utility function to analytically model the SVC-encoded multimedia applications and formulate the underlying optimization problem. Non-convexity of the optimization problem for such applications makes dual-based approaches incompetent, whereby achieving optimality proves quite challenging. We use a smooth approximation of the utility function to come up with a convex formulation and propose a dual-based distributed algorithm for rate allocation and bandwidth sharing in such scenarios. Numerical results are proposed as the support to the proposed rate control algorithm.**

I. Introduction

Over the past decades, data networks have witnessed a rapid proliferation in the usage of multimedia applications. Today, a plethora of such applications exists that include multimedia messaging, video telephony, and video conferencing over mobile TV, wireless and wired Internet video streaming, standard and high-definition TV broadcasting.

Emergence of these applications has led to new demands for QoS-provisioning in different networking scenarios. In order to tackle this issue, many technical challenges have to be addressed in the two areas of multimedia and networking.

A. Multimedia Transmission

For multimedia applications, the traffic characteristics of the underlying video stream usually demands for tight QoS requirements in terms of rate and delay. Throughput variation often occurs in many networking paradigms, ranging from wireless to wired ones. In wired networks such variations mainly stem from network congestion while in wireless networks this mainly occurs temporarily due to channel quality degradation caused by fading or shadowing. All of these make multimedia transmission over data networks quite challenging. On the other hand, a desirable key feature for multimedia transmission over multiuser environments is the ability to adapt the quality of the video signal on a per-user basis [1].

The essential remedy to both of the abovementioned issues, i.e. rate and quality adaptation, is to exploit video streams encoded in compliance with the Scalable Video Coding (SVC) standard [2], [3]. Scalable Video Coding (SVC) standard has been proposed as an extension to the famous H.264/AVC standard [4]. SVC features temporal, spatial and PSNR scalability of a decoded video signal through rate adaptation of the bit stream.

In this respect, from this point of view, SVC encoded-video streams are capable of adapting rate and quality so as to combat throughput variation and to exhibit several classes of quality for to be appropriate for heterogeneous clients.

B. Network Utility Maximization Approaches

Following the seminal work by Kelly *et al.* [5], the optimization flow control approach was proposed by Low *et al.* [6], in which the optimal rate allocation of a (wired) data network under elastic traffic was modeled and led to a dual-based distributed algorithm for rate allocation both in synchronous and asynchronous environments. Within the previous decade, the work by Low *et al.* [6] has been followed by the network research community and led to a general optimization framework known as Network Utility Maximization, abbreviated as NUM (see [7] and references therein). The underlying assumption of these works is that the network traffic is elastic, whose characteristics can be modeled by a continuous and strictly concave utility function [8]. Such utility functions, make the problem convex and thereby tractable for optimality analysis.

On the other hand, multimedia applications inherently admit inelastic flows which require tight performance guarantees in terms of rate and delay. For such flows, utility function is shown to be non-concave and even discontinuous [8]- [11]. Non-concavity and discontinuity of such utility functions yield non-convex and non-differentiable NUM problems, for which conventional optimization approaches proves quite incompetent. In particular, for multimedia applications supporting layered encoding schemes, the utility function is shown to admit a staircase-like function, as shown in Figure 1 in solid line [12]. Amongst such applications are streams complying

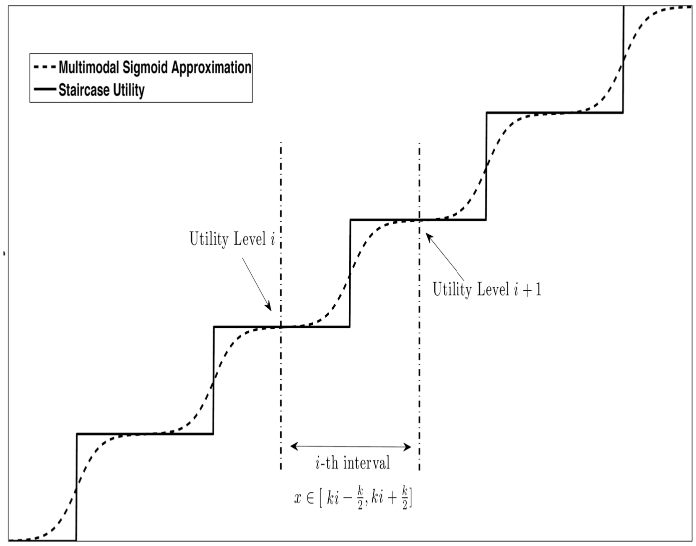

Fig. 1: Staircase Utility Function (solid line) and Multimodal Sigmoid Approximation (dashed line)

to SVC and FGS standards [13], both observed in the most of today's multimedia streaming applications. Therefore, as Figure 1 implies, SVC-encoded applications are associated with a non-concave and discontinuous utility function, whose generic form looks like the staircase-like function in Figure 1.

C. Our Contribution

Our contribution in this paper is two-fold. First, our focus is on modeling a smoothed and continuous utility function for SVC-encoded multimedia. Towards this, we introduce *multimodal sigmoid approximation* as a well-behaved utility function to be used for SVC-encoded multimedia. With the abovementioned approximation approach discontinuity of the staircase utility function will be remedied, however non-concavity problem still holds which later will result in a non-convex NUM. In order to come up with a convex formulation, we formulate the rate allocation optimization problem using the concept of *utility-proportional* optimization flow control proposed by Wang *et al.* [12] instead of the basic form of the NUM. Following this formulation allows us to devise a dual-based distributed rate allocation algorithm as the solution to the optimization problem. Numerical results present a proper validation of our endeavor in achieving a distributed solution.

D. Paper Organization

The rest of the paper is organized as follows. Section II reviews related work. Section III describes the system model. Section IV is devoted to approximate the ideal non-concave utility function and formulate the underlying optimization problem. Section V investigates the optimality condition and the optimal solution and presents a dual-based distributed algorithm as the solution to the optimization problem. Numerical results are presented in Section VI and VII concludes the paper and outlines some future directions.

II. Related Work

Providing quality of service for video streaming applications in wired and wireless networks has been a pivotal problem of many research communities. In particular, quality-of-service requirements in terms of rate and delay have been amongst of such research interests. Dealing with such challenges have led to several video adaptation schemes in video coding research community, as well as rate-based flow control schemes for inelastic flows in networking research community. Within the last decade, video adaptation techniques that are based on layered encoding schemes are mainly investigated in the category of scalable video coding [14], [15] and Fine Granularity Scalability (FGS) [13] techniques. Some rate-distortion (R-D) models of Fine-Granular Scalability (FGS) encoder (an accepted standard scalable video coding scheme for the video streaming profile in MPEG-4) is investigated in [16].

In the course of past years, a lot of recent studies have dealt with multimedia applications through modeling the traffic characteristics of such applications. Huang *et al.* [17], proposed a resource allocation solution for multiuser video streaming over cellular wireless networks. They developed a NUM framework with a resource pricing algorithm via previous well-established dual-based algorithms. The resource price is obtained in turn, is used to derive source content adaptation to each user, using video summarization techniques [18]. In [19], a content-aware distortion-fair networking framework with joint video source adaptation and network resource allocation is developed. A basic difference in this work is that an explicit utility function for sources is not considered. Instead, a content-aware time-varying utility function is chosen that is different per each frame as well as for each video content. Based on the idea of dropping less important frames when the network is congested, a distributed iterative algorithm is proposed to achieve min-max distortion fairness. The main superiority of this work is taking into account the special characteristics of video content such as dependency between frames.

Furthermore, several flow control schemes are proposed to tackle the problem of multimedia transmission over wired and wireless networks. In [20], a mechanism for dealing with loss in scalable video transmission over best-effort lossy network channels using forward-error correction (FEC) mechanisms is proposed. Also, a dynamic controller on the amount of FEC that maximizes the utility of scalable video is constructed. Media-friendly and TCP-friendly congestion control schemes for scalable video streams are investigated in [21]. It borrows well-established utility-based optimization flow control model as the underlying framework and adopts a two-fold time-scale approach, which can optimize the video quality in short term (multimedia-friendliness) and meet the TCP-friendliness in long term. Although network utility maximization for the scalable video coding is considered in [21], this work differs with ours in that we model the staircase utility function for SVC-encoded applications, which captures more efficiently the characteristics of such applications.

III. SYSTEM MODEL

We consider a network comprising a set of sources denoted by $\mathcal{S} = \{1, \ldots, S\}$ and a set of unidirectional links, denoted by $\mathcal{L} = \{1, \ldots, L\}$. Let x_s and c_l be the source rate for source s and capacity of link l, both in bps, respectively. Without loss of generality, we assume that the rate of source s is limited so as to certify

$$0 < m_s \leq x_s \leq M_s < \infty \tag{1}$$

where m_s and M_s denote the minimum and maximum rates for source s, respectively.

We associate with source s a *path*, i.e. a set of links $\mathcal{L}(s) \subseteq \mathcal{L}$, that determines the links that source s passes through. Similarly, we define $\mathcal{S}(l) \subseteq \mathcal{S}$, to be the set of sources traversing link l. For the sake of simplicity, we define the routing matrix as $\mathbf{R} = [R_{ls}]_{L \times S}$, where R_{ls} is defined as

$$R_{ls} = \begin{cases} 1 & \text{if source } s \text{ passes through link } l \\ 0 & \text{otherwise} \end{cases} \tag{2}$$

Thereafter, we will follow this notation, unless stated otherwise.

To simplify the analysis and not relying on any particular flow control and packet scheduling scheme, rate control is ideally accommodated by congestion in links.

We assume that source s, when submitting at rate x_s, attains a utility $U_s(x_s)$, which used to quantify its benefit, in terms of QoS, rate requirements, etc. Best Effort traffics, such as traditional file transfer, can be modeled using continuous and strictly concave utility functions, that make them tractable to be dealt with conventional optimization approaches [6]. On the other hand, Guaranteed Service streams, such as multimedia, demand for tight QoS requirements, in terms of rate, delay, jitter, etc. which cause their utility functions to be non-concave and even discontinuous. Such utility functions yield non-convex optimization problems that cannot be dealt with conventional optimization approaches.

As stated before, in this study we assume that each source s is in possession of a multimedia application which uses SVC-encoded video stream. This is equivalent to saying that each source s is capable of adapting the rate and quality of the video stream in accordance to the conditions charged by the network or to the user preferences. In particular, source s might choose a lower encoding scheme than the current one, so as to combat against network congestions, or might choose a higher one whenever additional bandwidth is available.

For multimedia applications supporting SVC standard, rate allocation is limited to distinct levels of utility, i.e. the utility function is increased only when a higher layer can be delivered due to increase in the available bandwidth. As stated before, the utility function of such applications can be ideally characterized using a non-concave and discontinuous utility function referred to as *staircase* utility function, which is shown in Figure 1 in solid line.

In order to deal with such utility functions, we use a smoothed approximation of them, the idea behind which is shown in Figure 1 in dashed line. In this figure, the curve in dashed line represents the smoothed approximation of the staircase function. In order to construct such a smoothed approximation, we divide its domain into non-overlapping intervals, so that a step transition occurs within the midpoint of each interval. The step transition i, i.e. the part of the curve in which utility function jumps from level i, (i.e. $U(x) = i$), to level $i+1$, (i.e. $U(x) = i+1$) is smoothed and approximated by a *sigmoid-like* function, whose point of inflection corresponds to $U(x) = \frac{i+1}{2}$.

Sigmoid-like functions have been well studied in the field of neural networks. The most commonly used form of sigmoid-like functions is the *logistic function* defined as

$$F(x, \alpha, \beta) = \frac{1}{1 + e^{-\alpha(x-\beta)}} \tag{3}$$

It is easy to show that β is the inflection point of F such that for $x < \beta$, F is convex, and for $x > \beta$ it is concave. Moreover, $\alpha > 0$ is a parameter that determines the sharpness of the curve. In order to efficiently model the behavior of a step transition, α must be chosen sufficiently large so as to effectively capture the sharp transition of an increase in the utility level.

Using the sigmoid-like function introduced above, we then represent the approximation shown in Figure 1 in dashed line. Recall the interval division of the domain introduced above. Then, for the step transition i, i.e. jump from $U(x) = i$ to $U(x) = i + 1$, we have,

$$\tilde{U}(x) = F(x, \alpha, ki) + i; \qquad x \in \left[ki - \frac{k}{2}, ki + \frac{k}{2}\right] \tag{4}$$

where $\tilde{U}(.)$ denotes the approximation to the original utility function and $[ki - \frac{k}{2}, ki + \frac{k}{2}]$ is the interval in which transition i occurs. It's worth mentioning that k is the required rate increase to advance the utility U by 1. Hence, $1/k$ can be thought of as the slope of the straight line passing through the midpoint of step transitions.

Combining all of the intervals and assuming that the least utility level is 1, we get

$$\tilde{U} = \begin{cases} F(x, \alpha, k) + 1 & x \in [k - \frac{k}{2}, k + \frac{k}{2}] \\ . & \\ . & \\ F(x, \alpha, ki) + i & x \in [ki - \frac{k}{2}, ki + \frac{k}{2}] \\ . & \\ . & \\ F(x, \alpha, kN) + N & x \in [kN - \frac{k}{2}, kN + \frac{k}{2}] \end{cases} \tag{5}$$

where it is assumed that the domain is divided into N equal intervals, corresponding to N encoding layers.

In statistics, a sigmoid-like function, which is in possession of a single point of inflection, is usually referred to as *unimodal function*. Our approximated staircase utility function is comprised of several sigmoid-like functions, and thereby has several points of inflection. Thus, it is a *multimodal function* as opposed to the unimodal case. In this respect, we refer to this approximation as the *multimodal sigmoid approximation*. The multimodal sigmoid approximation presented above is discontinuous in general; however, if α is chosen sufficiently large, discontinuity gap between the contiguous steps vanishes and thereby makes it continuous, as Figure 1 depicts.

Sources in the network may demand for different QoS requirements; hence, it makes sense that each source s, would advance its utility according to its own k_s factor, which may differ from the others. Therefore, the (approximated) utility function of source s is

$$\tilde{U}_s(x_s) = \tilde{U}(x_s, \alpha, k_s) \tag{6}$$

where $\tilde{U}$ is defined by (5) and k_s and α were omitted from the notation.

IV. Problem Formulation

We model the rate control for SVC-encoded applications as the solution to an optimization problem. As stated in the previous section, to deal with non-differentiable behavior of staircase utility functions of such multimedia applications, we introduced the multimodal sigmoid approximation. However, non-concavity of the aforementioned approximated utility results in a non-convex optimization problem.

In order to deal with non-concave approximated utility function, we use the *utility-proportional* optimization approach introduced by Wang *et al.* [12]. They proposed a novel application-oriented utility optimization formulation to address fairness and as a remedy to *convexify* specific forms of non-concave utility functions. By adopting utility-proportional formulation, the underlying optimization problem can be formulated as [12]:

$$\max_{\mathbf{x}\in\mathcal{X}} \sum_{s=1}^{S} \int_{m_s}^{x_s} \frac{w_s dz}{U_s(z)} \tag{7}$$

subject to:

$$\sum_s R_{ls} x_s \leq c_l; \quad l \in \mathcal{L} \tag{8}$$

where $\mathcal{X} = \mathcal{X}_1 \times \mathcal{X}_2 \times \ldots \times \mathcal{X}_S$ denotes the Cartesian product of all rate domains $\mathcal{X}_s = [m_s, M_s]$, and w_s is a positive weight factor assigned to source s to address its priority in rate allocation.

The constraint (8) states that the sum of rates passing through link l cannot exceed its capacity. Focusing on the SVC-encoded applications and recalling multimodal sigmoid approximation in Section III, we rewrite the optimization problem as

$$\max_{\mathbf{x}\in\mathcal{X}} \sum_{s=1}^{S} \int_{m_s}^{x_s} \frac{w_s dz}{\tilde{U}_s(z)} \tag{9}$$

subject to:

$$\sum_s R_{ls} x_s \leq c_l; \quad l \in \mathcal{L} \tag{10}$$

where $\tilde{U}_s(.)$ is defined by (5) and (6). Using elementary calculus, for the objective of (9) we get

$$\sum_{s=1}^{S} \int_{m_s}^{x_s} \frac{dz}{\tilde{U}_s(z)} = \sum_{s=1}^{S} \sum_{i_s=1}^{N} \int_{z\in\mathcal{D}_{i_s}} \frac{w_s}{F(z,\alpha,k_s i_s)+i_s} dz \tag{11}$$

where

$$\mathcal{D}_{i_s} = \{z | k_s i_s - \frac{k_s}{2} \leq z \leq \min(x_s, k_s i_s + \frac{k_s}{2})\} \tag{12}$$

Moreover, the optimal rate for source s and its corresponding layer are denoted by x_s^* and i_s^*, respectively.

We defer solving the optimization problem until the next section.

V. Optimal Rate Control

A. Convexity

We first investigate the convexity of problem (9)-(10). The result is stated in the following theorem.

Theorem 1: The problem (9)-(10) is strictly convex and admits a unique maximizer.

Proof: For proof see Appendix I. ■

B. Optimal Solution

In this subsection, we derive the optimality conditions. We start by writing the Lagrangian of (9). Using the standard optimization methods [22], the Lagrangian of the problem (9) can be written as

$$L(\mathbf{x},\lambda) = \sum_s \sum_{i_s=1}^{N} \int_{z\in\mathcal{D}_{s_i}} \frac{w_s}{F(z,\alpha,k_s i_s)+i_s} dz - \sum_l \lambda_l \left(\sum_s R_{ls} x_s - c_l\right) \tag{13}$$

where λ_l is the positive Lagrange multiplier associated with the corresponding constraint of link l and $\lambda = (\lambda_l, l \in \mathcal{L})$ is the vector of Lagrange multipliers.

According to Theorem 1, the problem (9)-(10) is strictly convex and admits a unique maximizer, denoted by $\mathbf{x}^* = (x_s^*, s \in \mathcal{S})$. Based on KKT Theorem [22], at the optimal point, the following conditions must be satisfied

$$\nabla_{\mathbf{x}} L(\mathbf{x},\lambda)|_{(\mathbf{x}^*,\lambda^*)} = \mathbf{0} \tag{14}$$

$$\sum_s R_{ls} x_s^* \leq c_l; \qquad l \in \mathcal{L} \tag{15}$$

$$\lambda_l^* \geq 0; \qquad l \in \mathcal{L} \tag{16}$$

$$\lambda_l^* \left(\sum_s R_{ls} x_s^* - c_l\right) = 0; \qquad l \in \mathcal{L} \tag{17}$$

where $\mathbf{0}$ is a vector, all of whose elements are zero.

Substituting (13) into (14) yields

$$\begin{aligned} \frac{\partial L}{\partial x_s} &= \frac{w_s}{F(x_s,\alpha,k_s i_s)+i_s} - \sum_l R_{ls}\lambda_l \\ &= \frac{w_s}{\frac{1}{1+e^{-\alpha(x_s-k_s i_s)}}+i_s} - \sum_l R_{ls}\lambda_l = 0 \end{aligned} \tag{18}$$

Doing some algebraic manipulations, for the optimal source rate we get

$$x_s^* = \left[k_s i_s^* - \frac{1}{\alpha}\log\left(\frac{\lambda^s}{w_s - \lambda^s i_s^*} - 1\right)\right]_{\mathcal{X}_s} \tag{19}$$

where

$$\lambda^s = \sum_l R_{ls}\lambda_l \tag{20}$$

and $[.]_{\mathcal{X}_s}$ is the projection operator on the $\mathcal{X}_s$.

Since the argument of the logarithm should be non-negative, i^*s must satisfy the following condition

$$\frac{\lambda^s}{w_s - \lambda^s i_s^*} > 1 \tag{21}$$

which results in

$$i_s^* = \lceil \frac{w_s - \lambda^s}{\lambda^s} \rceil \tag{22}$$

Although problem (9) can be separated among sources, its constraints will remain coupled across the links over the network. The coupled nature of such constrained problems, necessitates usage of centralized methods like *interior point method* which poses great computational overhead to the system [22] [23].

In order to come up with a distributed solution, we solve the problem through its dual problem. The dual problem is defined as

$$\min_{\lambda \geq 0} D(\lambda) \tag{23}$$

where $D(\lambda)$ is the dual function and is defined as the maximum of the Lagrangian over $\mathbf{x}$, i.e.

$$D(\lambda) = \max_{\mathbf{x} \in \mathcal{X}} L(\mathbf{x}, \lambda) \tag{24}$$

Problem (24) is an unconstrained optimization problem and is already solved by $\mathbf{x}^*$. Therefore, the dual function is given by

$$D(\lambda) = L(\mathbf{x}^*, \lambda) \tag{25}$$

where x_s^* is given by (19).

C. Solving Dual Problem

Now we are ready to solve the dual problem (23). According to the duality theory, dual problem is always convex regardless of Non-convexity of the primal problem. However, strict convexity of the primal problem (9) guarantees strong duality, i.e. assures that solving the dual results in the optimal solution of the primal [22] [23].

In order to obtain a distributed solution with low computational complexity, we will solve the dual problem using *gradient projection method* [23]. For an unconstrained minimization problem, the gradient projection method iteratively steps toward the opposite direction of the gradient of the objective of the problem. Therefore, for the dual problem (23), we get

$$\lambda^{(t+1)} = [\lambda^{(t)} - \gamma \nabla D(\lambda^{(t)})]^+ \tag{26}$$

or equivalently,

$$\lambda_l^{(t+1)} = \left[\lambda_l^{(t)} - \gamma \frac{\partial D(\lambda^{(t)})}{\partial \lambda_l}\right]^+ \tag{27}$$

where $\lambda^{(t)} = (\lambda_l^{(t)}, l \in \mathcal{L})$ is the value of λ at tth iteration step, γ is a constant step size and $[z]^+ = \max(z, 0)$.

According to Theorem 1, the objective of problem (9) is strictly concave, hence (24) is continuously differentiable ([24], pp. 669) with derivatives given by

$$\frac{\partial D(\lambda^{(t)})}{\partial \lambda_l} = c_l - \sum_s R_{ls} x_s^{(t)} \tag{28}$$

Substituting (28) into (27) yields

$$\lambda_l^{(t+1)} = \left[\lambda_l^{(t)} - \gamma \left(c_l - \sum_s R_{ls} x_s^{(t)}\right)\right]^+ \tag{29}$$

The two update equations (19) and (29) obtained above, form an iterative solution to the (9). In the next subsection, we will discuss the algorithmic aspects of such an iterative solution.

D. Optimal Rate Control Algorithm

In this subsection, we propose a distributed algorithm based on the iterative solution obtained above.

Optimal source rate equations, i.e. (19) and (22), and shadow price[1] update (29), derived in the previous section, can be used in conjunction with each other to form an iterative algorithm as the solution to the optimization problem (9)-(10).

For each time slot t (or iteration step t), the following key steps exist in the algorithm:

1) Each link l calculates its corresponding Lagrange multiplier (shadow price) for the next time slot, i.e. $\lambda_l^{(t+1)}$, based on its previous shadow price and its aggregate traffic in the current time slot.
2) Each source s calculates its rate based on the aggregate shadow price in its path.
3) Each source s transmits the packets based on the allocated rate.

The rate control algorithm can be described as follows. For each link l, shadow price λ_l is updated according to (29) and the new shadow price result is communicated to sources traversing this link. Each source s receives from the network the shadow prices for links on its path and calculates λ^s using (29), chooses a new source rate using (19) and (22), and communicates this new rate to all the links in its path. The procedures at the links and the video sources are repeated until the algorithm converges to the optimal video rates and optimal shadow prices.

The iterative rate control algorithm is listed as Algorithm 1.

[1]Lagrange multiplier is usually referred to as *shadow price* owing to the economic interpretation of its role to adjust the source rate [5].

Algorithm 1 Rate Control Algorithm for SVC Streams

Initialization

Initialize the following items:

1. Sets of sources and links including the routing matrix.
2. γ and c_l for $l \in \mathcal{L}$.
3. k_s, w_s for $s \in \mathcal{S}$.

Main Loop

Do until $\max_s |x_s^{(t+1)} - x_s^{(t)}| < \epsilon$

1. $\forall l \in \mathcal{L}$ Compute new link prices:
$$\lambda_l^{(t+1)} = [\lambda_l^{(t)} - \gamma \left(c_l - \sum_l R_{ls} x_s^{(t)} \right)]^+$$

2. $\forall s \in \mathcal{S}$ Compute new source rates as follows:
$$\lambda^{s(t)} = \sum_l R_{ls} \lambda_l^{(t)}$$
$$i_s^{(t+1)} = \lceil \frac{w_s - \lambda^{s(t)}}{\lambda^{s(t)}} \rceil$$
$$x_s^{(t+1)} = \left[k_s i_s^{(t+1)} - \frac{1}{\alpha} \log \left(\frac{\lambda^{s(t)}}{w_s - \lambda^{s(t)} i_s^{(t+1)}} - 1 \right) \right]_{\mathcal{X}_s}$$

where and $[.]_{\mathcal{X}_s}$ is the projection operator on the $\mathcal{X}_s$.

Output

Communicate source rates to the corresponding sources.

Algorithm 1. Rate Control Algorithm for SVC Streams

VI. Simulation Results

The proposed rate control algorithm is examined through extensive simulation experiments carried out using MATLAB. Numerous validation experiments have been performed for several network topology, however, for the sake of specific illustration, the two following results are only presented.

A. The Single Link Case

For the first scenario we consider a network with a single bottleneck link. There are five sources in the network, all passing through a shared link with capacity $c = 10$ Mbps. Different sources use different values of k_s as a result of different rate and quality requirements. Recall that k_s is the required rate increase for source s to advance the utility U_s by one. In this scenario we choose: $(k_1, \ldots, k_5) = (0.5, 0.7, 0.9, 1, 1.2)$. For the sake of illustration, utility functions corresponding to different values of k_s are displayed in Figure 2. The value of α for all sources is set to 4 and step size is chosen to be $\gamma = 0.02$. For simplicity, weight factors of all users are assumed to be equal. The rate allocation for this scenario is summarized in Table I.

Apart from the steady state rate allocation, one issue of our interest is the transient behavior of the algorithm. Such transient behaviors essentially implies the convergence properties of the algorithm and gives insight of how fast the algorithm converges towards the steady state regime. In order to achieve this goal, the evolution of source rates is depicted in Figure 3. Moreover, the evolution of the shadow price (λ_l) is depicted in Figure 4. Both figures reveal that the convergence is relatively fast and just as few as 20 iteration steps are needed to reach the steady state. It is worth mentioning that sources with larger k_s obtain greater rates. This fact is in contrast with the conventional network utility maximization (NUM), wherein sources with larger k_s would obtain smaller rates.

In fact, in conventional NUM formulations, smaller k_s results in a utility function whose value grows rapidly upon increase in the available bandwidth. Therefore, the optimization problem allocates available bandwidth so that utility functions attain larger rates in order to maximize the total utility as much as possible. However, the utility-proportional formulation used

TABLE I: Rate Allocation for Single Link Case

Source	k_s	x_s (Mbps)
1	0.5	1.28
2	0.7	1.73
3	0.9	2.12
4	1	2.34
5	1.2	2.74

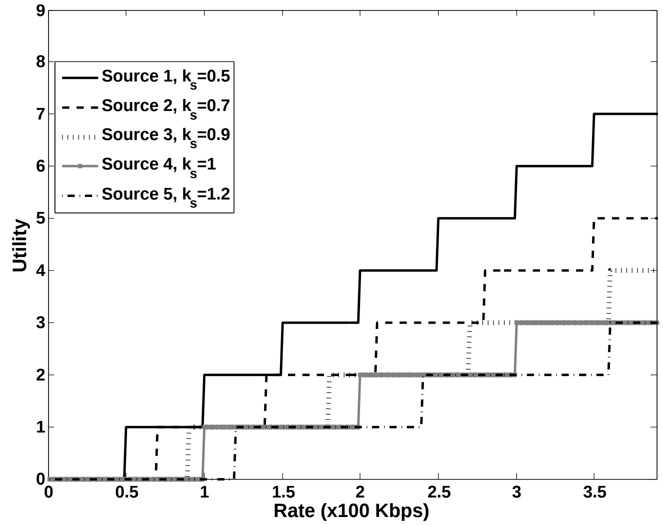

Fig. 2: Utility Functions for Different Values of k_s

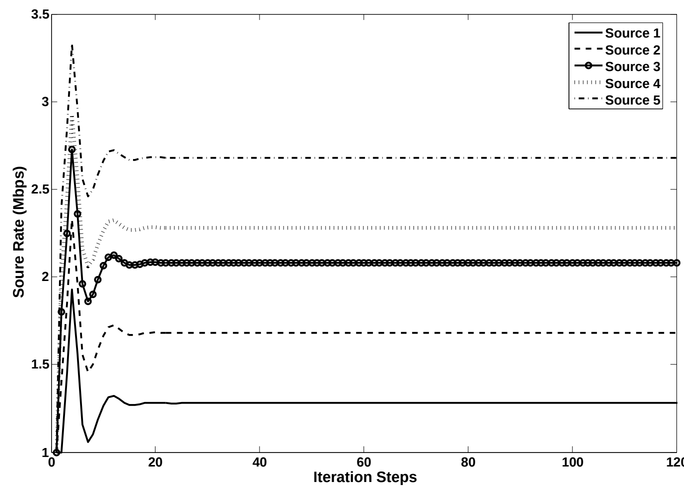

Fig. 3: Evolution of The Source Rates for The Single Link Topology.

here somehow regulates bandwidth sharing so that there are tolerable differences between steady state rates for different values of k_s.

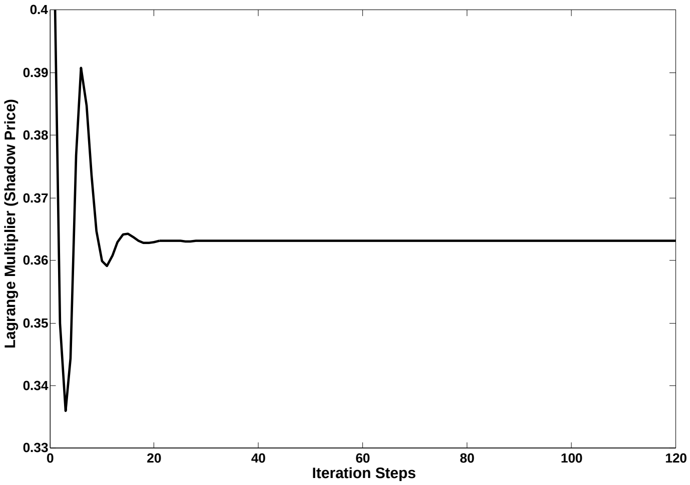

Fig. 4: Evolution of The Shadow Price for The Single Link Topology.

B. The Multiple Link Case

For the second scenario, we consider a network comprising of four sources as well as four links, whose topology along with link capacities are depicted in Figure 5. For this scenario we choose: $(k_1, \ldots, k_4) = (1, 1.2, 1, 0.9)$. α for all sources is set to 4 and step size is chosen to be $\gamma = 0.003$.

Rate allocation for this scenario is summarized in Table II. In order to compare the convergence behavior of this scenario, which consists of a rather complicated topology with respect to the latter case, the evolution of source rates is depicted in Figure 6. This figure shows that compared to the single link scenario, convergence is achieved at the expense of more iteration steps. While the achieved rate for some sources slightly differs after iteration 60, the steady state regime for others is achieved after iteration 110. The evolution of shadow prices is shown in Figure 7. As this figure presents, while links 1, 2, and 3 are assigned with a positive non-zero shadow price, that of link 4 is zero implying that this link is not a bottleneck and wouldn't be saturated.

VII. Conclusion

With the emergence of a plethora of applications demanding for multimedia applications, nowadays provisioning services

TABLE II: Rate Allocation for Multiple Link Case

Source	k_s	x_s (Mbps)
1	1	6.25
2	1.2	3.15
3	1	2.67
4	0.9	4.16

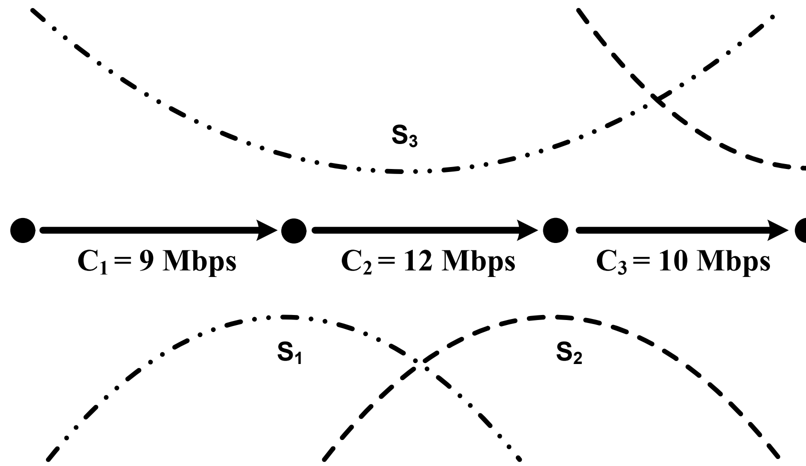

Fig. 5: Topology for The Second Scenario

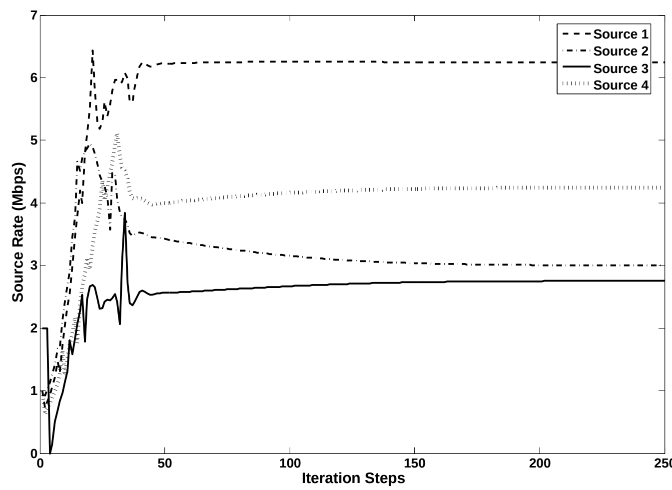

Fig. 6: Evolution of The Source Rates for The Multiple Link Topology.

with adaptive rate and quality proves quite inevitable. In this paper we addressed the problem of rate allocation for SVC-encoded multimedia applications. First, we elaborated to analytically model an approximated utility function for SVC compliant multimedia applications. Then, we proposed a convex optimization formulation for bandwidth sharing and rate control, which was solved indirectly through its dual using gradient projection method. This allowed us to devise a distributed algorithm that can be used to determine the optimal rate allocation in an iterative manner. The algorithm can be addressed in distributed scenarios, particularly in the context of ad-hoc networks, etc. Simulation results validated the effectiveness of the algorithm and verified its rapid convergence. A possible line for the continuation of this research would be going into more details of the SVC streams and related factors such as Qos/QoE. It proves quite challenging how these factors

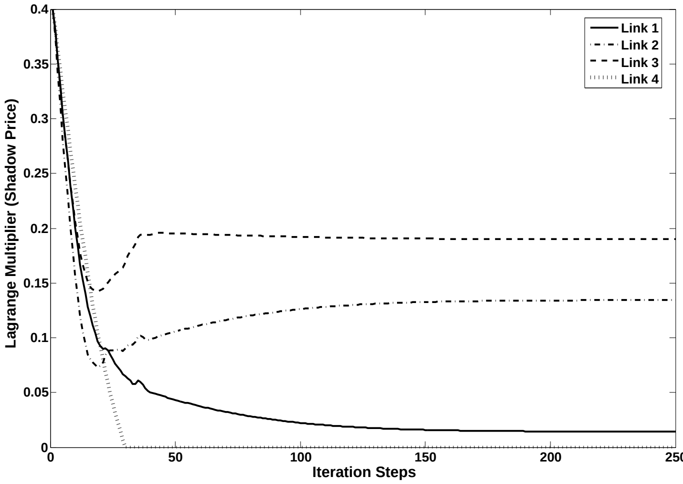

Fig. 7: Evolution of The Shadow Price for The Multiple Link Topology.

will affect the optimal rate control algorithm.

REFERENCES

[1] S. F. Chang and A. Vetro, "Video adaptation: concepts, technologies, and open issues," *In Proc. of the IEEE*, vol. 93, no. 1, pp. 148-158, 2005.

[2] *Advanced Video Coding for Generic Audiovisual Services,* ITU-T Rec. H.264 and ISO/IEC 14496-10 AVC, v3: 2005, Amendment 3: Scalable Video Coding.

[3] H. Schwarz, D. Marpe, and T. Wiegand, "Overview of the scalable video coding extension of H.264/AVC," *IEEE Trans. on Circuits and Systems for Video Tech.*, vol. 17, no. 9, pp. 1103-1120, Sep. 2007.

[4] *Advanced Video Coding for Generic Audiovisual Services,* ITU-T Recommendation H.264 and ISO/IEC 14496-10 AVC, v3: 2005.

[5] F. P. Kelly, A. Maulloo, and D. Tan, "Rate control for communication networks: shadow prices, proportional fairness and stability," *J. Oper. Res. Soc.*, vol. 49, pp. 237-252, Mar. 1998.

[6] S. H. Low, and D. E. Lapsley, "Optimal flow control, I: basic algorithm and convergence," *IEEE/ACM Trans. Netw.*, vol. 7, no. 6, pp. 861-874, Dec. 1999.

[7] M. Chiang, S. H. Low, A. R. Calderbank, and J. C. Doyle, "Layering as optimization decomposition: A mathematical theory of network architectures," *In Proc. of the IEEE*, vol. 95, no. 1, pp. 255-312, 2007.

[8] S. Shenker, "Fundamental design issues for the future Internet," *IEEE J. Sel. Areas Commun.*, vol. 13, no. 7, pp. 1176-1188, Sep. 1995.

[9] J. W. Lee, R. R. Mazumdar, and N. Shroff, "Non-convex optimization and rate control for multi-class services in the Internet," *IEEE/ACM Trans. Netw.*, vol. 13, no. 4, pp. 827-840, Aug. 2005.

[10] P. Hande, S. Zhang, and M. Chiang, "Distributed rate allocation for inelastic flows," *IEEE/ACM Trans. Netw.*, vol. 15, no. 6, pp. 1240-1253, Dec. 2007.

[11] C Luna, L Kondi, A Katsaggelos, "Maximizing user utility in video streaming applications," *IEEE Trans. on Circuits and Systems for Video Tech.*, vol. 13, no. 2, pp. 141-148, Feb. 2003.

[12] M. H. Wang, M. Palaniswami, and S. H. Low, "Application-Oriented Flow Control: Fundamentals, Algorithms and Fairness", *IEEE/ACM Trans. Netw.*, vol. 14, no. 6, pp. 1282-1291, Dec. 2006.

[13] W. Li, "Overview of fine granularity scalability in MPEG-4 video standard", *IEEE Trans. on Circuits and Systems for Video Tech.*, vol. 11, no. 3, pp. 301-317, Mar. 2001.

[14] J. Ohm, "Advances in Scalable Video Coding," *In Proc. of the IEEE*, vol. 93, no. 1, pp. 42-56, 2005.

[15] X. Zhu, T Schierl, T. Wiegand, and B. Girod, "Video Multicast over Wireless Mesh Networks with Scalable Video Coding (SVC)," *In Proc. VCIP*, San Jose, CA, 2008.

[16] M. Dai and D. Loguinov, "Analysis of rate-distortion functions and congestion control in scalable internet video streaming," *In Proc. ACM NOSSDAV*, pp. 60-69, 2003.

[17] J. Huang, Z. Li, M. Chiang, and A. K. Katsaggelos, "Joint source adaptation and resource pricing for multi-user wireless video streaming," *IEEE Trans. on Circuits and Systems for Video Tech.*, vol. 18, no. 5, pp. 582-595, May 2008.

[18] Z. Li, G. Schuster, A. Katsaggelos, and B. Gandhi, "Rate-distortion optimal video summary generation," *IEEE Trans. Image Processing*, vol. 14, no. 10, pp. 1550-1560, Oct. 2005.

[19] Y. Li, Z. Li, M. Chiang, and A. R. Calderbank, "Content Aware Distortion-Fair Video Streaming in Networks," *In Proc of IEEE GLOBECOM*, New Orleans, USA, 2008.

[20] S. R. Kang and D. Loguinov, "Modeling best-effort and FEC streaming of scalable video in lossy network channels," *IEEE/ACM Trans. Netw.*, vol. 15, no. 1, pp. 187-200, Feb. 2007.

[21] J. Yan, K. Katrinis, M. May, and B.Plattner, "Media-and TCP-friendly congestion control for scalable video streams," *IEEE Trans. on Multimedia*, vol. 8, no, 2, pp. 196-206, April 2006.

[22] S. Boyd, and L. Vandenberghe, *Convex Optimization*, Cambridge University Press, 2003.

[23] D. Bertsekas, *Nonlinear Programming*, Athena Scientific, 1999.

[24] D. Bertsekas, and J. Tsitsiklis, *Parallel and Distributed Computation: Numerical Methods*, Prentice-Hall, 1989.

APPENDIX I: PROOF OF THEOREM 1

To establish the convexity, we first obtain the Hessian of the objective function. Let us denote the objective of (9) by $\mathcal{U}(\mathbf{x})$. Focusing on the ith interval, the first derivative of the objective is given by

$$\frac{\partial \mathcal{U}}{\partial x_s} = \frac{w_s}{F(x_s, \alpha, k_s i_s) + i_s}; \quad x_s \in [k_s i_s - \frac{k_s}{2}, k_s i_s + \frac{k_s}{2}] \tag{30}$$

and for the second derivative term, we have

$$\frac{\partial^2 \mathcal{U}}{\partial x_s \partial x_r} = \begin{cases} -\frac{w_s F'(x_s,\alpha,k_s i_s)}{(F(x_s,\alpha,k_s i_s)+i_s)^2} & s = r \\ 0 & s \neq r \end{cases} \tag{31}$$

Now, substituting (3) into (31) yields

$$\begin{aligned} -\frac{w_s F'(x_s, \alpha, k_s i_s)}{(F(x_s, \alpha, k_s i_s) + i_s)^2} &= \frac{-w_s \frac{\alpha e^{-\alpha(x_s - k_s i_s)}}{(1+e^{-\alpha(x_s-k_s i_s)})^2}}{\left(\frac{1}{1+e^{-\alpha(x_s-k_s i_s)}} + i_s\right)^2} \\ &< 0 \end{aligned} \tag{32}$$

Thus, the Hessian of $\mathcal{U}(\mathbf{x})$ is a diagonal matrix, all of whose diagonal elements are negative, thereby making the Hessian negative-definite. Therefore, the objective function $\mathcal{U}$ is strictly concave.

Constraints (10) are comprised of affine functions, and thereby all of them are convex. We deduce that the optimization problem (9)-(10) is convex. Since the feasible set is compact, and the objective is strictly concave, the optimal solution exists and it is unique as a consequence of convexity.