This full text paper was peer reviewed at the direction of IEEE Communications Society subject matter experts for publication in the IEEE ICC 2009 proceedings

Prioritized Flow Optimization with Generalized Routing for Scalable Multirate Multicasting

Junni Zou[1], Hongkai Xiong[2,4], Li Song[2], Zhihai He[3], Tsuhan Chen[4]

[1]Key Laboratory of Specialty Fiber Optics and Optical Access Networks, Shanghai University, P.R.China
[2]Department of Electronic Engineering, Shanghai Jiao Tong University, P.R. China
[3]Department of Electrical and Computer Engineering, University of Missouri-Columbia, USA
[4] Dept. of Electrical and Computer Engineering, Carnegie Mellon University, USA

Abstract—**This paper addresses the performance optimization for scalable video coding and multicast over networks. Multi-path video streaming, network coding based routing, and network flow control are jointly optimized to maximize a network utility function defined over heterogeneous receivers. Importantly, contextual priors of scalable video layers are imposed on the flow routing optimization problem, seeking to guarantee the transmission cost for each layer in an incremental order and find jointly optimal multicast paths and associated rates. Through a primal decomposition and the primal-dual approach, a decentralized algorithm with two-level optimization update is developed to solve the target convex optimization problem. Numerical and simulation results validate the convergence and network performance of the proposed algorithm.**

Index Terms—**Multirate multicasting, utility maximization, scalable video coding, network coding, convex optimization.**

I. INTRODUCTION

Recently, multi-rate multicast has emerged as an important method for content distribution to adapt to different user requirements and time-varying network conditions of different receivers. From a source coding perspective, layered or hierarchical coding of source data, such as JVT/MPEG scalable video coding (SVC), allows video transmission and decoding at multiple bit rates with progressively improved video quality. It allows rate adaptation not only at the encoder/decoder, but also in intermediate network nodes while achieving highly efficient rate-distortion performance [1]. Optimizing the performance of scalable video coding and multicast over networks emerges as an important problem in content distribution and ubiquitous multimedia access. A SVC stream is normally represented in a compressed domain by a base layer and multiple enhancement layers with a flexible multi-dimension layer dependent graph structure, in terms of spatial resolution, temporal frame rate, and quality accuracy. For real-time SVC streaming with multirate multicasting, different SVC layers are transported in different IP multicast groups which are subscribed by heterogeneous receivers.

Rate control of video streaming has been studied extensively in the past [2-5]. Most of the existing schemes use predetermined distribution trees to improve the network throughput and overall video quality. Zhu et al [2] presented a rate adaption scheme for minimizing total distortion of multiple video streams with multi-path transmission. Schaar et al proposed a channel access scheme for scalable video streaming over wireless networks [3]. In this paper, we study layered utility maximization problems for communication networks where each receiver can have multiple alternative paths through the coded network (network coding) to receive subscribed SVC layers.

The first optimization model for the multirate multicast problem was studied by Kar et al. [4]. For maximizing the aggregate source utility, an optimization approach to flow control is presented in [5]. With the development of overlay networks, multiple routers between each source-destination pair are of interest. These flow control methods paid less attention to network information flow and source correlation.

Network coding is proven sufficient to achieve multicast capacity [6]. Chen et al in [7] developed two adaptive rate control algorithms by differentiating the networks with given coding subgraphs and without given coding subgraphs. An investigation of network utility maximization for network coding based multicasting was proposed in [8]. The authors in [9] proposed a LION algorithm to address the layered multicasting problem by adding network coding and multi-path constraints into the objective function. However, these methods do not involve either layered multirate multicasting or priority costs of layered multicasting groups.

In this paper, contextual priors of scalable video layers are imposed on the flow routing optimization problem, seeking to guarantee the transmission cost for each layer in an incremental order and find jointly optimal multicast paths and associated rates. Moreover, the proposed network coding can enhance network transmission performance and video streaming quality. Through a primal decomposition and the primal-dual approach, a decentralized algorithm with two-level optimization update is developed to solve the target convex optimization problem.

The work was partially supported by NSFC grants No. 60632040, No. 60736043, No. 60802019, National High Technology Research and Development Program of China (No. 2006AA01Z322) and funds from Shanghai Science and Technology Commission (No. 08220510900).

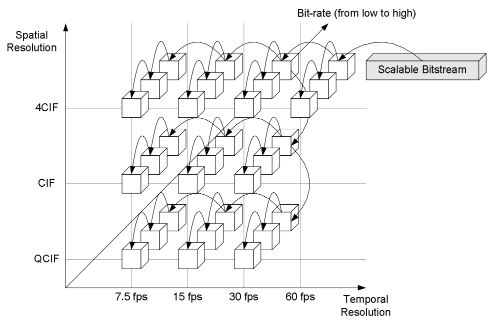

Fig. 1. Typical structure of scalable video bitstream with multiple-dimensions.

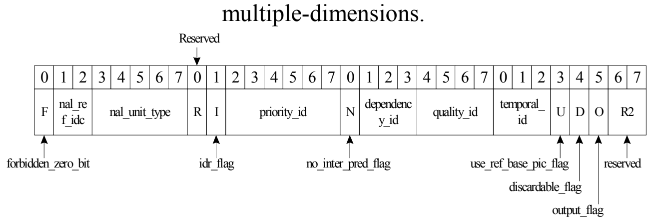

Fig. 2. 4 bytes SVC NAL unit header structure.

The rest of the paper is organized as follows. Sec. II presents network model and formulates the problem of resource allocation and performance optimization for scalable video coding and multicast over networks. Sec. III proposes a decentralized algorithm and proves its stability and convergence. Experimental results are presented in Sec. IV.

II. Problem Statement

A. Motivation

Initially, spatial and temporal layered coding techniques provide discrete achievable rate regions with different coarse resolutions and frame rates. Hereafter, fine granularity scalable (FGS) in quality scalability may preserve a smooth picture quality degradation along with the bandwidth variation. SVC with a full spatio-temporal and quality scalability could provide achievable rate set in a multi-dimension layer dependent graph structure, as shown in Fig. 1.

The coded bits in different layers are organized into NAL (Network Abstraction Layer) units preceded by 4 bytes header shown in Fig. 2. A series of NAL units encapsulating into RTP packets form a decodable entity or layer which is indicated by the syntax elements “dependency_id”, “temporal_id”, and “quality_id”. In practice, a media gateway or intermediate nodes in the network can tailor (extract) the in-coming RTP packet stream by a combined adaptation path along multiple scalability dimensions, to match the requirements of the outgoing RTP packet stream during the lifetime of a session. The “priority_id” syntax element has been assigned in NAL units to indicate a suggested adaptation path from the encoder.

A practical SVC-based multirate multicasting scenario is shown in Fig. 3, where each class is associated with one source-receivers cluster under given QoS requirements on both the bandwidth constraint and some other constraints such as

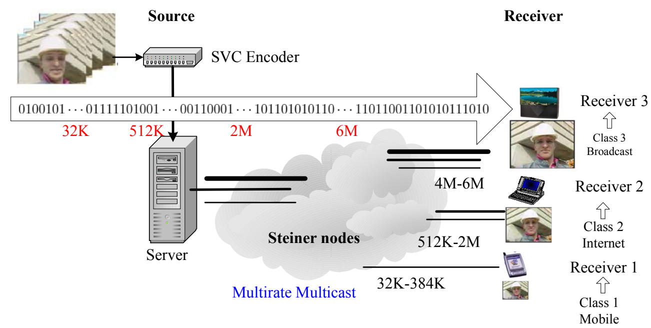

Fig. 3. The practical application scenario corresponding to SVC streaming network model.

delay or packet loss probabilities. The generated alternative paths for each class, must now be constrained to those that satisfy these other QoS requirements. Because one layer may need to coordinate multi-class receivers, network coding-based routing will alleviate the congestion at overlapping paths to different classes of receivers.

Once layer dependency and priority consideration in constructing multicasting path are not accounted, the higher layers may overwhelm the lower layers by low path costs and prices. In general, the packets of higher layers whose all packets of dependent layers are not available till playout deadline will have to be discarded, even if the bandwidth allocation for higher layers is optimally guaranteed to a maximized utility. This unexpected result obviously deviates from the original optimization objective.

B. Network Model

Consider a video distribution network, modelled as a directed graph $G(V,E)$, where V is the set of nodes and E is the set of weighted directed edges between nodes. The set V can be further divided into three disjoint subsets S, N and R, which represent source nodes, relay nodes and receiver nodes respectively. Let each edge $e \in E$ has a finite capacity of C_e.

Suppose the SVC stream is encoded into a set M of layers $\{l_1, l_2, \cdots, l_M\}$ at source node. Each layer m is distributed over a multicast group at transmission rate B_m. Assume there exists multiple alternative paths $P(r)$ from the source to receiver r. Let x_m^r denote the rate at which receiver r receives the data of layer m. Also let $x_{m,j}^r$ represent receiver r's flow rate on path j in layer m, and $f_{m,e}$ denote the bandwidth consumed on edge e in layer m. For each receiver r, we use a matrix $\mathbf{Z}^r$ to reflect the relationship between its paths and related edges. The (j,e) entry of $\mathbf{Z}^r$ is defined as:

$$z_{j,e}^r = \begin{cases} 1, & \text{if edge } e \text{ is included in path } j; \\ 0, & \text{otherwise.} \end{cases}$$

Generally, receiver r has multiple alternative paths to join the multicast group m, but not all these paths are optimal ones. Here we introduce a generic path cost function $\rho(\cdot)$ that is frequently used in packet routing applications [10]:

$$\rho(x_j) = \frac{x_j}{F_j - x_j} + d_j \cdot x_j$$

where x_j and F_j are flow rate and transmission capacity of the path j, d_j is the processing and propagation delay over path j.

To accurately measure the satisfaction perceived by a receiver from the bandwidth it obtained, each layer m is characterized by a utility function $U_m(\cdot)$ that is continuously differentiable, increasing and strict concave for the receiving rate.

C. Optimization Problem

In this study, we aim to find optimal routing paths with associated rates for which the total utility over all receivers with scalable multirate multicasting is maximized, meanwhile, guaranteeing the transmission cost for each layer in an incremental order.

This is a joint rate allocation and layered path scheduling issue. Mathematically, it can be formulated as follows:

$$\textbf{P1: } \text{maximize} \quad \sum_{r\in R}\sum_{m\in M} U_m\Big(\sum_{j\in P(r)} x^r_{m,j}\Big) \qquad (1)$$

subject to

1) $\sum_{j\in P(r)} z^r_{j,e} \cdot x^r_{m,j} \le f_{m,e},\ \forall e \in E,\ \forall m \in M,\ \forall r \in R;$

2) $\sum_{m\in M} f_{m,e} \le C_e,\ \forall e \in E;$

3) $\sum_{j\in P(r)} \rho(x^r_{m,j}) \le \sum_{j\in P(r)} \rho(x^r_{m+1,j}),\ \forall m,\ r;$

4) $b_m \le \sum_{j\in P(r)} x^r_{m,j} \le B_m,\ \forall m \in M;\ \forall r \in R;$

5) $x^r_{m,j} \ge 0,\ \forall j \in P(r);\ \forall m \in M,\ \forall r \in R.$

Constraints 1) reflect the relationship between information flow rate and physical flow rate on each edge where network coding is applied to information flows of the same video layer. Constraints 2) specify that the aggregate physical flow rates of different layers over each edge do not exceed the edge capacity. Constraints 3) ensure that the total path cost in each layer for each receiver is no more than the corresponding path cost of any higher layer. Here the path cost function $\rho(x^r_{m,j})$ is defined as:

$$\rho(x^r_{m,j}) = \frac{x^r_{m,j}}{B_m - x^r_{m,j}} + d^r_j \cdot x^r_{m,j} \qquad (2)$$

where d^r_j is the processing and propagation delay of receiver r over path j. Constraints 4) give the upper bound and lower bound of the receiving rate required for each layer.

Define

$\mathbf{x}^r = [x^r_{1,1}, \cdots, x^r_{1,P(r)}, x^r_{2,1}, \cdots, x^r_{2,P(r)}, \cdots, x^r_{M,1}, \cdots, x^r_{M,P(r)}]$, and $\mathbf{X} = [\mathbf{x}^1, \cdots, \mathbf{x}^R]^T$. Also let

$$\mathbb{C}^r = \Big\{\mathbf{x}^r \;\Big|\; x^r_{m,j} \ge 0 \text{ for all } m, j \text{ and } b_m \le \sum_{j\in P(r)} x^r_{m,j} \le B_m\Big\},$$

and $\mathbb{C}$ denote the Cartesian product of $\mathbb{C}^r (r \in R)$. Problem **P1** can be re-written as:

$$\textbf{P2}: \underset{\mathbf{X}\in\mathbb{C}}{\text{maximize}} \quad \sum_{r\in R}\sum_{m\in M} U_m\Big(\sum_{j\in P(r)} x^r_{m,j}\Big) \qquad (3)$$

subject to

1) $\sum_{j\in P(r)} z^r_{j,e} \cdot x^r_{m,j} \le f_{m,e},\ \forall e \in E,\ \forall m \in M,\ \forall r \in R;$

2) $\sum_{m\in M} f_{m,e} \le C_e,\ \forall e \in E;$

3) $\sum_{j\in P(r)} \rho(x^r_{m,j}) \le \sum_{j\in P(r)} \rho(x^r_{m+1,j}),\ \forall m,\ r;$

Clearly, Problem **P2** is feasible and there exists a unique optimal solution of **X** since the objective function is strictly convex and the constraint set is also convex. There are lots of polynomial-time centralized algorithms to solve **P2**, whereas, their requirements of coordination among all nodes are impossible in practice. In the subsequent section, we will provide a distributed and practical solution based on decomposition and duality theories.

III. Distributed Algorithm

A. Primal Decomposition

In general, we can use primal decomposition, a dual decomposition, or in combination to decompose an original large problem into a set of distributed sub-problems. The primal decomposition is suitable for problems with coupling variables, while the dual decomposition is a good choice for problems with coupling constraints [13]. If variables $f_{m,e}$ are fixed, Problem **P2** can be decoupled with respect to variables $x^r_{m,j}$. Based on this assumption, we use the primal decomposition approach and propose a two-level optimization procedure:

$$\textbf{P2a: } \underset{\mathbf{X}\in\mathbb{C}}{\text{maximize}} \quad \sum_{r\in R}\sum_{m\in M} U_m\Big(\sum_{j\in P(r)} x^r_{m,j}\Big) \qquad (4)$$

subject to

1) $\sum_{j\in P(r)} z^r_{j,e} \cdot x^r_{m,j} \le f_{m,e},\ \forall e \in E,\ \forall m \in M,\ \forall r \in R;$

2) $\sum_{j\in P(r)} \rho(x^r_{m,j}) \le \sum_{j\in P(r)} \rho(x^r_{m+1,j}),\ \forall m,\ r;$

$$\textbf{P2b: } \underset{\mathbf{f}\ge 0}{\text{maximize}} \quad \hat{U}_m(\mathbf{f}) \qquad (5)$$

subject to: $\sum_{m\in M} f_{m,e} \le C_e,\ \forall e \in E.$

It can be seen that Problem **P2a** performs a low-level optimization. It can be separated into a set of sub-problems, one for each combination of r, m and j, under the condition that **f** is fixed. Problem **P2b** performs a high-level optimization that is responsible for updating the coupling variable **f**. The objective value of the lower level optimization is locally optimal, which approximates to the global optimality using the result of the high-level optimization.

B. Two-level Optimization Update

To solve the low-level optimization problem, we define the Lagrangian dual of Problem **P2a** as:

$$\begin{aligned}\mathbf{L}(\mathbf{X},\mathbf{p},\mathbf{q}) = & \sum_{r\in R}\sum_{m\in M} U_m\Big(\sum_{j\in P(r)} x^r_{m,j}\Big) \\ & -\sum_{r\in R}\sum_{m\in M}\sum_{e\in E} p^r_{m,e}\Big(\sum_{j\in P(r)} z^r_{j,e}\cdot x^r_{m,j} - f_{m,e}\Big) \\ & -\sum_{r\in R}\sum_{m=1}^{M-1} q^r_m\Big[\sum_{j\in P(r)}\rho(x^r_{m,j}) - \sum_{j\in P(r)}\rho(x^r_{m+1,j})\Big]\end{aligned} \quad (6)$$

where $p^r_{m,e}$ and q^r_m are Lagrange multipliers. **P2a** is also a strictly convex problem that can be equivalently solved by solving its dual problem if the following Karush-Kuhn-Tucker (KTT) conditions are satisfied:

$$\begin{aligned}&\frac{\partial \mathbf{L}(\mathbf{X},\hat{\mathbf{p}},\hat{\mathbf{q}})}{\partial x^r_{m,j}}\bigg|_{x^r_{m,j}=\hat{x}^r_{m,j}} = 0,\ \forall j\in P(r),\ \forall m\in M,\ \forall r\in R.\\ &\hat{p}^r_{m,e}\Big[\sum_{j\in P(r)} z^r_{j,e}\cdot\hat{x}^r_{m,j} - f_{m,e}\Big] = 0,\ \forall e\in E,\ \forall m\in M,\ \forall r\in R;\\ &\hat{q}^r_m\Big[\sum_{j\in P(r)}\rho(\hat{x}^r_{m,j}) - \sum_{j\in P(r)}\rho(\hat{x}^r_{m+1,j})\Big] = 0,\ \forall m,\ r;\\ &\sum_{j\in P(r)} z^r_{j,e}\cdot\hat{x}^r_{m,j} - f_{m,e}\le 0,\ \forall e\in E,\ \forall m\in M,\ \forall r\in R;\\ &\sum_{j\in P(r)}\rho(\hat{x}^r_{m,j}) - \sum_{j\in P(r)}\rho(\hat{x}^r_{m+1,j})\le 0,\ \forall m,\ r;\\ &\hat{p}^r_{m,e}\ge 0,\ \forall e\in E,\ \forall m\in M,\ \forall r\in R;\\ &\hat{q}^r_m\ge 0,\ \forall m\in M,\ \forall r\in R;\end{aligned} \quad (7)$$

where $\hat{\mathbf{X}}$ and $(\hat{\mathbf{p}},\hat{\mathbf{q}})$ denote primal and dual optimal points respectively. In this paper, we propose the following primal-dual algorithm [11] to solve the low-level optimization problem. Unlike a dual method where the primal and the dual variables are updated alternatively, a primal-dual approach updates the primal and the dual variables simultaneously and moves together towards the optimal points asymptotically.

$$\begin{aligned}x^r_{m,j}(t+1) &= \Big[x^r_{m,j}(t) + \dot{x}^r_{m,j}\Big]^+\\ p^r_{m,e}(t+1) &= \Big[p^r_{m,e}(t) + \dot{p}^r_{m,e}\Big]^+\\ q^r_m(t+1) &= \Big[q^r_m(t) + \dot{q}^r_m\Big]^+\end{aligned} \quad (8)$$

where t is the iteration index, $\alpha(t)$, $\beta(t)$, and $\gamma(t)$ are positive step sizes, and $[\cdot]^+$ denotes the projection onto the set of non-negative real numbers. The partial derivatives of $\mathbf{X},\mathbf{p},$ and $\mathbf{q}$ are given by

$$\dot{x}^r_{m,j} = \begin{cases}\alpha(x^r_{m,j})\Big[U'_m(x^r_{m,j}) - \sum_{e\in E} z^r_{j,e}\cdot p^r_{m,e} - (q^r_m - q^r_{m-1})\cdot\rho'(x^r_{m,j})\Big], & \text{if } m=[2,M-1]\ ;\\ \alpha(x^r_{m,j})\Big[U'_m(x^r_{m,j}) - \sum_{e\in E} z^r_{j,e}\cdot p^r_{m,e} - q^r_m\cdot\rho'(x^r_{m,j})\Big], & \text{if } m=1,M\ .\end{cases}$$

$$\dot{p}^r_{m,e} = \beta(p^r_{m,e})\Big[\sum_{j\in P(r)} z^r_{j,e}\cdot x^r_{m,j} - f_{m,e}\Big]$$

$$\dot{q}^r_m = \begin{cases}\gamma(q^r_m)\Big[\sum_{j\in P(r)}\rho(x^r_{m,j}) - \sum_{j\in P(r)}\rho(x^r_{m+1,j})\Big], & \text{if } m<M\ ;\\ \gamma(q^r_m)\sum_{j\in P(r)}\rho(x^r_{m,j}), & \text{if } m=M\ .\end{cases} \quad (9)$$

During the updating steps in (8)-(9), we can consider $p^r_{m,e}$ as the congestion price at edge e for receiver r's bandwidth requirement in layer m. If the total demand $\sum_{j\in P(r)} z^r_{j,e}\cdot x^r_{m,j}$ exceeds the supply $f_{m,e}$, the price $p^r_{m,e}$ will rise, and decrease otherwise. q^r_m is the transmission cost for receiver r in layer m. If the cost in layer m exceeds that in layer $m+1$, raise q^r_m, otherwise reduce it. $x^r_{m,j}$ is the rate at which receiver r receivers data through path j in layer m. It will adjust with the aggregate congestion price $\sum_{e\in E} z^r_{j,e}\cdot p^r_{m,e}$ and cost q^r_m. All the updating steps are distributed and can be implemented at individual edges and receivers using only local information.

Next, we discuss how to adjust $\mathbf{f}$ to solve the high-level optimization Problem **P2b**. Consider $\hat{U}_m(\mathbf{f})$ may or may not be differentiable, we uniformly use the subgradient method to generate a sequence of feasible points of $\mathbf{f}$.

Suppose that $\hat{p}^r_{m,e}(f_{m,e})$ is the optimal Lagrange multiplier corresponding to the constraint $\sum_{j\in P(r)} z^r_{j,e}\cdot x^r_{m,j}\le f_{m,e}$ in Problem **P2a.** Let $\mathbf{f}_e = [f_{1,e},\cdots,f_{M,e}]$, and $\mathbf{f} = [f_1,\cdots,f_E]^T$. Also let

$$\mathbb{F}_e = \Big\{\mathbf{f}_e\ \Big|\ f_{m,e}\ge 0 \text{ for all } m \text{ and } \sum_{m\in M} f_{m,e}\le C_e\Big\},\ e\in E,$$

and $\mathbb{F}$ denote the Cartesian product of $\mathbb{F}_e (e\in E)$, then Problem **P2b** now can be solved with the following subgradient method:

$$f_{m,e}(t'+1) = [f_{m,e}(t') + \mu(t')\cdot\hat{p}_{m,e}(f_{m,e}(t'))]^{\mathbb{F}} \quad (10)$$

where $[\cdot]^{\mathbb{F}}$ denotes the projection onto the set $\mathbb{F}$ and $\hat{p}_{m,e}(f_{m,e}) \triangleq \sum_{r\in R}\hat{p}^r_{m,e}(f_{m,e})$.

$\hat{p}_{m,e}$ is the aggregate congestion price at edge e in layer m. If $\hat{p}_{m,e}$ has increased because of $\hat{p}^r_{m,e}$, which implies that the assigned capacity $f_{m,e}$ for edge e in layer m cannot meet the actual requirement of all receivers, then $f_{m,e}$ will be increased in the next step, else it will be decreased. The update of $f_{m,e}$ can be performed individually by each edge only with the knowledge of the sum of the congestion price $\hat{p}^r_{m,e}$, while the update of $\hat{p}^r_{m,e}$ simply uses the local information as mentioned above.

IV. RESULTS AND DISCUSSION

In this section, both numerical and simulation results are presented to show the performance of the proposed distributed

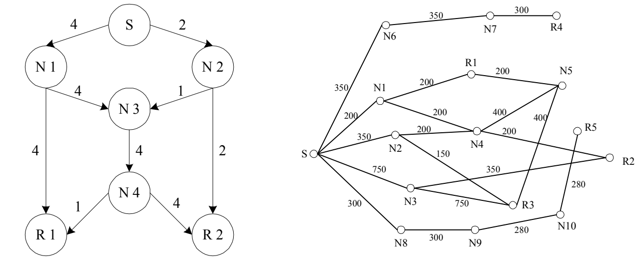

Fig. 4. Network topology associated with edge capacity, where (a) is a butterfly topology for numerical simulation, (b) is a general ISP access network for SVC streaming based simulation.

algorithm. We solved the two-level optimization model numerically for a classical butterfly topology and performed SVC streaming based simulations for a general ISP access network.

As seen in Fig.4 (a), S, N_i and R_i represent the source, the steiner nodes and the receivers respectively. The source rates are composed of three layers, with the base layer at rate 3 (data units/s), the first enhancement layer at rate 2 and the second enhancement layer at rate 1. The utility function is defined as: $U_m(x_m) = (M+1-m)\log(1+x_m)$, where $M+1-m$ can be viewed as weights associated with layer m.

Fig. 5 shows the evolution of allocated rate for each receiver in each layer, in the low-level optimization with a constant step size $\alpha(t) = 0.01$. Obviously, the whole rate allocation algorithm converges with a fast speed. For example, the base layer rate for R_1 reaches within 10% of its optimal value after 35 iterations and converges to 3.001 after 119 iterations. The first enhancement layer rate for R_2 reaches within 5% of its optimal value after 24 iterations and converges to 2.005 after 63 iterations.

Fig. 6 shows the achievable throughput on two receivers from shortest path tree, multicast capacity, LION algorithm [9] and the proposed algorithm. It is observed that video distribution with network coding offers a significant gain in throughput compared to that without network coding. And LION algorithm can achieve a multicast capacity of 5. Furthermore, using the proposed algorithm, R_1 and R_2 have successfully received two and three layers, which are just the maximal layer numbers they can subscribe to in terms of their respective max-flow capacity.

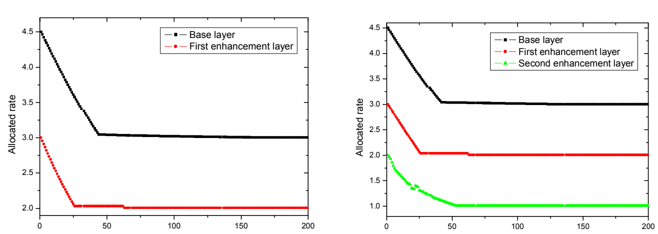

Fig. 5. Allocated rate for each receiver: (a) for R_1, (b) for R_2

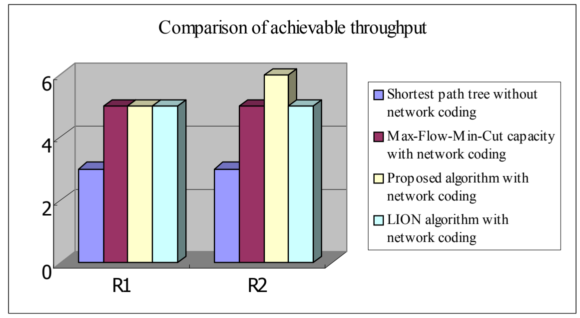

Fig. 6. Comparison of achievable throughput

To evaluate the received SVC quality, we present packet-level simulation results for a general access network in Fig. 4 (b), where the capacity (Kbps) is marked on each edge. This network model has 16 nodes (10 steiner nodes and 5 receivers) and 20 edges. The number of alternative paths for five receivers is $\{3,3,4,1,1\}$, and their maximum flows are $\{400, 550, 1300, 300, 280\}$. Five receivers are divided into three classes as shown in Fig. 3. Receiver 3 represents broadband broadcasting class, receiver 1 and 2 Internet class with moderate access capability, and receiver 4 and 5 mobile class with limited capacity.

The video source adopts Joint Scalable Video Model 9_10 reference codec of H.264/AVC extension standard, with two well-known test-sequences ("Foreman" and "Stefan") at frame rate of 30 fps, CIF resolution, and a GOP-length of 32 frames. They are encoded with 256 Kbps on the base layer, and 384Kbps, 512Kbps and 1024Kbps on the enhancement layers by FGS coding.

Table I compares the cost distribution of the shortest path algorithm, LION algorithm and the proposed algorithm calculated by the cost metric in Equation (2). It is observed that the proposed algorithm is the most efficient to make a global utility maximization over all receivers, because it maintains an overall balanced layered path routing at the minimum cost in an incremental order.

According to Equation (2), the cost in Table I can be regarded as linear related to the end-to-end delay. We vary the playout deadline for "Foreman" and "Stefan" streams from 200 ms to 500 ms, fixing the background traffic load to 15%. Comparison of average video quality in PSNR is shown in Table II. It is observed that the cost remains a direct proportion to the end-to-end delay. In the case of resource allocation by shortest path or LION scheme, the base layer packets for receiver 1 are dropped within a small playout deadline of 200ms owing to the high cost. As the playout deadline increases, larger packet delay can be tolerated. Once the playout deadline becomes 500ms, PSNR performance achieved by shortest path and LION scheme is similar to that obtained by the proposed algorithm.

V. CONCLUSION

This work investigates a network optimization for the

Table I. The cost distribution for all receivers

	R_1		R_2			R_3				R_4	R_5
	Layer 1	Layer 2	Layer 1	Layer 2	Layer 3	Layer 1	Layer 2	Layer 3	Layer 4	Layer 1	Layer 1
Shortest path	183.27	50.91	37.39	44.8	172.8	37.47	163.84	163.84	170.09	81.92	74.24
LION	161.98	81.97	55.74	40.36	58.92	85.18	30.15	52.36	-	81.92	74.24
Proposed algorithm	56.63	137.14	37.39	44.8	172.8	37.47	46.78	46.78	170.09	81.92	74.24

Table II. Received average video quality PSNR for Foreman and Stefan sequences, and the background traffic load is fixed at 15 %.

"Foreman" sequence, Playout deadline=200ms					
Average PSNR	R1	R2	R3	R4	R5
Shortest path	0	37.1	36	36	36
LION algorithm	0	38.22	38.22	36	36
Proposed algorithm	36	37.66	39.02	36	36

"Stefan" sequence, Playout deadline=200ms					
Average PSNR	R1	R2	R3	R4	R5
Shortest path	0	30.68	29.25	29.3	29.3
LION algorithm	0	31.81	31.81	29.3	29.3
Proposed algorithm	29.25	31.25	33.1	29.3	29.3

"Foreman" sequence, Playout deadline=500ms					
Average PSNR	R1	R2	R3	R4	R5
Shortest path	37.1	38.22	39.81	36	36
LION algorithm	37.1	38.22	38.22	36	36
Proposed algorithm	37.1	38.22	39.81	36	36

"Stefan" sequence, Playout deadline=500ms					
Average PSNR	R1	R2	R3	R4	R5
Shortest path	30.68	31.81	34.4	29.3	29.3
LION algorithm	30.68	31.81	31.81	29.3	29.3
Proposed algorithm	30.68	31.81	34.4	29.3	29.3

maximum achievable multicast capacity, where the network employs generalized multi-path and network coding based routing as well as flow control as an integral entity that jointly optimizes a global utility function over all receivers. By imposing the prior context of scalable video layers on convex optimization, we build lower layer path set with minimum path costs so as to preserve the inter-layer dependency from SVC structure. We develop a decentralized algorithm with two-level optimization update to solve the subproblems. Extensive numerical and simulation results validate the proposed algorithm under detailed critical performance factors.

REFERENCES

[1] 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 Technology, vol. 17, no. 9, pp. 1103-1120, Sept. 2007.

[2] X. Zhu, T Schierl, T. Wiegand, and B. Girod, "Video Multicast over Wireless Mesh Networks with Scalable Video Coding (SVC)", Proc. Visual Communication and Image Processing, VCIP 2008, Jan. 2008.

[3] M. van der Schaar, Y. Andreopoulos, and Z. Hu, "Optimized scalable video streaming over IEEE 802.11 a/e HCCA wireless networks under delay constraints," IEEE Trans. Mobile Computing, vol. 5, no. 6, pp. 755-768, June 2006.

[4] K. Kar, S. Sarkar, and L. Tassiulas, "Optimization based rate control for multirate multicast sessions," Proc. IEEE INFOCOM 2001, April 2001.

[5] S. Low and D. E. Lapsley, "Optimization flow control, I: Basic algorithm and convergence," IEEE/ACM Trans. Networking, vol. 7, pp. 861–874, Dec. 1999.

[6] R. Ahlswede, N. Cai, S.-Y. Li, and R. W. Yeung, "Network information flow," IEEE Trans. Inf. Theory, vol. 46, no. 4, pp. 1204–1216, Jul. 2000.

[7] L. Chen, T. Ho, S. H. Low, et al., "Optimization based rate control for multicast with network coding," Proceeding of IEEE Infocom, 2007.

[8] D. S. Lun, N. Ratnakar, M. Médard, et. al., "Minimum-cost multicast over coded packet networks," IEEE Trans. Inf. Theory, vol. 52, no. 6, pp. 2608-2623, Jun. 2006.

[9] J. Zhao, F. Yang, Q. Zhang, et.al., "LION: Layered overlay multicast with network coding," IEEE Transactions on Multimedia, vol. 8, no. 5, pp. 1021-1032, Oct. 2006.

[10] D.P. Bertsekas and J.N. Tsitsiklis, Parallel and Distributed Computation: Numerical Methods, Prentice-Hall, New Jersey, 1989.

[11] S. Shakkottai and R. Srikant, "Network Optimization and Control", in the series Foundations and Trends in Networking, vol. 2, no. 3, Now Publishers Inc, Boston - Delft, 2007.

[12] A. Nedic and D. P. Bertsekas, "Incremental Subgradient Methods for Nondifferentiable Optimization," SIAM J. on Optimization, vol. 12, 2001, pp. 109-138.

[13] D. Palomar and M. Chiang, "A tutorial on decomposition methods and distributed network resource allocation," IEEE Journal on Selected Areas in Communications, vol. 24, no. 8, pp. 1439-1451, Aug. 2006.