This full text paper was peer reviewed at the direction of IEEE Communications Society subject matter experts for publication in the IEEE ICC 2011 proceedings

Optimal Layered Multicast with Source Rate Adaption in Two-level Hierarchical Overlay Networks

Junni Zou[1], Leyang Li[1], Hongkai Xiong[2]
[1] Dept. of Communication Engineering, Shanghai University, Shanghai, 200072, China
[2] Dept. of Electronic Engineering, Shanghai Jiao Tong University, Shanghai, 200240, China

***Abstract*—Layered overlay multicast has emerged as an important solution for video streaming over heterogeneous network. In this paper, we formulate it into a joint network flow control and performance optimization problem where adaptive layer rates are determined through a greedy algorithm close to optimal. The overlay network is constrained by a practical two-level hierarchical overlay model where the bandwidth sharing is imposed on underlying edge-bottleneck links. To maintain less computational complexity and avoidance of global information, the M-layer maximization problem is transformed into multiple one-layer minimization subproblems. In turn, a fully distributed algorithm is developed by decomposing the one-layer minimization problem into both inequality constrained transportation subproblem and the shortest path subproblem by using Lagrangian duality. We demonstrate that the distributed algorithm converges quickly and attains effective performance in both static and dynamic networks.**

***Index Terms*—Layered multicast, overlay network, rate adaption, distributed algorithm**

I. Introduction

Unlike the conventional IP-layer multicast, application-layer multicast could overcome the deployment limitation and utilizes end-user's resources through overlay networks [1]. In view of the end-users' heterogeneity, multirate multicast was developed to adapt every client to its available bandwidth in a multimedia session. With the development of layered and scalable video coding, layered multicast as a variant of multirate multicast has been advanced to meet the volatile access of decodable data in a multirate-based progressive refinement [2]. In the layered multicast, the video stream is encoded into a sequence of progressively refinable layers, with each layer transmitted through a separate multicast session. From network performance perspective, resource allocation with layered overlay multicast should receive reasonable attention with optimal achievable rates.

There have been a lot of academic researches to address streaming video over layered multicast, which mainly focused on rate control such as protocol design for both congestion and error control [3]. Receiver-driven Layered Multicast (RLM) as a fundamental mechanism [4], adopted a probing mechanism called "join-experiment", which makes join/leave decisions of multicast groups based on observed packet losses. As an alternative approach, sender-driven multicast [5] advocated the source to adjust the number of layers and the bit rate of each layer dynamically according to the congestion feedback. However, this kind of end-to-end protocol exists difficulty in catching up with fast variations in the network and in coordinating receivers, leading to slow convergence and unfair bandwidth distribution among different sessions. Moreover, they constantly rely on additional network mechanisms to achieve enhanced performance and degrade the overall performance. From network performance perspective, optimization decomposition [6] has been casted towards a systematic understanding where the overall communication network is modeled by a generalized network utility maximization problem. In this sense, resource allocation may be analyzed and systematically designed as distributed solutions to some global optimization problems.

The first optimization model for the multirate multicast problem was studied by Kar et al. [7]. Since Ahlswede et al. [8] showed that network coding can achieve the capacity of single-source and multiple-terminal multicast, it has been paid a lot of attention in literature. Zhao et al. [9] addressed the layered multicast problem with network coding and multi-path constraints. However, they only formulated the objective function with fixed layered bit-rates and solved it by a heuristic approach. [10] urged receivers to get a non-consecutive layer or a partial layer, because the authors thought the undecodable data in the application layer might help other peers in the network layer. Besides the heuristic simulated annealing algorithm, it is also obvious that the application layer throughput could be drastically degraded at the cost of non-cumulative layering throughput. In [11], sender-driven layered multicast bit rates were calculated based on the max-flows of all receivers. The exhaustive-search method requires an exponentially growing computational complexity with the network heterogeneity.

For optimum overlay multicast, the existing researches mainly model the overlay as an independent network graph, in which the capacity share of physical links among overlay links are usually ignored [12]. Thus, the derived network throughput often deviates from its actual value. To the best of our knowledge, only the authors in [13] accurately describe the correlation between overlay and physical links by linear capacity constraints (LCC). The LCC model requires the explicit mapping of each overlay link to a physical path for constructing the maximum-bandwidth multicast tree, which is hard to implement in practice. In most realistic settings, it has been found that the congestion generally occurs at edge links [14]. This implies that the poor performance of the Internet arises primarily from constraints at the edges of the network. Within this paper, we incorporate the bandwidth share in the

The work has been partially supported by the NSFC grants No. 60632040, No. 60802019, and No. 60736043.

access network to model the overlay. In other words, the bottleneck link is assumed to be either the uplink of the sender or the downlink of the receiver.

We formulate the layered overlay multicast in a joint network flow control and performance optimization problem. To improve the bandwidth utilization as well as optimize the received video quality, adaptive layer rates are determined through a greedy algorithm close to optimal. To reflect the bandwidth sharing of the underlying bottleneck links, overlay network is modeled in the two-level hierarchical topology. In pursuit of less computational complexity and avoidance of global information, we convert the M-layer maximization problem to multiple one-layer minimization subproblems. The one-layer minimization problem is further decomposed into both inequality constrained transportation subproblem and the shortest path subproblem by using Lagrangian duality, and is solved by a fully distributed algorithm. Through simulations, we demonstrate that the distributed algorithm performs consistent with its centralized counterpart.

II. Network Modelling and Notations

A. Two-level Hierarchical Overlay Model

An overlay can be viewed as a hierarchical network consisting of two levels. The low level is an IP network including routers and physical links, while the high level is an application layer network constructed by end-users (overlay nodes) and overlay links. An overlay link is a virtual link that represents the unicast connection between two overlay nodes. Determined by IP routing algorithm, an overlay virtual link corresponds to a particular path of physical links in the low-level network. It is possible that multiple overlay links traverse common physical links, and share the underlying bandwidth.

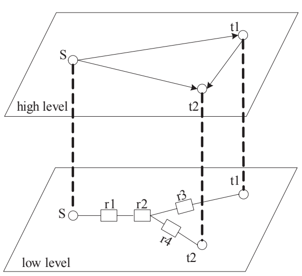

Fig. 1. Two-level overlay model

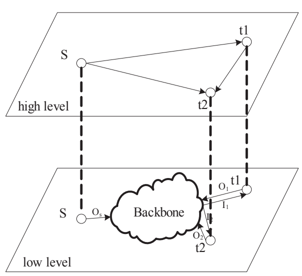

Fig. 2. Practical two-level overlay model

As Fig. 1, source node s, receivers t_1, t_2, routers $r_1 \sim r_4$, and involved links form a two-level overlay network. Given the IP routing, assume that overlay link (s, t_2) maps to a physical path $s \rightarrow r_1 \rightarrow r_2 \rightarrow r_4 \rightarrow t_2$, and link (t_1, t_2) maps to path $t_1 \rightarrow r_3 \rightarrow r_2 \rightarrow r_4 \rightarrow t_2$. Since these two overlay links share two physical bottleneck links, i.e., (r_2, r_4) and (r_4, t_2), their capacities are no longer independent, but correlated and constrained by the capacity of (r_2, r_4) and (r_4, t_2).

In the LCC model [13], the capacity of each overlay link is not labeled as a number. Instead, it is represented by a variable and a set of linear constraints used to specify the link share in the low-level network. Without the assumption of the global information, we focus on a reasonable two-level hierarchical overlay model where the bandwidth sharing is imposed on underlying edge-bottleneck links. Naturally, the low level of the overlay network is best modelled as a "cloud" model whose center represents the backbone, and each end-user connects to the cloud through a physical link. Each physical link with known characteristics provides a connection for the end-user to an Internet service provider (ISP).

The simplified overlay model is shown in Fig. 2. Supposing the uplink and downlink capacities of t_2 are, respectively, O_2 and I_2, the correlations of three overlay links with physical link (r_4, t_2) can be captured by the constraints $x_{(t_2,r_4)} \leq O_2$, and $x_{(s,t_2)} + x_{(t_1,t_2)} \leq I_2$. Here $x_{(t_2,r_4)}$ is the capacity variable for link (t_2, r_4), and so on. The complete correlation of the simplified model can be specified by the following matrix:

$$\begin{pmatrix} 1 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} x_{s,t_1} \\ x_{s,t_2} \\ x_{t_1,t_2} \end{pmatrix} \leq \begin{pmatrix} O_s \\ I_1 \\ O_1 \\ I_2 \\ O_2 \end{pmatrix}$$

B. Notations

Formally, the two-level overlay model can be formulated as consisting of:

1) A low-level (IP) graph $G_o(V_o, E_o)$, where V_o is the set of nodes, E_o is the set of access links; O_i and I_i represent the finite uplink and downlink capacities of node $i \in V_o$, respectively.

2) A high-level (overlay) directed graph $G(V, E)$, where $V = V_o$ and E is the set of virtual links. Each virtual link $(i, j) \in E$ corresponds to a low-level path $P(i, j) \in G_o$ from i to j. The set V can be further divided into three disjoint node sets, i.e., S, R for relay nodes and T for receiver nodes.

3) Correlation constraints. Since each path $P(i, j)$ will span node i's uplink and node j's downlink, the bandwidth share over node i's uplink and node j's downlink should be bounded by linear constraints $\sum_{\{j|(i,j)\in E\}} x_{ij} \leq O_i$ and $\sum_{\{i|(i,j)\in E\}} x_{ij} \leq I_j$, where x_{ij} represents the capacity variable of overlay link (i, j).

Suppose the video stream at the source is encoded into a set of M layers, with each layer m distributed over a multicast session m. B_m is the bit rate of layer m, $m = 1, 2, ..., M$. Limited by the max-flow capacity from s, each receiver t can subscribe to a proper number of layers. Let z_t^m represent whether receiver t can access to layer m. If it does, $z_t^m = 1$, else $z_t^m = 0$. Also, let f^{mt} be the information flow in session m from s to t via multi-paths and f_{ij}^{mt} as the flow rate of session m on virtual link (i, j).

III. Problem Statement

In this section, we illustrate an adaptive layer rates mechanism for the support of layered video multicast, and formulate the problem of optimal flow routing and performance optimization for layered video in overlay networks.

A. Adaptive Layer Rates Mechanism

Owing to the dynamics of the overlay network, layered video multicast with fixed bit rates is not sufficient to provide ideal bandwidth utilization or video quality for receivers with variable max-flow capacities. In this section, we target on the determination of an appropriate set of layer rates $\{B_m|m = 1, 2..., M\}$ to match the max-flow of the receivers in a best manner.

The set T is correspondingly partitioned into N subsets $T_1, T_2, ..., T_N$, such that:

1) The receivers in each subset T_p have identical max-flow F_p, $1 \leq p \leq N$. Here we use $|T_p|$ to denote the number of the receivers in T_p.

2) If $p > q$, $F_p > F_q$.

To ensure the fairness among receivers with the same max-flow capacity, we roughly denote the receiving rate of the receivers in the same subset T_p as R_p, i.e., $R_1 \leq R_2 \leq ... \leq R_N$. Generally, R_p can be calculated by:

$$R_p = \sum_{m=1}^{M} z_t^m \cdot B_m, \ \forall t \in T_p \tag{1}$$

The aggregate throughput Z of all receivers is then equal to

$$Z = \sum_{p=1}^{N} |T_p| \cdot R_p \tag{2}$$

Let $R_p = R_{p-1} + D_p$, where we set $R_0 = 0$, and $D_1 = R_1$, then R_p can also be: $R_p = \sum_{l=1}^{p} D_l$. It implies that the optimal layer rates set $\{B_m|m = 1, 2..., M\}$ can be uniquely determined by the set $\{D_p|p = 1, 2..., N\}$. Correspondingly, the formulation of Eq. 2 can be replaced by

$$Z = \sum_{p=1}^{N} |T_p| \cdot R_p = \sum_{p=1}^{N} |T_p| \cdot \sum_{l=1}^{p} D_l = \sum_{p=1}^{N} D_p \sum_{l=p}^{N} |T_l| \tag{3}$$

For a maximum Z, it has been proved that each element in $\{D_p\}$ is either zero or equal to $F_p - R_{p-1}$, meanwhile, the number of zero elements must be equal to $N - M$ [11]. Therefore, optimizing the layer rates $\{B_m\}$ is converted to finding out $N - M$ zero elements from $\{D_p\}$ under the constrains of network conditions. Here we adopt a low-complexity greedy algorithm to approximate the optimal solution. It iteratively selects one zero element in $\{D_p\}$ with which the overall throughput Z could has a maximum value, and stops until the zero elements reach $N - M$.

According to Eq. 3, we define the objective function of the greedy algorithm as follows:

$$F(D_1, D_2, ..., D_N) = \sum_{p=1}^{N} D_p \sum_{l=p}^{N} |T_l| \tag{4}$$

The details of the greedy algorithm are shown in Algorithm 1. In each round, an element i is moved from set P to P' when the objective function $F(D_1, D_2, ..., D_N)$ achieves the largest value at $D_i = 0$. Note that when calculating F , all the elements in $\{D_p\}$ with its index p falling into P' or equal to i are set to zero. The other elements are then recursively calculated by $D_p = F_p - \sum_{l=1}^{p-1} D_l$. The algorithm will go to an end once the number of zero elements in $\{D_p\}$ equals to $N - M$. Taking the M non-zero elements out of $\{D_p\}$ in an incremental order of index p, the optimal layer rates $\{B_m\}$ are achieved.

Algorithm 1 Greedy algorithm for layer rates optimization

Step 1. **Initialize** $P = \{1, 2, ..., N\}$, $P' = \varnothing$, and $D_p, p = 1, ..., N$ to some positive values.
Step 2. For each $i \in P$, calculate $F(D_1, ..., D_N)$, where $D_p = 0$, when $p \in P' \cup i$; $D_p = F_p - \sum_{l=1}^{p-1} D_l$, otherwise;
$f(D_i) = F(D_1, ..., D_N)$.
Step 3. $i' = \text{argmax}_{i \in P} f(D_i)$, Update P and P':
$P = P - i'$;
$P' = P' + i'$.
Step 4. If $|P| = M$ or $|P'| = N - M$, stop;
else go back to Step 2.
Step 5. Calculate $\{D_p|p = 1, 2..., N\}$, where $D_p = 0$, when $p \notin P$; $D_p = F_p - R_{p-1}$, when $p \in P$.

B. Problem Formulation

In this work, we aim at maximizing the aggregate end-to-end throughput for layered video distribution through the overlay network. For a reliable utilization of the overlay bandwidth, the hierarchical structure and correlation constrains are incorporated into a joint network flow control and performance optimization problem. Mathematically, it is formulated as:

$$\max \quad \sum_{t \in T} \sum_{m=1}^{M} z_t^m \cdot B_m \tag{5}$$

s.t. 1) $$\sum_{\{j|(i,j) \in E\}} f_{ij}^{mt} - \sum_{\{j|(j,i) \in E\}} f_{ij}^{mt} = \begin{cases} z_t^m \cdot B_m, & \text{if } i \in S; \\ -z_t^m \cdot B_m, & \text{if } i \in T; \\ 0, & \text{otherwise.} \end{cases}$$

$\forall m \in M, \ \forall t \in T, \ \forall i \in V.$

2) $z_t^m \geq z_t^{m+1}, \ \forall m \in M, \ \forall t \in T;$

3) $x_{ij}^m \geq f_{ij}^{mt}, \ \forall m \in M, \ \forall t \in T, \ \forall (i,j) \in E;$

4) $\sum_{m=1}^{M} \sum_{\{j|(i,j) \in E\}} x_{ij}^m \leq O_i, \ \forall i \in V;$

5) $\sum_{m=1}^{M} \sum_{\{j|(j,i) \in E\}} x_{ji}^m \leq I_i, \ \forall i \in V.$

Constraint 1) is the information flow balance equation. If $z_t^m = 1$, receiver t should have a quantity B_m of available bandwidth from s. Constraint 2) specifies the layer receiving order for the receivers. If receiver t could access to layer m, it must have already received all lower layers $1, 2, ..., m - 1$. Constraint 3) represents the relationship between information flow rate and physical flow rate on each link where network coding is applied to information flows of the same video layer. Here x_{ij}^m denotes the total bandwidth allocated on link (i, j) for session m. Constraint 4) and 5) are capacity constraints for underlying bottleneck links.

C. Distributed Algorithm

In pursuit of low computational complexity and decentralized implementation, we propose a distributed approximate algorithm. It sequentially allocates the bandwidth for each layer from low to high in a distributed manner, until there are no enough bandwidth left for the next layer. When constructing the video distribution meshes for each layer, its objective is to minimize the total bandwidth consumed at the current layer, so as to reserve more available bandwidth for the next higher layers. The one-layer optimization model is shown in Eq. 6.

$$\min \sum_{(i,j)\in E} x_{ij}^m \tag{6}$$

$$\text{s.t.}\quad 1) \sum_{\{j|(i,j)\in E\}} f_{ij}^{mt} - \sum_{\{j|(j,i)\in E\}} f_{ij}^{mt} = \begin{cases} B_m, & \text{if } i\in S; \\ B_m, & \text{if } i \in T; \\ 0, & \text{otherwise.} \end{cases}$$
$$\forall t\in T,\ \forall i \in V.$$
$$2)\ x_{ij}^m \geq f_{ij}^{mt},\ \forall (i,j)\in E,\ \forall t \in T;$$
$$3) \sum_{\{j|(i,j)\in E\}} x_{ij}^m \leq O_i^m,\ \forall i \in V;$$
$$4) \sum_{\{j|(j,i)\in E\}} x_{ji}^m \leq I_i^m,\ \forall i \in V.$$

Note that only the receivers whose max-flow in the residual network are larger than B_m can join session m. Constraint 3) and 4) are capacity correlation constraints, where O_i^m and I_i^m represent the capacity of node i's uplink and downlink in the residual network, respectively. After fulfilling the bandwidth allocation for each layer m, the link capacity is updated as follows:

$$O_i^{m+1} = O_i^m - \sum_{\{j|(i,j)\in E\}} x_{ij}^m \tag{7}$$

$$I_i^{m+1} = I_i^m - \sum_{\{j|(j,i)\in E\}} x_{ij}^m \tag{8}$$

To solve the one-layer optimization model in a fully distributed way, we relax constrain 2) with Lagrange multiplier λ_{ij}^{mt}, and formulate the following Langrangian:

$$L(\mathbf{x},\mathbf{f},\lambda) = \sum_{(i,j)\in E} x_{ij}^m (1 - \sum_{t\in T} \lambda_{ij}^{mt}) + \sum_{t\in T}\sum_{(i,j)\in E} \lambda_{ij}^{mt} f_{ij}^{mt} \tag{9}$$

The Lagrange dual function $G(\lambda)$ is given by

$$G(\lambda) = \min L(\mathbf{x},\mathbf{f},\lambda) \tag{10}$$

The Lagrange dual problem corresponding to the primal problem 6 is then given by: $\max_{\lambda\geq 0} L(\lambda)$. It is observed that the Lagrangian problem in 10 can be further decomposed into two subproblems [15]. One is an inequality constrained transportation problem:

$$\min \sum_{(i,j)\in E} x_{ij}^m (1 - \sum_{t\in T} \lambda_{ij}^{mt}) \tag{11}$$

$$\text{s.t.}\quad 1) \sum_{\{j|(i,j)\in E\}} x_{ij}^m \leq O_i^m,\ \forall i\in V;$$
$$2) \sum_{\{j|(j,i)\in E\}} x_{ji}^m \leq I_i^m,\ \forall i \in V;$$
$$3)\ x_{ij}^m \geq 0,\ \forall (i,j)\in E.$$

The other are $|T|$ shortest path problems:

$$\min \sum_{(i,j)\in E} \lambda_{ij}^{mt} f_{ij}^{mt} \tag{12}$$

$$\text{s.t.}\quad 1) \sum_{\{j|(i,j)\in E\}} f_{ij}^{mt} - \sum_{\{j|(j,i)\in E\}} f_{ij}^{mt} = \begin{cases} B_m, & \text{if } i\in S; \\ B_m, & \text{if } i \in T; \\ 0, & \text{otherwise.} \end{cases}$$
$$\forall t\in T,\ \forall i \in V.$$
$$2)\ f_{ij}^{mt} \geq 0,\ \forall (i,j)\in E.$$

At the k-th iteration of the subgradient algorithm, the lagrangian multiplier λ_{ij}^{mt} is updated by

$$\lambda_{ij}^{mt}(k+1) = \Big[\lambda_{ij}^{mt}(k) + \alpha(k)\Big(f_{ij}^{mt}(k) - x_{ij}^m(k)\Big)\Big]^+ \tag{13}$$

where α is a positive step size, and $[\cdot]^+$ denotes the projection onto the set of non-negative real numbers.

IV. SIMULATION RESULTS

This section presents simulation results to evaluate the performance of the proposed layered overlay multicast strategy. Our experiments are conducted over random network topologies. The average number of neighbors for each node is five, and each node has 0.5-1.5 Mbps of downlink capacity and 0.5-0.8 Mbps uplink capacity. Without a special specification, the number of receiver nodes in the topology is 20, and the video source is encoded into three scalable layers.

Firstly, the performance of the adaptive layer rates scheme based on the proposed greedy algorithm is evaluated and compared with the exhaustive-search method [11]. In each scenario, two metrics, i.e., the normalized aggregate throughput and the normalized running times are evaluated. The normalized aggregate throughput is defined as the ratio of the aggregate throughput achieved by the proposed greedy method to the optimal value from the exhaustive-search method; the normalized running times is defined as the ratio of running times spent by the proposed greedy method to that by the exhaustive-search method. Here the running times is referred to the number of times required for calculating the objective function $F(D_1, D_2, ..., D_N)$. Thus, the normalized running times can be calculated as follows:

Normalized running times $= \frac{(N-M)(N+M+1)/2}{N!/M!(N-M)!}$

Table I demonstrates the impact of the network heterogeneity (i.e., the number of the receiver classes N) on the performance of the greedy method. It can be found that, regardless of the network heterogeneity levels, the aggregate throughput achieved by the greedy method can closely approximate the global optimal value. On the other hand, the running times spent by the greedy method is significantly smaller than that of the exhaustive-search method. Also, the normalized running times drops sharply with the increase of N. It comes to the conclusion that the proposed greedy method is extremely applicable to the large-scale overlay networks.

Table II shows the effect of the video quality scalability (i.e., the number of the encoded video layers M) on the proposed method, where the receivers are divided into 20 classes. Namely, all the receivers are assumed to have different max-flow capacities. It is observed that with the increase of M , the normalized aggregate throughput is gradually improved, which means the solution by the greedy method draws more closely to the global optimum. That is because at smaller granularity of the quality scalability, receivers are more likely to fully saturate their link capacities and approximate their

TABLE I
IMPACT OF NETWORK HETEROGENEITY ($M = 3$)

	$N = 5$	$N = 10$	$N = 20$
Normalized network throughput	1.0	1.0	0.9923
Normalized running times	0.9	0.4083	0.1789

TABLE II
IMPACT OF VIDEO QUALITY SCALABILITY ($N = 20$)

	$L = 4$	$L = 5$	$L = 6$
Normalized network throughput	0.9892	0.9895	0.9898
Normalized running times	0.0413	0.0126	0.0048

max-flow rates. Moreover, the computational complexity of the greedy algorithm tremendously reduces with the number of video layers.

Fig. 3 presents the convergence speed of the proposed distributed algorithm in networks of different sizes. It can be seen that the whole distributed approach converges with a fast speed. Regardless of the network sizes, the number of iterations required for each layer to reach the convergence is within 140. And the total number of the iterations for three layers tends to be identical under large-scale network. Specifically, when the network scale is small (e.g., $|V| = 50$), the available bandwidth is insufficient for all the receivers to join in Layer 3. Thus, the rate allocation in Layer 3 is much simpler than in Layer 1 and Layer 2, which results in a small number of iterations compared to those of Layer 1 and 2. Conversely, under a large-scale network (e.g., $|V| = 250$), there would be enough bandwidth for delivering higher layers.

Fig. 4 displays the aggregate throughput of all the receivers under dynamic and static layer rates. For the static layer rates scheme, the rates for three layers are fixed to 256 Kbps, 128 Kbps, 128 Kbps. We can find that the proposed distributed algorithm with dynamic layer rates outperforms the one with static layer rates under all kinds of network scales. In contrast to static scheme, our adaptive layer rates increases the network throughput up to 115 %.

V. CONCLUSION

This paper formulates the layered overlay multicast into a joint network flow control and performance optimization problem, where adaptive layer rates are determined through a greedy algorithm close to optimal and overlay network is constrained by the two-level hierarchical topology with bandwidth sharing on edge-bottleneck links. For a fully distributed algorithm, we convert the M-layer maximization problem to multiple one-layer minimization subproblems while the one-layer minimization problem is further decomposed into both inequality constrained transportation subproblem and the shortest path subproblem through Lagrangian duality. It is validated that the distributed algorithm converges quickly and is highly applicable. In future work, we will further investigate the performance of the proposed distributed algorithm in dynamic settings.

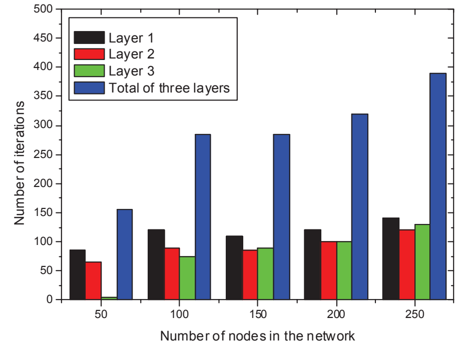

Fig. 3. Convergence speed vs. network scales

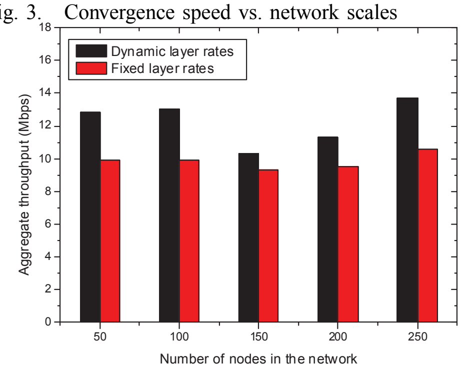

Fig. 4. Achievable throughput vs. network scales

REFERENCES

[1] S. Banerjee, B. Bhattacharjee, and C. Kommareddy, "Scalable Application Layer Multicast," *Proc. ACM SIGCOMM*, Pittsburgh, USA, Aug. 2002.
[2] B. Li and J. Liu, "Multirate video multicast over the internet: an overview," *IEEE Network*, vol. 17, no. 1, pp. 24-29, 2003.
[3] Q. Zhang, Q. Guo, Q. Ni, et al., "Sender-adaptive and receiver-driven layered multicast for scalable video over the Internet," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 15, no. 4, pp. 482-495, Apr. 2005.
[4] L. Cheng and M.R. Ito, "Receiver-Driven Layered Multicast Using Active Networks," *Proc. IEEE International Conf. Multimedia and Expo (ICME'03)*, pp. 501-504, July 2003.
[5] B. J.Vickers, C. Albuquerque, and T. Suda, "Adaptive multicast of multilayered video: rate-based and credit-based approaches," *Proc. IEEE INFOCOM*, pp. 1073-1083, 1998.
[6] M. Chiang, S. H. Low, A. R. Calderbank, and J. C. Doyle, "Layering as optimization decomposition: A mathematical theory of network architectures," *Proceedings of the IEEE*, vol. 95, no. 1, pp. 255-312, Jan. 2007.
[7] K. Kar, S. Sarkar, and L. Tassiulas, "Optimization based rate control for multirate multicast sessions," *Proc. IEEE INFOCOM 2001*, April 2001.
[8] 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.
[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] A. Gopinathan and Z. Li, "Optimal layered multicast with network coding: mathematical model and empirical studies," *Proceedings of 16th IEEE/ACM International Symposium on Modeling, Analysis, and Simulation of Computer and Telecommunication Systems*, Baltimore, USA, Sept. 2008.
[11] J. Si, B. Zhuang, A. Cai, "Optimal layered multicast using network coding," *Proceedings of the 4th workshop on Network Coding, Theory and Applications*, Hong Kong, China, Jan. 2008.
[12] Y. Chu, S. G. Rao, S. Seshan, and H. Zhang, "A case for end system multicast," *IEEE Journal on Selected Areas in Communications*, vol. 20, no. 8, pp. 1456-1471, Oct. 2002.
[13] Y. Zhu, B. Li, and K.Q. Pu, "Dynamic multicast in overlay networks with linear capacity constraints," *IEEE Transactions on Parallel and Distributed Systems*, vol. 20, no. 7, pp. 925-939, Jul. 2009.
[14] G. Kramer and G. Pesavento, "Ethernet Passive Optical Network (EPON): Building a Next-Generation Optical Access Network", *IEEE Communications Magazine*, vol. 40, no. 2, pp. 66-73, Feb. 2002.
[15] C. Wu, B. Li, "Optimal Rate Allocation in Overlay Content Distribution," *Proceedings of IFIP Networking 2007, Lecture Notes in Computer Science*, Springer, vol. 4479, pp. 678-690, Atlanta, USA, May 2007.