# Multicast Flow Aggregation in IP over Optical Networks

Yi Zhu, *Student Member, IEEE*, Yaohui Jin, *Member, IEEE*, Weiqiang Sun, *Member, IEEE*, Wei Guo, Weisheng Hu, *Member, IEEE*, Wen-De Zhong, *Senior Member, IEEE*, and Min-You Wu, *Senior Member, IEEE*

***Abstract*— It is widely believed that IP over optical networks will be a major component of the next generation Internet. However, it is not efficient to map a single multicast IP flow into one light-tree, since the bandwidth of an IP flow required is usually much less than that of a light-tree.**

**In this paper, we study the problem of multicast flow aggregation (MFA) in the IP over optical two-layered networks under the overlay model, which can be defined as follows: given a set of head ends (i.e. optical multicasting sources), each of which can provide a set of contents (i.e. multicast IP flows) with different required transmission bandwidth, and a set of requested content at the access routers (i.e. optical multicasting destinations), find a set of light-trees as well as the optimal aggregation of multicast IP flows in each light-tree.**

**We model MFA by a tri-partite graph with multiple criteria and show that the problem is NP-complete. Optimal solutions are designed by exploiting MFA to formulate an integer linear programming (ILP), with two parameters: the multicast receiving index $\alpha$ and the redundant transmitting index $\beta$. We also propose a heuristic algorithm. Finally, we compare the performance of MFA for different combination of $\alpha$ and $\beta$ via experiments and show our heuristic algorithm is effective for large-scale network in numerical results.**

***Index Terms*— Multicast flow aggregation, IP over optical network, overlay model, tri-partite graph, integer linear programming (ILP), heuristics**

## I. Introduction

MULTICASTING technology [1] has become increasingly important since many new applications such as content delivery, IP-TV, video conferences, and multiple-player gaming require the transmission of real-time multimedia from one source to many destinations. However, multicasting in IP layer [2] is not widely deployed in today's Internet, partially because there are still many technical issues such as scalability, reliability, security, and QoS control in current IP multicasting [3]–[5]. Recently, optical layer multicasting by using light-trees has been proposed to support point-to-multipoint communication in wide area networks [6]. The use of light-tree can better support bandwidth-intensive applications with guaranteed QoS, but there are also many challenging issues in optical layer multicasting [7]. In the data plane, work has focused on the design of multicast-capable optical crossconnects (MC-OXC) [8], [9]. In the network area, a lot of efficient algorithms have been developed for multicast routing and wavelength assignment [10], [11]. The extension to generalized multi-protocol label switching (GMPLS) has been presented for dynamic control of point-to-multipoint connections in optical networks [12]–[14].

Another critical and important issue is how to interwork the optical multicasting and the existing IP multicasting. In our previous efforts, we have proposed and demonstrated a new multicast-IP over light-tree network model [14]–[16]. The key idea is that a new optical network, with the capability of providing dynamic point-to-multipoint connections, replaces the conventional IP multicasting network in the core, while the edge remains an IP multicasting network. This hierarchical multicast architecture offers several attractive features. Firstly, it is compatible with the existing IP multicasting architecture since it is still IP based at the network edge. Therefore it is not necessary to change or upgrade existing client-side equipment. Secondly, it can support large-scale multicasting because the aggregation of IP sessions significantly reduces the burden of group management and multicast routing protocols. Thirdly, light-trees can provide better survivability for their inside aggregated IP flows [17], [18]. Finally, it can provide multicasting with improved QoS due to the fact that optical point-to-multipoint connections in the core network are circuit-switched with negligible delay and jitter.

Generally, the bandwidth of a light-tree is much higher than that required by many typical applications. It is not efficient to map a single multicast IP flow into one light-tree. In this paper, we study the problem of IP multicast flow aggregation (MFA) in the multicast capable optical networks, which can be defined as follows: given a set of head ends (i.e. optical multicasting sources), each of which can provide a set of contents (i.e. multicast IP flows) with different required transmission bandwidth, and a set of requested contents at the access routers (i.e. optical multicasting destinations), find a set of light-trees to accommodate the optimally aggregated multicast IP flows. We would like to point out that it is more consistent with the practical scenarios that the same content can be redundantly distributed in several head ends. Therefore,

Manuscript received April 15, 2006; revised December 3, 2006. This work was supported by Natural Science Foundation of China under Contract 60502004, 60602010, 60632010, and in part by Chinese 863 Program. Portions of this paper have appeared in IEEE GLOBECOM'05, St. Louis, MO, Nov. 2005

Yi Zhu was with Shanghai Jiao Tong University. He is now with the University of Texas at Dallas, Richardson, TX 75083 USA (e-mail: yxz053100@utdalls.edu).

Yaohui Jin, Weiqiang Sun, Wei Guo, and Weisheng Hu are with State Key Lab of Advanced Optical Communication System and Network, Shanghai Jiao Tong University, China (e-mail: {jinyh, wguo,sunwq, wshu}@sjtu.edu.cn).

Wen-De Zhong is with Network Technology Research Centre, School of EEE, Nanyang Technological University, Singapore (e-mail: ewdzhong@ntu.edu.sg).

Min-You Wu is with Computer Science Department, Shanghai Jiao Tong University, China (e-mail: mwu@sjtu.edu.cn).

Digital Object Identifier 10.1109/JSAC.2007.070613.

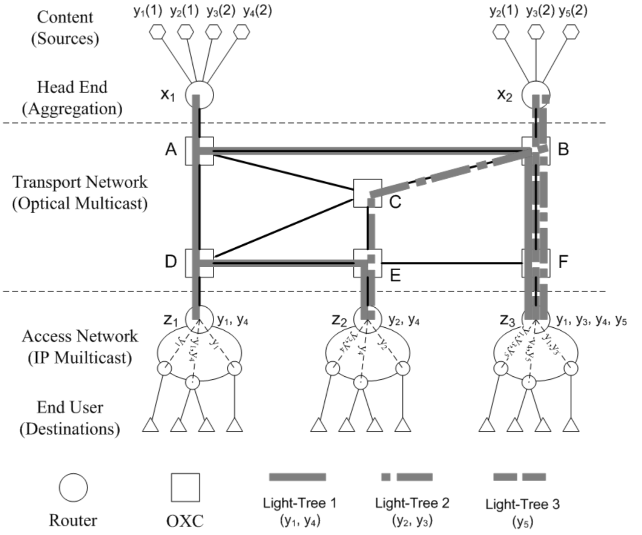


Fig. 1. Multicast flow aggregation in IP over optical two-layered network

the source of IP multicast flow is not given before optimal aggregation in the MFA problem, which is different from other work on aggregated multicasting and multicast grooming in the literatures. We model the problem with a tri-partite graph, and prove the NP-Completeness. We then develop an integer linear programming (ILP) for the MFA problem. We also propose a heuristic named least trees first (LTF) algorithm to solve the problem for large-scale networks with hundreds of contents. Numerical and simulation experiments are carried out to verify the effectiveness of our proposed model.

The rest of the paper is organized as follows. In section II, we introduce the concept of MFA and discuss its implementation in IP over optical networks. Then we present a tri-partite graph for the MFA problem and provide an overview of the related work. In section III, we develop the ILP for the MFA problem. In section IV, we propose a heuristic algorithm for the MFA problem. In section V, we present a few experimental results. Finally, section VI concludes the paper.

## II. Multicast Flow Aggregation

### A. Network model

Fig. 1 shows the network model considered in this paper. The head ends are multicasting sources which encode multimedia contents with multicast IP flows and then send such flows to the core network. The required bandwidth of a single multicast IP flow may vary, depending on its image quality and compression standard. For example, by using MPEG-2 technology, standard definition TV (SDTV) and high definition TV (HDTV) require nearly 6 Mb/s and 25 Mb/s transmission bandwidth respectively [19]. Note that the same content may be redundantly provided by several head ends in the network considered in this study, similar to the scenario of today's TV delivery networks where one channel may be distributed by several stations.

At the edges of the core optical network, there are two kinds of aggregation routers. At the head end side, the routers aggregate several multicast IP flows into one light-tree. At the access network side, the routers aggregate a large number of end users' requests originated at one residential area. After receiving the multicast contents via a light tree, each access network delivers the multicast contents through IP multicasting to the end users. The thin dashed lines in each access network indicate IP multicast flows, where the numbers on each dashed line denote the contents being delivered to a particular end user.

In the rest of this paper, for simplicity, we refer to the aggregation routers in the head ends and in the access networks as the heads and the tails respectively. There may be more than one transponders attached to the core optical network at one aggregation router. The transponders in the heads and the tails are the roots and the leaves of the light trees. The bit rate of one transponder is usually much higher than the required bandwidth of a single multicast IP flow. For example, the typical bit rates of Ethernet based interfaces are 100 Mb/s, 1 Gb/s and 10 Gb/s, while that of SONET/SDH based interfaces are 155 Mb/s, 622 Mb/s, 2.5 Gb/s and 10 Gb/s. Therefore, several multicast IP flows can be aggregated into one light-tree. In the network of figure 1, the bit rate of one transponder is assumed to be 3 units of bandwidth. The value in the parentheses following the content number denotes the required bandwidth for that content. If we employ the IP multicasting scheme only, each content corresponds to at least one IP multicast flow. However, if we use the IP over optical two-layer network model described above, only 3 light-trees are needed for 5 multicast IP flows in optical core network, which significantly reduces the number of multicasting trees.

### B. Implementation

In our network model, we require that: 1) the light-tree is constructed by using the shortest path tree, which means that every content is transmitted from head to one tail in the light-tree along the shortest path. 2) Both multicast IP flows and light-trees are one-to-many unidirectional connections. 3) A multicast IP flow cannot be split into two or more light-trees originated at a single head. 4) The transmitting and receiving transponders of a light-tree must be the same type with the same bit rate. 5) The intermediate OXC do not have grooming capability so that the maximum bandwidth of a light-tree is the bit rate of transponders at the aggregation routers.

Fig. 2 shows the implementation of multicasting flow aggregation in the "3TNet" project funded by Chinese "863" program, which employs IP over automatically switched optical networks (ASON) over dense wavelength division multiplexing (DWDM) 3-layer architecture. A field-trial has been carried out in Yangtze River Delta [20]. The interconnection of control plane between IP and ASON is based on the overlay model [21]. In the core optical networks, the OXC's in the data plane and GMPLS as well as user network interface (UNI) in the control plane have been extended to support point-to-multipoint (P2MP) connections [14], [15]. In the access networks, a tail control unit (TCU) collects the user requests information from each router via simple network management protocol (SNMP) and reports to a central scheduler. The scheduler calculates the optimal aggregation based on the content distribution and the user requests, and then sends the aggregation schemes to the corresponding head control units (HCU). The HCU is responsible for creating light-

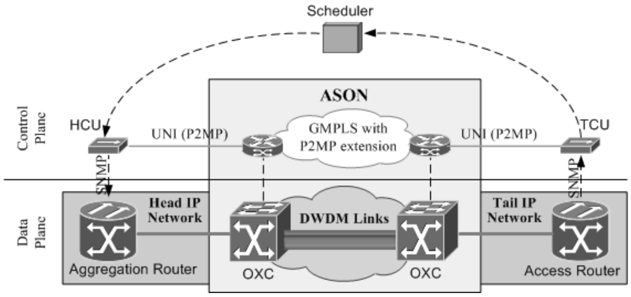


Fig. 2. Implementation of Multicast Flow Aggregation in IP over Optical networks

TABLE I

THE RESOURCE UTILIZATION OF MULTICAST IP FLOWS IN FIG. 1

| Multicast IP flow | Bandwidth | Source | Destinations | Resource utilization |
|---|---|---|---|---|
| $y_1$ | 1 | $x_1$ | $z_1$, $z_3$ | 3 |
| $y_2$ | 1 | $x_2$ | $z_2$ | 2 |
| $y_3$ | 2 | $x_2$ | $z_3$ | 2 |
| $y_4$ | 2 | $x_1$ | $z_1$, $z_2$, $z_3$ | 8 |
| $y_5$ | 2 | $x_2$ | $z_3$ | 2 |
| Sum | | | | 17 |

TABLE II

THE RESOURCE UTILIZATION OF LIGHT-TREES IN FIG. 1

| Light-tree | Bandwidth | source | destinations | Resource utilization |
|---|---|---|---|---|
| 1 | 3 | $x_1$ | $z_1$, $z_2$, $z_3$ | 12 |
| 2 | 3 | $x_2$ | $z_2$, $z_3$ | 9 |
| 3 | 3 | $x_2$ | $z_3$ | 3 |
| Sum | | | | 24 |

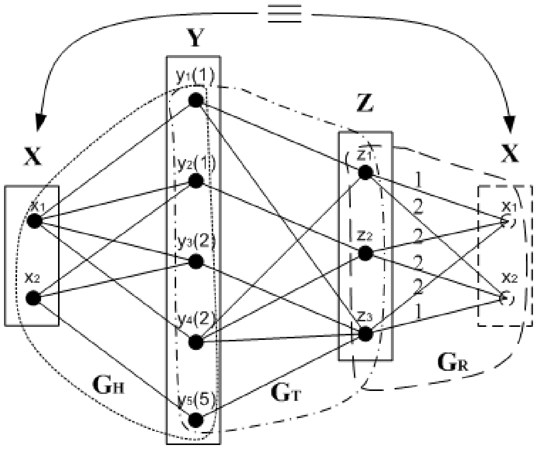


Fig. 3. Tri-partite graph for multicasting flow aggregation

tree in optical networks via UNI signaling and configuring aggregation in the routers of head ends via SNMP.

### C. Resource overhead

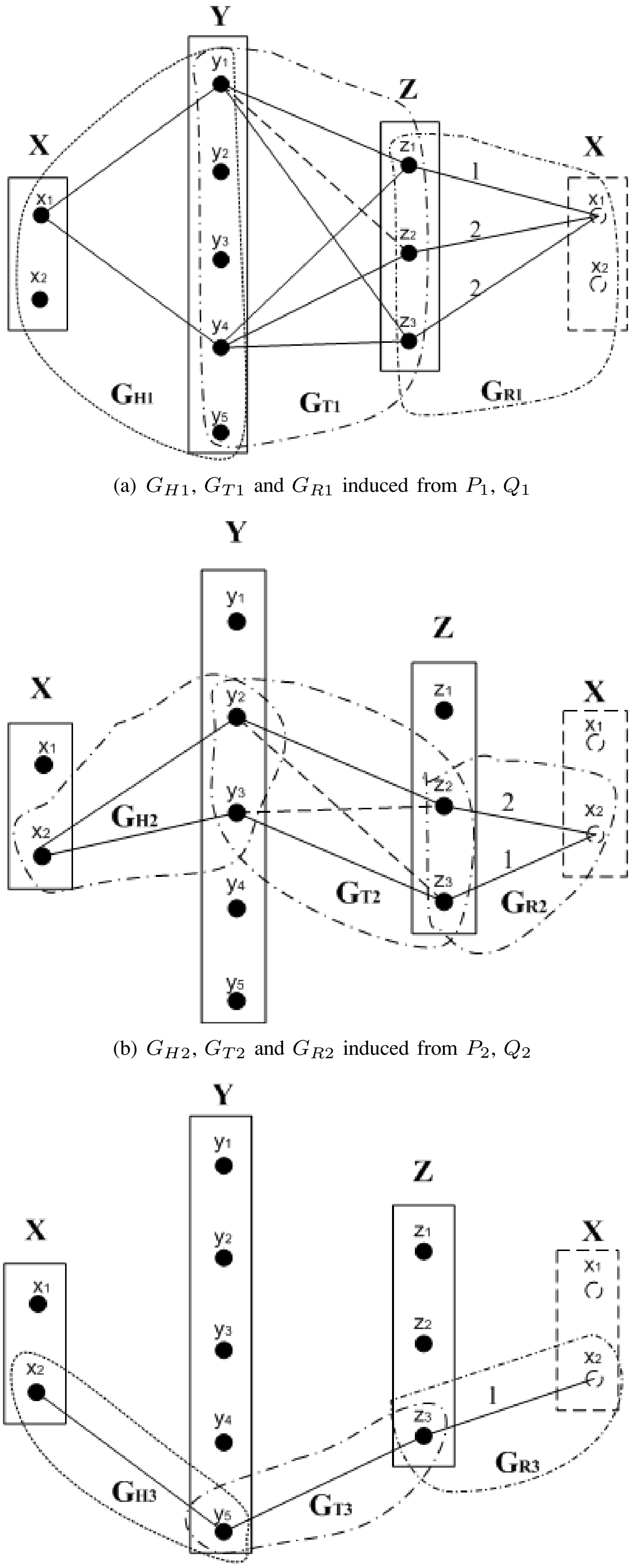
Although the MFA strategy provides many benefits for large-scale streaming media delivery, it may lead to extra resource overhead. We define the resource utilization of a tree is equal to the product of its bandwidth and all the links along the tree. For the case without MFA strategy, we consider that the core network is based on IP routers with the same topology as the optical network and the IP flows are transmitted along the same shortest path as the light-tree. For example, Tables I and II show the resource utilization of multicast IP flows and light-trees respectively in network of figure 1. 5 multicast IP flows would consume 17 units of bandwidth while 3 light-trees will do 24 units. The resource overhead of MFA is 7 units of bandwidth. This overhead comes from mismatch both in the heads and in the tails. In the heads, if the total bandwidth of aggregated IP flows is less than the transponder bandwidth, it will lead to head wastage, e.g. light-tree 3 wastes 1 unit in head $x_2$. In the tails, if one access router receives unwanted contents, it will lead to tail wastage, e.g. tails $z_2$and $z_3$ received unwanted contents $y_3$ and $y_2$ respectively in light-tree 2. The tail wastage is also called leaky match in other literatures [22] because the multicast IP flow cannot be perfectly matched into the aggregated tree. The objective of optimal MFA problem is to minimize the overall resource overhead.

### D. Tri-partite graph formulation

The MFA problem can be abstracted as a tri-partite graph $G(X \cup Y \cup Z, E_H \cup E_T \cup E_R)$, which is composed of three bipartite sub-graphs $G_H(X \cup Y, E_H)$, $G_T(Y \cup Z, E_T)$, and $G_R(X \cup Z, E_R)$, where the vertices $X$, $Y$ and $Z$ are the sets of heads, contents and tails, respectively, and the edges $E_H$, $E_T$ and $E_R$ are the relationship among them. As an example, figure 3 shows the tri-partite graph representation of figure 1. In order to illustrate the graph on the plane more clearly, we repeat the set $X$ with dashed line on the right side of $Z$ to show the subgragh $G_R$. For simplicity, we consider homogenous transponders whose bandwidth is t. The weight $b_m$ on the vertex $y$ is its required bandwidth. The weight $c_{x,z}$ on $E_R$ represents the length of the shortest path from head $x$ to tail $z$. Note that here we just use the length of the shortest path instead of network topology for the core optical network because our implementation is the overlay IP-over-Optical model, in which the detailed network topology information of optical layer should not be exposed to upper IP layer [21]. Such linear approximation partially reflects to the network routing due to the following reasons: 1) we do not know which head will be the root and which tails are the leaves before aggregating IP flows to light-tree; 2) given a light-tree, the sum of the shortest-path length from its root to all the leaves is the upper bound of its total links since the shortest-path light-tree may share some of links of the shortest paths. For example in figure 1, light-tree 1 consumes 4 links in the optical networks, while the sum of the shortest-path length is 5 links because link A-D is shared by $z_1$ and $z_2$.

Let $P = \{P_1, P_2, \ldots, P_k, \ldots\}$, where $P_k$ is a bipartite graph with $\{x_k \in X, Z_k \subseteq Z\}$, representing a root-leaves set of one light-tree, and $Q = \{Q_1, Q_2, \ldots, Q_k, \ldots\}$, where $Q_k \subseteq Y$ represents an aggregation group of contents in a light-tree $P_k$. Three complete bipartite graphs $G_{Hk}$ with $\{x_k,\ Q_k\}$, $G_{Tk}$ with $\{Q_k, Z_k\}$ and $G_{Rk}$ with $\{x_k, Z_k\}$ are induced from $P_k$ and $Q_k$ as shown in figure 4. We define the edge difference for the combination $P_k$ and $Q_k$.

$$d(Q_k) = |Q_k| \cdot |Z_k| - |E(G_T \cap G_{Tk})| \tag{1}$$



(a) $G_{H1}$, $G_{T1}$ and $G_{R1}$ induced from $P_1$, $Q_1$

(b) $G_{H2}$, $G_{T2}$ and $G_{R2}$ induced from $P_2$, $Q_2$

(c) $G_{H2}$, $G_{T2}$ and $G_{R2}$ induced from $P_2$, $Q_2$

Fig. 4. Induced graph from $P$ and $Q$

If $d(Q_k) = 0$, it is called perfect match; otherwise it is leaky match indicated by dashed lines in figure 4, which corresponds to the unwanted contents in the tails.

$P$ and $Q$ satisfy the following constraints:

a) in $G_H$, $\exists x_k \in X, \forall y \in \mathrm{Q_k}, (x, y) \in E_H$, which means all elements in $Q_k$ must connect to the same node $x_k$ in $G_H$;

b) the neighbor set of $Q_k$ must be equal to $Z_k$ in $G_T$, which means that the contents in $Q_k$ are the aggregated multicast flows that tails $Z_k$intend to receive;

c) $\sum_{y_m \in Q_k} b_m \leq t$, which means the total bandwidth of aggregated IP flows in one light-tree must be not greater than the effective bandwidth of a transponder.

*Definition 1:* The multicast flow aggregation (MFA) problem can be stated as follows:

**Given** a tri-partite graph $G(X \cup Y \cup Z, E_H \cup E_T \cup E_R)$ and t

**Find** an optimal combination of P and Q satisfies with the above constrains such that:

d) The total head wastages $\sum_k \left(t - \sum_{y_m \in Q_k} b_m\right)$must be as small as possible;

e) The total edge difference $\sum_k d(Q_k)$must be as small as possible;

f) The total length of light-trees $\sum_k (\sum_{x \in P_k, z \in Z_k} c_{x,z})$ must be as small as possible.

In appendix, we prove that the MFA problem is NP-complete.

### *E. Related work*

In IP networking research area, the MFA problem is related to aggregated multicast (AM) model [22]–[24]. The motivation of AM is to solve scalability and reliability issues in traditional IP multicast. The key idea is that, multiple IP multicast sessions is "forced" to share a single distribution tree in the core network so as to reduce the number of multicast states. Compared to the AM problem, the MFA problem considers the bandwidth constraint of a light-tree as well as the redundant distribution of contents in different head ends. Furthermore, instead of a dynamic scenario in the AM problem, the MFA problem is a static optimization problem that can help us to find a theoretical lower bound.

Recently, the multicast grooming (MG) problem has begun to attract attention in WDM networks, which is defined as follows: given a set of multicast sessions with various capacity requirements, satisfy all of the multicast sessions, and minimize the network cost at the same time [25]. To support grooming of multicast traffic in an optical network, the switch architecture must be enhanced with a grooming fabric [6]. The related research works of the MG problem can be categorized into: static optimization [26]–[29] and dynamic scenario [30], [31]. We note that the MG problem is different from the MFA problem in the following aspects: 1) the MG problem may have multi-hops in the optical network while the light-tree is transparent in the MFA network model; 2) the MG problem does not have tail wastage issue since multicast IP flows can be groomed at the intermediate nodes in the core network; 3) the roots of multicast sessions are given in the MG problem while they are not given before aggregation in the MFA problem since the contents are redundantly distributed in the different head ends.

| | |
|---|---|
| $N = \vert X\vert$ | the total number of heads |
| $M = \vert Y\vert$ | the total number of contents |
| $I = \vert Z\vert$ | the total number of tails |
| $\mathbf{S} := [s_{n,m}]_{N\times M}$ | the adjacency matrix represent $G_H(X\cup Y, E_H)$ |
| $\mathbf{R} := [r_{i,m}]_{I\times M}$ | the adjacency matrix represent $G_T(Y\cup Z, E_T)$ |
| $\mathbf{C} := [c_{n,i}]_{N\times I}$ | the shortest path matrix, whose element $c_{n,i}$ represents the length of the shortest path from head n to tail i |
| $\mathbf{B} := [b_m]_M$ | the bandwidth vector whose element bm is the required bandwidth of content m |
| $t$ | the bandwidth of a transponder |
| $K$ | the maximum possible number of light-trees to be set up |
| $W$ | the upper bound for $\sum_k d(\mathrm{Q_k})$ |
| $V$ | A very large integer number, is greater than max ($N$, $M$, $I$) |

| | |
|---|---|
| $\mathbf{H} := [h_{n,k}]_{N,K}$ | the matrix represents the relationship between the head and the light-trees, whose element $h_{n,k}$ is 1 if head n is the root of light-tree $k$, otherwise 0; |
| $\mathbf{F} := [f_{k,m}]_{K,M}$ | the matrix denotes the relationship between the multicasting contents and the light-trees, whose element $f_{k,m}$ is 1 if content m is aggregated in light-tree $k$, otherwise 0; |
| $\mathbf{L} := [l_{n,i,k,m}]_{N,I,K,M}$ | the matrix represents the relationship between the set P, Q and the light-trees, whose element $l_{n,i,k,m}$ is 1 if content m is aggregated in light-tree $k$ which originates at head n and terminates at tail $i$, otherwise 0; |
| $\mathbf{A} := [a_{n,i,k}]_{N,I,K}$ | the matrix indicates the relationship between the set P and the light-trees, whose element $a_{n,i,k}$ is 1 if the light-tree $k$ originates at head n and terminates at tail $i$, otherwise 0. |

## III. ILP Formulation

In this section, we present the ILP formulation for the MFA problem. First, we define some notations and variables.

- *Input parameters:*

The multicast receiving index $\alpha$ is defined in equation (2), which measures the multicasting degree of the receiving matrix:

$$\alpha = \begin{cases} \frac{\|\mathbf{RB}\|/\|\mathbf{B}\|-1}{I-1} & I > 1, \\ 1 & I = 1. \end{cases} \tag{2}$$

where $\|\mathbf{RB}\| = \sum_{i=1}^{I}\sum_{m=1}^{M} r_{i,m}b_m$ and $\|\mathbf{B}\| = \sum_{m=1}^{M} b_m$. $\alpha$ varies from 0 to 1. If $\alpha$ is 0, the light-tree is a unicast one, while if it equals to 1, the light-tree is a broadcast one. The redundant transmitting index $\beta$ is defined in equation (3), which measures the redundancy of transmitting matrix:

$$\beta = \begin{cases} \frac{\|\mathbf{SB}\|/\|\mathbf{B}\|-1}{N-1} & N > 1, \\ 1 & N = 1. \end{cases} \tag{3}$$

where $\|\mathbf{SB}\| = \sum_{n=1}^{N}\sum_{m=1}^{M} s_{n,m}b_m$. $\beta$ varies from 0 to 1.

- Variables of the ILP:

### *A. Constraints*

We now discuss constrains for the MFA problem.

- Head constraints

$$\frac{\sum_k \sum_i l_{n,i,k,m}}{V} \le s_{n,m} \tag{4}$$

It guarantees the constraint a) in the tri-partite graph model.

- Tail constraints

$$\sum_k \sum_n l_{n,i,k,m} \ge r_{i.m} \tag{5}$$

$$\frac{\sum_n \sum_i l_{n,i,k,m}}{V} \le f_{k,m} \tag{6}$$

$$\sum_n \sum_i l_{n,i,k,m} \ge f_{k,m} \tag{7}$$

$$l_{n,i,k,m_1} - l_{n,i,k,m_2} + 1 \ge \frac{f_{k,m_1} - f_{k,m_2} + 1}{V} \tag{8}$$

Equation (5) ensures that the results will not contradict to the constraints b) given in tri-partite graph modeling. Equations (6)-(8) states how to generate $G_{Tk}$. Equation (6) ensures that $f_{k,m} = 1$ if at least one tail receives content m through the light-tree $k$. If there is no tail receiving content m through such light-tree, Equation ( 7) ensures that such content is not aggregated in light-tree $k$.

We now explain equaiton (8) in detail with the following three cases:

1) $f_{k,m_1} = f_{k,m_2} = 0$: in this case, neither $m_1$ nor $m_2$ is aggregated into light-tree $k$. We will find the left hand side of Equation (8) is 1, and the right hand side is $1/V$, so the constraints are always satisfied.
2) $f_{k,m_1} = f_{k,m_2} = 1$: in this case, both $m_1$ and $m_2$ are aggregated into the light-tree $k$. The constraint given by Equation (8) guarantee that if tail $i$ receives one content, it should receives the other; otherwise tail $i$ receives neither of them.
3) $f_{k,m_1} = 1$, $f_{k,m_2} = 0$ or $f_{k,m_1} = 0$, $f_{k,m_2} = 1$: in this case, only one content is aggregated into the light-tree. $m_1$ and $m_2$ can be chosen from 1 to $M$ arbitrarily. Consider that $m_1$ is aggregated into the tree and then exchange the position of $m_1$ and $m_2$ that is just the second situation in this case. If we change the position of $m_1$ and $m_2$, i.e., only $m_2$ is aggregated, now $l_{n,i,k,m_2}$ can take either 0 or 1 while $l_{n,i,k,m_1} = 0$. In order to keep the unequal relationship, we add 1 and $1/V$ to the left hand side and the right hand side, respectively.

- Bandwidth constraints

$$\sum_m l_{n,i,k,m}b_m \le t \tag{9}$$

It guarantees the constraint c) in the tri-partroot-leaves setite graph model.

- Tree constraints

$$\frac{\sum_m l_{n,i,k,m}}{V} \le a_{n,i,k} \tag{10}$$

$$\sum_m l_{n,i,k,m} \ge a_{n,i,k} \tag{11}$$

$$\frac{\sum_i a_{n,i,k}}{V} \le h_{n,k} \tag{12}$$

$$\sum_i a_{n,i,k} \ge h_{n,k} \tag{13}$$

$$\sum_n h_{n,k} \le 1 \tag{14}$$

As a tree, we have to consider some constraints to guarantee the content-tree relationship, leaf-tree relationship and root-tree relationship, respectively.

1) Content-tree relationship: The contents must be aggregated into the corresponding existing trees. Equation (10) ensures that $a_{n,i,k}$ should be 1 if there exists one content to deliver. If no content will be delivered from the heads to the tails through the tree, Equation (11) ensures that the tree will not be set up.
2) Leaf-tree relationship: The tail must be a leaf of the corresponding existing tree. Equation (12) guarantees that when there exists a leaf of the tree that is $a_{n,i,k} > 1$, the head should set up such tree $h_{n,k} = 1$. On the other hand, if $h_{n,k} = 0$, no tail should add to the light-tree $k$ since head $n$ will not set up the tree. This is guaranteed by Equation (13).
3) Root-tree relationship: One tree must only have a single root since we consider a point-to-multipoint unidirectional connection. Equation (14) meets such condition by searching all the possible heads.

### B. Objective of ILP formulation

We can use the unified criteria of bandwidth to formulate objectives (d), (e) and (f)

The head wastage is

$$W_h = t\sum_n\sum_k h_{n,k} - \sum_k\sum_m f_{k,m}b_m \tag{15}$$

The tail wastage is

$$W_t = \sum_i\left(\sum_m\left(\sum_k\sum_n l_{n,i,k,m} - r_{i,m}\right)\times b_m\right) \tag{16}$$

The total path bandwidth for all the light-trees is

$$W_r = t\sum_n\sum_i\sum_k c_{n,i}\times a_{n,i,k} \tag{17}$$

There are many approaches [32] to solving multi-criteria problem and here we use an easy one - weighted sum model. The overall objective of the MFA problem is to minimize

$$\alpha_1\times W_h + \alpha_2\times W_t + \alpha_3\times W_r \tag{18}$$

As we discussed before, the overall objective of the MFA problem has to reflect minimization of the resource overhead. However, the topology information of optical network is not known by upper layer under the overlay model. So the objective function 18 should be a linear approximation of the resource overhead. We consider the tri-partite graph. When $G_H$ and $G_R$ are dense, many choices for $P_k$ should be considered for $Q_k$. In this case, $W_r$ will affect the objective more than the other two. When $G_T$ is sparse, $W_t$ will affect the objective more than the other two. As for $W_h$, it is just related to $Q_k$. when $\alpha$ is larger and $\beta$is smaller, $W_h$ will dominate the objective.

The following assignments capture such ideas:

$$\begin{aligned}\alpha_3 &= \beta\\ \alpha_2 &= 1-\alpha\\ \alpha_1 &= 1+\alpha-\beta\end{aligned}$$

## IV. A Heuristic Approach

While the ILP formulation is useful in providing insights into the nature of the problem, it may be hard to solve for large networks with hundreds of contents because of the NP-completeness of the original problem. In this section, we propose a heuristic algorithm, named least trees first (LTF), for large-scale problem. The heuristic has three phases.

Phase 1 chooses the first content for the empty new light-tree, and we use three rules for preferable content selection:

- *Rule 1.* Least number of heads contained first, which means that we choose the content that can be sent from the least head ends.
- *Rule 2.* Least number of tails received first, which means that we choose the content that minimum tails required to receive.
- *Rule 3.* Shortest path first, which means that we calculate the shortest path for all possible content sending from one head to all requested tails and then choose the shortest one.

Algorithm 1 shows how we use these three rules. It ensures that we choose the content most likely to aggregate with other contents, that is to say, phase 1 meets the demand of first criterion (d).

Phase 2 finds the candidate content to be aggregated into the tree. Note that phase 2 will do recursively. Algorithm 2 shows how we find such candidate. In every step, algorithm 2 chooses the content that minimizes the edge difference, so phase 2 guarantees the second criterion (e).

Phase 3 optimally aggregates candidate content into the light-tree. There are two cases for this aggregation:

- *Case 1* (total aggregation): if $d(Q_k)$ is smaller than $|Q_k|\cdot|Z_\mathrm{k}|$, then we just aggregate this contents. Otherwise,
- *Case 2* (partial aggregation): in this case, we will recursively remove the tail $z$ in $Z_y$ which does not receive $Q_k$.

Algorithm 3 will give more details about aggregation with these two cases. We consider the routing has the effect on our problem, so when we decrease the number of tails which only receive content $y_n$, we should also know that those tails will receive $y_n$ from some other trees. Therefore, Phase 3 guarantees the third criterion (f).

### A. LTF algorithm

Based on the three phases discussed above, we can get our LTF algorithm. Before giving the algorithm, it is worth to pointing out why we name this algorithm least trees first (LTF). Note that this algorithm tries to combine more contents into one tree. By doing so, the head wastage will be least. That is to say, the total number of light-trees finally set up in the network will be least. Now we propose the whole LTF algorithm as follows.

### B. Complexity of the heuristic

For step one, we need $O(|Y||X|)$ time to check Rule 1, $O(|Y||Z|)$ time to check Rule 2 and $O(|X||Y||Z|)$ time to check Rule 3, so at the worst case, we need $O(|X||Y||Z|)$ time for step 1. As for the second step, we need $O(|Y||Z|)$

**Algorithm 1** Find the first content for the light-tree $k$

```
Input: a tri-partite graph $G(X \cup Y \cup Z, E_H \cup E_T \cup E_R)$ that are represent by $S$, $R$ and $C$. All reminder content set $A$, $P_k = \emptyset, Q_k = \emptyset$.
Output: the combination of $P_k$ and $Q_k$ where there is one element in $Q_k$.
Begin
for each content is in set A do
    Calculate the number of heads contains such content in S
end for
Choose the content with minimum number of heads to join the set $Q_k$
if ($|Q_k| = 1$) then
    Call this content $y_m$, $Q_k = \{y_m\}$, $Z_k$ contains all tails which receive $Q_k$, $x_k$ is the head which contains $Q_k$ and have total shortest paths form $x_k$ to $Z_k$
else
    for (contents in $Q_k$) do
        Calculate the number of tails receive such content
    end for
end if
Drop all contents from $Q_k$ and choose the content with minimum number of tails to join the set $Q_k$
if ($|Q_k| = 1$) then
    Call this content $y_m$, $Q_k = \{y_m\}$, $Z_k$ contains all tails which receive $Q_k$, $x_k$ is the head which contains $Q_k$ and have total shortest paths form $x_k$ to $Z_k$
else
    for (contents in $Q_k$) do
        Calculate shortest path for transmitting such content from one head to all tails request it
    end for
    Choose one content with minimal value from the set Qk and drop the others, call this content $y_m$. $Z_k$ contains all tails which receive $Q_k$, $x_k$ is the head which contains Qk and have total shortest paths form $x_k$ to $Z_k$
end if
$A = A - Q_k$
End
```

time. It will take $O(|Z|)$ time when we consider step three. When we consider the complexity of the total algorithm, there are two circles in this algorithm. So the total complexity is $O(|Y|(|X||Y||Z| + |Y|(|Y||Z| + |Z|)))$ that is equal to $O(|X||Z||Y|^2 + |Z||Y|^3)$.

## V. Numerical Results and Discussion

In this section, we present some numerical examples to show that our ILP model can solve the small-scale problem very well while our heuristic algorithm can achieve good performance dealing with large-scale network with hundreds of contents. The ILP model was solved using CPLEX 7.0 [33].

### A. *Effects of the transponder bandwidth*

We first examine the effects of the different transponder bandwidth $t$. Let $N = 3$, $M = 16$ and $I = 5$, and matrices

**Algorithm 2** Find the Candidate Content for the Light-tree $k$

```
Input: a tri-partite graph $G(X \cup Y \cup Z, E_H \cup E_T \cup E_R)$ that are represent by $S$, $R$ and $C$, bandwidth vector $B$ and effective bandwidth of the transponder $t$. The reminder content set $A$, $Q_k$ and $P_k$
Output: candidate content $y_n$, $Z_y$, and $d(Q_k \cup y_n)$
Begin
$Y_c = \emptyset$
for all $y$ is in set A do
    if $\sum_{y_m \in Q_k} b_m + b_y \le t$ and $y \in X_k$ then
        $Y_c = Y_c \cup \{y\}$
    end if
end for
for all $y$ is in set $Y_c$ do
    Calculate $d(Q_k \cup y) = (|Q_c| + 1) \times |Z_k \cup Z_y| - |E(G_T \cap (G_{T_k}) \cup G_y)|$, $G_y$ is the subgraph of $G_T$ with all edges from $y$ to $Z_y$
end for
Choose one content with the minimal $d(Q_k \cup y)$, called as $y_n$, as the candidate content.
End
```

TABLE III

Effects of the Transponder Bandwidth

| $\mathbf{t}$(Mb/s) | $\mathbf{K_O}$ | $\mathbf{R_H}$ | $\mathbf{W_t}$(Mb/s) | $\mathbf{P_B}$ |
|---|---|---|---|---|
| 100 | 7 | 0.69 | 100 | 5.75 |
| 155 | 5 | 0.78 | 180 | 8.0 |
| 622 | 5 | 0.93 | 180 | 8.0 |
| 1000 | 5 | 0.95 | 180 | 8.0 |

$\mathbf{S}, \mathbf{R}, \mathbf{B}$ and $\mathbf{C}$ are given as follows:

$$\mathbf{S} = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 0 & 0 & 1 & 0 & 1 & 1 & 0 \\ 0 & 1 & 1 & 0 & 1 & 1 & 1 & 1 & 1 & 1 & 0 & 0 & 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 & 0 & 1 & 1 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

$$\mathbf{R} = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 0 & 1 & 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 1 & 1 & 1 & 0 & 1 & 1 & 1 & 1 & 1 & 0 & 1 & 1 & 1 \\ 0 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 0 & 0 & 1 & 1 & 1 & 1 & 0 & 0 \end{bmatrix}$$

$$\mathbf{B} = \begin{bmatrix} 6 & 25 6 & 6 & 6 & 25 25 6 & 25 25 6 & 6 & 6 & 25 6 & 6 \end{bmatrix}$$

$$\mathbf{C} = \begin{bmatrix} 2 & 2 & 4 & 4 & 4 \\ 5 & 3 & 3 & 3 & 2 \\ 1 & 4 & 2 & 3 & 2 \end{bmatrix}$$

Table III shows the results of the optimal number of light-trees $K_O$, the average head wastage ratio $R_H$, the tail wastage $W_t$and the average sum of path bandwidth $P_B$ for all the light-trees against the transponder bandwidth $t$ = 100Mb/s, 155Mb/s, 622Mb/s and 1000Mb/s, respectively. The average head wastage ratio is defined as the total head wastage over the product of $K_O$ and $t$. The average sum of path bandwidth ratio $P_B$ is defined as the total sum of path bandwidth for all light-trees over the product of $K_O$ and $t$. Obviously, the comparison of the average head wastage ratio and the average sum of path bandwidth ratio is more valid than that of the head wastage and the sum of path bandwidth. The results of $K_O$ are straight-forward since more multicast IP

**Algorithm 3** Optimally Aggregate Candidate Content into the Light-tree $k$

```
Input: a tri-partite graph $G(X \cup Y \cup Z, E_H \cup E_T \cup E_R)$ that are represent by $S$, $R$ and $C$, $Q_k$, $P_k$, $y_n$, $Z_y$ and $d(Q_k \cup y_n)$ which comes from Algorithm 2.
Output: the combination of $P_k$ and $Q_k$
Begin
$j = 0$
if $(d(Q_k) < |Q_k| \times |Z_k|)$ then
    $Q_k = Q_k \cup y_n$, $Z_k = Z_k \cup Z_y$ and $P_k = \{x_k, Z_k\}$;
    $A = A - y_n$
else
    for every $z$ in $Z_y$ which only receives $y_n$ do
        Drop $z$ from $Z_y$
        Calculate
        if $(d(Q_k) \cup y_n) < |Q_k||Z_k|$ and $|Q_k| + 2 > c_{x_k,Z}$ then
            $Q_k = Q_k \cup y_n$, $Z_k = Z_k \cup Z_y$ and $P_k = \{x_k, Z_k\}$
            Break;
        else
            if $Z_y \subseteq Z_k$ then
                Break;
            end if
        end if
    end for
end if
End
```

**Algorithm 4** LTF

```
Input: a tri-partite graph $G(X \cup Y \cup Z, E_H \cup E_T \cup E_R)$ that are represent by $S$, $R$ and $C$, bandwidth vector $B$ and effective bandwidth of the transponder $t$.
Output: the combination of $P$ and $Q$
Begin
$A = Y, k = 1$;
repeat
    $Q_k = \emptyset$;
    Call Algorithm 1 to find the first content for the light-tree $k$
    repeat
        Call Algorithm 2 to find the candidate content
        Call Algorithm 3 to aggregate candidate content
    until $Y_c = \emptyset$;
until $A = \emptyset$;
End
```

flows can be aggregated in one light-tree as the transponder bandwidth increases. Thus the total number of trees decreases. The head wastage ratio partially reflects that the network load decreases as the transponder bandwidth increases because the total bandwidth of the aggregated multicast IP flows increases much slower than that of the transponder. The tail wastage and the average sum of path bandwidth ratio remain unchanged when the transponder bandwidth $t$ is greater than 155Mb/s, which shows that the aggregation of multicast IP flows for each light-tree are the same. This is because the bandwidth of the transponder is greater than the total bandwidth needed by the multicast IP flows in every head. Consequently the aggregation of multicast IP flows and the trees to deliver such aggregated multicast IP flows are the same. The head wastage increases with the bandwidth of the transponder.

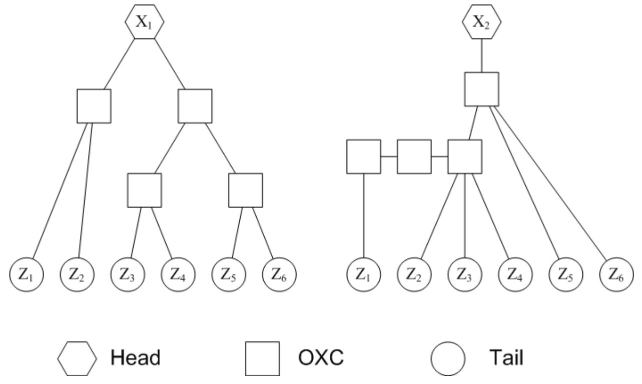


Fig. 5. Two specific multicasting trees from different two heads to all tails

### B. Verification of the heuristic

In this subsection, we will check the different combination of the parameter $\alpha$ and $\beta$ on the performance. By comparing the results obtained by the ILP model and the heuristic algorithm with the same inputs, we can verify the effectiveness of the heuristic.

Let $N = 2$, $M = 6$, and $I = 6$. The bit rate of transponder is 155Mb/s. We consider 3 typical types of contents whose encoded bandwidth are 2, 6, 25 Mb/s respectively. The content bandwidth vector $B$ is given as follows:

$$\mathbf{B} = \begin{bmatrix} 2 & 25 & 6 & 25 & 25 & 25 \end{bmatrix}$$

Fig. 5 shows the broadcast shortest path trees from two heads to all the tails, respectively. For simplicity, the length between any two neighboring nodes is assumed to be unity. Then we can set up the shortest path matrix $\mathbf{C}$ as follows:

$$\mathbf{C} = \begin{bmatrix} 2 & 2 & 3 & 3 & 3 & 3 \\ 5 & 3 & 3 & 3 & 2 & 2 \end{bmatrix}$$

We randomly generate the transmission matrix $\mathbf{S}$ and the receiving matrix $\mathbf{R}$ for different values of the parameter $\alpha$ and $\beta$ .

Table IV shows the number of light trees, the head wastage, the tail wastage obtained by the ILP model and the heuristic with the same inputs. The last columns of Table III give the objective function 18. In the table, we also compare the sum of path bandwidth for all the light-trees $W_r$ and the actual used link bandwidth $B_t$ of light-trees along with the given broadcast shortest path trees.

From the results in Table IV gained by ILP model, we can find that as $\alpha$ increases from 0 to 1, the average of head wastage $W_h$ decreases from 720 Mb/s to 150 Mb/s, the difference is nearly equal to the total bandwidth of 4 light-trees. At the same time, the optimal number of trees in the network also decreases as the same pace as the head wastage when $\alpha$ increases from 0 to 1. Secondly, the average tail wastage $W_t$ varies from 0 Mb/s at $\alpha$ = 1 to 184 Mb/s at $\alpha$ = 0.5. When $\alpha$= 0, the tails receive IP multicast flows quite different from each other, so the heads should deliver contents individually to the tails, resulting in the reduction in the tail wastage. When $\alpha$= 1, the tails receive almost all the multicast IP flows, so the heads should first guarantee that all required

TABLE IV

RESULTS OBTAINED BY THE ILP MODEL (I) AND THE HEURISTIC ALGORITHM (H)

| | | $K_O$ | | Wh(Mb/s) | | Wt(Mb/s)/d(Q) | | Wr(Mb/s)/Bt(Mb/s) | | Objective (Mb/s) | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| $\alpha$ | $\beta$ | I | H | I | H | I | H | I | H | I | H |
| 0 | 0 | 4 | 4 | 512 | 512 | 100/4 | 154/8 | 2480/2170 | 2480/2170 | 100 | 154 |
| 0 | 0.5 | 6 | 4 | 822 | 512 | 0/0 | 139/8 | 2170/2170 | 2325/2325 | 1446 | 1457.5 |
| 0 | 1 | 6 | 4 | 822 | 512 | 0/0 | 81/4 | 2170/2170 | 2790/2170 | 2170 | 2871 |
| 0.5 | 0 | 3 | 3 | 357 | 357 | 141/9 | 222/13 | 6510/3565 | 5115/3565 | 601 | 646.5 |
| 0.5 | 0.5 | 3 | 3 | 357 | 357 | 175/7 | 241/13 | 4650/2945 | 5115/3565 | 2769.5 | 2880 |
| 0.5 | 1 | 4 | 2 | 512 | 202 | 175/7 | 254/12 | 2170/1860 | 2790/2170 | 2436 | 3118 |
| 1 | 0 | 2 | 2 | 202 | 202 | 0/0 | 0/0 | 5890/3100 | 5890/3100 | 404 | 404 |
| 1 | 0.5 | 2 | 2 | 202 | 202 | 0/0 | 0/0 | 5890/3100 | 5890/3100 | 3248 | 3248 |
| 1 | 1 | 1 | 1 | 47 | 47 | 0/0 | 0/0 | 2170/1550 | 2170/1550 | 2217 | 2217 |

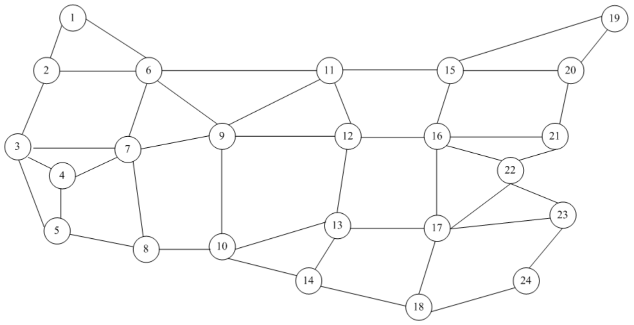


Fig. 6. 24-node mesh network

contents are delivered to the tails by aggregating and then sending them to the tails. When $\alpha$= 0.5, there is counterbalance between aggregating the contents and sending one content per tree, and consequently the average tail wastage is the largest at $\alpha$= 0.5. Thirdly, we can find that, for a given $\alpha$, the average $W_r$ decreases as $\beta$ increases, which partially reflects the actual resource consumed by the light-trees in optical network.

As for the results obtained by the heuristic algorithm in Table III, we can find they have the similar trends on $W_t$ and $W_r$ comparing with that done by the ILP model. The major difference between them is the number of trees and $W_h$ that is just because we try to combine as many contents into one tree as possible in LFT algorithm. When $\alpha$ is equal to 1, the ILP model will aggregate the contents as the same manner. That is why we obtain the same results at that point.

### *C. Large-scale design using the heuristic*

We adopt a 24-node mesh network, shown in Fig. 6, for the network-level simulation in which the maximum hop-distance is 6. We randomly choose 5 nodes and 10 nodes as the heads and tails respectively. 500 contents will be delivered from heads to tails. Among these contents, 60% require 25Mb/s, 30% require 6Mb/s and 2Mb/s for the others. We randomly generate the transmission matrix S and the receiving matrix R for $\alpha = 0.5$ and $\beta = 0.5$. For simplicity, we just assign length 1 to each link in the graph, that is to say, the shortest path is just based on the hop counts. We use two common bandwidth values 1Gb/s and 2.5Gb/s as t for the light-tree and do each experiment three times.

Table V shows the results obtained by the heuristic algorithm. In the network without the MFA strategy, it requires at least 500 multicast IP sessions for 500 contents. By introducing the MFA strategy, we can find that the number of light-trees is significantly reduced. It is consistent with the conclusion drawn by ILP that the number of light-trees decreases as transponder bandwidth increases. In addition, we find that content distribution and user requests have a strong impact on the results even for the same $\alpha$and $\beta$ . For example, when $t$ = $2.5G$b/s, the maximum number of tree is 17 that is almost twice of the minimum number of trees.

We also compare the resource utilization of multicast IP flow without the MFA strategy and light-tree the MFA strategy. The overhead ratio is defined as the difference of resource utilization over the resource utilization of light-tree. It does not change greatly over the experiments with the same $\alpha$ and $\beta$ . Note that, the transponder bandwidth does not affect the resource utilization of multicast IP flow. However, the resource utilization of light-tree increases as the transponder bandwidth increases. Consequently, the overhead ratio also increases. In the practical scenario, we have to choose proper transponder bandwidth to balance the number of light-trees and the average overhead ratio.

## VI. CONCLUSIONS

We proposed and investigated the multicast traffic aggregation (MFA) problem in IP over optical networks which is NP-complete and can be modeled by a tri-partite graph with multiple criteria. We have also developed an integer linear programming (ILP) model and a simple approach to dealing with multiple criteria. Although ILP model can do solve the small-scale problem perfectly, we propose the heuristic algorithm for the large-scale network with hundreds of contents. We have carried out numerical studies to verify the effectiveness of our model and heuristic algorithm. The experimental results show that 1) as the objective ruled, the model meets the main criteria very well at the extreme cases; 2) as the bandwidth of a transponder increases, the heads intend to aggregate more multicast IP flows into one light-tree while the tails still want to receive fewer unwanted flows, and there is a tradeoff between them; 3) heuristic algorithm gets the close results to the ILP model and can solve the large scale problem delivering 500 contents in the 24-node mesh network.

In our research, we use a linear approximation to deal with multicast routing issue in optical network and a simple weighted-sum model for multiple criteria optimization since we consider our IP over Optical networks is based on the overlay model. However, more work needs to be done such as taking the network topology into consideration under the

TABLE V
LARGE-SCALE NETWORK RESULTS WITH $\alpha = 0.5$ AND $\beta = 0.5$

| Transponder Bandwidth (Mb/s) | | 1000 | | | 2500 | | |
|---|---|---|---|---|---|---|---|
| Experiment | | 1 | 2 | 3 | 1 | 2 | 3 |
| Number of light-trees | | 23 | 20 | 26 | 17 | 9 | 14 |
| Resource Utilization (Mb/s) | IP | 199550 | 178500 | 212500 | 195500 | 173500 | 184500 |
| | LT | 501700 | 463525 | 463700 | 880250 | 573250 | 719075 |
| Overhead Ratio = (LT-IP) / LT | | 60.2% | 61.5% | 54.2% | 77.8% | 69.7% | 74.3% |
| Average Overhead Ratio | | 58.6% | | | 73.9% | | |

peer model, and other approaches to solving multiple criteria optimization.

## ACKNOWLEDGEMENT

We thank anonymous reviewers for their valuable comments and all the colleagues in the 3TNet project. Yi Zhu thanks Guobiao Zhou, Xiaodong Huang, and Jason Jue for their helpful discussion. We are grateful to Guowu Xie for help in the preparation of this manuscript.

## APPENDIX

The MFA problem can be divided into three sub-problems based on the distribution of the multicasting contents.

- Partition distribution sub-problem (MFA-P, $\beta = 0$): In this case, one multicast content is only provided by a single head, thus the node degree of $Y$ in $G_H$ is equal to 1;
- Mirroring distribution sub-problem (MFA-M, $\beta = 1$): In this case, any multicast content are provided by all the head end, thus $G_H$ is a complete bipartite sub-graph;
- Partially redundant distribution sub-problem (MFA-R, $0 < \beta < 1$): In this case, the node degree of Y in $G_H$ is not less than 1 while $G_H$ is a proper graph of the complete bipartite graph.

In order to prove the total MFA problem is NP-complete, we first prove the MFA-1, which refers to MFA problem with a single head, is NP-complete.

*Lemma 1:* MFA-1 is NP-Completeness.

*Proof:* Given one partition for the MFA-1 problem, we first check whether this partition satisfies with the contraints $\forall k = \{1, 2, \ldots, K\}$, $\sum_{y_m \in Q_k} b_m \leq t$ and $d(Q_k) \leq W$. It is easily to find that the check can be done in polynomial time.

We now show that the well-known BIN-PACKING problem is the special case of the MFA-1 problem. For simplicity, we set B as the all-1 vector. The decision problem for bin-packing can be stated as follows [34]:

*Given a set $E = e_1, e_2, \ldots, e_M$ and an positive integer K and value t, find the partition of the set $E' = E_1, E_2, \ldots, E_K$, such that $\sum_{e_j \in E_k} e_j \leq t$, $\forall j = 1, 2, \ldots, M. \forall k = 1, 2, \ldots, K$*

Obviously, if we let $E = Y$ and $G_T$ is a complete bipartite graph, bin-packing problem is the special case for MFA-1. ■

It is worth to point that the MFA-P, the MFA-M, and the MFA-R problems are degenerated into the MFA-1 problem for the single head. However, when we consider multiple heads, it is more complicated than MFA-1.

*Theorem 1:* The MFA problem is NP-complete.

*Proof:* In the case of the MFA-P problem, note that the content sets provided by any two heads are disjoint since $\beta = 0$, so the MFA-P problem with N heads is equivalent to N MFA-1 problems. In the case of the MFA-M problem, all contents can be provided by all heads since $\beta = 1$, MFA-M with N heads is equivalent to $N!$ MFA-1 problems. The complexity of MFA-R is between MFA-P and MFA-M that depends on $\beta$. Because the MFA-1 problem is NP-complete, the MFA problem is also NP-complete. ■

## REFERENCES

[1] P. Van Mieghem, G. Hooghiemstra, and R. van der Hofstad, "On the efficiency of multicast," *IEEE/ACM Trans. Networking (TON)*, vol. 9, no. 6, pp. 719–732, 2001.

[2] C. Diot, W. Dabbous, and J. Crowcroft, "Multipoint Communication: A Survey of Protocols, Functions, and Mechanisms," *IEEE J. Select. Areas Commun.*, vol. 15, no. 3, p. 277, 1997.

[3] C. Diot, B. Levine, B. Lyles, H. Kassem, and D. Balensiefen, "Deployment issues for the IP multicast service and architecture," *IEEE Network*, vol. 14, no. 1, pp. 78–88, 2000.

[4] B. Wang and J. Hou, "Multicast routing and its QoS extension: Problems, algorithms, and protocols," *IEEE Network*, vol. 14, no. 1, pp. 22–36, 2000.

[5] A. Striegel and G. Manimaran, "A survey of QoS multicasting issues," *IEEE Commun. Mag.*, vol. 40, no. 6, pp. 82–87, 2002.

[6] L. Sahasrabuddhe and B. Mukherjee, "Light trees: optical multicasting for improved performance inwavelength routed networks," *IEEE Commun.*, vol. 37, no. 2, pp. 67–73, 1999.

[7] G. Rouskas, "Optical layer multicast: rationale, building blocks, and challenges," *IEEE Network*, vol. 17, no. 1, pp. 60–65, 2003.

[8] W. Hu and Q. Zeng, "Multicasting Optical Cross Connects Employing Splitter-and-Delivery Switch," *IEEE Photon. Technol. Lett.*, vol. 10, no. 7, 1998.

[9] S. Yu, S. Lee, O. Ansell, and R. Varrazza, "Lossless Optical Packet Multicast Using Active Vertical Coupler Based Optical Crosspoint Switch Matrix," *J. Lightwave Technol.*, vol. 23, no. 10, pp. 2984–2992, 2005.

[10] M. Ali and J. Deogun, "Power-efficient design of multicast wavelength-routed networks," *IEEE J. Select. Areas Commun.*, vol. 18, no. 10, pp. 1852–1862, 2000.

[11] B. Chen and J. Wang, "Efficient routing and wavelength assignment for multicast in WDMnetworks," *IEEE J. Select. Areas Commun.*, vol. 20, no. 1, pp. 97–109, 2002.

[12] A. Banerjee, J. Drake, J. Lang, B. Turner, K. Kompella, and Y. Rekhter, "Generalized multiprotocol label switching: an overview of routingand management enhancements," *IEEE Commun. Mag.*, vol. 39, no. 1, pp. 144–150, 2001.

[13] S. Yasukawa, K. Sugisono, I. Inoue, and S. Urushidani, "Scalable multicast MPLS protocol for next generation broadband service convergence network," *2004 IEEE International Conference on Communications*, vol. 2, 2004.

[14] X. Wei, Y. Jin, G. Zhang, W. Sun, J. Sun, W. Guo, and W. Hu, "Demonstration of GMPLS-controlled dynamic point-to-multipoint trees in optical networks," *Optical Communication, 2005. ECOC 2005. 31st European Conference on*, pp. 29–30, 2005.

[15] W. Sun, Y. Jin, W. Hu, H. He, H. Luo, X., W. P., Guo, Y. Su, and L. Leng, "Prototype of demonstration of IP multicasting over optical network with dynamic point-to-multipoint configuration," *IEEE/OSA OFC 2005, OWG3*, 2005.

[16] Y. Zhu, Y. Jin, W. Sun, W. Guo, and W. Hu, "On topology-independent IP group aggregation in multicast capable optical networks," *Globecom'05*, 2005.

[17] N. Singhal and B. Mukherjee, "Protecting multicast sessions in WDM optical mesh networks," *J. Lightwave Technol.*, vol. 21, no. 4, pp. 884–892, 2003.

[18] N. Singhal, L. Sahasrabuddhe, and B. Mukherjee, "Provisioning of survivable multicast sessions against single link failures in optical WDM mesh networks," *J. Lightwave Technol.*, vol. 21, no. 11, pp. 2587–2594, 2003.

[19] B. Haskell, A. Netravali, and A. Puri, *Digital Video: An Introduction to Mpeg-2*. Springer, 1996.

[20] J. Wu and Y. Jin, "China High Performance Broadband Information Network (3TNET)," *APOC05, Shanghai, China*, 2005.

[21] G. Bernstein, J. Yates, and D. Saha, "IP-centric control and management of optical transport networks," *IEEE Commun.*, vol. 38, no. 10, pp. 161–167, 2000.

[22] J. Cui, M. Faloutsos, and M. Gerla, "An architecture for scalable, efficient, and fast fault-tolerant multicast provisioning," *IEEE Network*, vol. 18, no. 2, pp. 26–34, 2004.

[23] J. Cui, J. Kim, D. Maggiorini, K. Boussetta, and M. Gerla, "Aggregated Multicast–A Comparative Study," *Cluster Computing*, vol. 8, no. 1, pp. 15–26, 2005.

[24] A. Fei, Z. Duan, and M. Gerla, "Constructing shared-tree for group multicast with QoS constraints," *Global Telecommunications Conference, 2001. GLOBECOM'01. IEEE*, vol. 4, 2001.

[25] K. Zhu and B. Mukherjee, "A review of traffic grooming in WDM optical networks: Architectures and challenges," *Optical Networks Magazine*, vol. 4, no. 2, pp. 55–64, 2003.

[26] N. Singhal, L. Sahasrabuddhe, and B. Mukherjee, "Optimal Multicasting of Multiple Light-Trees of Different Bandwidth Granularities in a WDM Mesh Network With Sparse Splitting Capabilities," *IEEE/ACM Trans. Networking*, vol. 14, no. 5, pp. 1104–1117, 2006.

[27] A. Billah, B. Wang, and A. Awwal, "Multicast traffic grooming in WDM optical mesh networks," *Global Telecommunications Conference, 2003. GLOBECOM'03. IEEE*, vol. 5, 2003.

[28] H. Madhyastha, G. Chowdhary, N. Srinivas, and C. Siva Ram Murthy, "Grooming of multicast sessions in metropolitan WDM ring networks," *Computer Networks*, vol. 49, no. 4, pp. 561–579, 2005.

[29] R. Ul-Mustafa and A. Kamal, "Design and provisioning of wdm networks with multicast traffic grooming," *IEEE J. Select. Areas Commun.*, vol. 24, no. 4, pp. 37–53, 2006.

[30] X. Huang, F. Farahmand, and J. Jue, "Multicast Traffic Grooming in Wavelength-Routed WDM Mesh Networks Using Dynamically Changing Light-Trees," *J. Lightwave Technol.*, vol. 23, no. 10, pp. 3178–3187, 2005.

[31] A. Khalil, A. Hadjiantonis, G. Ellinas, and M. Ali, "Sequential and hybrid grooming approaches for multicast traffic in WDM networks," *Global Telecommunications Conference, 2004. GLOBECOM'04. IEEE*, vol. 3.

[32] R. Steuer, "Multiple Criteria Optimization: Theory, Computation and Application," 1985.

[33] [Online]. Available: http://www.ilog.com/products/cplex/

[34] [Online]. Available: http://www.nist.gov/dads/html/binpacking.html

**Yi Zhu** received the B.S and M.S degree in electricity engineering from Shanghai Jiaotong University, China, in 2003 and 2006, respectively. He is currently working toward the Ph.D. degree in computer science at the University of Texas, Dallas. His research is on multicast aggregation, traffic grooming, and complexity of optical network.

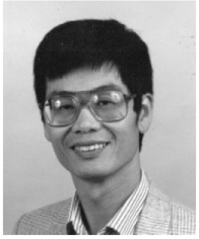
**Yaohui Jin** is a professor in the State Key Laboratory of Advanced Optical Communication Systems and Network, Shanghai Jiao Tong University, China. Prior to joining SJTU, he was a member of technical staff at Bell Labs Research China from 2000 to 2002. He served as the TPC member in many international conferences. He published more than 50 paper in technical journals and conferences. His research interests include optical networking, optical grid and switch scheduling.

**Weiqiang Sun** is currently a Lecturer at the State Key Laboratory on Fiber-Optic Local Area Networks and Advanced Optical Communication Systems, Shanghai Jiao Tong University. His research interests include Automatically Switched Optical Networks (ASON), optical multicast and TV distribution in overlay networks.

**Guo Wei** is an associate professor of state key lab of advanced optical communication system and network in Shanghai Jiao Tong University since 2003. Before she entered in SJTU, she was a senior engineer and a project manager of the Fiberhome Telecommunication Technologies CO., LTD from 2001-2003. She has over 50 publications published in technical journals and conferences. Her research interests include optical grid, network planning, and optimization algorithm.

**Weisheng Hu** is the Professor and Director of the State Key Laboratory on Fiber-Optic Local Area Networks and Advanced Optical Communication Systems, Shanghai Jiao Tong University. His interests are on generalized automatic switched optical network, and optical packet switching. He is the author or co-author of over 100 journal and conference papers.

**Wen-De Zhong** is an associate professor with School of Electrical and Electronic Engineering, Nanyang Technological University (NTU), Singapore. He received his Ph.D degree from the University of Electro-Communications, Tokyo in 1993. He has published more than 100 refereed journal and conference papers and has given several invited presentations at international conferences. He has served on organizing and/or TPC for numerous international conferences. His research interests include optical WDM systems and networks.



**Min-You Wu** is an IBM Chair Professor in the Department of Computer Science and Engineering at Shanghai Jiao Tong University. He serves as the Chief Scientist at Grid Center of Shanghai JiaoTong University. He is a research professor of the University of New Mexico, USA. His research interests include grid computing, wireless networks, sensor networks, overlay networks, multimedia networking, parallel and distributed systems, and compilers for parallel computers. He has published over 140 journal and conference papers in the above areas.