On the stability of multicast flow aggregation in IP over optical network for IPTV delivery

Xuan Luo (罗 萱), Yaohui Jin (金耀辉), Qingji Zeng (曾庆济),
Weiqiang Sun (孙卫强), Wei Guo (郭 薇), and Weisheng Hu (胡卫生)

State Key Lab of Advanced Optical Communication Systems and Networks, Shanghai Jiao Tong University, Shanghai 200240

Received January 7, 2008

The stable multicast flow aggregation (MFA) problem in internet protocol (IP) over optical network under the dynamical scenario is studied. Given an optical network topology, there is a set of head ends and access routers attached to the optical network, in which each head end can provide a set of programs (IP multicasting flows) and each access router requests a set of programs, we find a set of stable light-trees to accommodate the optimally aggregated multicast IP flows if the requests of access routers changed dynamically. We introduce a program correlation matrix to describe the preference of end users' requests. As the original MFA problem is NP-complete, a heuristic approach, named most correlated program first (MCPF), is presented and compared with the extended least tree first (ELTF) algorithm which is topology-aware. Simulation results show that MCPF can achieve better performance than ELTF in terms of stability with negligible increment of network resource usage.

OCIS codes: 060.4250, 060.4510.
doi: 10.3788/COL20080608.0553.

Internet protocol television (IPTV), a promising revenue generating application in the telcos' triple-play service portfolio, has been trialed and deployed by various operators and companies[1,2]. In the IPTV system, multicasting will play a key role in the delivery of high-quality video services to optimize the utilization of network resources. Since customers are used to viewing television programs without any appreciable jitter or delay, end-to-end quality of service (QoS) control is a must in the video service deployment. However, it is difficult to implement large-scale multicasting with hard QoS guarantees in packet-switched IP network due to its essential best-effort nature. To meet the stringent QoS and availability requirements of multimedia service, a novel hybrid packet/circuit multicasting network model has been proposed[3]. This hierarchical multicasting architecture, with optical multicasting (always implemented by the use of light-tree) in the core and IP multicasting at the edge, can not only greatly simplify QoS provisioning and traffic engineering but also preserve compatibility with the existing IP access networks. In China, 3TNet, a pioneering research initiative adopting the hybrid packet/circuit multicasting for large-scale IPTV-oriented streaming media delivery, has been carried out in Yangtze River Delta[4].

Generally, the bandwidth of one light-tree is much higher than that required by one single IP multicasting flow. Bandwidth efficiency can be remarkably increased by carefully aggregating multiple IP flows into one single light-tree, which is defined as multicast flow aggregation (MFA) problem[5]. The MFA problem, different from multicast grooming problem[6] and aggregated multicast problem[7], has been proven as NP-complete. A heuristic algorithm, named least tree first (LTF), has been proposed to minimize the resource usage for the large-scale network[5]. However, if we use the LTF algorithm in the dynamical scenario, we find that both the combination of the aggregated programs and the routes of light-trees are not stable, which may lead to degradation of quality of experience (QoE) for the end users, namely, large latency of channel zapping or service interruption. In this letter, we study the stable MFA problem under the dynamical scenario, which is formulated as follows: given an optical network topology, there is a set of head ends and access routers attached to the optical network, in which each head end can provide a set of programs (multicast IP flows) and each access router requests a set of programs, we find a set of stable light-trees to accommodate the optimally aggregated multicast IP flows if the requests of access routers are changed. To describe the preference of end user requests, we define a program correlation matrix. Then we develop a heuristics named most correlated program first (MCPF) to achieve the stable light-trees and the stable combinations of the aggregated IP flows as much as possible. Similar to the original MFA problem, optimal network resource utilization is still our objective. Simulation results show that MCPF can achieve better performance in terms of stability with negligible increment of network resource usage.

In hybrid packet/circuit multicasting network, we use a circuited-switched optical transport network, which consists of multicast capable cross-connects, as the core network instead of the traditional packet-switched IP core network. The structure of edge networks such as head ends, access networks as well as end users remains the same as the current IP network. Indeed, the head ends are multicasting sources which encode multimedia contents with IP multicasting flows and then send such flows to the core network. Note that, the same programs may be redundantly provided by several head ends in the network considered in this study, which is similar to the scenario of today's TV delivery networks where one channel may be distributed by several stations. As an example shown in Fig. 1, program p_2 is redundantly

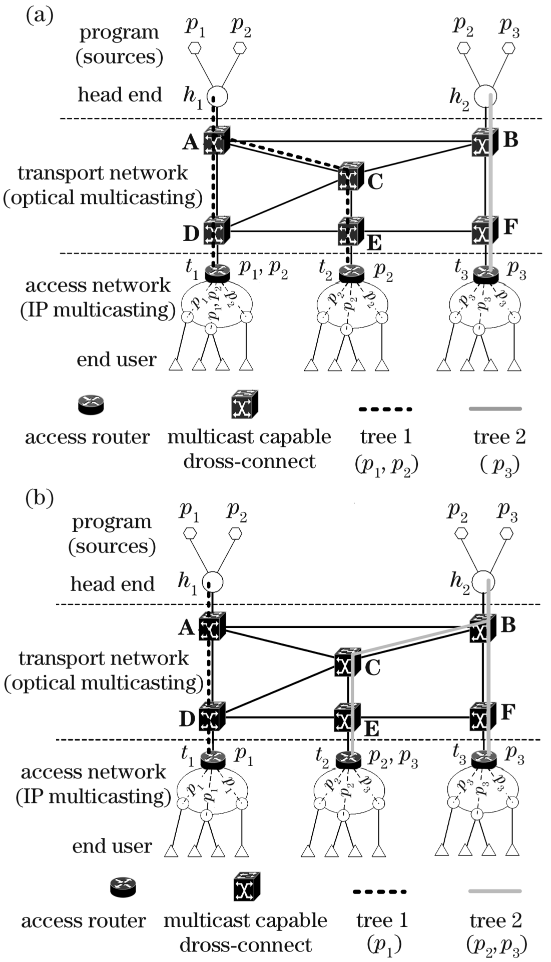

Fig. 1. Multicast flow aggregation in hybrid packet/circuit multicasting network. (a) Initial status; (b) current status.

distributed in head ends h_1 and h_2. In the optical network, several lower bit rate IP multicasting flows can be aggregated into one light-tree so that the bandwidth of a light-tree can be efficiently used. Unfortunately, even with the best carefulness, resource overheads always exist because the total bandwidth of aggregated IP flows may be less than that of one light-tree and access router may receive unwanted channels. In the access network side, access router aggregates a large number of end users' requests originated at one residential area and delivers the received programs through IP multicasting to end users. In the remaining of this paper, for simplicity, we refer to head ends as heads and the access routers as tails.

Figures 1(a) and (b) show the different MFA results when the requests of access router change dynamically. In the initial time, access router t_1 requires program p_1 and p_2, while access router t_2 requires program p_2. After certain time, access router t_1 only requires program p_1 and access router t_2 requires program p_2 and p_3. We can find that both the routes of light-trees and the combinations of aggregated programs have changed. Tables 1 and 2 summarize their differences.

We suppose that the end user requests have preference which could be based on the statistical end users' behaviors. A simple example is that many young men like to watch the news channels would also enjoy sport channels. Meanwhile, most housewives may refuse sport channels but prefer to soap opera channels. Therefore, we define a program correlation matrix $\mathbf{C}$ to describe such relationship. The element $c_{m,n}$ is the preference to request for program n when program m has been requested. Its value could be considered as costs assigned by carrier or content service providers for program aggregation. Table 3 shows an example of program correlation. Note that the program correlation matrix may not be symmetrical.

Table 1. Changes in the Route of the Light-Trees

Light-Tree	Route		Differences
	Initial	Current	
1	{**A**, {**D**, **E**}}	{**A**, {**D**}}	2
2	{**B**, {**F**}}	{**B**, {**E**, **F**}}	2

Table 2. Changes in the Combination of the Aggregated Programs

Light-Tree	Combination		Differences
	Initial	Current	
1	p_1 p_2	p_1	2
2	p_3	p_2 p_3	2

Table 3. An Example of Program Correlation

	News	Sport	Soap Opera
News	0	100	10
Sport	10	0	10
Soap Opera	1	1	0

In this letter, we make the following assumptions: 1) Both multicast IP flows and light-trees are one-to-many unidirectional connections. 2) A multicast IP flow cannot be split into two or more light-trees. 3) Sufficient link resources are available in the optical network, thus the light-trees are not blocked. 4) All transponder's bit rates are the same through the whole optical transport network.

We formulate the stable MFA problem as follows:

Given

a) The topology of optical network $\mathbf{G}(\mathbf{V}, \mathbf{E})$, where a vertex $v_i \in \mathbf{V}$ represents an optical cross-connect (OXC) and an edge $e_{i,j} \in \mathbf{E}$ denotes the transmission link connecting two OXCs v_i and v_j. $N = |\mathbf{V}|$.

b) The transponder bandwidth t.

c) The set of heads $\mathbf{H}$ attached to the optical network, thus $\mathbf{H} \subseteq \mathbf{V} \cdot J = |\mathbf{H}|$.

d) The set of tails $\mathbf{T}$ attached to the optical network, thus $\mathbf{T} \subseteq \mathbf{V} \cdot I = |\mathbf{T}|$.

e) The set of programs $\mathbf{P}$ with their required bandwidth $\mathbf{B} \cdot M = |\mathbf{P}|$.

f) The Boolean transmitting matrix $\mathbf{S}$ representing the relationships between the set $\mathbf{H}$ and $\mathbf{P}$, whose element $s_{j,m}$ is 1 if the source h_j contains the program p_m.

g) The program correlation matrix $\mathbf{C}$, where $c_{m,n}$ is a discrete number representing the program correlation from the program p_m to p_n.

h) The Boolean receiving matrix $\mathbf{R}$ representing the relationships between the set $\mathbf{T}$ and $\mathbf{P}$, whose element $r_{i,m}$ is 1 if the program p_m is requested by the tail t_i.

i) The changed Boolean receiving matrix $\mathbf{R}'$.

Find

$\mathbf{A} = \{\mathbf{A}_1, \mathbf{A}_2, \cdots, \mathbf{A}_k, \cdots\}$ is a result of the MFA problem, where $\mathbf{A}_k$ is a quadruple $(r_k, \mathbf{T}_k, \mathbf{P}_k, \mathbf{E}_k)$ representing one light-tree, in which $r_k \in \mathbf{H}$ is the root, $\mathbf{T}_k \subseteq \mathbf{T}$ is a set of the leaves, $\mathbf{P}_k \subseteq \mathbf{P}$ is a combination set of aggregated programs, and $\mathbf{E}_k \subseteq \mathbf{E}$ is a set of edges representing the route.

We use a matrix $\mathbf{L} = [l_{i,j,m}]_{N\times N\times M}$ to represent the route of each program after aggregation, whose element $l_{i,j,m}$ is 1 if the route of the program p_m includes the link $e_{i,j}$; otherwise 0.

$\mathbf{A}'$ and $\mathbf{L}'$ are the results of the MFA problem by using the same algorithm with the changed receiving matrix $\mathbf{R}'$.

With constraints:

a) All elements in $\mathbf{P}_k$ must be transmitted by the same head.

b) The total bandwidth of a combination in one light-tree must not be greater than the bandwidth of a transponder.

c) One program corresponds to one multicast IP flow.

d) The programs in $\mathbf{P}_k$ are the aggregated multicast flows that tails $\mathbf{T}_k$ intend to receive.

To minimize:

a) The difference between $\mathbf{A}$ and $\mathbf{A}'$.

b) Total network resource usage.

In the following section, we define some parameters for inputs and outputs.

The multicast receiving parameter α is defined as follows:

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

which measures the multicasting degree of the receiving matrix. In Eq. (1), $\|\mathbf{RB}\| = \sum\limits_{i=1}^{I}\sum\limits_{m=1}^{M} r_{i,m}b_m$ and $\|\mathbf{B}\| = \sum\limits_{m=1}^{M} b_m$. α varies from 0 to 1.

The redundant transmitting parameter β is defined as follows:

$$\beta = \begin{cases} \frac{\|\mathbf{SB}\|/\|\mathbf{B}\|-1}{J-1}, & J > 1 \\ 1, & J = 1 \end{cases}, \tag{2}$$

which measures the redundancy of transmitting matrix. In Eq. (2), $\|\mathbf{SB}\| = \sum\limits_{j=1}^{J}\sum\limits_{m=1}^{M} s_{j,m}b_m$. β varies from 0 to 1.

The receiving change parameter Δ is defined as follows:

$$\Delta = \frac{\sum\limits_{i}\sum\limits_{m}\left|r_{i,m} - r'_{i,m}\right|}{I \times M}, \tag{3}$$

which measures the dynamical change of the receiving matrix. Δ varies from 0 to 1.

We introduce a metric γ that is the change ratio of the program delivery routes, which can reflect the two kinds of changes in Tables 1 and 2.

$$\gamma = \frac{\sum\limits_{m=1}^{M}(\sum\limits_{i=1}^{N}\sum\limits_{j=1}^{N}(l_{i,j,m} \oplus l'_{i,j,m})) \times b_m}{\sum\limits_{m=1}^{M}(\sum\limits_{i=1}^{N}\sum\limits_{j=1}^{N}(l_{i,j,m} \vee l'_{i,j,m})) \times b_m}, \tag{4}$$

Where $\oplus$ and $\vee$ are Boolean XOR and OR operations respectively. γ aries from 0 to 1.

Although the MFA strategy provides many benefits for large-scale streaming media delivery, it may lead to extra resource overhead. We define three different network resource usages U_{IP}, U_{O}, and U_{R}. U_{IP} is for the all-IP case without MFA, U_{O} is the optical network resource usage with MFA, and U_{R} is the real consumption of network resource with MFA in IP over optical network. For the case without MFA strategy, Let $\mathbf{F} = \{\mathbf{F}_1, \mathbf{F}_2, \cdots, \mathbf{F}_m, \cdots, \mathbf{F}_M\}$, where $\mathbf{F}_m$ denotes the route of multicast IP flow p_m.

$$U_{\mathrm{IP}} = \sum_{m=1}^{M}(b_m \times |\mathbf{F}_m|), \tag{5}$$

$$U_{\mathrm{O}} = t \times \sum_{k}|\mathbf{E}_k|, \tag{6}$$

$$U_{\mathrm{R}} = \sum_{m=1}^{M}\left(b_m \times \sum_{i=1}^{N}\sum_{j=1}^{N} l_{i,j,m}\right). \tag{7}$$

The original MFA problem has been proven as NP-hard[5]. To minimize the network resource usage, a heuristics named least tree first (LTP) algorithm is proposed, whose objective is to make the total number of light-trees finally set up in the network as small as possible. However, we find that LTP will result in great instability on both the program combination and the route of light-trees in the dynamical scenario. The main reason for this instability lies in that the receiving matrix is used as one critical parameter in selecting candidate program in order to minimize edge difference with existing light-tree. Thus, the program combination will keep altering with the receiving matrix's change.

The motivation of our research is to achieve stability both on the route and the program combination of the light-trees considering dynamical users' requests. Similar to the MFA problem, optimal network resource utilization is still one of our objectives. For simplicity, we split this problem into two parts, program selection and routing, then, we can address them one by one.

To avoid the influence coming from the receiving matrix, we develop an algorithm, named most correlated program first, based on the program correlation model introduced above to fulfill the program selection. The key idea of the proposed algorithm is that programs having the most program correlation with each other will be selected to be aggregated into one light-tree, which means program correlations will play a key role in program selection instead of receiving matrix. Motivation behind our algorithm lies in the assumption that if one tail has requested program p_m, the tail would most probably request program p_n which has the largest program correlation index $c_{m,n}$. This assumption is more consistent with the practical scenarios, no matter whether the setup of program correlation is based on commercial strategy or statistical end user behaviors. Besides the stability of MFA, we also try to optimize the network resource utilization by combining more programs into one light-tree. That is to say, the total number of light-trees finally set up in the network will be the least.

There are many works that have been done on the multicast routing[8,9]. As the shortest path tree and Steiner tree are two of the most common approaches used, we would select one of them to be our choice. In Ref. [10], the stability of both the shortest path tree and Steiner tree are investigated. We find that the stability of Steiner tree is affected by the link weight distribution while that of the shortest tree is relatively better in most situations. Considering our objective, the shortest path tree algorithm is chosen to set up the light-trees. Based on the output of program selection, we need to choose a head as the source of the aggregated light-tree if there is more than one possible head existing. To achieve optimized network resource utilization, we use the Dijkstra shortest path algorithm to calculate routes from each possible head to all tails of the light-tree, respectively. Then we choose the head with minimum cost tree to be the source and create the light-tree. The detail of the MCPF algorithm is shown in Fig. 2.

The complexity of the program selection algorithm can be formulated to be $O(|\mathbf{H}||\mathbf{P}|^2 + |\mathbf{P}|^3 + |\mathbf{P}||\mathbf{T}|)$. We can know that it will take $O(|\mathbf{V}|\log|\mathbf{V}| + |\mathbf{E}|)$ time to run Dijkistra algorithm[9]. Then the complexity of routing approach is $O(|\mathbf{P}|(|\mathbf{V}|\log|\mathbf{V}| + |\mathbf{E}|))$. Consequently, the total complexity of MCPF algorithm is $O(|\mathbf{H}||\mathbf{P}|^2 + |\mathbf{P}|^3 + |\mathbf{P}||\mathbf{T}| + |\mathbf{P}||\mathbf{V}|\log|\mathbf{V}| + |\mathbf{P}||\mathbf{E}|)$.

In this section, we evaluate the performance of the proposed multicast flow aggregation algorithm according to the stability and network resource usage. The LTF algorithm is extended to be topology-aware to compare with the MCPF algorithm. We adopt the NSFNET network, as shown in Fig. 3, in our simulations, in which all links' lengthes are set to unity. The bit rates of transponders are set to 1 Gb/s. We randomly choose 5 nodes and 10 nodes as the heads and tails, respectively. 500 programs will be delivered from heads to tails. Among these programs, 40% programs require 25 Mb/s, 40% programs require 6 Mb/s, and 2 Mb/s for the others. The values of program correlations are set to discrete values of 1, 10 and 100 with the probabilities of 20%, 40%, and 40% respectively. We randomly generate the transmission matrix $\mathbf{S}$ and the receiving matrix $\mathbf{R}$ with $\alpha = 0.5$ and $\beta = 0.5$. After running both MCPF and ELTF algorithm, we change the receiving matrix $\mathbf{R}$ to $\mathbf{R}'$, where Δ is varied from 0.01 to 1. For each Δ, we repeat experiment 100 times to get the average value.

```
Input: $\mathbf{V}$, $\mathbf{E}$, $t$, $\mathbf{H}$, $\mathbf{P}$, $\mathbf{S}$, $\mathbf{C}$, $\mathbf{B}$, $\mathbf{T}$, $\mathbf{R}$
Output: $\mathbf{A}$
$\mathbf{X} \leftarrow \mathbf{P}$   $k = 0$
while $\mathbf{X} \neq \emptyset$ do
  $k{+}{+}$
  choose one program $p_m$ that least heads contained and
   with the least program correlation from other unsettled
   programs in $\mathbf{X}$ as the first channel to join the light-tree
   $\mathbf{A}_k$
  $\mathbf{X} \leftarrow \mathbf{X} - \{p_m\}$
  repeat
    $\mathbf{C} = \emptyset$
    select the program $p_n$, which has at least one same
     heads with light-tree $\mathbf{A}_k$ and requires bandwidth less
     than ($t$-bandwidth required by all programs in $\mathbf{P}_k$),
     as candidate program to join $\mathbf{C}$
    if ($|\mathbf{C}| > 1$) do
      choose one program $p_c$ which has the most total
       program correlations from all programs in $\mathbf{P}_k$ to
       itself to join the light-tree $\mathbf{A}_k$
    end if
    $\mathbf{X} \leftarrow \mathbf{X} - \{p_c\}$
  until could not find out candidate channel
  if there is more than one possible head existing for the
   light-tree $\mathbf{A}_k$ do
    use Dijkstra algorithm to calculate the shortest path
     tree from each possible head to all tails of light-tree
     $\mathbf{A}_k$
  end if
  choose the head with minimum cost as the source of
   the light-tree $\mathbf{A}_k$ and setup the light-tree along the
   shortest path tree to all tails of $\mathbf{A}_k$
end while
```

Fig. 2. MCPF algorithm.

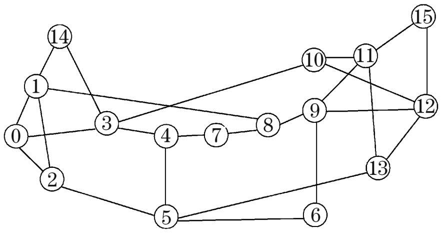

Fig. 3. NSFNET network topology.

Figure 4 presents the performance on the stability of MFA. We can see that the proposed algorithm performs very well in terms of the change ratio of the program delivery route that it is almost not influenced on receiving matrix's variation. Even when Δ equals to 1.0, γ still keeps around 10^{-3}.

We also observe the performance on the resource usage of two algorithms when α varies from 0.2 to 0.8. Figure 5 shows the statistical result about the relationship

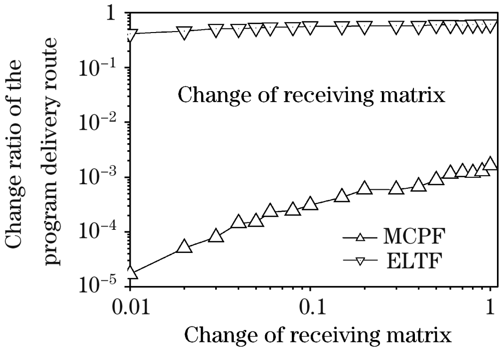

Fig. 4. Change ratio of the program delivery route versus change of receiving matrix.

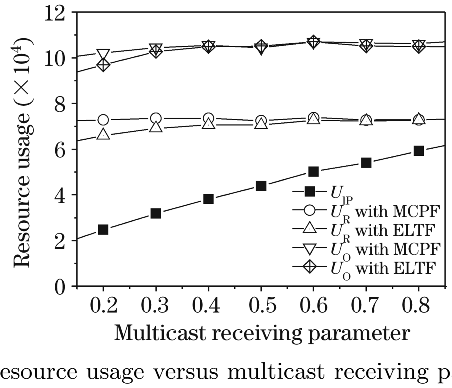

Fig. 5. Resource usage versus multicast receiving parameter.

between the resource usage and the multicast receiving parameter α. We can find that the resource overhead will continuously decline as the α increases, which is identical to the result of Ref. [5]. When the parameter α reaches to 0.5, the MCPF algorithm consumes almost the same network resource as the ELTF algorithm does. The increment can be negligible.

In conclusion, we focus on the stable MFA problem in IP over optical networks for digital TV transmission. We formulate the stable MFA problem and show that the original MFA problem is NP-complete. A program correlation matrix is introduced instead of the receiving matrix to complete program selection to avoid the instability along with the joining/leaving of end users. We try to combine programs into one light-tree as much as possible to reduce the number of aggregated light-trees. Simulation results show that the proposed algorithm keeps stable on the generated light-trees when Δ varies from 0 to 1, while it also achieves nearly the same network resource utilization as ELTF does.

This work was supported by the National Natural Science Foundation of China (No. 60502004) and the "863" Program of China. Y. Jin is the author to whom the correspondence should be addressed, his e-mail address is jinyh@sjtu.edu.cn.

References

1. C. Rossenhovel and J. Ganbar, EXCLUSIVE! Testing Cisco's IPTV infrastructure, www.lightreading.com /document.asp?doc_id=126173 (December 20, 2007).
2. T. Bertram, in *Proceedings of ECOC 2007* Plenary 2 (2007).
3. W. Sun, Y. Jin, W. Hu, H. He, X. Luo, P. Hu, W. Guo, Y. Su, and L. Leng, in *Proceedings of OFC 2005* OWG3 (2005).
4. Y. Jin, W. Hu, W, Sun, W. Guo, J. Wu, H. Li, J. Wang, M. Xu, Y. Li, L. Wei, G. Zhang, Y. Xu, H. Zhao, R. An, F. Yin, J. Wang, and X. Wei, in *Proceedings of ECOC 2007* **2,** 197 (2007).
5. Y. Zhu, Y. Jin, W. Sun, W. Guo, W. Hu, W. Zhong, and M. Wu, IEEE J. Sel. Areas Commun. **25,** 1011 (2007).
6. N. Singhal, L. H. Sahasrabuddhe, and B. Mukherjee, IEEE/ACM Trans. Networking **14,** 1104 (2006).
7. J. Cui, J. Kim, D. Maggiorini, K. Boussetta, and M. Gerla, in *Proceedings of IFIP Networking 2002* 1032 (2002).
8. M. Kodialam and T. V. Lakshman, IEEE/ACM Trans. Networking **11,** 676 (2003).
9. T. H. Cormen, C. E. Leiserson, and R. L. Rivest, Introduction to Algorithms (MIT, 1990).
10. P. van Mieghem and M. Janic, in *Proceedings of IEEE Infocom 2002* **2,** 1099 (2002).