國立臺灣大學資訊管理研究所

碩士論文

指導教授：林永松　博士

網路服務提供者提供具端對端頻寬限制與延遲限制之多傳輸率群播服務之路由及頻寬指定演算法

Multirate Multicast Routing and Capacity Assignment Algorithms with End-to-End Bandwidth and Delay Constraints for Network Service Providers

研究生：陳麒安　撰

中華民國九十二年七月

論 文 摘 要

本研究為網路服務提供者提供了一種群播網路設計方法，所要決定的是網路服務提供者應向網路提供者租用的頻寬大小以及提供具有服務品質保證的群播服務時所選擇之傳送路徑，我們考慮的服務品質包括了每個接收端的頻寬要求、端對端的最大平均延遲要求以及端對端的最大平均延遲變異要求。由於考慮提供的是多傳輸率的群播服務，所以一個群播群組內，各接收端要求的頻寬大小不同，針對此點，我們採用多層編碼機制將所傳輸的資料分層，爾後依據頻寬要求的大小決定所要傳輸的資料層多寡以達到路徑共用並且頻寬共用的優點。

研究中考慮了兩個網路服務提供者的經營方式，一為在網路容量均能負荷的前提下，網路服務提供者必須准予接受所有的服務品質保證的群播要求，於此策略下，服務提供者經由群播路由來最小化租用成本以達最大收益；另一則是網路服務者具有有限的線路租用預算，因此對於允許所有服務品質保證之群播要求有著不確定性的存在，於此策略下，網路服務提供者考慮使用允入控制以最大化允入之群播群組數量進而獲取最大收費，而所謂允入控制乃用於決定是否接受任何一組用戶群之服務要求，以達成確保其服務品質需求且不至於影響其他現存使用者服務品質的目標。

關鍵詞：服務品質保證、群播服務、最佳化、允入控制

THESIS ABSTRACT

GRADUATE INSTITUTE OF INFORMATION MANAGEMENT
NATIONAL TAIWAN UNIVERSITY
NAME: CHI-AN CHEN MONTH/YEAR: JANUARY 2003
ADVISER: YEONG-SUNG LIN

With the enormous growth, the Internet not only has the traffic demands increased but also the character of these IP applications. In particular, multimedia applications require a lot of bandwidth, and are very delay sensitive whether in the case of unicast or multicast. Nevertheless, the standard Internet Protocol (IP) based networks provide a best effort service. For the purpose of satisfying these multimedia applications, something better than "best effort" is required. The clients in pursuit of QoS must be assessed and if possible, improved upon.

In this thesis, the solution to the network service providers' decisions on how much capacity of network links they should lease from network providers and how they construct the paths for multicast routing with Quality-of-Service (QoS) guaranteed is proposed. The QoS requirements include bandwidth, end-to-end mean delay requirement, and end-to-end delay jitter requirement for each receiver of the multicast groups. And we are going to using some mechanism like 2-layered coding scheme in which the bandwidth of the signal as it passes through the network can be reduced in order to provide every receiver only with the bandwidth that it requests.

Two kinds of strategies are considered that network service providers may apply. One is to grant all the multicast requests in the assumption that network's total capacity could deal with no matter how many they are. Therefore, network service providers' aim is to minimize the

total cost of leasing network from network providers. The other strategy is when network service providers have budgets to lease the links. And this results in the uncertainty that not all requests can be accepted. In the consequence, service providers have to grant more as possible to maximize their revenue. Therefore, network service providers have to apply admission control mechanism to examine the network condition so as to permit the maximum number of request multicast groups by determining whether a request for multicast connection can be granted if the new requested QoS can be satisfied and the existing users' QoS would not be influenced in this strategy.

Keywords: Quality-of-Service, Multicast Service, Optimization, Admission Control

Table of Contents

List of Tables

List of Figures

Chapter 1 Introduction

1.1 Background

According to the contemporary developments in transmission and computing technologies, multimedia applications such as the teleconference and video on demand have already become achievable and will be comprehensively and commodiously used in the near future. Nevertheless, most of these applications necessitate large amount of bandwidth to deliver multimedia information to multiple destinations simultaneously. One possible method to meet this requirement is via multicasting.

Multicast stands for the transmission of data from one node (source node) to a selected multicast group of nodes (member nodes or destination nodes) in a communication network. Multicast routing takes advantage of trees, which we call multicast routing trees, through the network topology for transmissions to minimize resource usage such as cost and bandwidth by sharing links when transmitting data from one node to many destination nodes. The routing algorithm will only replicate at appropriate locations in order to arrive at all its destination nodes. A minimum cost multicast tree is also referred to as a Steiner tree. That is to say, a Steiner tree is to construct a minimum cost tree for a subset of the nodes in a network with fixed costs on the corresponding network links. The problem of determining a Steiner tree is known to be NP-complete[10].

IP Multicast traffic for a particular (source, destination group) pair is transmitted from the source to the receivers via a spanning tree that connects all the hosts in the group. Different IP Multicast routing protocols use different techniques to construct these multicast spanning trees; once a tree is constructed, all multicast traffic is distributed over it.

IP Multicast routing protocols generally follow one of two basic approaches, depending on the expected distribution of multicast group members throughout the network. The first approach is based on assumptions that the multicast group members are densely distributed throughout the network (i.e., many of the subnets contain at least one group member) and that bandwidth is plentiful. So-called “dense-mode” multicast routing protocols rely on a technique called flooding to propagate information to all network routers. Dense-mode routing protocols include Distance Vector Multicast Routing Protocol (DVMRP), Multicast Open Shortest Path First (MOSPF), and Protocol-Independent Multicast - Dense Mode (PIM-DM).

Currently, the multicasting backbone (MBone), which uses DVMRP for multicast routing, is one of the applications that have been developed rapidly on the Internet using IP multicasting technology.

The second approach to multicast routing basically assumes the multicast group members are sparsely distributed throughout the network and bandwidth is not necessarily widely available, for example across many regions of the Internet or if users are connected via ISDN lines. Sparse-mode does not imply that the group has a few members, just that they are widely dispersed. In this case, flooding would unnecessarily waste network bandwidth and hence could cause serious performance problems. Hence, “sparse-mode” multicast routing protocols must rely on more selective techniques to set up and maintain multicast trees. Sparse-mode routing protocols include Core Based Trees (CBT) and Protocol-Independent Multicast - Sparse Mode (PIM-SM) [18][19].

Furthermore, the current and future real-time applications such as teleconferencing, remote collaboration and distance education involve the transmission of multimedia information and therefore it is essential to satisfied quality-of-service constraints (such as bounded end-to-end delay, bounded delay-variation and bandwidth requirement). At the routing level, these two requirements are translated into the problem of determining a multicast tree, usually rooted at the source node and spanning the set of receiver nodes. The quality-of-service constraints typically impose a restriction on the acceptable multicast trees.

1.2 Motivation

With the enormous growth, the Internet not only has the traffic demands increased but also the character of these IP applications. In particular, multimedia applications require a lot of bandwidth, and are very delay sensitive whether in the case of unicast or multicast. Nevertheless, the standard Internet Protocol (IP) based networks provide a best effort service. For the purpose of satisfying these multimedia applications, something better than "best effort" is required. The clients in pursuit of QoS must be assessed and if possible, improved upon.

With the increased traffic on the Internet, it exceeds the network capacity and the service is not denied but rather degrades. Maybe some applications are able to stand such degradations however there's still a negative situation in the case of real-time applications which are sensitive to the traffic flow on the network. Increasing bandwidth may be a solution but when the traffic bursts occur, this will still not work. Furthermore, we cannot just create more and more bandwidth because, first, that would require tremendous amounts of investment and second, that would also be such wasteful as to excessively provision network resources. Therefore, a more intelligent solution would be to optimize usage of the available bandwidth. QoS mechanisms are designed to supply IP applications so that the network can distinguish traffic with strict requirements such as reliability, timing i.e. real-time multimedia traffic. The main intention of QoS is to achieve some level of predictability and control beyond the best effort service by.

Generally, QoS refers to the network element (e.g. operators, application, host or router etc.) commitment to providing and maintaining acceptable values of parameters or characteristics of user applications in order to satisfy the users' application requirements and expectations.

In certain cases the network operator may be able to guarantee (perhaps probabilistically) the QoS level a given user's application will receive. This is of particular concern for the continuous transmission of high-bandwidth video and multimedia information. Furthermore, in [6], it referred that QoS would be capable to:

- Supporting dedicated bandwidth
- Improving loss characteristics
- Avoiding and managing network congestion
- Shaping network traffic
- Setting traffic priorities across the network

But providing QoS guarantees is difficult in networks that offer "best effort" service, such as the Internet. IP makes no guarantees about when data will arrive, or how much data it can deliver. Therefore a lot of work has been carried out recently on how to add QoS support to the Internet service model. Examples of this include the intserv (Integrated Services) and diffserv (Differentiated Services) approaches.

Thus we try to provide the network service providers who lease the network from network providers with the consideration of QoS requirements and some mechanism of network resource control.

1.3 Literature Survey

1.3.1 QoS Routing

The QoS routing is a critical network function for the transmission and distribution of digitalized audio or video throughout the communication networks. It has two objectives: (1) finding routes that satisfy the QoS requirements and (2) making the efficient use of the network resources. Many extensive researches have been conducted on QoS routing issues recently. Overall, based on the way the state information is maintained, the existing QoS routing algorithms can be partitioned into three broad classes: (1) source routing; (2) distributed routing and (3) hierarchical routing algorithms. In [5], S. Chen and K. Nahrstedt did a thorough survey on these QoS routing algorithms. But they focused on network models in virtual circuit mode, which was connection oriented.

In [21], J. Kleinberg addressed an NP-Complete problem, which combined the selecting paths for routing and allocating bandwidth fairly among connections in the max-min sense. But their work was still more connection-oriented with single source.

In [15], Ghosh, Sarangan, and Acharya proposed a new distributed routing algorithm for QoS flows. The routing algorithm contains a new packet forwarding mechanism based on the QoS requirements of the connection. The two-level forwarding has a low overhead when compared to the flooding-based call setup. However, the algorithm only considers the bandwidth requirements, and other QoS requirements such as loss, delay, and jitter are also important that have to be considered in. Sufficient bandwidth cannot provide smooth video-on-demand service. It should control the delay and jitter under certain requirements.

In addition, this algorithm only focus on the unicast flows without considering multicast flows.

Therefore, in order to take account of the lack of end-to-end delay and delay jitter in the QoS routing problems, we propose a mathematical formulation for QoS routing including bandwidth, delay, and delay jitter in this thesis.

1.3.2 IP Multicast

In the current environment, the receivers are typically computers with a wide range of processing capabilities, possibly augmented by special purpose video processing hardware. As a result, some receivers can implement more complex decompression algorithms, at a higher frame rate or resolution than others. In addition, different receivers have different rate connections into the network. Data is sent from the source node and arrive at receiver nodes with different rate depending upon each receiver’ s bandwidth requirement. The range of connections to the Internet is from voice band modems of a few tens of kilobits per seconds for homes, up to OC3 rates of 155 megabits per second for several super computer centers. In a pay-per-view system, pricing can also be used to encourage receivers to limit the demands that they place upon the network. At present, most video broadcasts over the Mbone deliver the same signal to all of the receivers and operate conservatively so that all of the intended receivers can receive and decode the signal. In effect, everyone get the grade of service of the least capable receivers.

In order to provide every receiver only with the bandwidth that it requests, we have to reduce the bandwidth of the signal as it passes through the network. And M. Ghanbari and F.

Kishino et al, each used a two-layered coding scheme to extract critical video data. In [14], Ghanbari proposed a method to divide the bit stream generated by a conditional-replenishment interframe coding technique into two parts. The first part makes up the contents of the so-called 'guaranteed packets' and the second part constitutes the contents of the 'enhancement packets'. Guaranteed packets are transmitted in the guaranteed channel whereas enhancement packets are transmitted without any guarantee. And in [20], Kishino et al proposed a DCT layered coding technique, which separated the DCT coefficients into MSP's (most significant parts) and LSP's (least significant parts) where MSP packets take priority over the LSP packets.

Therefore this can be implemented by using a progressive coder or by converting between encoding formats. An example of a progressive coder is a Fourier transform coder in which the high resolution components and low resolution components are placed in different packets. The low resolution signal can be transmitted to all of the receivers and the high resolution components only th those that request them. Similarly, progressive intraframe coders can be designed to deliver 30, 15, or 5 frames per second, by marking the frames and not forwarding all of them along all of the branches.

Consequently, we can only consider the maximum request bandwidth of every group that passes through the link, and aggregate them to figure out how much link lease cost should be paid as a network service provider as illustrated in Figure 1-1.

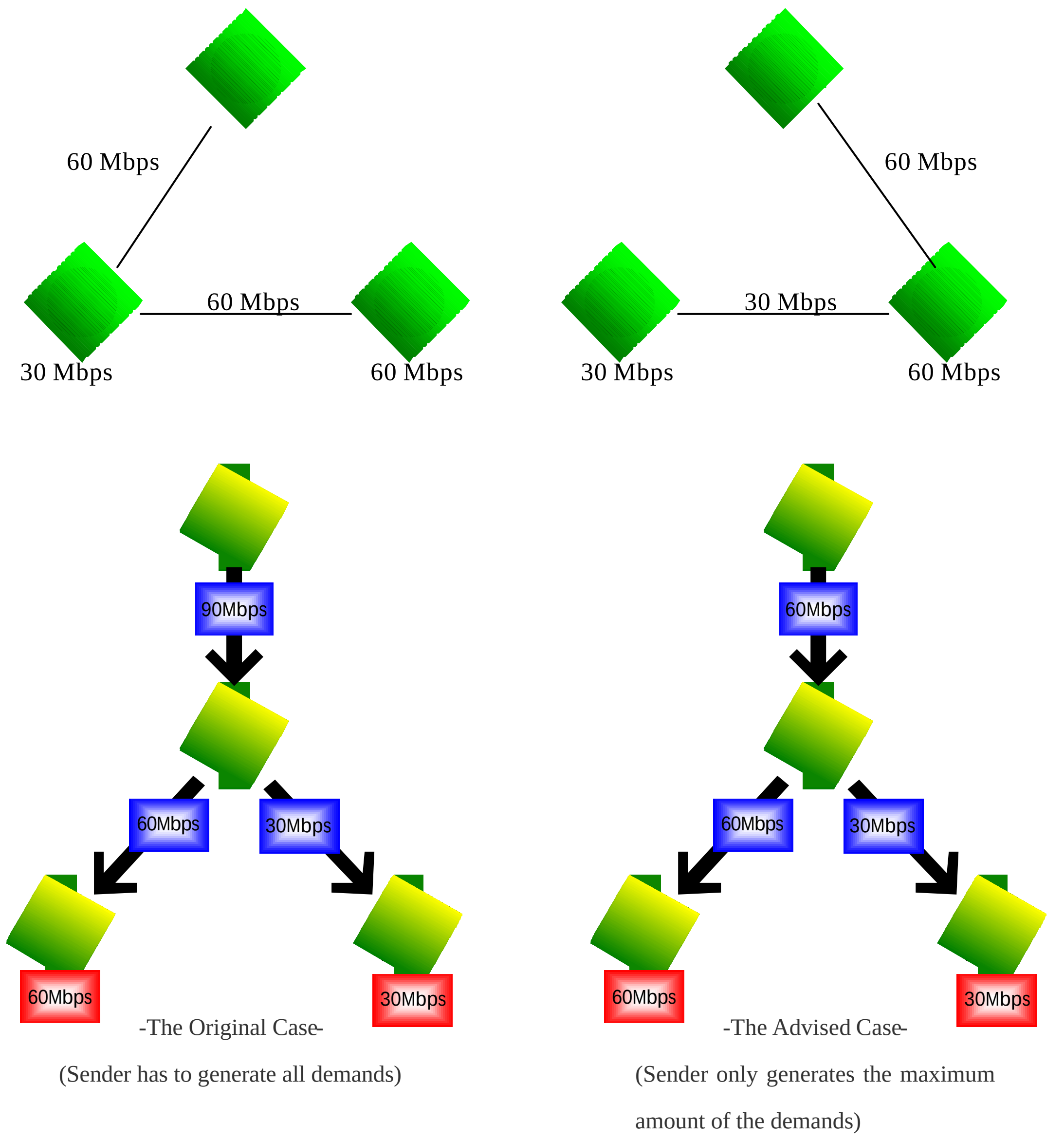

Figure 1-1 Multi-Layer Coding for Multicast with Multi-Rate Receivers

1.3.3 Admission Control

The objective of admission control is to regulate the operation of a network in such a way to ensure the uninterrupted service provision to the existing connections and at the same time to accommodate in an optimum way the new connection requests. From another point of view, admission control is also a preventive method to congestion control. This can be done by managing the available network resources and allocating them in an optimum way among the system users. Once a request is accepted, the required resources must be guaranteed [26].

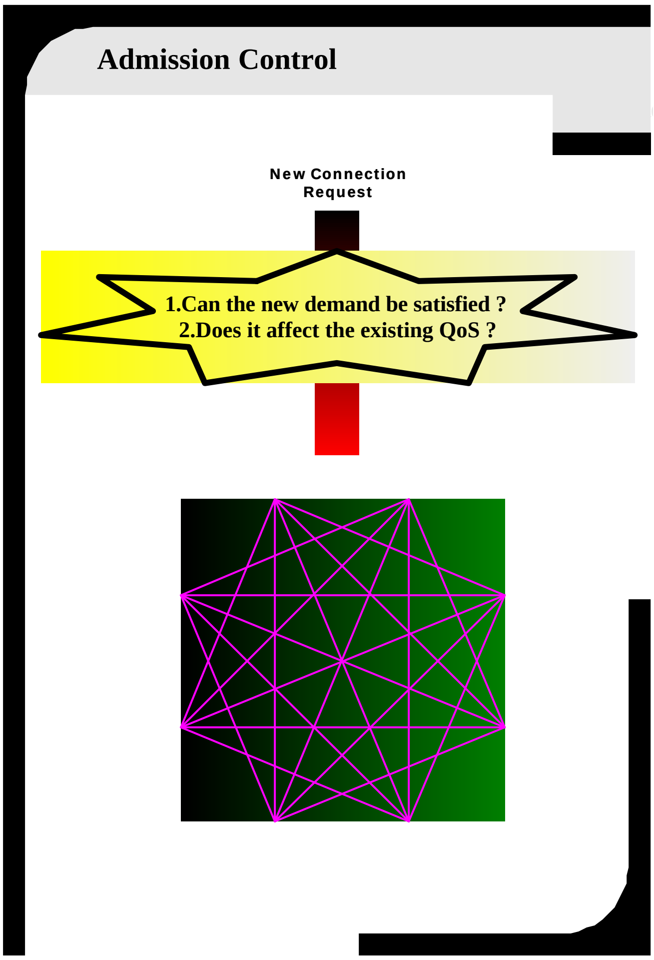

Figure 1-2 Admission Control

1.3.4 Pricing

In [4], usage-based pricing schemes can be classified into two general categories: static and dynamic pricing. Static pricing means that the price is set by the network provider based on observation and estimation from historical data and is independent of real-time network utilization. Advantages of static pricing are simplicity of implementation and predictability form the customer's point of view. While during congestions the bandwidth is especially scarce, efficient prices must reflect the current availability of resources. Therefore, this is the principal purpose of dynamic pricing which allows more formal optimization by taking into account the fluctuations in network utilization. The most common-used dynamic pricing scheme is a bidding price scheme because a lot of researchers argue that users should have the freedom to send traffic and show their willingness to pay for it. Mackie-Mason and Varian [24] propose a per-packet bidding price scheme called a "smart market" scheme. In this scheme, each user assigns a willingness to pay for each packet he sends to the communication network. The network will accept the packets that have a bidding price higher than the current cutoff price, which is calculated from the marginal congestion cost.

In [17], Honig and Steiglitz present a simple pricing policy containing two different entries: day price (or peak price) and night price (or off-peak price) in an attempt to achieve traffic smoothing. As users who want to transmit data during high network utilization periods will be charged more, some of them may choose to wait until a low network utilization period. By implementing this mechanism, network utilization can be distributed evenly over all time periods and very high peak utilization can be avoided.

To sum up, in order to preserve fairness and predictability from the customers' point of view with QoS service, we cannot allow the charge price changing in proportion to the utilization of network resources for the purpose of controlling resource overuse. Therefore, we take apart pricing and utilization control by pricing on QoS requirements such as bandwidth, delay and delay jitter requirements and additionally setting some link usage limits in our mathematical formulations.

1.4 Proposed Approachs

To solve the network service provider's capacity leasing problem over QoS constrained multicast routing is to get the optimal leased capacity for each link corresponding to each different QoS requirement and the budget constraint. And in each multicast group, for the purpose of providing every receiver only with the bandwidth that it requests, we can reduce the bandwidth of the signal as it passes through the network by using a progressive coder or by converting between encoding formats. Therefore we only consider the maximum request bandwidth of every group that passes through the link, and we aggregate them to figure out how much link lease cost should be paid as a network service provider. The problem we model is an optimization problem. It belongs to a nonlinear integer mathematical programming problem. We will apply the Lagrangean relaxation method and the subgradient method to solve the problems [3][9].

1.5 Thesis Organization

The remainder of the paper provides two strategies to a multicast problems and their mathematical formulations — network service providers' quality-of-service multicasting model. And the future work will be composed of the following chapters. Chapter 3 introduces the Lagrangean relaxation approach, which is applied to solve our problems. The Lagrangean relaxation is a powerful mathematical technique designed for large-scale linear programming problems in the 1970s. Chapter 4 describes that the efficiency and effectiveness of the algorithm will be evaluated by computational experiments using some possible network topologies.

Chapter 2 Problem Formulations

The problem we modeled is an admission control over a set of connection requests of multicast groups. A multicast group is a group requesting multicast connection, which has one sender and several receivers.

Each receiver in a multicast group has its own QoS requirements for the demands on the bandwidth, maximum delay and maximum delay variation. Except the constraints resulted from the guarantee of QoS requirements, we also set two limits of total link and total bandwidth in a multicast group to control the resource usage. And we also take the standpoint from a network service provider's point of view; the objective of our model is to maximize net profits, which are derived from the sum of revenue for admitting multicast groups with QoS requirements.

Therefore we try to construct minimum cost trees and to admit the most number of multicast groups. The cost to lease a network link is determined by the percentage of resource usage from admitting multicast groups which route through it. And the network service provider also has to decide how much capacity should be leased from the network provider constrained by the budget or the size of request groups.

We consider two strategies that the network service provider would apply. One is to grant all requested multicast groups. That is, with the presupposition of permitting all multicast requests, the network service provider tries to decide how much capacity to lease and how the multicast paths route through the network to minimize the total cost of leasing the network links. The other is constrained by the limit of budgets. Since the network service provider has limited budget, the leased capacity maybe not sufficient to allow all the multicast groups to transmit data files. Therefore, the provider has to admit requests to receive as more revenues as possible under the limited capacity.

2.1 Strategy I: Cost Down

2.1.1 Problem Descriptions

Problem assumptions:
1. Each buffer of network router is infinite. That is, the packet will not be discard in the network.
2. The average link delay can be decided by two parameters: aggregation link flows and link capacity.
3. The link delay function is a monotonically increasing function with respect to the aggregation flows.
4. The link delay function is a convex function with respect to aggregate flow or link capacity. But aggregation flow and link capacity jointly may not be a convex function.

Table 2-1: Problem Assumptions for Strategy 1

Given:

1. The network topology.
2. The end-to-end bandwidth requirements for receivers of multicast groups.
3. The end-to-end mean delay requirements for receivers of multicast groups.
4. The end-to-end mean delay variation requirements for receivers of multicast groups.
5. The limit of total links used by a multicast subtree.
6. The limit of total aggregate bandwidth used by a multicast group.

Objective:

To minimize the total cost of leasing the network links.

Subject to:

1. The QoS constraint guaranteed by limiting the end-to-end mean delay and delay jitter for each O-D pair of the multicast groups in the network.
2. The link resource occupation controlled by limiting the total number of used links and bandwidth for each multicast groups.
3. The tree structure constraints such as the number of links on a multicast subtree must exceed the minimum number of hops to the farthest destination node and the number of destinations.
4. For each link, the aggregate flow on it by admitting multicast groups must not exceed the capacity of it.

To determine:

1. The maximum capacity of each link to use.
2. The minimum total lease cost.
3. The topologies of the multicast sessions in the network.
4. The route of each O-D pair of each multicast group in the network.

Table 2-2. Problem Descriptions for Strategy 1

2.1.2 Notations

Given Parameters	
Notations	Descriptions
G	The set of user groups requesting for connection.
V	The set of nodes in the graph (network).
L	The set of real links in the graph (network).
A_l	The set of capacity choices for each link.
D_g	The set of destination of multicast group g.
P_{gd}	The set of paths that destination d of multicast group g may use.
d_{pl}	1 if link l is on path p, and 0 otherwise.
h_g	The minimum number of hops to the farthest destination node in multicast group g.
Q_{gd}	End-to-end bandwidth requirement for destination d of multicast group g.
$F_l(f_l, C_l)$	The average delay on link $l \in L$, which is a function of f_l and C_l.
$M_l(f_l, C_l)$	The average delay variation on link $l \in L$, which is a function of f_l and C_l.
L_{gd}	End-to-end mean delay requirement for destination d of multicast group g.
J_{gd}	End-to-end mean delay variation requirement for destination d of multicast group g.
H_g	The limit of total links used in the subtree for multicast group g.
R_g	The limit of total aggregate bandwidth used by multicast group g.
$e_l(C_l)$	The cost to use the link l, which is a function of C_l.

Table 2-3. Notations of Problem I - Given Parameters

Decision Variables	
Notations	Descriptions
x_{gpd}	1 if path is selected for group g destined for destination d and 0 otherwise.
y_{gl}	1 if link l is on the subtree adopted by multicast group g and 0 otherwise.
t_{gld}	1 if link l is used by destination d of multicast group g and 0 otherwise.
r_{gl}	The maximum data rate of the multicast group g on the link l.
f_l	The aggregate flow on link $l \in L$.
C_l (packets/sec)	The capacity of link l.

Table 2-4. Notations of Problem I - Decision Variables

2.1.3 Problem Formulation

Objective function (IP 1.1):

$$Z_1 \quad = \quad \min \sum_{l \in L} e_l\left(C_l\right) \qquad \text{(IP 1.1)}$$

subject to:

$x_{gpd} = 0$ or 1	$\forall d \in D_g,\ g \in G,\ p \in P_{gd}$	(1.1)
$y_{gl} = 0$ or 1	$\forall g \in G, l \in L$	(1.2)
$t_{gld} = 0$ or 1	$\forall g \in G, d \in D_g, l \in L$	(1.3)
$\sum_{p \in P_{gd}} x_{gpd} d_{pl} = t_{gld}$	$\forall g \in G, d \in D_g, l \in L$	(1.4)
$\sum_{d \in D_g} t_{gld} \le \lvert D_g \rvert y_{gl}$	$\forall g \in G, l \in L$	(1.5)
$\sum_{l \in L} y_{gl} \ge \max\{h_g, \lvert D_g \rvert\}$	$\forall g \in G$	(1.6)
$\sum_{p \in P_{gd}} x_{gpd} = 1$	$\forall d \in D_g, g \in G$	(1.7)
$t_{gld} Q_{gd} \le r_{gl}$	$\forall g \in G, d \in D_g, l \in L$	(1.8)
$r_{gl} \in [0, \max_{d \in D_g} Q_{gd}]$	$\forall g \in G, l \in L$	(1.9)
$\sum_{g \in G} r_{gl} = f_l$	$\forall l \in L$	(1.10)
$f_l \le C_l$	$\forall l \in L$	(1.11)

$$C_l \in A_l \qquad \forall l \in L \qquad (1.12)$$

$$\sum_{l \in L} t_{gld} F_l(f_l, C_l) \le L_{gd} \qquad \forall g \in G, d \in D_g \qquad (1.13)$$

$$\sum_{l \in L} t_{gld} M_l(f_l, C_l) \le J_{gd} \qquad \forall g \in G, d \in D_g \qquad (1.14)$$

$$\sum_{l \in L} r_{gl} \le R_g \qquad \forall g \in G \qquad (1.15)$$

$$\sum_{l \in L} y_{gl} \le H_g \qquad \forall g \in G. \qquad (1.16)$$

Constraint (1.1): The integer property constraint.

Constraint (1.2): The integer property constraint.

Constraint (1.3): The integer property constraint.

Constraint (1.4): If one link is selected for group g destined for destination d, it must also be on the path adopted by multicast group g for destination d.

Constraint (1.5): If one path is selected for group g destined for destination d, it must also be on the subtree adopted by multicast group g.

Constraint (1.6): The number of links on the multicast subtree adopted by the multicast group g is at least the maximum of h_g and the cardinality of D_g. The h_g and the cardinality of D_g are the legitimate lower bounds of the number of links on the multicast subtree adopted by the multicast group g.

Constraint (1.7): Exactly only one path is selected for any group g destined for its destination d.

Constraint (1.8): The data rate of the multicast group g on the link l is the maximum rate of group g 's O-D pairs which pass through the link l.

Constraint (1.9) : The range of r_{gl} is from 0 to the maximum bandwidth requested by the destination of multicast group g which route through the link l.

Constraint (1.10) : The aggregate flows on each link *l*.

Constraint (1.11) : The aggregate flows on each link l cannot exceed its physical capacity C_l.

Constraint (1,12) : The value of capacity on each link is a choice in a discrete value set.

Constraint (1.13) : The end-to-end average delay should be no longer than maximum allowable end-to-end average delay requirement for all users.

Constraint (1.14) : The end-to-end average delay jitter should be no longer than maximum allowable end-to-end delay jitter requirement for all users.

Constraint (1.15) : The total bandwidth assigned to multicast group g should not be larger than the limit as given.

Constraint (1.16) : The total number of network links assigned to multicast group g should not be larger than the limit as given.

2.2 Strategy II: Limited Budget

2.2.1 Problem Descriptions

Problem assumptions:
1. Each buffer of network router is infinite. That is, the packet will not be discard in the network.
2. The average link delay can be decided by two parameters: aggregation link flows and link capacity.
3. The link delay function is a monotonically increasing function with respect to the aggregation flows.
4. The link delay function is a convex function with respect to aggregate flow or link capacity. But aggregation flow and link capacity jointly may not be a convex function.

Table 2-5. Problem Assumptions for Strategy 2

Given:
1. The network topology.
2. The end-to-end bandwidth requirements for receivers of multicast groups.
3. The end-to-end mean delay requirements for receivers of multicast groups.
4. The end-to-end mean delay variation requirements for receivers of multicast groups.
5. The budget to lease the network links.
6. The limit of total links used by a multicast subtree.
7. The limit of total aggregate bandwidth used by a multicast group.
8. The charging price for receivers corresponding to their QoS requirements.

Objective:

To maximize the total profit.

Subject to:

1. The QoS constraint guaranteed by limiting the end-to-end mean delay and delay jitter for each O-D pair of the multicast groups in the network.
2. The link resource occupation controlled by limiting the total number of used links and bandwidth for each multicast groups.
3. The tree structure constraints such as the number of links on a multicast subtree must exceed the minimum number of hops to the farthest destination node and the number of destinations.
4. For each link, the aggregate flow on it by admitting multicast groups must not exceed the capacity of it.

To determine:

1. The maximum capacity of each link to use.
2. The maximum total profit.
3. The topologies of the multicast sessions in the network.
4. The route of each O-D pair of each multicast group in the network.

Table 2-6. Problem Descriptions for Strategy 2

2.2.2 Notations

Given Parameters	
Notation	Description
G	The set of all user groups requesting for connection.
V	The set of nodes in the graph (network).
L	The set of real links in the graph (network).
A_l	The set of capacity choices for each link.
T_g	The set of trees in the network for multicast/unicast group g.
t'_g	An artificial tree for group g with zero cost/revenue, and the link capacity of the tree is infinite.
T'_g	$T_g \cup \{t'_g\}$.
D_g	The set of destination of multicast group g.
P_{gd}	The set of paths that destination d of multicast group g may use.
p'_{gd}	The set of paths that destination d of multicast group g of the artificial tree t'_g.
P'_{gd}	$P_{gd} \cup \{p'_{gd}\}$.
d_{pl}	1 if link l is on path p, and 0 otherwise.
h_g	The minimum number of hops to the farthest destination node in multicast group g.
Q_{gd}	End-to-end bandwidth requirement for destination d of multicast group g.
$F_l(f_l, C_l)$	The average delay on link $l \in L$, which is a function of f_l and C_l.
$M_l(f_l, C_l)$	The average delay variation on link $l \in L$, which is a function of f_l and C_l.
L_{gd}	End-to-end mean delay requirement for destination d of multicast group g.
J_{gd}	End-to-end mean delay variation requirement for destination d of multicast group g.

H_g	The limit of total links used in the subtree for multicast group g.
R_g	The limit of total aggregate bandwidth used by multicast group g.
$k_{gd}\left(Q_{gd}, L_{gd}, J_{gd}\right)$	The price charging for destination d of the multicast group g, which is a function of Q_{gd}, L_{gd} and J_{gd}.
$e_l\left(C_l\right)$	The cost to use the link l, which is a function of C_l.
B	The total budget to lease the network links from network provider.

Table 2-7. Notations of Problem II- Given Parameters

Decision Variables	
Notations	Descriptions
z_g	1 if the multicast group g is admitted to the network and 0 otherwise.
x_{gpd}	1 if path is selected for group g destined for destination d and 0 otherwise.
y_{gl}	1 if link l is on the subtree adopted by multicast group g and 0 otherwise.
t_{gld}	1 if link l is used by destination d of multicast group g and 0 otherwise.
r_{gl}	The maximum data rate of the multicast group g on the link l.
f_l	The aggregate flow on link $l \in L$.
C_l (packets/sec)	The capacity of link l.

Table 2-8. Notations of Problem II - Decision Variables

2.2.3 Problem Formulation

Objective function (IP 2.1):

$$Z \quad = \quad \min \sum_{g \in G} \sum_{d \in D_g} -z_g k_{gd} (Q_{gd}, L_{gd}, J_{gd}) \qquad \text{(IP 2.1)}$$

subject to:

$x_{gpd} = 0 \text{ or } 1$	$\forall d \in D_g,\ g \in G,\ p \in P'_{gd}$	(2.1)
$y_{gl} = 0 \text{ or } 1$	$\forall g \in G, l \in L$	(2.2)
$t_{gld} = 0 \text{ or } 1$	$\forall g \in G, d \in D_g, l \in L$	(2.3)
$\sum_{p \in P_{gd}} x_{gpd} d_{pl} = t_{gld}$	$\forall g \in G, d \in D_g, l \in L$	(2.4)
$\sum_{d \in D_g} t_{gld} \le \left\|D_g\right\| y_{gl}$	$\forall g \in G, l \in L$	(2.5)
$\sum_{l \in L} y_{gl} \ge \max\{h_g, \left\|D_g\right\|\}$	$\forall g \in G$	(2.6)
$\sum_{p \in P'_{gd}} x_{gpd} = 1$	$\forall d \in D_g, g \in G$	(2.7)
$t_{gld} Q_{gd} \le r_{gl}$	$\forall g \in G, d \in D_g, l \in L$	(2.8)
$r_{gl} \in [0, \max_{d \in D_g} Q_{gd}]$	$\forall g \in G, l \in L$	(2.9)
$\sum_{g \in G} r_{gl} = f_l$	$\forall l \in L$	(2.10)
$f_l \le C_l$	$\forall l \in L$	(2.11)
$C_l \in A_l$	$\forall l \in L$	(2.12)

$$\sum_{l \in L} e_l(C_l) \leq B \qquad (2.13)$$

$$\sum_{l \in L} t_{gld} F_l(f_l, C_l) \leq L_{gd} \qquad \forall g \in G, d \in D_g \qquad (2.14)$$

$$\sum_{l \in L} t_{gld} M_l(f_l, C_l) \leq J_{gd} \qquad \forall g \in G, d \in D_g \qquad (2.15)$$

$$z_g = \sum_{p \in P_{gd}} x_{gpd} \qquad \forall g \in G, d \in D_g \qquad (2.16)$$

$$z_g = 0 \text{ or } 1 \qquad \forall g \in G \qquad (2.17)$$

$$\sum_{l \in L} r_{gl} \leq R_g \qquad \forall g \in G \qquad (2.18)$$

$$\sum_{l \in L} y_{gl} \leq H_g \qquad \forall g \in G. \qquad (2.19)$$

Constraint (2.1) : The integer property constraint.

Constraint (2.2) : The integer property constraint.

Constraint (2.3) : The integer property constraint.

Constraint (2.4) : If one link is selected for group g destined for destination d, it must also be on the path adopted by multicast group g for destination d.

Constraint (2.5) : If one path is selected for group g destined for destination d, it must also be on the subtree adopted by multicast group g.

Constraint (2.6) : The number of links on the multicast subtree adopted by the multicast group g is at least the maximum of h_g and the cardinality of D_g. The h_g and the cardinality of D_g are the legitimate lower bounds of the number of links on the multicast subtree adopted by the multicast group g.

Constraint (2.7) : Exactly only one path is selected for any group g destined for its destination d.

Constraint (2.8) : The data rate of the multicast group g on the link l is the maximum rate of group g's O-D pairs that pass through the link l.

Constraint (2.9) : The range of r_{gl} is from 0 to the maximum bandwidth requested by the destination of multicast group g which route through the link l.

Constraint (2.10) : The aggregate flows on each link l.

Constraint (2.11) : The aggregate flows on each link l cannot exceed its physical capacity C_l.

Constraint (2.12) : The value of capacity on each link is a choice in a discrete value set.

Constraint (2.13) : The cost of leasing network links should not be larger than the budget.

Constraint (2.14) : The end-to-end average delay should be no longer than maximum allowable end-to-end average delay requirement for all users.

Constraint (2.15) : The end-to-end average delay jitter should be no longer than maximum allowable end-to-end delay jitter requirement for all users.

Constraint (2.16) : If a group g which is not admitted to the network, the flag of group g : z_g must be 0, and select no paths.

Constraint (2.17) : The integer property constraint.

Constraint (2.18) : The total bandwidth assigned to multicast group g should not be larger than the limit as given.

Constraint (2.19) : The total number of network links assigned to multicast group g should not be larger than the limit as given.

Chapter 3 Lagrangean Relaxation

3.1 Strategy I: Cost Down

3.1.1 Solution Approach

By using the Lagrangean Relaxation method, we can transform the primal problem (IP1.1) into the following Lagrangean Relaxation problem (LR1.1) where Constraints (1.4), (1.5), (1.8), (1.10), (1.13), and (1.14) are relaxed.

3.1.2 Lagrangean Relaxation

With a vector of non-negative Lagrangean multipliers, a Lagrangean Relaxation problem of (IP1.1) is given by

Optimization problem (LR1.1):

$$Z_{D1.1}(b,e,q,p,t,m) = \min \sum_{l\in L} e_l\left(C_l\right) +$$

$$\sum_{g\in G}\sum_{d\in D_g}\sum_{l\in L} b_{gld}\Big(\sum_{p\in P_{gd}} x_{gpd}d_{pl} - t_{gld}\Big) +$$

$$\sum_{g\in G}\sum_{l\in L} e_{gl}\Big(\sum_{d\in D_g} t_{gld} - \left|D_g\right| y_{gl}\Big) +$$

$$\sum_{g\in G}\sum_{d\in D_g}\sum_{l\in L} q_{gld}(t_{gld}Q_{gd} - r_{gl}) +$$

$$\sum_{l\in L} p_l\Big(\sum_{g\in G} r_{gl} - f_l\Big) +$$

$$\sum_{g\in G}\sum_{d\in D_g} t_{gd}\Big[\sum_{l\in L} t_{gld}F_l(f_l, C_l) - L_{gd}\Big] +$$

$$\sum_{g\in G}\sum_{d\in D_g} m_{gd}\Big[\sum_{l\in L} t_{gld}M_l(f_l, C_l) - J_{gd}\Big] \qquad \text{(LR 1.1)}$$

subject to:

$x_{gpd} = 0$ or 1	$\forall d \in D_g,\ g \in G,\ p \in P_{gd}$	(1.1)
$y_{gl} = 0$ or 1	$\forall g \in G, l \in L$	(1.2)
$t_{gld} = 0$ or 1	$\forall g \in G, d \in D_g, l \in L$	(1.3)
$\sum_{l\in L} y_{gl} \ge \max\{h_g, \left\|D_g\right\|\}$	$\forall g \in G$	(1.6)
$\sum_{p\in P_{gd}} x_{gpd} = 1$	$\forall d \in D_g, g \in G$	(1.7)
$r_{gl} \in [0, \max_{d\in D_g} Q_{gd}]$	$\forall g \in G, l \in L$	(1.9)
$f_l \le C_l$	$\forall l \in L$	(1.11)
$C_l \in A_l$	$\forall l \in L$	(1.12)
$\sum_{l\in L} r_{gl} \le R_g$	$\forall g \in G$	(1.15)

$$\sum_{l \in L} y_{gl} \leq H_g \qquad \forall g \in G \qquad (1.16)$$

where $b_{gld}, e_{gl}, q_{gld}, p_l, t_{gd}, m_{gd}$ are Lagrangean multipliers and $e_{gl}, q_{gld}, t_{gd}, m_{gd} \geq 0$. To solve (LR1.1), we can decompose (LR1.1) into the following four independent and easily solvable optimization subproblems.

Subproblem 1.1: (related to decision variable x_{gpd})

$$Z_{Sub1.1}(b) = \min \sum_{g \in G} \sum_{d \in D_g} \sum_{l \in L} \sum_{p \in P_{gd}} b_{gld} d_{pl} x_{gpd}$$

subject to:

$$x_{gpd} = 0 \text{ or } 1 \qquad \forall d \in D_g, g \in G, p \in P_{gd} \qquad (1.1)$$

$$\sum_{p \in P_{gd}} x_{gpd} = 1 \qquad \forall d \in D_g, g \in G. \qquad (1.7)$$

The subproblem can be further decomposed into $|G||D_g|$ independent shortest path problems with nonnegative arc weights. Each shortest path problem can be easily solved by the Dijkstra's algorithm, the link cost is b_{gld}. If b_{gld} is negative, x_{gpd} is 1. Otherwise x_{gpd} is 0.

Subproblem 1.2: (related to decision variable y_{gl})

$$Z_{Sub1.2}(e) = \min \; -\sum_{g\in G}\sum_{l\in L} e_{gl}\left|D_g\right| y_{gl}$$

subject to:

$y_{gl} = 0$ or 1 $\quad \forall g \in G, l \in L \quad$ (1.2)

$\sum_{l\in L} y_{gl} \geq \max\{h_g, \left|D_g\right|\} \quad \forall g \in G \quad$ (1.6)

$\sum_{l\in L} y_{gl} \leq H_g \quad \forall g \in G\,. \quad$ (1.16)

The algorithm to solve (Subproblem 1.2) is stated as follows[1]:

Step 1. Compute max $\{h_g, |D_g|\}$ for multicast group g.

Step 2. Compute the number of negative coefficient $-e_{gl}\left|D_g\right|$ for all links on multicast group g.

Step 3. If the number of negative coefficient is greater than H_g for multicast group g, then assign [H_g] numbers of smallest negative coefficient of y_{gl} to 1 and 0 otherwise.

Step 4. If the number of negative coefficient is no greater than H_g but greater than max $\{h_g, |D_g|\}$ for multicast group g, then assign the corresponding negative coefficient of y_{gl} to 1 and 0 otherwise.

Step 5. If the number of negative coefficient is no greater than max $\{h_g, |D_g|\}$ for multicast group g, assign the corresponding negative coefficient of y_{gl} to 1. Then, assign [max $\{h_g, |D_g|\}$ − the number of negative coefficient of y_{gl}] smallest positive coefficients of y_{gl} to 1 and 0 otherwise.

Subproblem 1.3: (related to decision variable t_{gld}, f_l and C_l)

$$Z_{Sub1.3}(b,e,q,p,t,m)=$$
$$\min \sum_{l\in L}\Big[e_l(C_l)-\sum_{g\in G}\sum_{d\in D_g} b_{gld}t_{gld}+\sum_{g\in G}\sum_{d\in D_g} e_{gl}t_{gld}+\sum_{g\in G}\sum_{d\in D_g} q_{gld}t_{gld}Q_{gd}-p_l f_l+$$
$$\sum_{g\in G}\sum_{d\in D_g} t_{gd}t_{gld}F_l(f_l,C_l)+\sum_{g\in G}\sum_{d\in D_g} m_{gd}t_{gld}M_l(f_l,C_l)\Big]$$

subject to:

$t_{gld}=0$ or 1 $\quad\forall g\in G, d\in D_g, l\in L \quad$ (1.3)

$f_l\le C_l \quad \forall l\in L \quad$ (1.11)

$C_l\in A_l \quad \forall l\in L. \quad$ (1.12)

Subproblem 1.3 is complicated due to the coupling of b_{gld}, f_l and C_l. Since C_l is discrete, we can compute and compare all the results after $|A_l|$ iterations. Therefore, we only need to consider Subproblem 1.3 with the coupling of b_{gld} and f_l and it can be solved analytically [2] since C_l has been decided.

For each iteration, we first decompose Subproblem 1.3 into $|L|$ independent problems. For each link $l\in L$:

$$\min \left\{\sum_{g\in G}\sum_{d\in D_g}\left[(-b_{gld})+e_{gl}+q_{gld}Q_{gd}+t_{gd}F_l(f_l,C_l)+m_{gd}M_l(f_l,C_l)\right]t_{gld}\right\}+$$
$$-p_l f_l+e_l(C_l) \qquad \text{(Subproblem 1.3\_1)}$$

subject to $t_{gld}=0$ or 1, $f_l\le C_l$ and $C_l\in A_l$.

We define a set of break points of f_l as those points where $\left(-\sum_{g\in G}\sum_{d\in D_g} b_{gld}+\sum_{g\in G}\sum_{d\in D_g} e_{gl}+\sum_{g\in G}\sum_{d\in D_g} q_{gld}Q_{gd}+\sum_{g\in G}\sum_{d\in D_g} t_{gd}F_l(f_l,C_l)+\sum_{g\in G}\sum_{d\in D_g} m_{gd}M_l(f_l,C_l)\right)=0$ for each OD pair. These break points are sorted and denoted as $f_l^1, f_l^2, f_l^3,......, f_l^n$. We observe that when $f_l^i \le f_l \le f_l^{i+1}$, the value of t_{gld} remains constant for all OD pairs. Within the above interval, t_{gld} is 1 if $\left(-\sum_{g\in G}\sum_{d\in D_g} b_{gld}+\sum_{g\in G}\sum_{d\in D_g} e_{gl}+\sum_{g\in G}\sum_{d\in D_g} q_{gld}Q_{gd}+\sum_{g\in G}\sum_{d\in D_g} t_{gd}F_l(f_l,C_l)+\sum_{g\in G}\sum_{d\in D_g} m_{gd}M_l(f_l,C_l)\right)\le 0$ and is 0 otherwise.

Therefore, within an interval, the objective is only a function of f_l and minimum point within the interval can be found analytically. By extracting at most $|G||D_g|+1$ intervals, we can then find the global minimum point by comparing those local minimum points.

After all iterations, we can find the total minimum by comparing those global minimum points with different values of C_l.

Subproblem 1.4: (related to decision variable r_{gl})

$$Z_{Sub1.4}(q,p) = \min \sum_{g\in G}\sum_{l\in L}\left(-\sum_{d\in D_g} q_{gld} + p_l\right) r_{gl}$$

subject to:

$$r_{gl} \in [0, \max_{d\in D_g} Q_{gd}] \qquad \forall g \in G, l \in L \qquad (1.9)$$

$$\sum_{l\in L} r_{gl} \le R_g \qquad \forall g \in G. \qquad (1.15)$$

The algorithm to solve (Subproblem 1.4) is as follows:

Step 1. Sort all coefficients $-\sum_{d\in D_g} q_{gld} + p_l$ and set all values of r_{gl} to zero for all links on multicast group g.

Step 2. While the smallest coefficient is negative, set the corresponding r_{gl} as much as possible until constraint 1.15 can't be satisfied or the smallest coefficient is positive (including zero).

3.1.3 The Dual Problem and the Subgradient Method

According to the weak Lagrangean duality theorem [13], for any $e_{gl}, q_{gld}, t_{gd}, m_{gd} \geq 0$, $Z_{D1.1}(b, e, q, p, t, m)$ is a lower bound on $Z_{IP1.1}$. The following dual problem (D1.1) is then constructed to calculate the tightest lower bound.

Dual Problem (D1.1):

$$Z_{D1.1} = \max Z_{D1.1}(b, e, q, p, t, m)$$

subject to:

$$e_{gl}, q_{gld}, t_{gd}, m_{gd} \geq 0.$$

There are several methods to solve the dual problem (D1.1). Among them is the most popular method, the subgradient method, which is employed here [16]. Let a vector g be a subgradient of $Z_{D1.1}(b, e, q, p, t, m)$. Then, in iteration k of the subgradient optimization procedure, the multiplier vector is updated by $r^{k+1} = r^k + t^k g^k$. The step size t^k is determined by $t^k = d\frac{Z^h{}_{IP1.1} - Z_{D1.1}(r_k)}{\left\|g^k\right\|^2}$. $Z^h{}_{IP1.1}$ is the primal objective function value for a heuristic solution (an upper bound on $Z_{IP1.1}$). d is a constant, $0 < d \leq 2$.

3.2 Strategy II: Limited Budget

3.2.1 Solution Approach

By using the Lagrangean Relaxation method, we can transform the primal problem (IP2.1) into the following Lagrangean Relaxation problem (LR2.1) where Constraints (2.4), (2.5), (2.8), (2.10), (2.13), (2.14), (2.15) and (2.16) are relaxed.

3.2.2 Lagrangean Relaxation

For a vector of non-negative Lagrangean multipliers, a Lagrangean Relaxation problem of (IP2.1) is given by

Optimization problem (LR2.1):

$$
\begin{aligned}
Z_{D2.1}(b, e, q, p, f, t, m, s) = \min & \sum_{g\in G}\sum_{d\in D_g} -z_g k_{gd}(Q_{gd}, L_{gd}, J_{gd}) + \\
& \sum_{g\in G}\sum_{d\in D_g}\sum_{l\in L} b_{gld}\Big(\sum_{p\in P_{gd}} x_{gpd} d_{pl} - t_{gld}\Big) + \\
& \sum_{g\in G}\sum_{l\in L} e_{gl}\Big(\sum_{d\in D_g} t_{gld} - \left|D_g\right| y_{gl}\Big) + \\
& \sum_{g\in G}\sum_{d\in D_g}\sum_{l\in L} q_{gld}(t_{gld} Q_{gd} - r_{gl}) + \\
& \sum_{l\in L} p_l\Big(\sum_{g\in G} r_{gl} - f_l\Big) + \\
& f\Big[\sum_{l\in L} e_l(C_l) - B\Big] + \\
& \sum_{g\in G}\sum_{d\in D_g} t_{gd}\Big[\sum_{l\in L} t_{gld} F_l(f_l, C_l) - L_{gd}\Big] + \\
& \sum_{g\in G}\sum_{d\in D_g} m_{gd}\Big[\sum_{l\in L} t_{gld} M_l(f_l, C_l) - J_{gd}\Big] + \\
& \sum_{g\in G}\sum_{d\in D_g} s_{gd}\Big(z_g - \sum_{p\in P_{gd}} x_{gpd}\Big)
\end{aligned}
\qquad \text{(LR 2.1)}
$$

subject to:

$$x_{gpd} = 0 \text{ or } 1 \qquad \forall d \in D_g,\ g \in G,\ p \in P'_{gd} \tag{2.1}$$

$$y_{gl} = 0 \text{ or } 1 \qquad \forall g \in G, l \in L \tag{2.2}$$

$$t_{gld} = 0 \text{ or } 1 \qquad \forall g \in G, d \in D_g, l \in L \tag{2.3}$$

$$\sum_{l \in L} y_{gl} \geq \max\{h_g, |D_g|\} \qquad \forall g \in G \tag{2.6}$$

$$\sum_{p \in P'_{gd}} x_{gpd} = 1 \qquad \forall d \in D_g, g \in G \tag{2.7}$$

$$r_{gl} \in [0, \max_{d \in D_g} Q_{gd}] \qquad \forall g \in G, l \in L \tag{2.9}$$

$$f_l \leq C_l \qquad \forall l \in L \tag{2.11}$$

$$C_l \in A_l \qquad \forall l \in L \tag{2.12}$$

$$z_g = 0 \text{ or } 1 \qquad \forall g \in G \tag{2.17}$$

$$\sum_{l \in L} r_{gl} \leq R_g \qquad \forall g \in G \tag{2.18}$$

$$\sum_{l \in L} y_{gl} \leq H_g \qquad \forall g \in G \tag{2.19}$$

where $b_{gld}, e_{gl}, q_{gld}, p_l, f, t_{gd}, m_{gd}, s_{gd}$ are Lagrangean multipliers and $e_{gl}, q_{gld}, f, t_{gd}, m_{gd} \geq 0$. To solve (LR2.1), we can decompose (LR2.1) into the following five independent and easily solvable optimization subproblems.

Subproblem 2.1: (related to decision variable x_{gpd})

$$Z_{Sub2.1}(b, s) = \min \sum_{g \in G} \sum_{d \in D_g} \sum_{p \in P_{gd}} (\sum_{l \in L} b_{gld} d_{pl} - s_{gd}) x_{gpd}$$

subject to:

$$x_{gpd} = 0 \text{ or } 1 \qquad \forall d \in D_g,\ g \in G,\ p \in P'_{gd} \qquad (2.1)$$

$$\sum_{p \in P'_{gd}} x_{gpd} = 1 \qquad \forall d \in D_g, g \in G. \qquad (2.7)$$

The subproblem can be further decomposed into $|G||D_g|$ independent shortest path problems with nonnegative arc weights. Each shortest path problem can be easily solved by the Dijkstra's algorithm, the link cost is b_{gld}. If $b_{gld} - s_{gd}$ is negative, x_{gpd} is 1. Otherwise x_{gpd} is 0.

Subproblem 2.2: (related to decision variable y_{gl})

$$Z_{Sub2.2}(e) = \min \ -\sum_{g \in G} \sum_{l \in L} e_{gl} |D_g| y_{gl}$$

subject to:

$$y_{gl} = 0 \text{ or } 1 \qquad \forall g \in G, l \in L \qquad (2.2)$$

$$\sum_{l \in L} y_{gl} \ge \max\{h_g, |D_g|\} \qquad \forall g \in G \qquad (2.6)$$

$$\sum_{l \in L} y_{gl} \le H_g \qquad \forall g \in G. \qquad (2.19)$$

The algorithm to solve (Subproblem 2.2) is stated as follows[1]:

Step 1. Compute max $\{h_g, |D_g|\}$ for multicast group g.

Step 2. Compute the number of negative coefficient $-e_{gl}|D_g|$ for all links on multicast group g.

Step 3. If the number of negative coefficient is greater than H_g for multicast group g, then assign [H_g] numbers of smallest negative coefficient of y_{gl} to 1 and 0 otherwise.

Step 4. If the number of negative coefficient is no greater than H_g but greater than max $\{h_g, |D_g|\}$ for multicast group g, then assign the corresponding negative coefficient of y_{gl} to 1 and 0 otherwise.

Step 5. If the number of negative coefficient is no greater than max $\{h_g, |D_g|\}$ for multicast group g, then assign the corresponding negative coefficient of y_{gl} to 1. Then, assign [max $\{h_g, |D_g|\}$ − the number of negative coefficient of y_{gl}] numbers of smallest positive coefficient of y_{gl} to 1 and 0 otherwise.

Subproblem 2.3: (related to decision variable t_{gld}, f_l and C_l)

$$Z_{Sub2.3}(b,e,q,p,t,m)=$$

$$\min \sum_{l\in L}\Big[f\,e_l(C_l)-\sum_{g\in G}\sum_{d\in D_g} b_{gld}t_{gld}+\sum_{g\in G}\sum_{d\in D_g} e_{gl}\,t_{gld}+\sum_{g\in G}\sum_{d\in D_g} q_{gld}t_{gld}Q_{gd}-p_l f_l+$$
$$\sum_{g\in G}\sum_{d\in D_g} t_{gd}t_{gld}F_l(f_l,C_l)+\sum_{g\in G}\sum_{d\in D_g} m_{gd}t_{gld}M_l(f_l,C_l)\Big]$$

subject to:

$t_{gld}=0$ or 1	$\forall g\in G, d\in D_g, l\in L$	(2.3)
$f_l\le C_l$	$\forall l\in L$	(2.11)
$C_l\in A_l$	$\forall l\in L$.	(2.12)

Subproblem 2.3 is complicated due to the coupling of b_{gld}, f_l and C_l. Since C_l is discrete and finite, we can compute and compare all the results after $|A_l|$ iterations. Therefore, we only need to consider Subproblem 2.3 with the coupling of b_{gld} and f_l and it can be solved analytically [2] since C_l has been decided:

In an iteration, we first decompose Subproblem 2.3 into $|L|$ independent problems. For each link $l\in L$:

$$\min\ f\,e_l(C_l)-\sum_{g\in G}\sum_{d\in D_g} b_{gld}t_{gld}+\sum_{g\in G}\sum_{d\in D_g} e_{gl}\,t_{gld}+\sum_{g\in G}\sum_{d\in D_g} q_{gld}t_{gld}Q_{gd}-p_l f_l+$$
$$\sum_{g\in G}\sum_{d\in D_g} t_{gd}t_{gld}F_l(f_l,C_l)+\sum_{g\in G}\sum_{d\in D_g} m_{gd}t_{gld}M_l(f_l,C_l) \qquad \text{(Subproblem 2.3\_1)}$$

subject to $t_{gld}=0$ or 1, $f_l\le C_l$ and $C_l\in A_l$.

We define a set of break points of f_l as those points where

$$\left(-\sum_{g\in G}\sum_{d\in D_g} b_{gld}+\sum_{g\in G}\sum_{d\in D_g} e_{gl}+\sum_{g\in G}\sum_{d\in D_g} q_{gld}Q_{gd}+\sum_{g\in G}\sum_{d\in D_g} t_{gd}F_l(f_l,C_l)+\sum_{g\in G}\sum_{d\in D_g} m_{gd}M_l(f_l,C_l)\right)=0$$

for each OD pair. These break points are sorted and denoted as $f_l^1, f_l^2, f_l^3,......, f_l^n$. We observe that when $f_l^i \le f_l \le f_l^{i+1}$, the value of t_{gld} remains constant for all OD pairs. Within the above interval, t_{gld} is 1 if $\left(-\sum_{g\in G}\sum_{d\in D_g} b_{gld}+\sum_{g\in G}\sum_{d\in D_g} e_{gl}+\sum_{g\in G}\sum_{d\in D_g} q_{gld}Q_{gd}+\sum_{g\in G}\sum_{d\in D_g} t_{gd}F_l(f_l,C_l)+\sum_{g\in G}\sum_{d\in D_g} m_{gd}M_l(f_l,C_l)\right)\le 0$ and is 0 otherwise.

Therefore, within an interval, the objective is only a function of f_l and minimum point within the interval can be found analytically. By extracting at most $|G||D_g|+1$ intervals, we can then find the global minimum point by comparing those local minimum points.

After all iterations, we can find the total minimum by comparing those global minimum points with different values of C_l.

Subproblem 2.4: (related to decision variable r_{gl})

$$Z_{Sub2.4}(q,p)=\min \sum_{l\in L}\sum_{g\in G}\left(-\sum_{d\in D_g} q_{gld}+p_l\right)r_{gl}$$

subject to:

$$r_{gl}\in[0,\max_{d\in D_g} Q_{gd}] \qquad \forall g\in G, l\in L \qquad (2.9)$$

$$\sum_{l\in L} r_{gl}\le R_g \qquad \forall g\in G. \qquad (2.18)$$

The algorithm to solve Subproblem 2.4 is as follows:

Step 1. Sort all coefficients $-\sum_{d\in D_g} q_{gld} + p_l$ and set all values of r_{gl} to zero for all links on multicast group g.

Step 2. While the smallest coefficient is negative, set the corresponding r_{gl} as much as possible until constraint 2.18 can't be satisfied or the smallest coefficient is positive (including zero).

Subproblem 2.5: (related to decision variable z_g)

$$Z_{Sub2.5}(s) = \min \sum_{g\in G} \sum_{d\in D_g} \left[-k_{gd}(Q_{gd}, L_{gd}, J_{gd}) + s_{gd}\right] z_g$$

subject to:

z_g =0 or 1 $\quad \forall g \in G$. (2.17)

The Subproblem is to determine z_g . There are two cases to consider:

Case 1. If $\sum_{d\in D_g} \left[-k_{gd}(Q_{gd}, L_{gd}, J_{gd}) + s_{gd}\right] \geq 0$, then $z_g = 0$.

Case 2. If $\sum_{d\in D_g} \left[-k_{gd}(Q_{gd}, L_{gd}, J_{gd}) + s_{gd}\right] < 0$, then $z_g = 1$.

3.2.3 The Dual Problem and the Subgradient Method

According to the weak Lagrangean duality theorem [13], for any $e_{gl}, q_{gld}, f, t_{gd}, m_{gd} \geq 0$, $Z_{D2.1}(b, e, q, p, f, t, m, s)$ is a lower bound on $Z_{IP2.1}$. The following dual problem (D2.1) is then constructed to calculate the tightest lower bound.

Dual Problem (D2.1):

$$Z_{D2.1} = \max Z_{D2.1}(b, e, q, p, f, t, m, s)$$

subject to: $e_{gl}, q_{gld}, f, t_{gd}, m_{gd} \geq 0$.

There are several methods to solve the dual problem (D2.1). Among them is the most popular method, the subgradient method, which is employed here [16]. Let a vector g be a subgradient of $Z_{D2.1}(b, e, q, p, f, t, m, s)$. Then, in iteration k of the subgradient optimization procedure, the multiplier vector $r = (b, e, q, p, f, t, m, s)$ is updated by $r^{k+1} = r^k + t^k g^k$. The step size t^k is determined by $t^k = d \dfrac{Z^h{}_{IP2.1} - Z_{D2.1}(r_k)}{\left\| g^k \right\|^2}$. $Z^h{}_{IP2.1}$ is the primal objective function value for a heuristic solution (an upper bound on $Z_{IP2.1}$). d is a constant, $0 < d \leq 2$.

Chapter 4 Getting Primal Feasible Solutions

By using Lagrangean relaxation and the subgradient method as our tools to solve these problems, we can get not only a theoretical lower bound of primal feasible solution, but also some hints to help us get our primal feasible solution under each solving dual problem iteration.

But we cannot guarantee that the results of Lagrangean dual problems will be a feasible solution to the primal optimization problem, since there are some constraints relaxed by Lagrangean relaxation.

If the decision variables calculated happen to satisfy the relaxed constraints, then a primal feasible solution is found. Otherwise, the modification on such infeasible primal solutions must be necessary and have to be made to obtain primal feasible solutions.

4.1 Heuristics for Strategy 1: Cost Down

To acquire primal feasible solutions for the cost-down model, solutions to the Lagrangean Relaxation problems are considered:

- The solution set of $\{x_{gpd}\}$ obtained after solving (Subproblem 1.1) may not be a feasible solution to the primal problem (IP 1.1) because of the capacity constraint being relaxed. There may be some links that the aggregate flows on them violate the capacity constraint.
- The solution set of $\{y_{gl}\}$ obtained after solving (Subproblem 1.2) may not be a feasible solution. The solution set possibly cause the union of $\{y_{gl}\}$ not to be a tree.
- The solution sets $\{t_{gld}\}$, $\{f_l\}$, $\{C_l\}$ obtained after solving (Subproblem 1.3) may not be a feasible solution. Since the QoS constraints are relaxed. The total delay from the source to the destination may exceed the user's delay requirement.

Therefore, it is necessary to apply additional heuristics so as to obtain a primal feasible solution. In this section, the detail of the heuristics described.

4.1.1 Multicast Routing Subproblems

Heuristic 4.1.1

The set of $\{x_{gpd}\}$ is used to decide the routing of each OD pair in a multicast group. To solve routing subproblems, we can use the method of solving shortest path problems like Dijkstra's algorithm to find the shortest path of each OD pair, or we can use the method of solving minimum spanning tree problem like Prim's or Sollin's algorithms to find a multicast tree for a multicast group. And while solving the Lagrangean relaxation dual problem, we have multipliers related to ach OD pair and each link. By using them, we can avoid using highly loaded link as some links underestimated or overestimated. So, there are several options about how to decide which multipliers to represent the arc weights of the links. And we use three kinds of multipliers to assign the relative arc weight as follow:

i. For each multicast group, link *l*'s arc weight is equal to $\sum_{d \in D_g} b_{gld}$.

ii. For each multicast group, link *l*'s arc weight is equal to $\sum_{d \in D_g} \{ b_{gld} \times$ (Each OD pair's Traffic Demand)}.

iii. For all multicast groups, link *l*'s arc weight is equal to p_l .

Furthermore, we can also use the set of $\{C_l\}$ generated by the result of (Subproblem 1.3) to only take into account the links whose capacity is not zero after solving (Subproblem 1.3) and once again apply three types of arc weights on them.

Therefore, we have 6 possible heuristics to setup arc weights and 2 type of routing (shortest path and minimum spanning tree) as follows:

Algorithm 4.1.1

Step 1.	*Option 1:* Take into account all links. *Option 2:* Take into account only the links with nonzero capacity after solving (Subproblem 1.3).
Step 2.	Decide a set of arc weights of the links: *Option 1:* For group g, link l's arc weight $= \sum_{d \in D_g} b_{gld}$. *Option 2:* For group g, link l's arc weight $=$ $\sum_{d \in D_g} \{ b_{gld} \times$ (Each OD pair's Traffic Demand)$\}$. *Option 3:* Link l's arc weight $= p_l$.
Step 3.	*Option 1:* Run Dijkstra algorithm to determine the routing of each path for the OD pairs of each multicast group. *Option 2:* Run Sollin's algorithm to determine the routing of each multicast group.

Table 4-1. Solving the Routing Subproblem of Model 1

And we give notations to represent each heuristic as follows:

	Option in Step 1	**Option in Step 2**	**Option in Step 3**
BETA+SP	1	1	1
BETA+MST	1	1	2
TD+SP	1	2	1
TD+MST	1	2	2
PI+SP	1	3	1
PI+MST	1	3	2
Cl+BETA+SP	2	1	1

Cl+BETA+MST	2	1	2
Cl+TD+SP	2	2	1
Cl+TD+MST	2	2	2
Cl+PI+SP	2	3	1
Cl+PI+MST	2	3	2

Table 4-2. Notations of 12 Routing Heuristics for Algorithm 4.1.1

4.1.2 Capacity Assignment Subproblems

Heuristic 4.1.2

After deciding each OD pair's route, we can calculate the aggregate flows on each link according to the OD pairs, which pass through it by aggregating their traffic demands. Since knowing each link's traffic flow, we should figure out the optimum assignment of link capacity, which makes the leasing cost minimum under the QoS constraints (Constraint 1.13, 1.14) and the capacity constraint (Constraint 1.11). Aimed at the characteristics about the discrete distribution of each link capacity, we can relax it and suppose that they are continuously distributed. Since the leasing cost, delay and delay jitter functions are convex, it's a convex programming problem, which can be solved by using the penalty method or the Lagrangean multiplier method.

Algorithm 4.1.2

Step 1.	Calculate each link's traffic flow by aggregating those passing OD pairs' bandwidth requirements.
Step 2.	Use the Lagrangean multiplier method to find the setting of capacity for each link minimizing the total leasing cost with the QoS and capacity constraints relaxed, and calculate the total leasing cost.

Table 4-3. Solving the Capacity Assignment Subproblem of Model 1

4.2 Heuristics for Strategy 2: Limited Budget

To acquire primal feasible solutions for the cost-down model, solutions to the Lagrangean Relaxation problems are considered:

- ❖ The set of $\{x_{gpd}\}$ obtained after solving (Subproblem 2.1) may not be a feasible solution to the primal problem (IP 2.1) because of the capacity constraint being relaxed. There may be some links that the aggregate flows on them violate the capacity constraint.
- ❖ The set of $\{y_{gl}\}$ obtained after solving (Subproblem 2.2) may not be a feasible solution. The solution set possibly cause the union of $\{y_{gl}\}$ not to be a tree.

- The solution sets $\{t_{gld}\}$, $\{f_l\}$, $\{C_l\}$ obtained after solving (Subproblem 2.3) may not be a feasible solution. Since the QoS constraints and budget constraint are relaxed. The total delay from the source to the destination may exceed the user's delay requirement, and total leasing cost may exceed the budget.
- The set of $\{z_g\}$ obtained after solving (Subproblem 2.5) may not be a feasible solution. The set of $\{z_g\}$ may be not consistent with the set of $\{x_{gpd}\}$ acquired after solving (Subproblem 2.1) because of violating (Constraint 2.16).

Therefore, it is necessary to apply additional heuristics so as to obtain a primal feasible solution. In this section, the detail of the heuristics described.

4.2.1 Multicast Routing Subproblems

Heuristic 4.2.1

The set of $\{x_{gpd}\}$ is used to decide the routing of each OD pair in a multicast group. To solve routing subproblems, we can use the method of solving shortest path problems like Dijkstra's algorithm to find the shortest path of each OD pair, or we can use the method of solving minimum spanning tree problem like Prim's or Sollin's algorithms to find a multicast tree for a multicast group. And while solving the Lagrangean relaxation dual problem, we have multipliers related to ach OD pair and each link. By using them, we can avoid using highly loaded link as some links underestimated or overestimated. So, there are several options about how to decide which multipliers to represent the arc weights of the links. And we use three kinds of multipliers to assign the relative arc weight as follow:

i. For each multicast group, link l's arc weight is equal to $\sum_{d \in D_g} b_{gld}$.

ii. For each multicast group, link l's arc weight is equal to $\sum_{d \in D_g} \{b_{gld} \times$ (Each OD pair's Traffic Demand)$\}$.

iii. For all multicast groups, link l's arc weight is equal to p_l.

Furthermore, we can also use the set of $\{C_l\}$ generated by the result of (Subproblem 1.3) to only take into account the links whose capacity is not zero after solving (Subproblem 1.3) and once again apply three types of arc weights on them.

Therefore, we have 6 possible heuristics to setup arc weights and 2 type of routing (shortest path and minimum spanning tree) as follows:

Algorithm 4.2.

Step 1.	*Option 1:* Take into account all links. *Option 2:* Take into account only the links with nonzero capacity after solving (Subproblem 1.3).
Step 2.	Decide a set of arc weights of the links: *Option 1:* For group g, link l's arc weight $= \sum_{d \in D_g} b_{gld}$. *Option 2:* For group g, link l's arc weight $=$ $\sum_{d \in D_g} \{b_{gld} \times$ (Each OD pair's Traffic Demand)$\}$. *Option 3:* Link l's arc weight $= p_l$.

Step 3.	*Option 1:* Run Dijkstra algorithm to determine the routing of each path for the OD pairs of each multicast group. *Option 2:* Run Sollin's algorithm to determine the routing of each multicast group.

Table 4-4. Solving the Routing Subproblem of Model 2

And we use the notations the same as (Algorithm 4.1.1):

	Option in Step 1	**Option in Step 2**	**Option in Step 3**
BETA+SP	1	1	1
BETA+MST	1	1	2
TD+SP	1	2	1
TD+MST	1	2	2
PI+SP	1	3	1
PI+MST	1	3	2
Cl+BETA+SP	2	1	1
Cl+BETA+MST	2	1	2
Cl+TD+SP	2	2	1
Cl+TD+MST	2	2	2
Cl+PI+SP	2	3	1
Cl+PI+MST	2	3	2

Table 4-5. Notations of 12 Routing Heuristics for Algorithm 4.2.1

4.2.2 Multicast Admission Control Subproblems

Heuristic 4.2.2

After deciding each OD pair's route, we have to think about which groups should be admitted. At first, we can use the result of (Subproblem 2.5); those groups whose $z_g = 1$ are admitted first. Then calculate their aggregate flows on each link according to their OD pairs. Knowing each link's traffic flow, we use the Lagrangean multiplier method with the QoS and capacity constraints relaxed to check whether admitting these groups cause the violation of the budget constraint or not. If not, we can put more multicast groups to the networks. Otherwise, we should drop some admitted groups by some specific order we define. Here we use the subgradients of z_g used in (Subproblem 2.5) to represent the order. Adding groups starts from the group whose subgradient is the smallest, while dropping from the largest.

Algorithm 4.2.2

Step 1.	Take into account the groups with $z_g = 1$ after (Subproblem 2.5) and calculate each link's according flow by aggregating their passing OD pairs' bandwidth requirements.
Step 2.	Use the Lagrangean multiplier method with the QoS and capacity constraints relaxed to check whether admitting these groups will violate the budget constraint. If there is no budget-constraint violation, go to Step 4. Otherwise, go to Step 3.

Step 3.	Take into account all the groups with $z_g = 1$. Sum up and sort their subgradients ($\sum_{d \in D_g} \left[s_{gd} - k_{gd}(Q_{gd}, L_{gd}, J_{gd}) \right]$). Drop the group with the largest subgradient, and go to Step 2.
Step 4.	Take into account all the groups with $z_g = 0$. Sum up and sort their subgradients ($\sum_{d \in D_g} \left[s_{gd} - k_{gd}(Q_{gd}, L_{gd}, J_{gd}) \right]$). Add the group with the smallest subgradient and run budget checking by Lagrangean relaxation. If passed, admit the group (set $z_g = 1$). Keep on trying to add until all the groups with $z_g = 0$ are tested.
Step 5.	Calculate the total revenue while $z_g = 1$.

Table 4-6. Solving the Admission Control Subproblem of Model 2

Chapter 5 Computational Experiments

For the purpose of showing the difference between the results from our Lagrangean relaxation method and other primal heuristics, we implement two simple algorithms to compare with our heuristics. With the comparison of the results, we can not only examine the quality of the primal heuristics, but also get some implications from the Lagrangean multipliers to find a feasible solution.

5.1 Simple Algorithm for Strategy 1: Cost Down

Algorithm 5.1

Step 1.	Let each link's arc weight equal to (1 / the sum of its connected nodes' degrees).
Step 2.	According the link set metrics, every user group construct its multicast tree.
Step 3.	Calculate the aggregate flow on each link according to the multicast requirements. And set each link capacity to its aggregate flow plus one capacity unit.

Step 4.	Check all OD pairs. If the delay constraint or delay jitter constraint is violated. Find the link with the most OD pairs crossed by and increase its link capacity until the OD pairs' QoS constraints are satisfied.

Table 5-1. Simple Algorithm for Model 1

5.2 Simple Algorithm for Strategy 2: Limited Budget

Algorithm 5.2

Step 1.	Let each link's arc weight equal to (1 / the sum of its connected nodes' degrees).
Step 2.	According the link set metrics, every user group construct its multicast tree.
Step 3.	Calculate the aggregate flow on each link according to the multicast requirements. And set each link capacity to its aggregate flow plus one capacity unit.
Step 4.	Check the leasing cost. If it satisfies the budget constraint, begin to do the adding tests. Otherwise, drop the user group as the sequence in their group ID until satisfying the budget constraint.

Table 5-2. Simple Algorithm for Model 2

5.3 Assumptions, Parameters, and Cases

Number of Nodes	6~21
Cost Unit	5
Number of Iteration	10000
Maximum Unimprovement Counter	100
Begin to Tune	300
Initial Upper Bound	Cost of Leasing Maximum Capacity
Initial Scalar of Step Size	2
Test Platform	Windows 2000, 2G Hz CPU, 1G RAM

Table 5-3. Command Testing Parameters for Model 1

Number of Nodes	6~21
Cost Unit	10
Number of Iteration	2000
Maximum Unimprovement Counter	80
Begin to Tune	100
Initial Upper Bound	0
Initial Scalar of Step Size	2
Test Platform	Windows 2000, 2G Hz CPU, 1G RAM

Table 5-4. Command Testing Parameters for Model 2

These models and algorithms is written in ANSI C and is compiled by Microsoft® Visual C++ 6.0.

For the model 1 (Strategy 1), it is flexible to reduce the number of iterations in program in some special cases. In the implementation, $Z^h_{IP1.1}$ is initially chosen as the maximum leasing cost when all links choose the maximum capacity choice. The choice of the initial values of the multipliers is 1.0, except 0.0 for the multipliers b_{gld} .

For the model 2 (Strategy 2), it is also flexible to reduce the number of iterations in program in some special cases. In the implementation, $Z^h{}_{IP2.1}$ is initially chosen as 0, which means the worst case of rejecting all user groups. The choice of the initial values of the multipliers is 0.1.

We have tested the algorithms on three networks – Mesh, GTE, and NSF with 9, 12, and 14 nodes. These topologies are as shown in figure 5.1, 5.2, and 5.3. Representative results have been selected for the purpose of demonstration.

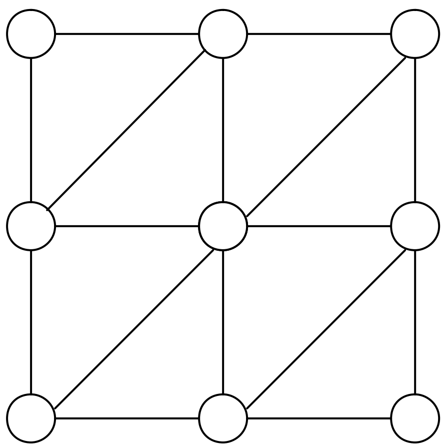

Figure 5-1 Mesh Network: 9 nodes, 16 links

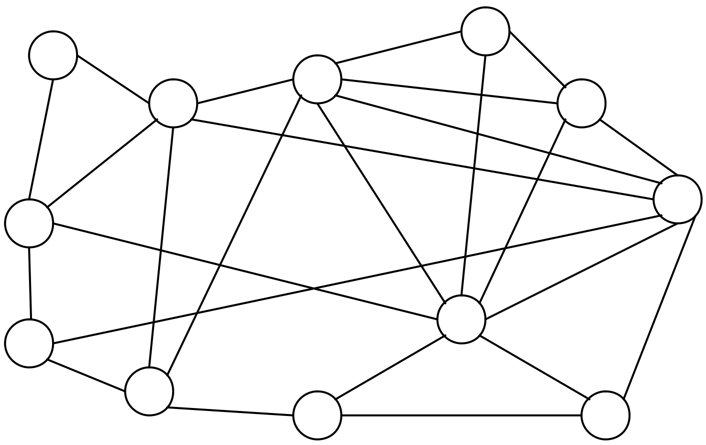

Figure 5-2 GTE Network: 12 nodes, 25 links

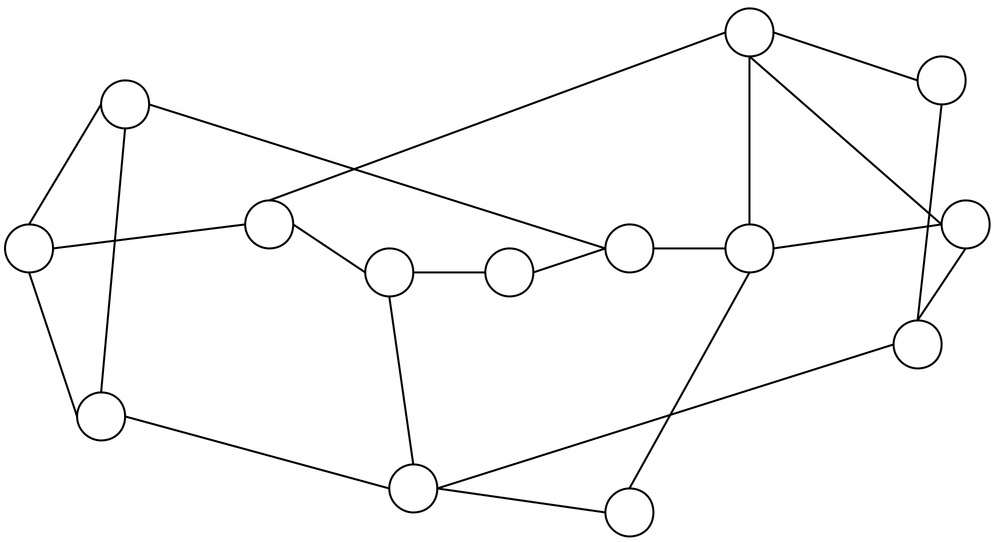

Figure 5-3 NSF Network: 14 nodes, 21 links

5.4 Heuristic Comparisons for Getting Primal Feasible Solutions

Before the implementation of the experiments, there are 12 heuristics for each model to determine the multicast routing, and it's only necessary to find the better one to do the experiments. Therefore, we run two cases by the 12 heuristics, and choose the heuristic, which have the better result.

5.4.1 Model 1: Cost Down

Case 1:

Mesh Network (9 nodes, 16 links)
Number of Requested Multicast Group: 6
Range of Requested Minimum Bandwidth: 1 ~ 5 (Mbps)
Range of Requested Maximum Delay: 1 ~ 10 (sec)
Range of Requested Maximum Delay Jitter: 1 ~ 5 (sec)
Number of Iterations: 10000
Test Result:

	UB	LB	Gap
Cl+PI+SP	98.00	82.649300	18.57%
Cl+TD+SP	98.00	82.485489	18.81%
Cl+BETA+SP	100.00	82.449066	21.29%
PI+MST	156.00	83.786812	86.19%
PI+SP	162.00	83.706108	93.53%
BETA+SP	170.00	84.121361	102.09%
TD+SP	176.00	83.690132	110.30%
BETA+MST	268.00	84.316429	217.85%
TD+MST	270.00	83.983856	221.49%

Table 5-5. Case 1 of 12 Heuristic Comparisons for Model 1

And we find out when we use option 2 in Algorithm 4.1.1, then we cannot apply Sollin's method to find a minimum spanning tree since some nodes could be unconnected after the result of (Subproblem 1.3).

Case 2:

GTE Network (12 nodes, 25 links)
Number of Requested Multicast Group: 12
Range of Requested Minimum Bandwidth: 10 ~ 30 (Mbps)
Range of Requested Maximum Delay: 0.5 ~ 3.5 (sec)
Range of Requested Maximum Delay Jitter: 0.5 ~ 1.5 (sec)
Number of Iterations: 20000
Test Result:

	UB	LB	Gap
Cl+PI+SP	6235.00	5404.723145	15.36%
Cl+TD+SP	6235.00	5334.316895	16.88%
Cl+BETA+SP	6235.00	5258.529297	18.57%
PI+MST	10740.00	5715.932617	87.90%
PI+SP	11110.00	5596.846680	98.50%
BETA+SP	11210.00	5488.998047	104.23%
TD+SP	11320.00	5476.857414	106.68%
BETA+MST	15145.00	5491.164263	175.80%
TD+MST	15320.00	5479.723248	179.57%

Table 5-6. Case 2 of 12 Heuristic Comparisons for Model 1

From the pretest results, it can be obviously observed that Cl+PI+SP heuristic is the better choice of 12 ones. That is, we take into account only the links with nonzero capacity after solving (Subproblem 1.3), let link l's arc weight $= p_l$, and run Dijkstra's algorithm to determine the routing of each multicast group.

5.4.1 Model 2: Limited Budget

Case 1:

Mesh Network (9 nodes, 16 links)
Budget: 3000
Number of Requested Multicast Group: 9
Range of Requested Minimum Bandwidth: 1 ~ 10 (Mbps)
Range of Requested Maximum Delay: 1 ~ 10 (sec)
Range of Requested Maximum Delay Jitter: 1 ~ 3 (sec)
Test Result:

	UB	LB	Gap
PI+MST	-4691.11	-4784.066406	1.98%
Cl+PI+SP	-4691.11	-4881.750488	4.06%
Cl+TD+SP	-4691.11	-4881.750488	4.06%
Cl+BETA+SP	-4691.11	-4881.750488	4.06%
PI+SP	-4691.11	-4881.750488	4.06%
TD+SP	-4691.11	-4881.750488	4.06%
BETA+SP	-4691.11	-4881.750488	4.06%
TD+MST	-4284.44	-4744.225586	10.73%
BETA+MST	-4284.44	-4749.038086	10.84%

Table 5-7. Case 1 of 12 Heuristic Comparisons for Model 2

Case 2:

GTE Network (12 nodes, 25 links)
Budget: 5000
Number of Requested Multicast Group: 12
Range of Requested Minimum Bandwidth: 5 ~ 20 (Mbps)
Range of Requested Maximum Delay: 1 ~ 10 (sec)
Range of Requested Maximum Delay Jitter: 1 ~ 3 (sec)
Test Result:

	UB	LB	Gap
PI+MST	-7583.70	-7675.221680	1.21%
Cl+PI+SP	-7583.70	-7675.961426	1.22%
Cl+TD+SP	-7583.70	-7675.961426	1.22%
Cl+BETA+SP	-7583.70	-7675.961426	1.22%
PI+SP	-7583.70	-7675.961426	1.22%
TD+SP	-7583.70	-7675.961426	1.22%
BETA+SP	-7583.70	-7675.961426	1.22%
TD+MST	-6989.63	-7669.408203	9.73%
BETA+MST	-6989.63	-7689.936035	10.02%

Table 5-8. Case 2 of 12 Heuristic Comparisons for Model 2

From the pretest results, it can be obviously observed that PI+MST heuristic is the better choice of 12 ones. That is, we take into account all links, let link *l*'s arc weight $= p_l$, and run Sollin's algorithm to determine the routing of each multicast group.

On the whole, the considered links sifted by (Subproblem 1.3 & 2.3) have better effects with lower upper bounds. But in the model 2, the best choice is to consider all links and Sollin's algorithm is applied instead of Dijkstra's. It is because admission control doesn't necessarily accept all user groups in the model 2. And under this condition, the minimum spanning tree algorithm has the better performance than the shortest path spanning tree algorithm.

5.5 Experiment Results

5.5.1 Experiment Results of Strategy 1

Case 1

Network topology: Mesh Network

Number of Requested Multicast Group：6

Traffic Range	Delay Range	Delay Jitter Range	SA	LR	Lower Bound	Gap
3.0 ~ 10.0	1.0 ~10.0	1.0 ~3.0	695	350	350	0.00%
10.0 ~ 30.0	0.5 ~5.5	0.5 ~1.5	2125	1030	1021	0.88%

Table 5-9. The Result of Case 1 for Model 1

Case 2

Network topology: Mesh Network

Number of Requested Multicast Group：9

Traffic Range	Delay Range	Delay Jitter Range	SA	LR	Lower Bound	Gap
3.0 ~ 10.0	1.0 ~10.0	1.0 ~3.0	970	480	480	0.00%
10.0 ~ 30.0	0.5 ~5.5	0.5 ~1.5	3175	1520	1520	0.00%

Table 5-10. The Result of Case 2 for Model 1

Case 3

Network topology: Mesh Network

Number of Requested Multicast Group：12

Traffic Range	Delay Range	Delay Jitter Range	SA	LR	Lower Bound	Gap
3.0 ~ 10.0	1.0 ~10.0	1.0 ~3.0	1445	760	760	0.00%
10.0 ~ 30.0	0.5 ~5.5	0.5 ~1.5	4445	2235	2235	0.00%

Table 5-11. The Result of Case 3 for Model 1

Case 4

Network topology: GTE Network

Number of Requested Multicast Group：9

Traffic Range	Delay Range	Delay Jitter Range	SA	LR	Lower Bound	Gap
3.0 ~ 10.0	1.0 ~10.0	1.0 ~3.0	1385	800	615	30.16%
10.0 ~ 30.0	0.5 ~5.5	0.5 ~1.5	4670	2555	1262	102.53%

Table 5-12. The Result of Case 4 for Model 1

Case 5

Network topology: GTE Network

Number of Requested Multicast Group：12

Traffic Range	Delay Range	Delay Jitter Range	SA	LR	Lower Bound	Gap
3.0 ~ 10.0	1.0 ~10.0	1.0 ~3.0	2050	1080	841	28.37%
10.0 ~ 30.0	0.5 ~5.5	0.5 ~1.5	6615	3370	1461	130.73%

Table 5-13. The Result of Case 5 for Model 1

Case 6

Network topology: GTE Network

Number of Requested Multicast Group : 20

Traffic Range	Delay Range	Delay Jitter Range	SA	LR	Lower Bound	Gap
3.0 ~ 10.0	1.0 ~10.0	1.0 ~3.0	3550	1775	1343	32.16%
10.0 ~ 30.0	0.5 ~5.5	0.5 ~1.5	11185	5320	2001	165.82%

Table 5-14. The Result of Case 6 for Model 1

Case 7

Network topology: NSF Network

Number of Requested Multicast Group : 7

Traffic Range	Delay Range	Delay Jitter Range	SA	LR	Lower Bound	Gap
3.0 ~ 10.0	1.0 ~10.0	1.0 ~3.0	1575	520	519	0.19%
10.0 ~ 30.0	0.5 ~5.5	0.5 ~1.5	4725	1575	1293	21.79%

Table 5-15. The Result of Case 7 for Model 1

Case 8

Network topology: NSF Network

Number of Requested Multicast Group : 14

Traffic Range	Delay Range	Delay Jitter Range	SA	LR	Lower Bound	Gap
3.0 ~ 10.0	1.0 ~10.0	1.0 ~3.0	2640	850	849	0.11%
10.0 ~ 30.0	0.5 ~5.5	0.5 ~1.5	8275	2575	2380	8.17%

Table 5-16. The Result of Case 8 for Model 1

Case 9

Network topology: NSF Network

Number of Requested Multicast Group：21

Traffic Range	**Delay Range**	**Delay Jitter Range**	**SA**	**LR**	**Lower Bound**	**Gap**
3.0 ~ 10.0	1.0 ~10.0	1.0 ~3.0	3905	1330	1330	0.00%
10.0 ~ 30.0	0.5 ~5.5	0.5 ~1.5	12170	4080	3549	14.97%

Table 5-17. The Result of Case 9 for Model 1

Although some results are not quite good. But if we increase the number of iterations, the result is being better. Like the case 4 for the model 1 with the higher QoS requirements, we increase the number of iterations to twenty thousand. And the result is as follows:

Case 4

Network topology: GTE Network

Number of Requested Multicast Group：9

Traffic Range	**Delay Range**	**Delay Jitter Range**	**SA**	**LR**	**Lower Bound**	**Gap**	**Iteration**
10.0 ~ 30.0	0.5 ~5.5	0.5 ~1.5	4670	2555	1262	102.53%	10000
10.0 ~ 30.0	0.5 ~5.5	0.5 ~1.5	4670	2555	1360	87.87%	20000

Table 5-18. Comparison with Different Numbers of Iterations

5.5.2 Experiment Results of Strategy 2

Case 1

Network Topology: Mesh Network

Number of Requested Multicast Group : 6

Traffic Range	Delay Range	Delay Jitter Range	Budget	Admitted Groups	SA	LR	Lower Bound	Gap
3.0 ~ 10.0	1.0 ~10.0	1.0 ~3.0	400	2	-1214	-1214	-2550	110.06%
			800	5	-2559	-2559	-2569	0.42%
10.0 ~ 30.0	0.5 ~5.5	0.5 ~1.5	1600	2	-2181	-2181	-2875	31.85%
			3200	6	-2875	-2875	-2886	0.39%

Table 5-19. The Result of Case 1 for Model 2

Case 2

Network topology: Mesh Network

Number of Requested Multicast Group : 9

Traffic Range	Delay Range	Delay Jitter Range	Budget	Admitted Groups	SA	LR	Lower Bound	Gap
3.0 ~ 10.0	1.0~10.0	1.0 ~3.0	800	4	-3242	-3288	-4127	25.53%
			1600	9	-4131	-4131	-4146	0.36%
10.0 ~ 30.0	0.5 ~5.5	0.5 ~1.5	3200	6	-4438	-4438	-4809	8.35%
			6400	9	-4790	-4790	-4828	0.80%

Table 5-20. The Result of Case 2 for Model 2

Case 3

Network topology: Mesh Network

Number of Requested Multicast Group：12

Traffic Range	Delay Range	Delay Jitter Range	Budget	Admitted Groups	SA	LR	Lower Bound	Gap
3.0 ~ 10.0	1.0~10.0	1.0 ~3.0	800	4	-2742	-3021	-5411	79.14%
			1600	12	-5425	-5425	-5424	0.17%
10.0 ~ 30.0	0.5 ~5.5	0.5 ~1.5	4000	10	-5078	-5255	-5795	10.29%
			4800	12	-5772	-5772	-5806	0.58%

Table 5-21. The Result of Case 3 for Model 2

Case 4

Network topology: GTE Network

Number of Requested Multicast Group：9

Traffic Range	Delay Range	Delay Jitter Range	Budget	Admitted Groups	SA	LR	Lower Bound	Gap
3.0 ~ 10.0	1.0~10.0	1.0 ~3.0	1200	5	-4237	-4679	-6681	42.80%
			2000	9	-6662	-6662	-6708	0.69%
10.0 ~ 30.0	0.5 ~5.5	0.5 ~1.5	3600	5	-4304	-4786	-6754	41.14%
			4000	6	-4304	-5021	-6759	34.63%

Table 5-22. The Result of Case 4 for Model 2

Case 5

Network topology: GTE Network

Number of Requested Multicast Group：12

Traffic Range	Delay Range	Delay Jitter Range	Budget	Admitted Groups	SA	LR	Lower Bound	Gap
3.0 ~ 10.0	1.0~10.0	1.0 ~3.0	2000	8	-7228	-7453	-9287	24.61%
			2800	12	-9243	-9243	-9322	0.86%
10.0 ~ 30.0	0.5 ~5.5	0.5 ~1.5	6800	10	-8272	-8542	-9306	8.94%
			7600	12	-9189	-9189	-9321	1.44%

Table 5-23. The Result of Case 5 for Model 2

Case 6

Network topology: GTE Network

Number of Requested Multicast Group : 20

Traffic Range	Delay Range	Delay Jitter Range	Budget	Admitted Groups	SA	LR	Lower Bound	Gap
3.0 ~ 10.0	1.0~10.0	1.0 ~3.0	3200	15	-11598	-12068	-14099	16.83%
			4000	20	-14022	-14022	-14140	0.84%
10.0 ~ 30.0	0.5 ~5.5	0.5 ~1.5	10800	18	-13751	-14022	-14869	6.04%
			12000	20	-14669	-14669	-14915	1.68%

Table 5-24. The Result of Case 6 for Model 2

Case 7

Network topology: NSF Network

Number of Requested Multicast Group : 7

Traffic Range	Delay Range	Delay Jitter Range	Budget	Admitted Groups	SA	LR	Lower Bound	Gap
3.0 ~ 10.0	1.0~10.0	1.0 ~3.0	800	3	-1661	-2698	-4228	56.67%
			1600	7	-4226	-4226	-4251	0.59%
10.0 ~ 30.0	0.5 ~5.5	0.5 ~1.5	3200	5	-2677	-3454	-4133	19.65%
			4000	7	-3371	-4106	-4141	0.85%

Table 5-25. The Result of Case 7 for Model 2

Case 8

Network topology: NSF Network

Number of Requested Multicast Group : 14

Traffic Range	Delay Range	Delay Jitter Range	Budget	Admitted Groups	SA	LR	Lower Bound	Gap
3.0 ~ 10.0	1.0~10.0	1.0 ~3.0	2400	10	-6840	-7929	-9772	23.24%
			3200	13	-9018	-9726	-9804	0.80%
10.0 ~ 30.0	0.5 ~5.5	0.5 ~1.5	7600	10	-7369	-8082	-9627	19.11%
			8800	13	-8477	-9213	-9677	5.03%

Table 5-26. The Result of Case 8 for Model 2

Case 9

Network topology: NSF Network

Number of Requested Multicast Group : 21

Traffic Range	Delay Range	Delay Jitter Range	Budget	Admitted Groups	SA	LR	Lower Bound	Gap
3.0 ~ 10.0	1.0~10.0	1.0 ~3.0	3600	15	-10459	-11278	-13086	16.04%
			4400	21	-12585	-13024	-13126	0.78%
10.0 ~ 30.0	0.5 ~5.5	0.5 ~1.5	10800	17	-10626	-11677	-13099	12.18%
			12000	20	-12056	-12634	-12975	2.70%

Table 5-27. The Result of Case 9 for Model 2

For comparing the influences of QoS items on the leasing cost, we have the following tests to show which item of the QoS requirements has the largest influence on the leasing cost.

Network topology: Mesh Network

Number of Requested Multicast Group : 6

Traffic Range	Delay Range	Delay Jitter Range	SA	LR	Lower Bound	Gap
3.0 ~ 10.0	1.0 ~10.0	1.0 ~3.0	695	350	350	0.00%
3.0 ~ 10.0	0.5 ~5.5	1.0 ~3.0	725	350	350	0.00%
3.0 ~ 10.0	1.0 ~10.0	0.5 ~1.5	720	350	350	0.00%
3.0 ~ 10.0	0.5 ~5.5	0.5 ~1.5	730	350	350	0.00%
10.0 ~ 30.0	0.5 ~5.5	0.5 ~1.5	2125	1030	1021	0.88%
10.0 ~ 30.0	1.0 ~10.0	0.5 ~1.5	2115	1030	1030	0.00%
10.0 ~ 30.0	0.5 ~5.5	1.0 ~3.0	2120	1030	1030	0.00%
10.0 ~ 30.0	1.0 ~10.0	1.0 ~3.0	2090	1030	1030	0.00%

Table 5-28. The Result of Influence Test of QoS Items on the Leasing Cost

With the results of SA for each experiment, we observe that the bandwidth requirement has the most significant influence on the leasing cost among all QoS requirements. And the following is another test:

Network topology: Mesh Network
Number of Requested Multicast Group : 12

Traffic Range	**Delay Range**	**Delay Jitter Range**	**SA**
3.0 ~ 10.0	1.0 ~10.0	1.0 ~3.0	1445
3.0 ~ 10.0	0.5 ~5.5	1.0 ~3.0	1495
3.0 ~ 10.0	1.0 ~10.0	0.5 ~1.5	1470
3.0 ~ 10.0	0.5 ~5.5	0.5 ~1.5	1495
10.0 ~ 30.0	0.5 ~5.5	0.5 ~1.5	4445
10.0 ~ 30.0	1.0 ~10.0	0.5 ~1.5	4420
10.0 ~ 30.0	0.5 ~5.5	1.0 ~3.0	4445
10.0 ~ 30.0	1.0 ~10.0	1.0 ~3.0	4395

Table 5-29. Another Result of Influence Test of QoS Items on the Leasing Cost

The reason is about the way to calculate average delay and average delay jitter. Because the fluctuation of bandwidth requirement is larger than delay and delay jitter requirement. When users need more bandwidths, the extent of increasing the capacity is large. But the time length of delay that normal people can stand can't be too much and so is its fluctuation.

We also can observe that the lower bounds in the model 2 are almost the same under the same QoS requirement. This is because the value of the lower bound is calculated by solving subproblems. In subproblems, the budget constraint has been relaxed. We can accept all requested user groups since there's no budget constraint. Therefore, under the same QoS requirements, the model admits the same number of user groups.

5.5.3 Result Discussion

Up to present, we have got the experiment data above and other further experiments are still being designed and tested. We have tested three topologies with the different numbers of requesting Groups. Compared with the results of algorithm SA, our Lagrangean-based algorithm LR has achieved improvements in the above three topologies, which can show the effectiveness of Lagrangean relaxation approach.

The reason that LR works better than SA is that LR has multipliers to provide hints about the extent of constraint violating. And it can help to make decision variables more effectively and accurately. SA's hints are rare, since it only uses the degree of nodes to reflect the link's importance and runs add/drop heuristics by group ID numbers. Furthermore, LR can improve the result iteration by iteration; since the result of SA is the same no matter how many iterations it runs.

Chapter 6 Summary and Future Works

6.1 Summary

With the enormous growth, the Internet not only has the traffic demands increased but also the character of these IP applications. In particular, multimedia applications require a lot of bandwidth, and are very delay sensitive whether in the case of unicast or multicast. For the purpose of satisfying these multimedia applications, something better than "best effort" is required. The clients in pursuit of QoS must be assessed and if possible, improved upon. And in this thesis, we stand a point of view from network service providers. Thus we propose the algorithms to find the solution to the network service providers' decisions on how much capacity of network links they should lease from network providers and how they construct the paths for multicast routing with Quality-of-Service (QoS) guaranteed.

In this thesis, we consider capacity assignment, end-to-end quality-of-service routing, and admission control problems jointly and use a mathematical description to model this overall optimization problem. By applying Lagrangean relaxation and subgradient method on it, we can relax some complicated constraints and divide the primal problem into several subproblems. Then we can acquire some information to support our heuristics moving toward the better direction by solving the decomposed subproblems. We implement the algorithm and test three well-known network topologies. In terms of performance, our Lagrangean Relaxation based solution has more significant improvement than simple heuristics. And finally, not only we implement an algorithm for constructing a QoS-constrained network with the minimum leased capacity, but also our algorithm is a realization of QoS-constrained multicast routing and capacity assignment by applying optimization-based technique and the effectiveness of the algorithms is quite good.

6.2 Future Works

The multicast environment we concern here is long-term and static. Thus, we can consider the dynamic condition of that. That is, to do the real-time admission control over the multicast networks. And also, we can do rerouting when there comes new requested multicast groups to gain more revenue.

Admission control we use is to fully coordinate the multicast traffic. That is, we take into account an entire user group at one time. When some QoS requirement of the user in a requested group cannot be satisfied, we can just reject the whole group. After concerning all requested user groups, we fully admit them as possible. But it could be possible that there is still redundant capacity. And if we can ignore the user with higher requirements and let other ones in that group to use the multicast network, then we can use the leased capacity more effectively than before. That is, if we implement partial admission-control to coordinate the network traffic, we can admit more usage and raise the revenue.

References

[1] **顏宏旭著,** ***電腦網路與後勤網路之規劃與容量管理*****, 國立台灣大學資訊管理學研究所博士學位論文, 民國九十年六月。**

[2] **宋之揆著,** ***服務品質保證限制下群播開放最短路徑優先路由演算法之研究*****, 國立台灣大學資訊管理學研究所碩士學位論文, 民國九十一年六月。**

[3] Ahuja, R. K., Magnanti, T. L. and Orlin, J. B., "Network Flows—Theory, Algorithms, and Applications," Prentice Hall, ISBN0-13-027765-7, 1993.

[4] Breker, L. P. "A Survey of Network Pricing Schemes," *Proceedings of the 8th symposium on computer science*, University of Saskatchewan, 1996.

[5] Chen S. and Nahrstedt K., "An Overview of Quality-of-Service Routing for the Next Generation High-Speed Network: Problems and Solutions," *IEEE Network, Special Issue on Transmission and Distribution of Digital Video,* November/December 1998.

[6] Cisco Technical Report, "Introduction: Quality of Service Overview," *Cisco's QoS features*, June 1999.

[7] Cornujols, G., Fisher, M. L. and Nemhauser, G. L. "Location of Bank Accounts to Optimize Float: An Analysis Study of Exact and Approximate Algorithms," *Management Sci.*, vol. 23, pp. 789-810, 1977.

[8] Fisher, M., "Optimal Solution of Scheduling Problems Using Lagrange Multipliers: Part I.," *Operations Res.*, vol. 21, pp. 1114-1127, 1973.

[9] Fisher, M., "The Lagrangean Relaxation Method for Solving Integer Programming Problems," *Management Science*, vol. 27, pp. 1-18, 1981.

[10] Garey, M. R. and Johnson, D. S., "Computers and Intractability—a Guide to the Theory of NP-Completeness," Freeman: New York, 1979.

[11] Gavish, B., "Topological Design of Centralized Computer Networks: Formulations and Algorithms," *Networks*, vol. 12, pp. 355-377, 1982.

[12] Gavish, B. and Hantler, S.L, "An Algorithm for Optimal Route Selection in SNA Networks," *IEEE transactions on Communications*, vol. 31, no. 10, pp. 1154-60, October 1983.

[13] Geoffrion, A. M., "Lagrangean Relaxation and Its Uses in Integer Programming," *Math. Programming Study*, vol. 2, pp. 82-114, 1974.

[14] Ghanbari, M., "Two-Layered Coding of Video Signals for VBR Networks," *IEEE JSAC*, vol. 7, no. 5, pp. 771-781, June 1989.

[15] Ghosh, D., Sarangan, V. and Acharya, R., "Quality-of-Service Routing in IP Networks," *IEEE Transactions on Multimedia*, vol. 3, Issue 2, June 2001.

[16] Held, M. and Karp, R.M. "The Traveling Salesman Problem and Minimum Spanning Trees: Part I," *Operations Res.*, vol. 18, pp. 1138-62, 1970.

[17] Honig, M. L., Steiglitz, K., "Usaged-Based Pricing for Packet Data Generated by a Heterogeneous User Population," *IEEE Infocom' 95*, vol. 2, Boston, MA, 1995.

[18] Johnson, V. and Johnson, M., "Introduction to IP Multicast Routing," Stardust Technologies, Inc, 1997.
http://www.ipmulticast.com/community/whitepapers/introrouting.html

[19] Kenji, F., Masataka, O. and Katsuo, I., "Integration of Multicast Routing and QoS Routing," *The 10th Annual Internet Society Conference*, pp. 18-21, July 2000.

[20] Kishino, F., Manabe, K., Hayashi, Y. and Yasuda, H., "Variable Bit-Rate Coding of Video Signals for ATM Networks," *IEEE J. Select. Areas Commun.*, vol.7, pp.801-806, June 1989.

[21] Kleinberg, J., Rabani, Y., and Tardos, E., "Fairness in Routing and Load Balancing," in *the 40th Annual Symposium on Foundations of Computer Science*, pp. 568-578, 1999.

[22] Lin, F. Y. S. and Wang, J., "A Near-Optimal Minimax Routing Algorithm in Networks with Single-Path Routing," *IEEE GLOBECOM '93*, Houston, Texas, November 1993.

[23] Lin, F. Y. S. and Yee, J. R., "A Distributed Multiplier Adjustment Method for the Virtual Circuit Routing Problem," *ORSA J. Computing*, summer 1992.

[24] Mackie-Mason, J. K. and Varian, H. R. "Pricing the Internet," in Kahin, B., and Keller, J., editors, "Public Access to the Internet," Prentice-Hall, 1994.

[25] Mirzajan, A., "Lagrangean Relaxation for the Star-Star Concentrator Location Problem: Approximation Algorithm and Bounds," *Networks*, vol. 15, no. 1, pp. 20, 1985.

[26] Perros, H. G. and Elsayed, K. M., "Call Admission Control Schemes: A Review," *IEEE Communications Magazine*, pp. 82-91, November 1996.

簡　　歷

姓　　名：陳麒安

出 生 地：台灣省宜蘭縣

出 生 日：中華民國六十八年十一月九日

學　　歷：八十六年九月至九十年六月

國立台灣大學資訊管理學系

九十年九月至九十二年七月

國立台灣大學資訊管理學研究所