Chapter 1
Introduction

Routing optimization and resource allocation on Internet protocol (IP) networks have received a considerable amount of attention in the recent literatures. In [6] and [7], Mulyana and Killat proposed a hybrid genetic algorithm for the routing optimization problems. In [10] Ramakrishnan and Rodrigues discussed different forms of shortest-path routing optimization. In [3], K. Gopalan and T. Chiueh presented a method where large networks are broken down into smaller areas and computed individually before being reunited.

While most research has focused on resource allocation for communication resources, resource allocation decisions [8] on networks are concerned with the allocation of limited bandwidth so as to achieve the best systems performances. Resource allocation in the multi-service communication networks presents a very important problem in the design of the future multi-class Internet [9]. The major motivation for the research in this field lies in the necessity for structural changes in the way the Internet is designed. The current Internet offers a single class of 'best-effort' service; however, the Internet is changing.

Different from traditional separated network, an All-Internet protocol (All-IP) network carries most services such as voice, video, and data by an integrated packet-switching network [2]. This revolutionary network not only reduces network deployment and operation costs, but also makes it possible to offers more effective and more economy new time sensitive services that are not possible on the traditional separated networks.

New time sensitive services such as voice and video may severely suffer from three major quality of service (QoS) problems in a packet switching network, including long delay time, jitter, and packet loss [11]. These services require a better and more reliable network performance. Moreover, these services require firm performance guarantees from the network where certain resources should be reserved for them. In this regard, we term a network path with an end-to-end performance guarantee as a reserved path [9]. A reserved path is a unicast path, such as one that carries VoIP (Voice over IP) data between two endpoints. Since the focus of this study is on network resource provisioning, we assume the reserved paths that customers request from the provider last for a relatively long period of time.

The problem of optimal resource allocation for satisfying an end-to-end QoS requirement in a network of schedulers is usually addressed by partitioning the end-to-end QoS requirement into local requirements and then solving them in each individual node [11]. Hence, the problem of further mapping traffics' local QoS requirements to their resource allocations in a single network node is of practical importance. A number of works [10] recently use end-users' utility as the maximizing objective for resource allocation schemes. From the service provider's point of view, the optimal resource allocation scheme for their revenue maximization is essential to QoS requirements.

Consider a core network where there are m classes of sessions defined by the system managers. The objective is to provide an end-to-end QoS scheme in order to meet a basic requirement in terms of costs. In Wang [13], he presented a fair resource allocation approach for QoS routing on All-IP networks. His model is introduced in the following.

Define G $=\langle N, L\rangle$ as an oriented graph, where N and L denote the set of nodes and the set of links in the network respectively. Let k_e be the marginal cost of bandwidth for each link $e \in L$, and B be the total budget. Denote by θ_i the bandwidth allocated to each i and r_i the bandwidth requirements of class i. Let $f_i(\theta_i) = \ln(\theta_i / r_i)$ be the satisfaction function of each connection for each class i, and $w_i \in (0, 1)$ represents the weight of each class i, for i=1, 2, ..., m, where $\sum_{i=1}^{m} w_i = 1$. Denote by x_e a decision variable of the total bandwidth allocated to the link e. Meanwhile, let

$$x_i(e) = \begin{cases} 0, & \text{if link } e \text{ is not used by class } i \\ 1, & \text{if link } e \text{ is used by class } i \end{cases} \quad (1)$$

Define the satisfaction of the network as $w_1 \times f_1(\theta_1) + w_2 \times f_2(\theta_2) + \Lambda + w_m \times f_m(\theta_m)$ and the cost of the network as $\sum_{e \in L} k_e x_e$ [8].

The approach of Wang [13] for QoS is written by a mixed integer nonlinear programming model described in the following:

$$\text{Maximize} \quad w_1 \times f_1(\theta_1) + w_2 \times f_2(\theta_2) + \Lambda + w_m \times f_m(\theta_m) \quad (2)$$

$$\text{Subject to} \quad \sum_{e \in L} k_e x_e \le B \quad (3)$$

$$\sum_i x_i(e) \theta_i E[\lambda_i] = x_e, \forall e \in L \quad (4)$$

$$\sum_{e \in I_j} x_e = \sum_{e \in O_j} x_e, \forall e \in L \quad (5)$$

$$j \in N$$

$$x_e \ge 0, \forall e \in L$$

$$x_e \le U_e, \forall e \in L \quad (6)$$

$$x_i(e) \in \{0,1\}$$

$$\theta_i \ge r_i$$

where $E[\lambda_i]$ is the average number of connections for class i, U_e is the maximum bandwidth of link *e*, I_j is denoted for a set of links flowing into node *j,* and O_j is the set of paths flowing out from node *j,* $j \in N$. In the model above, it includes the bandwidth allocated to each link and the bandwidth allocated to each class simultaneously. The quality of service on networks is defined by satisfaction functions which are simply written in terms of bandwidth required by the users on the network. The objective function (2) in terms of satisfaction is written as the sum of weighted utilization of each class. The budget constraint is shown in (3). Equation (4) involves three variables where $x_i(e)$ is defined in (1). Equation (5) defines the flow balance at each node. Equation (6) restricts the bandwidth at each link under a maximum capacity. Obviously, the model reveals a NP hard problem since an optimal path is determined in this model. In order to reach the goal while avoiding the computational difficulty, we separate (2)-(6) into two phases. There are two different objectives discussed in the first phase: one focused on maximizing satisfaction, another one focused on minimizing cost. Based on optimum paths derived from the first one, the second phase determines a balance of total bandwidth required for each class.

This two-phase model will assist managers to estimate the quality of service on networks and to select paths for meeting minimum requirements. By giving the confidence interval of the satisfaction, we use simulation to investigate problems arising in random environments. Moreover, we discuss how to obtain *Q*% of total bandwidth requirement for budget allocation.

In this thesis, we shall propose a two-phase approach to solve (2) by adopting the **Maximum Satisfaction Method** (**MSM)** and the **Minimum Cost Method (MCM).** The remainder of this thesis is organized as follows. In Chapter 2, we focus on how to maximize the satisfaction function with a total budget. In order to reach the goal and to avoid wasting the resource, we propose the **MSM** which includes two phases: **Maximum Bandwidth Model**

(**MBM**) and **Best Recourse Allocation Model** (**BRAM**). In Chapter 3, we focus on how to minimize the cost with which we can satisfy *Q*% of total bandwidth requirement. Meanwhile we discuss how to maximize the satisfaction with a total budget. Like MSM, we propose **MCM** which includes two phases: **Minimum Cost Model** (**MCML**) (different from **MBM**, **MCML** focuses on cost saving) and **BRAM**. In Chapter 4, we compare **MSM** with **MCM** which is studied in Chapter 3 and Chapter 4 respectively. In Chapter 5, we will draw a conclusion and suggest some future studies.

The contributions of our studies are:

1. Propose a two-phase approach which assists the system managers to find the bandwidth and satisfy *Q*% of total bandwidth requirement.

2. Propose two methods which solve a budget allocation problem on All-IP networks that was previously formulated as a mixed integer nonlinear programming problem in [13].

3. Present a unified model of the routing optimization problem with end-to-end QoS guarantees which maximize the satisfaction function and save cost simultaneously.

The structure of our study is presented in Figure 1.1.

The structure of the study

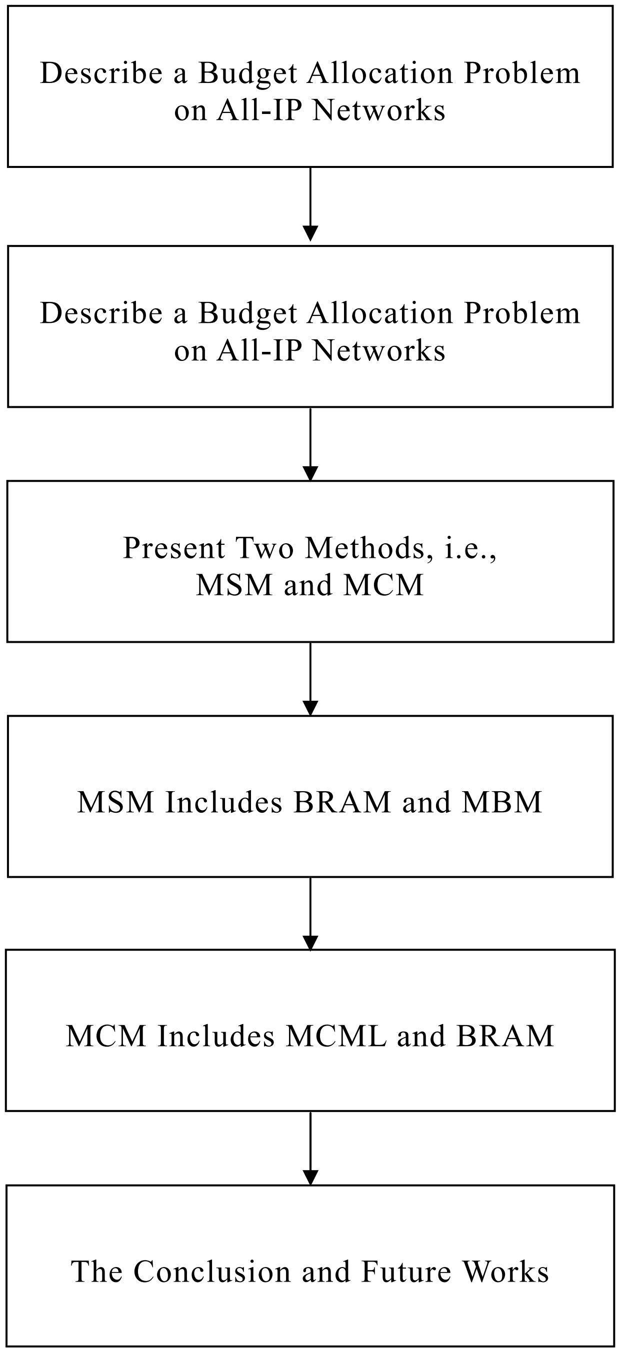

Figure 1.1 The structure of the study