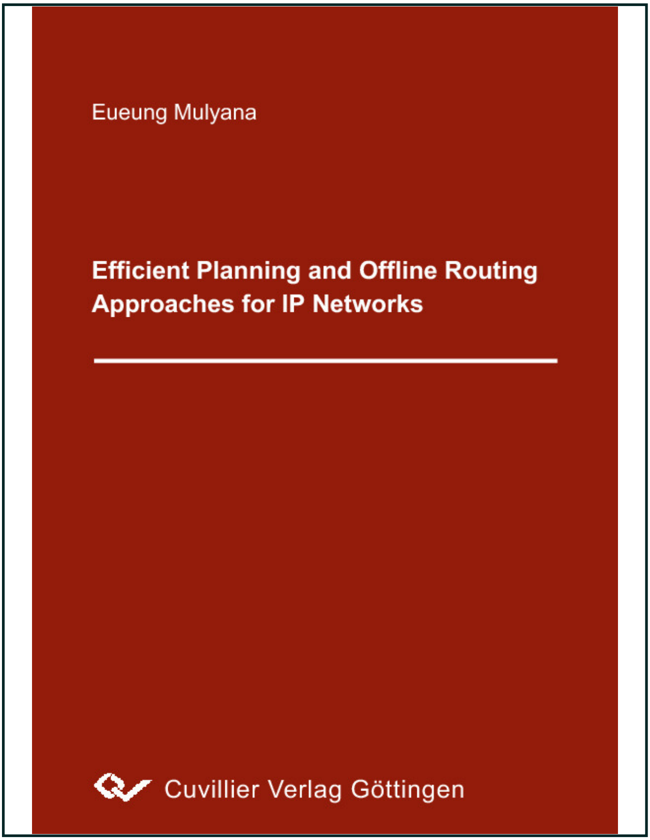
Eueung Mulyana (Autor)

Efficient Planning and Offline Routing Approaches for IP Networks



https://cuvillier.de/de/shop/publications/2250

Chapter 1

Introduction

In recent years, the Internet has evolved to the most dominant communication network carrying diverse applications including those, which are traditionally served by dedicated networks, such as voice and video services. Such different applications certainly require different levels of Quality of Service(QoS), on which the early IP networks[1] unfortunately were not focused. The Internet was not designed to guarantee a particular degree of performance but it was created with *best effort* service in mind where connectivity was the most important issue [QPS$^+$03].

For these reasons, the Internet has continuously been a topic of research and has been enhanced accordingly. Today, to a certain extent, the Internet has proven its important role in providing efficient data-centric and multi-service applications, so that it is believed to be the underlying platform for future communication networks, the Next Generation Networks(NGNs) [ALM$^+$01]. Generally, when dealing with performance issues, the corresponding research is termed Traffic Engineering (TE), which is basically composed of two aspects: (i) performance *evaluation*, which encompasses the application of technology and scientific principles to the measurement, characterization and modeling of (Internet) traffic; and (ii) performance enhancement and *optimization*, which covers the issues of controlling traffic according to performance requirements, while utilizing network resources economically and reliably [ACWX02].

The optimization aspects can be achieved though *capacity management* and *traffic management*. The first includes routing control and network resource dimensioning (e.g. bandwidth, buffer and computational resources). The last includes: nodal traffic control functions such as admission control, traffic conditioning, queue management, scheduling; and other functions that regulate traffic flow through the network or that arbitrate access to network resources between different packets or between different traffic streams. Furthermore, these optimization aspects can also be viewed from a control perspective:

[1]IP stands for Internet Protocol, the network layer protocol on which the Internet is based

preventive (*offline*) vs. *reactive* (*online*). In the first case, the traffic engineering control system takes preventive action to obviate predicted unfavorable future network states. In the second case, the control system responds correctively and adaptively to events that have already transpired in the network [ACWX02].

This dissertation addresses planning and management issues in IP networks. Specific focus is put on *routing*, which is one of the most significant functions that have to be performed by the Internet to fulfill its basic task: proper information exchange between communicating nodes. Thus, here the term "Traffic Engineering" is always referred to as "routing control". As such a control function can operate at different levels of temporal resolution, ranging from short (e.g. miliseconds) to intermediate (e.g. days or weeks) level, we limit the scope of this dissertation to the latter case i.e. to medium-term, or even to long-term[2], control activities, where the corresponding computation is performed *offline*. The challenge of research in this area is to control and to steer traffic through the network in the most effective way while satisfying some requirements (e.g. performance, minimization of cost).

1.1 Contributions

Today's IP networks are *diverse*, in the sense that many network operators apply different instruments in order to meet QoS and other requirements in their networks. Thus, traffic engineering is sometimes unique for each type of network. This dissertation is mainly concerned with offline routing control in diverse IP networks. The notion "diverse IP networks" refers to networks with different routing technologies such as the *classical* Interior Gateway Protocols (IGPs)[3], Multi-Protocol Label Switching (MPLS) or hybrid IGP/MPLS[4]. Our contributions are primarily in the aspects outlined in the following Subsections 1.1.1 through 1.1.3.

1.1.1 Traffic Engineering

At first, we propose a novel hybrid Genetic Algorithm (GA) to deal with the trafic engineering problem in the *classical* IP networks. The algorithm combines a population-based search capability in GA with a simple individual-based search heuristic, that simulates the behavior of network's administrators when they try to reroute traffic on/to a certain link. The work in this area has been published in [MK02a], [MK02b] and [MK05a].

Afterwards, a traffic engineering approach for several *transitional* IP networks is pre-

[2]We will discuss different time-scales for network planning and management in Chapter 2.

[3]In this dissertation, the terms "*classical* IP networks" and "IGP networks" are used synonymously.

[4]Hybrid IGP/MPLS networks will also be called as *transitional* IP networks.

sented. The basic idea was to establish a few explicit routing paths by making use of MPLS, instead of changing link-metric values as in pure IGP networks. Some results for and the comparison between various hybrid IGP/MPLS schemes are also given. This work has been reported in [MK03], [MK04b] and [MK04c].

At last, we investigate the impact of partial (non-linear) demand increase and develop a methodology to decide when and how reoptimization should be performed. Two methods for reoptimization based on local search frameworks are suggested. The work is published in [MK04a].

1.1.2 Multi-Class IP/MPLS Networks

Routing in multi-class IP/MPLS networks is much more flexible than in a pure IGP network: (i) routing can be implemented on a per service-class basis; and (ii) both shortest path and source routing are possible to be deployed. In this regard, we propose an offline traffic engineering approach for the problem of per-class unsplittable routing in IP/MPLS networks to specifically address per-class *over-provisioning* requirements. Such *per-class* over-provisioning is a simple, practical and less expensive means for providing QoS. Furthermore, we also consider the problem of dimensioning of such a network under several different routing schemes. Novel mathematical formulations and the some heuristic frameworks for solving these problems are also given. This work is reported in parts in [MK05c] and [MSK05].

1.1.3 Demand Uncertainty

To obtain accurate demand information between node-pairs in a network is becoming more and more difficult, particularly as the network size grows. In such a situation, taking traffic variations explicitly into account when making routing decisions, may provide a better performance predictability. In this regard, we propose: (i) several simple traffic uncertainty models based on information of outgoing/incoming traffic from/to each node in a network; (ii) a flexible traffic model, addressing a situation where demands are composed of both fixed and uncertain parts. The corresponding approach for routing control under such demand conditions is also presented. This work has been published in parts in [MK05b] and [MZK05].

1.2 Outline

This dissertation is structured as follows.

Chapter 2 describes some fundamental notions for network planning and reviews the optimization approaches, focusing on those, which are intensively used for solving the problems in the subsequent chapters. It first addresses the basic terminology of networks and network planning. Then, it discusses several optimization approaches, covering linear programming and some heuristic frameworks.

In Chapter 3, we present a compact overview of routing in IP networks. This embraces the classical hop-by-hop destination based shortest path routing, routing via label switching in MPLS enabled IP networks, and also the more flexible class-based routing in IP networks applying MPLS and service differentiation.

Chapter 4 reveals our novel approach for solving the problem of traffic engineering in classical and transitional IP Networks. The latter is referred to as combined routing scheme in classical IP networks, where some nodes are MPLS enabled. Furthermore, this chapter also discusses the impact of partial demand increase on network utilization and presents a simple policy and two reoptimization approaches to deal with the issue.

Chapter 5 deals with the problem of offline routing control and with the joint problem of routing and dimensioning for multi-class IP/MPLS networks. In all problems, we particularly emphasize over-provisioning constraints, since they are of paramount importance for providing a good quality of service in the network. Moreover, the resilience aspect is also considered, by simultaneously planning backup paths for routing under network failures.

Chapter 6 is devoted to routing optimization under demand uncertainty. At first, the corresponding demand models are introduced, the impact on link occupancy is explained and the corresponding link load calculation is derived. At last, we also introduce the concept of partially uncertain demands to address a situation where traffic is composed of both fixed and uncertain parts, providing flexibility to deal with common practical cases, where only a subset of the necessary information can precisely be determined.

Chapter 7 gives a summary of this dissertation and points out some directions for futher research.

Chapter 2

Network Planning and Optimization

This chapter is devoted to the introduction of some basic issues related to network planning and optimization. We first review some fundamental notions which are intensively used throughout this dissertation. Afterwards, a general network planning and management framework is presented, providing a clear view to the role of the specific problems addressed in the following chapters. In the last section, we discuss several optimization approaches and especially focus on those being used for solving the problems presented in this thesis.

2.1 Terminology

A communication network consists of equipment interconnected by transmission media, allowing communication entities (users) to exchange information, which may be voice, graphics, video or data. A public communication network connects a large collection of users, that are typically distributed over different geographical locations. The telephone network is probably the most historical example of a such public network, that exists since several decades. In recent years, the global computer network (i.e. the Internet) plays an increasing role and has become a standard platform for the current and future multi-media communication. From a functional point of view, networks can be subdivided into *access* and *backbone*(*core*) networks. Access networks are connected directly to customers, while a backbone network joins all access networks together. As our focus is on backbone networks only, in this dissertation the term "network" always means "backbone network".

A communication network is an object with a certain structure (often called network *topology*) and with a set of attributes. The topology can be viewed as a *graph* which consists of *nodes* connected by *links*. The attributes describe the network's status and its

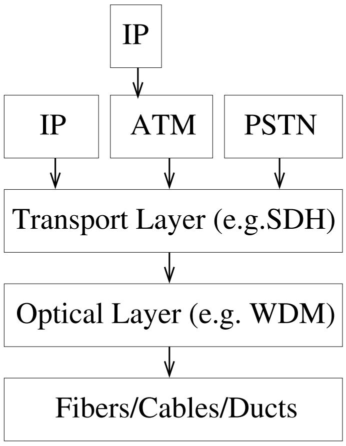


Figure 2.1: An example of multi-layer network architecture

specific configuration e.g. link capacities or routing parameters. A communication network carries communication *traffic* from ingress (source) to egress (destination) nodes. This traffic can be thought of as an aggregation of individual customers' traffic, which is connected to a common pair of ingress and egress nodes. Since for cost and efficiency reasons, a network does not always provide a point-to-point physical connection between node pairs, the network *resources* have to be shared for all traffic in the network. These resources may be given in terms of transport bandwidth on the links, switching capacity or forwarding resources at the nodes.

Traffic *demands* for a whole network can be pictured as a *traffic matrix*, in which each element of the matrix specifies the traffic volume between any two nodes in that network. In order to fulfil these demands, the corresponding traffic has to be *routed* through one or more paths connecting the ingress and egress nodes. The amount of traffic associated with a route can be thought of as a *flow*. It is obviously clear, that for the purpose of traffic routing there shall be sufficient resources available in the network. Therefore, the need for network planning and management is becoming apparent, since network resources are limited and correspond directly to factors such as investment or operational cost.

Apart from the functionality to exchange information, communication networks may completely differ from each other. Differences can exist, for instance, in communication protocols and thus in nodal equipment and transmission technology. Moreover, due to the use of digital technology, it is common that a network is working on the top of another. Figure 2.1 gives such a multi-layer architecture[1]. It shows among other things,

[1] The notion "Transport Layer" in the figure is not to be confused with OSI Transport Layer. Here, it is used for the context of general networking i.e. for carrying services with lower data-rates (usually) over long distances.

that IP networks can make use of either ATM (Asynchronous Transfer Mode) or directly SDH (Synchronous Digital Hierarchy) networks. SDH networks in turn can use WDM (Wavelength Division Multiplexing) networks to deliver services[2]. A similar layering architecture can also be deployed for the PSTN (Public Switched Telephone Network). In regard to this multi-layer architecture, it is useful to distinguish two types of networks: (i) *traffic networks*, where demands are stochastic in nature (e.g. packet, voice or high speed on-demand circuit); they have also the switching/routing capability to handle short-lived requests on-demand; and (ii) *transport networks*, which provide high-data rate services that are required to be set up on a semi-permanent or permanent basis. Note that in such a multi-layer architecture, each layer has its own definition of traffic, link capacity and node functionality [PM04]. This dissertation is dealing with planning and management issues in IP networks, which can be categorized as traffic networks. For dimensioning problems, as those to be addressed in Chapter 5, we also need to consider the underlying transport network, since transport granularities are different from one type of transport network to another.

2.2 Network Planning and Management

Network Planning and Management (NPM) addresses all activities related to the network development and evolution. There are basically three NPM activities, which correspond to different time-scales [PM04][3]:

- long-term (months to years) activities to design or extend the network in order to meet demands and requirements for a long period of time. These include for example: topology design, node and link dimensioning, capacity expansion, routing and resilience planning.
- medium-term (days to weeks) activities, which cover a list of actions to achieve the convergence towards the established long-term plans. Routing control (offline) can also be seen as a medium-term activity, at which routing is reconfigured to meet service requirements or to obtain a better network usage.
- short-term (real-time to hours) activities, which incorporate real-time operations such as packet level operations (marking, scheduling, policing, buffer management), restoration or online routing control.

Figure 2.2 shows a generic interaction model for the NPM activities both for traffic and transport networks. Long and medium-term NPM activities consider the current and fore-

[2]We refer to [Tan03] for a comprehensive overview of these different types of networks.

[3]Different and coarser time granularities can also be used, in particular when planning focuses on transport networks as in [Rob99] and [DDT$^+$00].

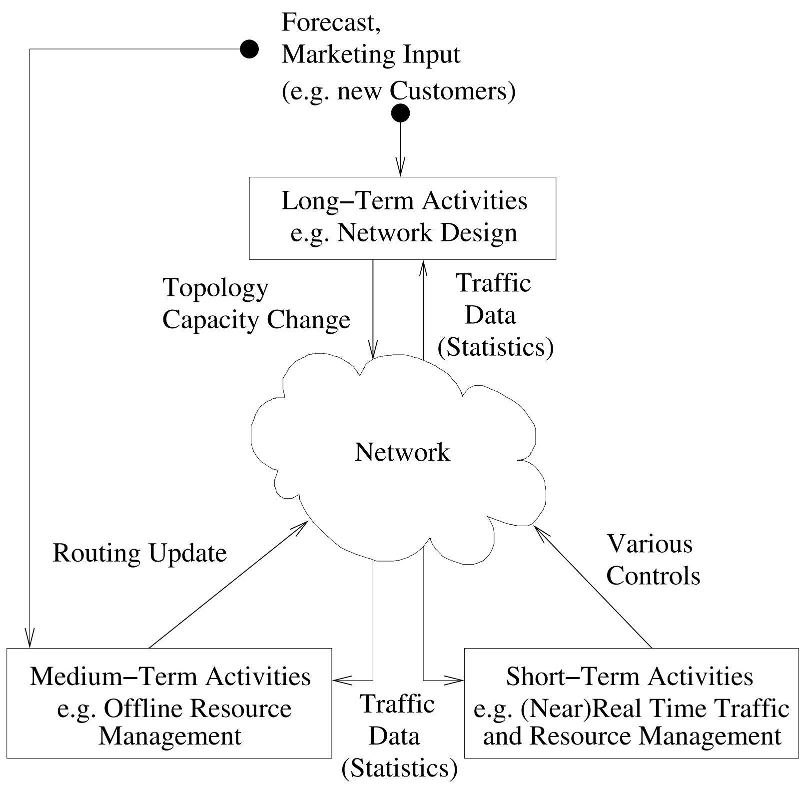


Figure 2.2: A generic model for Network Planning and Management

cast information regarding customers, services and other strategical issues. The outcomes of the process are dependent on particular details of the activities. These can be a topological change, link capacity upgrades, routing updates or even a complete new network design. For operational networks, current network status and traffic information are also considered by all NPM activities. They may possibly trigger a new management action e.g.: (i) if the current situation deviates significantly from the forecast in the case of long and medium term activities; or (ii) if a specific event such as congestion occurs in the case of short term management activities.

The issues addressed in this work are mainly related to long and medium-term NPM activities in IP Networks. The first issue is about offline traffic engineering, which can be viewed as a medium-term management problem, in order to balance traffic, to avoid congestion or in general to efficiently provision network's links for providing a desired QoS level. In this context, the network has to satisfy the current transmission demands with the already installed capacities without additional capital investment. The second issue is about network dimensioning problem, which is usually handled as a long-term NPM activity. For these reasons, if it is not stated explicitly, in this dissertation we use