University of Alberta

Library Release Form

Name of Author: Dongfang Zheng

Title of Thesis: Dynamic Optimization of Virtual Path System with Stochastic Traffic

Flows in Satellite ATM Networks

Degree: Master of Science

Year this Degree Granted: 2003

. . . .

Dongfang Zheng

2nd Floor,
Electrical and Computer Engineering
Research Facility (ECERF)
University of Alberta
Edmonton, AB
Canada, T6G 2V4

Date: . . . .

University of Alberta

Dynamic Optimization of Virtual Path System with Stochastic Traffic Flows in Satellite ATM Networks

By

Zheng, Dongfang

A thesis submitted to the faculty of graduate studies and research in partial fulfillment of the requirements of the degree of **Master Science.**

Department of Electrical and Computer Engineering

Edmonton, Alberta

Fall 2002

University of Alberta

Faculty of Graduate Studies and Research

The undersigned certify that they have read, and recommend to the Faculty of GraduateStudies and Research for acceptance, a thesis entitled **Dynamic Optimization of Virtual Path System with Stochastic Traffic Flows in Satellite ATM Networks** submitte by Dongfang Zheng in partial fulfillment of the requirements for the degree of **Master of Science**.

. . . .
Dr. M. Mandal (Supervisor)

. . . .
Dr. B. Cockburn

. . . .
Dr. J. Harms

Date: . . . .

Acknowledgements

I would like to thank my parents. Every time in my life, they are always the source of my power and courage.

During the past two years, Dr. M. Mandal and Dr. X. Liu have given me countless advice, guidance, and encouragement. I would like to appreciate their supervisory efforts.

I would also like to thank all members of Multimedia Computing and Communications Laboratory in the Department of Electrical and Computer Engineering, University of Alberta. When talking to them, I always obtained good ideas and help.

I would further like to thank all my friends in China. Their friendship gives me so much happiness and I am deeply missing them.

Last, I would like to thank *Canadian Institute for Telecommunications Research* (CITR), Canada Space Agency and *National Science and Engineering Research Council of Canada* (NSERC) for their financial support.

Abstract

The exponential growth of multimedia applications demands high-speed, broadband global information infrastructure. The *Asynchronous Transfer Mode* (ATM) is a preferred transmission and switching technique for B-ISDN. ATM has played a major role in current terrestrial B-ISDN infrastructure. A satellite ATM network has been envisioned as the next information super-skyway, and has been included in the global information infrastructure. In *Satellite ATM* (SATM) networks, the various traffic patterns and dynamically changing topology call for effective network traffic control mechanisms to reduce congestion and utilize network resource efficiently. On network level, *Virtual Path* (VP) distribution is a critical issue in ATM traffic control domain.

In this thesis, three schemes have been proposed: Firstly, a new *Discrete Time Dynamic Virtual Path Allocation* (DTDVPA) is proposed with consideration of periodically time varying topology for *Inter-satellite Link* (ISL) based SATM networks. The DTDVPA allows the setting up off-line of a time-dependent virtual topology for online dynamic virtual path selection. Secondly, an improved network flow model, *Virtual Path Flow Assignment* (VPFA) model, is proposed with non-linear objective function to solve the VP optimization problem in ATM networks. The VPFA model extends the network modeling concepts from link-node incidence to path-node incidence. Finally, considering the stochastic nature of ATM traffic demands, sample values of traffic demands is used to estimate the values of virtual path capacity instead of using mean value. The *Weighted Scenario Tracking* (WST) approach is used here to solve this *Stochastic Programming* (SP) problem.

Table of Contents

List of Figures

List of Tables

List of Acronyms

AAL	ATM adaptation layer
ABR	Available Bit Rate
ATM	Asynchronous Transfer Mode
B-ISDN	Broadband Integrated Services Digital Network
BT	Burst Tolerance
CA	Capacity Assignment
CAC	Call Admission Control
CBR	Constant bit rate
CDV	Cell Delay Variation
CDVT	Cell Delay Variation Tolerance
CFA	Capacity and Flow Assignment
CLP	Congestion Loss Priority
CLR	Cell Loss Rate
CPCS	Common part convergence sub-layer
CS	Convergence sub-layer
DSL	Digital Subscriber Line
DSU	Digital Service Unit
DTDVPA	Discrete Time Dynamic Virtual Path Allocation
DT-PSS	Discrete-Time Path Sequence Selection
DT-VTS	Discrete-time Virtual Topology Setup
FA	Flow Assignment
GEO	Geostationary Satellites
GES	Gateway earth station
GFC	Generic Flow Control
HEC	Header error-control
HN	Here-and Now
ISL	Inter-satellite Link
LEO	Low Earth Orbit
MC	Multi-Commodity
MCR	Minimum Cell Rate
MCTD	Maximum Cell Transfer Delay
MEO	Medium Earth Orbit
MMF	Multimode fiber
MP	Mathematical programming
NCC	Network Control Center
NNI	Network-to-network interface
OBP	Onboard processing
OD	Origin and destination
OSI	Open System Interconnect Reference Model
PCR	Peak Cell Rate
pdf	Probability density function
PDF	Probability Distribution Function
PDU	Protocol data unit

PMD	Physical medium-dependent
PT	Payload Type
QoS	Quality of Service
SATM	Satellite ATM
SCR	Sustainable Cell Rate
SDU	Service data units
SONET/SDH	Synchronous Optical Network/Synchronous Digital Hierarchy
ST	Scenario Tracking
STP	Shielded twisted-pair
SAR	Segmentation and reassembly
SMDS	Switched multi-megabit data service
SSCS	Service specific convergence sub-layer
SP	Stochastic Programming
TC	Transmission-convergence
TDM	Time-division multiplexing
TNL	Terrestrial network link
UBR	Unspecified Bit Rate
UT	User terminal
UDL	Up- /down-link
UNI	User-to-network interface
VBR	Variable Bit Rate
VC	Virtual Channel
VCI	Virtual Channel Identifier
VCC	Virtual Channel Connection
VP	Virtual Path
VPI	Virtual Path Identifier
VPC	Virtual Path Connection
VPFA	Virtual Path Flow Assignment
WS	Wait and See
WST	Weighted Scenario Tracking

Chapter 1

Introduction

The *Broadband Integrated Services Digital Network* (B-ISDN) technology is an extension of the ISDN and is developed to support a large variety of services such as Video Telephony, Video Surveillance, Interactive Multimedia applications and High Definition Television. A switching technique known as ATM has been chosen as the transfer mode for B-ISDN. ATM combines the strengths of circuit switching and packet switching techniques, and is therefore capable of transporting all kinds of information and supporting each type of service. ATM can be mainly characterized by the following features [1]:

1. Fixed length packets: The information is transported in fixed length packets called ATM cells. Each ATM cell has 53 bytes, with a 5-byte header and a 48-byte user data field.

2. Virtual connections: ATM allows multiple logical connections to be multiplexed over a single physical interface. All cells belonging to one connection follow the same route through the network (connection-oriented transfer).

3. Statistical multiplexing: The traffic sources can generate cells asynchronously and at different rates. There is no fixed time relation between the input and output cells of the multiplexer. Statistical multiplexing allows the network resources to be flexibly and efficiently shared among all connections, favoring time-varying bit rates and multimedia services.

4. Service integrated and independent network: With ATM, all kinds of speech, video, and data services can be provided by a single network. Moreover, future services with unknown characteristics will be supported without any modification of the network.

5. Quality of service negotiation: During call setup a traffic contract is negotiated between the source and network, containing the relevant traffic parameters of the requested service. The negotiation between the source and network allows, on one hand, estimation of the network resources needed by the source, and ensures, on the other hand, that the resources are not underutilized.

6. Reduced control functionality: ATM is designed for high-rate physical links with very low bit error rates; therefore, no error control is applied in the ATM layer.

VP and VC

As a modern communication technology, the ATM has introduced several novel approaches. One of the most important concepts is the *virtual path* (VP) [2]. While the data is organized into cells and transmitted on *virtual channel connections* (VCC), a *virtual path connection* (VPC) can be used to carry a bundle of VCCs. This way, the function of setting up a path through the network needs to executed only once for all VCCs subsequently using that path. By reserving capacity on VPCs, the VCCs can be established simply and quickly. Furthermore, the task of managing VCCs is also significantly simplified with VPCs since multiple VCCs are grouped into units of VPCs. Appropriate use of VPs can significantly reduces the processing cost, since the network management is able to focus on a small number of connection groups instead of a large number of individual connections. On the other hand, because of the capacity reservation for VP, the link capacity utilization decreases, and causes the decrease of the throughput of the entire network. This is caused by relatively long-term VPC adjustment compared to VCC. Therefore, there is always a tradeoff between the network call throughput and the processing load on the signaling system [3]. When capacity is reserved on VPCs, it is desirable to dynamically adjust the allocation to track changes in network traffic flows. Therefore, high resource utilization can still be achieved.

The key to achieving high network efficiency in an ATM network is appropriate management of the VP pool and dynamic adjustment of the VP allocation to track changes in network traffic flows. VP pool management involves both issues of VPC capacity allocation and VPC routing. VP pool management offers adaptability to traffic variation and to network failures through dynamic resource management [2]. An optimal VP system can be obtained by solving VP routing and VP capacity allocation jointly. However, in general the joint optimization is highly complex. Even in its simplest setting, the VP distribution problem contains non-linearity in the objective function, and is very difficult to solve analytically [4]. Therefore, most available solutions are based on the fixed VP topology or fixed VP capacity.

A VP can be regarded as an abstraction of traffic flows in networks. By nature, the VP distribution is a flow-based optimization problem, and hence in most VP optimization schemes, the *flow assignment* (FA) model [5] is generally used. The conventional FA model is based on the link-node incidence, which is suitable for connectionless paradigms. In ATM networks, since the dominant traffic is connection-oriented, the conventional FA model may not be suitable for network flow modeling.

SATM Networks

In the satellite world, there are three kinds of satellites: *Geostationary Earth Orbit* (GEO), *Medium Earth Orbit* (MEO) and *Low Earth Orbit* (LEO). MEO and LEO satellites are also called non-geostationary satellites. Compared to GEO satellite system, the constellation of MEO or LEO satellite system offers a significantly smaller round-trip delay between earth and space segment. Since the LEO system has the smallest round-trip delay compared to MEO and GEO systems, it is the preferred transport system for broadband satellite data communication networks.

In LEO system, we need more satellites to provide global coverage, and the meshed satellite network topology in the sky changes with time. To achieve unrestricted connectivity, and smaller and cheaper user terminals with lower transmission delay, the ability to switch individual data streams onboard the satellite is desired [6]. These satellites are called *Onboard Processing* (OBP) satellites. Among several proposed SATM architectures, OBP SATM full meshed architecture provides the most flexible and advanced functionalities [6]. These OBP satellites are connected by *Inter-satellite Link* (ISL) to form a dynamically meshed sub-network in space. This meshed sub-network performs ATM traffic switching, flow, and congestion control functions, and supports connection setup, teardown, and QoS requirements [6].

The ISL sub-network is indeed a full function ATM network in the sky. Due to the permanent mobility of the entire LEO satellite network, in particular, ISL sub-networks in space are subject to permanent topological changes. The large number of satellites and the dynamics of the ISL sub-networks make the VP optimization in SATM more difficult. In this thesis, we propose a Discrete Time Dynamic Virtual Path Allocation to dynamically adjust VP system to adapt to dynamics of ISL sub-networks.

Stochastic Traffic

When flow models are used to optimize VP system in ATM, the external traffic demand (or input flow) can be obtained by describing incoming calls in terms of bandwidth under certain blocking rate. However, ATM supports six service categories, which have very different traffic paradigms and may exist simultaneously. On the other hand, when VCs are aggregated to VPs, statistical multiplexing decreases the required bandwidth to accommodate the traffic. This twofold complexity makes it difficult to efficiently model the traffic flows in ATM. We note that real traffic flows in ATM are not only dynamic but also stochastic in nature.

For traffic control in SATM systems with ISL sub-network, as well as the problem caused by stochastically changing traffic flows, the periodically time varying topology gives rise to new challenges for VP distribution mechanism.

1.1 Outline of the Thesis

In this thesis, we propose a new flow model based on path-node incidence. This model incorporates a non-linear objective function in terms of the path utilization ratio. As an extension of the link-node incidence, the path-node incidence describes the network based on its path-node relations. It is more suitable for connection-oriented networks like ATM. Besides the new flow model, we propose a Discrete Time Dynamic Virtual Path Allocation to dynamically adjust VP system to adapt to dynamics of ISL sub-networks. Finally, Stochastic Programming methodology is introduced to solve VP optimization problem. The Weighted Scenario Tracking method is proposed for VP optimization.

This thesis is organized as follows: In Chapter 2, background knowledge about ATM, satellite communication systems, and Stochastic Programming is reviewed. In Chapter 3, the Discrete Time Dynamic Virtual Path Allocation is proposed. In Chapter 4, the new Virtual Path Flow Assignment model is presented and compared with classic Flow Assignment model. Two prototype networks are used for performance evaluation, and the simulation results are reported. In Chapter 5, Scenario Tracking, One of Stochastic Programming approaches, is introduced into VP optimization. The Weighted Scenario Tracking model is proposed, and compared with simple mean value approach. The simulation results are presented. In Chapter 6, the conclusions and future work are included. Possible future applications of the VPFA model is discussed here.

1.2 Main Contributions

- The DTDVPA scheme has been proposed for VP optimization in SATM networks.
- A new VPFA model has been proposed for VP optimization in ATM networks [7].
- The WST scheme has been proposed to introduce SP methodology into VP optimization in ATM networks [8].

Chapter 2

Background Review

In this Chapter, we review some background works. The fundamentals of ATM are reviewed first in section 2.1, followed by the review of VP optimization in section 2.2. The review of satellite communications and SATM networks is then presented in section 2.4. Finally, the introduction to stochastic programming is presented in section 2.5.

2.1 Fundamentals of ATM Networks

In this section, some basic knowledge about ATM is reviewed. Some keywords used in literature are presented below:

In modern application of telecommunications, digital technology is widely adopted to transfer information through a variety of networks to the end users. One typical application is ISDN. Generally speaking, there are two forms of ISDN service [9].

Narrowband ISDN: Narrowband ISDN is digital service with the transport rates of 1.544Mbs (T1) or less. The following services are included in Narrowband ISDN.

Circuit Switched Voice: Circuit Switched Voice service provides digital voice service with many of the capabilities of a business Centrex over a 4-wire ISDN *Digital* Subscriber Line (DSL).

Circuit Switched Data: Circuit Switched Data service provides end-to-end digital service, data or video information are transferred over the public telecommunication network. By using special processing in out-of-band signaling, ISDN set up and maintain data connections.

Low Speed Packet: ISDN connections are transmitted over packaged information by D channel at 16Kbs X.25 connection.

High Speed Packet: ISDN connections are equipped with two B channels. Circuit Switched Voice, Circuit Switched Data, or High Speed Packet Service are transferred by B channels with 64 Kbs. Moreover, one or both the 64Kbs B channels are permanent virtual circuit to the packet network and able to provide high speed packet service by 64Kbs X.25 connection.

Broadband ISDN Service: Broadband ISDN service is applied to digital services with higher rates like more than 1.544 Mbs. These digital services include in the form of Frame Relay, SMDS, or ATM. The Broadband ISDN services support wide range of transfer rates such as from 25 Mbs up to the Gigabit. Broadband ISDN Service therefore makes the high quality of the digital service possible and is considered as the service of the future. Currently it is being applied to the many applications of the Information Super Highway.

The developments of latest digital facilities such as fiber optics help to significantly improve the performance of high speed broadband services and eliminated some network operations such as, X.25 with extensive overhead, error correction and flow control performed at intermittent points. Indeed, these functions have been handed to the upper layer on an end-to-end basis.

Asynchronous Transfer Mode (ATM)

ATM was selected as the target architecture for basic transport of heterogeneous traffic (data, voice, and video) within the scope of B-ISDN [10] by CCITT in 1990. For most ATM applications, the data rates are from 45 Mbps up to 1.5 Gbps and beyond. The cell size is fixed as 53 bytes. The details will be included in the following sections.

2.1.1 ATM Overview

ATM is a cell-switching and multiplexing technology that combines the benefits of circuit switching (guaranteed capacity and constant transmission delay) with those of packet switching (flexibility and efficiency for intermittent traffic). It provides scalable bandwidth from a few megabits per second (Mbps) to many gigabits per second (Gbps). Because of its asynchronous nature, ATM is more efficient than synchronous technologies, such as *time-division multiplexing* (TDM). The outstanding advantage of this technology is its ability to support a high switching rate, because switching is done purely at a hardware level without the assistance from software interface. Additionally, by using fixed length cells (53 bytes), the queuing delay can be decreased.

ATM Cell Basic Format

ATM transfers information in fixed-size units called cells. Each cell consists of 53 bytes. The first 5 bytes contain cell-header information, and the remaining 48 contain the "payload" (user information). Small fixed-length cells are well suited to transferring voice and video traffic because of the low delay.

ATM Devices

An ATM network is made up of an ATM switch and ATM endpoints. An ATM switch is responsible for cell transit through an ATM network. The job of an ATM switch is well defined: it accepts the incoming cell from an ATM endpoint or another ATM switch. It then reads and updates the cell-header information and quickly switches the cell to an output interface toward its destination. An ATM endpoint (or end system) contains an ATM network interface adapter. Examples of ATM endpoints are workstations, routers, *digital service units* (DSUs), and LAN switches.

ATM Network Interfaces

An ATM network consists of a set of ATM switches interconnected by point-to-point ATM links or interfaces. ATM switches support two primary types of interfaces: *user –to-network interface* (UNI) and *network-to-network interface* (NNI). The UNI connects ATM end systems (such as hosts and routers) to an ATM switch. The NNI connects two ATM switches. Depending on whether the switch is owned and located at the customer's buildings or publicly owned and operated by the telephone company, UNI and NNI can be further subdivided into public and private UNIs and NNIs. A private UNI connects an ATM endpoint and a private ATM switch, while a public UNI connects an ATM endpoint or private switch to a public switch. A private NNI connects two ATM switches within the same private organization. A public one connects two ATM switches within the same public organization.

ATM Cell-Header Format

An ATM cell header can be one of two Formats: UNI or the NNI. The UNI header is used for communication between ATM endpoints and ATM switches in private ATM networks. The NNI header is used for communication between ATM switches. Unlike the

UNI, the NNI header does not include the *Generic Flow Control* (GFC) field. Additionally, the NNI header has a *Virtual Path Identifier* (VPI) field that occupies the first 12 bits, allowing for larger trunks between public ATM switches.

ATM Cell-Header Fields

In addition to GFC and VPI header fields, several others are used in ATM cell-header fields. Generic Flow Control (GFC)---Provides local functions, such as identifying multiple stations that share a single ATM interface. This field is typically not used and is set to its default value. Virtual Path Identifier (VPI)---In conjunction with the VCI, identifies the next destination of a cell as it passes through a series of ATM switches on the way to its destination. *Virtual Channel Identifier* (VCI)---In conjunction with the VPI, identifies the next destination of a cell as it passes through a series of ATM switches on the way to its destination. *Payload Type* (PT)---Indicates in the first bit whether the cell contains user data or control data. If the cell contains user data, the second bit indicates congestion, and the third bit indicates whether the cell is the last in a series of cells that represent a single AAL5 frame. *Congestion Loss Priority* (CLP)---Indicates whether the cell should be discarded if it encounters extreme congestion as it moves through the network. If the CLP bit equals 1, the cell should be discarded in preference to cells with the CLP bit equal to zero. *Header Error Control* (HEC) calculates checksum only on the header itself.

ATM Reference Model

The ATM architecture uses a logical model to describe the functionality it supports. ATM functionality corresponds to the physical layer and part of the data link layer of the

OSI reference model. The ATM reference model is composed of the following planes, which span all layers [11][12]:

Control---This plane is responsible for generating and managing signaling requests.

User--- This plane is responsible for managing the transfer of data.

Management--- This plane contains two components:

Layer management manages layer-specific functions, such as the detection of failures and protocol problems. Plane management manages and coordinates functions related to the complete system. The ATM reference model is composed of the following ATM layers:

Physical layer----Similar to the physical layer of the OSI reference model, the ATM physical layer manages the medium-dependent transmission.

ATM layer----Combined with the ATM adaptation layer, the ATM layer is roughly analogous to the data link layer of the OSI reference model. The ATM layer is responsible for establishing connections and passing cells through the ATM network. To do this, it uses information in the header of each ATM cell.

ATM adaptation layer (AAL)---Combined with the ATM layer, the AAL is roughly analogous to the data data-link layer of the OSI model. The AAL is responsible for isolating higher-layer protocols from the details of the ATM processes.

Finally, the higher layers residing above the AAL accept user data, arrange it into packets, and hand it to the AAL.

The ATM physical layer has four functions: bits are converted into cells, the transmission and receipt of bits on the physical medium are controlled, ATM cell boundaries are tracked, and cells are packaged into the appropriate types of frames for the physical medium. The ATM physical layer is divided into two parts: the *physical medium-dependent* (PMD) sub-layer and the *transmission-convergence* (TC) sub-layer. The PMD sub-layer provides two key functions. First, it synchronizes transmission and reception by sending and receiving a continuous flow of bits with associated timing information. Second, it specifies the physical media for the physical medium used, including connector types and cable. Examples of physical medium standards for ATM include *Synchronous Optical Network/Synchronous Digital Hierarchy* (SONET/SDH), DS-3/E3, 155 Mbps over *multimode fiber* (MMF) using the 8B/10B encoding scheme, and 155 Mbps 8B/10B over *shielded twisted-pair* (STP) cabling. The TC sub-layer has four functions: cell delineation, HEC sequence generation and verification, cell-rate decoupling, and transmission-frame adaptation. The cell delineation function maintains ATM cell boundaries, allowing devices to locate cells within a stream of bits. HEC sequence generation and verification generates and checks the header error-control code to ensure valid data. Cell-rate decoupling maintains synchronization and inserts or suppresses unassigned ATM cells to adapt the rate of valid ATM cells to the payload capacity of the transmission system. Transmission frame adaptation packages ATM cells into frames acceptable to the particular physical-layer implementation.

ATM Adaptation Layer

The purpose of the *ATM Adaptation Layer* (AAL) is to adapt the *protocol data units* (PDUs) passed down from the higher layer onto ATM cells. As the higher-level PDUs may in general be of an arbitrary size, so one of the two sub-layers in the AAL is responsible for *segmentation and reassembly* (SAR) of the higher layer PDUs. The other sub-layer, the *convergence sub-layer* (CS), is responsible for packaging the higher layer PDU with any additional information required for the adaptation necessary and offering an interface to the B-ISDN user. As mentioned previously, there are many and varied requirements for applications using the B-ISDN and as there are several AAL protocols defined [11][12].

AAL Type 1

This type of AAL is normally used by *Constant Bit Rate* (CBR) services. The functions performed by this AAL are:

1) Segmentation and reassembly of user information.
2) Handling of cell delay variation (jitter).
3) Handling of cell reassembly variation.
4) Handling of lost and ministered cells.
5) Source clock frequency recovery at receiver.
6) Source data structure recovery at receiver.
7) Monitoring and handling of AAL-PCI bit errors.
8) Monitoring and (possibly) correcting the bit errors in the user information field.
9) For circuit emulation, monitoring and maintenance of end-to-end QoS.

AAL Type 2

This type of AAL would be used with *Variable Bit Rate* (VBR) traffic. This type is not well defined and it seems possible that it may be merged with AAL Type 1 in the future. Some of its functions are similar to ALL Type 1:

1) Segmentation and reassembly of user information.

2) Handling of cell delay variation (jitter).

3) Handling of lost and mis-inserted cells.

4) Source clock frequency recovery at receiver.

5) Monitoring and handling of AAL-PCI bit errors.

6) Monitoring and (possibly) correcting the bit errors in the user information field.

7) It also has the additional functions:

8) Handle *service data units* (SDUs) from a variable bit rate source.

9) Transfer timing information between source and destination.

10) Notify the higher layers of uncorrectable errors in AAL.

AAL Type 3/4

There was once separate Type 3 and Type 4 AALs, but they have now been merged. This AAL is now intended to support both *Available Bit Rate* (ABR) and *Unspecified Bit Rate* (UBR) services. In this AAL, the convergence sub-layer is split into two, the *service specific convergence sub-layer* (SSCS) and the *common part convergence sub-layer* (CPCS). The SSCS is application dependent, i.e. it could be for a VBR video application. The CPCS is responsible for constructing PDUs that can be sent to the other end user. There are two modes of operation of ALL Type 3/4: message mode and streaming mode. The message mode is intended for using framed data where the AAL-SDU is a logical unit of data with respect to the B-ISDN user. It allows the transport of a single AAL-SDU in one or (optionally) more than one CS-PDU. The CS-PDU may be then further spilt into several SAR-PDUs. The AAL-SDU can be of an arbitrary size.

In streaming mode, the AAL-SDUs are of fixed size and one or more of them may be transported in a single CS-PDU. Each AAL-SDU is delivered in a separate SAR-PDU.

In both cases, the SAR sub-layer provides error detection and both these modes can offer the following operational procedures:

Assured operation. Flow control and retransmission of missing or erroneous AAL-SDUs. Flow control restricted to point-to-point connections at the ATM layer and point-to-multipoint flow control possible.

Non-assured operation. No retransmission of missing or erroneous SAR-PDUs. Optionally deliver erroneous PDUs to user. Allow flow control for point-to-point connections but not point-to-multipoint.

This AAL type also provides multiplexing at the SAR sub-layer.

AAL Type 5

This AAL type provides similar services as AAL Type 3/4, but has a reduced overhead when compared to AAL Type 3/4. It is intended for use by VBR sources with timing relationship between source and destination. It has identical modes and operational procedures as AAL Type 3/4. The difference is that this AAL does not provide the AAL Type 3/4 multiplexing capability. The SAR sub-layer accepts only AAL-SDUs that are an integer multiple of 48 octets. So, it would be possible for this AAL to offer an efficient cell based interface to the B-ISDN user.

AAL Types 1 and Type 2 are not used much. AAL Type 3/4 has been chosen for the provision of the Bellcore switched multi-megabit data service (SMDS), the European version of which is the connectionless broadband data service (CBDS). AAL Type 5 has been selected by the IETF to provide IP services over ATM.

VPC and VCC

The VCC establishes an end to end connection and the connection request must pass through all the nodes between sender and receiver. The combined address space of VPI and VCI at the UNI interface is 24 bits, thus any two end users can establish 2^{24} connectons between them. However, VCC is only identified by the port and VCI number given to it by the network switch at the UNI interface. Since the VCI field is 16 bits, so a total of 65536 VCC can be established between two end users[13].

The VPC establishes permanentt/semi-permanent connections based on VPI value in the cell header. Once a VPC is established between two node pairs, called VP terminator, a VCC between these node pairs can be established by assigning the next available VCI number within the established VPC. Now the VCC between the VP terminators is identified by both VPI and VCI values in the cell header. Thus the VCC associated with a particular VPI value of a VPC will be switched at the transit nodes through which VPC passes between the VP terminators. Then at UNI interface between the two end users 256 VPC is established and each VPC can carry up to 65536 VCC, so totally 2^{24} connections can be established between the end users.

The advantage is that the VCC is established without performing routing functions at transmit nodes between the VP terminators. Thus a VPC can bundle VCCs and route these connections together, reducing the network operation and connections setup costs. If 65536 VCCs are routed through a single VPC, these can be identified by a single VPI label in the routing table of transit nodes. A link failure between switches will require only to find an alternative path for a single VPC and all the 65536 VCC will be rerouted by a single operation. The survivabilty of a network is greatly improved. Since the QoS of VPC must be met , so it is suggested to group the similar traffic VCCs in a single VPC. It can also reduce the the management cost[14].

Another case is that a VCC can be connected as a concatenation of VPCs and physical links. The VPC assign capacity and physical route between VP terminators. The routing for VPC is different to the case mentioned before. In this case only physical link can be facilitated. A VP topology or layout is a set of VPC established over the physical topology. The capacity of virtual path can be deterministic or statistical. If it has deterministic capacity, each VPC is multiplexed on a physical link based on their peak bandwidth requirement; while in statistical case, the VPC are multiplexed based on the computed effective bandwidth. The deterministic mechanism will reserve the peak bandwidth for all the VPCs sharing a physical link, so if some VPCs are idle on the assigned capacity, it can not be used by other VPCs even in the same physical link, this will result poor bandwidth utilization, as a result may possible higer the blocking rate. It can be improved by increasing the capacity of a physical link. If the VPC multiplexed statistically, the VPCs can share the capacity with other VPCs in a link, this may lower the bandwidth fragmentation, improve the utilizaton rate of limited bandwidth, the shortage of statistical method is that it may complicate the process of maintaining QoS for VPCs and VCCs within [14].

2.2 VP optimization in ATM

In this section, the ATM traffic control and VP optimization is reviewed. Some VP optimization schemes are also reviewed here.

2.2.1 ATM Traffic Attributes and Services

The traffic attributes of ATM networks specify the behavior of traffic and can be represented as [15]:

1) Peak Cell Rate (PCR)
2) Cell Delay Variation Tolerance (CDVT)
3) Sustainable Cell Rate (SCR)
4) Burst Tolerance (BT)
5) Minimum Cell Rate (MCR)
6) The QoS can be specified by the following parameter:
7) Cell Delay Variation (CDV)
8) Maximum Cell Transfer Delay (MCTD)
9) Mean Cell Transfer Delay (Mean CTD)
10) Cell Loss Rate (CLR)

However, not all the traffic type needs all the above traffic and QoS parameters. So a set of QoS classes were constructed, which group the traffic type based on their traffic character. The following are defined by ATM forum:

Constant Bit Rate: The traffic using a fixed amount of bandwidth during the connection. This traffic requires a QoS guarantee on the delay and loss rate. Using PCR and CDVT parameters can specify the traffic. Circuit emulation and some multimedia applications can use this service.

Variable Bit Rate: This specifies a peak rate and a sustained or average rate for the connections. Some VBR traffic is delay sensitive and some not. The ATM Forum defined VBR (RT) for *real time traffic* and VBR (NRT) for *non-real time traffic*. For VBR (RT) the QoS guarantees include delay and loss rate. MPEG transmission used in videoconference is an example. For VBR (NRT) the QoS guarantee includes only loss rate. It can be used by multimedia mail or Frame Relay Internetworking. Both traffic types can be characterized by SCR and BT.

Available Bit Rate: The available bit rate traffic does not require guaranteed bandwidth, but rather use the available bandwidth. However, it requires QoS guarantee on cell loss rate. In order to implement ABR, a feedback mechanism is needed from the network. The traffic can be specified by PCR and MCR. This service can be used to carry Internet traffic like email and file transfer and etc.

Unspecified Bit Rate: The traffic using UBR service needs no guarantee on the delay, loss and bandwidth. The service is the simplest to implement and can be used to support LAN traffic.

2.2.2 ATM Traffic Control

In ATM, the goal of traffic control is to achieve good ATM network efficiency while meeting the users' QoS requirements. This goal implies two conflicting objectives: high network efficiency; user's QoS guaranteed. The heterogeneous traffic in ATM makes this goal more difficult to achieve.

Generally, there are two types of traffic control, which are known as: flow control and congestion control. Flow Control ensures that the sender's data rate never exceeds the receiver's data rate. On the other hand, Congestion Control tries to reduce the possibility of congestion within the network. Congestion Control can be classified to two categories: Preventive Control and Reactive Control [9]. While reactive control is based on the feedback mechanism, preventive control is based on network provisioning.

In ATM, traffic control can be implemented at three different levels: cell level, call level, and network (or traffic flow) level [9].

Cell-transfer Control

This level of control provides functions to regulate the flow of cells.

Call Control

This level facilitates functions to control the flow of calls across a network. Relevant functions correspond to *Call Admission Control* (CAC) and path selection for call setup requests.

Network Control

This level includes network-wide controls like routing control, VP distribution.

On network level, VP distribution is a critical topic. VP can be regarded as an abstraction of traffic flows in networks. When network flows are optimized, network congestion can be prevented potentially. When combining with call and cell level traffic control mechanisms, traffic control objectives can be achieved.

In order to achieve high network efficiency in ATM networks, we must appropriately manage the VP pool, which involves both VPC capacity allocation and VPC routing, to track changes in network traffic flows. VP pool management offers adaptability to traffic variation and to network failures through dynamic resource management [4]. An optimal VP system can be obtained by solving VP routing and VP capacity allocation jointly. However, in general the joint optimization is highly complex. Even in its simplest setting, the VP distribution problem contains non-linearity in the objective function, and is very difficult to solve analytically [2]. Therefore, most available solutions are based on a two-phase approach: find capacities with the fixed routing or find routings with fixed capacities. The two-phase approach has two basic forms: iterative or non-iterative. The former solves the two-phase models recursively until the error between the updated solution and the last solution is less than a pre-specified tolerance. The recursive process will converge because there are only a finite number of optimal path flows. The latter, the non-iterative form, tries to establish a set of equations in terms of both capacities and flows. The order of these equations is usually not high, typically being a quadratic system, thus many well-developed algorithms can be adopted to solve it. Note that the non-iterative form differs from the joint approach mentioned above. The joint approach is in nature a one-phase approach.

2.2.3 Review of VP Optimization Schemes

In ATM networks, Given a network topology and traffic demands, the establishment of the system of virtual paths, and the assignment of connections to them so that network performance is optimized, entails a number of problems: How to select VP terminator pairs? What is the optimal amount of capacity to be reserved for each VP? How to choose the VP routes between the terminators? How to distribute traffic among the virtual paths (flow assignment) and which call routing algorithm to use to realize the desired assignment [16]? All these problems have to be solved by network management entity.

Based on a given physical network, virtual networks can be established on it. As discussed in Chapter 2, VP is an abstraction of traffic flow. Therefore, the VP distribution problem indeed is a virtual network optimization problem on flow level. Below VP level, VCC can be regarded as individual call, so VCCs form a call level virtual network above the VP virtual network.

In ATM traffic management, the VP distribution problem is critical. VP distribution problem includes two parts: i) VP routing (VP topology), ii) VP capacity allocation. In nature, these two parts are jointed together, and optimal solution can be obtained by solving VP routing and VP capacity allocation jointly. However, the joint optimization is very complex. Even in its simplest setting the VP distribution problem contains nonlinearities in the objective function, and is very difficult to solve analytically. Therefore, most available solutions are based on fixed VP topology or fixed VP capacity.

The VP distribution problem has been studied by numerous researchers. These researchers have tried to solve the VP distribution problem based on different objectives. Anerousis et al. [3] proposed a scheme to allocate bandwidth to VPCs based on a given fixed VP topology. This scheme can satisfy nodal constraints on the call processing load and blocking constraints for each *origin and destination* (OD) pair, and support multiple traffic classes. Here, the basic idea is that the more capacity we reserve on VPs, the lower the network call throughput. In their scheme, an optimal solution is obtained in terms of balancing network call throughput and the processing load on the signaling system. The scheme is complex and computationally expensive.

Chlamtac et al. [16] presented an algorithm to find a system of VP routes for a given set of VP terminators and VP capacity demands. By considering that the link capacity is limited in ATM networks, they proposed a scheme to find an optimal VP topology in terms of minimizing the maximum link load in the network. The basic idea is if traffic is distributed in a way that reduces the maximum link load, the possibility of congestion can be decreased and network robustness can be increased. Based on a given set of VP terminators and VP capacity demands, the algorithm can find an optimal VP topology while minimizing the load and reducing congestion on individual links.

Ahn et al. [13] stated a VP layout design method. The method is proposed to solve the VP routing problem based on a flow optimization algorithm with heuristic approaches. The algorithm can find an optimal VP topology in terms of carrying maximal percentage of offered traffic while minimizing call blocking probability, call setup, switching, and transmission costs. Guidelines for the design of robust VP layouts and the efficient establishment of VCs are also presented. In this VP layout design scheme, only one traffic class is considered.

Ryu et al. [17] proposed a heuristic design algorithm for VP distribution problem based on equivalent bandwidth concept. An equivalent bandwidth of a set of VCs multiplexed on a VP is defined as an amount of the bandwidth required to achieve a desired QoS. In their scheme, fluid flow approximation and stationary approximation methods [18] are used to calculate the equivalent bandwidth. In the scheme, they have tried to solve this jointed VP distribution problem In terms of minimizing network cost. In this scheme, only one class of traffic is considered, and all VCs are considered to be grouped into VPs.

Liu et al. [19] have proposed a generic capacity and flow assignment (GCFA) flow model based on path-node incidence for SATM networks. The proposed GCFA model can be used to solve VP distribution problem by assuming all traffic is carried on VP. In this model, the link capacity is assumed changeable. Another important feature is that SP methods have been used to solve this optimization problem. The wait and see (WS) approach has been investigated in their works. Even though the GCFA model is proposed for SATM networks, it is more like a generic model for general ATM networks. The dynamics of satellite ISL sub-networks is not considered in the scheme.

Since VPC can be regarded as abstraction of traffic flows in ATM networks, VP distribution is a flow-based optimization problem. Therefore, in most VP optimization schemes, the *flow assignment* (FA) [5] model is generally used. The classic FA model is described in the following section.

2.3 Review of Network Models

In network optimization and modeling area, three generic models are generally used. We give a brief review of these models in this section. A communication network can be expressed by a graph G = (V, A), where V is the set of nodes (vertices) and A is the set of arcs (or links, edges). Most packet communication networks can be regarded as an augmented graph with two types of indices: the operating index, expressed by T, e.g. packets delay or the number of packets in the system, generally speaking, network operation cost; the capital index, expressed by D, e.g. capacities cost. In a communication network, two important entities must be considered: link flow λ_i and link transmission capacity C_i. The physical meaning of λ_i is the traffic arrival rate in link i, expressed as data units per second, and C_i has the same dimension as λ_i. Accordingly, to a given network topology, the external traffic requirements, and the constraints of λ_i and C_i, the following three models can be developed:

Capacity Assignment (CA) model:

Given λ_i

Minimize T

Adjust C_i

Constraint D

Flow Assignment (FA) model:

Given C_i

Minimize T

Adjust λ_i

Capacity and Flow Assignment (CFA) model:

Minimize T

Adjust C_i and λ_i

Constraint D

Several important issues may be discovered with further analysis of these generic models as flow:

- Whether a model is CA, FA, or CFA, depends on what the design variables are, rather than what variables are presented in the objective function.

- An objective function should be able to reflect the effect of design variables; generally speaking, all design variables should be integrated into the objective function.

We note that the CA and FA models are special cases of the CFA model. Conventional formulations of all these models are based on the link-node incidence. However, link flows $\lambda_{i.}$ may further be expressed by path flows X_k. Usually the total number of X_k is greater than the total number of $\lambda_{i.}$ Accordingly, the design variables are X_k and the model under consideration becomes a *multi-commodity* (MC) problem. Most B-ISDN paradigms need to be characterized by the MC formulation.

2.4 Overview of SATM Networks

ATM, a high speed switching technique, has played a crucial rule in current terrestrial information infrastructure. A satellite ATM network has been envisioned as the next information super-skyway, and has been included in the Global Information Infrastructure. The growing interest in the interconnection of satellite and ATM networks is based on the following reasons [20]:

- By using satellite, ATM service can be provided over a wide geographical area, including remote, rural, urban, and inaccessible areas.
- Satellite communication systems have a global reach with very flexible bandwidth-on-demand capabilities. This significant strength of satellite communications ideally matches the main characteristics of ATM networks that provide bandwidth-on-demand and multimedia services.
- Satellites offer flexibility in terms of network configuration and capacity allocation to different sites that use ATM networks in various geographical areas.
- Satellites provide broadcast and multipoint-to-multipoint capabilities as well as fast network setup that can be useful in deploying multipoint-to-multipoint communication of ATM networks.
- New user can be easily be added to the system by simply installing the ATM stations at customer premises. As a result, possible network expansions will be a simple task.
- Satellites can act as a safety valve for optical fiber ATM networks. Fiber failure, or network congestion problem can be recovered easily by routing traffic through a satellite channel on a demand basis.

2.4.1 Three Types of Satellites

In satellite world, there are three types of satellites:

***Geostationary* (GEO) Satellites**

Geostationary satellites are placed in an equatorial orbit approximately 36,000 km above the earth, such that the satellites orbit synchronously with the rotating earth and seem to be fixed in the sky. Because of their high altitude, global coverage (excluding the polar regions) can be provided by just three geostationary satellites. However, due to the large distance to the satellites, GEO satellites have a very high signal attenuation and a large propagation delay (0.5s round-trip delay from one earth station to another and back), which may be annoying in a telephone conversation.

GEO satellites have been in use for more than 35 years. Today, there are more than 180 GEOs in commercial use, and most of them are used for fixed high data rate services and broadcasting. GEO systems are not suitable for mobile services, because the compensation of high free space propagation losses requires large satellite antennas. GEO systems for multimedia and real-time applications also suffer from high signal attenuation and large delay.

Medium Earth Orbit (MEO) Satellites

MEO satellites are placed around 10,000km in the sky to avoid the large signal attenuation and delay of the GEO satellites. For MEO satellites, a global coverage still needs a small number (e.g. 10) of satellites.

***Low Earth Orbit* (LEO) Satellites**

LEO satellites stay at 700-1500km altitude sky. However, a large number of satellites are needed to cover the earth's surface. On the other hand, LEO satellites have lower weight and are less complex than GEO satellites. LEO satellites have orbit periods of roughly 2 hours, and thus are non-geostationary. LEO satellites have the lowest signal attenuation and smallest propagation delay. Therefore, it is suitable for mobile services because small satellite antenna can be used. Since its lowest delay, real-time and multimedia applications can be supported, and it is the preferred satellite system for SATM.

2.4.2 SATM system architectures

There are mainly two types of SATM architectures:

Bent-Pipe SATM Architectures

In this category, no processing is done at the ATM layer or above on the satellites, and can be implemented using today's satellites. All protocol processing is performed on the ground at *user terminal* (UT), *gateway earth stations* (GES), and the *Network Control Center* (NCC). These architectures provide the fastest way of deploying ATM to remote locations and interconnecting ATM islands by using today's satellites. Although the lack of on-board processing and switching prevents full integration between a satellite network and the terrestrial ATM network, the advantages of ubiquitous coverage, inherent broadcast capability, cost effectiveness for long distances, suitability for providing mobile services, and flexibility in network configuration and capacity allocation, make these architectures a viable complement to high s peed terrestrial networks. These network architectures are further divided into three categories according to the type of mobility supported by the satellite network [20]:

1. Fixed ATM Network Architecture
2. Terminal Mobility Supporting ATM Network Architecture

3. Network Mobility Supporting ATM Network Architecture

Onboard Processing SATM Architectures

To achieve unrestricted connectivity, and smaller and cheaper user terminals with lower transmission delay, the ability to switch individual data streams onboard the satellite is desired. Depending on usage scenarios and system setup, these architectures can be divided into three architectures:

1. ATM Network Access – in this architecture, low-speed satellite links are used to connect remote ATM hosts to a terrestrial network.
2. ATM Network Interconnectivity – in this architecture, the satellite is an ATM node interconnecting multiple terrestrial ATM networks via high speed links.
3. Mesh ATM – in this architecture, several satellites form an ATM network in the sky via inter-satellite links. This architecture provides both ATM network access and network interconnectivity.

The shift in satellite communications paradigm from carrying traditional voice and data traffic to more time-critical and bandwidth-sensitive multimedia traffic makes the LEO satellites with OBP architecture the most promising one. Within three proposed OBP architectures, the full mesh SATM architecture provides the most flexible and advanced functionality. In this architecture, all satellites form a full meshed network in the sky, and they are connected by ISLs. Because these satellites have full ATM switching functionalities onboard, they form an ATM switching sub-network in space. This meshed sub-network performs ATM traffic switching, flow, and congestion control functions, and supports connection setup, teardown, and QoS requirements [6].

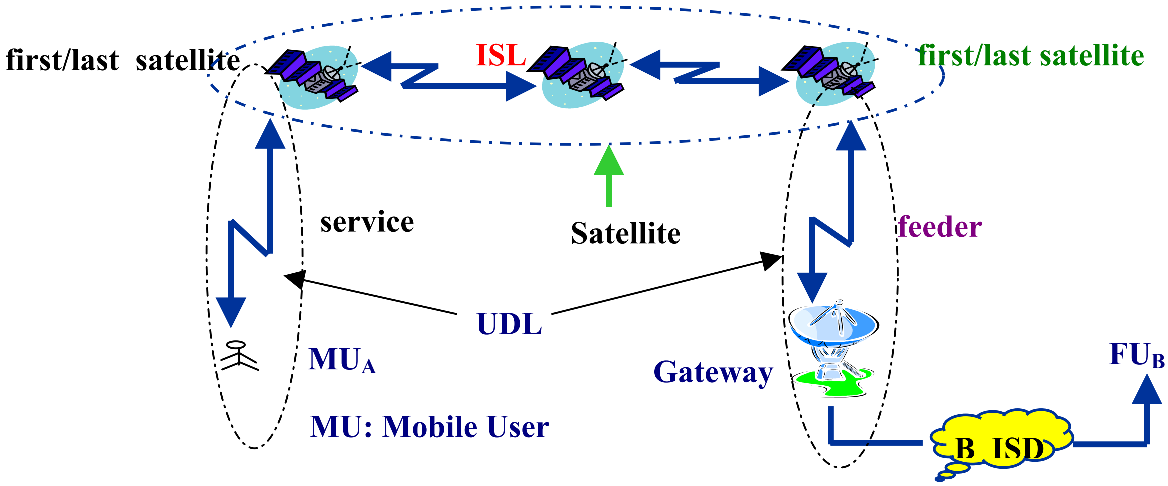


Figure 2-1 Segments of a typical end-to-end connection in ISL-based LEO satellite systems.

Figure 2-1 shows a typical segmentation of end-to-end connection in LEO satellite communication systems with ISLs [1]. The end-to-end connection between a mobile user (MU_A) and a remote fixed user (FU_B) comprises three segments. The *up- /down-link* (UDL) segment incorporates the service link and the feeder link. The ISL segment forms a dynamically meshed sub-network in space. The *terrestrial network link* (TNL) segment connects all terrestrial users together.

Since the LEO satellite covers smaller areas, a large number of LEO satellites are required for establishing a global communication networks. Due to the permanent mobility of the entire satellite network, in particular, ISL sub-networks in space are subject to permanent topological changes. The large number of satellites and the dynamics of the ISL sub-networks in LEO systems make the operation of such systems entails many new challenges to be tackled, especially at the networking level [21]. Consequently, the network traffic control task requires highly sophisticated approaches to provide the predominant connection-oriented services with acceptable QoS for the end users and efficient use of network resources.

In SATM, the ISL sub-network topology change means that some ISLs changes by time, so the delay and cost of the links change correspondingly, or even some ISL disappears when satellites approach polar regions [21]. This variance of the ISL sub-networks significantly increases the complexity of connection-oriented network operation. Some sophisticated traffic control mechanisms have to be employed to achieve high network resource utilization while meeting end users' QoS. For traffic control in ISL sub-networks, as well as the problem caused by stochastically changing traffic flows, the periodically time varying topology gives rise to new challenges for VP distribution mechanism.

In this thesis, the optimization of VP distribution in satellite environment is investigated, and stochastic programming methodology is introduced into the VP optimization. About stochastic programming, more details are shown in the following section.

2.5 Introduction to Stochastic Programming

All network optimization problem can be abstracted as a mathematical programming (MP) problem:

$$\text{minimize } f(X,a)$$

$$\text{subject to } g(X,b) \geq 0$$

where f: $R^n \rightarrow R^1$, g: $R^n \rightarrow R^m$. X is called the decision variable. $f(X)$ is called the objective function and $g(X)$ the constraint function. The feasible region of X is represented by the region enclosed with $g(X) = 0$. Generally, both $f(X)$ and $g(X)$ are expressed analytically in terms of X and a parameter set $\Omega = \{a, b\}$. The MP methodology assumes that all parameter in Ω are deterministic.

Stochastic Programming is an extension of the classical *Mathematical Programming* (MP) that deals with deterministic parameters. The SP extends parameter space to include stochastic parameters. There are two types of problems in stochastic mathematics [22]. The first type assumes that the *probability density function* (pdf) of anticipation is known with certainty, while the second type recognizes the uncertainty of the pdfs of *probability distribution functions* (PDFs). The SP theory has been developed primarily to deal with the first type problem.

A stochastic programming problem can be described as follows:

Minimize $f(X, a_d, a_u)$ (2-1)

subject to $g(X, b_d, b_u) \geq 0$ (2-2)

where: f: $R^n \rightarrow R^1$, g: $R^n \rightarrow R^m$

X: decision variable.

$f(x)$: objective function.

$g(x)$: constraint function.

The feasible region of X is represented by the region enclosed with $g(X) = 0$. Usually, both $f(X)$ and $g(X)$ are expressed analytically in terms of X and a parameter set $\Omega = \{a_d, b_d, a_u, b_u\}$. The a_d and b_d are deterministic parameters while a_u and b_u are uncertain parameters. A particular realization of the uncertain parameters a_u and b_u is called a scenario of the given problem.

Generally, the stochastic programming problems are difficult to solve in real world. Hence, in many cases, the mean values of uncertain parameters are used to solve the problem. In this way, a SP problem is converted to a standard MP problem. In traditional telecommunication networks, this method is generally used, and generates good estimates. However, in ATM, because of the complexity of various traffic types, this simple method cannot provide good estimates.

Within many proposed approaches for solving SP problems, one of the representative approaches to solve SP problems is the *here and now* (HN) method [23]. The philosophy of this method is to make a decision before all scenarios are observed. However, instead of using the moments like means and/or variances, the HN approach integrates the distributions themselves into the formulation. In practice, most network problems are analyzed with simulation techniques rather than analytical models. In some cases, we even cannot give out the distributions of the stochastic parameters. Therefore, the HN approach is hard to realize in practice. However, in situations where the traffic parameters closely follow conventional distributions, the HN method often provides more system insights with less computational work [7].

Another approach is called *Wait and See* (WS), in which the decision variable X is chosen after making an observation for each $\omega \in \Omega$. This means that a set of samples of realized ω is available, and optimization is performed for each sample. Finally, based on a set of optimal solutions, the probability distribution of the optimal solutions is estimated with hypothesis testing techniques. However, we note that the derivation of

PDFs of stochastic parameters is intractable in most applications. Thus the moments of stochastic variables are generally sought instead. The disadvantage of this approach is no single optimal solution provided. For most network optimization problem, we need a single optimal solution, but this approach cannot provide this single optimal solution.

Another alternative approach is the *Scenario Tracking* (ST) [24], in which the stochastic programming problem is converted to a number of standard mathematical programming sub-problems. It takes scenarios (samples) of stochastic parameters, and based on these scenario values, a number of scenario sub-problems are constructed. Optimization then is performed for these scenario sub-problems. Because these sub-problems are MP problems, many well-developed algorithms can be used to solve them. Finally, based on all scenario solutions, a coordinating model is constructed, and another time optimization is performed to track the scenario solutions to get single optimal solution. The ST approach can track all scenario solutions as closely as possible while still maintaining feasibility. In this thesis, the ST approach is chosen to solve the SP optimization for VP system in SATM, as it does not need analytical formulas of distributions of stochastic parameters and can find a single, feasible solution [24]. More detailed information about ST is shown in Chapter 5.

2.6 Summary

In this chapter, we have first reviewed the fundamentals of ATM networks, and VP optimization. In addition, we have reviewed a few of VP optimization schemes in this chapter. Some fundamentals of SATM networks are overviewed here, followed by introduction to stochastic programming.

Chapter 3

Discrete Time Dynamic Virtual Path Allocation

In this chapter, the dynamics of ISL sub-network in SATM networks is analyzed, and then the proposed discrete time dynamic virtual path allocation is presented.

3.1 Dynamics of ISL sub-network in SATM systems

Since the LEO satellite covers smaller areas, a large number of LEO satellites are required for establishing a global communication networks. Due to the permanent mobility of the entire satellite network, in particular, ISL sub-network in space is subject to permanent topological changes. The large number of satellites and the dynamics of the ISL sub-network in LEO systems make the operation of such systems entails many new challenges to be tackled, especially at the networking level [21]. Consequently, the network traffic control task requires highly sophisticated approaches to provide the predominant connection-oriented services with acceptable QoS for the end users and efficient use of network resources.

As examples, the Iridium system and Teledesic are overviewed here. The Iridium system is the first operational system employing ISLs, and it's based on LEO satellites. It's used to provide narrowband personal communications with voice as the strong primary service [25]. Compared to it, the Teledesic system may be regarded as even more forward-looking, already envisaging broadband services and incorporating ATM-like operation. As stated by the company, the Teledesic system will build "Internet in the Sky" [26]. It's used to provide broadband services and support multimedia applications.

Since the Iridium system is the only operating satellite system employing ISLs and based on LEO system, it will be used as example here for demonstration.

In Iridium-like systems, the relative movement of satellites and user location areas on earth leads to a complex combined spatial and time-variant traffic pattern collected and delivered by every single satellite in its footprint [27]. Figure 3-1 shows the 66-satellie constellation with six quasipolar orbits. The specific orbit pattern results in a longitude "seam" encountering between two neighboring orbits where satellites are moving in opposite directions. With respect to this seam, the constellation comprises two hemispherical areas of co-rotating orbits, each extending from the North Pole to the South Pole. The polar view on the Iridium orbits in Fig 3-2 illustrates this phenomenon [21].

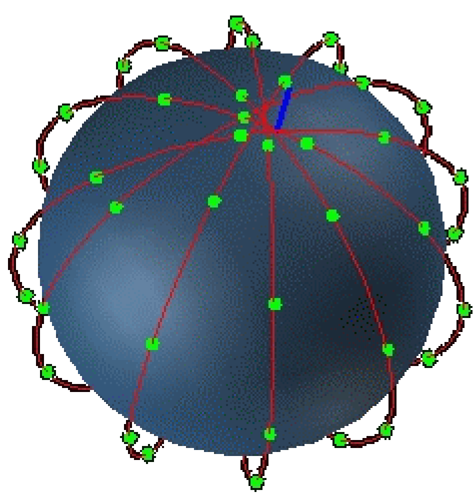


Figure 3-1. Satellite constellation of IRIDIUM system

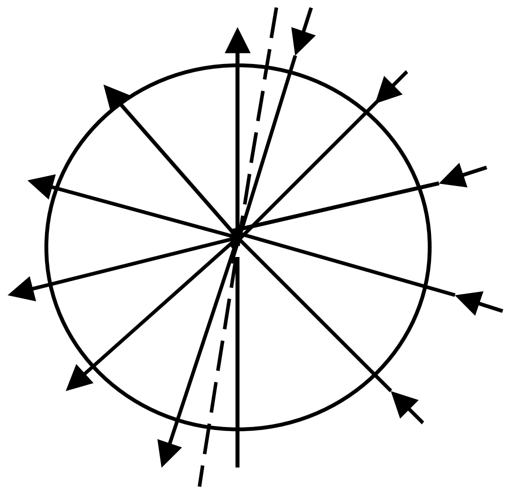
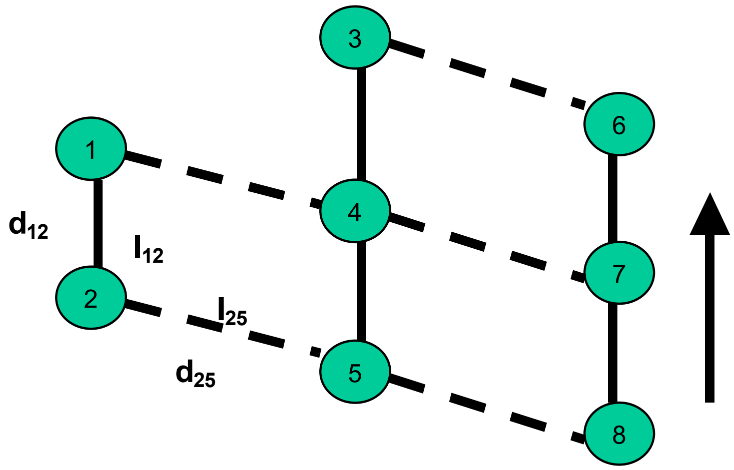
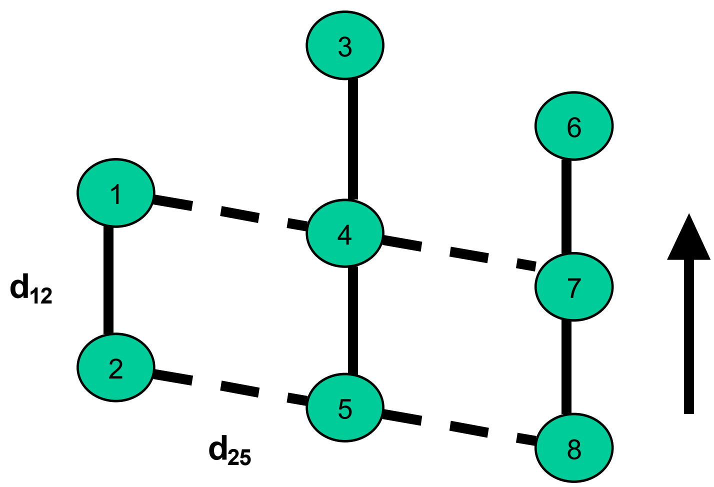


Figure 3-2. Schematic polar view on co- and counter-rotating orbits and the "seam," with arrows indicating the orbital direction of satellite movement

The time varying ISL topology of the system consists of all instantaneously existing direct links between pairs of satellites. In the ISL sub-network, there are two kinds of ISLs: *intra-orbit* ISLs connecting successive satellites in the same orbit plane, and *inter-orbit* ISLs connecting satellites in adjacent co-rotating orbits, as shown in Figure 3-3. Intra-orbit ISLs have fixed distance and antenna pointing, whereas inter-orbit ones are subject to continuous variations of both [1]. Besides the continuous distance changes on co-rotating inter-orbit ISLs there is also a discrete-time contribution to the ISL topology dynamics: inter-orbit ISLs are deactivated in polar regions, meaning on/off switching of links. This twofold (continuous-time and discrete-time) variance of the ISL sub-networks significantly increases the complexity of connection-oriented network operation (i.e., ATM).



(a)



(b)

Figure 3-3. A prototype ISL sub-network topology; (a) topology at time T_0, nodes 3 and 6 are approaching the polar region, (b) topology at time T_1, nodes 3 and 6 are in the polar region. Solid lines denote the intra-orbit ISLs, and dashed lines denote the inter-orbit ISLs. The arrow shows the satellites moving direction.

Considering a prototype ISL sub-networks in Fig 3-3: at time T_0, there is an inter-orbit ISL between node 3 and node 6. After a period of time, at time T_1, the node 3 is in polar region, so the inter-orbit ISL between node 3 and node 6 is deactivated. It can also be noticed that all inter-orbit ISLs between different orbit plane change in distance while all intra-orbit ISLs keep fixed in distance:

I_{xy} : ISL between two satellites x and y.

d_{xy} : length of ISL I_{xy}.

d_{xy} = constant (if x and y are in the same orbit plane)

I_{12} is an intra-orbit ISL whereas I_{25} is an inter-orbit ISL

d_{12} at T_0 = d_{12} at T_1; d_{25} at $T_0 \neq d_{25}$ at T_1.

As stated above, this twofold variance of the ISL sub-network gives rise to a big challenge for VP optimization in SATM networks. Fortunately, the LEO/MEO "physical" topology dynamics is periodically deterministic, with the orbit period of the satellite constellation (100 min for Iridium). Therefore, one possible way is to analytically construct formulas to describe the dynamics of ISL sub-network, and then combine these formulas into VP optimization model. However, the large number of satellites and quite complex satellite movements make this task very difficult to implement. Even though we can incorporate the complex constellation dynamics into VP optimization model, the model will be too complex to be solved.

In this thesis, we choose another approach to conquer the ISL sub-network dynamics. This approach discretizes the continuous dynamics of ISL sub-network, and solves VP optimization for each discrete moment. The details about this approach are presented in following section.

3.2 Discrete Time Dynamic Virtual Path Allocation Scheme

The VP optimization scheme stated here is indeed developed for Iridium-like dynamic ISL constellation, and not just Iridium concept. In SATM system employing

ISLs as presented in the previous section, we face the challenge of a dynamic multiple network topology. In order to utilize the precious satellite network resource efficiently while meeting users' QoS requirements, VPs should be distributed in the network adaptively to the dynamics of the ISL sub-network. This is a challenge for traffic control in SATM, and has to be tackled by specialized VP optimization schemes.

The LEO/MEO "physical" topology dynamics is periodically deterministic, with the orbit period of the satellite constellation. Therefore, for a specific constellation, we can set up a time-dependent optimized virtual path topology for online operation. Werner et al. [1,21] proposed a dynamic routing mechanism for SATM networks. The mechanism exploits the deterministic time-dependent topology to complete the discrete-time virtual topology setup (DT-VTS) procedure. After the procedure, an offline database storing all discrete-time virtual topology is established for the operation of the discrete-time path sequence selection (DT-PSS). The DT-PSS module is used to select path sequences over successive time intervals from the virtual topology provided by DT-VTS. Finally, for every pair of *origin and destination* (OD) nodes, there are predefined paths to connect them, and the information is stored in a database for online operation. This scheme is designed to solve routing problem, and at a specific time, for a specific OD pair, there is only one path to connect them.

However, from traffic control point of view, there may be more than one VP established for an OD pair in the ISL sub-network. These VPs form a virtual network above the physically network. We have to distribute traffic in this virtual network optimally by using VP optimization methods, so high network resource utilization can be achieved while meeting end users' QoS requirements. Based on this idea, a new discrete time dynamic virtual path allocation (DTDVPA) is presented to establish an optimal VP system within ISL sub-network while adaptive to dynamics of ISL sub-network.

In DTDVPA, we propose to exploit the deterministic time-dependent topology to optimize VP system in the ISL sub-network. For a given physical network, a virtual network topology can be set up, specifically for this discussion, a VP topology can be set up. Based on the idea from the work of Werner et al. [1,21], we can get K topologies for an orbit period of the satellite constellation by taking K topology snapshots of the ISL sub-network at different time. For every physical topology, we then do VP optimization to get an optimal VP system. These VP systems are stored in a database for online operation. The whole scheme is an offline process, and can be finished prior to operation. The snapshot process is shown in Fig 3-4.

In Fig 3-2, (a) can be considered as a topology snapshot taken at time T_0 while (b) can be considered as a topology snapshot taken at time T_1. For (a), there is a time tag T_0 on it, and VP optimization methods can be used to get VP system VT_0. For (b), there is also a time tag T_1 on it, and VP optimization methods can be used to get VP system VT_1. Finally, the two VP topologies at time T_0 and T_1 are stored in a database for online operation. During online operation, when time approaches T_0, the corresponding VP system VT_0 is checked by network management entity to distribute VPs in the ISL sub-network. When time approaches T_1, network management entity redistribute VPs in the ISL sub-network according to the corresponding VP system VT_1.

T : period of topology dynamics

K : number of topology snapshots during time T.

$t = T/K$, each of the snapshots at $T_k = kt$, $k = \{0, \ldots, K-1\}$

G(k) : network graph at time T_k (presents the physical topology at time T_k)

VT_k : VP topology at time T_k (virtual topology)

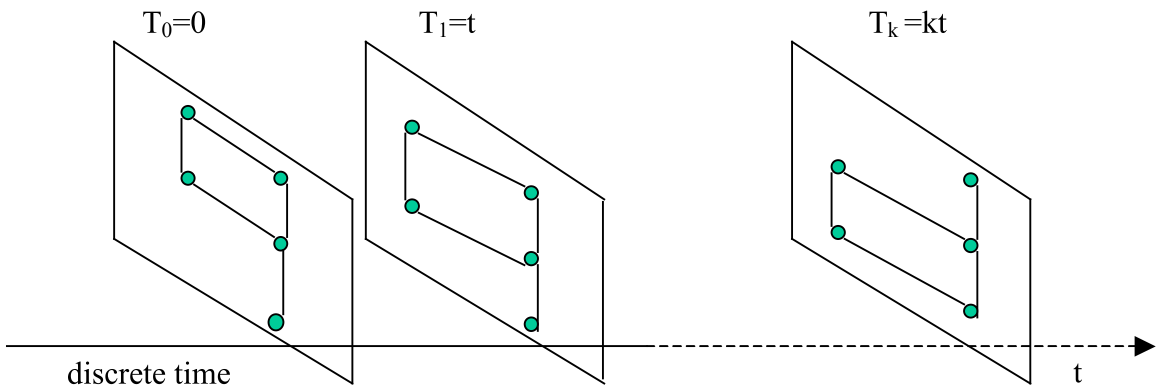


Figure 3-4. Schematic of snapshot of discrete time topology

The DTDVPA process can be summarized as following steps:

1. Take snapshots of the ISL sub-networks of the LEO satellite system. Each snapshot represents the topology of the ISL sub-networks at different time.
2. Optimize the VP system for every snapshot, and store all these VP systems in a database.
3. During operation phrase, the network management entity checks VP system in the database based on current time, and then adjusts the operating VPs according to the stored system.

3.3 Summary

In this chapter, the dynamics of ISL sub-network in SATM systems is presented. The dynamic topology makes the VP optimization difficult in SATM networks. We have proposed the DTDVPA scheme to simplify the VP optimization in SATM networks.

Chapter 4

Virtual Path Flow Assignment Model

In the last chapter, we considered the dynamics of ISL sub-network in SATM networks and proposed the DTDVPA scheme to simplify the VP optimization. For each discrete time snapshot, we considered it as a fixed topology VP optimization problem. In this chapter, we will discuss the flow models used for VP optimization. Firstly, the classic FA model is presented and analyzed. Because of the dominant connection-oriented service in ATM networks, a new VPFA model is proposed for VP optimization. The new model is based on the path-node incidence and incorporates a non-linear objective function. Finally, the VPFA model is compared with the classic FA model.

4.1 The Classic Flow Assignment Model

Before present the FA model, some basic concepts used in graph theory are introduced here for network modeling. A physical network can be defined as a graph, G = (N, A), where N is a finite nonempty set of nodes, and A is a collection of pairs of distinct nodes [4]. We note that A is indeed a set of all links in the network. In Fig. 4-1, a simple network is shown to illustrate the sets A, N and paths as follows:

N = {1, 2, 3, 4}

A = {(1,2), (1,3), (1,4), (2,3), (2,4), (3,4)}

Paths: {1,2}, {1,3,2}, {1,3,4,2}, {1,4,2}, {1,4,3,2} are incident on OD pair (1,2)

Links: (1,3), (2,3) are incident on path {1,3,2}

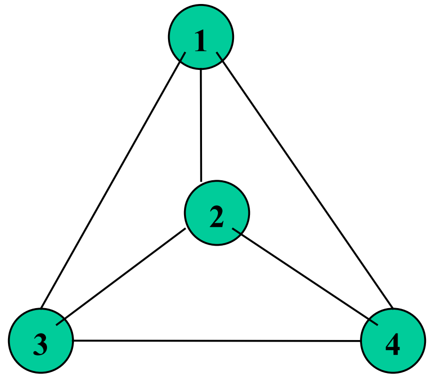


Figure 4-1 A Prototype Network

The FA model [5] typically used in ATM VP optimization is presented as follows:

Minimize $$\sum_{\pi \in W} \sum_{p \in P_w} D_p(x_p) \tag{4-1}$$

subject to

$$\sum_{p \in P_w} x_p = \gamma_w \tag{4-2}$$

$$\sum_{p \in Q_i} x_p \leq C_i \tag{4-3}$$

$$x_p \geq 0 \tag{4-4}$$

$$\forall p \in P_w, w \in W, i \in L$$

where x_p is the design variable.

Notations used in above equations are shown as follows:

W: the set of all OD pairs.

P_w: the set of all paths that connect a particular OD pair w.

Q_i: the set of all paths that pass link i.

x_p: the flow of path p.

$D_{p::}$ the flow cost function of path p.

$C_{i::}$ the capacity of link i.

γ_w: the input flow for OD pair *w*.

L: the set of all links.

The FA model is based on the link-node incidence, which is suitable for traditional data networks. The model describes networks based on link-node relations. In Eq. (4-3), the left side of the equation is link flow expressed as path flows. The generalized cost function D_p in Eq. (4-1) can be expressed in different forms, such as physical distance, unit traffic flow cost, utilization ratio, and the degree of congestion. By extending D_p to different forms, we can get variants of the FA model, and different optimization objectives can be achieved. Hence, we refer to this FA model as *generic flow assignment* (GFA) model. For traditional data networks, connectionless service is dominant, so the network can be described by the relations between the links and the nodes. Therefore, the above link-node incidence based GFA model is suited in traditional data networks.

In ATM networks, the above GFA model is generally used for VP optimization. For example, a variant of the FA model is used in [4] by extending the FA model to including additional features. A flow-based optimization model is also used in [13]. In fact, this flow-based optimization model is a variant of the GFA model. The GFA model is based on the link-node incidence, which is suitable for traditional data networks.

For ATM networks, connection-oriented traffic is dominant, and hence the link-node relation may not be suitable for network modeling. The GFA needs to be modified for ATM network modeling.

4.2 Proposed VP Flow Assignment Model

ATM networks operate on end-to-end connections that can be treated as paths connecting OD pairs. The link-node relation used in the FA model may not be suitable for network modeling. Hence, the formulation coupled with the link-node incidence needs to be extended to the one based on the path-node incidence. With such a formulation associated to OD pairs explicitly, a connection-oriented network can be described directly [7]. In other words, an ATM network can be described by path-node relation rather than link-node relation. For example, when a new connection needs to be set up, the *call admission control* (CAC) has to check if there is enough bandwidth available in VPs connecting the OD pair. If there is enough bandwidth available, the new connection can be accepted; otherwise, the connection request will be rejected. Obviously, the network throughput and congestion are better described by the relations between VP capacity and VP flow than the relations between physical link flow and link capacity.

By changing the modeling concept based on link entities to path entities, we refer the modeling to a virtual network induced from the physical network. In Eq. (4-1) of the FA model, D_p can be expressed in different forms, such as physical distance, unit traffic flow cost, utilization ratio, and the degree of congestion. When we extend the link-node relation in the physical network to the path-node relation in the virtual network, the generalized flow cost function D_p can still be used [19]. Based on the path-node incidence, an induced *virtual path flow assignment* (VPFA) model is proposed as follows:

Minimize $$\sum_{\pi \in W} \sum_{p \in P_w} d_p(\frac{x_p}{G_p}) x_p$$ (total flow cost in the network) (4-5)

Subject to

$$\sum_{p \in P_w} x_p = \gamma_w$$ (the assigned path flows between OD w satisfy the input flows) (4-6)

$$x_p \le G_p$$ (path flow cannot exceed path capacity) (4-7)

$$\sum_{p \in P\pi} G_p \le C_i$$ (the total path capacity passing i cannot exceed its link capacity) (4-8)

$$x_p \ge 0$$ (path flow cannot be negative) (4-9)

$$\forall p \in P_w, w \in W, i \in L$$

where x_p and G_p are design variables. Notations used in the above equations are as follows:

W: the set of all OD pairs.

P_w: the set of all paths that connect a particular OD pair w.

Q_i: the set of all paths that pass link i.

x_p: the flow of path p.

$C_{i::}$ the capacity of link i.

γ_w: the input flow for OD pair w.

L: the set of all links.

d_p : the unit flow cost of path p.

G_p : the capacity of path p.

$\frac{x_p}{G_p}$: the path utilization factor.

In this model, x_p is treated as the VP flow and G_p is treated as the VP capacity. The generalized cost function D_p is expressed in terms of weighted delay. The unit path flow cost d_p can be derived from unit link flow cost. In terms of delay, the unit link flow cost is mainly determined by physical distance of the link. Hence, we can express d_p as the physical distance of path p between an OD pair. The main contribution of the VP flow model is that it introduces two concepts: VP flow and VP capacity. By introducing these two concepts, the VP flow model becomes a *capacity and flow assignment* (CFA) model on logical network level, but on physical network level, it still remains an FA model.

Note that Eq. (4-5) incorporates the path utilization factor and the unit path flow cost d_p, in the VP flow model. The path utilization factor x_p/G_p introduces the degree of congestion into the objective function. Obviously, when x_p/G_p approaches 1, the VP flow is close to the VP capacity. Hence, the delay and the possibility of congestion in this VP increase. By incorporating the utilization factor x_p/G_p into the objective function, we distribute the traffic between an OD pair on all available VPs evenly, and decrease the possibility of congestion. In this model, because the path utilization factor x_p/G_p is kept small while minimizing the total flow delay, the resulting VP system can be resilient to input traffic changes. Therefore, possibility of congestion is likely to be decreased.

Since traffic is distributed evenly on all available VPs between OD pairs, the highest link load can be minimized, and the traffic can be distributed evenly on all links. In ATM networks, the higher the maximum load on any specific link, the more catastrophic may be the effect of the failure of the link. If traffic can be distributed evenly on all links and highest link load can be reduced, the network will have high robustness. Therefore, the resulting VP system has high robustness to physical link failure.

4.3Performance Evaluation of the VPFA Model

The performance of the proposed VPFA model is presented in this section. A small prototype network and a medium size prototype network are chosen for the simulation. In the simulation, we assume that for every OD pair, there is a maximum of two VPs. These two VPs are the shortest and the second shortest paths to connect the OD pair. After simulation, each VP will be assigned a capacity. When a VP is assigned a capacity of zero, the VP virtually does not exist.

The VP systems resulting from the GFA model and the corresponding VP system from VPFA model are compared. For evaluating the performance, the flow cost function D_p in GFA model takes the form of d_p (the unit flow cost of path p), which is same as d_p used in the VPFA model. Therefore, the objective function of GFA model is expressed in terms of the total delay in the network.

Based on the features of the VPFA model, two criteria are used in the performance evaluation. Firstly, redundant capacity assigned in VP system is compared to observe if the VP system is resilient to external traffic changes. Secondly, the highest link load is compared, and the evenness of traffic distribution in the network is compared. We use the link utilization ratio as the indicator of the link load in the simulation. Correspondingly, the standard deviation of link utilization ratio is used to indicate the spreading of the traffic load in the networks. The lower the standard deviation, the higher is the evenness of traffic distribution. After applying one set of external traffic demands on the models, we compare the two models based on above two criteria. In order to compare these two models under different traffic conditions, we use a random number generator to generate uniformly distributed random numbers as external traffic demands, and compare the two models based on standard deviation of link utilization ratio.

The VP optimization for the GFA and VPFA model has been carried out by using the General Algebraic Modeling System (GAMS). Note that the VPFA model is a non-linear model, and the corresponding optimization needs non-linear programming methods. GAMS is a high-level modeling system for mathematical programming problems, and it is capable of solving large-scale linear and non-linear programming models (see Appendix A). GAMS consists of a language compiler and a stable of integrated high-performance solvers [28].

We use two prototype networks for the performance evaluation. The small size network is presented first for tutorial purpose, and it is followed by the medium size network simulation.

4.3.1 Small Size Prototype Network

A prototype network with 4 nodes, 4 links and 6 OD pairs is shown in Fig. 4-2.

In the network, we consider at most two VPs for each OD pair, so there are 11 paths, which are shown in Table 4-1. Table 4-1 shows the all paths associated with OD pairs and corresponding path cost. Table 4-2 shows the path and link incidence. For example, for OD pair (1,2), there are two paths {1,2} and {1,4,2}, which are the first two entries in

Table 4-1. For link (1,2), there are several paths passing through it, such as {1,2,3}, {1,2,4} and so on, which are shown in the first column in Table 4-2. Table 4-3 shows a few selected network parameters that we are using in the simulation.

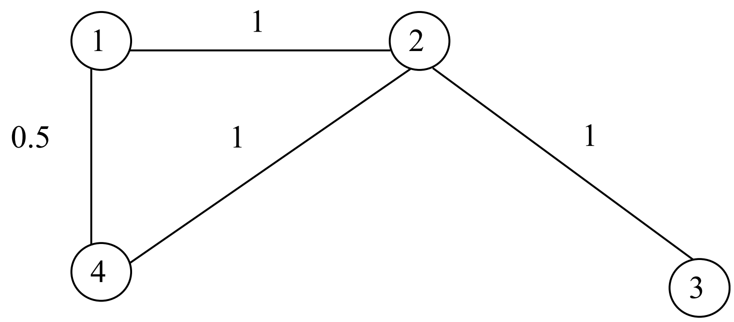


Figure 4-2 A prototype network with 4 nodes, the number above each link denotes the length of this link.

Table 4-1 OD Pairs, Paths and Their Costs.

OD Pair Index	OD Pair	Path Index	Path and Nodes	Path Cost (Distance)
1	(1,2)	P1	{1,2}	1
		P2	{1,4,2}	1.5
2	(1,3)	P3	{1,2,3}	2
		P4	{1,4,2,3}	2.5
3	(1,4)	P5	{1,4}	0.5
		P6	{1,2,4}	2
4	(2,3)	P7	{2,3}	1
5	(2,4)	P8	{2,4}	1
		P9	{2,1,4}	1.5
6	(3,4)	P10	{3,2,4}	2
		P11	{3,2,1,4}	2.5

Table 4-2 Links and paths incidence. "*" is used to indicate the corresponding path passing the link.

Path	Link1 (1,2)	Link2 (2,3)	Link3 (2,4)	Link4 (1,4)
P1	*			
P2			*	*
P3	*	*		
P4		*	*	*
P5				*
P6	*		*	
P7		*		
P8			*	
P9	*			*
P10		*	*	
P11	*	*		*

Table 4-3 Network parameters used in the optimization.

d_p $p = 1,\ldots 11$			C_i , $i = 1,..4$			γ_w , $w = 1,..6$	
d_1	1		C_1	35		γ_1	13
d_2	1.5		C_2	38		γ_2	10.2
d_3	2		C_3	20		γ_3	13
d_4	2.5		C_4	40		γ_4	8
d_5	0.5					γ_5	10.2
d_6	2					γ_6	10.2
d_7	1						
d_8	1						
d_9	1.5						
d_{10}	2						
d_{11}	2.5						

GFA Model

For the prototype network, the GFA model can be formulated as follows:

Minimize $$\sum_{p=1}^{11} d_p x_p \qquad (4\text{-}10)$$

Subject to the following constraints:

the following constraints are obtained from Eq. 4-2

$x_1 + x_2 = \gamma_1$

$x_3 + x = \gamma_2$

$x_5 + x_6 = \gamma_3$

$x_7 = \gamma_4$

$x_8 + x_9 = \gamma_5$

$x_{10} + x_{11} = \gamma_6$

the following constraints are obtained from Eq. 4-3

$x_1 + x_3 + x_6 + x_9 + x_{11} \leq C_1$

$x_3 + x_4 + x_7 + x_{10} + x_{11} \leq C_2$

$x_2 + x_4 + x_6 + x_8 + x_{10} \leq C_3$

$x_2 + x_4 + x_5 + x_9 + x_{11} \leq C_4$

the following constraints are obtained from Eq. 4-4

$x_p \geq 0\ ,\ \ p = 1,...11$

The Eq. 4-10 is optimized using the parameter values given in Table 4-3. After converting the optimization problem into GAMS and running, we get the optimized network flows as shown in Table 4-4.

Table 4-4 Optimized flows for GFA model.

x_1	13.00
x_2	0
x_3	10.20
x_4	0
x_5	13.00
x_6	0
x_7	8.00
x_8	9.80
x_9	0.40
x_{10}	10.2
x_{11}	0

The minimized objective function $\sum_{p=1}^{11} d_p x_p$ =78.70 (with d_p as shown in Table 4-3)

When the GFA Model is used for VP optimization, the optimized network flows are considered optimized VP capacity assignment. It is observed that the flows in Path 2 and 6 are zero. Therefore, all traffic of OD pairs (1,2) and (1,4) are carried by VP1 and VP5 respectively. Here, there is no distinction between VP capacities and VP flows. In other words, VP capacities and VP flows are considered same. Therefore, VP capacities are only optimized to meet external traffic demands γ_w, so there are no redundant capacities assigned to accommodate external traffic changes. Consequently, the resulting VP system is not resilient to traffic changes. In addition, because of the nature of the shortest path algorithm, most traffic flows are distributed on several shortest paths. Therefore, the network traffic is not distributed evenly in the networks.

VPFA Model

For the prototype network shown in Fig. 4-2, the VPFA Model can be formulated as follows:

Minimize $$\sum_{p=1}^{11} d_p\left(\frac{x_p}{G_p}\right)x_p \qquad (4\text{-}11)$$

Subject to the following constraints:

the following constraints are obtained from Eq. 4-6

$x_1 + x_2 = \gamma_1$

$x_3 + x = \gamma_2$

$x_5 + x_6 = \gamma_3$

$x_7 = \gamma_4$

$x_8 + x_9 = \gamma_5$

$x_{10} + x_{11} = \gamma_6$

the following constraints are obtained from Eq. 4-7

$G_1 + G_3 + G_6 + G_9 + G_{11} \le C_1$

$G_3 + G_4 + G_7 + G_{10} + G_{11} \le C_2$

$G_2 + G_4 + G_6 + G_8 + G_{10} \le C_3$

$G_2 + G_4 + G_5 + G_9 + G_{11} \le C_4$

the following constraints are obtained from Eq. 4-8

$x_p \le G_p$, $p = 1,...11$

the following constraints are obtained from Eq. 4-9

$x_p \ge 0$, $p = 1,...11$

The Eq. 4-11 is optimized using the parameter values given in Table 4-3. After converting the optimization problem into GAMS and running, we get the optimized VP capacity and VP flows shown in Table 4-5.

Table 4-5 Optimized VP capacities and VP flows for VPFA model.

Path flow			Path capacity	
x_1	13		G_1	17.44
x_2	0		G_2	0
x_3	10.2		G_3	13.96
x_4	0		G_4	0
x_5	13		G_5	36.41
x_6	0		G_6	0
x_7	8		G_7	11.18
x_8	10.2		G_8	10.74
x_9	0		G_9	0
x_{10}	7.78		G_{10}	9.26
x_{11}	2.42		G_{11}	3.59

The minimized objective function $\sum_{p=1}^{11} d_p(\frac{x_p}{G_p})x_p$ =59.47 (with d_p as shown in Table 4-3)

It is observed that in most cases, the path utilization factor x_p / G_p is low for high traffic load VPs. For example, Paths 5 and 6 have path utilization factor 13/36.41 and 8/11.18 respect. We have put path utilization factor in the objective function, so when the external traffic demands γ_w are below the saturated limit, the VP flows will be kept as low as possible while VP capacities will be assigned in the manner that all link capacities are occupied. In networks, the saturated traffic limit means the maximum amount that can be accommodated by the networks. In this way, the traffic can be distributed in the network evenly, and each VP is assigned some redundant capacity. Therefore, the assigned VP is resilient to external traffic changes. When external traffic goes up, the VP system can accommodate certain amount of extra traffic. For example, Path 5 connects OD pair (1,4). When γ_3 goes from 13 up to 36.41, the extra 23.41 units of traffic can be accommodated by Path 5. In addition, it is observed that the low cost paths have high traffic load. This result is expected, because traffic should be carried through less costly path as much as possible.

In the network, the higher the maximum load on any specific link, the more catastrophic may be the effect of the failure of the link. Therefore, if traffic can be distributed evenly and highest link load can be reduced, the network can have high robustness. In the following part, we compare the traffic distribution of these two models by using path utilization ratio and standard deviation of the ratio.

Comparison of Traffic Distribution in GFA Model and VPFA Model

For presentation convenience, we denote the VP system resulted from GFA model as VP-GFA, and accordingly, the VP system resulted from VPFA model is denoted as VP-VPFA.

Traffic distribution in VP-GFA

Based on the optimal traffic flows in Table 4-4, the link utilization ratio can be calculated and shown in Table 4-6.

Table 4-6 Link utilization ratio in VP-GFA.

	Link1 (1,2)	Link2 (2,3)	Link3 (2,4)	Link4 (1,4)
Flows	23.6	28.4	20.0	13.4
Capacity	35	38	20	40
Utilization ratio	0.674	0.747	1	0.335

Mean of utilization ratio = 0.689

Standard deviation of utilization ratio = 0.238

Traffic distribution in VP-VPFA

Based on the optimal traffic flows in Table 4-5, the link utilization ratio can be calculated and shown in Table 4-7.

Table 4-7 Link utilization ratio in VP-VPFA.

	Link1 (1,2)	Link2 (2,3)	Link3 (2,4)	Link4 (1,4)
Flows	25.62	28.4	17.98	15.42
Capacity	35	38	20	40
Utilization ratio	0.732	0.747	0.899	0.385

Mean of utilization ratio = 0.691

Standard deviation of utilization ratio = 0.188

In the GFA model result, link 4 is used for less than half of its capacity while link3 is occupied in full. However, in the VP Assignment Model result, every link is utilized in an even manner, and the highest link load is reduced. We choose standard deviation of link utilization to express the spreading of utilization, so the lower standard deviation, the higher evenness of traffic distribution. The standard deviation of utilization ratio in GFA model is much higher that in the proposed VP Assignment model. The result shows that the traffic is distributed evenly in the VP Assignment model.

In order to demonstrate that the proposed model distributes traffic more evenly in more general cases, we change the external traffic demand, and then observe the distribution of traffic in the resulting VP systems. Every link load is still expressed in terms of link utilization ratio, and the standard deviation is used to compare if traffic is distributed evenly in the network.

We use a random number generator to generate discrete random numbers between 2 and 11. In the network, there are six OD pairs, so we generate six groups of random numbers independently shown in table 4-8. We combine the data from table 8 and data from table 3, and then run GAMS to get optimal VP assignment for every data set for both GFA and VPA models. The standard deviation of link utilization for every data set is shown in Table 4-9 for comparison between GFA model and the proposed VP Assignment model. The corresponding graph is shown in Figure 4-3.

Table 4-8 Arbitrarily chosen 20 sets external traffic demand. N is set number, γ_w is the external traffic demand.

N	γ_1	γ_2	γ_3	γ_4	γ_5	γ_6
1	7	6	2	6	11	9
2	11	6	2	11	2	8
3	8	4	8	8	2	11
4	11	6	6	3	8	5
5	7	4	3	9	11	3
6	2	8	8	8	9	8
7	7	5	6	6	5	11
8	2	2	3	10	5	2
9	2	11	2	6	4	6
10	6	3	6	7	2	9
11	4	5	8	7	4	8
12	11	4	3	7	7	10
13	8	8	4	8	2	2
14	4	2	9	8	5	11
15	9	9	2	4	3	4
16	8	4	10	9	5	9
17	5	8	5	3	8	5
18	10	5	4	11	11	8
19	11	7	2	6	5	8
20	8	10	7	4	9	6

Table 4-9 Standard deviation comparison of 20 sets of optimal VP assignment.

N	GFA	Proposed CFA		N	GFA	Proposed CFA
1	0.338	0.151		11	0.153	0.094
2	0.219	0.217		12	0.269	0.158
3	0.164	0.127		13	0.151	0.107
4	0.164	0.143		14	0.241	0.102
5	0.216	0.114		15	0.171	0.151
6	0.245	0.096		16	0.157	0.089
7	0.229	0.107		17	0.172	0.112
8	0.126	0.083		18	0.298	0.155
9	0.202	0.179		19	0.224	0.211
10	0.151	0.109		20	0.183	0.147

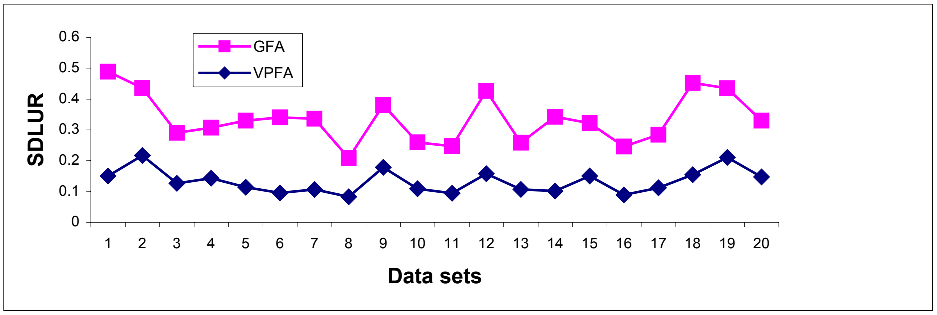


Figure 4-3 Standard deviations of link utilizations ratio (SDLUR) of VP-GFA and VP-VPFA.

It is observed that for all 20 cases, our proposed VPFA model has lower standard deviation. The result shows that the proposed VPFA model can distribute traffic more evenly in the network than the GFA model. Obviously, we can get the conclusion that the VP system resulting from the proposed VPFA Model has higher robustness than the VP system resulting from the GFA Model, because the traffic is distributed evenly and the highest link load is reduced. The comparison result is also graphically shown in Fig. 4-3.

4.3.2 Medium Size Prototype Network

In order investigate the proposed VPFA model on large-scale problem, we apply the GFA model and the proposed VPFA model on a medium size prototype network, which has 10 nodes and shown in Fig. 4-4.

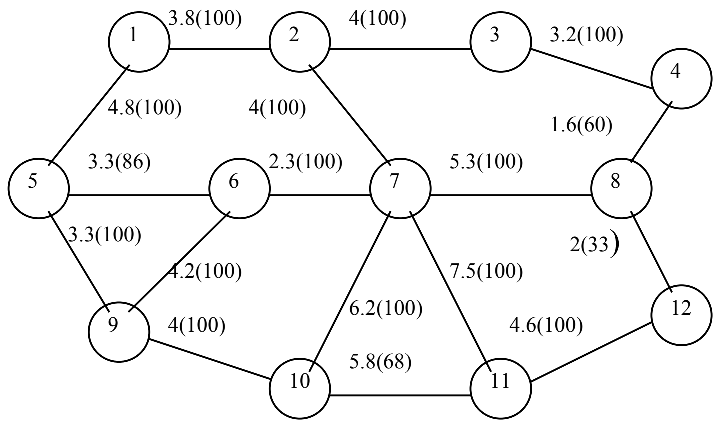


Figure 4-4 A medium size prototype network with 12 nodes. The numbers over each link denote distance (capacity).

For each link, the corresponding link capacity and link distance are shown in the figure. Only 10 OD pairs are considered out of total 45 possible OD pairs, and we assume that for every OD pair, there is a maximum of two VPs. These two VPs are the two shortest paths to connect the OD pair. Therefore, there are totally 20 paths in consideration. After optimization, each VTP will be assigned capacity. When some VPs are assigned capacity of zero, these VPs virtually do not exist. Table 4-10 shows the all paths associated with OD pairs and corresponding path cost. Table 4-11 shows the path and link incidence. Table 4-12 shows some network parameters that we are using in the simulation.

Table 4-10 OD Pairs, Paths and Their Costs.

OD Pair Index	OD Pair	Path Index	Path and Nodes	Path Cost (Distance)
1	(1,7)	P1	{1,2,3}	7.8
		P2	{1,5,6,7}	10.4
2	(2,5)	P3	{2,1,5}	8.6
		P4	{2,7,6,5}	9.6
3	(3,10)	P5	{3,2,7,10}	14.2
		P6	{3,4,8,7,10}	16.3
4	(4,2)	P7	{4,3,2}	7.2
		P8	{4,8,7,2}	10.9
5	(5,10)	P9	{5,9,10}	7.3
		P10	{5,6,9,10}	11.5
6	(6,11)	P11	{6,7,11}	9.8
		P12	{6,9,10,11}	14
7	(7,9)	P13	{7,6,9}	6.5
		P14	{7,10,9}	10.2
8	(8,11)	P15	{8,12,11}	6.6
		P16	{8,7,11}	12.8
9	(9,1)	P17	{9,5,1}	8.1
		P18	{9,6,5,1}	12.3
10	(12,10)	P19	{12,11,10}	10.4
		P20	{12,8,7,10}	13.5

Table 4-11 Links and paths incidence, * is used to indicate the corresponding path passing the link.

Link / Path	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
1	*		*														
2		*		*	*												
3	*	*															
4			*	*	*												
5			*						*	*							
6									*		*	*	*				
7										*	*						
8			*									*	*				
9						*		*									
10				*			*	*									
11					*												*
12					*		*	*								*	
13					*		*										
14								*	*								
15														*	*		
16													*				*
17		*				*											
18				*			*										
19															*	*	
20									*				*	*			

Table 4-12 Network parameters used in the optimization.

C_i, $i = 1,..17$			γ_w, $w = 1,..6$	
C_1	100		γ_1	50
C_2	100		γ_2	50
C_3	100		γ_3	50
C_4	86		γ_4	50
C_5	100		γ_5	50
C_6	100		γ_6	50
C_7	100		γ_7	50
C_8	100		γ_8	50
C_9	100		γ_9	50
C_{10}	100		γ_{10}	50
C_{11}	100			
C_{12}	60			
C_{13}	100			
C_{14}	33			
C_{15}	100			
C_{16}	68			
C_{17}	100			

We perform the optimization for the GFA model and the proposed VPFA model. After converting the optimization problem into GAMS and running, we get optimal result as follows:

GFA Model

For the prototype network in Fig. 4-4, the GFA model can be formulated as follows:

$$\text{Minimize} \quad \sum_{p=1}^{20} d_p x_p \tag{4-12}$$

Subject to the following constraints:

the following constraints are obtained from Eq. 4-2

$x_1 + x_2 = \gamma_1$

$x_3 + x = \gamma_2$

$x_5 + x_6 = \gamma_3$

$x_7 + x_8 = \gamma_4$

$x_9 + x_{10} = \gamma_5$

$x_{11} + x_{12} = \gamma_6$

$x_{13} + x_{14} = \gamma_7$

$x_{15} + x_{16} = \gamma_8$

$x_{17} + x_{18} = \gamma_9$

$x_{19} + x_{20} = \gamma_{10}$

the following constraints are obtained from Eq. 4-3

$x_1 + x_3 \leq C_1$

$x_2 + x_3 + x_{17} \leq C_2$

$x_1 + x_4 + x_5 + x_8 \leq C_3$

$x_2 + x_4 + x_{10} + x_{18} \leq C_4$

$x_2 + x_4 + x_{11} + x_{12} + x_{13} \leq C_5$

$x_9 + x_{17} \leq C_6$

$x_{10} + x_{12} + x_{13} + x_{18} \leq C_7$

$x_9 + x_{10} + x_{12} + x_{14} \leq C_8$

$x_5 + x_6 + x_{14} + x_{20} \leq C_9$

$x_5 + x_7 \leq C_{10}$

$x_6 + x_7 \leq C_{11}$

$x_6 + x_8 \leq C_{12}$

$x_6 + x_8 + x_{16} + x_{20} \leq C_{13}$

$x_{15} + x_{20} \le C_{14}$

$x_{15} + x_{19} \le C_{15}$

$x_{12} + x_{19} \le C_{16}$

$x_{11} + x_{16} \le C_{17}$

the following constraints are obtained from Eq.4-4

$x_p \ge 0$, $p = 1,...20$

The optimized network flows shown in Table 4-13.

Table 4-13 Optimized flows.

x_1	50.00		x_{11}	50.00
x_2	0.00		x_{12}	0.00
x_3	50.00		x_{13}	50.00
x_4	0.00		x_{14}	0.00
x_5	50.00		x_{15}	33.00
x_6	0.00		x_{16}	17.00
x_7	50.00		x_{17}	50.00
x_8	0.00		x_{18}	0.00
x_9	50.00		x_{19}	50.00
x_{10}	0.00		X_{20}	0.00

The minimized objective function $\sum_{p=1}^{20} d_p x_p$ =4430.4 (with d_p as shown in Table 4-10)

When the GFA Model is used for VP optimization, the optimized network flows are considered optimized VP capacity assignment. It is observed that the flows in Paths 2 and 4 are zero. Therefore, all traffic of OD pairs (1,7) and (2,5) are carried by VP1 and VP3 respectively. Because VP capacities and VP flows are considered same, there are no redundant capacities assigned to accommodate external traffic changes. Consequently, the resulted VP system is not resilient to traffic changes. In addition, because of the nature of the shortest path algorithm, network traffic is not distributed evenly in the networks.

VPFA Model

For the prototype network shown in Fig. 4-4, the VPFA Model can be formulated as follows:

Minimize $$\sum_{p=1}^{20} d_p(\frac{x_p}{G_p})x_p \quad (4\text{-}13)$$

Subject to the following constraints:

the following constraints are obtained from Eq. 4-6

$x_1 + x_2 = \gamma_1$

$x_3 + x = \gamma_2$

$x_5 + x_6 = \gamma_3$

$x_7 + x_8 = \gamma_4$

$x_9 + x_{10} = \gamma_5$

$x_{11} + x_{12} = \gamma_6$

$x_{13} + x_{14} = \gamma_7$

$x_{15} + x_{16} = \gamma_8$

$x_{17} + x_{18} = \gamma_9$

$x_{19} + x_{20} = \gamma_{10}$

the following constraints are obtained from Eq. 4-7

$G_1 + G_3 \leq C_1$

$G_2 + G_3 + G_{17} \le C_2$

$G_1 + G_4 + G_5 + G_8 \le C_3$

$G_2 + G_4 + G_{10} + G_{18} \le C_4$

$G_2 + G_4 + G_{11} + G_{12} + G_{13} \le C_5$

$G_9 + G_{17} \le C_6$

$G_{10} + G_{12} + G_{13} + G_{18} \le C_7$

$G_9 + G_{10} + G_{12} + G_{14} \le C_8$

$G_5 + G_6 + G_{14} + G_{20} \le C_9$

$G_5 + G_7 \le C_{10}$

$G_6 + G_7 \le C_{11}$

$G_6 + G_8 \le C_{12}$

$G_6 + G_8 + G_{16} + G_{20} \le C_{13}$

$G_{15} + G_{20} \le C_{14}$

$G_{15} + G_{19} \le C_{15}$

$G_{12} + G_{19} \le C_{16}$

$G_{11} + G_{16} \le C_{17}$

the following constraints are obtained from Eq. 4-8

$x_p \le G_p$, $p = 1,...20$

the following constraints are obtained from Eq. 4-9

$x_p \ge 0$, $p = 1,...20$

The optimized VP capacity and VP flows shown in Table 4-14.

Table 4-14 Optimized VP capacities and VP flows.

Path flow			Path capacity	
x_1	40.954		G_1	42.286
x_2	9.046		G_2	12.453
x_3	50.000		G_3	57.714
x_4	0		G_4	0
x_5	26.721		G_5	35.673
x_6	23.279		G_6	35.673
x_7	40.772		G_7	64.327
x_8	9.228		G_8	22.041
x_9	50.000		G_9	71.204
x_{10}	0		G_{10}	0
x_{11}	50.000		G_{11}	57.714
x_{12}	0		G_{12}	0
x_{13}	29.833		G_{13}	29.833
x_{14}	20.167		G_{14}	28.654
x_{15}	29.814		G_{15}	32.204
x_{16}	20.186		G_{16}	42.286
x_{17}	19.197		G_{17}	28.796
x_{18}	30.803		G_{18}	70.167
x_{19}	50.000		G_{19}	67.796
X_{20}	0		G_{20}	0

The minimized objective function $\sum_{p=1}^{11} d_p(\frac{x_p}{G_p})x_p = 3488.76$ (with d_p as shown in Table 4-10)

It's observed that the resulted VP system from the VPFA model is resilient to traffic changes, because for each VP, they are assigned some redundant capacities while VP flows meet external traffic demands. However, the VP system resulted from the GFA model is not resilient to external traffic changes because the VP capacities are assigned equal to VP flows, and there are no redundant capacities. When traffic is distributed evenly in network and highest link load is reduced, the network has high robustness to link failure. We show the path utilization ratios for two VP systems resulted from GFA model and VPFA model in Table 15 and 16. We choose standard deviation of link utilization to express the spreading of traffic distribution, so the lower standard deviation, the higher evenness of traffic distribution.

Comparison of Traffic Distribution in GFA Model and VPFA Model

For presentation convenience, we denote the VP system resulted from GFA model as VP-GFA, and accordingly, the VP system resulted from VPFA model is denoted as VP-VPFA.

Traffic distribution in VP-GFA

Based on the optimal traffic flows in Table 4-13, the link utilization ratio can be calculated and shown in Table 4-15.

Table 4-15 Link utilization ratio in VP-GFA.

Link	Link flow	Link capacity	Utilization ratio
1	100	100	1
2	100	100	1
3	100	100	1
4	0	86	0
5	100	100	0
6	100	100	0
7	50	100	0.5
8	50	100	0.5
9	50	100	0.5
10	100	100	1
11	50	100	0.5
12	0	60	0
13	17	100	0.17
14	33	33	1
15	83	100	0.83
16	50	68	0.74
17	67	100	0.67

Mean of utilization ratio = 0.671

Standard deviation of utilization ratio =0.348

Traffic distribution in VP-VPFA

Based on the optimal traffic flows in Table 4-14, the link utilization ratio can be calculated and shown in Table 4-16.

Table 4-16 Link utilization ratio in VP-VPFA.

Link	Link flow	Link capacity	Utilization ratio
1	90.954	100	0.910
2	78.242	100	0.782
3	76.904	100	0.769
4	39.849	86	0.463
5	88.879	100	0.889
6	69.197	100	0.692
7	60.636	100	0.606
8	70.167	100	0.702
9	70.167	100	0.702
10	67.493	100	0.675
11	64.051	100	0.641
12	32.507	60	0.542
13	52.692	100	0.527
14	29.814	33	0.903
15	79.814	100	0.798
16	50.000	68	0.735
17	67	100	0.702

Mean of utilization ratio = 0.708

Standard deviation of utilization ratio = 0.125

In the GFA model result, link13 is used for less than one third of its capacity while link2, link3, and link10 are occupied in full. However, in the VP Assignment Model result, every link is utilized in an even manner, and the highest link load is reduced. We choose standard deviation of link utilization to express the spreading of utilization, so the lower standard deviation, the higher evenness of traffic distribution. The standard deviation of utilization ratio in GFA model is much higher that in the proposed VP

Assignment model. The result shows that the traffic is distributed evenly in the VP Assignment model.

In order to demonstrate that the proposed model distributes traffic more evenly in more general cases, we change the external traffic demand, and then observe the distribution of traffic in the resulting VP systems. Each link load is still expressed in terms of link utilization ratio, and the standard deviation is used to compare if traffic is distributed evenly in the network. We use a random number generator to generate 20 sets of external traffic demands ranging from 39 to 58. Because there are 10 OD pairs in the network, we generate 10 groups of random numbers independently (20 sets, each set including 10 numbers) as shown in Table 4-17.

Table 4-17 Arbitrarily chosen 20 sets external traffic demand. N is set number, γ_w is the external traffic demand.

N	γ_1	γ_2	γ_3	γ_4	γ_5	γ_6	γ_7	γ_8	γ_9	γ_{10}
1	39	56	46	53	55	58	53	49	55	58
2	42	56	47	49	39	41	46	51	46	39
3	46	45	58	56	54	53	41	50	55	40
4	49	43	42	48	43	47	39	48	40	47
5	54	54	45	47	55	42	55	46	46	58
6	58	51	48	45	40	49	55	51	57	54
7	42	53	40	41	47	50	47	58	44	47
8	48	39	56	58	56	46	58	47	42	45
9	58	49	43	42	55	44	47	50	54	54
10	40	45	55	45	52	50	46	48	41	58
11	54	44	56	51	50	46	50	46	39	51
12	57	57	52	55	57	51	57	58	54	54
13	49	50	50	53	51	43	46	48	51	54
14	53	51	45	41	39	50	44	39	49	51
15	52	42	48	39	39	51	39	46	50	58
16	52	57	58	55	48	56	40	50	58	52
17	51	48	54	52	57	40	41	40	45	45
18	56	54	58	39	57	50	46	57	50	47
19	42	53	43	56	55	49	39	45	56	45
20	56	44	52	53	56	45	41	54	43	50

Table 4-18 Standard deviation comparison of 25 sets of optimal VP assignment.

N	GFA	Proposed CFA		N	GFA	Proposed CFA
1	0.323	0.145		11	0.304	0.126
2	0.325	0.124		12	0.275	0.142
3	0.278	0.108		13	0.333	0.115
4	0.317	0.129		14	0.349	0.141
5	0.341	0.145		15	0.345	0.150
6	0.288	0.150		16	0.273	0.131
7	0.321	0.156		17	0.324	0.102
8	0.284	0.130		18	0.279	0.140
9	0.326	0.145		19	0.321	0.113
10	0.334	0.140		20	0.308	0.126

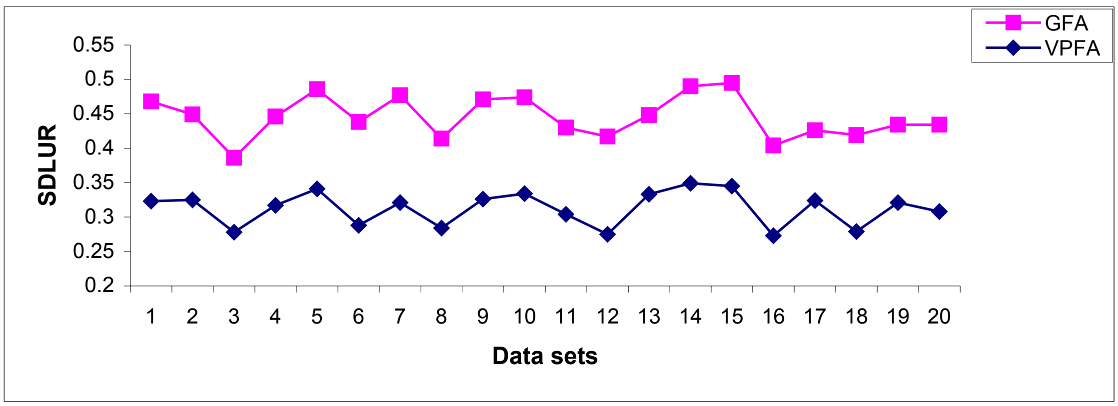


Figure 4-5 SDLUR of the VP-GFA and the VP-VPFA for middle network.

The standard deviation of link utilization for every data set is shown in Table 4-18 for comparison between GFA model and the proposed VP Assignment model. The corresponding graph is shown in Fig. 4-5. It is observed that for all 20 cases, our proposed VPFA model has lower standard deviation. The result shows that the proposed VPFA model can distribute traffic more evenly in the network than the GFA model. Obviously, we can get the conclusion that the VP system resulting from the proposed VPFA Model has higher robustness than the VP system resulting from the GFA Model, because the traffic is distributed evenly and the highest link load is reduced.

4.4 Summary

In this chapter, the proposed VPFA model is presented and compared with GFA model. The simulation shows that the proposed VPFA model has following features:

- ***The resulting VP system can be resilient to external traffic changes.***
- ***The resulting VP system is robust to physical link failure.***

Chapter 5

Weighted Scenario Tracking Model

In this chapter, the scenario tracking is introduced for VP optimization. A weighted scenario tracking scheme is proposed here.

5.1 Introduction to Scenario Tracking

When flow models are used to optimize VP allocation in ATM, the input flow γ_w in Eq. (4-2) can be obtained by aggregating incoming calls in terms of bandwidth under certain blocking rate. However, ATM supports various traffic types, such as CBR, VBR, ABR, and UBR, which have different characteristics and may exist simultaneously. In real networks, it is difficult to describe all types of traffic by equivalent bandwidth. In addition, when VCs are aggregated to VPs, statistical multiplexing decreases the required bandwidth to accommodate the traffic. This twofold complexity makes it difficult to describe the incoming calls in terms of equivalent bandwidth demand. Different applications and high bursty traffic make traffic flows entering networks dynamic and stochastically changing. Therefore, traditional mean value method or equivalent bandwidth method may not provide valid solution.

In the paper, we abstract the VP optimization as a *stochastic programming* (SP) problem. As we state in Chapter 2, the ST method does not need analytical formulas of distribution of stochastic parameters and can find a single optimal solution. These features make ST method feasible and flexible for VP optimization when comparing with other approaches. Therefore, we choose the ST method for stochastic optimization [24]. The ST approach is introduced as follows:

As we denote in Chapter 2, a stochastic programming problem can be described as follows:

Minimize $f(X, a_d, a_u)$ (2-1)

Subject to $g(X, b_d, b_u) \geq 0$ (2-2)

Where: f: $R^n \rightarrow R^1$, g: $R^n \rightarrow R^m$

X: decision variable.

$f(x)$: objective function.

$g(x)$: constraint function.

Let a_s and b_s represent a particular joint realization of uncertain parameters, a_u and b_u, respectively. For each scenario $s \in S \equiv \{\text{set of all scenarios}\}$, the SP formulations, expressed by (2-1) and (2-2), reduces to a deterministic problem shown below, which is referred to as the *scenario sub-problem*:

Minimize $f(X, a_d, a_s) = y_s$ (5-1)

Subject to $g(X, b_d, b_s) \geq 0$ (5-2)

For each scenario, there is a corresponding probability p_s. In such an environment, a solution x to a stochastic nonlinear system $g(X, b_d, b_s) \geq 0$ is said to be feasible if it minimizes:

$$\sum_s p_s[\|\min\{0, g(X, b_d, b_s)\}\|] \quad (5\text{-}3)$$

Accordingly, a coordination model can be constructed as:

$$\text{Minimize} \quad \sum_{s} p_s[\| f(X, a_d, a_s) - y_s \| + \| \min\{0, g(X, b_d, b_s)\} \|] \qquad (5\text{-}4)$$

The coordinating model tracks the scenario solutions as closely as possible while still maintaining feasibility, and this model is referred as tracking model [24].

In summary, the ST approach to stochastic programming has two stages:

Stage1: Compute a solution to the (deterministic) problem under all scenarios.

Stage2: Solve a coordinating or tracking model to find a single, feasible solution.

5.2 Weighted Scenario Tracking

In this thesis, we apply the ST scheme to GFA model (we present in Chapter 4.1) to solve the VP optimization problem. Here, the input flow γ_w is assumed to be a stochastic variable. Scenario values of γ_w are taken, and for every scenario, solve the scenario sub-problem, which is indeed a deterministic problem. After we get solutions for all scenarios, construct a coordinating model based on all scenarios to track the scenario solutions as closely as possible while still maintaining feasibility. Finally, the optimal solution is calculated based on the tracking model. The tracking model consists of two parts: "*objective value*" and "*feasibility penalty*" cost function.

OBJECTIVE VALUE

The objective value cost function is used to track all scenario solutions by taking summation of difference between the final objective value and every scenario objective value.

Let the final optimal objective value be denoted by $G = \sum_{\pi \in W} \sum_{p \in P_w} dp(xp)$

$$Z_{obj} = | G - ys | \tag{5-5}$$

where x_p is the design variable; n: the number of OD pairs in W; S: the set of all scenarios; y_s: objective function value at scenario s; the remaining notations are same as defined in section Chapter 4.

We can see that ys is the objective function value at scenario s. This objective value part shown in Eq. (5-5) is used to track all scenario solutions to find the final single solution by minimizing summation of difference between the final solution and every scenario objective value. This part is to track the scenario solutions as closely as possible.

FEASIBILITY PENALTY

This cost function is used to reflect the feasibility into the tracking model by summating the negative distance between the final solution and each scenario input flow.

Let the lowest difference between the final solution and scenario input flow is denoted by $D_s = \min[((\sum_{p \in P_w} x_p) - \gamma_{ws}), w = 1..n]$

$$Z_{penalty} = | \min[\ 0, D_s\ ] | \tag{5-6}$$

where x_p is the design variable; n: the number of OD pairs in W; S: the set of all scenarios; γ_{ws}: the input flow for OD pair w at scenario s; the remaining notations are same as defined in Chapter 4.

In Eq. (5-6), D_s is used to determine the lowest difference (which can be negative or positive) between the final solution and every scenario input flow γ_{ws}. Note that in networks, only when path flows between an OD pair cannot meet external input traffic flows ($\sum_{p \in P_w} x_p < \gamma_{ws}$), the feasibility of the flow assignment is violated. Hence, we count only these negative distances here. We then compare the lowest value D_s with 0 to get the summation of all negative distances. Finally, we get absolute value of this summation. By doing this, we reflect the feasibility into tracking model by minimizing the largest negative distance between final solution and every scenario input flow γ_{ws}. By adding the feasibility penalty into the tracking model, we are trying to maintain the feasibility while tracking the scenario solutions.

The idea of ST method is to construct a tracking model to track the scenario solutions as closely as possible while still maintaining feasibility. Hence, the complete tracking model can be constructed as follows by combining Eqs. (5-5) and (5-6):

Minimize: $$\sum_s ps[aZ_{obj} + bZ_{penalty}] \tag{5-7}$$

Subject to: $$\sum_{p \in Q_i} x_p \le C_i, \quad ys = \sum_{\pi \in W} \sum_{p \in P_w} dp(x_{ps})$$

$$\forall p \in P_w, w \in W, i \in L, s \in S$$

The fully extended form of Eq. (5-7) is shown as follows:

$$\sum_s ps[a\underbrace{|\sum_{\pi \in W} \sum_{p \in P_w} dp(xp) - ys|}_{Z_{obj}} + b\underbrace{|\min[0, \min[(\sum_{p \in P_w} x_p - \gamma_{ws}), w = 1..n]]|}_{Z_{penalty}}] \tag{5-8}$$

where p_s: probability of scenario s; a: weighting factor of objective value tracking; b: weighting factor of feasibility tracking.

The proposed *weighted scenario tracking* (WST) as shown in Eq. (5-7) has two weighting factors a and b in the tracking model, which give us flexibility to adjust the tracking performance. The factor a is used to adjust the weight of Z_{obj}, which reflects the influence of the summation of difference between the final solution and every scenario objective value. The factor b is used to adjust the weight of $Z_{penalty}$, which reflects the influence of the summation of the negative distance between final solution and every scenario input flow γ_{ws}. By adjusting factors a and b, we change the weight of objective value part and feasibility part in the tracking model to meet different objectives.

5.3 Performance Evaluation

We apply the WST method to GFA model, and then compare the resulting VP system with the VP system obtained from mean values. In our research, we assume that the external traffic demands, expressed in terms of bandwidth, change randomly. A random number generator is used to generate 10 sets of traffic demands with Poisson distribution. The mean value is set to 50. Each group corresponds to external traffic demand of an OD pair, and each group includes 64 numbers corresponding to 64 scenarios. In WST approach, the more scenarios we use, the higher accuracy we obtain. However, if we choose a large number of scenarios, the computation will take long time and the tracking model will be difficult to solve. Hence, in our experiment, we arbitrarily choose only 64 scenarios. The medium size network shown in Fig. 4-4 is again used for the simulation. The network parameters used are shown in Table 5-1.

Table 5-1 Network parameters used in the optimization

C_i , $i=1,..17$			γ_w , $w=1,..6$ (Mean value)	
C_1	100		γ_1	50
C_2	100		γ_2	50
C_3	100		γ_3	50
C_4	86		γ_4	50
C_5	100		γ_5	50
C_6	100		γ_6	50
C_7	100		γ_7	50
C_8	100		γ_8	50
C_9	100		γ_9	50
C_{10}	100		γ_{10}	50
C_{11}	100			
C_{12}	60			
C_{13}	100			
C_{14}	33			
C_{15}	100			
C_{16}	68			
C_{17}	100			

Tracking Model:

$$\sum_{s} ps[a|\underbrace{\sum_{\pi\in W}\sum_{p\in P_w} dp(xp) - ys|}_{Z_{obj}} + b|\underbrace{\min[\,0, \min[(\sum_{p\in P_w} x_p - \gamma_{ws}), w = 1..n]]}_{Z_{penalty}}|]$$

Weighting factors a, b =1

Initially, *a* and *b* in the tracking model are set as 1. For the ST tracking method, firstly, we perform 64 times optimization for 64 scenarios based on the GFA model. Secondly, Based one the tracking model, we perform another time of optimization to get the optimal solution. The optimal solution is compared with the solution obtained from applying mean values for external traffic demands. The comparison is shown in Table 5-2.

Table 5-2 Comparison of results obtained from ST approach and mean value approach

VP capacity (ST approach)			VP capacity (mean value)	
x_1	51.121		G_1	50.000
x_2	0		G_2	0
x_3	48.789		G_3	50.000
x_4	0.571		G_4	0
x_5	46.141		G_5	50.000
x_6	0		G_6	0
x_7	52.934		G_7	50.000
x_8	0		G_8	0
x_9	47.159		G_9	50.000
x_{10}	0		G_{10}	0
x_{11}	54.085		G_{11}	50.000
x_{12}	0		G_{12}	0
x_{13}	45.344		G_{13}	50.000
x_{14}	3.171		G_{14}	0
x_{15}	33.000		G_{15}	33.000
x_{16}	16.302		G_{16}	17.000
x_{17}	51.205		G_{17}	50.000
x_{18}	0.963		G_{18}	0
x_{19}	51.127		G_{19}	50.000
X_{20}	0		G_{20}	0

The external traffic demands are random variables. At some instances, the VTPs may not accommodate external traffic demands, and hence certain amount of traffic will be rejected while the network still has spare capacity. In order to compare the resilience of the solutions obtained from the WST approach and the mean value approach, we compare the failure instances at which the VPs between an OD pair fail to accommodate the entering external traffic demand. The higher the failure instances, the lower the resilience to randomness of external traffic demands.

In order to compare the failure times, we generate Poisson numbers of mean value at 50 as external traffic demands, and then compare the failure frequency of the two solutions. We use failure times of mean value approach minus failure times of ST approach as comparison parameter. The comparison result is shown in Table 5-3. The result is also graphically shown in Fig 5-1.

Table 5-3 Comparison of failure times of mean value approach and ST approach.

Number of instances	32	64	96	128	160	192	224	256
Failures of mean	144	310	462	608	756	883	1027	1190
Failures of ST	143	305	457	600	744	865	1007	1166
Difference of failures (mean – ST)	1	5	5	8	12	18	20	24

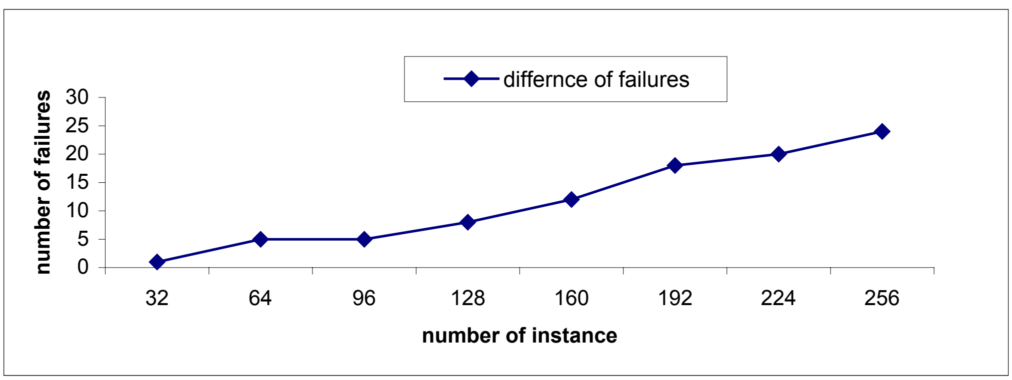


Figure 5-1 the difference of failure times of mean value approach and ST approach.

It is observed that the ST approach has fewer failures than the mean value approach. The above comparison also shows that the higher the number of instances, the larger the difference. When we have more instances, if the ST approach has higher accuracy to express the randomness of external traffic, the failure difference between the ST approach and mean value approach should increase. The above comparison shows the same trend.

Weighting Factors a, b Adjustment

However, for most cases, the improvement is less than 2%. The improvement is trivial, considering the computation expense. It is observed that we add two weighting factors in the tracking model: *a* and *b*. The *a* is used to adjust the weight of objective values in the model, and the *b* is used to adjust the weight of feasibility in the model. Because we want to have less network failures, we could increase *b* and decrease *a* to achieve it. We try different combinations of *a* and *b*, and compare the failure instances. The comparison is based on 128 scenarios, and the comparison result is shown in Table 5-4. Fig. 5-2 graphically shows the result.

Table 5-4 Comparison of failures for different *b*/*a* ratios (128 scenarios)

b/*a*	80	100	120	140	160	180	200
Failures of mean	608	608	608	608	608	608	608
Failures of ST	545	484	487	475	465	444	271
Difference of failures (mean – ST)	63	124	121	133	143	164	337

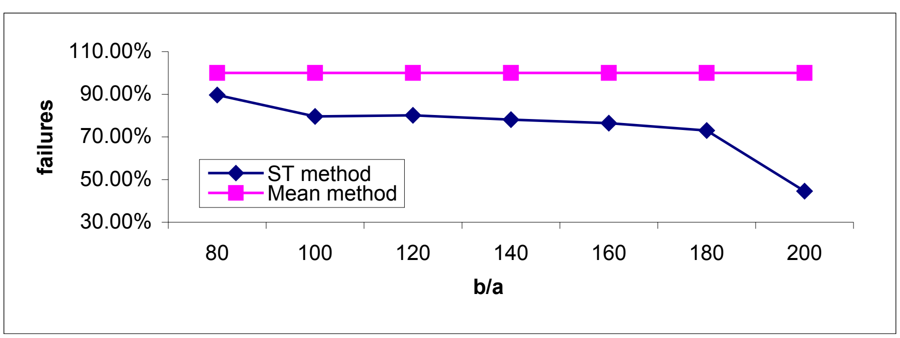


Figure 5-2 Comparison of failures for different *b*/*a* ratios.

Based on the above simulation, it is observed that the WST approach shows great flexibility on solving stochastic programming problem. Based on different objectives, we can construct different tracking model. By adding two weighing factor *a* and *b* into the tracking model, we provide flexibility for users to change the weight of objective value part and feasibility part in the tracking model to meet their objectives. The ratio of *b/a* can be increased in order to obtain a higher resilience to external traffic fluctuation.

5.4 Summary

In this chapter, the weighted scenario tracking scheme is proposed for VP optimization. It is compared with the mean value approach. The simulation results show that the WST approach shows better performance on VTP optimization than mean value approach in terms of fewer failures to accommodate external traffic changes.

Chapter 6

Conclusions and Future Work

6.1 Conclusions

ATM has played a major role in current terrestrial B-ISDN infrastructure. A satellite ATM network has been envisioned as the next information super-skyway, and has been included in the Global Information Infrastructure. In SATM networks, the virtual path distribution plays an important role for high network efficiency while meeting users' QoS requirements. Because of the dynamics of the *Inter-satellite Link* (ISL) sub-network and stochastically changing traffic flows, the optimization of VP distribution becomes a difficult problem.

This thesis has been concerned about the dynamic virtual path optimization in satellite ATM networks with stochastic traffic flows. The main contributions of this thesis are as follows:

1. A new *discrete time dynamic virtual path allocation* (DTDVPA) scheme has been proposed to establish an optimal VP system within ISL sub-network while adaptive to dynamics of ISL sub-network. The basic idea is to exploit the fact that all topology dynamics is periodically deterministic. The DTDVPA allows the setting up off-line of a time-dependent virtual topology for online dynamic virtual path selection.

2. A network flow model is proposed for dynamically managing VP system in ATM networks. The VPFA model is based on path-node incidence and introduces two concepts: the VP flow and the VP capacity. On logical network level, the

proposed model is a capacity and flow assignment model while on physical network level, it remains a FA model. Because the VP utilization ratio is incorporated into the non-linear delay function, the resulting VP system has redundant capacity to accommodate extra traffic, and hence the VP system is resilient to external traffic changes. In addition, the highest link load is decreased and traffic is distributed evenly in the network, and hence the VP system is robust to physical link failure.

3. The stochastic programming methodology has been introduced for VP optimization in SATM networks. A weighted scenario tracking (WST) approach is simulated and compared with the mean value method. It is observed that the VP system resulting from the proposed approach has higher resilience to external traffic changes. By introducing two weighing factors *a* and *b* into the tracking model, we provide users flexibility to change the tracking model to meet different objectives.

6.2 Future Work

The VPFA model that has been proposed in this thesis for ATM networks can be extended to use for general connection-oriented networks, because the general concepts used here can be extended for any connection-oriented networks. For example, the VPFA model can be extended to use in IP with IntServ and MPLS networks [29].

Traditional IP networks only provide the *best-effort service* and there is no QoS support. However, most real-time applications call for the guaranteed service. To accept this challenge, in the last decade, a new IP architecture, *Integrated Services* (IntServ) was proposed. IntServ [30] is a per-flow based framework with dynamic resource reservation and provides the end-to-end QoS. Based on IntServ, IP networks evolved to support connection-oriented service. A typical IntServ network uses the *Resource Reservation Protocol* (RSVP) [31] as the signalling protocol to reserve network resources. With reservation, a *virtual traffic path* (VTP) is established for upcoming traffic. These VTPs need to be optimized to achieve high network efficiency while satisfying the QoS requirements.

On the other hand, *Multiprotocol Label Switching* (MPLS) tries to separate the forwarding component from the control component in the IP networks. This way, MPLS may have a significant potential to shorten the time taken to develop and deploy new routing functionality in IP networks [32]. Better price/performance characteristics in router, seamless integration IP over ATM, and good scalability make MPLS a promising networking technology since its emergence. In MPLS, traffic is transmitted through the *label switched path* (LSP). Similar to a VP in ATM, an LSP is an abstraction of network flows in MPLS. LSP distribution on the network flow level is a critical issue for high network efficiency and guaranteed QoS requirements.

The VP in ATM, LSP in MPLS, and end-to-end virtual traffic path in IP with IntServ are all virtual traffic paths in networks. For these connection-oriented networks, these virtual traffic paths are abstraction of network flows. Therefore, by considering features of these connection-oriented networks, the proposed VPFA model can be extended for VTP optimization in IP with IntServ and LSP optimization in MPLS networks.

The VPFA model is a network flow level model, where all traffic is abstracted as network flow, and the arrival process and queuing process are ignored. In many applications, the arrival process and queuing process have to be taken into account. Therefore, the VPFA model can be extended to include arrival process and queuing process to solve more complicated problems.

Bibliography

1. E. Lutz, M. Werner, and A. Jahn, *Satellite Systems-for Personal and Broadband Communications*, Springer-Verlag Berlin Heidelberg, 2000.
2. R. G. Addie, J. L. Burgin, and S. L. Sutherland, "B-ISDN Protocol Architecture," *Proc. Globecom 88*, Paper 22.6.1, Florida, December 1988.
3. N. Anerousis, A. A. Lazar, "Virtual Path Control for ATM Networks with Call Level Quality of Service Guarantees," *IEEE/ACM Transaction on Networking*, VOL 6, NO.2, April 1998.
4. P. Jiyapanichkul, and J. Lambert, "Optimal Resource Management with Dynamic Virtual Paths in ATM Networks," *Proc. of IEEE International Conference on Communications*, ICC. V3, pp. 1728-1732, 1998.
5. D. Bertsekas, and R. Gallager, *Data Networks*, 2nd edition, Prentice Hall, NJ, 1992.
6. C-K Toh and V. O .K .Li, "Satellite ATM Network Architectures: An Overview," *IEEE Network*, pp. 61-71, September/October 1998.
7. D. Zheng, X. Liu, and M. Mandal, "Virtual Path System Optimization in ATM Networks with Stochastic Traffic Flows," *Proc. of SPIE: Network Design and Management*, Vol. 4909, Shanghai, China, Oct 2002.
8. D. Zheng, X. Liu, and M. Mandal, "Efficient Virtual Path Allocation for Stochastic Traffic in ATM Networks," submitted to ICC 2003, Anchorage, AK, USA, May 11-15, 2003
9. P.S. Neelakanta, *ATM Telecommunication – Principles and Implementation*, CRC Press, Boca Raton, 2000.
10. D.E. McDysan and D.L. Spohn, *ATM Theory and Applications*, McGraw-Hill, Inc.,

New York, NY, 1998.

11. G. Kesidis. *ATM Network Performance*. (second edition), pp 2, 2000.

12. O. kyas. *ATM Networks* (second edition), pp 151-177,1996

13. S. Ahn, R. Tsang, S-R. Tong, and D. Du, "Virtual Path Layout Design on ATM Networks," *Proc. IEEE INFOCOM '94*, pp.192-200, 1994.

14. J. Jiau, "Study of integrated bandwidth allocation and multiplexing strategy for VP-based ATM networks," *Proceedings of the 14th Annual Int'l Phoenix Conf. On Computers and Communications*, pp226-232, 1995.

15. A. Alleles. *ATM internetworking*. Cisco Systems, Inc. report, 1995

16. I. Chlamtac, A.Farago, and T. Zhang, "Optimizing the System of Virtual Path," *IEEE/ACM Trans, Networking*, Vol. 2, no.6, pp. 581-587, Dec, 1994.

17. B. H. RYU, H. OHSAKI, M. MURATA, and H. MIYAHARA, "Design Algorithm for Virtual Path Based ATM Networks," *IEICE Trans. Commun*., Vol. E79-B, NO.2 Feb. 1996.

18. R. Guerin, H. Ahmadi, and M. Naghshineh, "Equivalent capacity and its application to bandwidth allocation in high-speed network," *IEEE J. Select. Areas in Commun*., Vol.9, no.7, pp.968-981, Sep. 1991.

19. Y. Li, D. Zheng, X. Liu, and M. Mandal, "Novel Congestion Control Techniques for Satellite ATM Networks," *Proc. of SPIE: Network Design and Management*, Vol. 4909, Shanghai, China, Oct 2002.

20. Ian F. Akyildiz and S-H. Jeong, "Satellite ATM Networks: A Survey," *IEEE Communications Magazines*, pp30-43, July 1997.

21. M. Werner, "A Dynamic Routing Concept for ATM-Based Satellite Personal

Communication Networks," *IEEE Journal on Selected Areas in Communications*, pp. 1636-1648, October 1997.

22. G. Tintner, "The Pure Theory of Production under Technological Risk and Uncertainty", *Econometrica* 9 (1941), pp. 298-304.
23. M. Dempster, *Stochastic programming*, Academic Press, NY, 1980.
24. D. Dembo, "Scenario Optimization," *Annals of Operations Research*, Vol. 30, pp. 63-80, 1991.
25. K. Maine, C. Devieux, and P. Swan, "Overview of IRIDIUM Satellite Network," Motorola Satellite Communications Division, USA. ISBN 0-7803-2636-9
26. Teledesic web site www.teledesic.com.
27. M. Werner, "Analysis of system connectivity and traffic capacity requirements for LEO/MEO S-PCN's," in Mobile and Personal Commun., pro. 2nd Joint COST 227/231 Workshop, E. Del Re, Ed., Florence, Italy, Apr. 1995, Elsevier, pp. 183-204.
28. A. Brooke, D. Kendrick, A. Meeraus, and R. Raman, *GAMS, A User's Guide*, GAMS Development Corporation, 1998.
29. D. Zheng, X. Liu, and M. Mandal, "Virtual Traffic Path optimization in Connection-oriented Networks with Stochastic Traffic," submitted to *Journal of Network and Systems Management*, Nov 2002.
30. R. Braden, D. Clark, and S. Shenker, "Integrated services in the Internet architecture: an overview." Request for Comments (Informational) 1633, Internet Engineering Task Force, June 1994.
31. R. Braden, Ed., L. Zhang, S. Berson, S. Herzog, and S. Jamin, "Resource

ReSerVation protocol (RSVP) – version 1 functional specification," Request for comments (Proposed Standard) 2205, Internet Engineering Task Force, September 1997.

32. B. Davie and Y. Rekhter, *MPLS Technology and Applications*, pp. 4-15, Morgan Kaufmann Publishers, 2000.

Appendix A

Introduction to *General Algebraic Modeling System* (GAMS)

GAMS is a high-level modeling system for mathematical programming problems and is capable of solving large-scale linear and non-linear programming models [28]. The GAMS has incorporated ideas drawn from relational database theory and mathematical programming and has attempted to merge theses ideas to suit the needs of strategic modelers. Relational database theory provides a structured framework for developing general data organization and transformation capabilities. Mathematical programming provides a way of describing a problem and a variety of methods for solving it.

The GAMS model representation is in a form that can be easily read by people and by computers. The GAMS program itself is the documentation of the model. The GAMS system is designed so that models can be solved on different types of computers with no change.

GAMS programs consist of one or more statements (sentences) that define data structures, initial values, data modifications, and symbolic relationships (equations). Symbols must be declared as to type before they are used, and must have values assigned before they can be referenced in assignment statements.

In GAMS system, there are two most common ways of organizing programs. The first style places the data first, followed by the model and then the solution statements. In this style of organization, the sets are placed first. Then the data are specified with parameters, scalar, and table statements. Next the model is defined with the variable, equation declaration, equation definition, and model statement. Finally the model is solved and the results are displayed. The second style emphasizes the model by placing it before the data. This is a particularly useful order when the model may be solved repeatedly with different data sets. There is a separation between declaration and definition.

There are five basic GAMS data types and each symbol or identifier must be declared to belong to one of the following groups: sets, variables, parameters, equations, acronyms and models.

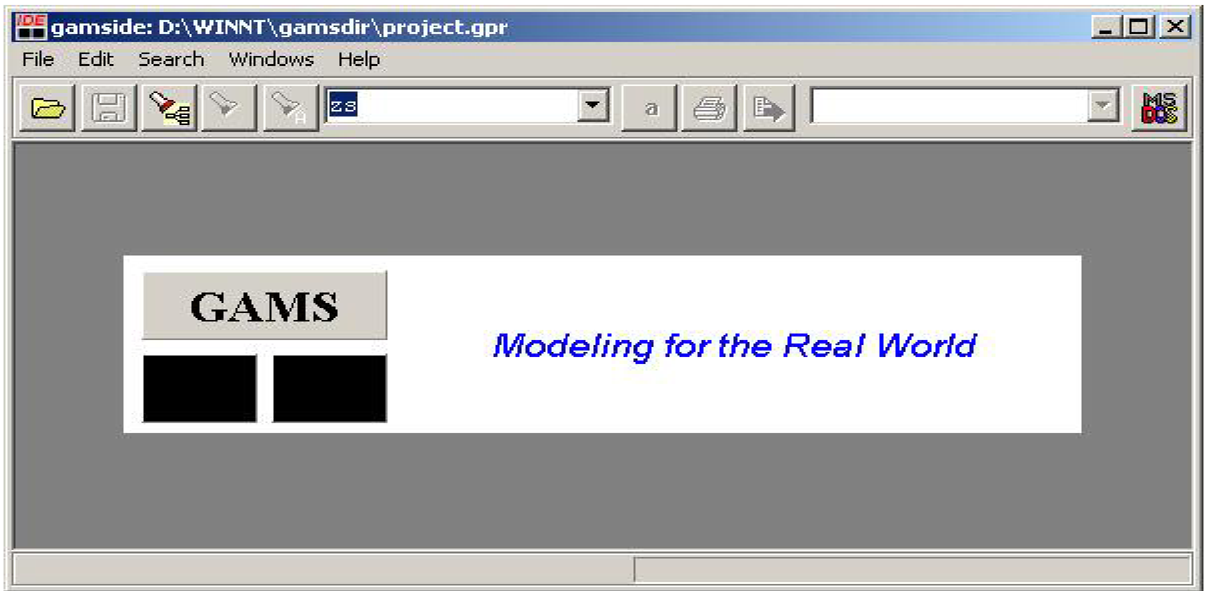
Sets are fundamental building blocks in any GAMS model. They allow the model to be succinctly stated and easily read. GAMS allows sets with up to 10 dimensions. In GAMS, a variable is the GAMS name for what are called ‘decision variables’ by industrial Operations Research practitioners. They are the entities whose values are generally unknown until after a model has been solved. A GAMS variable, like all other identifiers, must be declared before it is referenced.

In GAMS, equations are the names for the symbolic algebraic relationships that will be used to generate the constraints in the model. An equation must be declared before it can be used. In GAMS, all equations are collected into groups and labeled by using the model statement so that they can be solved. Once the model has been put together through the model statement, one can now attempt to solve it using the solve statement. GAMS itself does not solve the problem, but passes the problem definition to one of a number of separate solver programs. GAMS comes with a number of solvers that can be chosen by users.

A GAMS model will be first compiled, and then run by using one of a number of solvers.

Appendix B

In GAMS IDE system, a typical working window is as follows:



Users have to generate a new GAMS file with extension "gams," and then they can edit the program in the above IDE environment. After finishing edit, they can compile it. After revising all errors, they can run it and get the output file with extension "lst." In the following section, we give out the compiled codes for GFA and VPFA models.

GAMS codes for GFA model and VPFA model

The following codes for GFA model and VPFA model are based on the medium size prototype network.

GFA Model

```
Sets
        p  all paths in the network /1*20/
        w  all OD pairs in the network /1*10/
        i  all links in the network /1*17/
```

Parameters

```
d(p)   the unit flow cost of path p in cases
 /    1   7.8
      2   10.4
      3   8.6
      4   9.6
      5   14.2
      6   16.3
      7   7.2
      8   10.9
      9   7.3
      10  11.5
      11  9.8
      12  14.0
      13  6.5
      14  10.2
      15  6.6
      16  12.8
      17  8.1
      18  12.3
      19  10.4
      20  13.5   /

r(w)   external traffic demand for every OD pair in cases
 /    1  50
      2  50
      3  50
      4  50
      5  50
      6  50
      7  50
      8  50
      9  50
      10 50  /

c(i)   capacity for every link in cases
 /    1   100
      2   100
      3   100
      4   86
      5   100
      6   100
      7   100
      8   100
      9   100
```

```
          10  100
          11  100
          12  60
          13  100
          14  33
          15  100
          16  68
          17  100    /

      Table pr(w,p) indicator used to indicate if a path p connects OD pair w
          1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20
      1       1 1 0 0 0 0 0 0 0  0  0  0  0  0  0  0  0  0  0  0
      2       0 0 1 1 0 0 0 0 0  0  0  0  0  0  0  0  0  0  0  0
      3       0 0 0 0 1 1 0 0 0  0  0  0  0  0  0  0  0  0  0  0
      4       0 0 0 0 0 0 1 1 0  0  0  0  0  0  0  0  0  0  0  0
      5       0 0 0 0 0 0 0 0 1  1  0  0  0  0  0  0  0  0  0  0
      6       0 0 0 0 0 0 0 0 0  0  1  1  0  0  0  0  0  0  0  0
      7       0 0 0 0 0 0 0 0 0  0  0  0  1  1  0  0  0  0  0  0
      8       0 0 0 0 0 0 0 0 0  0  0  0  0  0  1  1  0  0  0  0
      9       0 0 0 0 0 0 0 0 0  0  0  0  0  0  0  0  1  1  0  0
      10      0 0 0 0 0 0 0 0 0  0  0  0  0  0  0  0  0  0  1  1  ;


      Table pc(i,p) indicator used to indicate if a path p passes link i
          1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20
      1       1 0 1 0 0 0 0 0 0  0  0  0  0  0  0  0  0  0  0  0
      2       0 1 1 0 0 0 0 0 0  0  0  0  0  0  0  0  1  0  0  0
      3       1 0 0 1 1 0 0 1 0  0  0  0  0  0  0  0  0  0  0  0
      4       0 1 0 1 0 0 0 0 0  1  0  0  0  0  0  0  0  1  0  0
      5       0 1 0 1 0 0 0 0 0  0  1  1  1  0  0  0  0  0  0  0
      6       0 0 0 0 0 0 0 0 1  0  0  0  0  0  0  0  1  0  0  0
      7       0 0 0 0 0 0 0 0 0  1  0  1  1  0  0  0  0  1  0  0
      8       0 0 0 0 0 0 0 0 1  1  0  1  0  1  0  0  0  0  0  0
      9       0 0 0 0 1 1 0 0 0  0  0  0  0  1  0  0  0  0  0  1
      10      0 0 0 0 1 0 1 0 0  0  0  0  0  0  0  0  0  0  0  0
      11      0 0 0 0 0 1 1 0 0  0  0  0  0  0  0  0  0  0  0  0
      12      0 0 0 0 0 1 0 1 0  0  0  0  0  0  0  0  0  0  0  0
      13      0 0 0 0 0 1 0 1 0  0  0  0  0  0  0  1  0  0  0  1
      14      0 0 0 0 0 0 0 0 0  0  0  0  0  0  1  0  0  0  0  1
      15      0 0 0 0 0 0 0 0 0  0  0  0  0  0  1  0  0  0  1  0
      16      0 0 0 0 0 0 0 0 0  0  0  1  0  0  0  0  0  0  1  0
      17      0 0 0 0 0 0 0 0 0  0  1  0  0  0  0  1  0  0  0  0  ;

Variables
      x(p) virtual path flow
      z    total delay in network ;
```

```
Positive variable x;


Equations
        delay           define objective function
        demand(w)       satisfy demand of OD pair w
        capacity(i)     observe capacity limit at link i;

delay ..  z =e= sum(p, d(p)*x(p));

demand(w) ..  sum(p, pr(w,p)*x(p)) =e= r(w);

capacity(i) ..  sum(p, pc(i,p)*x(p)) =l= c(i);


Model network /all/;

Solve network using lp minimizing z;


Display x.l;

parameter lf(i) link flow on link i;
         lf(i) = sum(p, pc(i,p)*x.l(p));
Display lf;

parameter ratio(i) link utilization ratio;
         ratio(i) = lf(i)/c(i);
Display ratio;

parameter mean mean value of link utilization;
         mean = sum(i, ratio(i))/17;
Display mean;

parameter std standard deviation of link utilization;
         std = sqrt(sum(i, (ratio(i)-mean)*(ratio(i)-mean))/17);
Display std;
Display z.l;
```

VPFA model

```
Sets
        p  all paths in the network /1*20/
```

```
      w  all OD pairs in the network /1*10/
      i  all links in the network /1*17/

Parameters

      d(p)    the unit flow cost of path p in cases
       /     1   7.8
             2   10.4
             3   8.6
             4   9.6
             5   14.2
             6   16.3
             7   7.2
             8   10.9
             9   7.3
             10  11.5
             11  9.8
             12  14.0
             13  6.5
             14  10.2
             15  6.6
             16  12.8
             17  8.1
             18  12.3
             19  10.4
             20  13.5   /

      r(w)    external traffic demand for every OD pair in cases
      /
             1        50
             2        50
             3        50
             4        50
             5        50
             6        50
             7        50
             8        50
             9        50
            10        50     /

      c(i)    capacity for every link in cases
       /     1   100
             2   100
             3   100
             4   86
             5   100
```

6 100
7 100
8 100
9 100
10 100
11 100
12 60
13 100
14 33
15 100
16 68
17 100 /

Table pr(w,p) indicator used to indicate if a path p connects OD pair w

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	
1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
2	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
3	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
4	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	
5	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	
6	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	
7	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	
8	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	
9	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	
10	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	;

Table pc(i,p) indicator used to indicate if a path p passes link i

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	
1	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
2	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	
3	1	0	0	1	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	
4	0	1	0	1	0	0	0	0	0	1	0	0	0	0	0	0	0	1	0	0	
5	0	1	0	1	0	0	0	0	0	0	1	1	1	0	0	0	0	0	0	0	
6	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	1	0	0	0	
7	0	0	0	0	0	0	0	0	0	1	0	1	1	0	0	0	0	1	0	0	
8	0	0	0	0	0	0	0	0	1	1	0	1	0	1	0	0	0	0	0	0	
9	0	0	0	0	1	1	0	0	0	0	0	0	0	1	0	0	0	0	0	1	
10	0	0	0	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	
11	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	
12	0	0	0	0	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	
13	0	0	0	0	0	1	0	1	0	0	0	0	0	0	0	1	0	0	0	1	
14	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	1	
15	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	1	0	
16	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	1	0	
17	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	1	0	0	0	0	;

```
Variables
        x(p) virtual path flow
        g(p) virtual path capacity
        z    total delay in network ;

Positive variable x, g;

        g.lo(p) = 0.000001;

Equations
        delay           define objective function
        demand(w)       satisfy demand of OD pair w
        capacity(i)     observe capacity limit at link i
        pathlimit(p)        observe capacity limit at path p;

delay ..  z =e= sum(p, d(p)*x(p)*x(p)/g(p));

demand(w) ..  sum(p, pr(w,p)*x(p)) =e= r(w);

capacity(i) ..  sum(p, pc(i,p)*g(p)) =l= c(i);

pathlimit(p) ..  g(p)-x(p) =g= 0;

Model network /all/;

Solve network using nlp minimizing z;

Display x.l, g.l ;

parameter lf(i) link flow on link i;
         lf(i) = sum(p, pc(i,p)*x.l(p));
Display lf;

parameter ratio(i) link utilization ratio;
         ratio(i) = lf(i)/c(i);
Display ratio;

parameter mean mean value of link utilization;
         mean = sum(i, ratio(i))/17;
Display mean;

parameter std standard deviation of link utilization;
         std = sqrt(sum(i, (ratio(i)-mean)*(ratio(i)-mean))/17);
Display std;
```

WST based on GFA model

Sets

p all paths in the network /1*20/
w all OD pairs in the network /1*10/
i all links in the network /1*17/
s all scenarios of stochastically changing external traffic demands /1*64/

Parameters

d(p)   the unit flow cost of path p in cases

```
 /   1   7.8
     2   10.4
     3   8.6
     4   9.6
     5   14.2
     6   16.3
     7   7.2
     8   10.9
     9   7.3
     10  11.5
     11  9.8
     12  14.0
     13  6.5
     14  10.2
     15  6.6
     16  12.8
     17  8.1
     18  12.3
     19  10.4
     20  13.5   /
```

c(i)   capacity for every link in cases

```
 /   1   100
     2   100
     3   100
     4   86
     5   100
     6   100
     7   100
     8   100
     9   100
     10  100
     11  100
     12  60
```

13 100
14 33
15 100
16 68
17 100 /

r(w) external traffic demand
/ 1 58
2 41
3 50
4 56
5 56
6 58
7 39
8 47
9 56
10 41 /

pdf(s) probability of senario s
/ 1 0.004622675
2 0.000139573
3 0.041617548
4 0.000652771
5 0.004955364
6 0.084759865
7 0.000817386
8 0.000438225
9 0.123810464
10 0.010980947
11 0.000364963
12 0.018549567
13 0.001249798
14 0.034868418
15 0.001109583
16 5.66434E-05
17 0.007530221
18 0.014886199
19 0.005208016
20 4.93087E-06
21 0.026712204
22 0.022168123
23 0.002475925
24 0.01099229
25 0.02130397
26 4.58682E-05
27 0.000640145

28	3.57473E-05	
29	0.005792914	
30	0.001015379	
31	0.002429978	
32	9.21978E-08	
33	0.00551032	
34	0.004566904	
35	0.000871254	
36	0.041346124	
37	0.003990249	
38	5.37731E-05	
39	0.006080366	
40	0.028165881	
41	0.003998571	
42	8.24191E-05	
43	0.013535015	
44	0.001432691	
45	0.056446994	
46	0.000736308	
47	0.001925202	
48	0.007364147	
49	0.022861184	
50	0.004699201	
51	1.93355E-05	
52	0.000559267	
53	0.000100127	
54	0.01895947	
55	0.000601143	
56	0.001994485	
57	0.008125434	
58	0.000485961	
59	0.002833602	
60	0.001853163	
61	0.00255103	
62	0.00310126	
63	0.01335999	
64	0.013087917	/

Table t(w,s) external traffic demand for every OD pair in all scenarios

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1	52	41	46	38	47	52	69	66	53	60	63	63	62	55	46	49	47	44	49	63
2	53	43	48	32	51	49	57	50	51	49	62	48	55	44	61	32	59	50	53	42
3	46	46	52	52	48	44	54	64	46	45	56	50	58	49	42	46	46	48	42	41
4	56	42	50	55	54	45	44	53	53	50	54	45	47	58	53	42	53	61	55	42
5	45	59	40	43	44	45	44	42	49	55	38	53	48	52	50	50	47	44	48	28
6	52	30	53	49	63	55	55	57	43	52	47	46	50	58	50	37	59	54	55	58

```
7      63 48 48 48 54 46 46 50 50 49 44 49 54 50 60 39 56 46 46 54
8      59 45 52 54 59 50 56 49 51 46 45 57 50 53 64 61 42 55 58 44
9      44 45 58 52 49 53 48 58 50 43 48 52 59 52 57 55 58 52 50 53
10     43 46 49 45 59 51 56 55 53 61 60 53 36 51 46 51 48 57 36 58

+      21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40
1      57 56 41 48 56 54 49 44 50 46 59 42 53 45 56 44 47 33 46 46
2      51 54 49 42 42 51 47 45 50 59 55 36 43 42 51 56 58 55 55 55
3      54 51 45 53 48 50 55 54 60 57 54 61 54 45 35 47 46 52 45 51
4      45 49 46 40 51 28 55 48 59 45 55 53 50 58 51 43 54 57 58 43
5      48 46 49 48 47 50 48 40 47 49 41 63 53 59 42 53 61 36 48 58
6      48 46 45 51 61 49 53 39 58 49 47 78 48 52 49 53 47 50 52 48
7      57 57 54 55 44 37 57 75 57 41 38 57 51 48 61 53 39 64 57 43
8      50 57 51 42 52 48 34 50 47 56 40 53 49 46 47 51 55 58 45 54
9      41 47 68 41 50 50 64 51 58 62 51 43 67 55 51 47 46 50 62 48
10     54 58 55 50 51 38 55 42 53 38 48 62 56 61 61 56 43 53 56 45

+      41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60 61 62 63 64
1      55 43 58 52 54 44 44 56 44 60 46 43 58 50 35 56 53 48 43 54 39 48 47 51
2      47 70 43 50 49 46 43 53 45 52 46 46 53 57 60 45 49 49 57 42 54 54 47 49
3      40 42 53 58 47 59 62 47 54 49 52 50 40 45 52 63 57 62 47 43 57 53 53 50
4      62 59 51 41 41 53 47 59 55 60 55 53 47 56 44 45 51 56 45 52 49 50 52 42
5      40 50 52 55 49 47 46 49 49 51 51 48 61 45 42 56 55 60 42 52 57 55 49 46
6      50 50 37 49 55 46 52 50 42 47 52 54 63 54 53 50 45 59 53 53 41 58 39 57
7      52 53 48 42 47 37 42 59 52 54 63 55 39 53 47 37 47 62 53 57 49 43 48 45
8      54 51 50 46 48 62 51 59 55 51 54 39 39 50 45 51 36 56 64 52 52 42 42 59
9      46 35 52 61 55 44 52 53 48 59 67 67 43 45 62 43 49 42 58 41 60 46 56 42
10     55 54 47 63 52 60 36 57 58 40 30 61 57 40 46 48 43 44 47 35 57 65 58 55  ;


 Table pr(w,p) indicator used to indicate if a path p connects OD pair w
        1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20
 1      1 1 0 0 0 0 0 0 0  0  0  0  0  0  0  0  0  0  0  0
 2      0 0 1 1 0 0 0 0 0  0  0  0  0  0  0  0  0  0  0  0
 3      0 0 0 0 1 1 0 0 0  0  0  0  0  0  0  0  0  0  0  0
 4      0 0 0 0 0 0 1 1 0  0  0  0  0  0  0  0  0  0  0  0
 5      0 0 0 0 0 0 0 0 1  1  0  0  0  0  0  0  0  0  0  0
 6      0 0 0 0 0 0 0 0 0  0  1  1  0  0  0  0  0  0  0  0
 7      0 0 0 0 0 0 0 0 0  0  0  0  1  1  0  0  0  0  0  0
 8      0 0 0 0 0 0 0 0 0  0  0  0  0  0  1  1  0  0  0  0
 9      0 0 0 0 0 0 0 0 0  0  0  0  0  0  0  0  1  1  0  0
 10     0 0 0 0 0 0 0 0 0  0  0  0  0  0  0  0  0  0  1  1  ;

 Table pc(i,p) indicator used to indicate if a path p passes link i
        1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20
 1      1 0 1 0 0 0 0 0 0  0  0  0  0  0  0  0  0  0  0  0
```

```
         2      0 1 1 0 0 0 0 0 0  0  0  0  0  0  0  0  1  0  0  0
         3      1 0 0 1 1 0 0 1 0  0  0  0  0  0  0  0  0  0  0  0
         4      0 1 0 1 0 0 0 0 0  1  0  0  0  0  0  0  0  1  0  0
         5      0 1 0 1 0 0 0 0 0  0  1  1  1  0  0  0  0  0  0  0
         6      0 0 0 0 0 0 0 0 1  0  0  0  0  0  0  0  1  0  0  0
         7      0 0 0 0 0 0 0 0 0  1  0  1  1  0  0  0  0  1  0  0
         8      0 0 0 0 0 0 0 0 1  1  0  1  0  1  0  0  0  0  0  0
         9      0 0 0 0 1 1 0 0 0  0  0  0  0  1  0  0  0  0  0  1
         10      0 0 0 0 1 0 1 0 0  0  0  0  0  0  0  0  0  0  0  0
         11      0 0 0 0 0 1 1 0 0  0  0  0  0  0  0  0  0  0  0  0
         12      0 0 0 0 0 1 0 1 0  0  0  0  0  0  0  0  0  0  0  0
         13      0 0 0 0 0 1 0 1 0  0  0  0  0  0  0  1  0  0  0  1
         14      0 0 0 0 0 0 0 0 0  0  0  0  0  0  1  0  0  0  0  1
         15      0 0 0 0 0 0 0 0 0  0  0  0  0  0  1  0  0  0  1  0
         16      0 0 0 0 0 0 0 0 0  0  0  1  0  0  0  0  0  0  1  0
         17      0 0 0 0 0 0 0 0 0  0  1  0  0  0  0  1  0  0  0  0  ;

Variables
         x(p)   virtual path flow
         z      total delay in network  ;

Positive variable x;


Equations
         delay           define objective function
         demand(w)       satisfy demand of OD pair w
         capacity(i)     observe capacity limit at link i ;

delay ..  z =e= sum(p, d(p)*x(p));

demand(w) ..  sum(p, pr(w,p)*x(p)) =e= r(w);

capacity(i) ..  sum(p, pc(i,p)*x(p)) =l= c(i);


Model network /all/;

Parameter lf(i) link flow on link i;
parameter ratio(i) link utilization ratio;
parameter mean mean value of link utilization;
Parameters  a(w,s), b(p,s), y(p,s);
parameter st standard deviation;
parameter std(s) standard deviation of link utilization for senario s;
parameter f(s) used for swapping data for std;
parameter zt(s) objective value for senario s;
```

```
parameter k(s) used for swapping data for zt;



scalar j;

for (j=0 to 63 by 1,
    a(w,s) = t(w,s+j);
    r(w) = a(w,'1');
    Solve network using lp minimizing z;
    b(p,'1') = x.l(p);
    y(p,s+j) = b(p,s);
    lf(i) = sum(p, pc(i,p)*x.l(p));
    ratio(i) = lf(i)/c(i);
    mean = sum(i, ratio(i))/17;
    st = sqrt(sum(i, (ratio(i)-mean)*(ratio(i)-mean))/17);
    f('1') = st;
    std(s+j) = f(s);
    k('1') = z.l;
    zt(s+j) = k(s);
   );

Display x.l;
Display y;
Display std;
Display zt;

scalar   aw "weighting factor" /0.1/
       bw "weighting factor" /10/;



Variable
       xs(p)  stochastic programming result for virtual path flow on path p
       zs     tracking model value;

Positive variable xs;

Equations
       delays   coordination model
       demands(w)     satisfy demand of OD pair w
       capacitys(i)   observe capacity limit at link i for tracking model;

delays .. zs =e= sum(s, pdf(s)*(aw*abs(sum(p, d(p)*xs(p))-zt(s))+
bw*abs(min(0,smin(w,(sum(p, pr(w,p)*xs(p))- t(w,s)))))));

capacitys(i) ..  sum(p, pc(i,p)*xs(p)) =l= c(i);
```

```
Model scenarios /delays, capacitys/;

Solve scenarios using dnlp minimizing zs;

Display xs.l;
```