MPsLS: A Novel Switching Scheme Improving Per-flow QoS for Time-sensitive Applications

by

Jun YANG

submitted to
Japan Advanced Institute of Science and Technology
in partial fulfillment of the requirements
for the degree of
Doctor of Philosophy

Supervisor: Professor Yasushi HIBINO

School of Information Science
Japan Advanced Institute of Science and Technology

September 2006

Abstract

In this paper, the author proposes a new data forwarding scheme, named Multi-Protocol synchronous Label Switching (MPsLS), suits for the integrated data network. MPsLS is a novel data forwarding scheme for the core areas of the integrated data network. It provides an interface between layer-3, the network layer, and layer-2, the data link layer. It successfully integrates layer-2 synchronous frame transfer, similar to DTM, with asynchronous packet transfer based on label switching.

In the MPsLS network, data are transferred in cyclic mode with a constant period ($\tau = 125\mu s$) called a frame. A frame consists of a number of slots with the same size (512 bits). A few slots in the header part of a frame are used as control slots, which dedicate to communicate with neighboring nodes and exchange the network control information for routing and setting up connections.

The remaining slots in a frame are data slots, which carry the application data and temporary control message. A slot carries only a segment of a layer-3 packet. Two-bits tag is introduced on the header of each data slot to distinguish the types of the slots.

MPsLS has two types of channels with different characteristics on data transfer, named appointed channels and filler channels.

Appointed channels of MPsLS provides connection-oriented service. Time-sensitive traffic is transferred synchronously on specified appointed channel. Since the number and the positions of the appointed channel slots are fixed, time-sensitive application flow can be identified by checking the positions in the frame and referring the corresponding channel table.

While non time-sensitive traffic is forwarded pseudo-asynchronously by filler channels which shares slots in the frame. The number and the positions of the filler channel slots are variable, thus, an additional 32 bits switching label like MPLS following tag bits is introduced to a filler slot in order to identify non time-sensitive flows.

The appointed channel has two possible connection modes, exact synchronous connections and less strict synchronous connections. Although less strict synchronous connections lead to a little longer delay than exact synchronous connections, but the delay values of the time-sensitive applications can be controlled within a given offset range of the slots positions according to QoS requirement.

Since the number and positions of each appointed channel slot are reserved to per-flow, and the traffic belonging to distinct time-sensitive applications is isolated each other. Besides, the appointed channels are dynamically set up at the beginning of a session and are kept all the time during the session. Therefore, it naturally provides per-flow QoS guarantees for the time-sensitive application.

On the other hand, the introduction of tag bits and additional switching label to filler slots makes the filler channel possible to use not only free slots which are not reserved for time-sensitive applications, but also to use temporary idle slots which are reserved for time-sensitive applications. When the slots are temporarily idle, the slots in that positions can transfer non time-sensitive traffic, though the appointed channels for time-sensitive traffic have priority to non time-sensitive traffic. Therefore, non time-sensitive traffic is transferred at best-effort model on filler channels. The use of the idle slots has not influence upon the performance of the time-sensitive applications, while the utilization of network resources is improved remarkably.

Therefore, MPsLS successfully combines the advantages of synchronous transfer mode and flexibility of label switching technology. It provides capabilities of guaranteeing QoS for time-sensitive applications and maintaining high network resource utilization.

keywords: QoS, switching, synchronization, frame, slot, appointed channel, filler channel, exact synchronization channel, less strict synchronization channel

Acknowledgments

First of all, the author would like to express his heartful thanks to his principal advisor Professor Yasushi Hibino of Japan Advenced Institute of Science and Techology for his constant encouragement and kindly guidance during this work. Professor Hibino is very warmhearted, he not only taught the author how to become a better researcher, but also helped the author to learn the culture of Japan and people's habits. The author hopes and looks forward to continued collaboration with Professor Hibino in the future research.

The author would hope to express his great thanks to the examination committee members, Professor Yasushi Hibino, Professor Hiroshi Esaki, Professor Yoichi Shinoda, Associate Professor Yasuo Tan and Associate Professor Kiyofumi Tanaka for their friutful comments.

The author would like to thank his advisor of minor research Professor Mineo Kaneko of Japan Advanced Institute of Science and Technology for his helpful discussions and suggestions on minor research.

The author also wishes to express his thanks to Assistant Sugawara, and the other members of Hibino Lab. of Japan Advanced Institute of Science and Technology for their helps during the research.

The author is grateful to all the friends and colleagues for spending his time with them as a graduate student at Japan Advanced Institute of Science and Technology.

The author acknowledges the support of "Fostering Talent in Emergent Research Fields" in Special Coordination Funds for Promoting Science and Technology by Ministry of Education, Culture, Sports, Science and Technology.

The author devotes his earnest gratitude to his parents for their love.

The author feels deeply indebted to his dear wife Jingzhi Guan and son Zhengyu Yang for their love and understanding through the graduate years. They gave their unconditional support even if that meant to sacrifice the time they spent together. Without their support and encouragement, it would have been impossible to finish this research.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Aims of Research

Over the past decades, many communication networks dedicated to various application media have been developed with the advance of computer and digital information technologies. For example, telephone networks (ISDN) [1, 2, 3] for voice communication., cable TV networks (Broad Band-ISDN) for video communication [4, 5, 6, 7, 8, 9], and data networks (packet networks such as IP networks) [10, 11] for text communication.

The fact of various communication networks existing separately results in the huge wastage of the network resources. To reduce the cost of building and maintaining dedicated networks for the variety of media, the best method is integrating voice, video, and plain data communications into a public network. However, the challenge of integrated multimedia network is how to guarantee quality of service for real-time multimedia applications (such as voice, video-conferencing) that require guarantees in strict delay and delay jitter, while to maintain good scalability and high resource utilization to be adapts for rapidly growing application traffic. Therefore, the aim of this research is to develop a new switching technology which adapts to future integrated multimedia service networks.

1.2 Quality of Service

Quality of Service (QoS) can be defined in a variety of ways and includes a diverse set of service requirements [12, 13, 14, 15, 16, 17, 18], such as performance, availability, reliability, security, etc. All these service requirements are important aspects of an network offering QoS levels.

In this dissertation, the author will take a more performance-centric view of network QoS and focus primarily on the issues in providing performance guarantees.

1.2.1 Main Parameters Evaluating Performance

Typical performance metrics defining network QoS are bandwidth, delay, jitter, and loss ratio.

- **End-to-End Delay**

  Delay refers to the latency of information data on a transmission medium and/or in transmission devices from a transmission node to a reception node.

End-to-end delay means the total latency from a source node to a destination node along the transmission path.

The delays typically consist of two components: a fixed delay and a variable delay.

The fixed delay arises on the transmission links when the signal propagates and in the originating and terminating endpoints where the signal is processed. Fixed delays depend on the current state of technology.

Variable delays result primarily from the queuing and processing of the packets. They depend on the traffic intensity, the characteristics of the network route, including the number of hops (nodes), the type and speed of links.

- **Jitter**

  Jitter means the distortion of information arrival timing influenced by traffic flow state as it is propagated through the network. Information arrival timing varies from its original reference timing depends on flow state, and information does not arrive consecutively or on a timely basis at its destination.

  In packet-switched networks, jitter is a distortion of the inter-packet arrival times compared to the inter-packet times of the original transmission, and also, referred to as delay variance.

  This distortion give serious damage to multimedia traffic.

  A typical method for suppressing jitter is to add a fixed buffer delay that exceeds the maximum variable delay time. This can be successfully introduced into local area networks and corporate intranets with minimal jitter. However, wide area networks cannot introduce the additional fixed delay because the increased delay that were the sum of maximum variable delay along the path would be too great for communication purposes. Thus, the decision of a fixed delay value or buffer size is a trade-off between acceptance of large delays and tolerance of jitter arising.

- **Bandwidth**

  Bandwidth is the maximum transmission capacity of a connection, which is theoretical value. Available bandwidth or throughput is the practical value because some negative factors shrinking the theoretical bandwidth can cause transmission delay that deteriorates service quality.

- **Packet Loss Ratio**

  Packet loss rate is defined as a ratio of the number of lost packets to the total number of transmitted packets. There are several reasons for packet loss: 1) transmission impairments, 2) excess delay and 3) congestion. Transmission impairments are likely to occur and cause packet loss if there are physical problems in the transmission equipment comprising a network. Excess delay can lead to packet loss if the delay exceeds the "time-to-live" (TTL) value of the packet. If the TTL value is exceed, then the packet is discarded.

  Another frequent source of packet loss is congestion. Packets travel in an IP network with a hop-by-hop methodology. At each hop, a router must read the packet header and get the destination information (address) of the packet, and then determine the next hop router.

For the "bursty" nature of data communications, the router queue occasionally grows extremely large when the traffic flocks to an output port. When the router queue is filled up, it falls into the congestion state. A common decongestion algorithm for the routers is to dispose packets in the queue. Thus all packets in the queue are discarded to ease congestion state according to the service discipline for packet delivery.

- **Multicasting Capability**

  Multicasting is an efficient way delivering information to a select group of destinations on the Internet or an internal network, much like the conference call among a select group of people. Instead of sending same messages to the individual recipients, a single message is sent to a multicast group, which includes all the recipient who want to participate in the multicast session.

1.2.2 How to Guarantee QoS

The main parameters from the user view affecting QoS are delay, jitter, loss ratio and so on. Those parameters are, in principle, determined by more basic network characteristics, such as communication network architecture, speed of processing units, bandwidth of the link, and so on. The latter two depend on current technology level.

In this paper, the author mainly analyzes the performances of his proposed communication network architecture.

In theory, in order to guarantee excellent QoS, two capabilities are required, i.e., path reservation and resource reservation (including for both processing and transmitting).

Path reservation is an indispensable capability, because the transmission through distinct routes have different propagation time and arrival time interval through distinct path varies.

Although fluctuation of arrival time interval can be hidden with the buffer at the destination node, growth of delay is inevitable. Therefore, path reservation capability enables that all data are transmitted along a designated path during a session. Otherwise, jitter or delay will be generated.

Resource reservation is another necessary capability. This requires that resource can be allocated dynamically at any moment if necessary.

Presently, since all units comprising a network are used in time-division multiplexing manner, when the data arrive the processing or transmitting unit and they can not obtain momentary enough resource, they must wait till the processing or transmitting units to be idle. This also leads to the growth of delay or jitter.

So-called guaranteeing QoS for time-sensitive applications does not pursue removing completely variable delay or jitter. It is enough that the delay and jitter satisfy the QoS requirements.

As developing information technologies, the applications requires more strict QoS emerges. In order to satisfy those special requirement, it is necessary to develop new network technologies that have capability of providing good performances, as short variable delay and small jitter as possible.

Strictly to say, although it is very difficult to completely remove variable delay or jitter of time-division multiplexing. We can accomplish a purpose at core area in the network. It is worthwhile, because degradation of QoS performance mainly arises at core area.

A feasible scheme such as synchronous channels is for a specific mechanism assigning application traffic flow to specified channels and allocating resource for that flow.

1.3 Main Contribution of This Dissertation

In this paper, a novel switching scheme, MPsLS, is proposed, which can be used to forward data in integrated data network. MPsLS combines the advantages of synchronous transfer mode and flexibility of label switching technology. It enhances MPLS concept, uses fixed slots to forward synchronously time-sensitive application flows and uses floating position slots to forward the non real-time application flows. Therefore, MPsLS has capabilities of both guaranteeing the QoS required for time-sensitive application flows and maintaining high network resource utilization.

1.4 Organization

The rest of this dissertation is organized as follows:

Chapter 2 describes background information, introduces two prominent network models, datagram model and virtual circuit model including ATM. The former is the foundation of today's Internet, but only provides best-effort service. Besides, three QoS improvement schemes, Integrated Services, Differentiated Services and Multiple Access Protocol over SONET/SDH, have been described. The latter is considered as great revolution of telecommunication industry research into packet network with huge efforts to exploit the advantages of multiplexing circuits. ATM uses fixed-size cell, which enables high QoS by means of fine-grain scheduling policies, and reduces the contention between traffic flows. However, due to the essential features of asynchronous transmission, the resource contention and the congestion of the packets or cells are inevitable because of the stochastically arriving process of the packets or cells, which leads to degradation of the application QoS.

Furthermore, a synchronous transfer mode, DTM, is described, which provides complete QoS guarantee for specific applications. It has, however, drawbacks of lower channel utilization and longer switching delay.

Chapter 3 proposes a novel approach called "Multi-Protocol synchronous Label Switching (MPsLS)" to improve QoS performance of per-flow. Firstly the framework and architecture of MPsLS are explained, then its implementation details at ingress and core nodes are introduced.

Chapter 4 analyzes the access performance and capability of network providing appointed channels to carry time-sensitive applications.

Analytical and numerical solutions are obtained to demonstrate the main performance of MPsLS network, based on the assumption model and the parameters. The results show that less strict synchronous channels have high calling success probability, at the same time, over 80% bandwidth can be used to support appointed channel service.

Chapter 5 discusses the delay performance when the application passing MPsLS cloud, and derives of the analytical solutions of local delay at ingress and core nodes. Furthermore, this chapter gives derivation of an approximate expression of estimating edge-to-edge delay. The results illustrate that MPsLS has smaller delay than DTM.

Chapter 6 discusses the effect of idle appointed slot positions being used or not to carrying non time-sensitive traffic temporarily on average delay of non time-sensitive traffic. The results show that MPsLS scheme when idle appointed slot positions being used to carrying non time-sensitive traffic can improve the throughput and, at the same time, reduce the average delay of non time-sensitive application traffic.

Chapter 7 compares the maximum delay and jitter in MPsLS with the values in other QoS improvement schemes based on asynchronous transfer models, DiffServ and CBR of ATM; also compares the calling loss probability of non time-sensitive traffic in MPsLS with that in synchronous model, DTM.

Finally, Chapter 8 summarizes the conclusions of the dissertation, and simply discusses other MPsLS performances, such as transmission loss, large bandwidth and multicast capability.

Chapter 2

QoS Requirement and Current QoS Improving Schemes

Currently, there are many kinds of the communication networks for the variety of services, such as telephone networks for voice, cable TV networks for video, and data networks for text.

The typical telephone networks utilizes circuit switching technology, where a dedicated channel (circuit) between two end points is set up when they need to communicate each other. The channel remains busy for the entire duration of the call, whether the channel is actually used or not. The network using circuit switching has capability of transmitting analog or digital information without distortion and inflection, and can provide best service quality. However the utilization (and throughput) of the links are low because of feature of the dedicated channels.

The data networks, such as Internet, carry information in packets, where original information (a message) is separated into segments and is contained into packets. A packet is switched and transmitted through the network. The packets are reassembled at the destination to reconstruct the original information. In packet switching networks, the resources are shared by all applications, at switch nodes, the packets are processed at certain scheduling mechanisms, simple FCFS, Round Robin and so on, therefore, they have superior flexibility and high resource utilization.

Recently, since the advance of computer and digitization technologies, new multimedia develop rapidly, the wide-spreading of information presents a high requirement to current communication networks, namely, the networks must provide QoS guarantees to time-sensitive applications, and maintain high resource utilization to adapt to explosively increasing of the digital information.

In this chapter, we will introduce several typical models that can improve the network performances.

This chapter is organized as follows. Section 2.1 firstly introduces the types of main multimedia and QoS sensitivities. Section 2.2 describes the datagram model, and three QoS improving schemes, IntServ, DiffServ and MAPOS. Section 2.3 introduces virtual circuit modeles, MPLS and ATM model. Although the purpose of MPLS itself is not to improve QoS, but it can be achieved the better QoS when it is combined with other QoS schemes. ATM model can provide better QoS than datagram modeles. Section 2.4 explains a typical synchronous transfer model, DTM. At last, Section 2.5 summarizes this chapter.

2.1 Types of Multimedia Information and Their QoS Sensitivities

The term multimedia information implies that the information is consisted of more than one types of media, for instance, text, audio, images, video, and animation. Each single media has its basic characteristics. Text text media commonly has small size, but it requires very high accuracy. Audio/video media has larger data size, but it allows certain degree of error or loss of data. Animation media has not only very large size, but also requires strict time delay.

The multimedia consist of well-selected media according to different features of single medium.

The multimedia applications are classified into either time-sensitive or non time-sensitive. Time-sensitive multimedia require either hard or soft bounds on the end-to-end delay/jitter, while non time-sensitive media have not any strict constraints on delay, but may have rigid constraints on error.

Wide range of applications have different QoS sensitivities. The QoS values for some typical applications are summarized in Table 2.1.

Table 2.1: QoS Requirements of Typical Applications

Application Type	Bandwidth requirement	Sensitivities		
		Loss	Delay	Jitter
Voice	Very low	Medium	High	High
E-commerce	Low	High	High	Low
E-mail	Low	High	Low	Low
Telnet	Low	High	Medium	Low
Web-browsing	Medium	Medium	Medium	Low
File transfers	High	Medium	Low	Low
Video conferencing	High	Medium	High	High
IPTV	High	Medium	High	High

To satisfy the QoS requirements for different types of applications, the transmission networks need to provide the performances, such as high bandwidth, low loss probability, small delay and jitter, and high resource utilization.

2.2 Datagram Mode

The current Internet is the representative network, in which packet switched Internet Protocol (IP) [10, 11] is applied. IP network uses datagram model, and supports only one basic service: best-effort datagram, which does not provide any guarantee of reliabile delivery or strict timeliness. This behavior was acceptable when most of the data carried over the Internet were textual and non real-time in nature. Almost all efforts to make a reliable communication medium for text were concentrated on the Internet.

However, as the Internet evolves into a global commercial infrastructure, there are growing needs to provide more powerful services than best-effort such as QoS guaranteed services. They are necessary of transferring multimedia applications, such as IP telephony, video-conferencing, and remote diagnostics. Therefore, providing these services in the Internet has been one of the major challenges in the network research.

Now, several typical schemes have been deployed [16, 19], for example, IntServ, Diff-Serv and MAPOS.

2.2.1 Integrated Services Model

Integrated Services (IntServ) [20, 21, 22, 23, 24] is an Internet service model in order to support real-time applications over the Internet, which was developed by the Integrated Services working group in the Internet Engineering Task Force (IETF).

It is usually used with Resource Reservation Protocol(RSVP) [25, 26, 27, 28] together. RSVP can reserve the resource to per-flow at all network nodes along the path from source to destination. Therefore, it requires applications to notify their traffic characteristics beforehand and to inform reserved resources satisfing them to the intermediate network routers.

Accordingly, if the requested resources are available, the routers reserve them and send back a positive acknowledgment to the source, which can then start sending data. On the other hand, if sufficient resources are not available at a router in the path, the request is turned down and the request has to try again in some time.

Intserv includes three service classes: Guaranteed service, Controlled Load service and Best-effort service.

Guaranteed Service

The guaranteed service class provides maximum end-to-end delay bound, because in network, maximum queuing and routing delay of packet can be approximately computed by the flow state, and the flow state is determined by the the token bucket size and the data rate. Therefore, the maximal queuing delay can be controlled by setting the bucket parameters to each application flow.

Controlled Load Service

The controlled load service provides better performance than the traditional best-effort service. It is only assumed that the end-to-end behavior of the aggregate traffic flow can be determined under loaded conditions. The load is defined by the admission control algorithm according to the measurements of the current aggregate network load.

The network does not give any quantitative guarantees for each flow. Therefore, it only ensures that adequate bandwidth and packet processing resources to handle the flows are available, but it does not gaurantee specific target values for delay or loss.

Thus, controlled-load service merely provides a service that the flows are transmitted under lightly loaded conditions that makes the flows experience little or no congestion loss.

Best-Effort Service

The best-effort service class does not provide any admission control to the flows, thus, there are no QoS guarantees that the network provides.

2.2.2 Differentiated Services Model

Differentiated services (DiffServ) model [29, 30, 31] is proposed by diffserv working group in the IETF for the Internet, which handles traffic at aggregate levels rather than the IntServ approach of handling individual flows.

DiffServ puts complex processing in the edge routers, and core routers are simply forwarding and scheduling the already classified data.

In differentiated services domains, the routers are divided into boundary nodes and interior nodes, the former performs certain edge functions like admission control, packet classification and traffic conditioning. The admission control algorithm limits the number of flows that are admitted into the diffserv domain. For any admitted flow, the packets arrived the DS domain is classified and marked as belonging to one of the service classes, called "Behavior Aggregates (BA)". Each such behavior aggregate is assigned a distinct 8-bit codeword, called the DS code-point. The later only performs packet forwarding. When a packet with a particular DS code point arrives at this node, it is forwarded to the next hop, according to some pre-defined rule associated with the packet class. Such pre-defined rules are called Per-Hop Behaviors (PHB's). Aggregates are grouped into per-hop behaviours (PHBs), which are marked in the DiffServ code point (DSCP). The DSCP is located in the first six bits of the IP Type of Service field. The typical models are Expedited Forwarding and Assured Forwarding.

Expedited Forwarding PHB

Expedited Forwarding (EF) [32], also known as premium service, which provides the services of the departure rate of a traffic class being equal or exceed the configured rate. therefore, a bound on delay and jitter can be obtained.

To achieve this service, the departure rate of the aggregate on each outgoing link must be greater than or equal to the sum of maximum arrival rates on incoming links. The traffic forwarding process is based on the queueing methods, such as simple priority queues (PQ), weighted round robin queue scheduling (WRR), and class based queues (CBQ).

EF does not deal with individual user flows, but rather the aggregates of them. Therefore, no bounds are placed on individual flows.

Assured Forwarding PHB

Assured Forwarding (AF) [33, 34] is primarily for applications which require reliability better than best-effort service. AF guarantees minimum bandwidth and buffer space to each aggregated class. But the flow rate beyond the defined rate is allowable, however, this part of traffic is marked with high drop precedence, in the case of congestion, these packets are dropped at priority. To achieve this service, assured queues (AQ) are used and managed through Random Early Discard(RED).

2.2.3 Multiple Access Protocol over SONET/SDH (MAPOS)

Multiple Access Protocol over SONET/SDH (MAPOS) [101, 102, 103, 104, 105, 106, 107, 108, 109], proposed by NTT Laboratories, is a data link layer protocol of HDLC-like frame over SONET/SDH. MAPOS encapsulates a upper layer variable length packet such as an IP packet in a HDLC-like frame and sends it in SONET/SDH payloads. In an SONET/SDH frame, HDLC-like frames are separated by a HDLC flag sequence "01111110". IP addresses are mapped into HDLC addresses and nature of connection-less is turned over.

Therefore, weakness of asynchronous packet switching is inherited and it is unavoidable to develop queuing delay according to traffic condition, although high-speed switching is conducted by HDLC addresses, shorter format than IP address, and efficient data transmission is performed on SONET/SDH frames.

2.2.4 The Disadvantages of Improvement Schemes Based on Datagram Mode

Although the models above all have ability to improve QoS for time-sensitive applications in IP networks, they can not satisfy the requirement of the high quality multimedia applications due to their obvious disadvantages:

IntServ

IntServ uses RSVP to make per-flow reservations at routers along a network path. At early of IP network development, due to the limitation of router's speed, it suffers scalability problem, since the routers have to maintain per-flow state for every flow that is passing through the router, which can lead to huge overhead. Moreover RSVP is a soft-state protocol, which means that the router state has to be refreshed at regular intervals. This increases traffic overhead. But at present, with the development of router technologies, the problem caused by reserving states for per-flow has no existance, on the contrary, it has become the common measure of QoS guarantee for high quality multimedia applications.

Its main disadvantage is that the congestion of the packets and contention for the processing or transmitting resource are unavoidable due to the randomness of the packets arriving, which usually leads to increasing of delay and loss of the packets.

DiffServ

At early, DiffServ model is intended to address scalability and complexity. By not requiring per-flow state to be stored in the routers, and no complex signaling protocols.

Therefore, DiffServ is highly scalable and relatively less complex. In DiffServ only the edge routers have to maintain per-flow states, it is difficult to provide quantitative QoS to individual flows. It is primarily designed for and used by ISPs, and is not too useful (or even intended) for end users. Additionally, DiffServ does not offer any receiver control, and the congestion and contention for the resource caused by stochastic arriving process still exist.

Accordingly, the improving schemes based on datagram mode, although, have high resource utilization, they can not remove the jiiter, and the delay also is kept on a high level because of the stochastic arrival process of the packets and variable length packets.

2.3 Virtual Circuit Modes

The typical virtual circuit models include MPLS and ATM.

2.3.1 Multiprotocol Label Switching Model

Multiprotocol Label Switching (MPLS) [35, 36, 37, 38] is an Internet Engineering Task Force (IETF)-specified framework that provides for the designation, routing, forwarding and switching of traffic flows through the network. Its initial goal of label based switching was to bring the speed of Layer 2 switching to Layer 3. Label based switching methods allow routers to make forwarding decisions based on the contents of a simple label, rather than by performing a complex route lookup based on destination IP address.

Since the label switching can be easily done in hardware, it results in very high speeds. But with the advancement of high speed Layer 3 switching technology, such as Layer 3 switches (ASIC-based routers) have been able to perform route lookups at sufficient speeds to support most interface types, the speed of MPLS is no longer perceived as the main benefit. The current advantages are connection-oriented and extensively usage, due to the feature of connection-oriented, makes it possible of Traffic Engineering (TE) [39, 40], while it can use to different Layer 2 transports, such as Ethernet, ATM and Frame Relay, because the MPLS label is situated between the Layer 2 and Layer 3 headers.

Similar to Diffserv, an MPLS network is consisted of boundary nodes, called Label Edge Routers (LER), and interior nodes, called Label Switching Routers (LSR). When packets enter an MPLS domain, a label edge router (LER) assigns a short fixed-length label to them. Then, the packets are forwarded through a series of label switched routers (LSR), till the destination. The entire path is called as a label switched path (LSP). Therefore, the most important implementation is label assignment, which is realized by a label distribution protocol, the simple Label Distribution Protocol (LDP) uses existing IP routing information to discover the LSRs, then establishes paths through an MPLS network by distributing the appropriate labels.

To obtain improved transmission performance, two complex label distribution protocols are proposed.

CR-LDP

Constraint-based [41] Routing over Label Distribution Protocol CR-LDP was developed over Label Distribution Protocol (LDP), by adding traffic engineering capabilities. In

CR-LDP, the label request is sent from the ingress to egress LER through either traditional routing protocols or explicitly stating the path, when the request passing through intermediate LSRs, the resource reservation is made before forwarding the request, if the request arrives the egress LER, it means the path satisfying the resource requriment is found, a label mapping message is sent back towards the ingress LER containing a new LSP label and information regarding the reservation just made, and a new LSP label is created, and the forwarding table is updated for the new LSP. Therefore, CR-LDP is considered as resource reservation based on the sender's requiring.

RSVP-TE

RSVP-TE [42, 43] is an extension to RSVP that includes mechanisms for MPLS traffic engineering. Being different with CR-LDP, RSVP-TE is considered as resource reservation of the receiver oriented. When the path request message is sent to the egress LER from ingress LER, a requested reservation (RESV) message is formulated, then a new LSP label is attached to the RESV message and returned back through the reverse path. In intermediate LSRs, the label is renewed hop by hop, the forwarding table also is updated according to received new label, till the ingress LER. Then, a new LSP is created.

Since it runs over a raw IP transport, it has mechanisms present to account for message loss. Additionally, its soft state nature requires periodic refreshes to keep reservations from being removed. Similar to RSVP, reservations requests are receiver oriented. To make a new LSP, a PATH message is sent from the ingress to egress LER similar to that of CR-LDP. The PATH message is traversed through all intermediate nodes using existing routing protocols, explicitly specified paths, or partial paths. Upon reaching the egress LER, a RESV message is formulated containing a flow descriptor describing the requested reservation. After making its own reservation, a new LSP label is attached to the RESV message and returned back through the reverse path. Intermediate LSRs will attempt to make any reservations, update the forwarding table with the label received, and attach a new LSP label. The ingress LER will do the same, without the need to attach a new label.

Like DiffServ, MPLS uses the concept of Forward Equivalence Class (FEC) to provide differential treatment to different media types. In theory, it is possible to define a separate FEC to per-flow or for every media type, however, which can result in large overhead.

Combining MPLS with DiffServ and IntServ [44, 45, 46, 47], better QoS can be obtained than simple DiffServ or IntServ.

2.3.2 ATM

Asynchronous Transfer Mode (ATM) [48, 49, 50, 51, 52, 53, 54] is a set of international interface and signaling standards defined by the International Telecommunications Union-Telecommunications (ITU-T) Standards Sector standard for cell relay wherein information for multiple service types, such as voice, video, or data. ATM transfers information in small, fixed-size cells. Unlike other packet technologies, ATM is a connection-oriented, virtual circuit-based packet switching and multiplexing technology. Because the cells are all the same size, cell delay at ATM switches is more predictable and manageable.

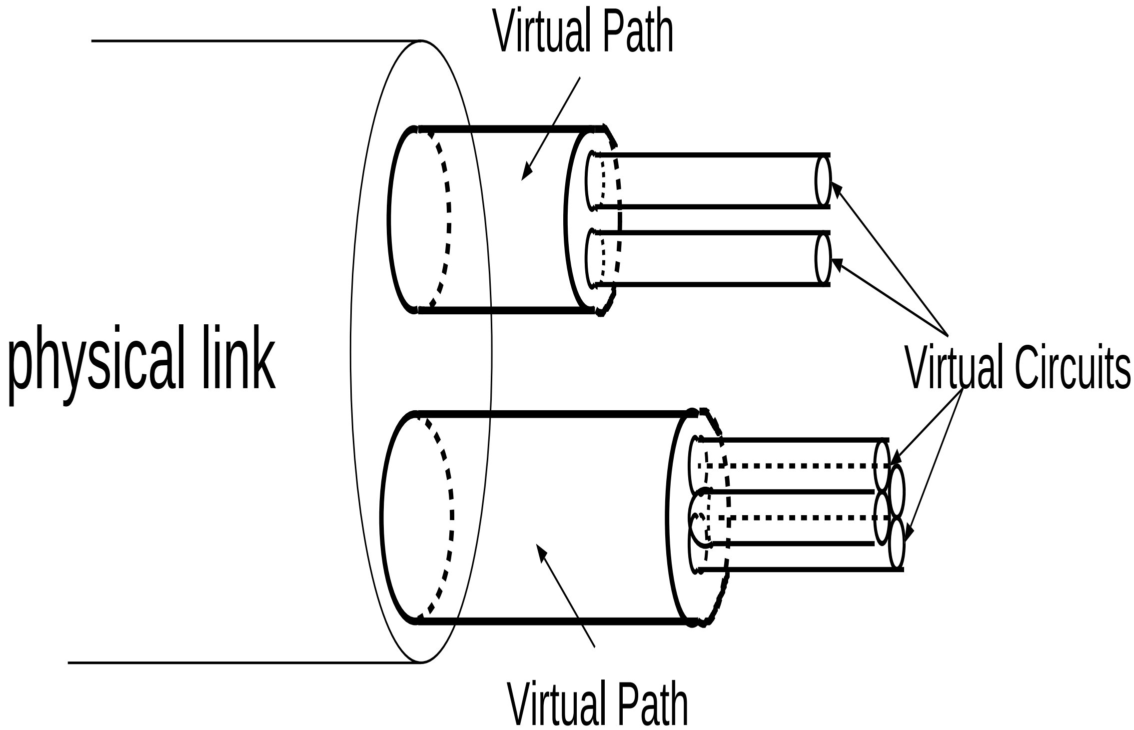

Figure 2.1: Virtual Circuits and Paths

In contrast to connectionless transmission protocols, ATM is connection-oriented. An ATM connection is a transmission path made up of a number of Virtual Paths, each of which contains a number of Virtual Channels. Therefore, a virtual channel connection is the basic unit, which carries a single stream of cells. A virtual path connection is a collection of virtual circuits, which is shown in Figure 2.1. A connection between ATM end services is defined by a virtual path identifier (VPI) and a virtual circuit identifier (VCI).

When a network device, such as a router, wants to send cells into an ATM network, it requests a virtual circuit. That request is passed from the initial ATM switch to other ATM switches in the network, with each switch determining whether it can handle the request. If all switches along the path can accommodate the request, the virtual circuit is established. If not, the request must be repeated. A Virtual Channel only exists for the duration of the call, bandwidth is only allocated for that call therefore it is very efficient.

An ATM network uses virtual paths internally for the purpose of bundling virtual circuits together between switches. Two ATM switches may have many different virtual circuit connections between them, belonging to different users. These can be bundled by the two ATM switches into a virtual path connection. This can serve the purpose of a virtual trunk between the two switches. This virtual trunk can then be handled as a single entity by, perhaps, multiple intermediate virtual path cross connects between the two virtual circuit switches.

Virtual circuits can be statically configured as permanent virtual circuits (PVCs) or dynamically controlled via signaling as switched virtual circuits (SVCs). They can also be point-to-point or point-to-multipoint, thus providing a rich set of service capabilities.

SVCs are the preferred mode of operation because they can be dynamically established, thus minimizing reconfiguration complexity.

Service Classes in ATM

Due to the diversity of traffic in network, such as low jitter tolerance traffic, real-time variable bit rate jitter intolerant traffic, non-real-time VBR low latency traffic, best-effort traffic and so on, in order to satisfy different service requirements, the ATM traffic management specifications define five categories of service [56, 57, 58]: constant bit rate (CBR), real-time variable bit rate (rt-VBR), non-real-time VBR (nrt-VBR), availble bit rate (ABR) and unspecified bit rate (UBR).

- **Constant Bit Rate(CBR)**

  CBR primarily to accommodate time-sensitive applications with stable bit rate. It specifies a fixed bit rate(usually peak cell rate) so that data is sent in a steady stream. Therefore, it is mainly used to emulate the time-division multiplexed circuit switching connections.

- **Variable Bit Rate(nrt-VBR and rt-VBR)**

  VBR is a popular choice for voice and videoconferencing data, it provides a specified throughput capacity but data is not sent evenly. At the same time, it adds the ability to statistically oversubscribe user traffic, traffic is managed by the network according to a guaranteed sustained cell rate (SCR) or a peak cell rate (PCR), and a maximum burst size. CBR cannot be exceeded, but the guaranteed bandwidth is a lower SBR. Further, VBR consists of real-time VBR and non-real-time VBR.

- **Available Bit Rate(ABR)**

  ABR provides a guaranteed minimum capacity and but allows data to be bursted at higher capacities when the network is free. It has a minimum cell rate option and several flow control mechanisms to provide some level of guaranteed bandwidth and data transmission integrity. A PCR has also been defined for ABR service, which can be set at the maximum information rate of the end user terminal interface. ABR is a very promising category, due to it can be maintained without wasting resources.

- **Unspecified Bit Rate**

  UBR provides a best-effort service, which makes no guarantees about traffic bandwidth or latency. This is used for applications, such as file transfer, that can tolerate delays.

  It is important to note, however, that just because data is not guaranteed for delivery by some ATM service classes doesn't mean it won't arrive. Rather, higher layer service applications like TCP will request a re-transmission, and eventually all the data or application should arrive.

Traffic Management in ATM

When an ATM circuit is set up, each switch is informed of the traffic class of the connection, then, to avoid the congestion and guarantee service quality, strict bandwidth provisioning is done by service level agreements (SLAs) at the virtual circuit (VC) and virtual path (VP) levels. The SLA consists of several elements: peak cell rate (PCR), sustained cell rate (SCR), maximum burst size (MBS), and permissible cell delay variation (CDVT).

The PCR specifies the absolute maximum bandwidth to be apportioned to a VC. The sum of PCR of all the VCs on a link could potentially exceed the link bandwidth. The traffic rate averaged over a long period of time is limited to the SCR, though it may spike over short time intervals. The traffic rate spike is limited to the PCR and its time duration is limited by the MBS. MBS is specified in number of bytes or number of cells. Determining the PCR, SCR and MBS enables a fine level of control over bandwidth apportioning in ATM. Based on the SLAs and its parameters, main control methods [59, 60, 61, 62] include:

- **Virtual Circuit Routing and Call Admission**

  Most ATM networks supporting SPVPs, SPVCs, and SVCs use the Private Network Node Interface or Private Network-to-Network Interface (PNNI) protocol. PNNI uses the same shortest path first algorithm used by OSPF and IS-IS to route IP packets to share topology information between switches and select a route through a network. PNNI also includes a very powerful summarization mechanism to allow construction of very large networks, as well as a call admission control (CAC) algorithm that determines whether sufficient bandwidth is available on a proposed route through a network to satisfy the service requirements of a VC or VP.

- **Traffic Policing**

  After the virtual paths and circuits established, a traffic contract is reached between the application and management module. Traffic contracts are usually maintained by the use of "Policing", which includes Shaping (a combination of queuing and marking of cells) and scheduling.

  Shaping Traffic shaping is usually done at the entry point to an ATM network and attempts to ensure that the cell flow will meet its traffic contract. The shaper maintains a pool of soft queues per traffic category to buffer cells for future transmission. This pool of per traffic category soft queues is replicated for each egress port. Among the soft queues in the pool, one queue in each category is marked as the current queue. The current queue pointer in every category is updated to point to the next soft queue in the pool after every max queue depth cell periods. To maintain network performance, traffic policing is used to police virtual circuits against their traffic contracts. If a circuit is exceeding its traffic contract, the network can either drop the cells or mark the Cell Loss Priority (CLP) bit (to identify a cell as discardable further down the line).

  Scheduling Besides shaping, scheduling is used to manage output queues at each ATM nodes, the scheduler module runs periodically, once every cell period, to service all the soft queues associated with an ATM egress port. This ensures that cell bursts are effectively spread in time, thus ensuring SCR conformance.

Usually, the egress module does a priority round-robin dequeue from the set of current queues for each physical egress port. The CBR queue has the highest priority followed by rtVBR, which has a priority higher than nrtVBR, which in turn has a higher priority than the UBR. The CBR queue is completely drained before visiting any lower priority queues. This behavior ensures that CBR and rtVBR cells are processed with the least possible latency.

2.3.3 The Disadvantages of Virtual Circuit Models

Although virtual circuit models can improve QoS further comparing to state-less datagram transfer mode, they still have following disadvantages:

MPLS

MPLS itself is not a QoS technology. Rather, it introduces a networking environment that is capable of transporting different traffic over a common infrastructure, while being able to enable QoS effectively. Thus, it is only a QoS-enabling technology, and provides a flexible solution for QoS deployment and management.

ATM CBR Service Model

Now, ATM has been accepted all over the world as the transfer mode used for B-ISDN (Broadband Integrated Services Digital Networks), which can handle any kind of information i.e. voice, data, image, text and video in an integrated manner. ATM provides a good bandwidth flexibility and can be used efficiently from desktop computers to local area and wide area networks.

However, at core nodes each switch forwards the CBR cells only by means of priority queue to schedule them, therefore, in the strict sense, the conflict of the cells with same priority can not be removed because of the stochastic arriving process of the cells from different links. Namely, the jitter still exists.

In addition, the overhead caused by cell header reaches to near 10%.

2.4 Synchronous Transfer Mode

In principle, asynchronous tranfer leads to the congestion and the contention for the resource due to the stochastically arriving process of the packets or cells on the case of time-devision multiplexing, therefore, it is inevitable of increasing delay or jitter, which can lower the service quality. It is well known that the synchronous scheme is a feasible approach avoiding the congestion and contention for the resource because synchronous transfer can specify arrived flow traffic, and avoid the randomness of the packets or cells arriving.

Dynamic synchronous Transfer Mode (DTM) [78, 79, 80, 81, 82, 83, 84, 85, 86], DTM is a typical synchronous transport network technology developed by Swedish company, Netinsight AB (`http: //www.netinsight.se`). It is a true broadband network architecture based on a synchronous fast circuit switching solution with dynamic reallocation of resources.

2.4.1 DTM Networks Structure

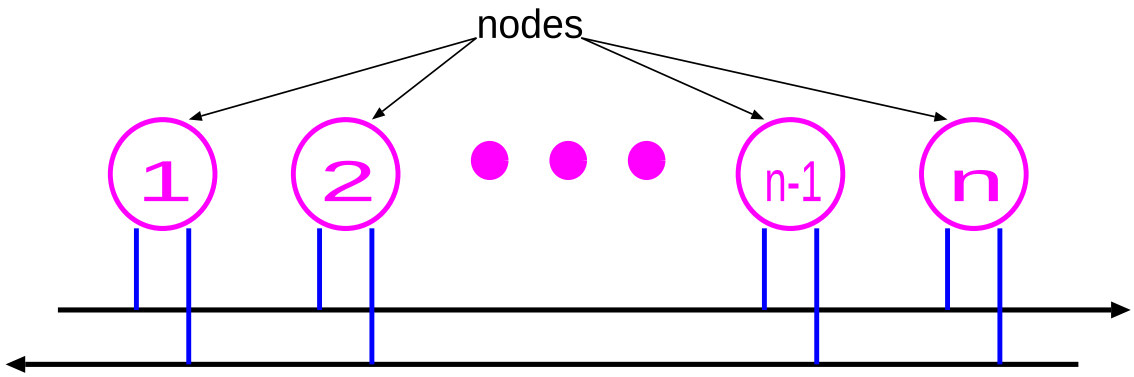

Figure 2.2: Dual Bus Topology

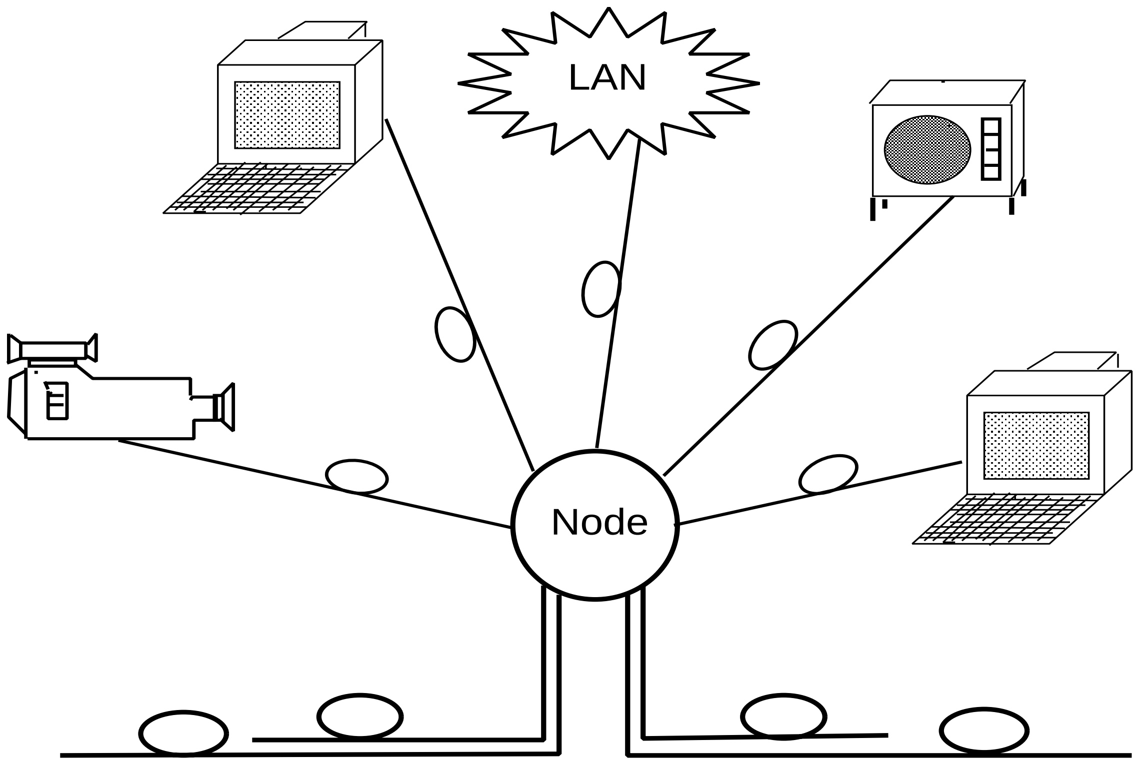

Figure 2.3: Hosts Connected to DTM Network through a Node

The basic topology structure of DTM network is the dual bus on which a set of nodes are connected. A dual bus is a pair of optical fibers where each fiber is used for data transmission in one direction. It is illustrated in Figure 2.2. For each node, there are several hosts can be connected to it. The hosts can be super computers, multimedia workstations, HDTV cameras and monitors, and local area networks, the connection is shown in Figure 2.3. Since the DTM protocols and medium access technique can, in principle, be used on any types of networks. Besides the dual bus, the topologies can be

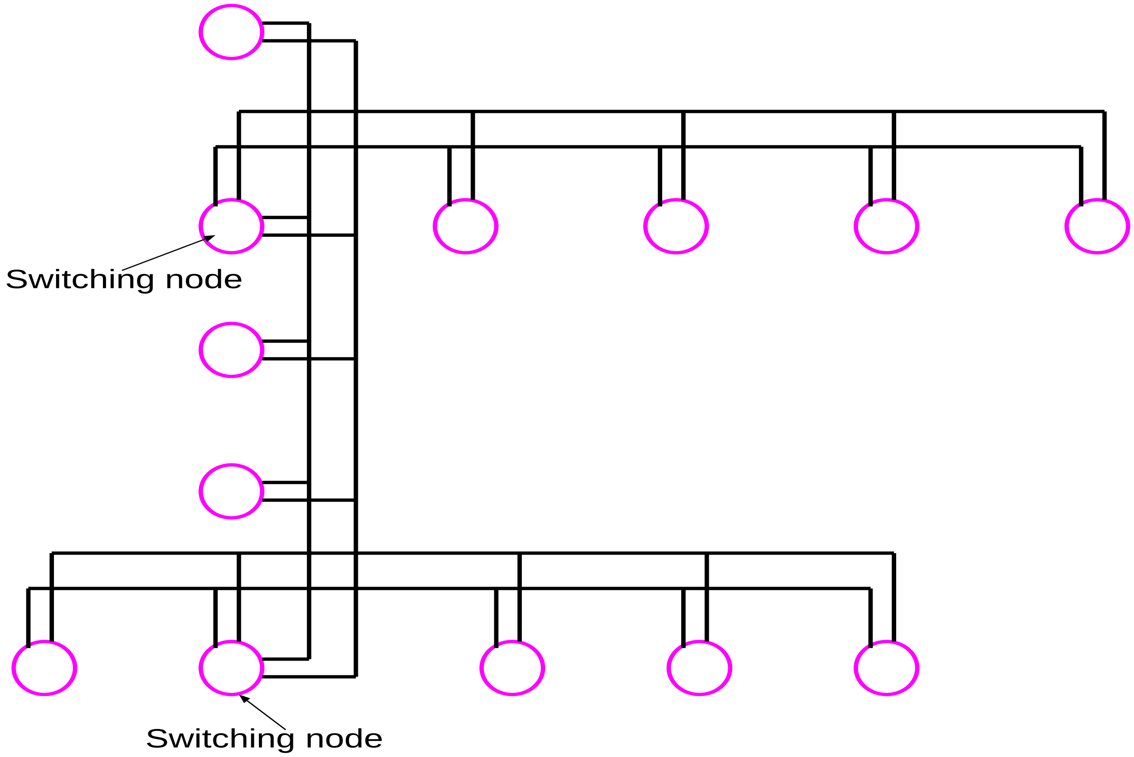

Figure 2.4: Mesh Topology

other types, such as a ring or mesh structure. The mesh structure is interconnected by several dual buses through a few switching nodes, as shown in Figure 2.4. In order to communicate for hosts on different dual buses, the channels on the fibers from the sender to the receiver have to be linked together into a route. Each switching node on the route is responsible for creating an outgoing channel on its bus. When a node has created the outgoing channel, the node starts switching by copying slots from incoming to outgoing channel.

2.4.2 DTM Channels

In DTM networks, the bitstream on a DTM fiber is organized into a flow of contiguous slots, where each slot contains 64 bits and the slots are grouped in 125 microsecond long cycles. The bit rate is determined by the number of slots in a cycle, so one slot corresponds to a bit rate of 512 kbps. By allocating a different numbers of slots, the transmission rate for a channel can be changed in steps of 512 kbps.

The structure of the frame is shown in Figure 2.5, the slots are divided into control and data slots. The control slots are statically allocated to a node and are used for signalling. Every node has at least one control slot allocated to it. The data slots are used for data transmission and each slot is always owned by a node. A node is only allowed to send the data in owns slots. The ownership of the slots is controlled by a distributed algorithm, where the nodes can request slots from other nodes. The algorithms for slot distribution between the nodes affect the network performance.

DTM nodes transfer the data flows, slot by slot, from the incoming to the outgoing

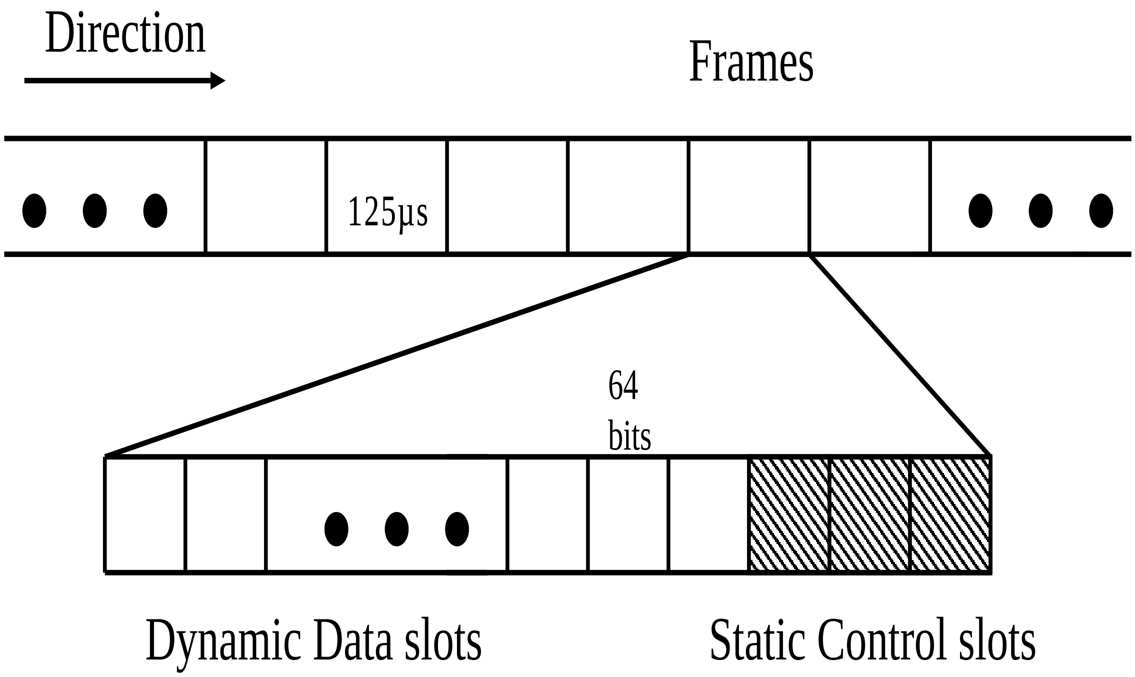

Figure 2.5: DTM Cycle Structure

side of the fiber. A node only reads the data in the slots that it owns in incoming fiber, and write the data into the slots in outgoing one. The DTM fiber access protocol [87, 88, 89, 90, 91, 92, 93] ensures that a node does not write data into slots that already carry data. A node does not have to read a slot to see if it is free to use. Likewise, a node does not have to read a slot to decide what data to write into it.

2.4.3 Merits of DTM

DTM is a natural circuit switched technology, based on the same principles with telephony networks. The main merits are:

- **Constant Delay**

  At each switching node, the delay is constant, and equals cycle period, thus, the delay from ingress to egress of DTM networks is predictable.

- **Arbitrary Channel Sizes**

  In DTM, the channels can have an arbitrary size in steps of 512 kbps up to full link bandwidth. Therefore, the channels can be provisioned to fit the service requriment at minimum limitation, which dramatically increases bandwidth utilization. Besides, the channel is exclusive between the sender and a destination for each flow, which separates the traffic from other traffic flows.

- **Non hierarchical switching**

The channels can be switched arbitrarily between network links, without considering network hierarchies. Links can be setup as desired to build large network structures, there are no hierarchy limitations of the size of the channels being switched.

- **Signaled end-to-end provisioning**

  Out-band signaling protocol handles the setup of the channels through the network. Only requiring identification of end points, then the channels automatically find their paths through the network according to the provision. .

- **Multicast**

  DTM channels can be point-to-multipoint, a resource can accommodate media services of high quality to a large number of receivers.

- **Routers Replaced Partly by Switches**

  DTM networks use switches instead of routers in a part of connection points, because when a channel has been established between sender and destination, the forwarding process of the traffic in each crossing is only implemented based on the switching of the slot positions, needs not read and analyse each packet's address, the less complex function can be realized by the switches replacing the routers.

2.4.4 The Disadvantages of DTM

Although, DTM has many excellent inherent characteristics, such as structure simplicity and avoiding computation-intensive policing, queuing, buffering and control mechanisms, specially, dynamic allocation of bandwidth improves channel utilization. However, for the time-sensitive applications, it is not enough to maintain high utilization, because if implementing dynamic allocation of bandwidth for each time-sensitive dialogue within session period, there is a danger of failing to establish channel for some dialogues, if the channel is maintained all the time during the lift time of whole session, thus, due to the idle slots allocated to active applications can not be shared by other flows, which leads to the remarkable decrease of utilization of the channels, on the other hand, the fact of the frames being buffer one frame period at switching nodes, although removing the jitter of delay during transmission, but increasing the delay.

In addition, in DTM the data link topologies can be bus, ring or mesh structures, however, the mesh structure is only based on bus, it does not support complex mesh structure such as star-shaped links. On the bus, the slots have low utilization since whole bus is dedicated to only one communication at moment even if more than one communication areas are not overlapped.

2.5 Summary

In the first part of this chapter, we overviewed the datagram model and virtual circuit models based on asynchronous transmission, in the former, the data traffic is transmitted by the packets with variable size. Each one includes an address tag in its header showing the destination. When a packet passes through the network, its address is read at the router nodes, then is forwarded to the next hop, one by one, till the destination. It supports best-effort service. in addition, three schemes to improve QoS for time-sensitive applications, Integrated Services (Intserv) model, Differentiated Services (Diffserv) model and MAPOS technology are introduced. In the latter, the data traffic is transmitted based on virtual circuit, which obviously shortens the routing time. Specially, ATM mode uses the cells with same size replacing variable size packets, the advantages comparing to datagram model are, on the one hand, the small and same size cells reduce the uncertainty of data traffic, it makes traffic management easier, on the other hand, the small cell header and determined virtual path increase the forwarding speed and decrease the delay.

However, all improvement schemes still maintain the inherent shortcomings of asynchronous communication network, due to using time-division multiplexing, when several traffic flows unite into a larger stream in a switching node, the packets or cells in each individual stream will mix and be forwarded. The confluence process of the streams inevitably leads to congestion of the packets, even the discard of the packets which overflow the buffer queue in a moment due to stochastic arriving process of the packets.

In the second part of this chapter, we introduced typical synchronous transfer mode, DTM. DTM performs synchronous circuit switching model, and is designed to fully utilize the almost unlimited capacity of optical fiber, it achieves better characteristics, including constant delay, almost zero delay variation, dynamic allocation of bandwidth, and full traffic isolation between channels. In nature, DTM has the capability of providing QoS guarantee for time-sensitive applications. However, there is the shortcoming of lower channels utilization, to obtain better quality of service further, in the next chapter, we will present our solution to provide better services in the integrated network.

Chapter 3

The Architecture of Multi-Protocol synchronous Label Switching Networks

In the past few years, several effective QoS improvement schemes for multimedia applications, such as video-conferencing, telephony, and video-on-demand, into public networks, have been proposed, such as IntServ, DiffServ, CBR model of ATM and DTM.

IntServ and DiffServ are schemes based on traditional IP technology, which adopt the priority queues to schedule traffic packets corresponding to an output link in order to assure QoS for time-sensitive flows, however, on the case of time-division multiplexing, the congestion of the packets and contention for network resource is inevitable because the arrival of the packets is undetermined process on asynchronization transmission.

CBR class service model of ATM improves the performance by using small and fixed size cells, at core switches, the priority queues are used to schedule cells, therefore, the contention among cells with same priority from different links also exists, which leads to little increasing of the delay, besides, the introdution of cell header generates about 10% ($5/53 = 9.4\%$) overhand.

DTM provides circuit switching emulation service, however, due to implementing frame-level synchronization at switching nodes, it leads to little longer delay, in addition, since the channel reserved for an application can not been shared by other applications even if the slots are idle during the life time of the session, which results in low utilization of the fibers.

Due to there still exist some shortcomings in present schemes, to obtain better QoS for time-sensitive multimedia applications, we propose a novel forwarding scheme, Multi-Protocol synchronous Label Switching(MPsLS), which is combination of both MPLS and synchronization transfer. MPsLS inherits the advantages of DTM, which reserves the channels all the time during the life of the connection, and it isolates the traffic on different channel, which means that the traffic in one channel do not disturb that in the other one, therefore, it naturely can guarantee QoS for time-sensitive applications. On the other hand, the idle slots can be used to temporarily carry filler slots due to the introduction of tag bits and MPLS header in filler slots, which remarkably improves the resource utilization of integrated network comparing to DTM.

This chapter is organized as follows. The basic architecture of MPsLS and transfer principle are introduced in Sections 3.1 and 3.2, then the control elements also are introduced in Section 3.3, after that the implementation processes at ingress and non ingress nodes are described in Sections 3.4 and 3.5, respectively. Sections 3.6 simply introduces the link between MPsLS network and non MPsLS networks, and convertion process of the transfer modes. At last, we summarize this chapter in Section 3.7.

3.1 Architecture of MPsLS

MPsLS is an extension of layer-2 (the data link layer) using asynchronous transfer model. It successfully integrates synchronous frame transfer similar to DTM and asynchronous transfer similar to ATM into new layer-2, therefore, MPsLS supports synchronization and asynchronization two kinds of different transfer mode simultaneously, synchronous transfer is realized by using position-fixed data piece in the frames, asynchronous transfer is realized by variable positions data piece with multi-protocol label, the main characteristics of MPsLS networks are as following:

3.1.1 Network Topology

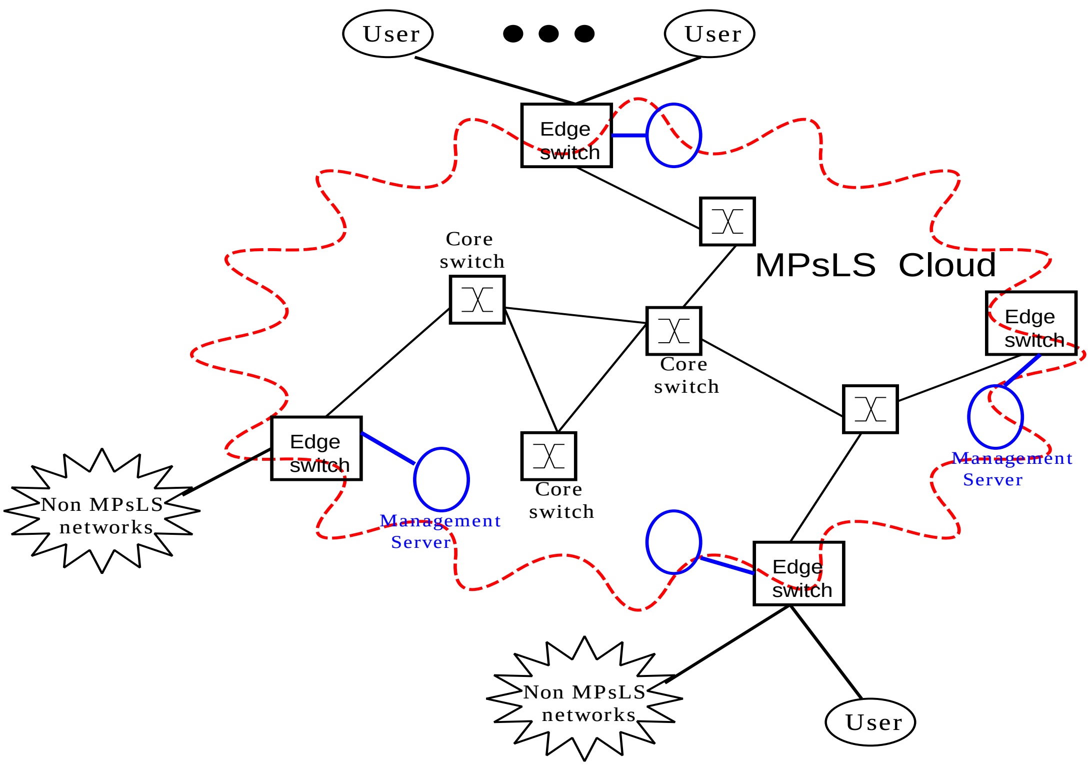

Figure 3.1: The MPsLS Network Topology

A MPsLS network is the network where the data traffic is transmitted at MPsLS switching model, MPsLS network is usually used to a backbone, the users are linked to MPsLS network directly or through LAN, the topology is shown in Figure 3.1, in MPsLS network the main equipments are switching nodes and link fibers. The switching nodes are divided into two types, core nodes and edge nodes (or switches). The former constitutes the backbone links, which connects only with other MPsLS nodes, the latter connects the core nodes with the nodes outside of the MPsLS cloud and provide the functions of entering to or exiting from the MPsLS network. The entering edge nodes are called ingress nodes and the exiting edge nodes are called egress nodes. Each port of MPsLS nodes has a unique HDLC address, which is identified by the address. Each link fiber connects two nodes at its two ends, it is supposed to be a base band or WDM optical fiber.

3.1.2 Layer Architecture of MPsLS Network

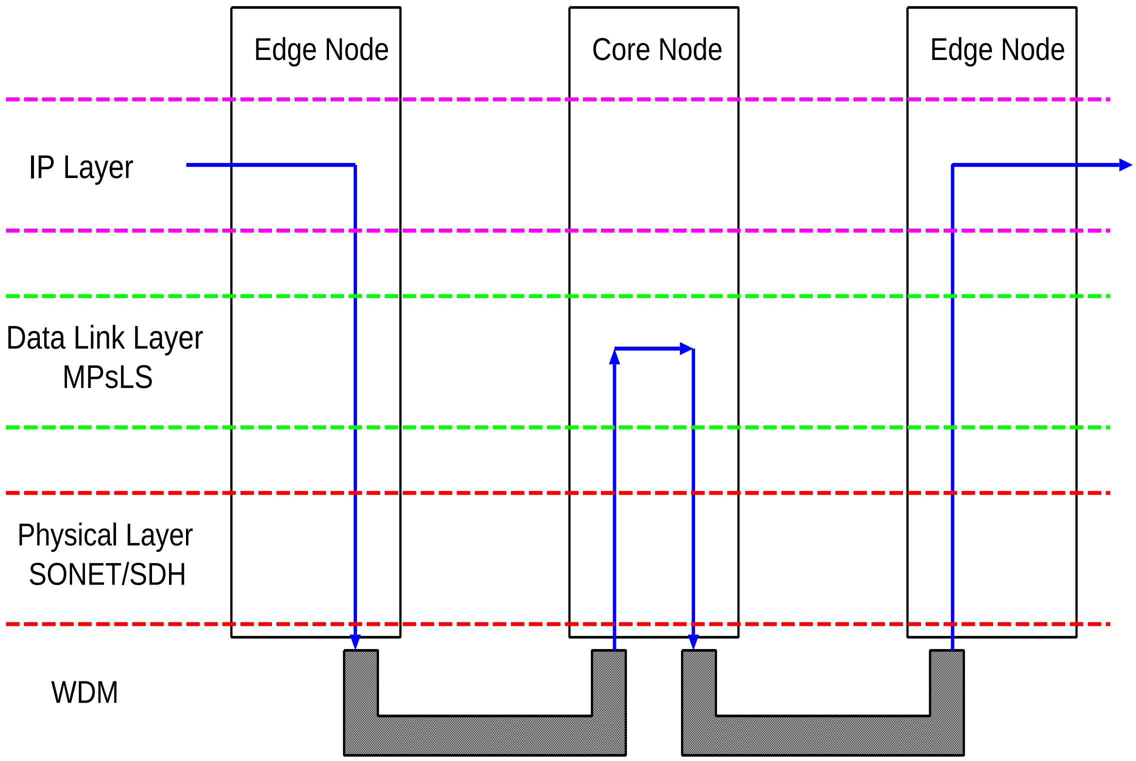

Figure 3.2: Layer Architecture of MPsLS Network

The layer architecture of MPsLS network is shown in Figure 3.2. In data link layer (or MPsLS layer), the data need to be segmented and assembled into slots, then further assembled into MPsLS frame, because the transmission of the data is done at frame pattern, but at switching nodes, the forwarding is implemented at slot level synchronous mode.

The layer under MPsLS layer is the physical layer, which is responsible for the ultimate transmission of data over network communications media. It operates with data in the form of bits that are sent from the Physical layer of the sending (source) device and received at the Physical layer of the destination device. In MPsLS networks, the physical layer bases on SONET/SDH structure, the network layer over MPsLS layer also uses IP protocol.

3.2 Transfer Principle of MPsLS

In MPsLS networks, the applications are divided into two types, non time-sensitive applications and time-sensitive applications. The former is insensitive to the delay and jitter, such as file transfers and E-mail service. The latter is sensitive to the delay and jitter, it is possible to lead to noticeable degradation of service quality with the increase of delay and jitter, such as live voice and video data transmission.

3.2.1 Service Classes

For different applications, MPsLS provides two classes of services, QoS guarantee class and best-effort services class. The former is used to transfer time-sensitive applications on appointed channels; the latter is used to transmit non time-sensitive applications on filler channels. Furthermore, although MPsLS only provides two classes services, which also can satisfy the QoS requirements for different applications, because the service level can be limberly configured by reserving different number of slots according to the QoS requirement: if reserving the channels at peak bit rate, the applications can receive best service quality; if equiping longer queues for the applications, the service quality with larger delay but smaller loss ratio can be obtained; on the contrary, equiping shorter queues, networks provide smaller delay but higher loss ratio service quality.

In addition, to guarantee the quality of service, and avoid the impact of traffic from different time-sensitive applications during transmission, MPsLS uses isolated channels to transmit time-sensitive applications traffic, while uses shared filler channel to transmit all non time-sensitive traffic.

3.2.2 Structure of Frames

In the MPsLS network, data are transferred in cyclic mode with a constant period called a frame. Each frame is divided into a number of slots with the same size. And each slot carries only a segment of a layer-3 packet. The structure of the frame is shown in Figure 3.3. Each frame is classified into two fields, a fixed control slots field and a data slots field. The fixed control slots field is located in the header part of the frame, the number of the slots in each frame is decided by the number of the neighbouring nodes, those slots are used dedicately to communicate with neighbouring nodes and to exchange the network control information for routing and setting up connections. The data slots field is in the remaining part, and which carry the application data and temporary control message.

The slots carrying time-sensitive data are called appointed slots, because which only use some appointed positions within the channel region; those carrying non time-sensitive data are called filler slots, which can utilize any free slots and idle positions reserved for appointed channels.

Let the period of frames be $125\mu s$, and the length of each slot be 512 bits, then, the size of the frame is determined completely by the link rate of a fiber or a lambda link. For usual optical fibers or lambda paths of WDM (Wavelength Division Multiplexing) with the bandwidths of 5Gbps to 20 Gbps, the number of the slots within each frame is 1200 to 4800.

The bandwidth of the appointed channels including only one slot in a frame is,

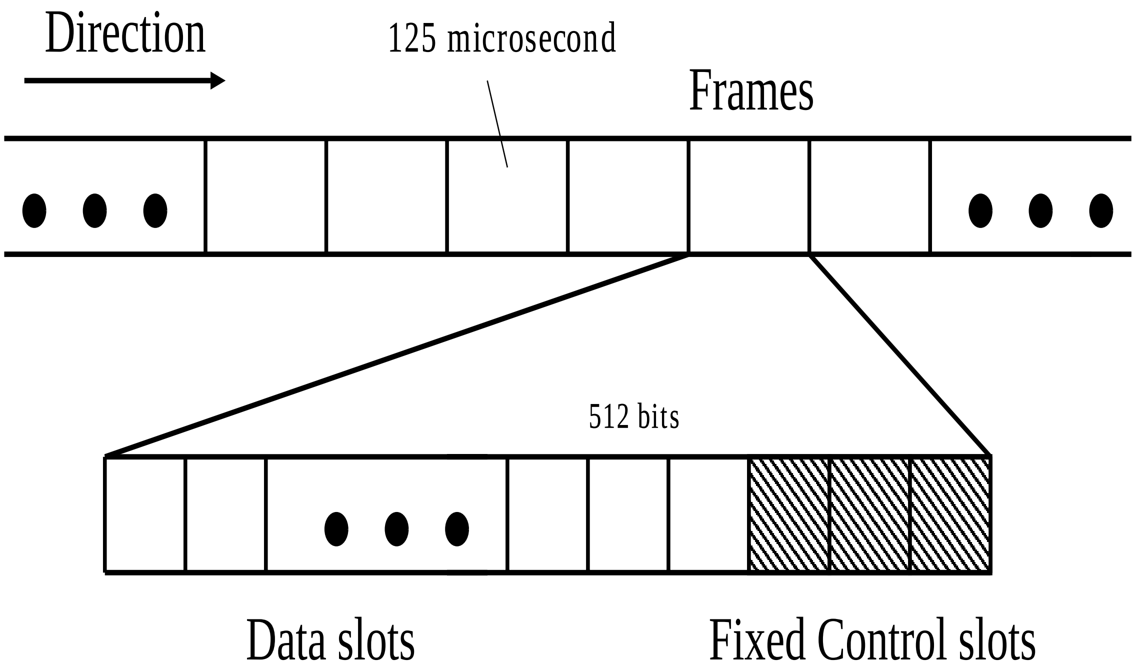

Figure 3.3: The Structure of the Frame

$$b = \frac{L}{\tau} = \frac{510bits}{125\mu s} = 4.08Mbps \tag{3.1}$$

where L is the length of the data in a slot carrying effective application data, which equals $512 - 2$(length of tag bits)= 510 bits. τ is the frame period.

3.2.3 Formats of Slots and Identification

In MPsLS network the data are classified into three types except fixed control message, dynamic control message for time-sensitive application, time-sensitive application data and non time-sensitive data.

The types of data carried by slots are distinguished by the tag bits located in the header, the formats of the slots are shown in Figure 3.4. Because the number and positions of the appointed slots carrying time-sensitive traffic are determined, the slots belonging to different application flows can be completely identified by referring to the channel table and checking only the corresponding tag bits, the rest parts after 2-bits tag bits are payload; but for non time-sensitive traffic, the positions of the filler slots are not fixed, it needs an additional label header like MPLS to identify different flows, therefore, the next 32 bits following the tag bits provide a switching label like MPLS, after that, carrying the payload, and the slots not carrying any effective data are called free slots.

However, the introduction of tag bits and a label header can cause the overhead, the overhead ratios for the appointed slots and the filler slots are $2/512 = 0.4\%$, and $(32+2)/512 = 6.6\%$, respectively.

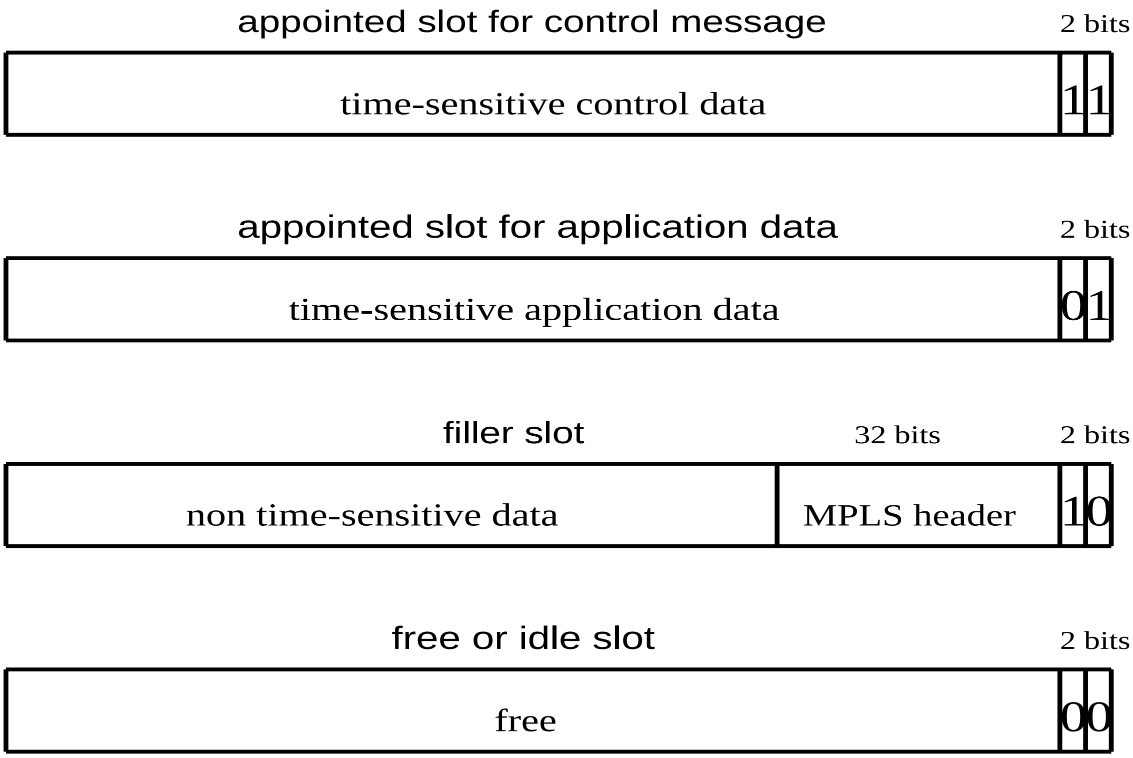

Figure 3.4: The Formats of Different Slots

3.2.4 Slot Switching

During forwarding process, the positions of the slots carrying time-sensitive traffic are fixed in the frames on each linking section, which are determined when the connection is set up, and are reserved during life time of the connection, thus, the data are transmitted synchronously in different frames; while the slot positions carrying non time-sensitive traffic are variable with the temporary state of the traffic, therefore, the transmission of the data is asynchronous. Since those reserved slot positions are used at priority by the time-sensitive traffic, the performances of time-sensitive applications are predicable, the service quality can be guaranteed according to their own parameter requirements, while non time-sensitive applications receive only best-effort service.

The forwarding process is illustrated in Figure 3.5. In incomming frames, if the slots carry time-sensitive traffic, the data are forwarded to corresponding reserved positions in outgoing frames, otherwise, the non time-sensitive traffic data are firstly sent to an queue according to FIFO mechanism, then filled in any free or idle slot position in the outgoing frame.

3.2.5 Channel Models

MPsLS channels are divided into two types: appointed channels and filler channels. The former is used to transfer time-sensitive applications, its connection path is determined, bandwidth and the positions of the slots in the frame are also fixed; the latter is used to transfer non time-sensitive applications, only its connection path is determined, although average bandwidth is also determined, but the instantaneous value is variable according

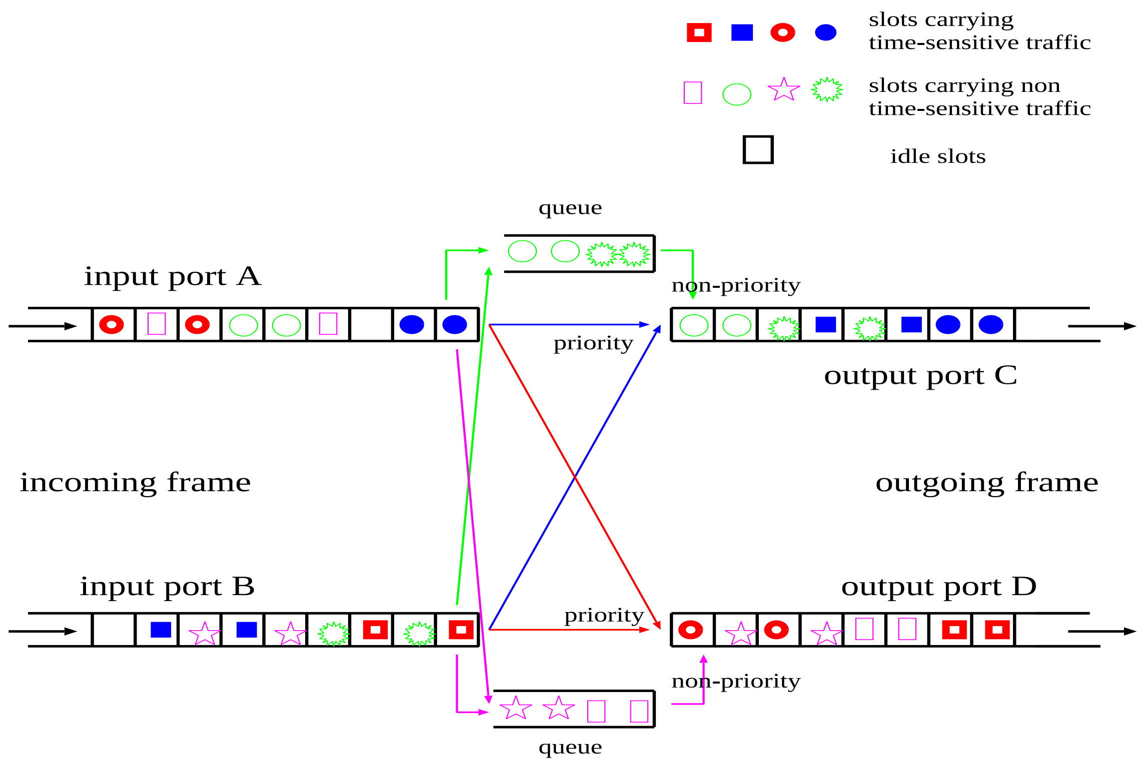

Figure 3.5: The Principle of MPsLS

to traffic state in the network.

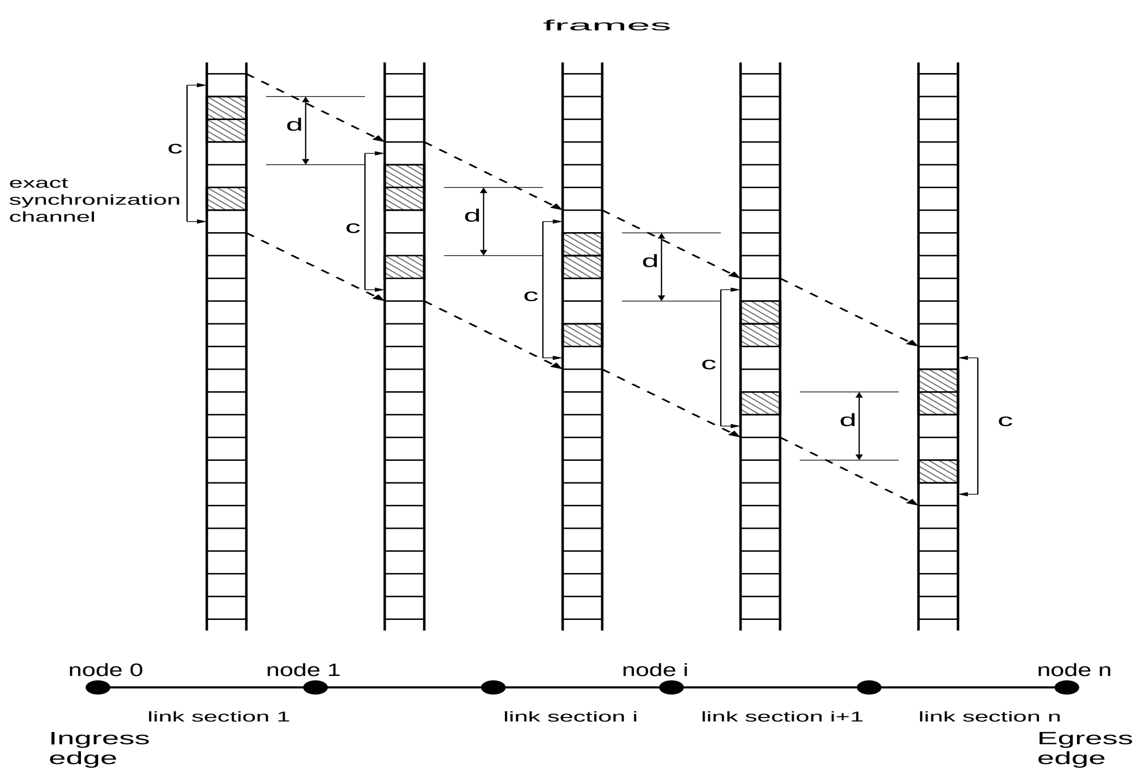

Figure 3.6: Exact Synchronous Appointed Channels

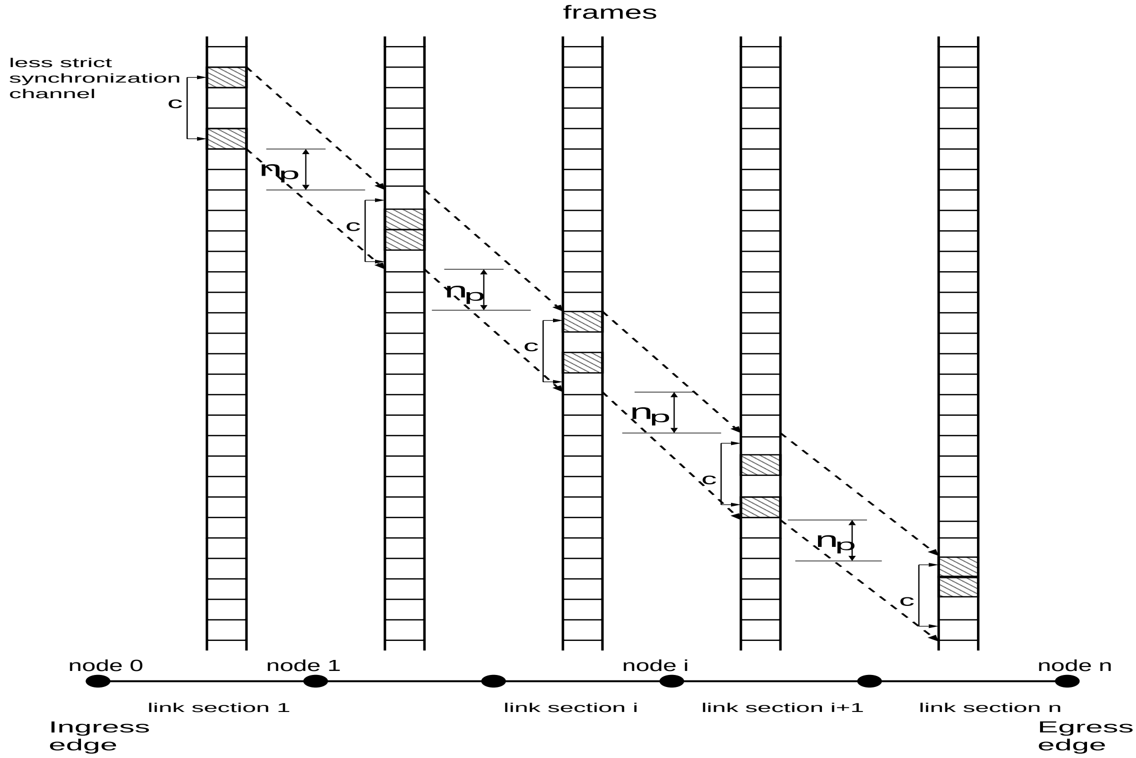

Figure 3.7: Less Strict Synchronous Appointed Channels

Furthermore, the appointed channels are divided into two sub-types, exact synchronization and less strict synchronization channels.

The exact synchronization channel is shown in Figure 3.6, where all slots consist of the channel belong to a strip areas, the area at each different link section along the path has same offset range c, the distribution of the slot positions in each offset range is fixed, and d indicates the lag of the corresponding slot positions between two neighbouring link sections, and the values equal the minimum latency time, n_p, which includes pipeline processing, buffer and reorder process.

The less strict synchronization channel is shown in Figure 3.7, where all slots consisting of the channel also belong to a strip areas, the area at each different link section along the path has same offset range c, the difference with the exact synchronization channel is that the lags between the header and tailer of the offset ranges in any two neighbouring link sections is constant, the value is n_p. The distribution of the slot positions in each offset range is variable, but within an offset range c, used slot positions are fixed all the time during a session.

Which type of channels should be set up, it is determined according to the QoS requirements, when the slots are forwarded passing theough core nodes, those on the exact synchronization channel have less delay than that on less strict synchronization channels, however, in general, the exact synchronization channels have smaller calling success probability than less strict synchronization channels, besides, the values also depend on the the hops between two edge nodes, required channel bandwidth and current utilization.

From the description on exact synchronous and less strict synchronous channels, it is clear that the strip ranges for both types of channels are no-overlapped even if the widths of offset ranges are same.

3.2.6 Idle Slots in Appointed Channels Utilized by Filler Slots

In MPsLS networks all time-sensitive traffic is transferred on appointed channels, for each appointed channel the number and positions of slots are reserved for per-flow during the life time of the whole session, even if momentary gaps or silence. While all non time-sensitive traffic is forwarded by asynchronous filler slots with multi-protocol label, for non time-sensitive applications the number and positions of slots used are not reserved, since tag bits and the MPLS label are introduced in filler slots, by which the appointed slots can be easily differed from the filler slots. Therefore, the filler slots can use not only free slots which are not reserved for time-sensitive applications, but also temporary idle slots which are reserved for time-sensitive applications, but on these channels the time-sensitive traffic is priority to non time-sensitive traffic, therefore, the use of the idle slots does not impact the performance of the time-sensitive applications, however, the utilization is improved remarkably, while the non time-sensitive applications receive best-effort service.

Figure 3.8 indicates the case of the idle slots reserved for appointed channel being used to carry filler slots temporarily. In Figure 3.8, an appointed channel includes three slots, in some frame periods, if three appointed slots arrived, all three positions are used to carry appointed slots, if only one or two appointed slots arrived, the same number of positions are used to carry appointed slots, left ones are temporarily used to carry filler slots.

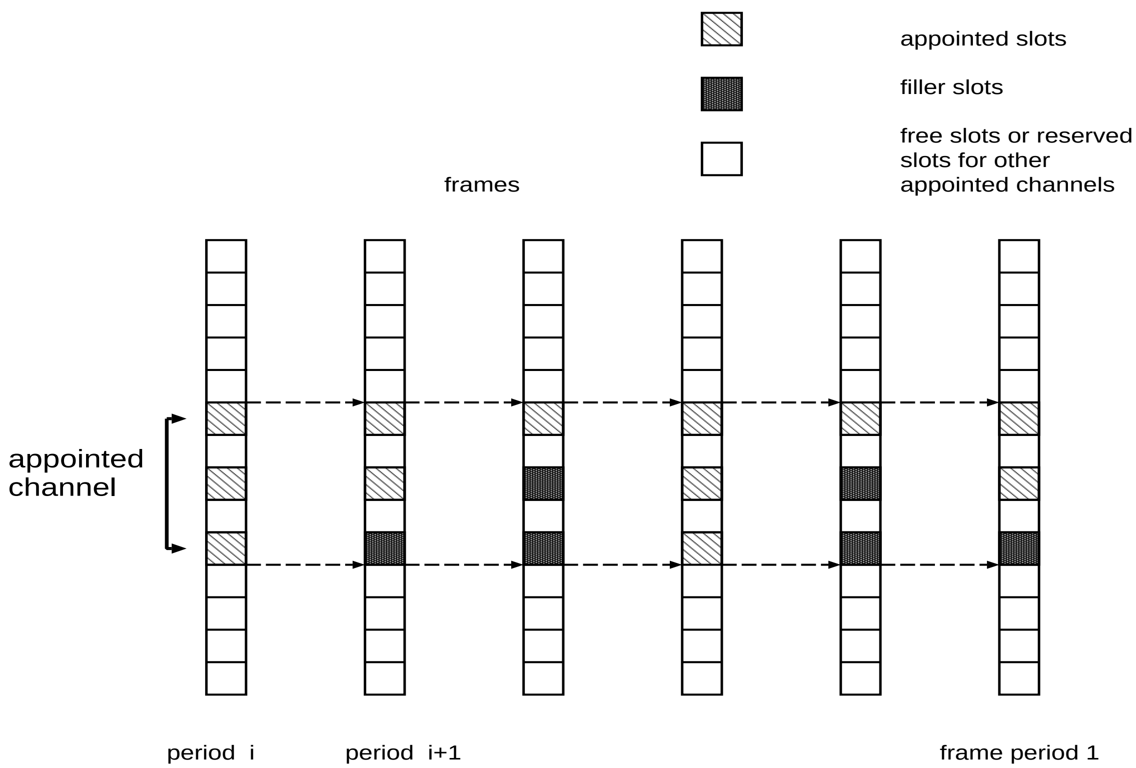

Figure 3.8: The Reserved Slots are Used by Filler Slots Temporarily When They are Idle During the Life Time of Appointed Channel Connection

This mechanism, on the one hand, isolates all appointed channels, avoides the impact of the time-sensitive traffic each other, on the other hand, makes it possible of the appointed channels being shared by appointed slots and filler slots, but appointed slots are priority to filler slots, therefore, MPsLS provides QoS guarantees for time-sensitive applications, also maintains high resource utilization simultaneously.

3.3 Connection Establishment

In MPsLS cloud, all connection routes between edge to edge are pre-determined, the information is kept in link route information database, while the circuit switching at core nodes is based on data forwarding database.

3.3.1 Link Route Information Database

In MPsLS network, there is a management database which is dedicatedly responsible for providing link information, the deployment may be centralized server(a server is shared by several edge nodes), separated server (a server is dedicated to an edge node) or imbedded module (management part is imbedded in edge node).

When an user wants to communicate with others, it sends out a request, where includes the type of the application and necessary parameters, for time-sensitive applications, the parameters include average bit rate, peak bit rate and required average delay; for non time-sensitive applications, needs only average bit rate.

When the message arrives the edge MPsLS node, MPsLS node sends out a route request message to management server or module, where the message is analyzed, and the route from ingress edge MPsLS node to egress node is obtained.

3.3.2 Data Forwarding Database

In each node of MPsLS network, there are several control information databases which are used to manage slots switching.

Channel Switching Table The channel switching table is table-driven, and which recodes the information of each channel. It tells the node the output port of each channel and the slot positions consisting of the channels within the frame. The channel switching table is only used for channel switching. It is updated when each appointed channel is set up or torn down. For each arrived slot, the node firstly examines its tag bits, if it is an appointed slot, then turn to consult channel switching table to find corresponding outgoing port and the slot positions within outgoing frame.

Label Switching Table In MPsLS network, non time sensitive traffic is transmitted by filler slots, the label switching table is used to record the information of filler slots forwarding, it is also updated when each filler channel is set up or torn down. For each arrived slot, firstly the tag bits are examined, if it is filler slot, the label situated in the header after tag bits is consulted and swapped, then, sent to corresponding queue to wait to be transmitted.

3.3.3 Establishing Connection and Channel

For time-sensitive application, the edge MPsLS node at sender side of the applocation sends out the channel request message along the route hop by hop till the egress node, if the requirement is satisfied, the channel is set up, otherwise, the request is refused. For non time-sensitive application, the ingress edge MPsLS node sends out the request message, if only there is enough left bandwidth, the channel is set up, otherwise, the request is refused.

Furthermore, the setup processes of appointed channels are as following: firstly MPsLS edge node inspects the channel switching table, if there is enough free slot positions within an offset range c, from a to $a+c$, records the tailer position $a+c$ of the offset range c, then, sends out the request message to second hop, which include the positions $a+c$ and value c, then, this node investigates if or not there is enough free slot positions within the new range from $a+c+n_p$ to $a+n_p+2\cdot c$, and so on, hop by hop, till the egress edge node, if free slot positions can be found at all sections along the path, the channel is set up, otherwise, do another attempt by selecting a new offset range at ingress link section, which is not overlapped with that used before, namely, replacing a with $a+c'$, where $c' > c$, and repeatting the process above. If given number of attempts are all failed, calling request of this application is rejected.

3.4 Architecture of Edge Node

The ingress edge nodes of MPsLS provide more functions than core nodes, not only construct the frames, but also establish connections, differ from the types of service, and segment and reassemble the layer-3 packets.

3.4.1 Establishing Connections

MPsLS networks are connection-oriented, thus, the channels must be established firstly before transferring data.

For a time-sensitive application, when the application request message arrives to the ingress edge node from the users, MPsLS edge node firstly consults the management server or management module to obtain the route between source and destination edge nodes of MPsLS, then, sends out channel request message hop by hop along the route until the egress nodes. If all resources and time constraints can be satisfied according to the QoS requirement, the appointed channel is set up, otherwise, the connection is refused. For a non time-sensitive application, if only enough bandwidth can be met, the connection can be established, otherwise, the request is rejected.

3.4.2 Segmentation and Reassembly of Packets

All information in the IP networks is transferred in the form of variable length packets. But, in the MPsLS area information is transferred in fixed length slots. Since the length of most packets is longer than a slot, the packet must be segmented into short data pieces at ingress nodes according to their traffic type. The packets carrying time-sensitive applications are segmented into data pieces of 510 bits length, then insert two bits tag signal in the header of each data piece, and consist of the MPsLS slot. Those carrying non time-sensitive applications are segmented into data pieces with length of 478 bits, then insert two bits tag signal and 32 bits MPLS header label.

If MPsLS network is directly linked to some special real-time equipments which send out the data at bit stream, MPsLS edge node can also segment the bit stream into data pieces so that the packetization process is cut.

At egress nodes the slots are handled after the whole frame has arrived; the slots of frames are reassembled into packets or bit stream according to their original order.

For the packets whose size is not integral times of the slot length, packet segmentation may cause a decline in the channel utilization, due to last slot carrying the tailer of the packets may be not full.

3.4.3 Frame Level Synchronization

At all output ports of each ingress node, the frames are sent out at frame level synchronization, it is shown in Figure 3.9.

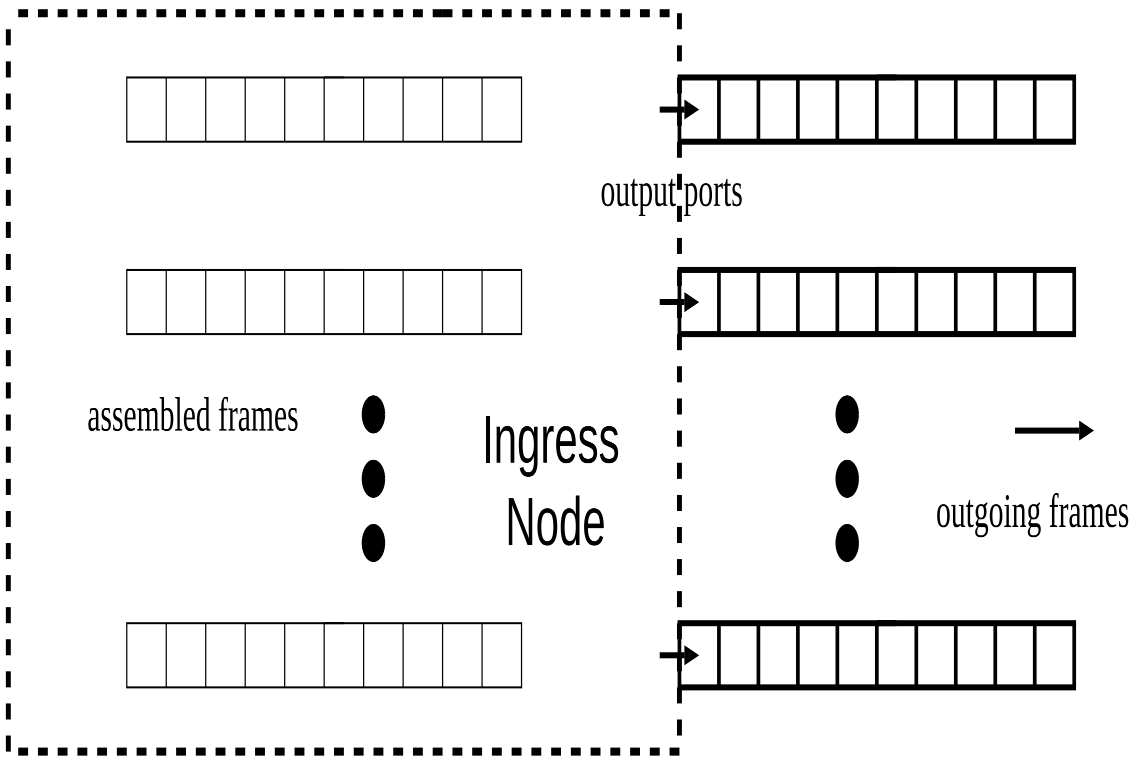

Figure 3.9: Frame Level Synchronization

3.4.4 Sharing Appointed Channels

Because the minimum bandwidth of the channels with one slot in a frame is 4Mbps, if reserving such a large channel for an application with small bit rate, such as 64kbps network phone, although which does not waste network resource, since the idle slots can be temporarily used by filler slots, but it must result in the reduction of number of time-sensitive application flows transmitted at a moment. To increase the quantity of time-sensitive application flows in the network, we suggest to use bundle technology, namely, let several small bit rate real-time application streams share a channel if their edge nodes(both ingress and egress nodes) are same.

For the channels shared by several sub-flows, former 6 bits next tag bits are used to identify the flows, thus, the maximum number of flows sharing a channel is 63, which mean that maximum period of each independent application sub-flow becomes 63τ, namely, a slot including $L' = 512 - 2 - 6 = 504$ bits is transmitted in cycle for each sub-flow in interval of 63 frame periods. The minimum bandwidth of the sub-flows is,

$$b_{sub} = \frac{L'}{63\tau} = \frac{504bits}{63 \times 125\mu s} = 64kbps \tag{3.2}$$

3.5 Architecture of Core Node

Data forwarding in core nodes of MPsLS is done at slot level synchronization.

3.5.1 Slot Level Synchronization

Due to the long linking distances, there is a noticeable fiber buffer effect, therefore, at non ingress nodes, the arrived frames through different ports must be non frame level synchronous, which is illustrated in Figure 3.10, thus, only slot level synchronization forwarding is implemented at all input ports of each non ingress node. But all output ports of each non ingress node still implement more strict frame level synchronization transmission.

To achieve slot level synchronization on the case of fiber buffer effect, the little time offset less than one clock time among data flows from different links is regulated with fiber loops and that less than a slot time is adjusted with shift registers.

3.5.2 Structure of Reorder Buffer and Implementation

For less strict synchronous channels, it is possible that the slots in incoming frames arrive early, but depart late in outgoing frames, which requires to adjust the orders of the appointed slots. In this paper, we use reorder buffer to realize the function. The structure of reorder buffer is shown in Figure 3.11, it is consisted of a memory block, which consists of several storing units, the length of each unit is 512 bits, which equals the length of the slots. At every slot period, the slot in the header of reorder buffer is read out, and filled in the outgoing frame, at the same time, the rest slots in reorder buffer shift forward one unit.

Assume that the header address of the reorder buffer is m, for a slot, if the lag between the slot positions in incoming frame and in outgoing frame is n_{lag}, it is sent to unit $n_{lag}+m$ in reorder buffer after processed, then, it shifts forward an unit every slot period, till reachs to the header of the reorder buffer after n_{lag} periods, at last, it is read out from the header of reorder buffer, and filled into the outgoing frame. Specially, for exact synchronous slots, the lag between the slot positions in incoming frame and in outgoing frame just equals the processing time n_p, it means that processed slot is just sent to header unit of reorder buffer.

Within the strip range with width of c, since the slot in any position in the former link section can utilize any one position in the later link section, specially, if the channel

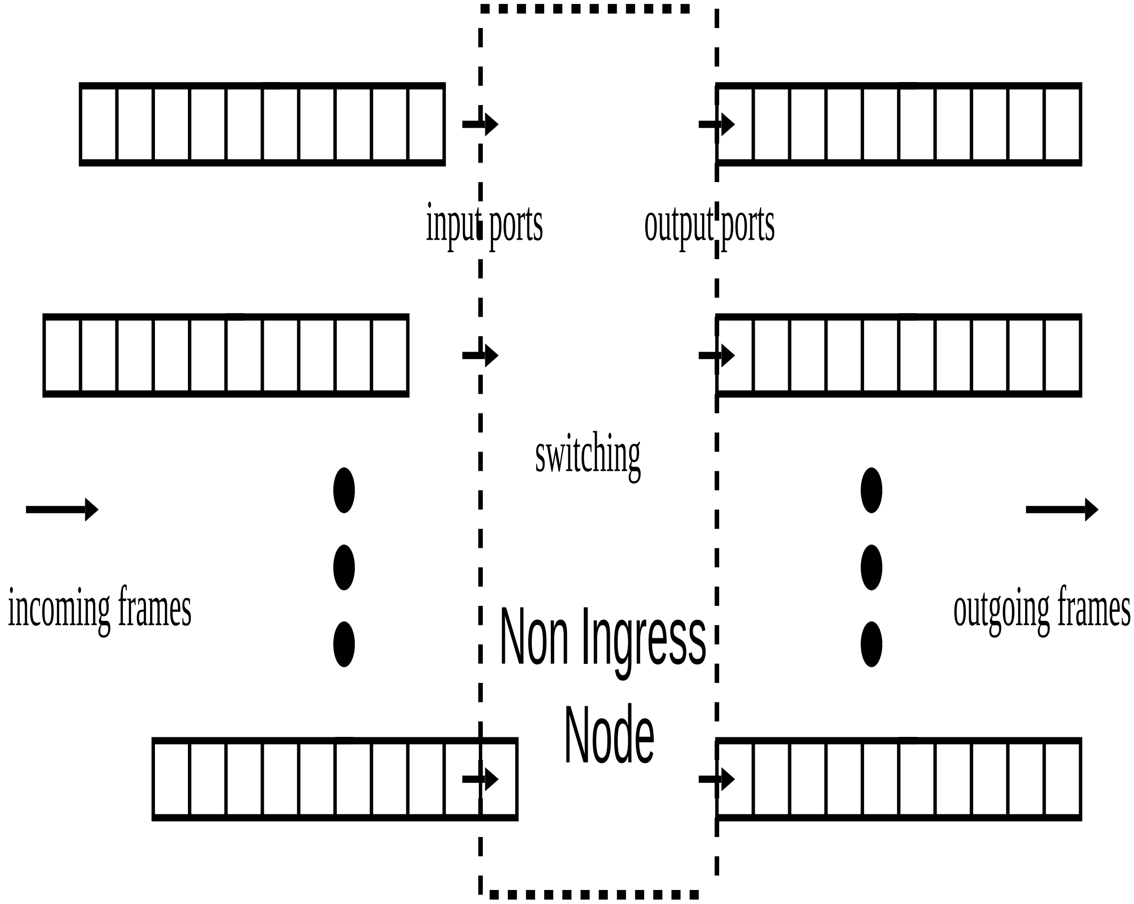

Figure 3.10: Slot Level Synchronization

including only one slot, and in the former section it is located in the header position, while in the later section it just is located in the tailer position, thus, it is necessary of large reorder buffe which accommodates enough space to satisfy the long waiting time, the minimum size is $2 \cdot c$ units. For different time-sensitive applications, the c is variable with the QoS requirements, let the size of the reorder buffer be 400 units, the allowable maximum offset range, or the width of the strip is 200 slots.

3.5.3 Forwarding Process of Appointed Slots and Constructing Out-going Frames

The forwarding process of appointed slots is shown in Figure 3.12, in this figure, A_{pp0}, A_{pp1} and A_{pp2} are slots from different application flows, assume that A_{pp1} is exact synchronous, while A_{pp0} and A_{pp2} are less strict synchronous, therefore, processed A_{pp1} is just situated in header of reorder buffer when it enters reorder buffer, the interval, n_p, is minimum latency time inluding buffer, pipeline processing and reorder process. A_{pp1}, A_{pp0} and A_{pp2} address their positions in reorder buffer according to lags between the positions in incoming frame and outgoing frame.

Figure 3.13 shows the process of constructing the outgoing frame. At each output port of the core nodes, only one slot is filled in current position of outgoing frame in each

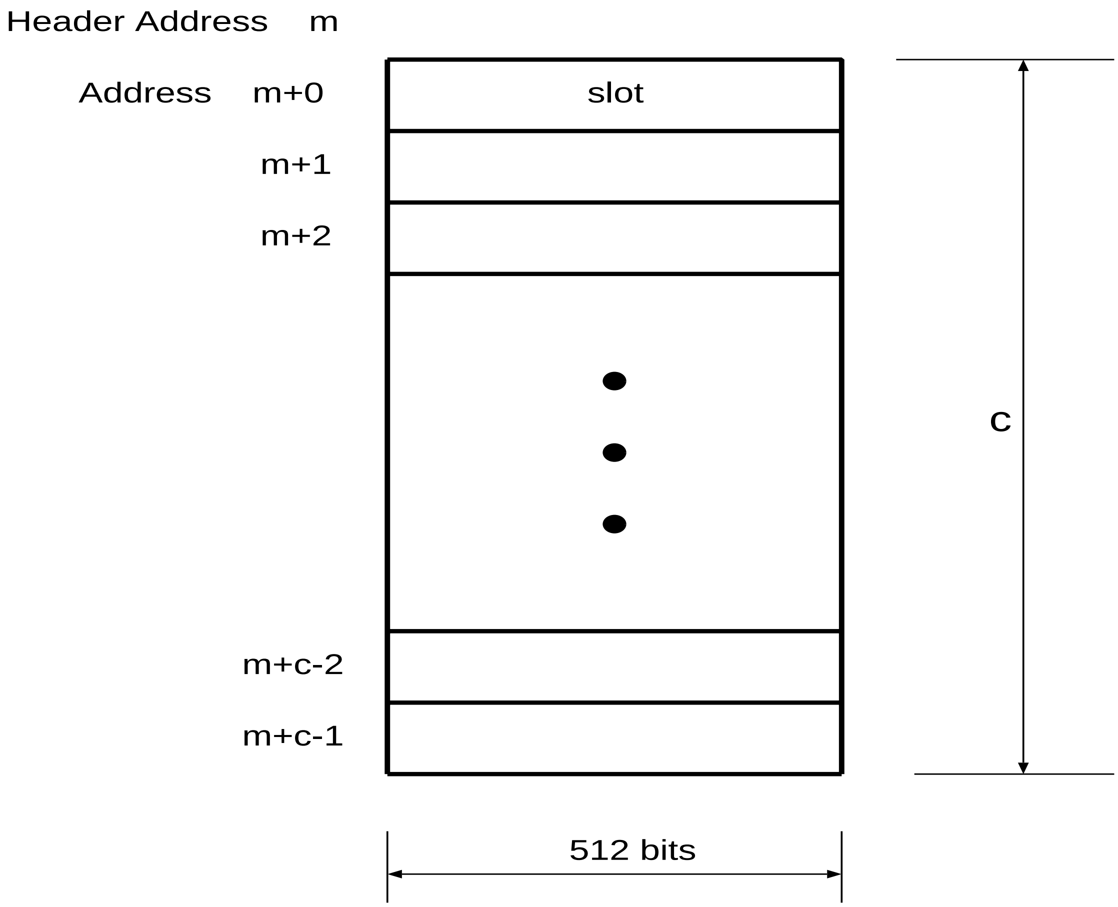

Figure 3.11: The Structure of Reorder Buffer

slot period, the slot may be appointed or filler, if there is an appointed slot in the header of the reorder buffer, waiting to be sent out, it is read out at priority, otherwise, a filler slot will be read out from the header of queue storing filler slots, if the queue is free, the corresponding slot in outgoing frame is idle. Then, similar process is done at every slot period one by one.

Furthermore, the scenario of slots forwarding process at core nodes is illustrated in Figure 3.14.

3.5.4 Pipeline Processing

All incoming slots are forwarded at slot level synchronization. The forwarding process consists of scanning slots in a frame, identifying slot types, looking up the channel tables or label switching tables, and assembling new outgoing frames. It is difficult to finish these sub-processes within only one synchronous slot period. In order to reduce the latency time during forwarding process, we suggest using slot level pipelining. One or two processing steps are implemented in each pipeline stage. The usage of the pipeline can lower the performance requirement for the nodes. The number of pipeline stages is determined by the software and hardware resources.

Furthermore, two different types of pipelines are used for appointed channel slots and

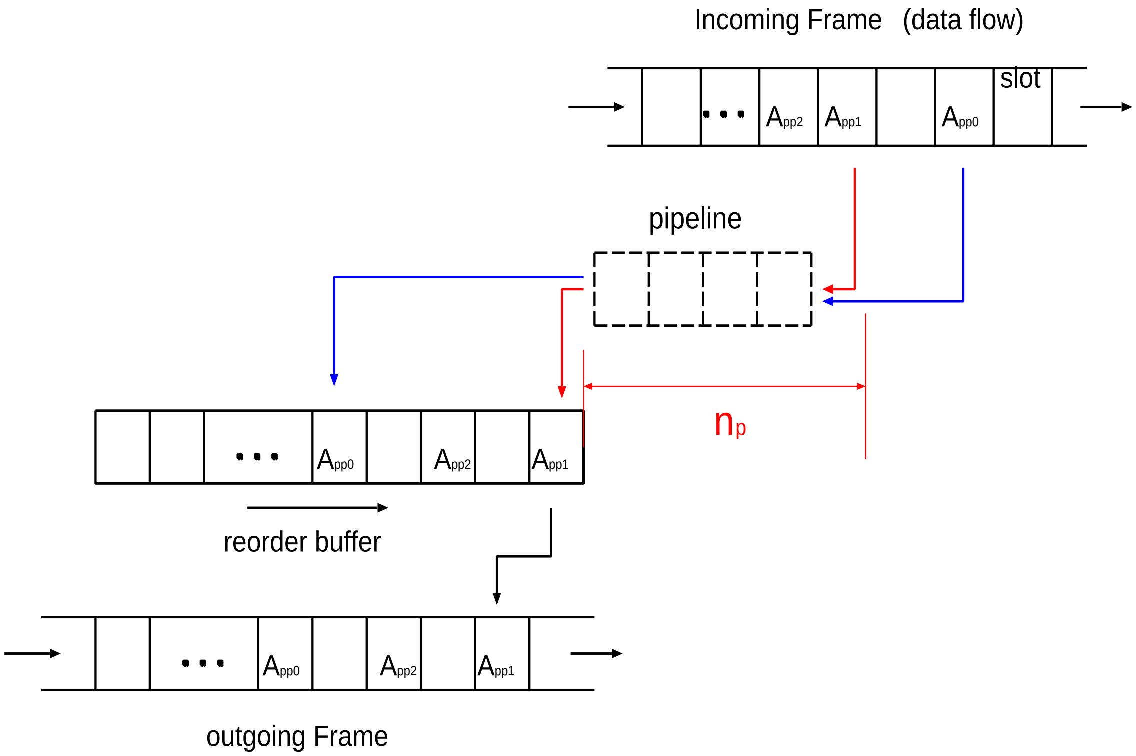

Figure 3.12: Forwarding Process of Appointed Slots

filler channel slots, respectively, which is easy to implement by hardware, and specialized hardware for different slot formats can gain higher speed that common hardware.

Figure 3.15 illustrates the pipeline process of appointed channel slots. In the figure, A_{pp} and F_{iller} indicate slot types, appointed slots or filler slots. Number $0, 1, 2, \cdots$ mean the order of the slots in incoming frames.

At an input port, the slot in incoming frame is sent to buffer in order in each slot period, then the type of the slot in buffer is indentified by checking tag bits, if the slot is appointed, it is sent to pipeline for appointed slots, the pipeline includes several stages, the table look up stage TL, the position finding stage PF, and the data copying stage DC.

In TL stage looking up the appointed channel table, finding the output port, in the PF stage calculating the position in reorder buffer, and then, in the DC stage copying data to the proper position in reorder buffer situated in output port.

In the case of the exact synchronization mode, the slot from the buffer just arrives the header position of reorder buffer, it is put away in next slot period, and filled in outgoing frame.

While in the less strict synchronization mode, in PF stage calculating the position in reorder buffer, then, the data in the buffer are copied to the proper position in reorder buffer, later, the slot is shifted one by one at each slot period, till the header of reorder buffer, after that, it is filled in outgoing frame as that in the exact synchronization model.

If the slot is filler, the processing is done by other pipeline dedicated for filler slots, which is shown in Figure 3.16, the pipeline includes the label switching table look up stage TL, the header modifying stage HM, and forwarding to output queue stage FQ.

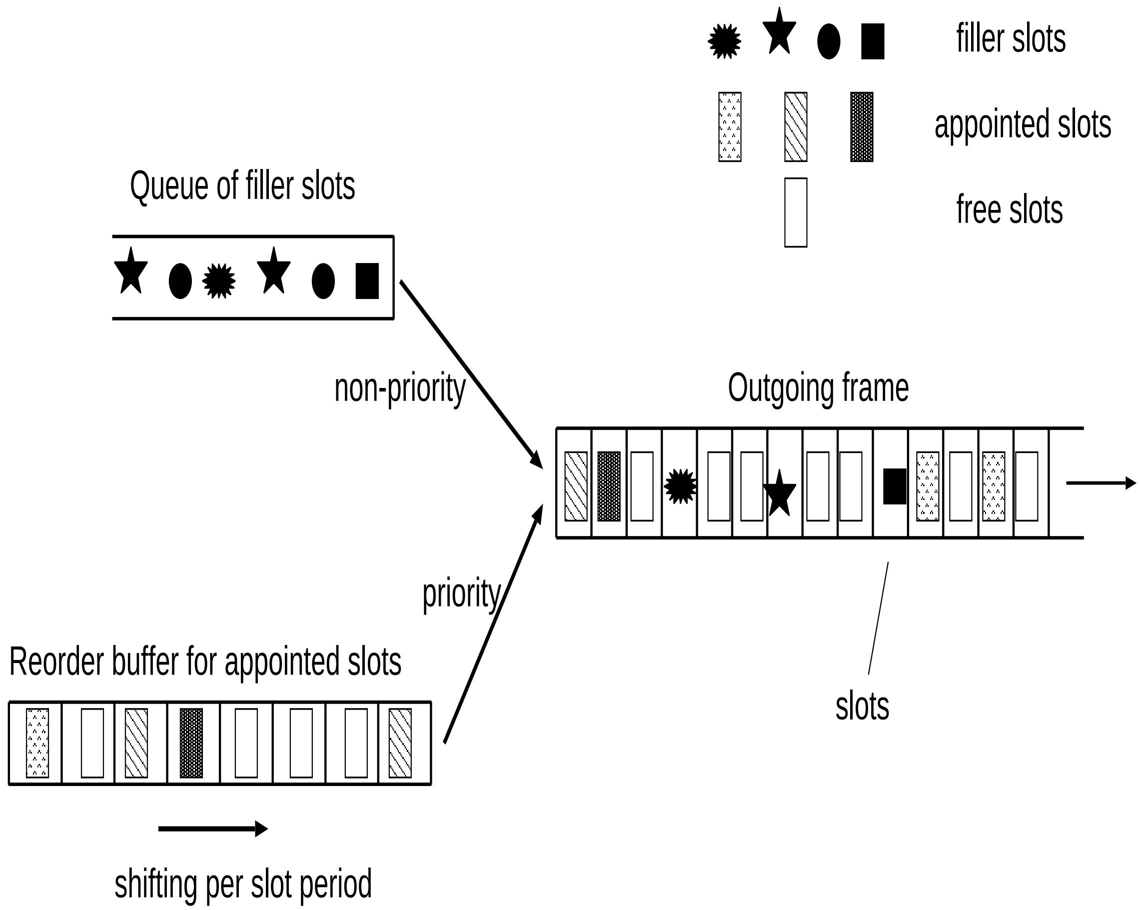

Figure 3.13: Constructing Outgoing Frame

Firstly, in TL stagethe looking up the label switching table, after that, in HM stage modifying the label header of filler channel slots, at last, the slot in the buffer is sent to queue in the corresponding output port in stage FQ.

The complexity of control message forwarding is resolved with non pipeline software. The control slots at the header part of a frame (the outband control slots), and other slots carrying control information (the inband control slots), may be handled by software.

3.6 Application of MPsLS

An MPsLS network is a network implementing MPsLS protocol, it performs MPsLS switching in layer 2, while supports upper layer multi-protocol including IP, TCP, UDP and so on, therefore, MPsLS is closed network liking Frame Relay, ATM.

MPsLS can link directly with other networks implementing different protocols, such as Frame Relay, ATM, as shown in Figure 3.17. Namely, MPsLS supports multi-protocol at edge nodes, edge MPsLS nodes have not only frame processing function which is speculiar in MPsLS, but also the functions of MPLS edge node, therefore, MPsLS edge node can automatically identify the multi-protocol packets and convert the transfer models between MPsLS and others, but in the core area, only MPsLS transfer mode is performed.

The applications have two access models into MPsLS network: 1) directly link with

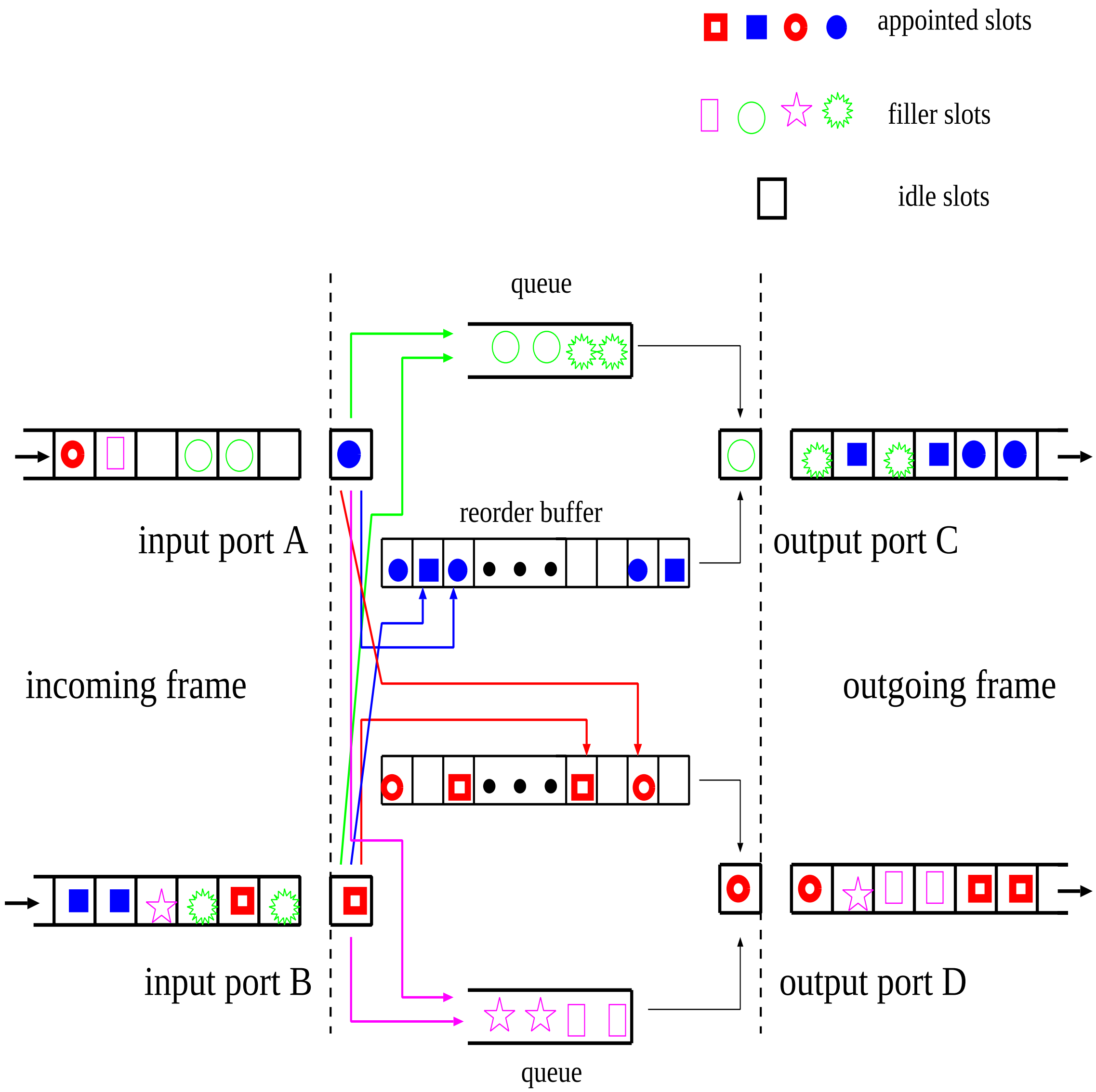

Figure 3.14: The Scenario of Slots Forwarding

MPsLS edge node through special hardware, such as Telephony, or common equipments, such as computer and workstation, 2) indirectly through non-MPsLS networks.

In 1)., the whole network is a simple MPsLS network, the MPsLS edge node needs to support only basic frame function, (this type of edge node can be called simple MPsLS edge node or MPsLS edge node), the route to destination node across MPsLS area can be found by investigating the management module of MPsLS edge node, then the sender side of the application data flow sends out a resource request message, if the application data flow is time-sensitive, in the message, some stream characteristics, such as bandwidth and delay requirement, are included, the parameter values can be identified by digital item adapter similar to Digital Item Adaptation (DIA) engine [113] of MPEG-21 standard.

The mapping of the stream parameters to connection channel is

$$N_{slot} = \lfloor \frac{A^p}{L_{slot}} \rfloor \tag{3.3}$$

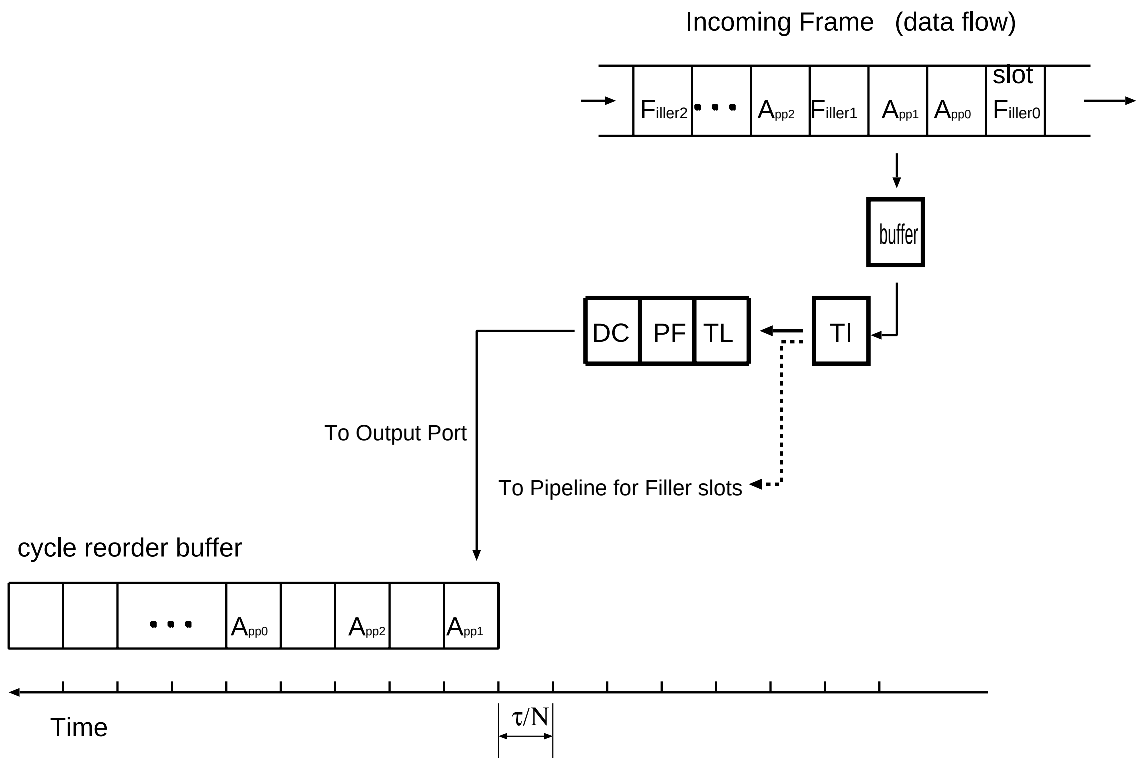

Figure 3.15: The Pipeline Implementation of Appointed Slots

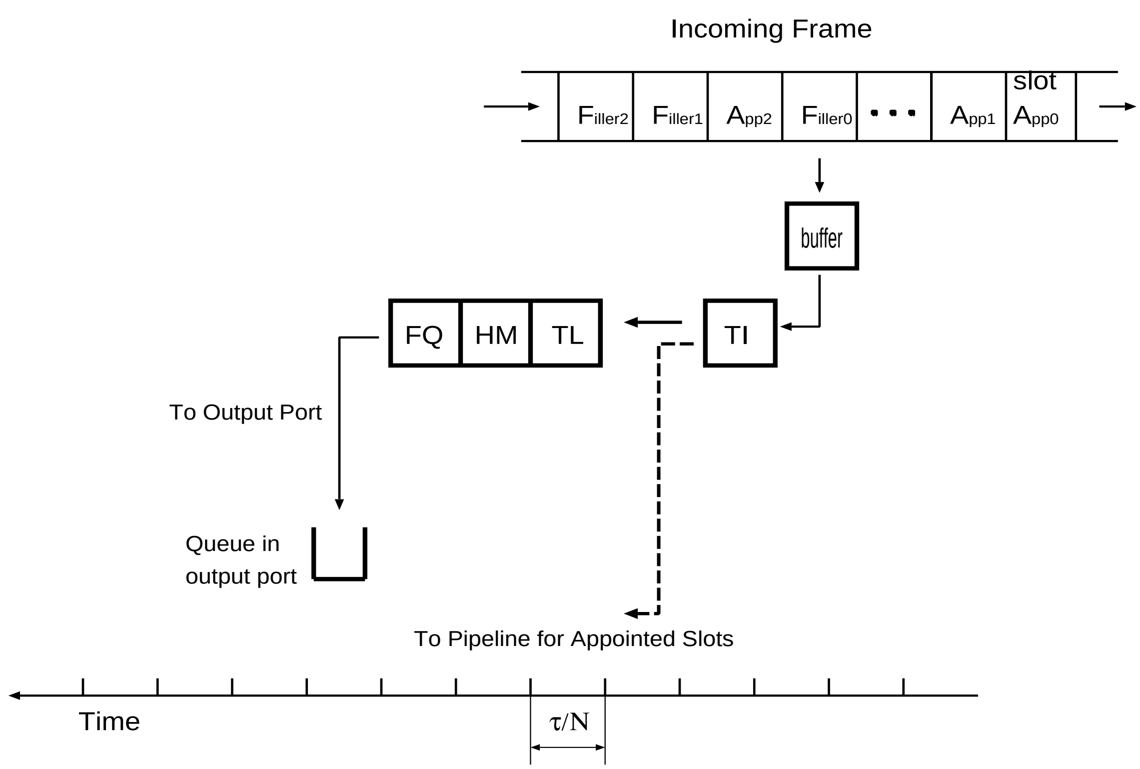

Figure 3.16: The Pipeline Implementation of Filler Slots

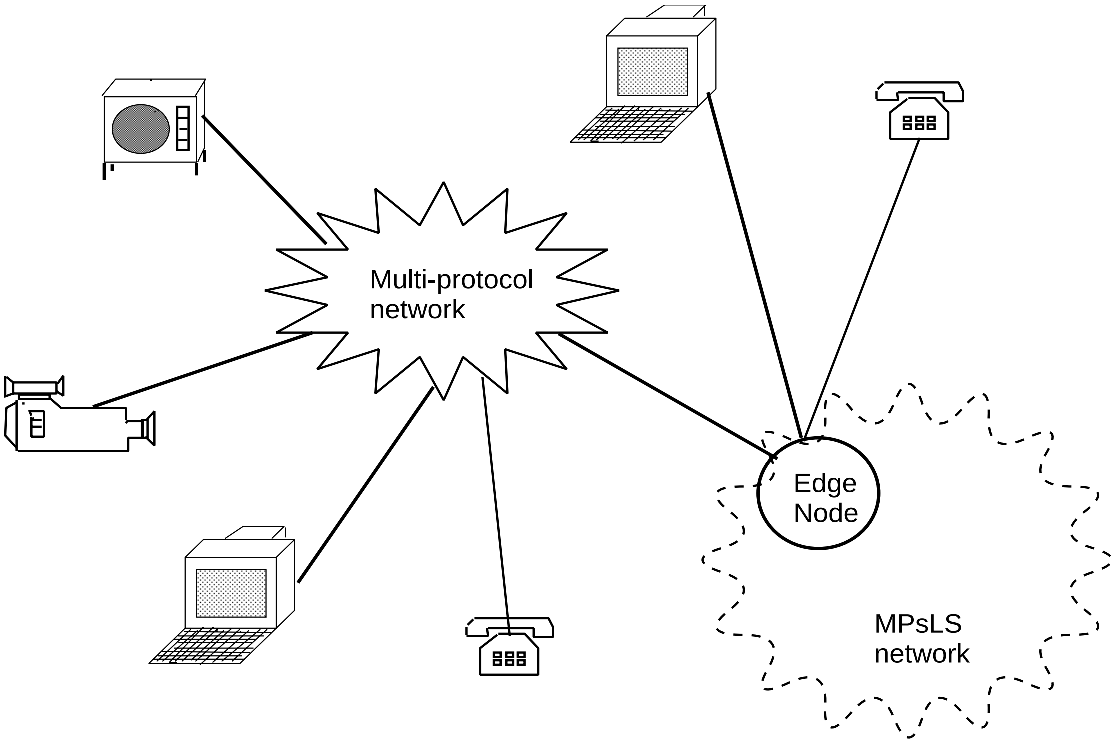

Figure 3.17: Access Model of MPsLS Network

where $\lfloor\star\rfloor$ denotes an integer not smaller than $\star$, L_{slot} is the bandwidth corresponding to a slot, A^p is peak bit rate of the application flow A, N_{slot} is the number of required slots.

And

$$c_{offset} = \lceil \frac{D_{max}}{\tau} \cdot \frac{L_{frame}}{2 \cdot N_{hops}} \rceil \tag{3.4}$$

where $\lceil\star\rceil$ denotes an integer number not larger than $\star$, c_{offset} is the width of the strip including the channel, N_{hops} is the hops number, τ and L_{frame} are period and length of the frame, respectively, D_{max} is maximum variable delay except the access delay at ingress node, since the access delay can be considered as constant, half of the frame period, if the channel is reserved according to the peak bit rate. Therefore, a pair of stream parameters (A^p,D_{max}) can be mapped into channel parameters (N_{slot}, c_{offset}).

In 2)., the MPsLS edge node needs to support both functions of MPsLS edge node and MPLS edge node, (this type of edge node can be called gateway), if the message arriving MPsLS edge node is a IP packet, IP-header is investigated as in MPLS edge node, then IP destination address included in it is mapped into equivalent MPsLS address by using subnet mask, then consult the mapping table, find the MPsLS edge node directly linking the destination user, finally, looking up the MPsLS connection table, obtain all hops along the suitable route. All IP packets will be considered as non time-sensitive applications, which are carried by filler channel passing through MPsLS area.

If the traffic is from MPLS or ATM, when the request message applying for virtual path arrives the edge MPsLS node, the parameters including in the message are investigated as in MPLS edge node, then the virture path like in MPLS or ATM is replaced by the MPsLS channels in MPsLS cloud, at this case, the whole MPsLS area can be considered as a blackbox for all outside applications, specially, if the request message includes QoS requirement, the channels across the MPsLS area will be set up as appointed channels, otherwise, as filler channel. After setting up the virtual path or channel, the application data transmission is performed at MPLS mode and MPsLS mode in outside and inside of MPsLS cloud, respectively. There is a convertion process at linking gateway.

Accordingly, for the applications directly linking to MPsLS network, the determined service QoS can be guaranteed when the traffic across the MPsLS cloud; while for the applications linking to MPsLS network through other networks, the QoS depends on the added effects across different areas.

3.7 Summary

This chapter mainly described the architecture of MPsLS networks, furthermore, main characteristics, several important control elements and implementation process of data forwarding at MPsLS edge and core nodes are explaned carefully. Due to using synchronous channels to transmit time-sensitive applications traffic, avoiding the contention for outgoing slot positions and congestion of arrived appointed slots, in theory, MPsLS can provide better QoS for time-sensitive applications, besides, due to introducing tag bits, the idle slots reserved for appointed channels can be utilized by filler slots carrying non time-sensitive applications traffic, therefore, MPsLS can maintain high resource utilization, simultaneously. In later four chapters, we will quantificationally analyze the performance of MPsLS scheme.

Chapter 4

Access Performance of Time-sensitive Applications and Percentage of Appointed Slots in Frames

In MPsLS networks, all appointed channels are connection-oriented, not only the quantity of used slots consisting of the channel is determined, but also their positions in the frames are fixed during transmission, thus, both conditions of enough bandwidth and suitable slot positions must be satisfied for establishing an appointed channel, otherwise, the channel can not be set up.

Therefore, the access performance is one of important parameters to evaluate MPsLS network. Besides the dynamic access performance, the statical parameter, the percentage of appointed slots to total slots in frames is also very important for evaluating the performance, because the appointed slots can transfer both time-sensitive and non time-sensitive traffic (if time-sensitive traffic arrives, which is transmitted at priority, if not, non time-sensitive traffic is transmitted temporarily). Therefore, the larger the ratio of appointed slots to total slots in frames is, the better the flexibility of MPsLS network is.

In this chapter, we discuss both dynamic access performance of time-sensitive applications, and the percentage of the appointed slots in frames.

4.1 Access Performance of Time-sensitive Applications

For each access request, only if the number of free slots is enough and the slot positions are also satisfying the requirement, the new channel is allowable to be set up, then the application traffic can be transmitted, therefore, we can use calling success probability to express the access performance in MPsLS network. So-called calling success probability is the probability of an application obtaining enough free and suitable slot positions for establishing an appointed channel corresponding to once time request attempt within a strip range along a path.

According to the mechanism of setting up appointed channels, all slots consisting of a channel are limited to given strip range, which is the searching area for each calling attempt.

Assume that the width of the strip range corresponding to once attempt is c, the number of slots required for setting up one or more time-sensitive applications is x, then the calling success probability, P_{succ}, on once attempt within this strip range, satisfies,

$$P_{succ} = p(i \geq x) \tag{4.1}$$

where i is the number of free slots along a path, p is the probability.

4.1.1 Simplified Model on Distribution State of Used Appointed Slots

If the network is completely time-sensitive traffic free, namely, there is not any time-sensitive traffic passing through the network, thus any new calling attempt for appointed channels with the bandwidth less than the width of the strip range is always successful, because all slot positions can be allocated to the channel according to completely same slot positions distribution in all link sections, on this case, P_{succ} always equals 1, this is only a special case.

But generally, there is time-sensitive traffic in the network, namely, some slot positions in the strip range have been reserved for other time-sensitive applications as appointed channels, therefore, we need to give a model to describe the distribution state of time-sensitive traffic before theoretical anlysis.

Since MPsLS supports general network topology of star-shaped mesh structure, on this case, a switch node has several input and output ports, a link fiber connects two switches on its two ends. We consider three nodes shown in Figure 4.1, the links $X \rightarrow Y$ and $Y \rightarrow Z$ are two of many fiber links.

If a path passes through link sections $X \rightarrow Y$ and $Y \rightarrow Z$, due to there exists time-sensitive traffic, some slot positions in the frames passing two sections have been partly reserved by other appointed channels, thus the calling success probability of setting up a new channel does not strictly equal 1, the value is determined by the distribution state of used slots and the number of used slots (or utilization). We suppose that the used slot positions in different link sections are completely stochastic and independent, the reason is that: for any two neighbouring link sections, $X \rightarrow Y$ and $Y \rightarrow Z$, due to the joint switching node Y has several input and output ports, respectively, the traffic passing through sections $X \rightarrow Y$ is not always passing $Y \rightarrow Z$, and vice versa, therefore, the most of traffic passing through sections $X \rightarrow Y$ and $Y \rightarrow Z$ is unrelated, only small part of traffic passes through both sections simultaneously. Furthermore, even if for the related traffic, the relationship of the slot positions used in both sections is variable with the types of the channels (for example, if the channels are exact synchronous, each slot has constant lag value, otherwise, whole channel has a lag, the value of each slot is variable). Therefore, it is reasonable of assuming that the slot positions used in different sections are independent and stochastic, furthermore, the states (used or not) of each slot position in strip range can be decided by binomial distribution.

Based on the general network structure and distribution feature of used slot positions, we will discuss dynamic calling success probability on different utilizations to two types of appointed channel models, exact synchronous and less strict synchronous channels.

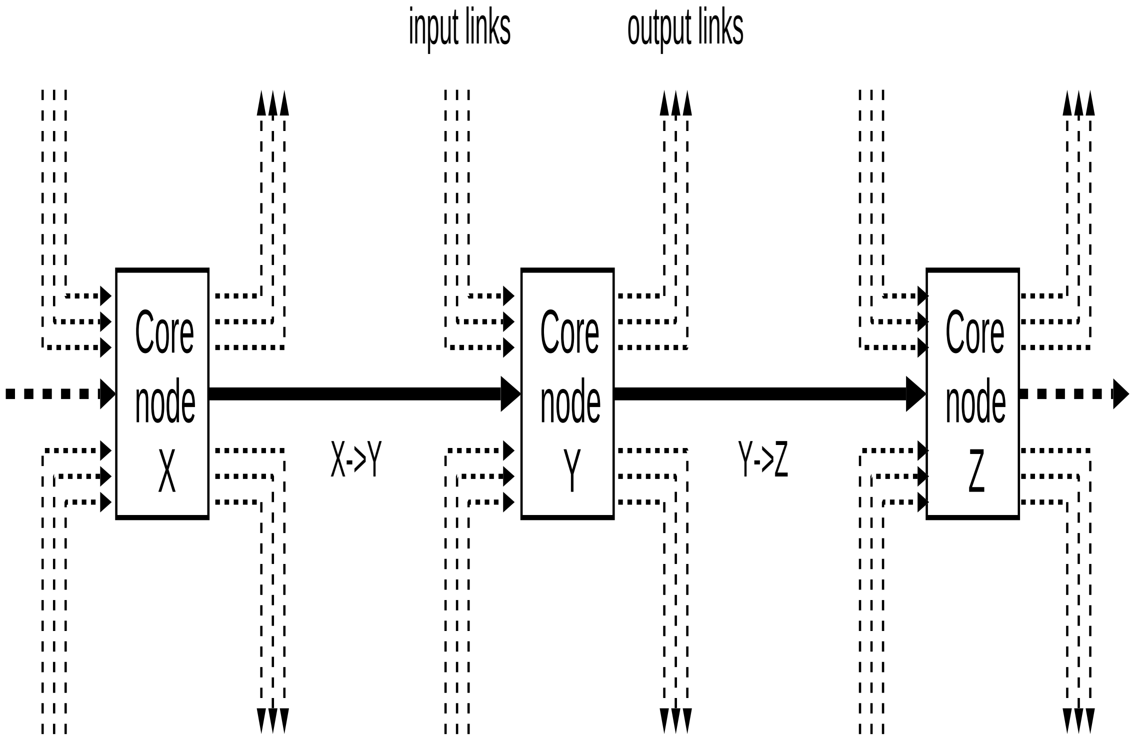

Figure 4.1: The Common Linking Structure at Core Nodes

4.1.2 Exact Synchronization Channels

An exact synchronous channel is shown in Figure 4.2, where the distribution of the slot positions is completely same in different link sections, the lag between adjacent offset ranges is constant, the value equals minimum latency time, n_p, which includes pipeline processing, reorder processing and so on.

Analytic Solution

Within the strip range, any slot position in an offset range has only one corresponding position in each other offset range, assume that the width of the strip range is c, it means that the offset range in each link section is c, from ingress edge to egress edge, if the application data flow passes through h hops, besides, the usage state of each slot is completely independent and stochastic within c, and the values of the utilizations of any slot positions on the different linking sections equal ρ_1, ρ_2, $\cdots$, ρ_h, respectively.

Then, for any one slot position, the probability of this position being free on all sections along the path is,

$$P_s = \prod_{i=1}^{h}(1 - \rho_i) \tag{4.2}$$

On the contrary, the probability of it being non free is,

$$P_{no} = 1 - \prod_{i=1}^{h}(1 - \rho_i) \tag{4.3}$$

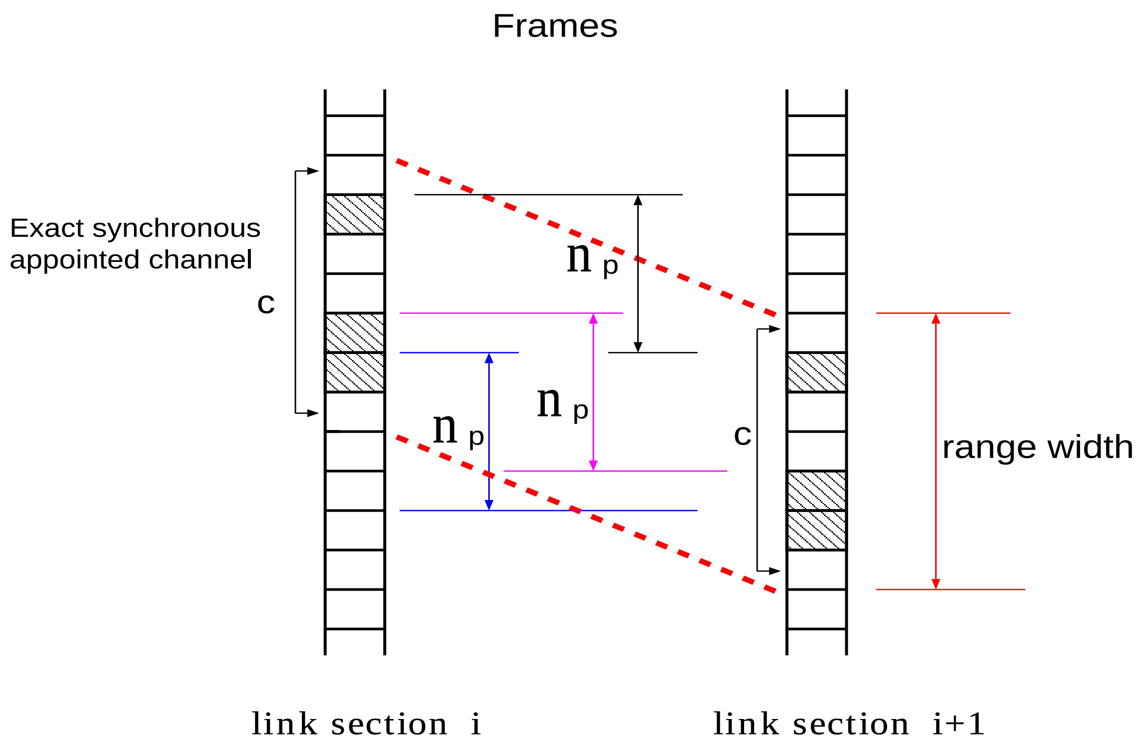

Figure 4.2: The Slots Positions Distribution of Exact Synchronous Appointed Channels

Therefore, for each calling attempt, the calling success probability of a new channel with x slots is,

$$P_{succ} = \sum_{j=x}^{c} {}_cC_j \cdot (\prod_{i=1}^{h}(1-\rho_i))^j \cdot (1 - \prod_{i=1}^{h}(1-\rho_i))^{c-j} \tag{4.4}$$

To simplify the calculation, let $\rho = \rho_1 = \rho_2 = \cdots = \rho_h$, then

$$P_{succ} = \sum_{j=x}^{c} {}_cC_j \cdot (1-\rho)^{h \cdot j} \cdot (1-(1-\rho)^h)^{c-j} \tag{4.5}$$

Calculation Results

Currently, the mean distance between two users in Internet is about 15 hops [97], since the access of the users into Internet usually passes by local networks, thus, the average hops between edge MPsLS nodes in backbone area should be less than 15 hops.

Here, to keep stringency, we still assume that the average distance between two edge MPsLS nodes is 15 hops, then, the calling success probability for requesting a new appointed channel on different utilization can be calculated by equation (4.5). Since the average hops is 15, we calculate the value on the conditions of the hops equal 5, 10, 15 and 30, the number of the slots consisting of the channel is 1, 5 and 25, respectively. On different $c = 20, 50, 100, 200, 400$ and 500, the calculation results are illustrated in Figures 4.3 to 4.14:

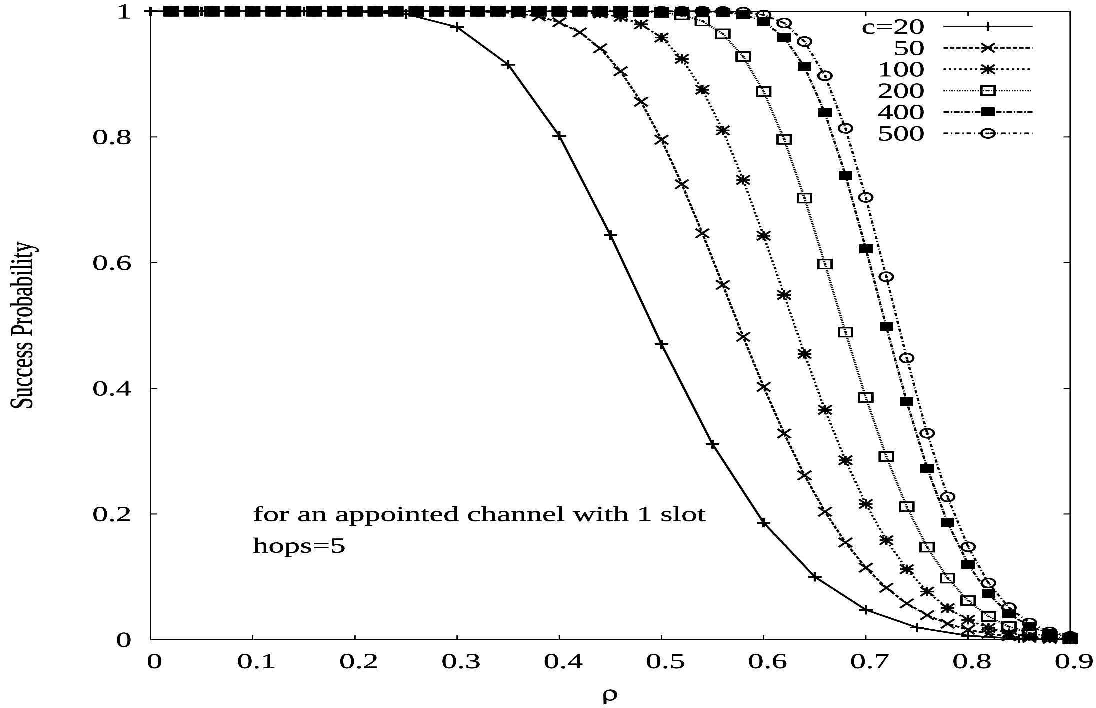

Figure 4.3: The Calling Success Probability for Setting up an Exact Synchronization Channel with 1 Slot, When Hops=5

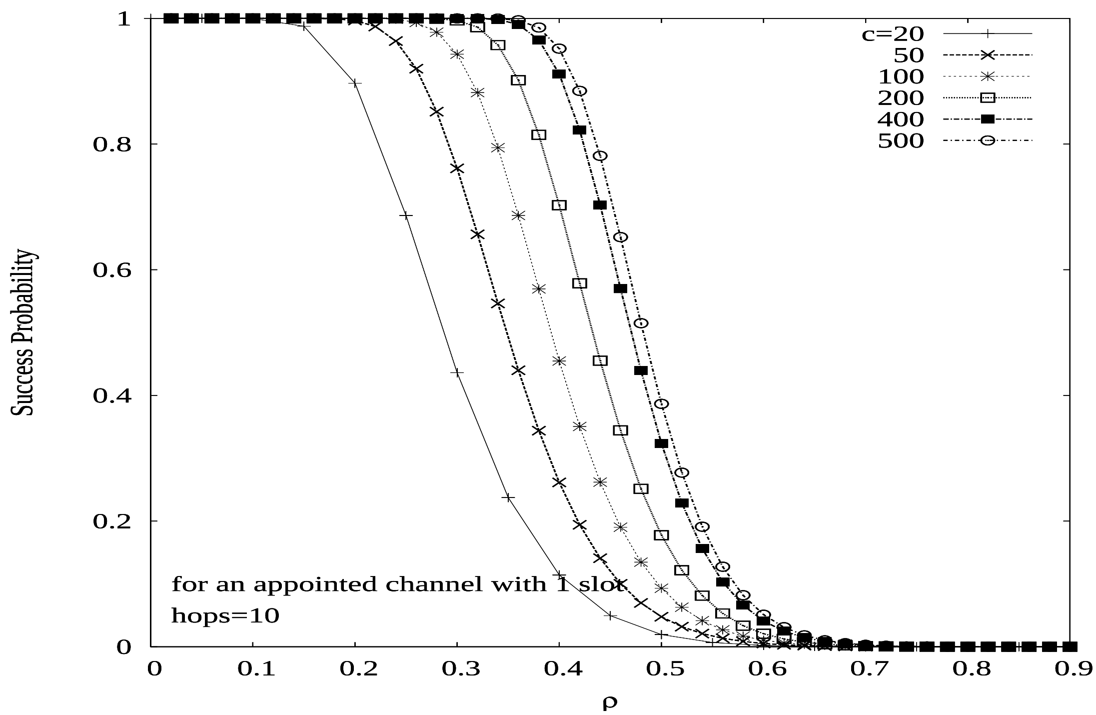

Figure 4.4: The Calling Success Probability for Setting up an Exact Synchronization Channel with 1 Slot, When Hops=10

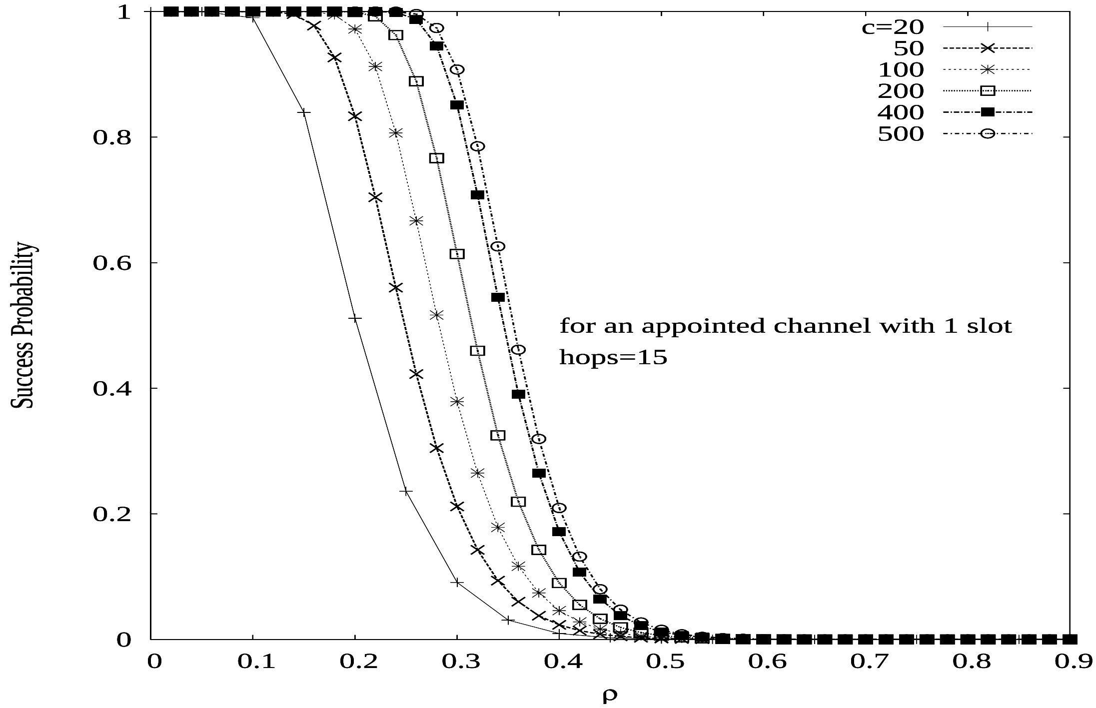

Figure 4.5: The Calling Success Probability for Setting up an Exact Synchronization Channel with 1 Slot, When Hops=15

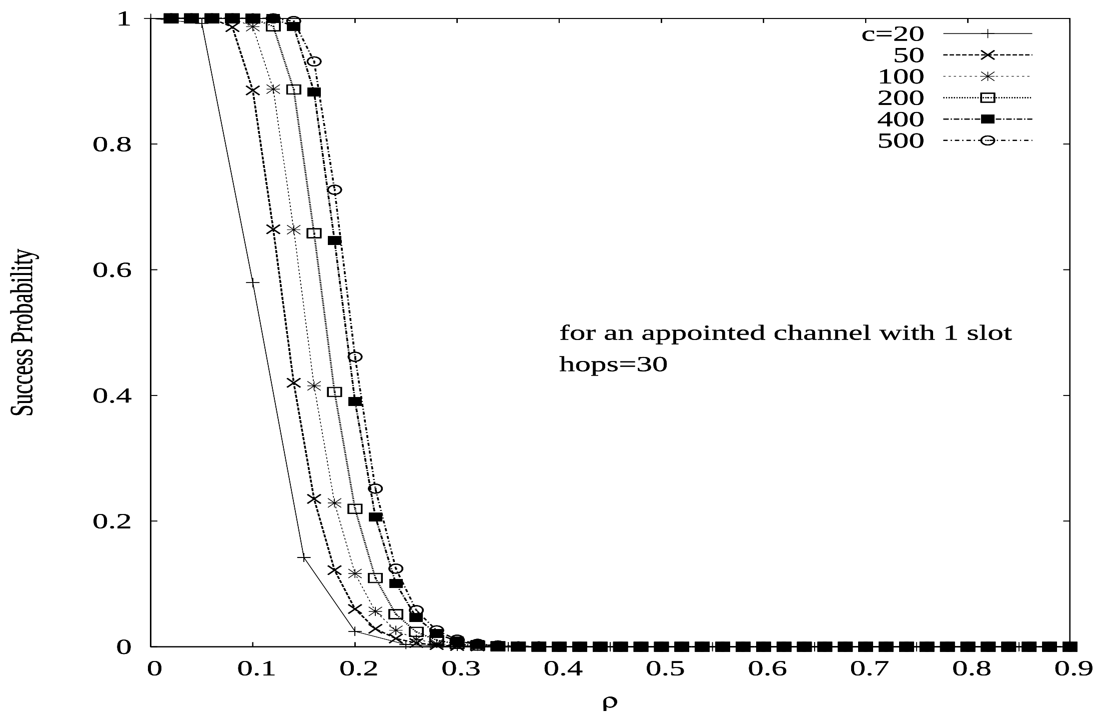

Figure 4.6: The Calling Success Probability for Setting up an Exact Synchronization Channel with 1 Slot, When Hops=30

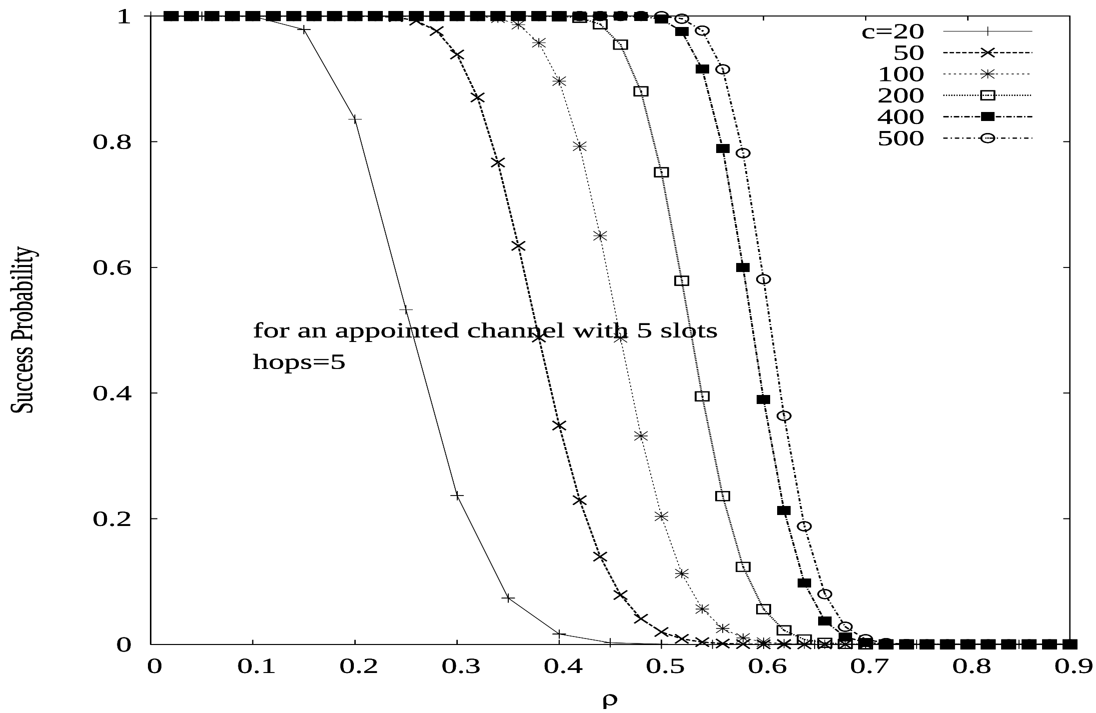

Figure 4.7: The Calling Success Probability for Setting up an Exact Synchronization Channel with 5 Slots, When Hops=5

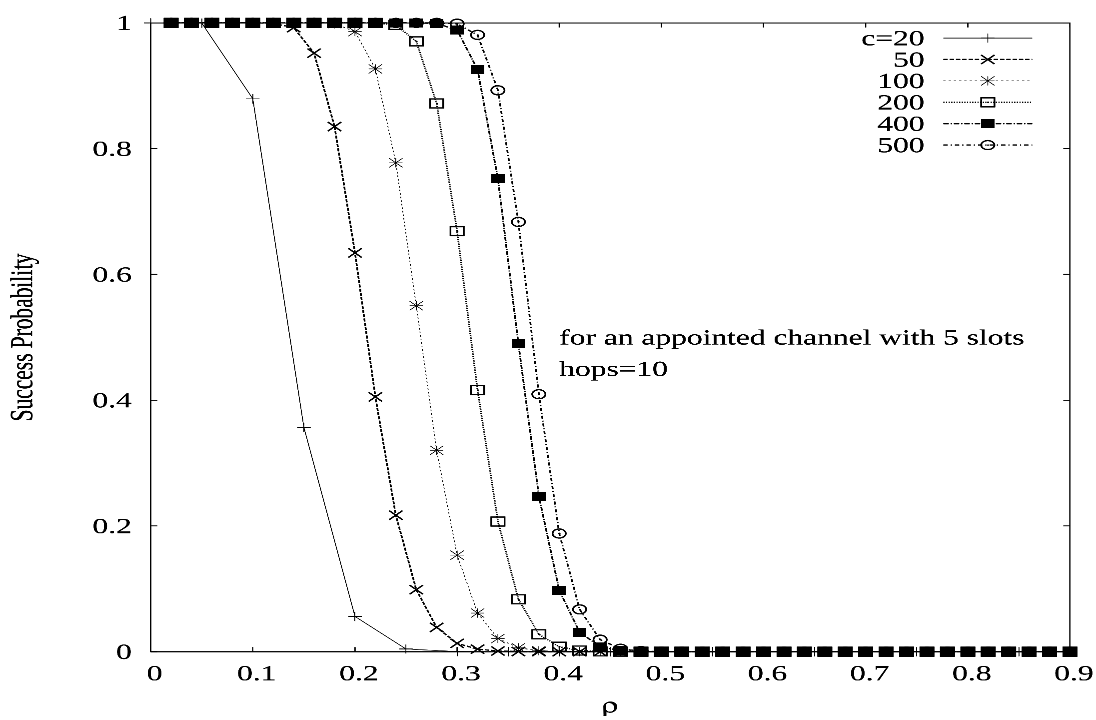

Figure 4.8: The Calling Success Probability for Setting up an Exact Synchronization Channel with 5 Slots, When Hops=10

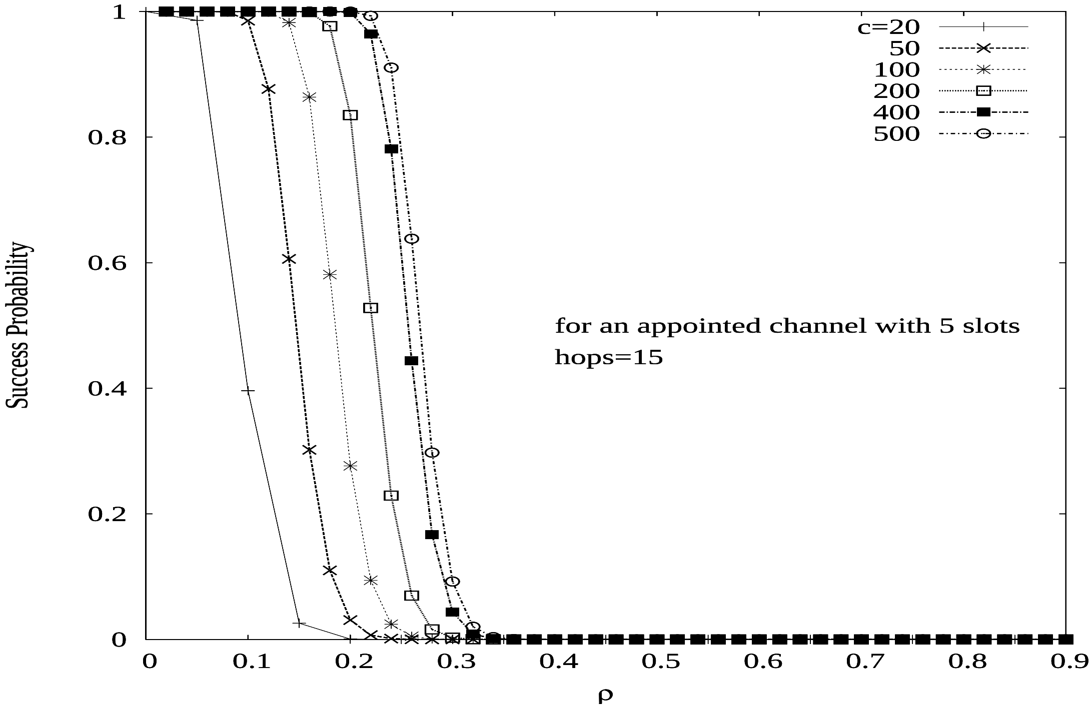

Figure 4.9: The Calling Success Probability for Setting up an Exact Synchronization Channel with 5 Slots, When Hops=15

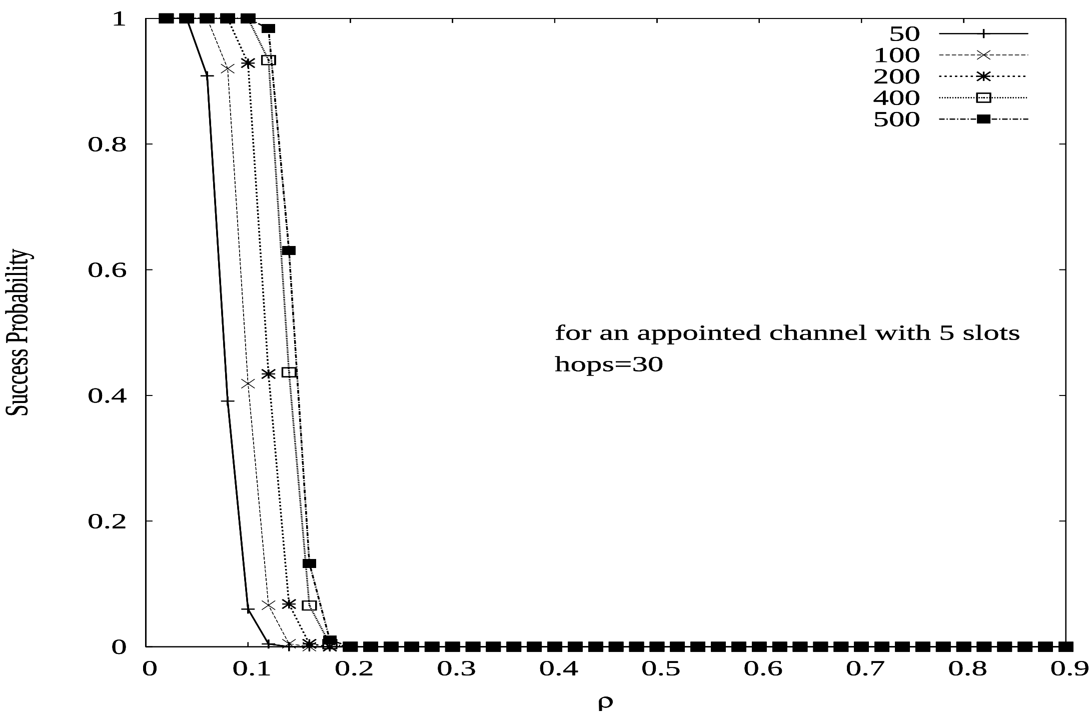

Figure 4.10: The Calling Success Probability for Setting up an Exact Synchronization Channel with 5 Slots, When Hops=30

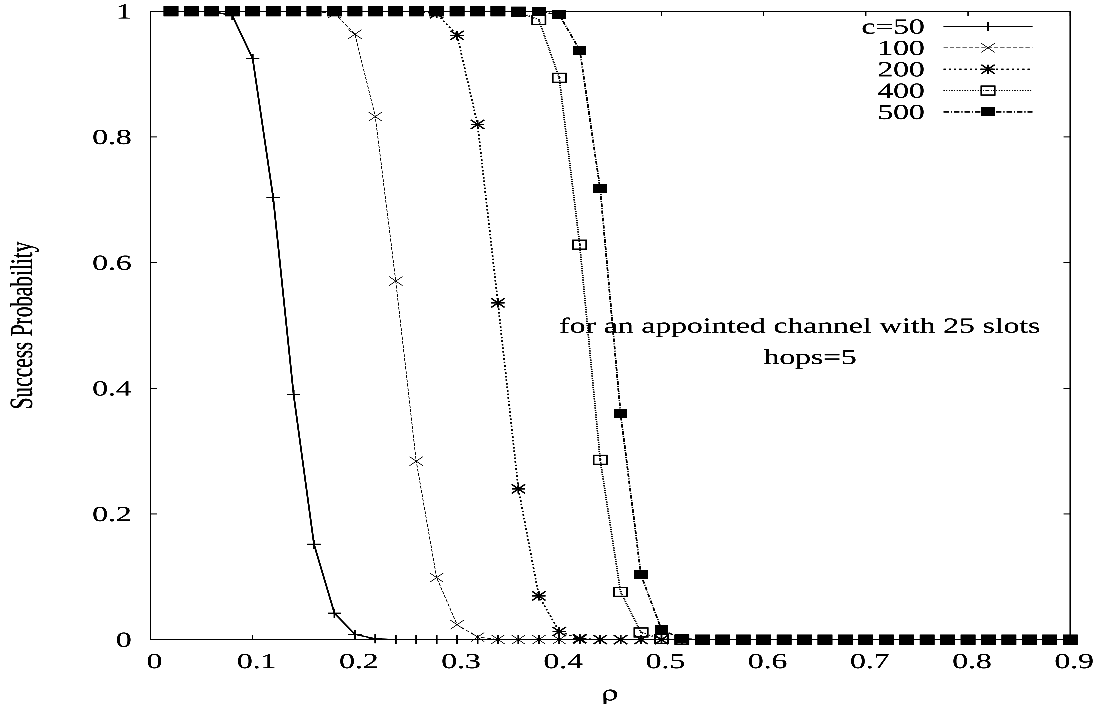

Figure 4.11: The Calling Success Probability for Setting up an Exact Synchronization Channel with 25 Slots, When Hops=5

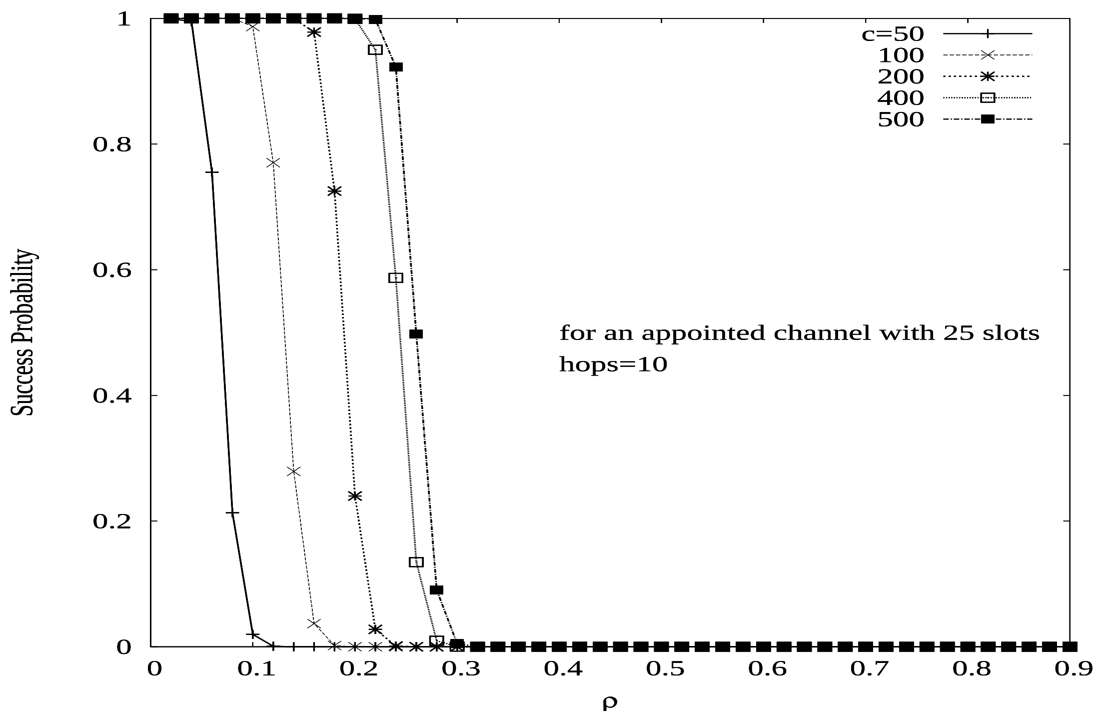

Figure 4.12: The Calling Success Probability for Setting up an Exact Synchronization Channel with 25 Slots, When Hops=10

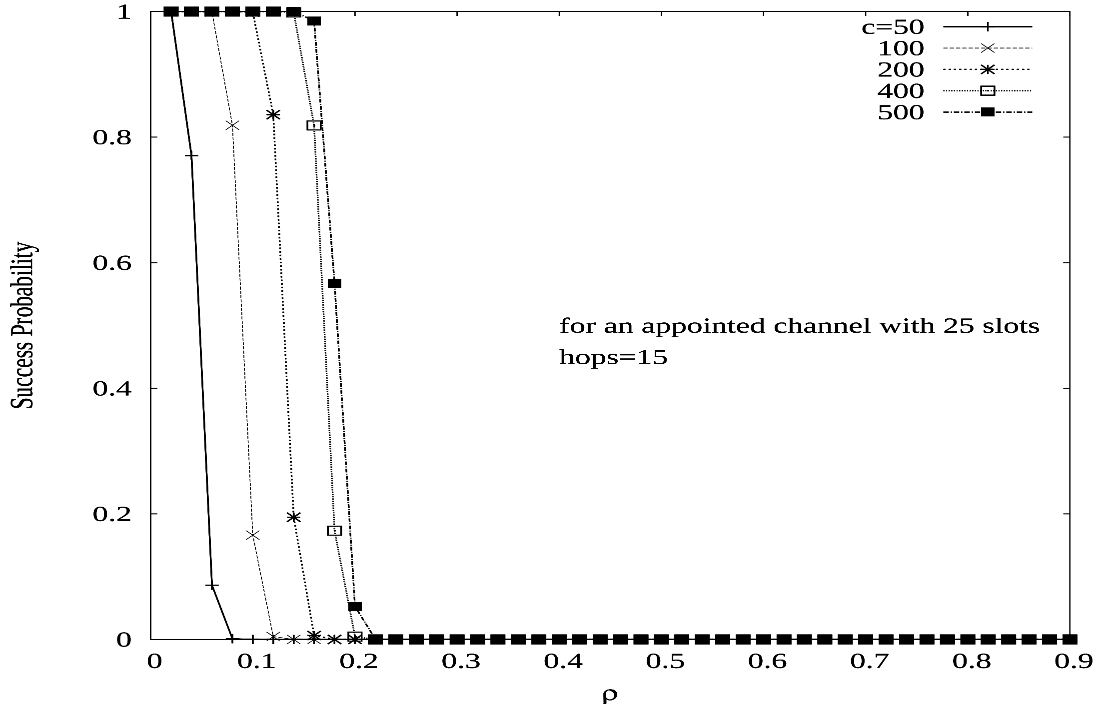

Figure 4.13: The Calling Success Probability for Setting up an Exact Synchronization Channel with 25 Slots, When Hops=15

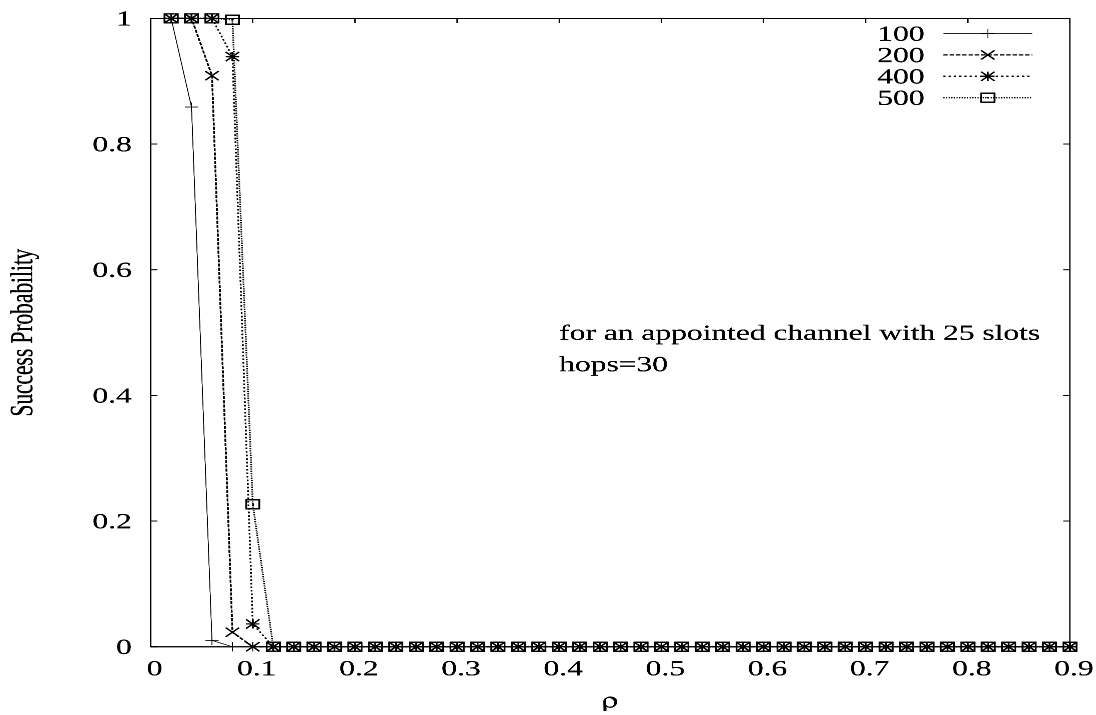

Figure 4.14: The Calling Success Probability for Setting up an Exact Synchronization Channel with 25 Slots, When Hops=30

The results show that: (1) there exists an area where the calling success probability changes very quickly with the increasing of the utilization. On the condition of low utilization, there is high calling success probability, when the utilization increases to a value, the calling success probability drops sharply. (2) the hops have a larger affect to the calling success probability. It means that calling success probability is remarkably different for different path even on the conditions of same utilization and setting up a new channel with same bandwidth. (3) the bandwidth of new channel also has some degree of impact to the calling success probability, comparing corresponding figures on same hops, it is clear that the channel bandwidth has larger impact on the small hops, and has smaller affect on the large hops. (4) calling success probability is sensitive with the offset range, c, when c is small; while it is insensitive when c is large, specially, when c is larger than 200.

As a whole, the exact synchronous channels is established easily on the conditions of only small hops, small bandwidth, low utilization and large offset range.

4.1.3 Less Strict Synchronization Channels

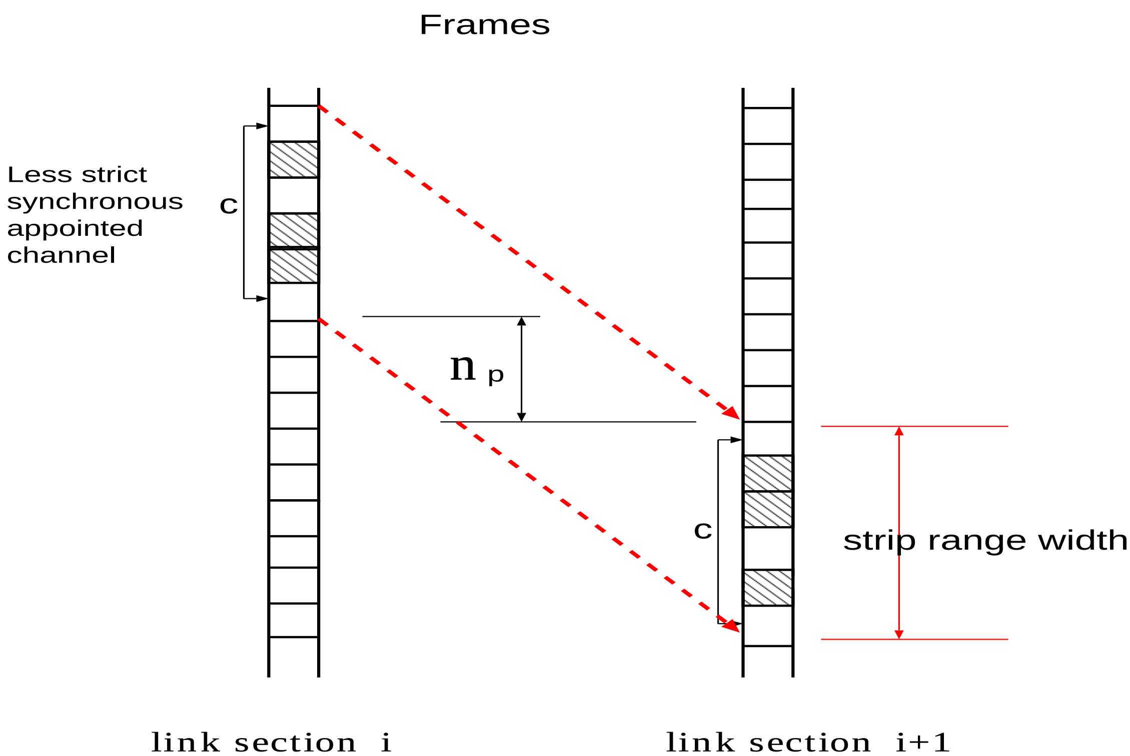

Figure 4.15: The Slots Positions Distribution of Less Strict Synchronous Appointed Channels

For less strict synchronous channels, the distribution of used slot positions within strip range, is variable in different link sections, it is shown in Figure 4.15, but the lag between tailer slot position in former section and header slot position in later section is constant, n_p, which indicates necessary minimum latency time measured at slot periods, it includes pipeline processing, reorder processing and so on, thus, the calling success probability can be calculated as following:

Analytic Solution

Because the offset ranges in different link sections are not overlapped, any slot positions consisting of a channel in former link sections can use any combination of slot positions within the offset range of later link sections, therefore, calling success probability in any one link section within the strip range is

$$p_1 = \sum_{i=x}^{c} {}_cC_i (1-\rho)^i \cdot \rho^{c-i} \tag{4.6}$$

where x is the required slot number by the new channel. If there is h hops along whole path, the channel calling success probability is

$$P_{succ} = p_1^h \tag{4.7}$$

Calculation Results

By using equation (4.7), the calculation results of the calling success probability on less strict synchronous model are shown in Figures 4.16 to 4.27. In these figures, the horizontal axis indicates utilization, and the vertical axis illustrates the calling success probability for setting up a new channel.

The curves indicate that: (1)there exists a narrow range where the changing of the calling success probability is very sharp with the increasing of the utilization. (2) the impact of the bandwidth of new channel to the calling success probability is very strong in the case of same range c. (3) the hops is insensitive, it hardly affects the calling success probability. (4) the calling success probability is very sensitive to c, when the utilization is about 0.9, the calling success probability is almost 100% for the new channel with 1 slot if $c > 50$, while for the new channel with 5 slots and 25 slots, c needs to be larger than 200 and 400, respectively. (5) in contrast to the results on exact synchronous model, the less strict synchronous model can obtain larger calling success probability on same conditions of utilization, hops and strip range c.

Furthermore, if implementing at most three times attemps on different non overlapped strip ranges when the first or the second times attempt is failed, the calling success probability can reach to

$$P_{succ,three} = P_{succ} + (1 - P_{succ})P_{succ} + (1 - P_{succ})^2 \cdot P_{succ} \tag{4.8}$$

where P_{succ} is success probability, $1 - P_{succ}$ is failure probability. Based on the analysis above, substituting P_{succ} in equation (4.7) into equation (4.8), we can obtain $P_{succ,three}$ values. Specially, when P_{succ} is not less than 0.8, then $P_{succ,three}$ can exceed $0.8 + 0.2 \times 0.8 + 0.2^2 \times 0.8 = 99.2\%$. Therefore, even on the worse case of hops=30, 25 slots channel and utilization being 0.8, if let maximum strip range, c, equal 200, the calling success probability almost reaches to 100%.

4.2 Percentage of Appointed Slots in Frames

In general, calling success probability of exact synchronous channel is smaller, the appointed channels in MPsLS network mainly exist at less strict synchronous model. In

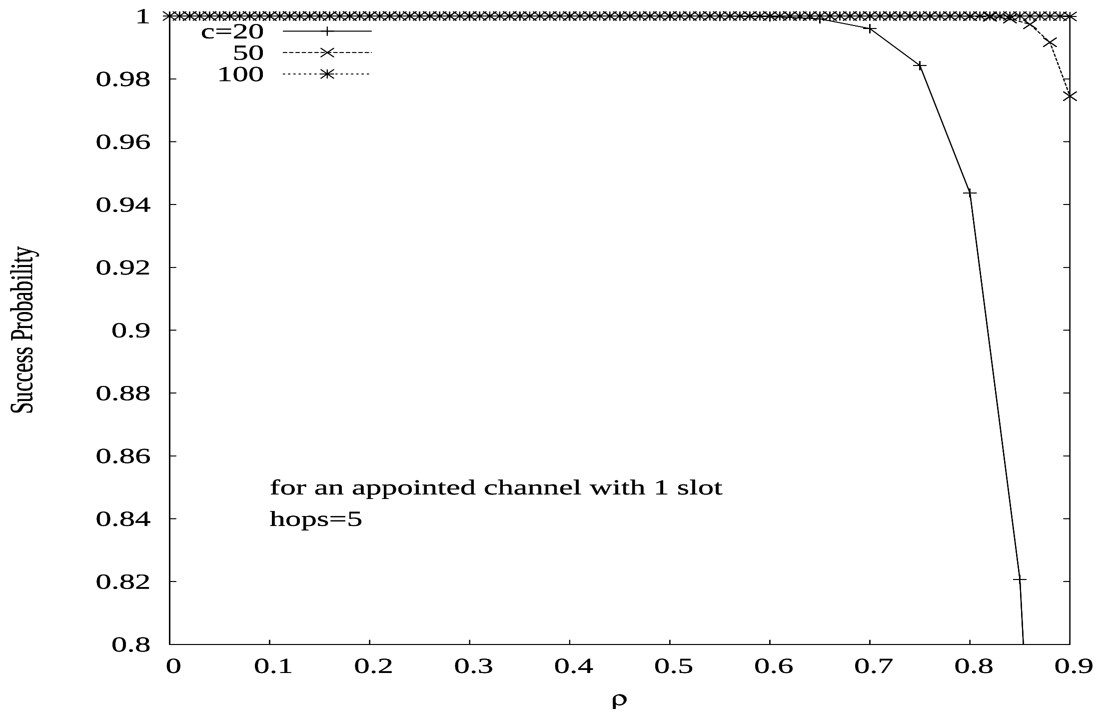

Figure 4.16: The Calling Success Probability for Setting up Less Strict Synchronous Appointed Channel with 1 Slot, When Hops=5

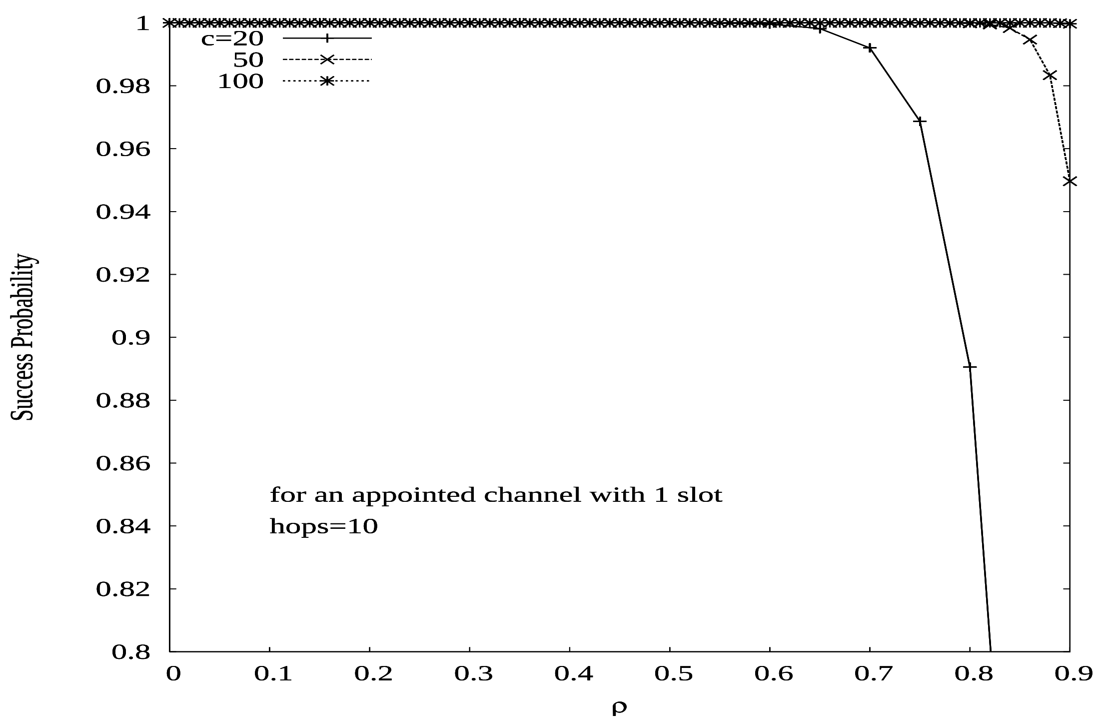

Figure 4.17: The Calling Success Probability for Setting up Less Strict Synchronous Appointed Channel with 1 Slot, When Hops=10

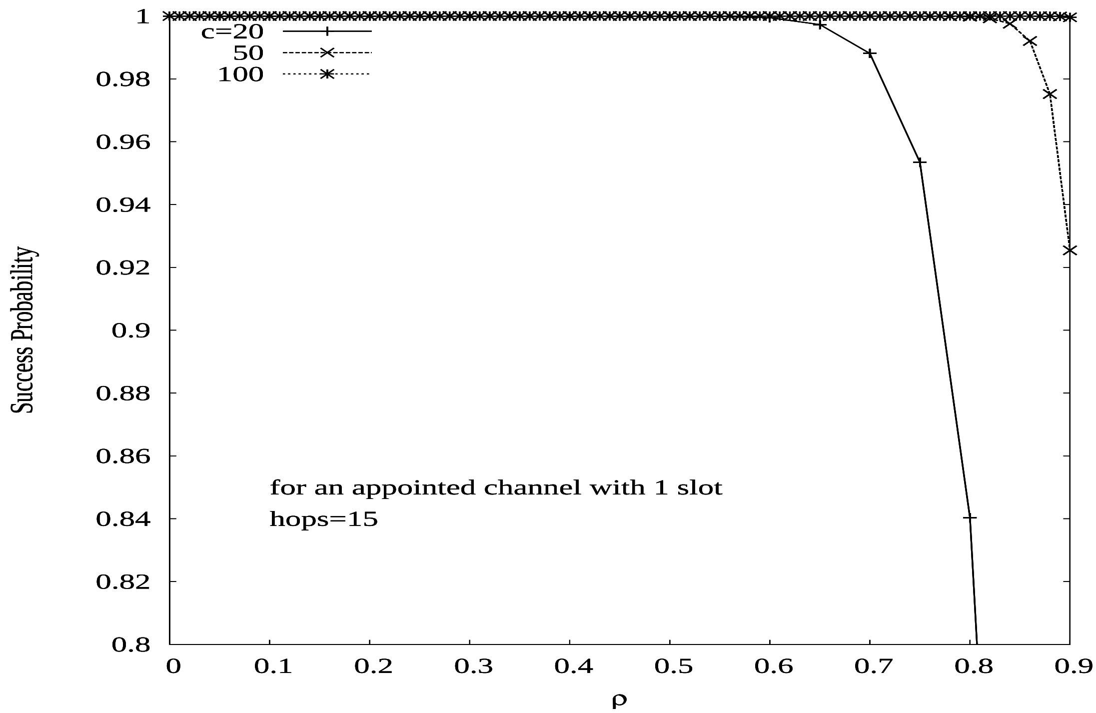

Figure 4.18: The Calling Success Probability for Setting up Less Strict Synchronous Appointed Channel with 1 Slot, When Hops=15

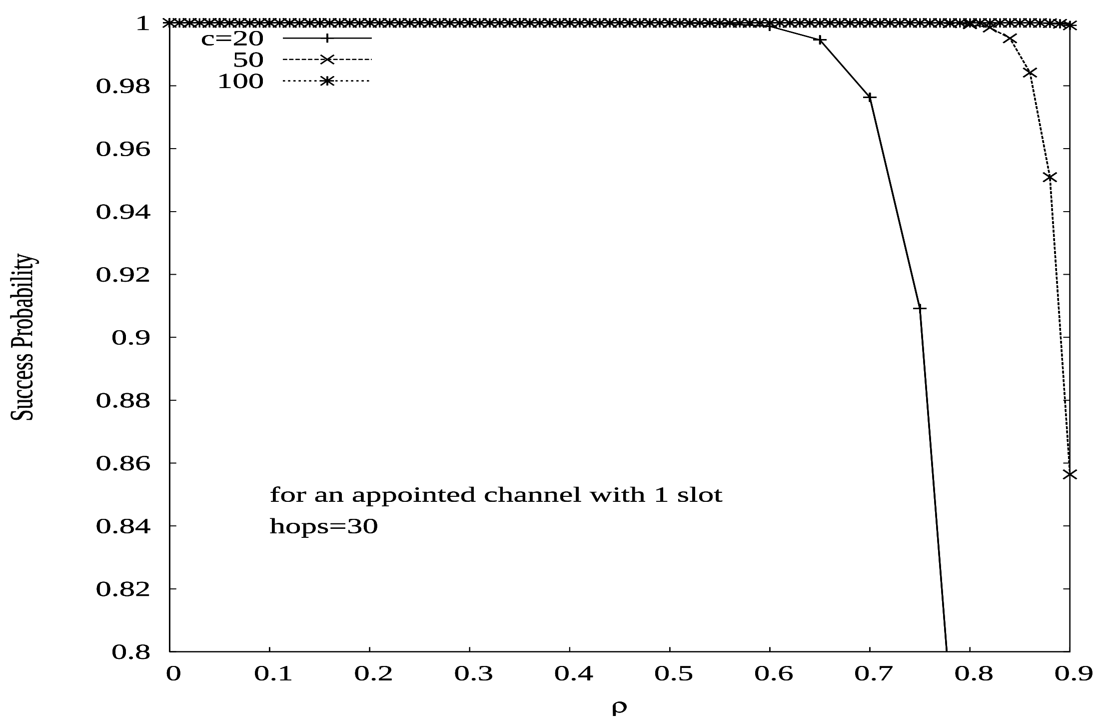

Figure 4.19: The Calling Success Probability for Setting up Less Strict Synchronous Appointed Channel with 1 Slot, When Hops=30

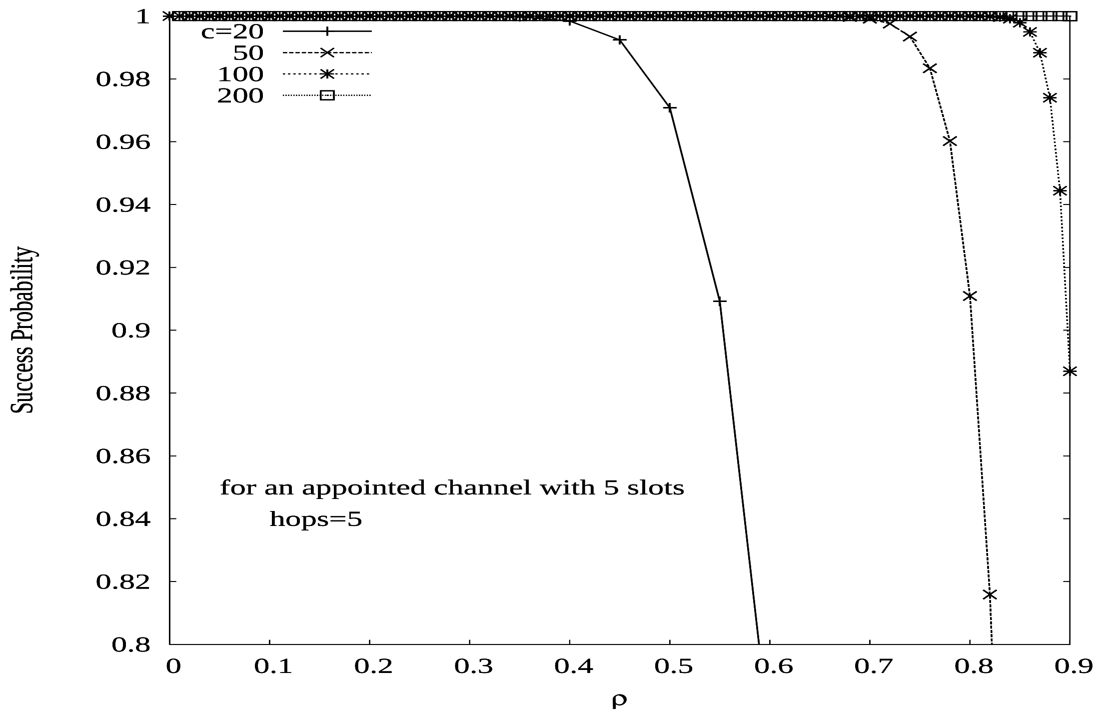

Figure 4.20: The Calling Success Probability for Setting up Less Strict Synchronous Appointed Channel with 5 Slots, When Hops=5

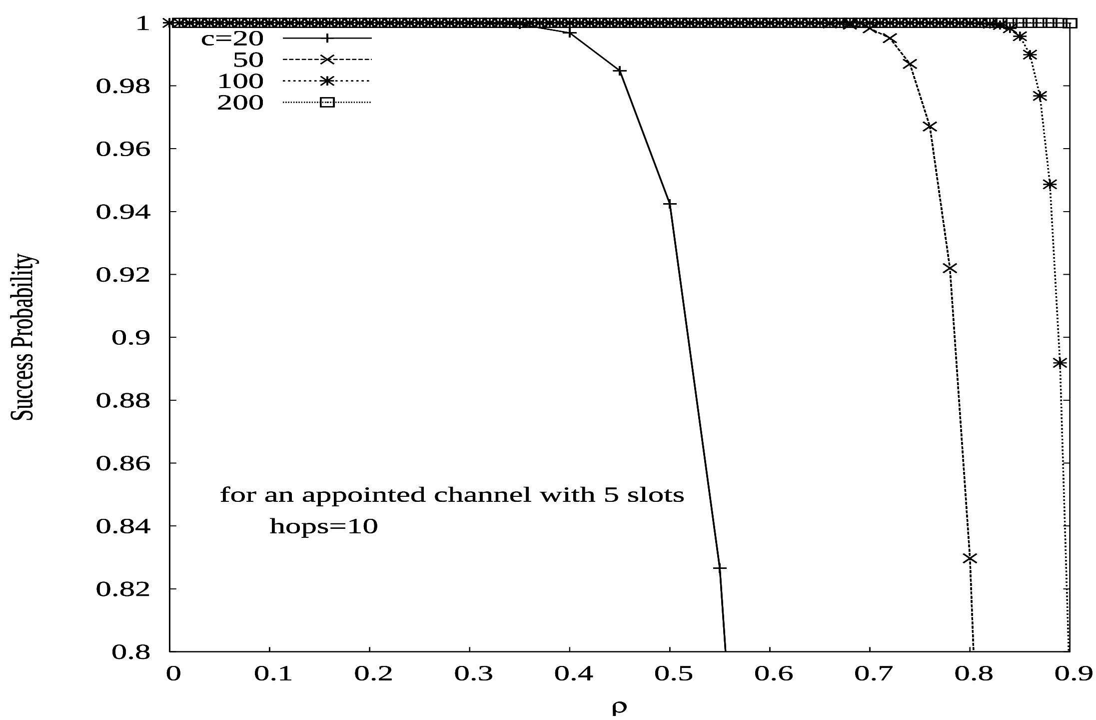

Figure 4.21: The Calling Success Probability for Setting up Less Strict Synchronous Appointed Channel with 5 Slots, When Hops=10

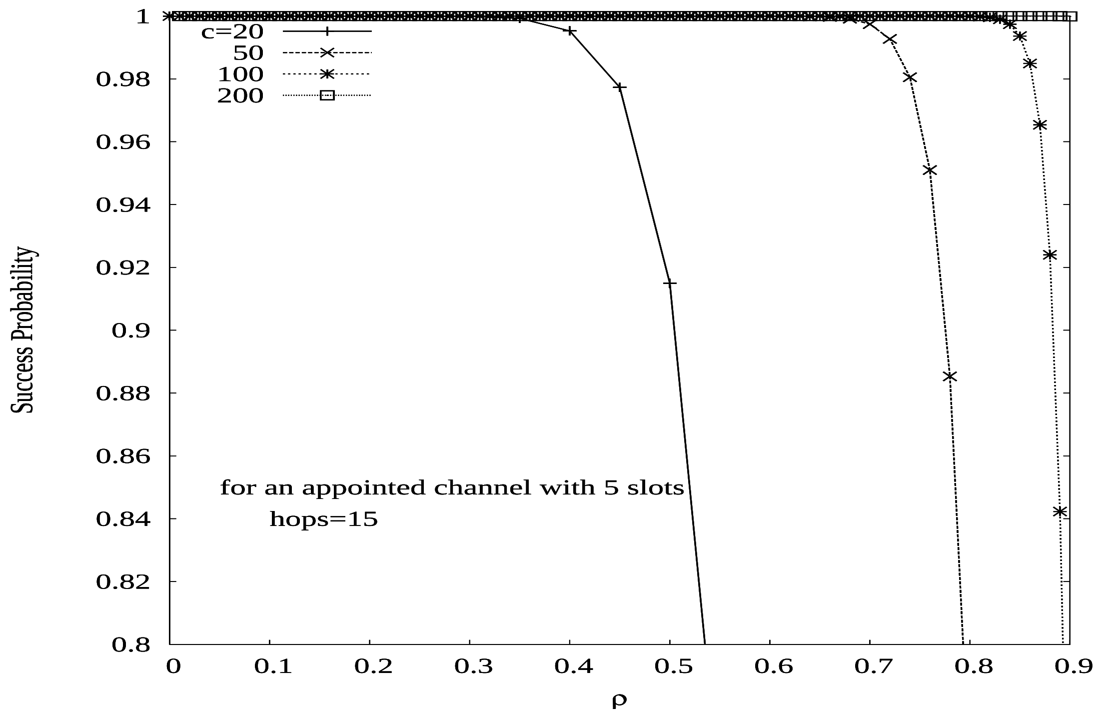

Figure 4.22: The Calling Success Probability for Setting up Less Strict Synchronous Appointed Channel with 5 Slots, When Hops=15

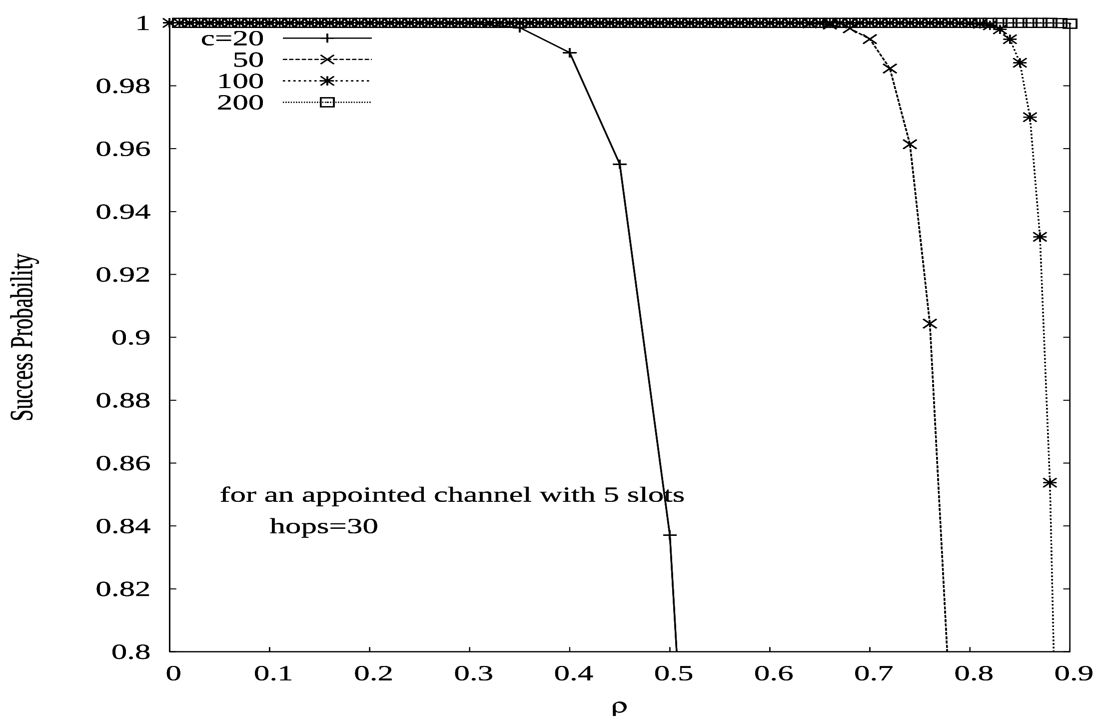

Figure 4.23: The Calling Success Probability for Setting up Less Strict Synchronous Appointed Channel with 5 Slots, When Hops=30

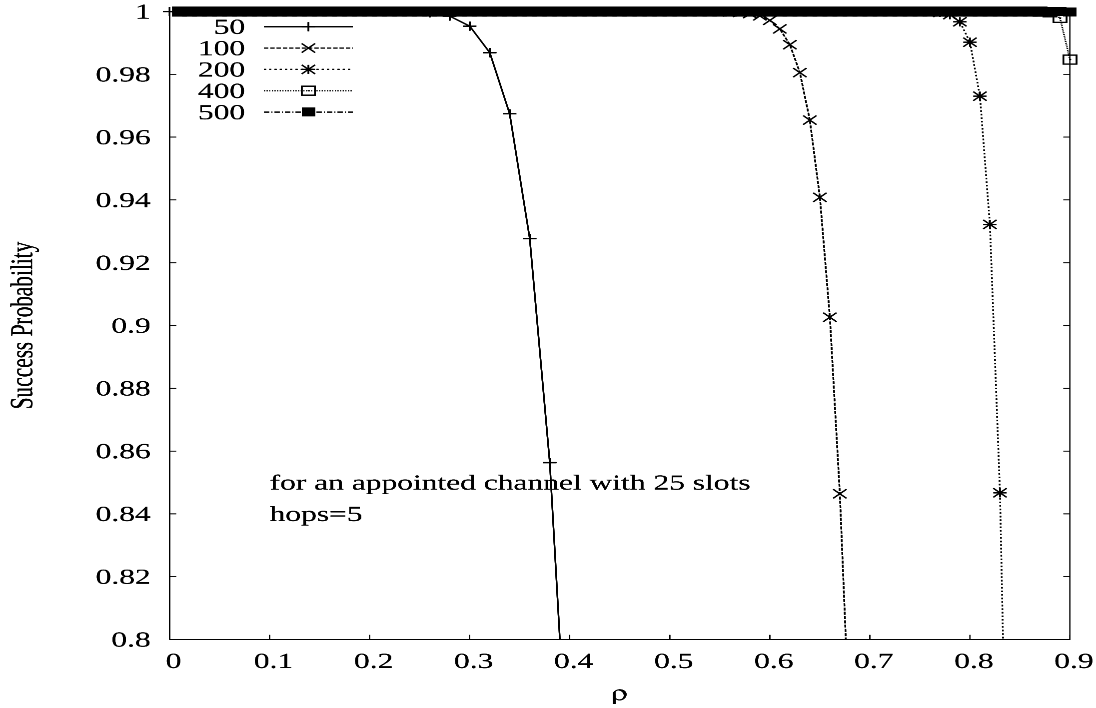

Figure 4.24: The Calling Success Probability for Setting up Less Strict Synchronous Appointed Channel with 25 Slots, When Hops=5

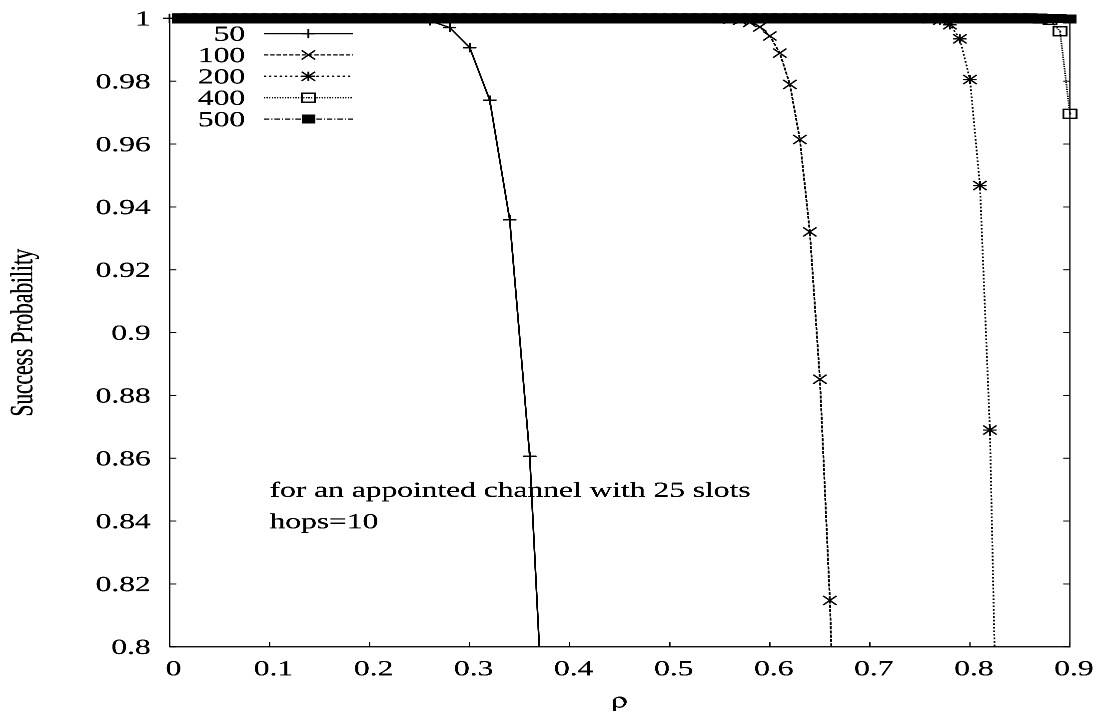

Figure 4.25: The Calling Success Probability for Setting up Less Strict Synchronous Appointed Channel with 25 Slots, When Hops=10

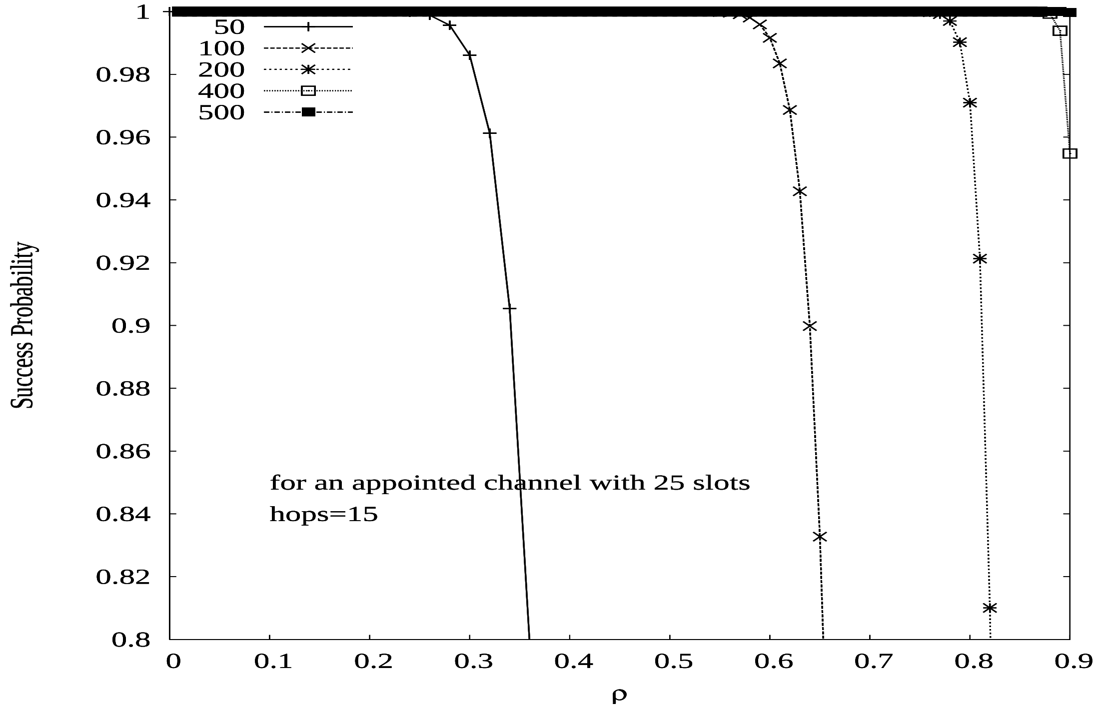

Figure 4.26: The Calling Success Probability for Setting up Less Strict Synchronous Appointed Channel with 25 Slots, When Hops=15

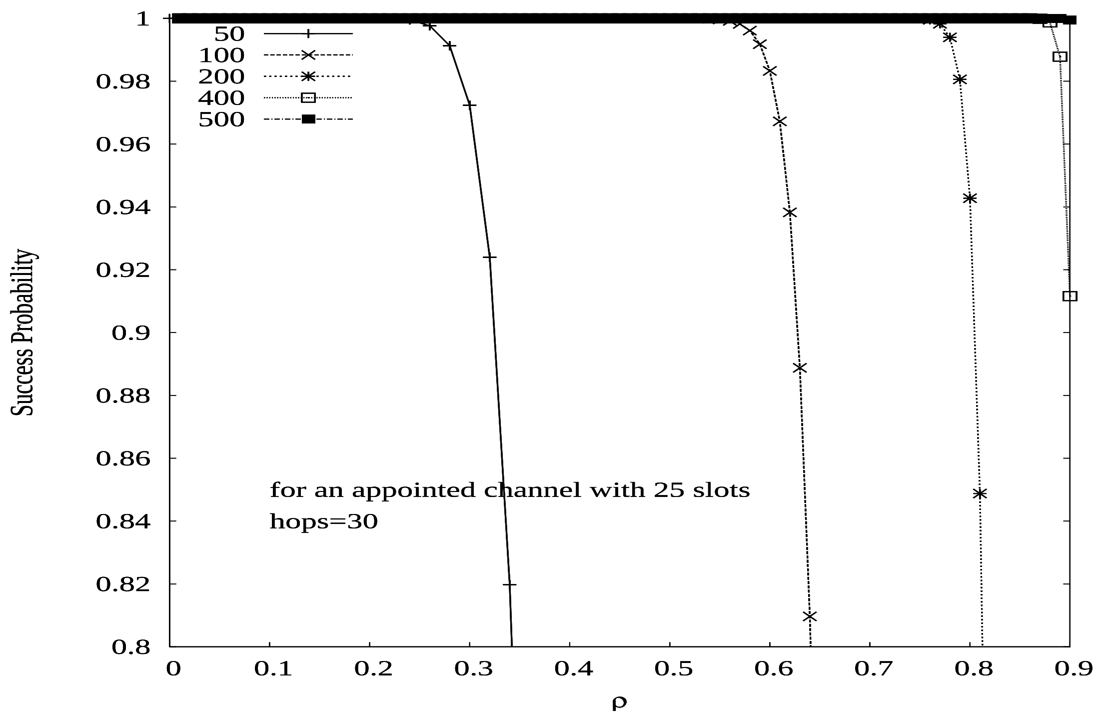

Figure 4.27: The Calling Success Probability for Setting up Less Strict Synchronous Appointed Channel with 25 Slots, When Hops=30

this chapter, we will analyze the capability of providing appointed channel service on less strict synchronous model.

4.2.1 Analytic Expression

Within a strip range with width of c between two edge MPsLS nodes, let the average number of the free slots which can be used to set up appointed channels be $\overline{N}_{app,\rho}$ when the utilization of each slot position equals ρ, thus the potential capability dynamically providing appointed channels utilizing residual bandwidth can be expressed by p_ρ,

$$p_\rho = \frac{\overline{N}_{app,\rho}}{c} \tag{4.9}$$

p_ρ illustrates the ratio of new appointed slots in residual bandwidth to total strip bandwidth.

Furthermore, the total ratio of appointed slots to total slots within strip range c satisfies

$$R_\rho = \rho + p_\rho = \frac{\overline{N}_{app,\rho} + \rho \cdot c}{c} \tag{4.10}$$

R_ρ indicates the capability of network providing appointed channel service, it includes two parts, having been used part of each slot position (the current utilization) and the potential part of each slot position (residual utilization).

Within a strip range c, the probability with j free slot positions is

$$p_{1,j} = {}_cC_j \cdot (1-\rho)^j \cdot \rho^{c-j} \tag{4.11}$$

the probability with more than j free slot positions is

$$p_{1,>j} = \begin{cases} \sum_{i=j+1}^{c} {}_cC_i \cdot (1-\rho)^i \cdot \rho^{c-i} & j < c \\ 0 & j = c \end{cases}$$

therefore, the probability with j free slot positions through whole connection path with h hops is

$$p_j = \begin{cases} \sum_{m=1}^{h} {}_hC_m \cdot p_{1,j}^m \cdot p_{1,>j}^{h-m} & j < c \\ p_{1,c}^h & j = c \end{cases}$$

furthermore, the average number of the free slots which can be used to set up appointed channels,

$$\overline{N}_{app,\rho} = \sum_{j=1}^{c} j \cdot p_j \tag{4.12}$$

Then, substituting equation (4.12) into (4.10), we obtain

$$R_\rho = \frac{\sum_{j=1}^{c} j \cdot p_j + \rho \cdot c}{c} \tag{4.13}$$

The function R_ρ exists a minimum, it means that the MPsLS network can ensure the ratio of appointed slots transmitting time-sensitive applications to total strip range being not less than this minimum.

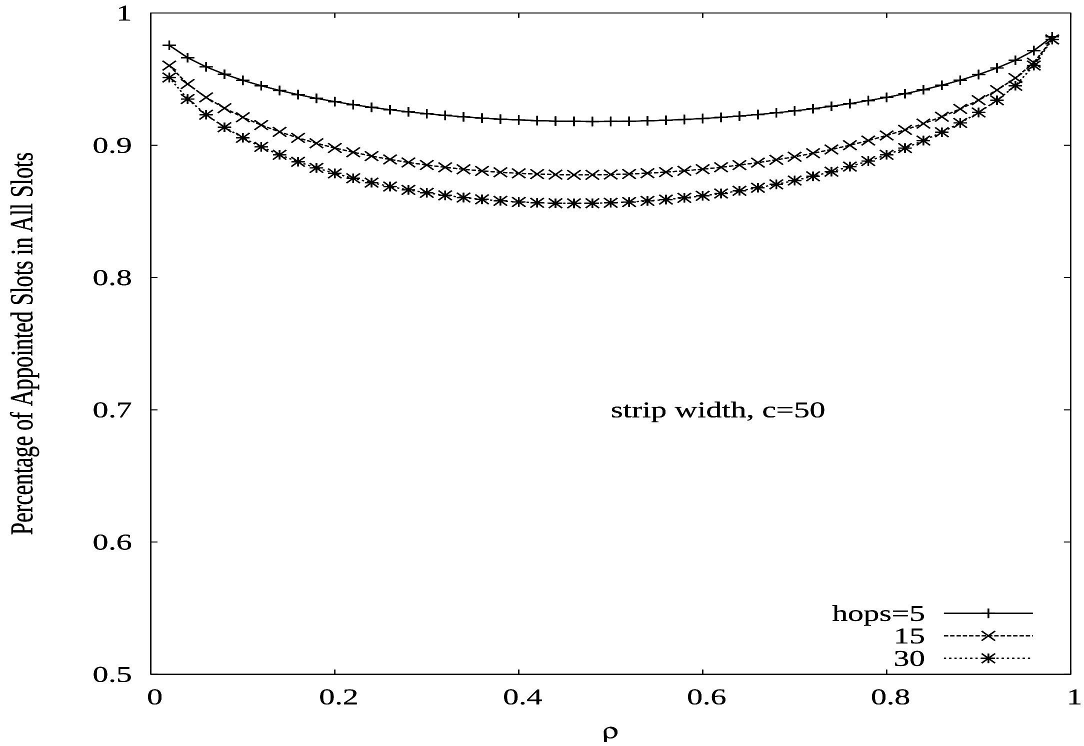

Figure 4.28: Percentage of Appointed Slots within Each Offset Range, when c=50

The minimum value can be obtained by solving following equation

$$R_{\rho}^{'} = 0 \tag{4.14}$$

or drawing the R_ρ curves, then find the minimum from the figure.

4.2.2 Calculation Results

To clearly describe the variation of R_ρ function, in this paper, we draw the R_ρ variable curves with the utilization on different $c = 50$ and 200, and hops=5,15 and 30.

The results of the percentage of appointed slots calculated by equation (4.13), are shown in Figures 4.28 and 4.29. In these figures, the horizontal axis indicates current utilization, and the vertical axis illustrates the ratio of appointed slots to all slots within the offset range. There is a minimum at each curve, on the left hand side of the minimum, the current utilization is small, but more slots can be found to establish new appointed channels in residual bandwidth; on the right hand side of the minimum, the current utilization is large, but less slots can be found to establish new appointed channels in residual bandwidth.

Besides, the relationship between the minimum of the ratio of appointed slots to total slots in strip range, c, is shown in Figure 4.30, the figure indicates that the minimum of the ratio decrease with the increasing of hops, and increase with the increasing of the range c. But there is sharp gradient when $c < 200$, which means that the value is sensitive with c in this area, but when $c > 200$, the value is insensitive with c.

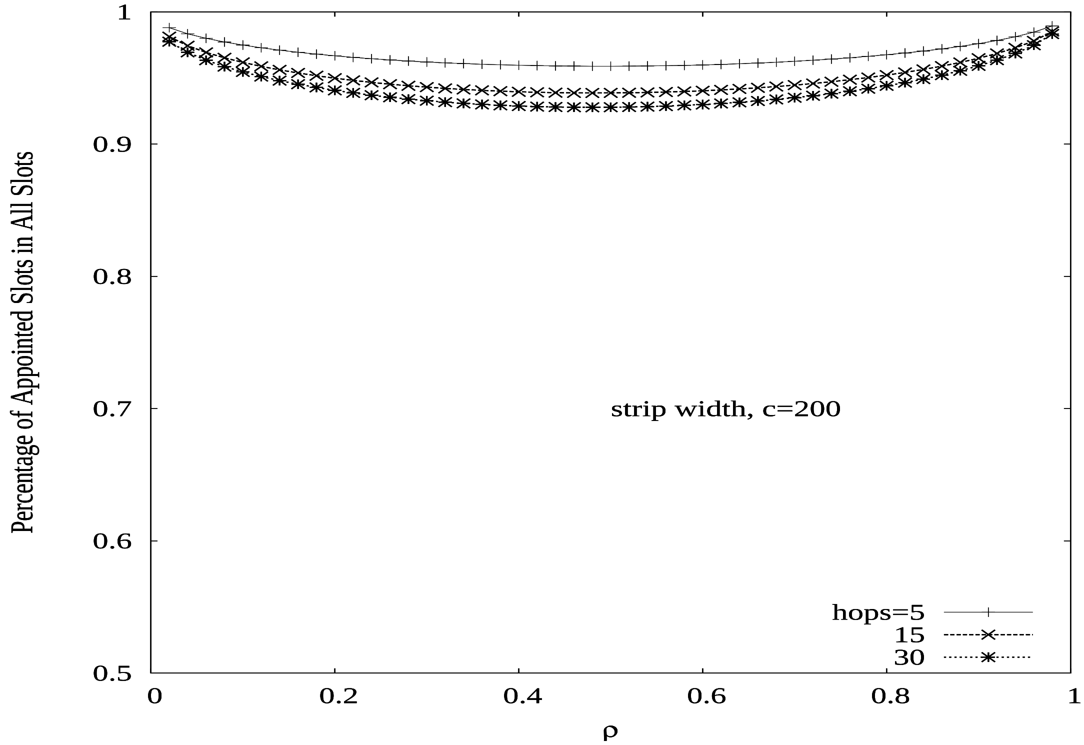

Figure 4.29: Percentage of Appointed Slots within Each Offset Range, when c=200

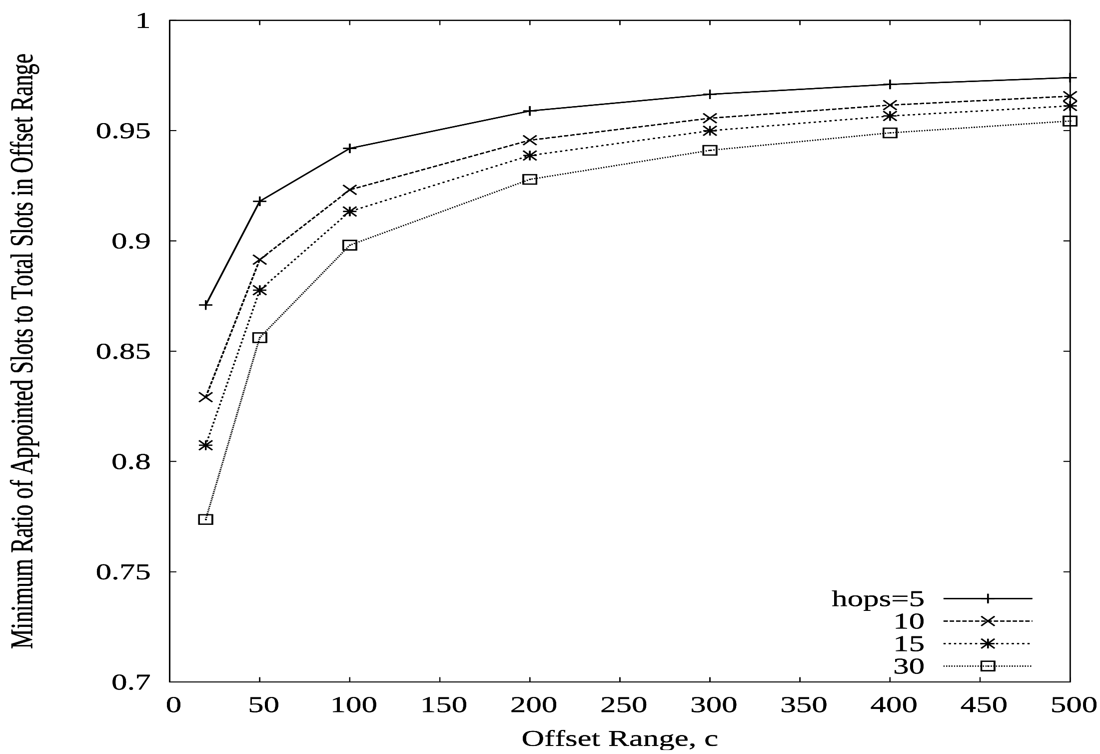

Figure 4.30: Relationship between Ratio of Appointed Slots to Total Slots within Offset Range and c

Usually, the value is larger than 80%, if $c \geq 20$, except on the worst case of $hops = 30$ and $c = 20$, the value is little smaller than 80%. If $c \geq 200$, the value exceeds 90%, therefore, we conservatively consider that each strip range can provide over 80% bandwidth to establishing appointed channel.

Because the total fiber bandwidth can divided into many non overlapped strip ranges with same width c, the conclusion in each strip range can extend to whole fiber bandwidth area.

4.3 Summary

In the first part of this chapter, we discuss the access performance of MPsLS network. For different parameters ρ, $hops$ and c, the calculation results illustrate that exact synchronous channels have small calling success probability, it is difficult to set up the channel through large number of hops; less stric synchronous channels have large calling success probability on even worst case. Specially, for the new channels with variable slots, 1, 5 and 25, even the utilization reaches to 0.9, the almost 100% calling success probability can be obtained if c is over 50, 200 and 400, respectively.

Anymore, for each calling request, if we implement three times attempts in non overlapped strip ranges, the dynamic calling success probability can be further improved. If only the success probability of each attemp reaches to 0.8, the total calling success probability after three times attempts can be near 100%, where it needs only $c = 200$.

In the second part of this chapter, analytical solution and calculation results show that within each strip range (also can extend to whole fiber bandwidth), MPsLS network can averagely provide over 80% slots to establish appointed channels, specially, if let $c = 200$, the percentage can exceed 90%.

Chapter 5

Delay Analysis

The delay is an important factor affecting the service quality, usually which is used to evaluate the network performance.

In MPsLS network, time-sensitive applications are transmitted on appointed channels, the channels are allocated and reserved to per-flow, the traffic from one application does not impact that from other applications, therefore, the QoS of each flow can be completely determined by the traffic state and network resource assigned to the flow, such as arrival process of the traffic, channel bandwidth and position distribution of the slots in the frame.

The transmission of the non time-sensitive applications is different with that of time-sensitive applications, all non time-sensitive application flows share large filler channel, the resource of the filler channel is not ensured, it is allowed to use any free slot positions and idle slot position on the appointed channels which are utilized by corresponding appointed slots at priority, thus, not only the traffic from different non time-sensitive application flows is impacted each other, but also is affected by the traffic state of time-sensitive applications.

In addtion, the delay tends to be separated into the constant parts(e.g. propagation delay) and the variable parts(e.g. queue delay) which depends on forwarding mechanism and traffic state. In this chapter, we carefully discuss the variable delay according to different tranfer modes, exact synchronous and less strict synchronous modes of appointed channels, and filler channels mode, respectively, which mainly arising at MPsLS nodes is caused by processing and queueing of application data .

The remainder of this chapter is organized as follows. Sections 5.1 and 5.2 analyze the variable delay of application traffic passing through core and edge nodes of MPsLS network, respectively. And Section 5.3 proposes an expression of estimating whole delay from edge-to-edge nodes. Then Section 5.4 summarizes this chapter.

5.1 Forwarding Delay at Core Nodes

At core nodes, variable delay is caused by buffering, processing, and reordering slots, we call it forwarding delay.

5.1.1 On Exact Synchronization Channel

Waiting Queue Assume that a core node has m input and m output ports, m input ports share every output port fairly, and vice verse. And each input port carries same amount of traffic, and traffic obeys same arrival distribution.

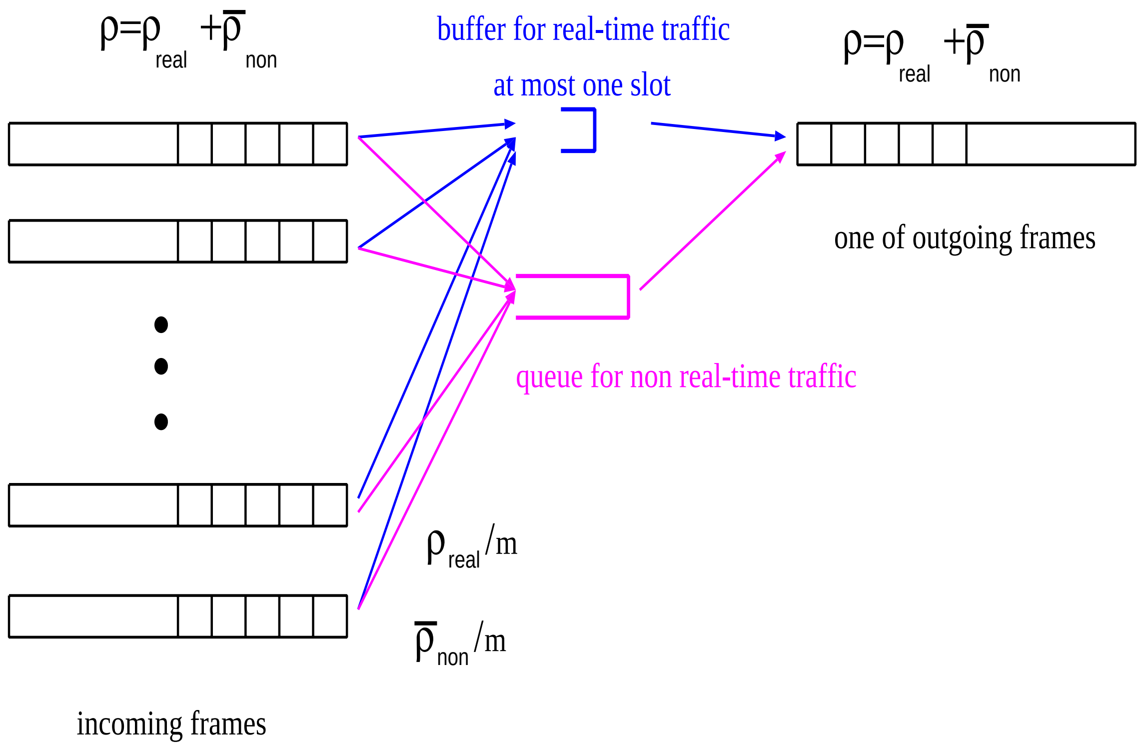

Figure 5.1: Forwarding Process of Slots at Core Nodes

The forwarding process of the slots at core nodes is illustrated in figure 5.1, if the total average utilization of time-sensitive traffic is ρ_{real} and that of non time-sensitive traffic is $\overline{\rho}_{non}$, but the average value is ρ_{non} when there is not time-sensitive traffic. The traffic of each input port contributing to any one output port is same, the values of the average bit rates are ρ_{real}/m and $\overline{\rho}_{non}/m$, respectively.

During a synchronization slot period, each time-sensitive traffic queue corresponding to an output port receives at most one appointed slot, while each non time-sensitive traffic queue receives at most m filler slots when there is not appointed slot arriving, or only atmost $m-1$ filler slots when there is an appointed slot arriving, and only one slot is sent out, therefore, the sum of both queue lengths satisfies,

$$L_{k+1} = L_k - B_k + A_k \tag{5.1}$$

where L_k (or L_{k+1}) is sum of length of both time-sensitive traffic and non time-sensitive traffic queues in kth (or $k+1$th) synchronization slot period.

A_k is the number of the slots that arrive in kth slot period, which includes time-sensitive and non time-sensitive traffic slots, the utilizations are ρ_{real} and ρ_{non}, respectively. B_k is a function denoting the number of slots sent out in kth slot period, it satisfies

$$B_k = \begin{cases} 1 & L_k \geq 1 \\ 0 & L_k = 0 \end{cases}$$

Let $L_k^* = L_k - B_k$, thus the sum of both queue lengths L consists of an imbedded Markov chain. L^* and A are independent of each other, the probability generating function(*pgf*) of L satisfies

$$L(z) = \frac{x_0 \cdot (z-1) \cdot A(z)}{z - A(z)} \tag{5.2}$$

where $A(z)$ is the *pgf* of arrival distribution A, x_0 is the probability when the sum of both queue length equals zero, and

$$x_0 = 1 - A'(z)|_{z=1}. \tag{5.3}$$

From equation (5.2) and (5.3), we obtain

$$\overline{L} = L'(z)|_{z=1} = [\frac{A''(z)}{2(1 - A'(z))} + A'(z)]|_{z=1}. \tag{5.4}$$

Since arrival distribution A follows conditional binomial distribution in a synchronization slot period, $A(z)$ satisfies

$$\begin{aligned} A(z) &= \sum_{i=0}^{m-1} P(A_{non} = i | A_{real} = 1) \rho_{real} z^{i+1} \\ &+ \sum_{i=0}^{m} P(A_{non} = i | A_{real} = 0)(1 - \rho_{real}) z^i \end{aligned} \tag{5.5}$$

where

$$\begin{aligned} &P(A_{non} = i | A_{real} = 1) \\ &\quad = {}_iC_{m-1} \cdot (\frac{\rho_{non}}{m})^i \cdot (1 - \frac{\rho_{non}}{m})^{m-1-i} \rho_{real} \end{aligned} \tag{5.6}$$

and

$$\begin{aligned} &P(A_{non} = i | A_{real} = 0) \\ &\quad = {}_iC_m \cdot (\frac{\rho_{non}}{m})^i \cdot (1 - \frac{\rho_{non}}{m})^{m-i} (1 - \rho_{real}). \end{aligned} \tag{5.7}$$

Substituting equations (5.6) and (5.7) into (5.5), then calculating the first and second order derivatives of (5.5), we obtain

$$A'(z)|_{z=1} = \rho_{real} + \rho_{non} - \frac{\rho_{real}\rho_{non}}{m} \tag{5.8}$$

and

$$\begin{aligned} &A''(z)|_{z=1} \\ &\quad = \frac{(m-1)\rho_{non}(2\rho_{real} + \rho_{non} - \frac{2\rho_{real}\rho_{non}}{m})}{m}. \end{aligned} \tag{5.9}$$

Substituting equations (5.8) and (5.9) into (5.4), we have

$$\begin{aligned}\overline{L} &= \rho_{real} + \frac{\rho_{non}(1-\rho_{real})}{m} \\ &+ \frac{(1-\frac{\rho_{non}}{2})(1-\frac{1}{m})\rho_{non}}{1-\rho_{real}-\rho_{non}+\frac{\rho_{real}\rho_{non}}{m}}.\end{aligned} \tag{5.10}$$

The first term of equation (5.10), ρ_{real}, equals the average queue length of time-sensitive traffic, L_{real},

$$L_{real} = \rho_{real} \tag{5.11}$$

and the rest is the average queue length of non time-sensitive traffic, L_{non}, therefore,

$$\begin{aligned}L_{non} &= \frac{\rho_{non}(1-\rho_{real})}{m} \\ &+ \frac{(1-\frac{\rho_{non}}{2})(1-\frac{1}{m})\rho_{non}}{1-\rho_{real}-\rho_{non}+\frac{\rho_{real}\rho_{non}}{m}}\end{aligned} \tag{5.12}$$

Mean Waiting Delay Since there is only at most one slot stored in a queue for the time-sensitive traffic and its average length, L_{real}, equals ρ_{real}, the number of slot periods waiting in the queue is,

$$W_{real} = 1 \qquad\qquad (or\frac{L_{real}}{\rho_{real}}) \tag{5.13}$$

For non time-sensitive traffic, since average arriving rate is $\overline{\rho}_{non} = \rho_{non}(1-\rho_{real}/m)$ according to equation (5.8), and equation (5.12), we can obtain the number of slot periods waiting in queue for each connection,

$$\begin{aligned}W_{non} &= \frac{L_{non}}{\overline{\rho}_{non}} \\ &= \frac{(1-\rho_{real})}{m-\rho_{real}} \\ &+ \frac{(1-\frac{\rho_{non}}{2})(m-1)}{(1-\rho_{real}-\rho_{non}+\frac{\rho_{real}\rho_{non}}{m})(m-\rho_{real})}\end{aligned} \tag{5.14}$$

Simulation Results The calculation results from equations (5.11), (5.12), and (5.14), and the simulation results are shown in Figures 5.2 and 5.3 on the conditions of total utilization, $\rho = \rho_{real} + \overline{\rho}_{non} = 0.9$, and number of input ports, $m = 20$.

Figure 5.2 shows the mean queue length of time-sensitive and non time-sensitive traffic. The curve indicates that the queue length of non time-sensitive traffic decreases rapidly as the utilization of time-sensitive traffic increases, while that of time-sensitive traffic is proportional to the utilization of time-sensitive traffic.

Figure 5.3 indicates the mean waiting time in queue for non time-sensitive traffic.

We also calculated the results of mean queue length and mean waiting time for both time-sensitive and non time-sensitive traffic when $\rho = \rho_{real} + \overline{\rho}_{non} = 0.5$, they are shown in

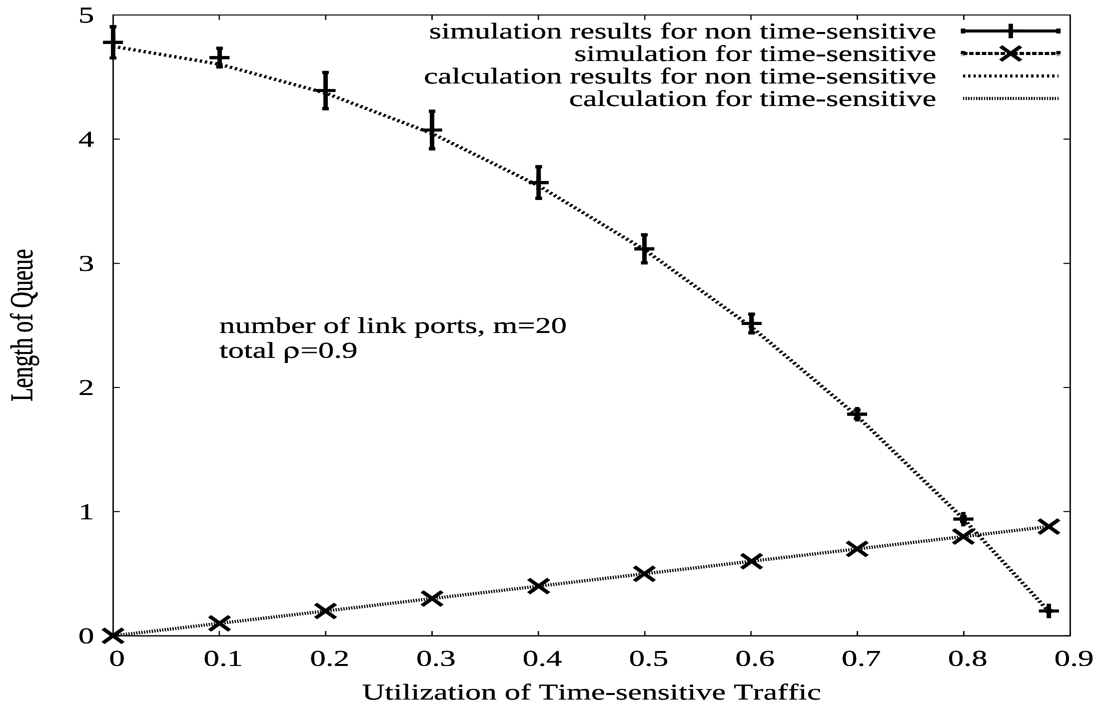

Figure 5.2: Mean Queue Length to Utilization (in Exact Synchronization Model)

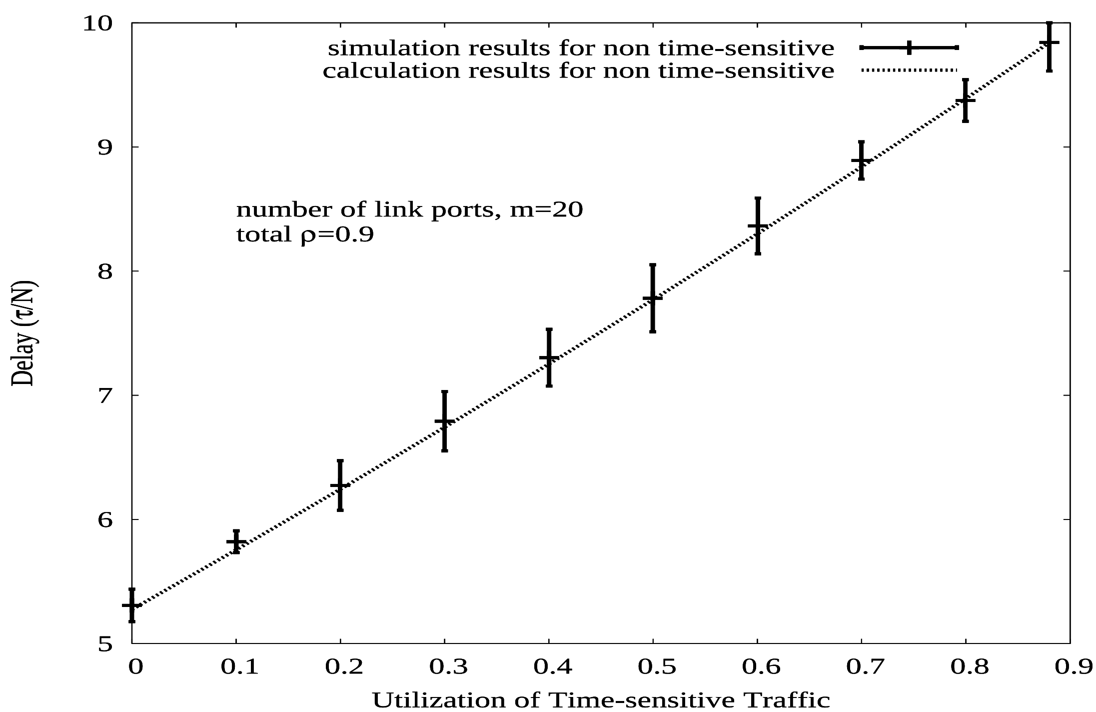

Figure 5.3: Mean Waiting Delay to Utilization (in Exact Synchronization Model)

figures 5.4 and 5.5, the results illustrate that the variation of the queue length and delay of non time-sensitive traffic with the time-sensitive traffic utilization is almost linear. Comparing to corresponding values in Figures 5.2 and 5.3 when $\rho = 0.9$, the queue length and delay remarkably decrease with the drop of total utilization ρ in network.

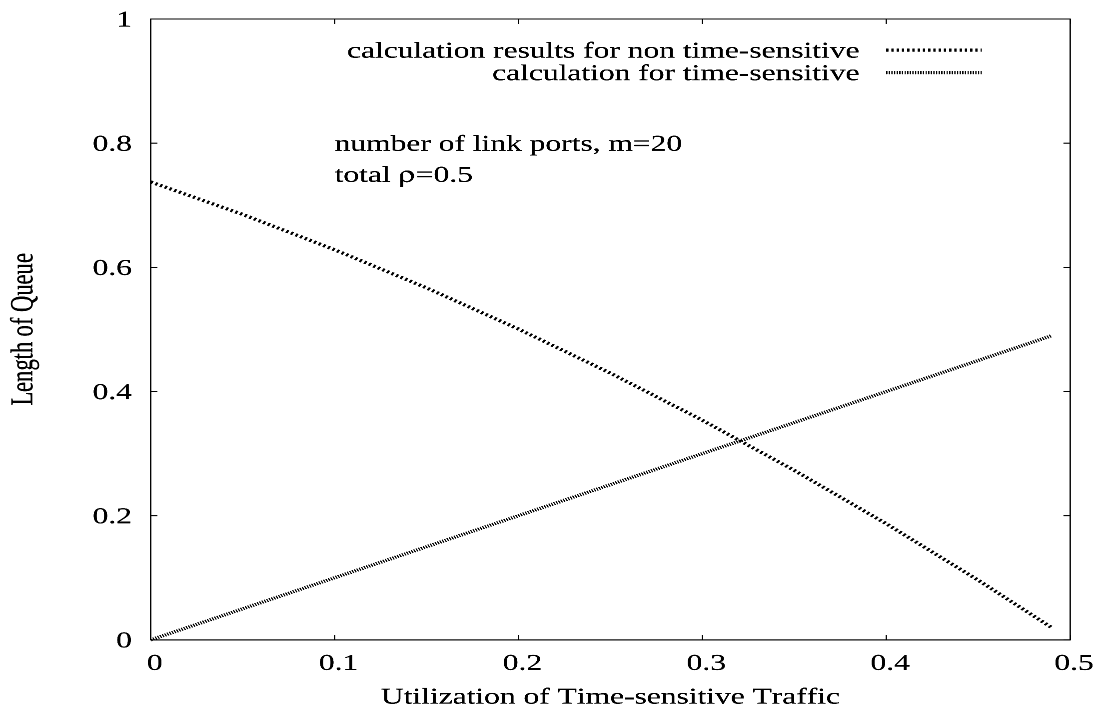

Figure 5.4: Mean Queue Length to Utilization (in Exact Synchronization Model)

Besides, from Figures 5.2 to 5.5, we can see that, although the queue length of non time-sensitive traffic decreases as the utilization of time-sensitive traffic increases, the mean waiting delay grows significantly. This is because time-sensitive traffic is transferred at priority to non time-sensitive traffic.

The long waiting delay, however, is acceptable for non time-sensitive applications.

Impact of m We also investigate the influences of m and time-sensitive traffic to mean delay of non time-sensitive traffic, when $\rho = 0.9$. The variation of the mean delay of non time-sensitive traffic with m is shown in Figure 5.6. The results indicate that on same time-sensitive traffic, the mean delay of non time-sensitive traffic increases with the increasing of m, but when $m > 5$, the variation is very slow.

5.1.2 On Less Strict Synchronization Channel

In the less strict synchronization model the arrived appointed slots need to sojourn longer time in reorder buffer than that in the exact synchronization model.

Because the channels are reserved exclusively and there is no contention for the same output slot. Therefore, the average sojourn time equals

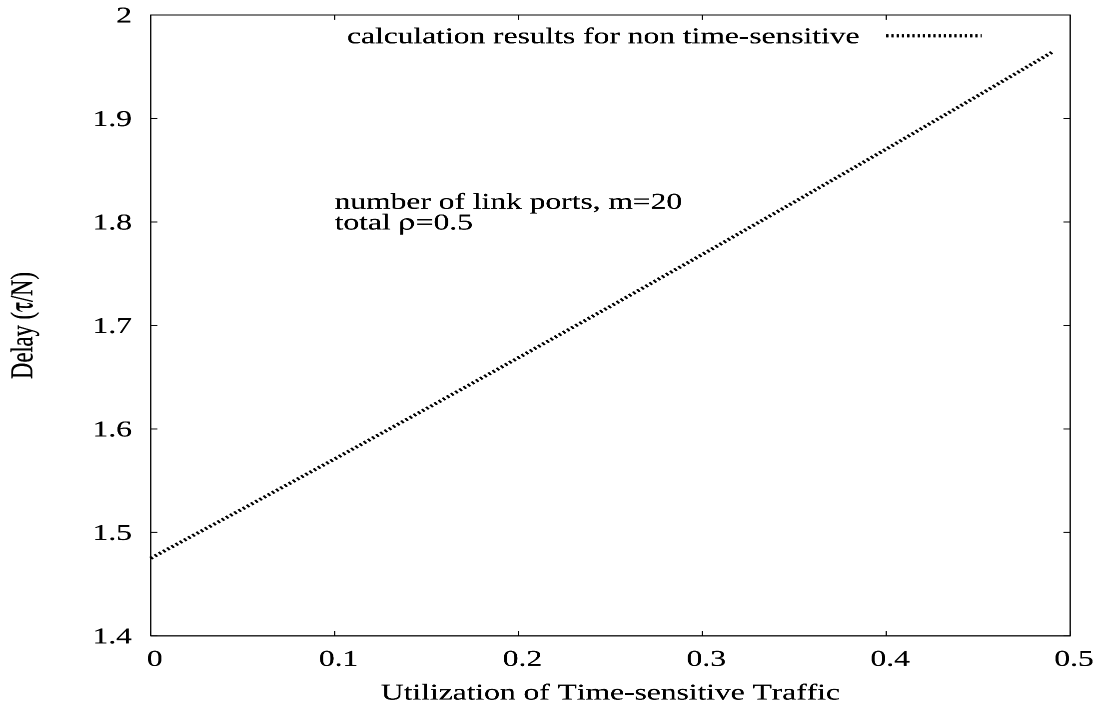

Figure 5.5: Mean Waiting Delay to Utilization (in Exact Synchronization Model)

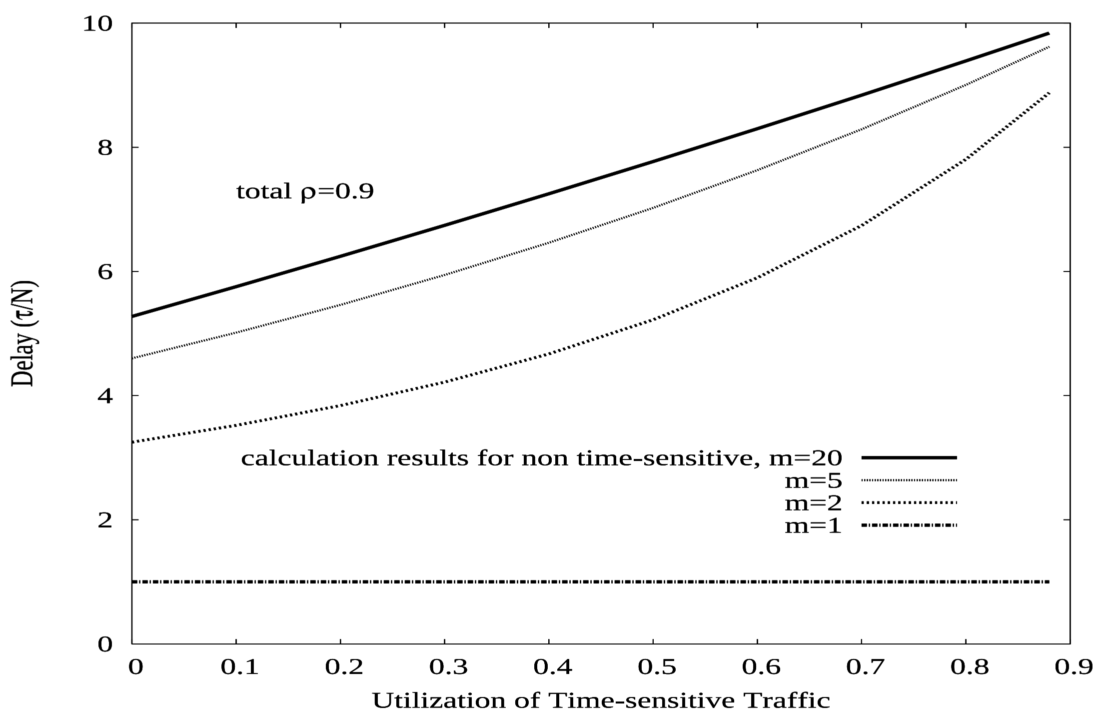

Figure 5.6: The Impact of m to Mean Waiting Delay of Non Time-sensitive Traffic

$$W'_{real} = c \tag{5.15}$$

where c is the width of the strip range. Therefore, average sojourn time is proportional to the offset range c. it can fall in allowable range to assure required QoS by controlling c.

We also have investigated the influence of the offset range of appointed channel to waiting delay of non time-sensitive application flows.

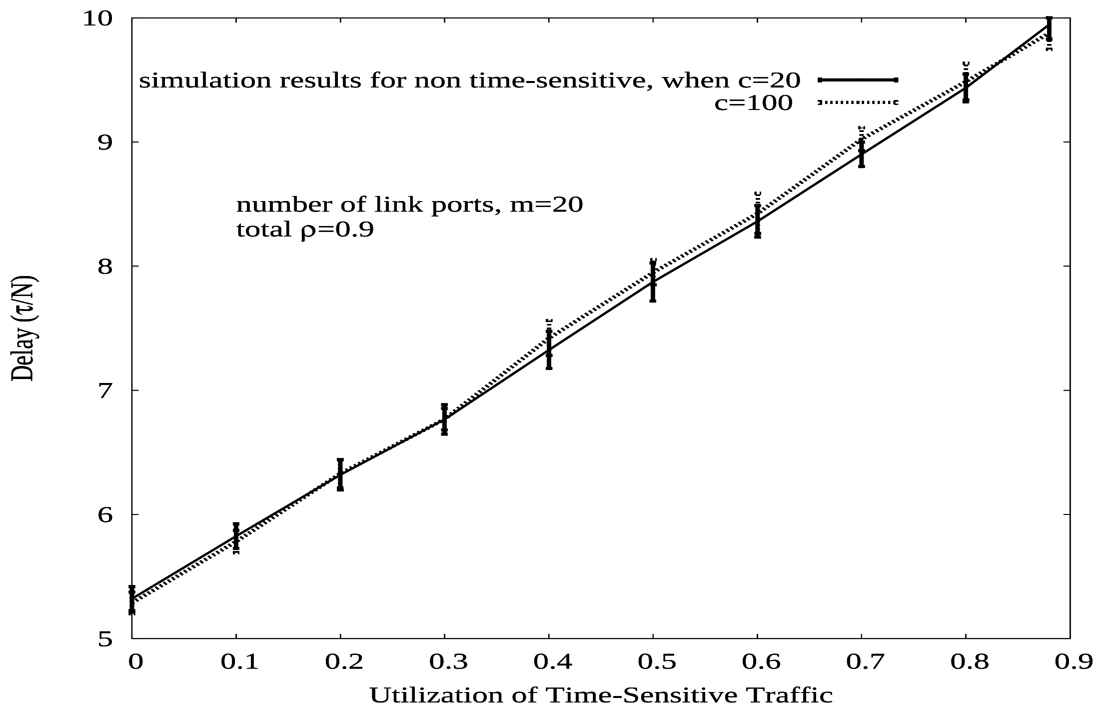

Figure 5.7: Mean Waiting Delay to Utilization (in Less Strict Synchronization Model)

Figure 5.7 exhibits the simulation results of the mean waiting delay for non time-sensitive traffic when c=20 and 100, and m=20, comparing to the cures in Figures 5.7 and 5.3, which mean not only the types of the appointed channel, but also size of the offset range have hardly influence to mean delay of non time-sensitive traffic when m is large. The reason is that although the existance of c can cause little bulk effect comparing to exact synchronous model(in less strict synchronous model, for example, a case may take place in special slot periods, at that time, maximum slot number which needs to be processed exceeds m, reach to m+1, namely, m filler slots arriving plus a waiting appointed slot), but the inflence becomes weak with the increase of m. Thus, mean delay of non time-sensitive traffic when there are less strict synchronous channels can be approximated by the value when appointed channels are all exact synchronous when m is large.

5.1.3 General Expression of Total Forwarding Delay for Different Types of Applications

From the analysis and simulation results in sections 5.1.1 and 5.1.2, time-sensitive traffic transmitted in the less strict synchronization model hardly causes additional delay for non time-sensitive traffic compared to that when all time-sensitive traffic is transmitted in the exact synchronization model.

Therefore, for any application flow j, the total switching delay at each core node is represented by

$$D_{core,j} = W_j \cdot \frac{\tau}{N} + p \cdot \frac{\tau}{N} + \frac{\tau}{2N}. \tag{5.16}$$

The first term is the waiting delay obtained in previous subsections, the second term is processing delay(including pipeline, buffer), where p is the sum of both number of pipeline stages and buffers, and the last term means synchronization delay for slot level synchronization.

For time-sensitive traffic in the exact synchronization model, $W_j = W_{real} = 1$; in the less strict synchronization model, $W_j = W'_{real} = c$. For both models, p equals the sum of both number of pipeline stages and buffers processing appointed slots, p_{app}.

For non time-sensitive traffic on filler channels, p equals the sum of both number of pipeline stages and buffers processing filler slots, p_{filler}, and $W_j = W_{non}$, which can be approximately calculated by equation (5.14) on the case of appointed channels being less strict synchronous.

5.2 Mean Delay at Ingress Nodes

At the ingress edge nodes, firstly long packets are split to data segments with slot size, then are filled in corresponding positions of a large frame. Therefore, the entering delay at ingress edge nodes includes splitting packets, assembling, and sending out frames.

Due to implementing frame level cycle at edge nodes, assume that the splitting delay of packets, and assembling delay of the frame is very short comparing to frame periods, therefore, we need only consider the waiting delay in the queues. For cycling transmission, there is additional waiting queue delay for both time-sensitive and non time-sensitive traffic, because although the appointed channels are reserved for time-sensitive traffic, stochasticly arrived slots needs to wait own slot positions, while, similarly, filler slots also need to wait free slot or idle appointed slot positions.

5.2.1 Mean Entering Delay

For simplicity, we assume that the arrival process of the slots is Poisson arrival, and all packet have same size which equals the size of a slot. If the arrival process of jth application flow is A_j. In any two neighboring discrete frame periods, k and $k+1$, the following equation is satisfied for jth flow,

$$L_j^{k+1} = [L_j^k - n_j]^+ + A_j^k \tag{5.17}$$

where,

$$[L_j^k - n_j]^+ = \begin{cases} L_j^k - n_j & L_j^k \geq n_j \\ 0 & L_j^k < n_j \end{cases}$$

the queue length, L_j, consists of an imbedded Markov chain. For this type of group service queue, Bailey gave the following solution based on the probability generating function (*pgf*) in [99],

$$L_j(z) = \frac{A_j(z)[n_j - E(A_j)](z-1)}{z^{n_j} - A_j(z)} \prod_{k=1}^{n_j-1} \frac{z - z_k}{1 - z_k} \tag{5.18}$$

where $L_j(z)$ is a *pgf* of queue length L_j, $A_j(z)$ is a *pgf* of the arrival process A_j, and z_k, $k = 0, 1, 2, \cdots n-1$, are the complex roots of characteristic equation, $z^{n_j} - A_j(z)$, specially, $z_0 = 1$.

The mean queue length is

$$\overline{L}_j = L_j(z)'|_{z=1} = \frac{Var(A_j)}{2[n_j - E(A_j)]} + \frac{E(A_j)}{2} - \frac{n_j - 1}{2} + \sum_{k=1}^{n_j-1} \frac{1}{1 - z_k} \tag{5.19}$$

where $E(A_j)$ and $Var(A_j)$ are expectation and variance of the arrival process A_j, respectively.

Since the arrival A_j follows the Possion distribution, substituting the relation, $E(A_j) = Var(A_j)$, to equation (5.19), we obtain the Bailey's result [99],

$$\overline{L}_j = \frac{n_j - [n_j - E(A_j)]}{2[n_j - E(A_j)]} + \sum_{k=1}^{n_j-1} \frac{1}{1 - z_k} \tag{5.20}$$

From the *pgf* of the characteristic equation, $A_j(z) = exp[\lambda_j \tau (z-1)]$, we obtain,

$$z_k = w_k \cdot e^{\frac{\lambda_j \tau (z_k - 1)}{n_j}} \tag{5.21}$$

where λ_j is the average arrival rate of A_j, z_k is the k^{th} root, and w_k is the k^{th} root of complex number 1, $w_k = exp(2\pi i \cdot k/n)$, $k = 0, 1, \cdots n-1$, especially when $k = 0$, then $w_0 = 1$.

We can get numerical solutions by the following iteration expression,

$$z_k^{m+1} = e^{\frac{2\pi i \cdot k - \lambda_j \tau (1 - z_k^m)}{n_j}} \tag{5.22}$$

Finally, since the system satisfies PASTA (Poisson Arrivals See Time Averages), thus, the mean waiting time at ingress node can be calculated by

$$D_{ingress,j} = \frac{\overline{L}_j}{\lambda_j} - \frac{\tau}{2} \tag{5.23}$$

Calculation Results The results of numerical calculation from equation (5.23) are shown in Figure 5.8 when reserved slots for per-flow equal 1, 5 and 25, and the frame length is 2400 slots.

Figure 5.8 clearly indicates that the mean waiting delay decreases as the number of reserved slots in a frame increases at the same utilization.

Therefore, the greater the reserved bandwidth is, the stronger the ability to smooth the arrival randomness is. When the number of reserved slots per-flow reaches to 5, the mean delay is nearly $3\tau/2$ even if $\rho = 0.9$. At worst case of number of slots being 1, the mean delay is only 5τ if the utilization is $\rho = 0.9$.

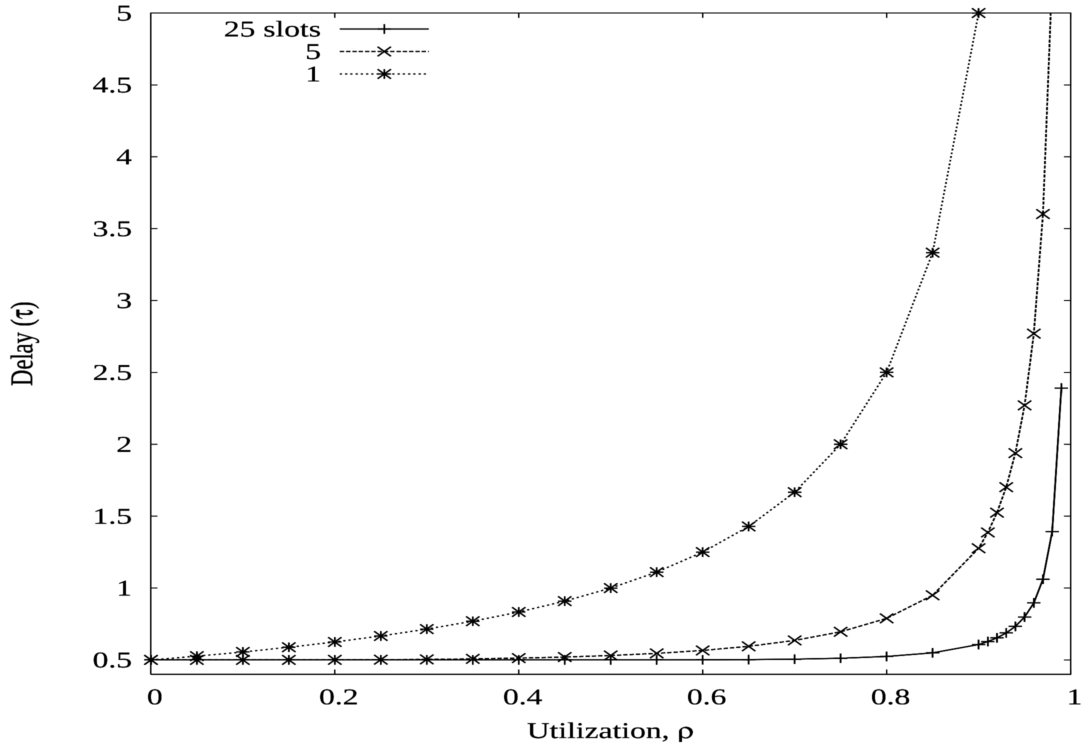

Figure 5.8: Mean Waiting Delay to Utilization at Ingress Node

Simulation Results For non time-sensitive traffic, strictly to say, the delay can not be calculated by equation (5.23), because the number of the slots in different frame periods is variable, the value is related to prior time-sensitive slots, therefore, we investigate the mean delay of non time-sensitive traffic when the mean bit rates of time-sensitive flows are changable at binomial distribution within reserved channel area. On the cases of reserved slots are 50% and 90% to total slots in frames, the relationship between mean delay and utilization of reserved channel is shown in figure 5.9,

The simulation results show that the mean delay of non time-sensitive traffic is near constant, and equals 0.5τ, because in theory the value is affected by prior time-sensitive traffic, however, the larger filler channel has strong power to reduce the impact, since the bandwidth of filler channel at least reaches to 10% of total bandwidth of the link fiber, which natually remove the impact of the variation of time-sensitive traffic.

5.2.2 Delay of Sending and Receiving Slots by Turn at Edges

At the ingress and egress edge nodes, the frames are sent and received as a whole, respectively, at ingress node the slots in the frame are transmitted one by one, each slot must wait its order in the frame buffer; at egress node the decomposing process of the frame must be done till whole frame arriving in the frame buffer, thus, the mean waiting delay is(for all appointed slots):

$$\overline{W}_w = \overline{W}_i + \overline{W}_e \approx \tau \qquad (5.24)$$

where W_w satisfies $0 < W_w < 2\tau$. Because for a given appointed channel, usually the

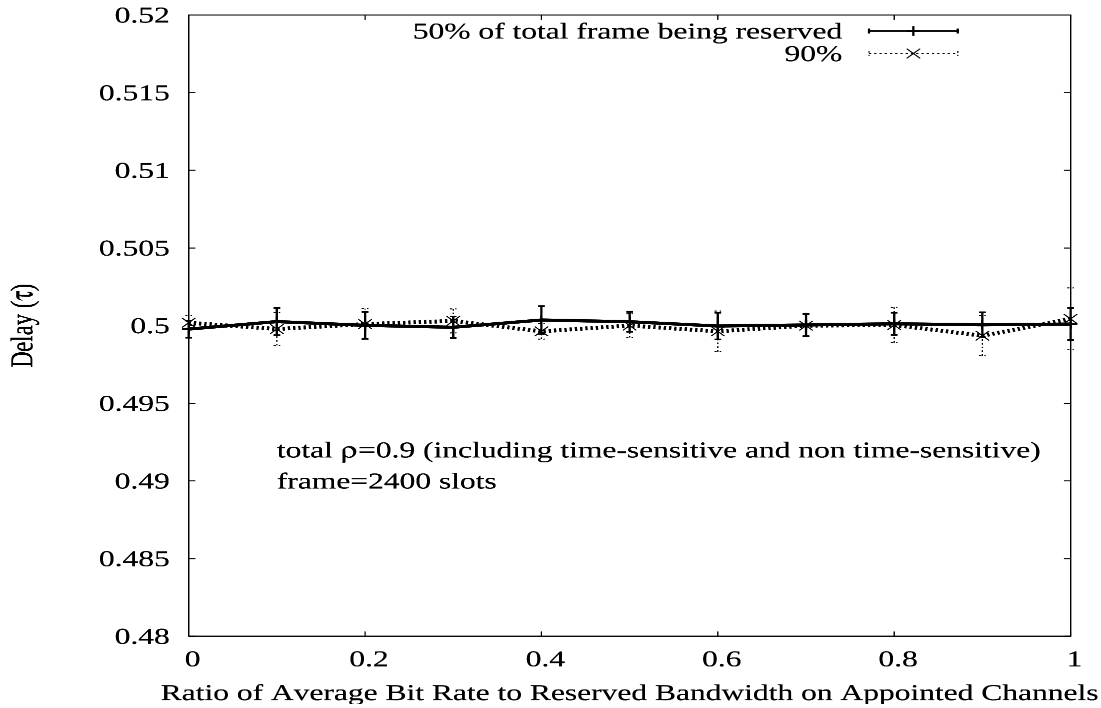

Figure 5.9: Mean Waiting Delay of Non Time-sensitive Traffic to Utilization at Ingress Node

entering frame at ingress node and departing frame at egress node are not synchronous in frame level, the slot positions in the frame of egress node are different with those in the frame of ingress node, the value is variable with different paths. To estimate the value simply, assume that the positions in the departing frame at egress node are random, thus, $\overline{W}_e = 0.5\tau$; similarly, because the positions in the entering frame depend on the channel distribution, hence, $\overline{W}_i = 0.5\tau$. Furthmore, if the positions of used appointed slots are in the place near to header, $0 < W_i < 0.5\tau$, on the contrary, if the positions is near to tailer, $\tau > W_i > 0.5\tau$. Therefore, the delay can be slightly adjusted by allocating the positions to appointed channels in entering frame.

5.3 Edge-to-Edge Mean Delay

For any core nodes in the MPsLS cloud, if traffic is independent with nodes and forwarding process is uniform, the mean delay for any traffic flow between ingress and egress edge nodes can be described as

$$\overline{D}_j = W_w + D_{ingress,j} + h \cdot D_{core,j} \tag{5.25}$$

where $D_{ingress,j}$ and $D_{core,j}$ are delays of flow j at ingress edge and any one core node, respectively, h is the number of hops along the path between ingress and egress edge nodes, and the first term of the right-hand side in equation (5.25), W_w, denotes the sum of the delays when sending and receiving slots at the edge nodes, which equals τ.

For time-sensitive application flows Since the channels are reserved at peak bit rate, $D_{ingress,j} = \tau/2$, thus,

- in the exact synchronization model:

$$D_{core,j} = (p_{app} + 1.5) \cdot \frac{\tau}{N} \to 0 \tag{5.26}$$

thus,

$$\overline{D}_{real(E)} = 1.5\tau + h \cdot (p_{app} + 1.5) \cdot \frac{\tau}{N} \approx 1.5\tau \tag{5.27}$$

- in the less strict synchronization model:

$$D_{core,j} = (p_{app} + c + 1.5) \cdot \frac{\tau}{N} \to c \cdot \frac{\tau}{N} \tag{5.28}$$

therefore,

$$\overline{D}_{real(L)} = 1.5\tau + h \cdot (p_{app} + c + 1.5) \cdot \frac{\tau}{N} \approx 1.5\tau + h \cdot c \cdot \frac{\tau}{N}. \tag{5.29}$$

For non time-sensitive application flows All filler channels are shared, and usually the number of usable total slots is larger than 500. Furthermore, assume that no congestion arises at the ingress nodes (or $\rho_{non} < 0.9$), therefore, at ingress edge node,

$$D_{ingress,j} \approx \frac{\tau}{2} \tag{5.30}$$

while at core nodes,

$$D_{core,j} = (p_{filler} + 0.5 + W_{non}) \cdot \frac{\tau}{N} \to W_{non} \cdot \frac{\tau}{N} \tag{5.31}$$

therefore,

$$\overline{D}_{non} \approx 1.5\tau + h(p_{filler} + 0.5 + W_{non}) \cdot \frac{\tau}{N} \approx 1.5\tau + h \cdot W_{non} \cdot \frac{\tau}{N} \tag{5.32}$$

Therefore, we can estimate the edge-to-edge mean delay for an application connection path by the calculation results from the equations (5.27), (5.29) and (5.32). The value is determined by the path, reserved bandwidth and utilization on the channel. Once the parameters and path are determined according to QoS requirement, the mean delay can be guaranteed.

Furthermore, at ingress node, both types of traffic has same average delay, $\tau/2$, but at core nodes, the exact synchronous model provides smallest delay, and in the non congestion state, the average delay of non time-sensitive traffic is even smaller than that of time-sensitive traffic transmitted on less strict synchronous model, because all non time-sensitive traffic shares large residual bandwidth. However, since network is a larger system, the congestion can not be avoided, in the congestion state, both exact synchronous and less strict synchronous models can still provide assure service, the delay does not vary with the traffic state, while the delay of filler slots will increase drastically, but it is acceptable for non time-sensitive applications.

5.4 Summary

In this chapter, we analyzed the variable delay when the traffic is transferred along the connection path between ingress edge and egress edge in MPsLS cloud. And propose an approximate expression to estimate the total delay. For the time-sensitive applications on less strict synchronous channels, on worst case of $c = 200$, $N = 2400$ and $h = 30$, the total delay can reach to $\overline{D}_{real(E)} < 4.5\tau < 0.6ms$.

Chapter 6

Maintaining High Utilization of Resource in MPsLS

In MPsLS network, the resource is reserved at per-flow for all time-sensitive application flows, each flow uses different bandwidth and synchronous slots according to own QoS requirements.

The non time-sensitive applications are different with time-sensitive applications, which all share large left bandwidth except for the parts used by all appointed channels, the performances are affected by time-sensitive applications.

In this chapter, we simply discuss the contribution of the usage of idle appointed slot positions to raise resource utilization and to reduce delay of non time-sensitive traffic.

The remainder of this chapter is organized as follows. Section 6.1 analyzes the resource utilization on both idle appointed slot positions being used or not. Section 6.2 discusses the influence of using idle appointed slot positions to non time-sensitive application delay. Then, Section 6.3 summarizes this chapter.

6.1 Utilization of Channels and Throughput for Non Time-sensitive Traffic

In our MPsLS scheme, the appointed channels are reserved for time-sensitive applications at per-flow, while the filler channel is shared by all non time-sensitive applications. Therefore, not only appointed channels and filler channel have different channel utilizations, but also the utilizations of the appointed channels for different time-sensitive application flows are variable.

Assume that the total link rate of a fibre is B bps, the period of a frame is τ seconds, and the length of a slot is L_{slot} bits, thus the bandwidth of a channel with a slot in a frame, b, is

$$b = \frac{L_{slot}}{\tau} \tag{6.1}$$

and, the length of a frame, N, is

$$N = \lceil \frac{B \cdot \tau}{L_{slot}} \rceil \tag{6.2}$$

where $\lceil\star\rceil$ denotes an integer number not more than $\star$. To simplify, we assume that $(B \cdot \tau)/L_{slot}$ is integer, thus

$$N = \frac{B \cdot \tau}{L_{slot}} \tag{6.3}$$

6.1.1 For Time-Sensitive Flows

Generally, the reserved channel bandwidths for time-sensitive applications have relations with their bit rate, assume that there are i time-sensitive application flows being transferred simultaneously. The average bit rates of those flows are A_j $(j = 1, 2, \cdots i)$. S_j and P_j $(j = 1, 2, \cdots i)$ are reserved bandwidths and peak bit rates corresponding to the flow A_j, respectively.

Thus, the number of reserved slots, n_j $(j = 1, 2, \cdots i)$ is defined by the following equation,

$$n_j = \lfloor \frac{A_j \cdot \tau}{L_{slot}} \rfloor + N_{over,j} \qquad j = 1, 2, \cdots i \tag{6.4}$$

where $\lfloor\star\rfloor$ denotes an integer not smaller than $\star$, and $N_{over,j}$ is the number of over-reserving slots. Accordingly, $S_j = n_j \cdot b$, and their utilizations satisfy

$$\rho_j = \frac{A_j}{S_j} = \frac{A_j \cdot \tau}{L_{slot} \cdot n_j} \qquad j = 1, \cdots i. \tag{6.5}$$

Specially, when the bandwidths of the applications are reserved at peak bit rate, then

$$n_j^p = \lfloor \frac{P_j \cdot \tau}{L_{slot}} \rfloor \qquad j = 1, 2, \cdots i. \tag{6.6}$$

and the channel utilization is

$$\rho_j^p = \frac{A_j \cdot \tau}{L_{slot} \cdot n_j^p} \qquad j = 1, 2, \cdots i. \tag{6.7}$$

6.1.2 For Non Time-Sensitive Flows

Assume that there are numbers of non time-sensitive application flows besides i time-sensitive application flows. The sum of average bit rates is V_{non}. All non time-sensitive flows share a large filler channel, the average bandwidth available for those non time-sensitive flows is B_{non}.

If idle bandwidth reserved for appointed channels can be used by filler slots temporarily, the usable whole bandwidth of filler channel is

$$B_{non} = B - \sum_{j=1}^{i} A_j \tag{6.8}$$

where B is the total bandwidth of a link fibre.

The average utilization of filler channel, ρ_{non}, is

$$\rho_{non} = \frac{V_{non}}{B_{non}} \tag{6.9}$$

Otherwise, if idle bandwidth reserved for appointed channels can not be used by filler slots temporarily, the usable bandwidth is only

$$B'_{non} = B - \sum_{j=1}^{i} n_j \cdot b \tag{6.10}$$

Similarly, the utilization of filler channel is

$$\rho'_{non} = \frac{V_{non}}{B'_{non}} \tag{6.11}$$

Comparing equations (6.9) and (6.11), since $A_j \leq n_j \cdot b$ even if $N_{over,j} = 0$, thus, $\rho'_{non} \geq \rho_{non}$, if $N_{over,j} \neq 0$, then $\rho'_{non} \gg \rho_{non}$.

Therefore, if total bit rate of non time-sensitive flows is constant, then when idle slot positions on appointed channels are used by filler slots temporarily, the filler channel has small utilization because the usable bandwidth is large. While when idle slot positions on appointed channels are not used by filler slots temporarily, the filler channel has large utilization. On the contrary, if total bit rate of non time-sensitive flows is variable, when idle slot positions on appointed channels are used by filler slots temporarily, the link can provide large throughput for non time-sensitive applications.

6.1.3 Total Resource Utilization of Link Fiber

Besides increasing the throughput for non time-sensitive traffic, the usage of idle appointed slots can also improve the resource utilization of total link fiber. If the state of time-sensitive application and non time-sensitive application flows is as same as that described in sections 6.1.1 and **??**, the utilization of link fiber is

$$\rho = \frac{V_{non} + \sum_{j=1}^{i} A_j}{B} \tag{6.12}$$

When the idle appointed slots are used temporarily, since the maximum throughput of non time-sensitive traffic can reach to $B - \sum_{j=1}^{i} A_j$, thus, the maximum utilization of link fiber can be near 1. But when the idle appointed slots are not used temporarily, the maximum throughput of non time-sensitive traffic is $B - \sum_{j=1}^{i} n_j \cdot b$, because $n_j \cdot b \geq A_j$, if usually existing some $N_{over,j} \neq 0$, then $\sum_{j=1}^{i} n_j \cdot b \gg \sum_{j=1}^{i} A_j$, therefore, $\rho \ll 1$.

6.2 Impact of Using Idle Appointed Slots to Delay of Non Time-sensitive Traffic

The usage of idle appointed slots not only improve the resource utilization and throughput of non time-sensitive traffic, but also reduce the delay of non time-sensitive traffic. In this part, we investigate the delay of non time-sensitive traffic at core and ingress nodes on the cases of reserved bandwidths for time-sensitive traffic being 40% and 80%, and the average utilization of appointed channels being 0.5 and 0.9, respectively.

6.2.1 At Core Node

On both cases of the idle appointed slot positions being used or not by filler slots, the mean delays are simulated, besides, when the idle appointed slots being used, the mean delay can also be calculated by equation (5.14). The variation curves of the mean delay with utilization are shown in figures 6.1- 6.4. The results indicate that (1) when the idle appointed slot positions are not used, the delay of non time-sensitive traffic has not relations with the utilization of appointed channels, but when the idle appointed slot positions are used, the delay of non time-sensitive traffic decreases with the decreasing of the utilization of appointed channels. (2) the usage of idle appointed slot positions can improve the capability of carrying non time-sensitive traffic. (3) the congestion area of the non time-sensitive traffic is deferred if the idle appointed slots are used, the congestion area means that the delay drastically varies with the utilization.

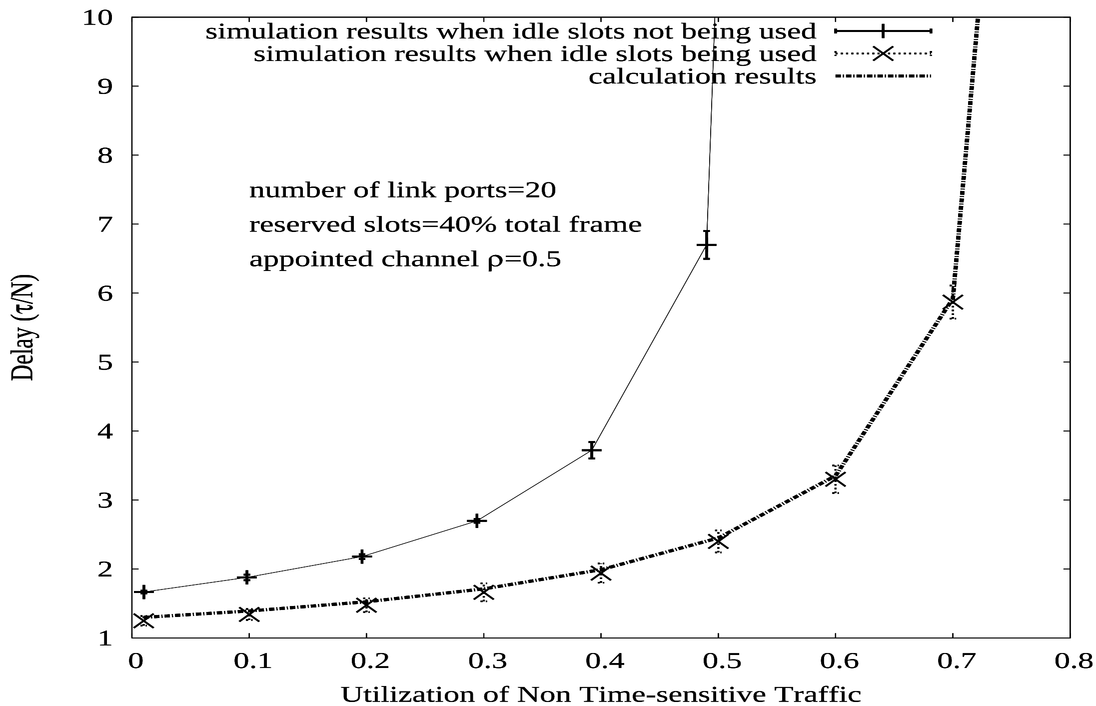

Figure 6.1: Mean Waiting Delay of Non Time-sensitive Traffic at Core Node

6.2.2 At Ingress Node

Similar to analysis in section 6.2.1, on both cases of the idle appointed slot positions being used or not, the mean delays can be calculated according to equation (5.23), furthermore, the simulation is done when the idle appointed slot positions being used, the curves are shown in Figures 6.5- 6.8. The results show that (1) the congestion area of the non time-sensitive traffic arrives later if the idle appointed slots are used, where the delay drastically varies with the utilization. (2) the average delay almost equals $\tau/2$ if only in non congestion state.

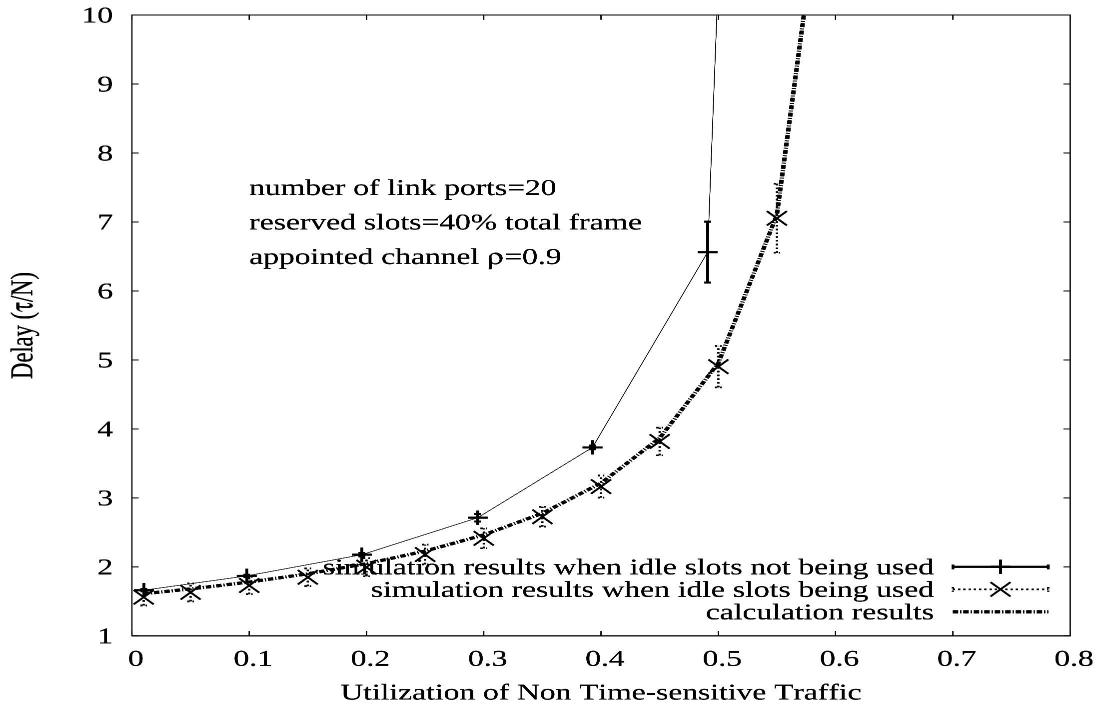

Figure 6.2: Mean Waiting Delay of Non Time-sensitive Traffic at Core Node

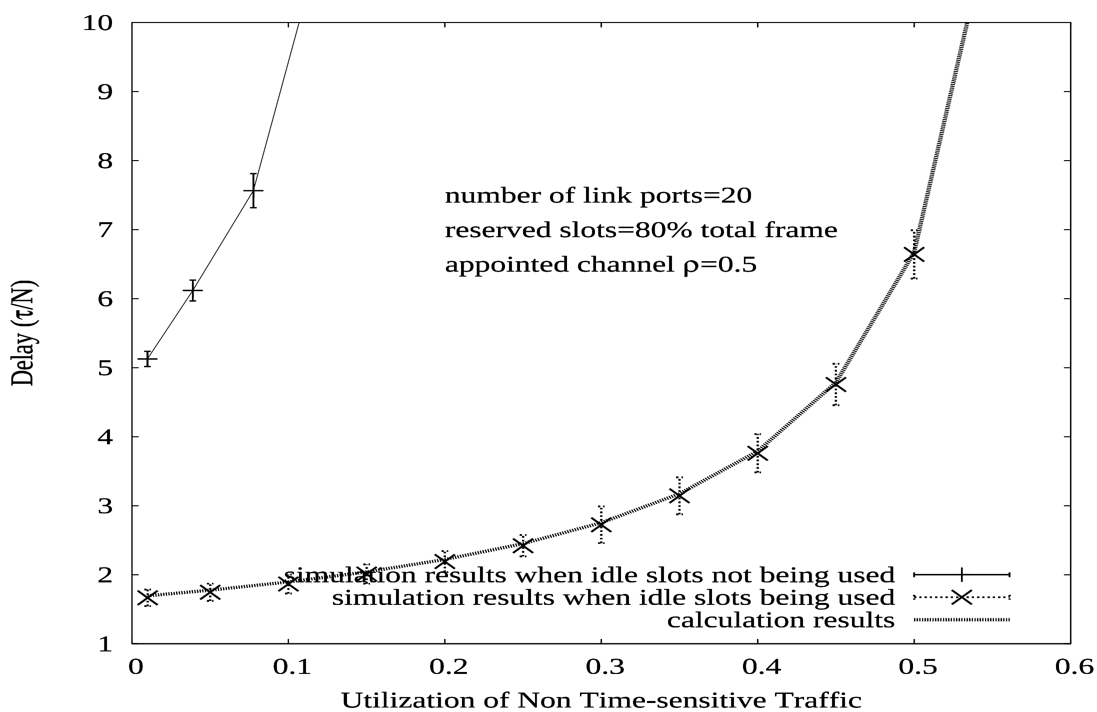

Figure 6.3: Mean Waiting Delay of Non Time-sensitive Traffic at Core Node

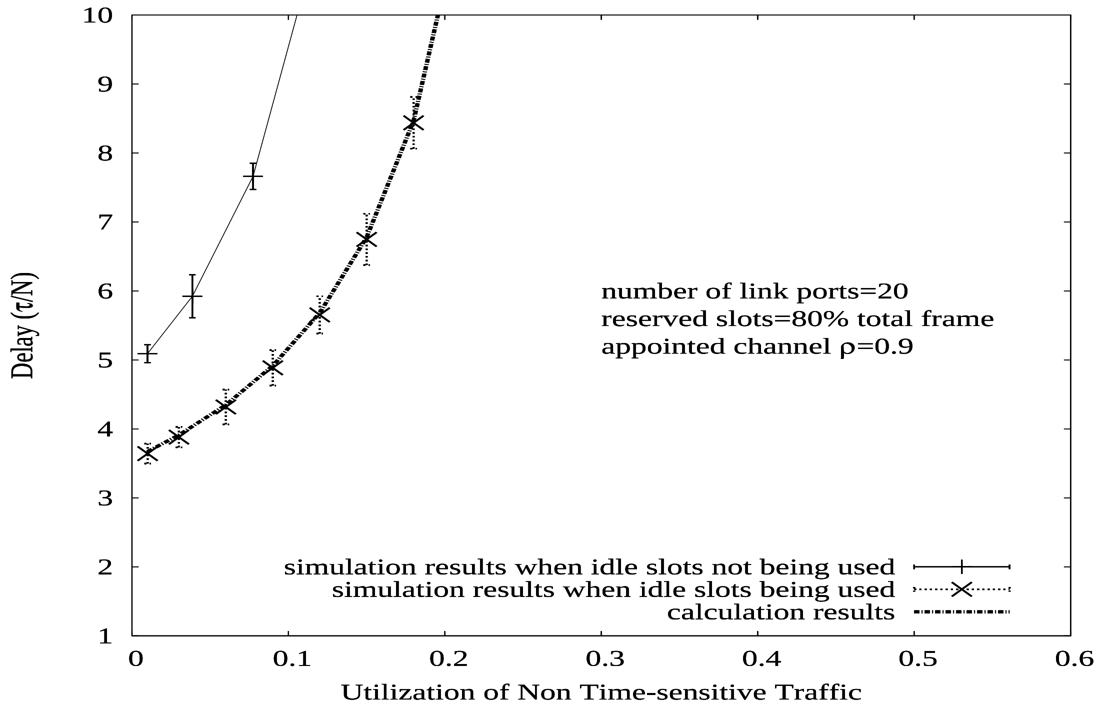

Figure 6.4: Mean Waiting Delay of Non Time-sensitive Traffic at Core Node

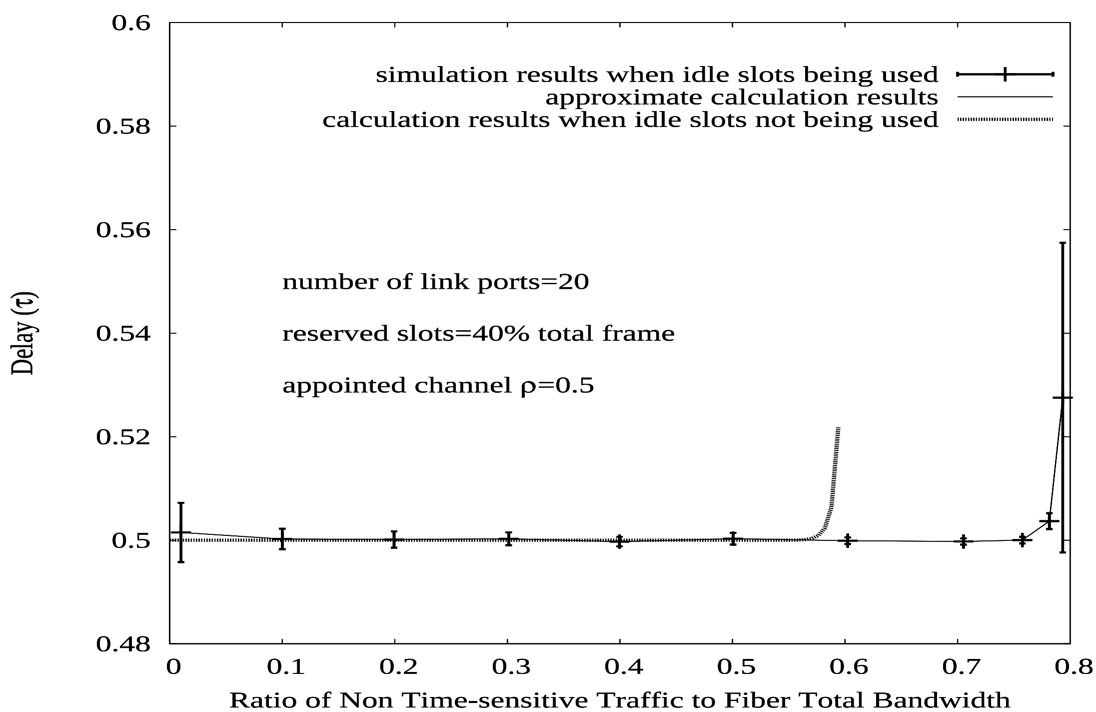

Figure 6.5: Mean Waiting Delay of Non Time-sensitive Traffic at Ingress Node

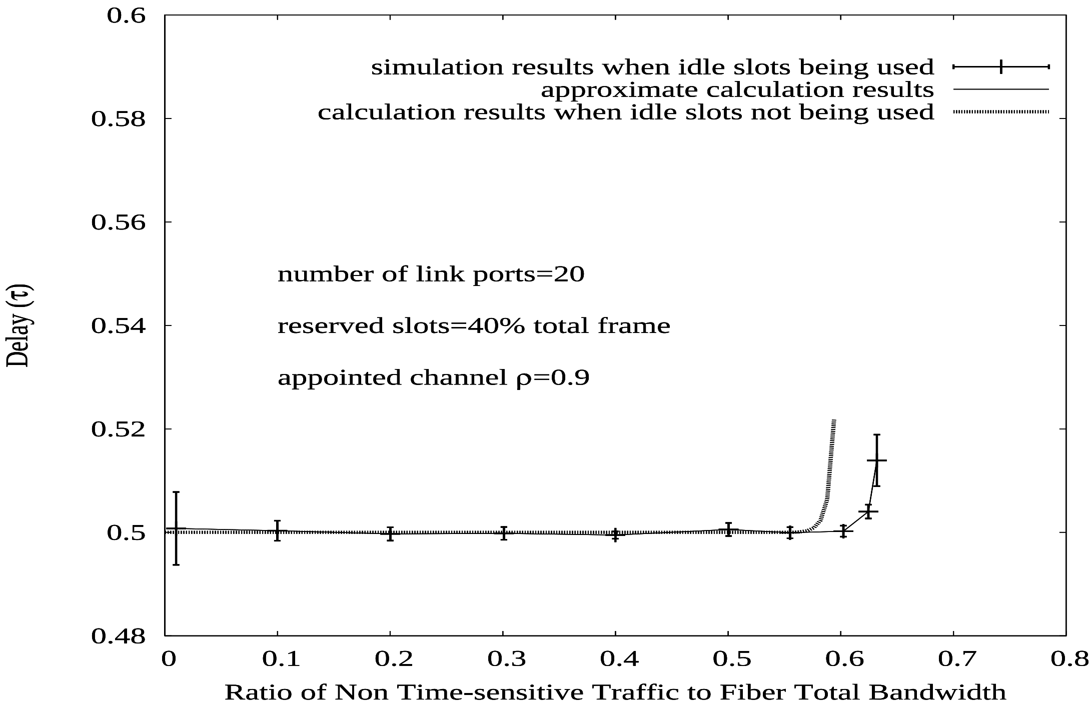

Figure 6.6: Mean Waiting Delay of Non Time-sensitive Traffic at Ingress Node

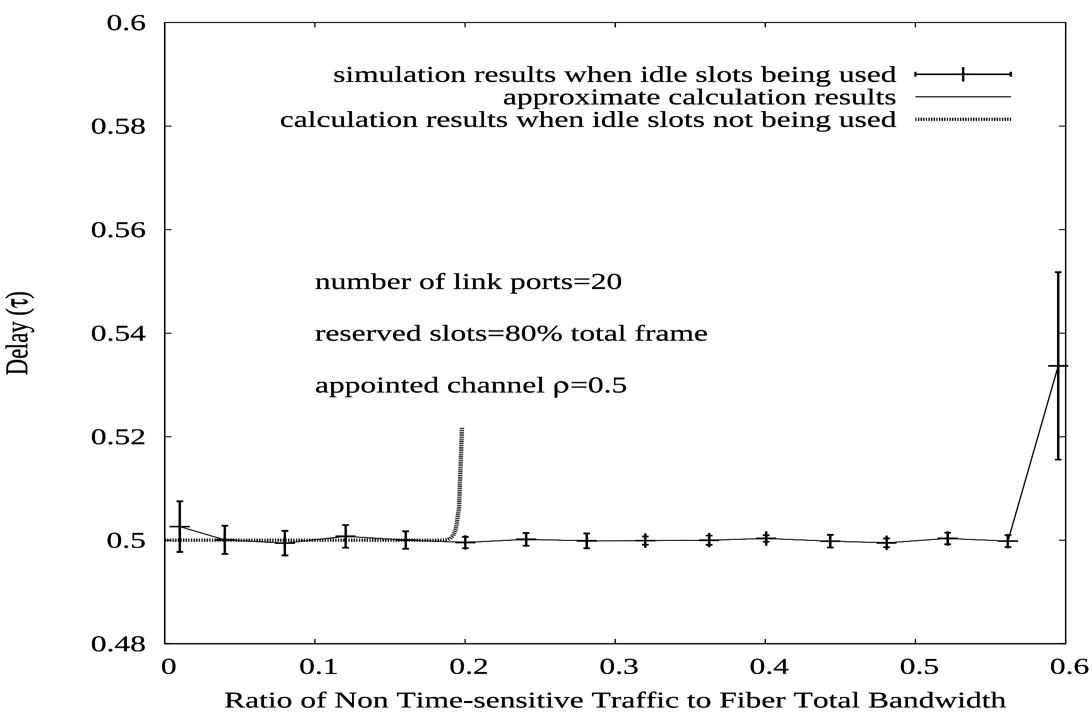

Figure 6.7: Mean Waiting Delay of Non Time-sensitive Traffic at Ingress Node

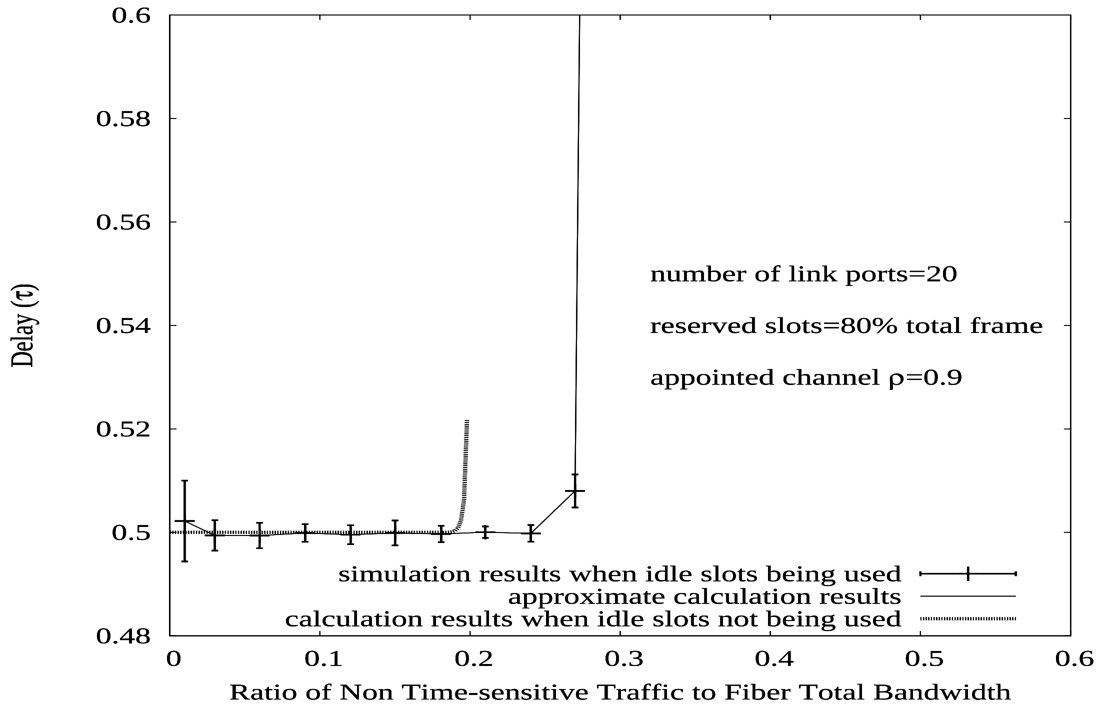

Figure 6.8: Mean Waiting Delay of Non Time-sensitive Traffic at Ingress Node

6.3 Summary

In this chapter, we simply analyzed the impact of the usage of idle appointed slot positions to resource utilization and average delay of non time-sensitive traffic, the simulation and calculation results show that MPsLS scheme not only improves resource utilization, but also reduces the average delay of non time-sensitive traffic, specially, defers the arriving of congestion area.

Chapter 7

Comparison with Other Switching Schemes

In a large network, the bottleneck tends to arise at core nodes, at core nodes, on the one hand, the transmission of the traffic needs to slow down or temporarily stops in order to investigate the message included in the data, on the other hand, there is larger utilization at core area than edge area, the confluence process of the traffic may leads to temporary congestion, which causes the degradation of the service quality.

As a new switching scheme, the main advantage of MPsLS is that in core area it can specify the time-sensitive application flow traffic based on synchronous channels, and reduce the randomness of the arrival process, thus, MPsLS can remove congestion and contention for the transmission channel and processing resource at core node, accordingly, it has a potential of providing better service quality to time-sensitive applications.

It is very difficult to compare the performance of two different schemes comprehensively, in this chapter, we mainly compare delay performance of MPsLS at core node with that in other typical schemes, such as DiffServ, ATM and DTM, which are considered as effective schemes to improve QoS for time-sensitive applications currently. At the same time, we also compare the calling loss performance in MPsLS with that in other schemes.

The remainder of this chapter is organized as follows. Section 7.1 analyzes the delay and jitter performance in different schemes based on the simulation results. Section 7.2 simply discusses the calling loss. Then Section 7.3 summarizes this chapter.

7.1 Delay and Jitter

The delay and jitter are the important parameters to evaluate the network transmission performance, although several network models deployed currently can improve QoS, the delay performances are different.

7.1.1 Datagram Model

In usual IP network, the basic transfer mode is datagram model, in this part, we will simulate the delay in both best-effort and DiffServ models, respectively by assuming that the arrival of the flows is Poisson process, the size of the packets are variable, the average length of the packets is $1/2 \times 1500$ bytes.

7.1.1.1 Simple Best-effort Model

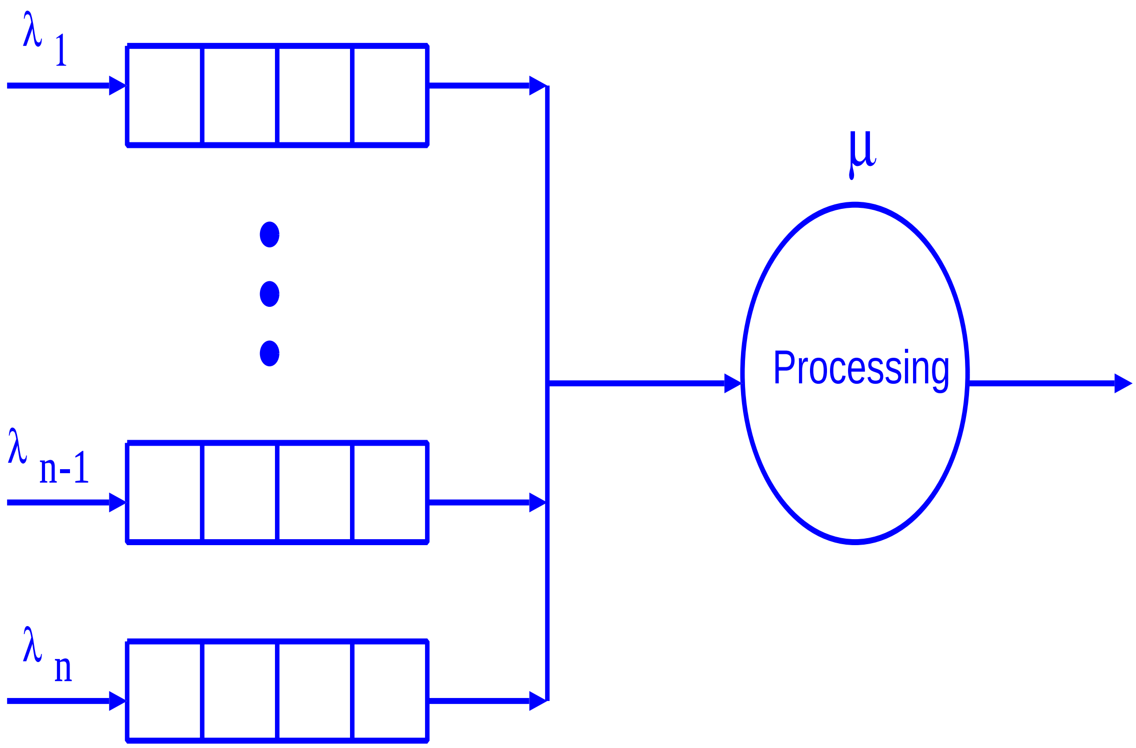

Figure 7.1: Service Model in Usual IP Network

At a core node, the service model of simple best-effort is illustrated in figure 7.1, the service time satisfies exponential distribution, if processing speed of the node is 10Gbps, let the processing time of 512 bits (length of slots) be one unit time, thus average processing time for a packet is about $1/2 \times 1500 \times 8/512 = 12$ unit time, the system time of the packets in service system on different utilization can be obtained with the variation of the time interval of the packets arriving. The results are shown in figure 7.2, where the horizontal axis indicates the utilization, and the vertical axis illustrates the ratio of system time to unit time, the simulation results show that although the average delay is very small, there exists lager maximum delay even if the utilization is small, which means that queue increasing or temporary congestion is inevitable even on the small utilization case, when the utilization is about 0.7, the maximum delay of the packets exceeds 400 unit time. While for MPsLS model, although average value is larger than that in IP network on the case of small utilization, the maximum delay is controlled by allowable offset range, even if the utilization is 1, specially, when $c = 200$, the maximum delay is 400 unit time; when $c = 50$, the maximum delay is only 100 unit time. Therefore, MPsLS can avoid the worst case, but the cost is the increasing of the average delay on the small utilization.

7.1.1.2 Differentiated Service Model with Priorities

If the application flows are classified at different priorities, the service model is shown in figure 7.3, furthermore, the implementation processes are divided into two types, preemptive and non-preemptive, the former means that the priority is absolute, the implemen-

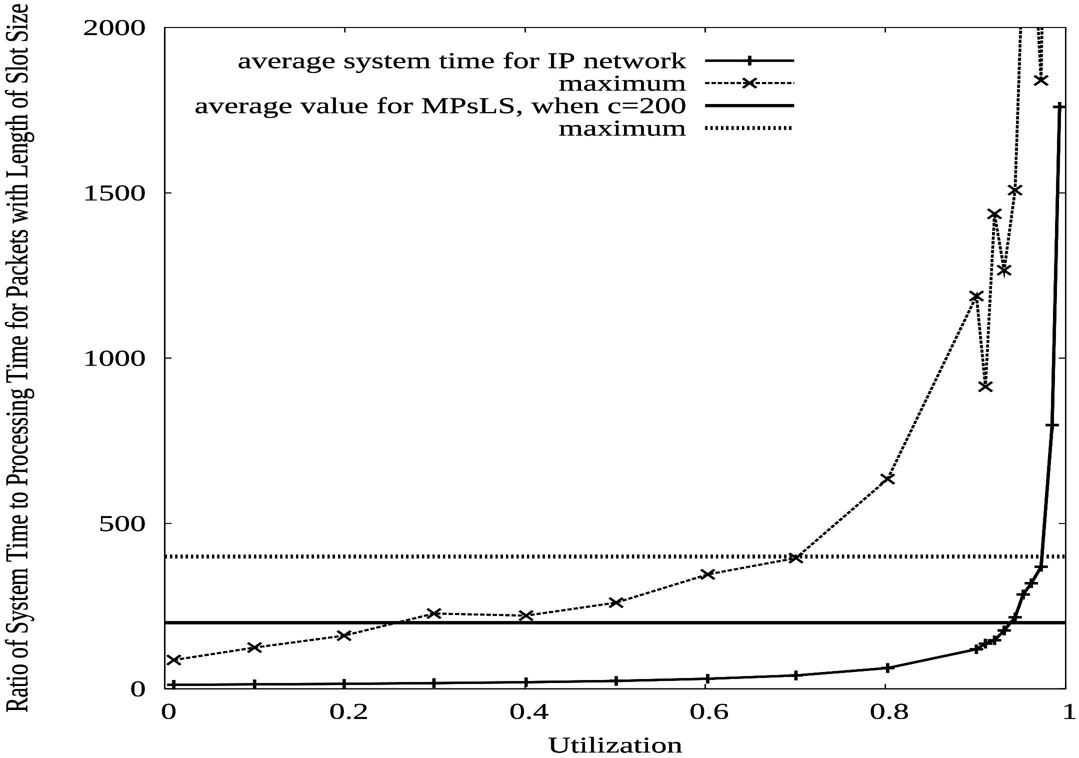

Figure 7.2: Relationship between System Time and Utilization

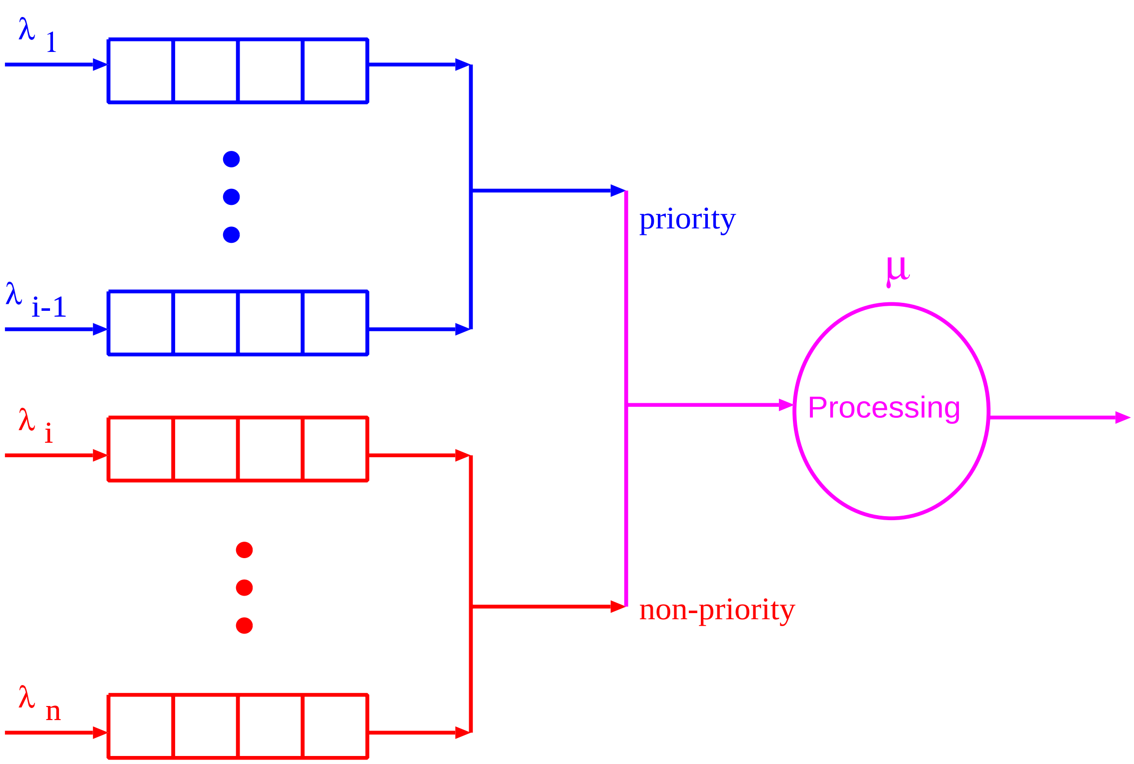

Figure 7.3: Service Model with Priority Classes in Usual IP Network (such as DiffServ)

tation of non priority class service can be interrupted when the traffic with priority class arrives; the latter means that the implementation of non priority class service can not be interrupted, although the former obtains smaller delay for priority service class than the latter, it leads to large loss ratio and much worse performance to non priority class traffic, therefore, usually non-preemptive priority class mechanism is used, in this part, the relationships of maximum system time and jitter with utilization are simulated for non-preemptive priority class service.

When the Ratio of Time-sensitive Traffic to Total Traffic being Constant

Assume that the ratio of the priority class traffic to total traffic is constant, for non-preemptive, the variation of the average system time of both priority and non priority class service with total utilization are shown in Figure 7.4, at the same time maximum system time and maximum jitter of priority service class traffic are shown in Figure 7.5 when the ratio equals 1/3.

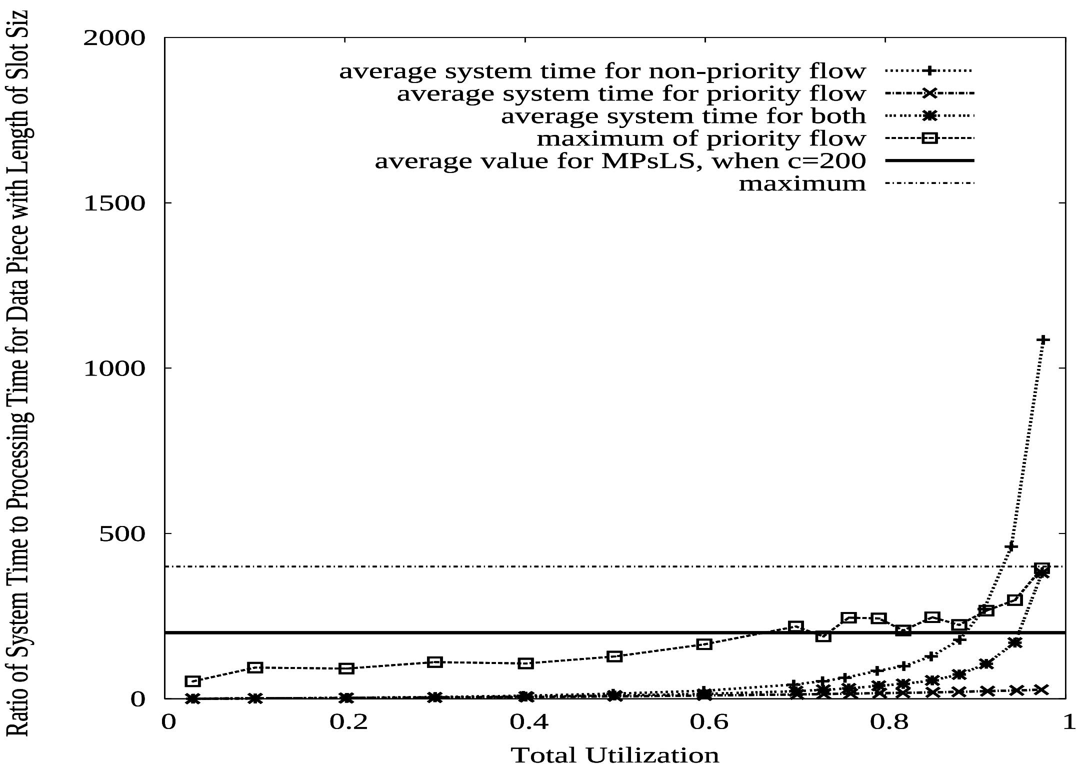

Figure 7.4: Relationship between System Time and Total Utilization

Figure 7.4 indicates that the applications with priority service class can obtain smaller average delay, therefore, which can provide better service quality than that in non differentiated service model. While the average delay of non priority service class traffic is enlarged, it is allowable. Even for priority service class traffic, the queue increasing or temporary congestion is inevitable, the maximum delay exceeds 200 when total utilization $\rho > 0.7$.

Besides, Figure 7.5 shows that the jitter still exists, the maximum value is about 100 unit time.

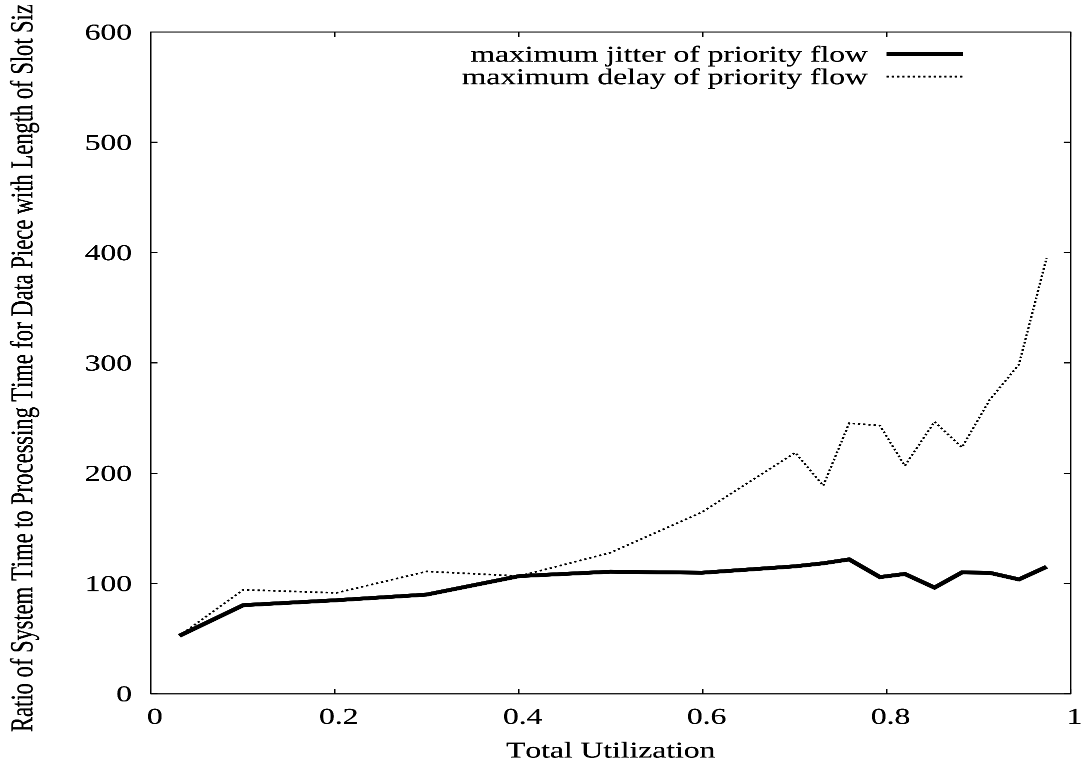

Figure 7.5: Relationship between Maximum Delay Jitter of Priority Class Traffic and Total Utilization

When Total Utilization being Constant

If assume that the total utilization of both the priority class traffic and non priority class traffic is constant, for non-preemptive, the variation of the average system time of both priority and non priority class service with the utilization of priority class traffic are shown in Figure 7.6, while maximum system time and maximum jitter of priority service class traffic are shown in Figure 7.7 when the total utilization equals 0.9.

The results illustrate that the maximum delay and jitter of priority service class traffic increase with the increasing of the utilization of priority service class traffic when the total utilization is constant 0.9, if the utilization of priority class traffic reaches to about 0.75 (while the utilization of non priority class traffic is about 0.15), the maximum delay will exceed 400 unit time, the maximim jitter is near 100 unit time.

7.1.2 Virtual Circuit Mode

ATM is a typical virtual circuit model, which supports five types of service classes, among them, CBR (constant bit rate service) is usually used to transfer time-sensitive applications, which has the ability of providing better QoS.

In CBR service class, the application traffic is shaped according to peak cell rate(PCR) at edge node of ATM network, the shaping process is done by the algorithms, such as leaky bucket or dual leaky bucket. At core nodes, the cells are forwarded based on priority scheduling mechanism, since CBR service class has priority to others, the traffic is not impacted by other types of traffic. Logicaly, the service model of CBR can be abstracted

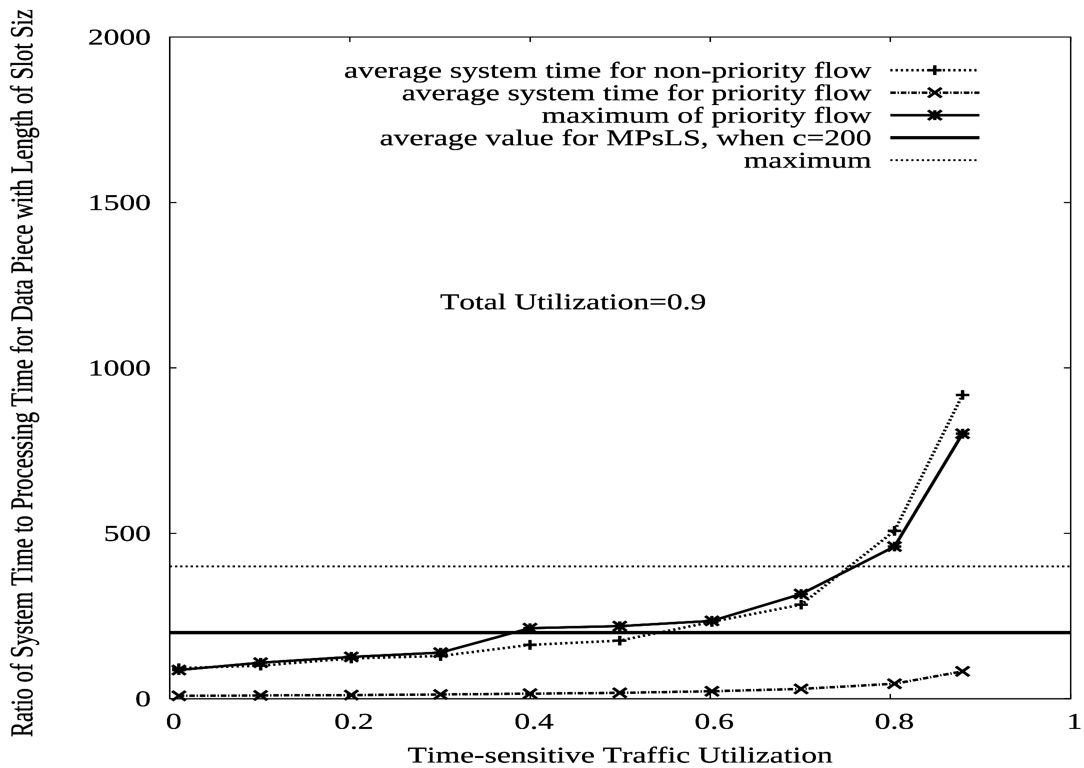

Figure 7.6: Relationship between System Time and Total Utilization

into Figure 7.8 if ignoring the short switching delay.

However, in constrast to switching delay of MPsLS, the scheduling delay also needs to be considered. In the past few years, many researches on the scheduling algorithms have been done, the most basic one is Round Robin among all CBR channels.

Because the flows may have different bandwidth, it is difficult to ensure all cells of the flows being forwarded at fixed time interval even if implementing cell level Round Robin scheduling, which results in the little delay increasing and jitter. Even on the ideal case of all flows having same average bit rate, thus, the cell level Round Robin scheduling can be easily implemented, obviously, the bandwidth can be strictly guaranteed, however, on the one hand, the temporary bit rate may be variable with the time, the arrival of the cells does not accurately match with the scheduling time assigned to it, which leads to delay or jitter, on the other hand, even if the bit rate is constant, it is also impossible of the arrived cells accurately matching with the scheduling time due to the asychronous arrival process, besides, more than one cells may arrive simultaneously. Therefore, to strictly say, CBR can not completely remove the jitter, the value is determined by the scheduling mechanism, because usually the value is small, which is main reason of CBR channels being considered as dedicated channels liking private lines by most users and service providers.

In this part, we simulate the maximum jitter when the arriving process is Poisson. Assume that there are four time-sensitive application flows with same average and peak bit rates, thus, maximum utilization of each flow is 1/4, if the scheduling is implemented fairly to four flows at cell level cycle, and the distribution of time interval between the cells satisfies exponential distribution, but the minimum time interval between neighboring

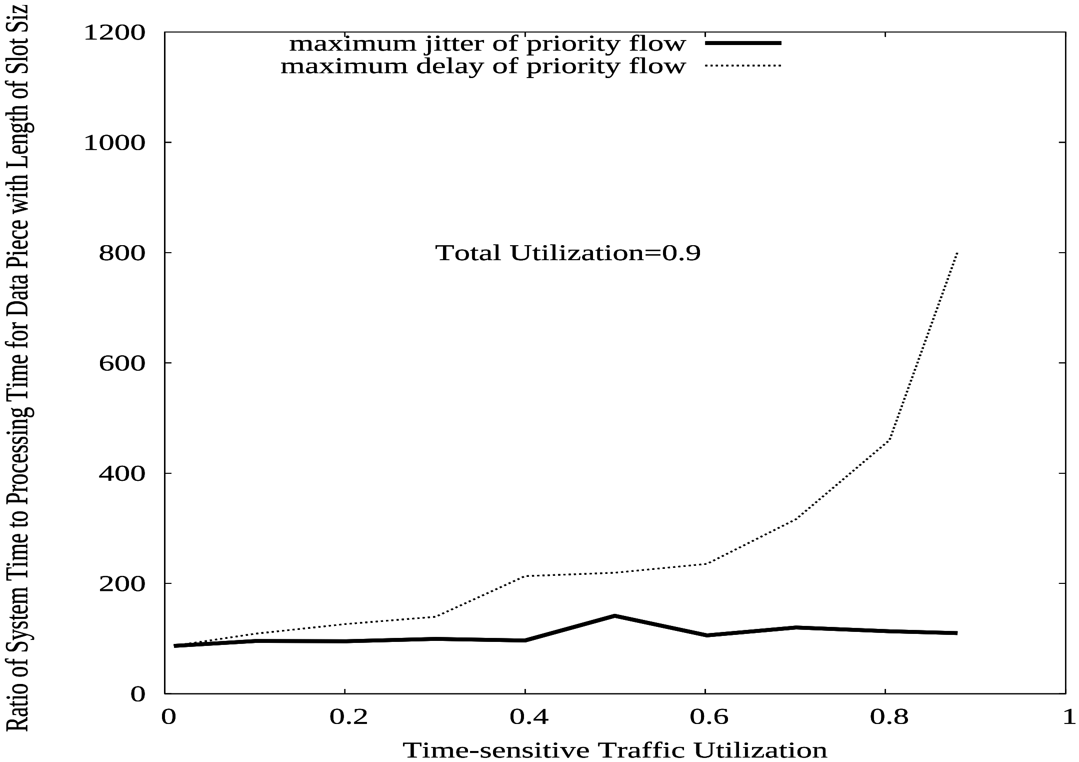

Figure 7.7: Relationship between Maximum Delay Jitter and Utilization of of Priority Class Traffic

cells is four cells processing time, namely, peak bit rate is $B/4$, if reserving bandwidth of $B/4$ to each flow, the simulation results of maximum jitter are shown in Figure 7.9.

The results indicate that maximum jitter is near 4 cells processing time, besides, the average jitter is non-zero, although the value decreases with the increasing of the utilization.

In fact, the number of the CBR channels is not only 4, the maximum jitter drastically increases with the increasing of the number of the CBR channels, we consider another special case, if there are 1000 CBR flows with same average and peak bit rate, the scheduling period for each channel is 1000 cells processing time, thus the maximum jitter will reach to 1000 cells processing time (or $1000 \times 53 \times 8/512 = 828$ unit time). Therefore, in CBR class service, the jitter also exists, the value is related to the number of the total CBR channels and the scheduling mechanism.

The data in figure 7.9 illustrate that the average jitter decreases with the increasing of the utilization, accordingly, a new question appears, when the application flows are constant bit rate flows, if or not the average jitter is always zero? the answer is ‘no‘, we consider a special case of all flows being constant bit rate, assume that there are four constant bit rate flows, the bit rates are 1/12, 2/12, 3/12 and 4/12, respectvely, then the ideal cell level scheduling is implemented liking figure 7.10, we can see that for some cells the jitter still exists at least in one flow, therefore, even on special case of all flows being constant bit rate, it does not guarantee that the jitter is always zero for all flows. Usually, the value of the jiiter deponds on not only the flow state itself, but the flow states of other applications.

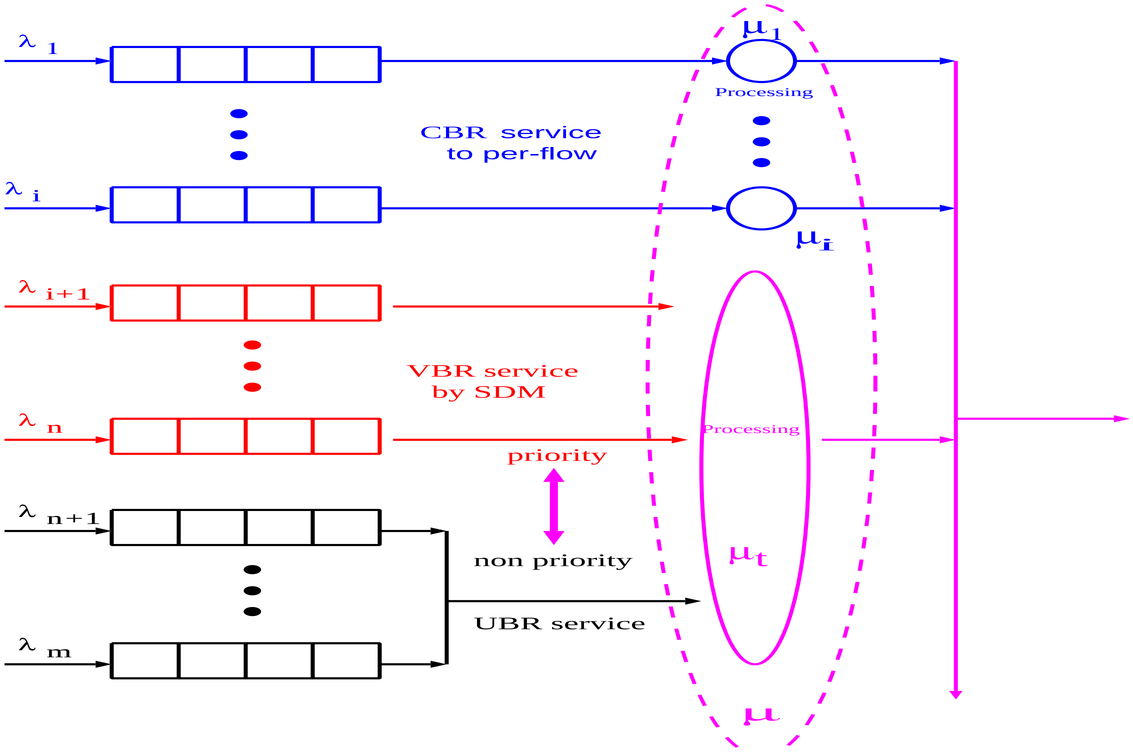

Figure 7.8: Logic Service Model in ATM Network

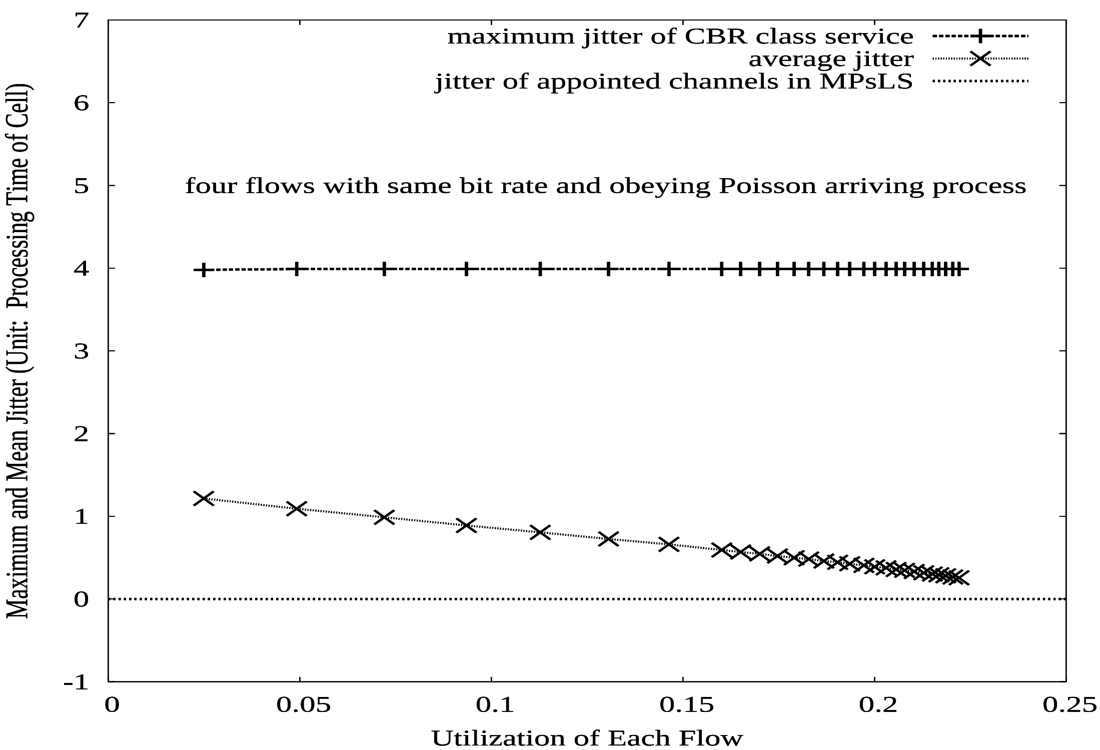

Figure 7.9: Maximum Jitters in MPsLS and CBR Class Service

For MPsLS model, since the slot positions are fixed during each session, thus, the jitter is 0 when the slots pass through core nodes, it is realized by specifying the arrival process of the traffic based on the synchronous channel.

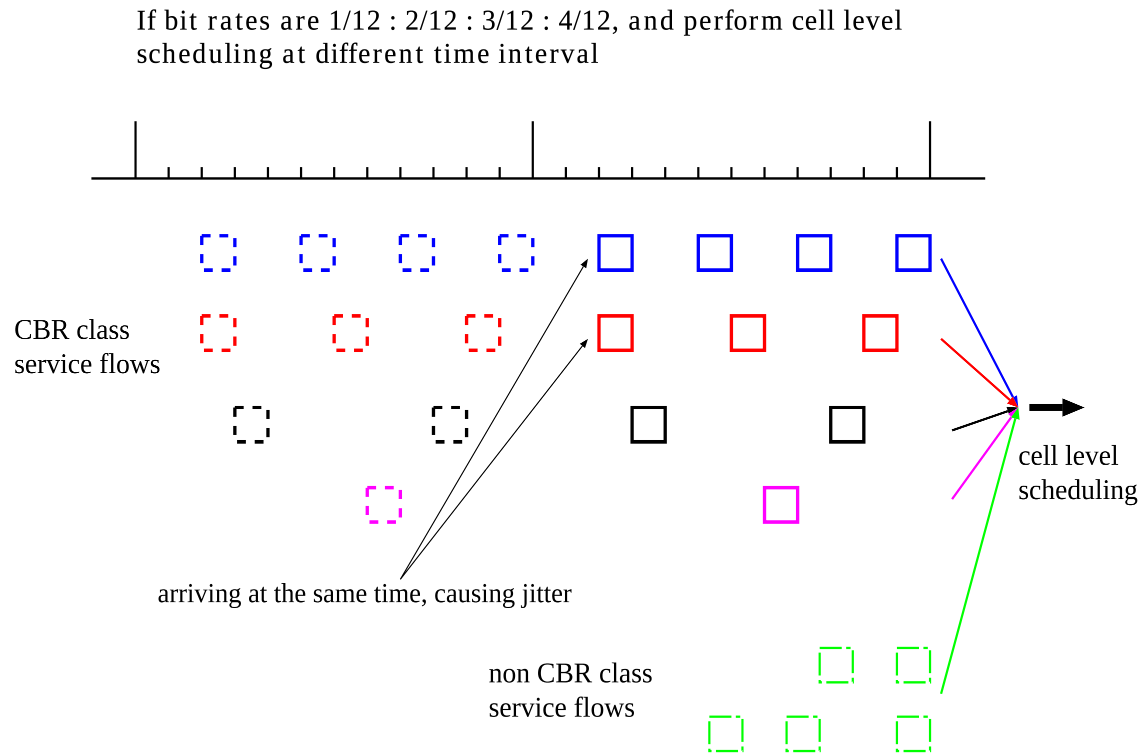

Figure 7.10: Example of Jitter Existing Even for Ideal Constant Bit Rate Flows

7.1.3 DTM and MPsLS

DTM is a completely synchronous mode, all traffic is transmitted on synchronous channels, while MPsLS is a concrescence mode, in which the time-sensitive traffic is transmitted on synchronous channels, and non time-sensitive traffic is transmitted by pseudo-synchronous channel. Therefore, although both modes are different, if network resource and segmentation parameters, such as total fiber bandwidth, frame period and length of the slots, are same, the delay performance for the time-sensitive traffic also completely same on both transmission modes, because the slot positions in each link section are reserved and fixed, the jitter is zero when the traffic passes through a core node.

Besides, in DTM network, although the jitter is also 0 similar to in MPsLS at core nodes, the switching delay at switch nodes equals one frame periods (or 125 μs=2400 unit time), it is much larger than that in MPsLS, the value in the latter is determinted by the QoS requirement, the maximum delay is double size of offset range, usually the value is less than 400 unit time (when c=200).

7.1.4 Qualitative Evaluation

Based on the analysis and simulation results to delay and jitter, the qualitative comparison of the dealy and jitter in different schemes is shown in figure 7.1.

Table 7.1: Transmission Performance in Different QoS Improvement Schemes

Transfer Mode / Performance Evaluation / Type of Application	Asynchronous modes			Synchronous mode	
	DiffServ	MPLS	ATM	MPsLS	DTM
Time-sensitive applications	Fair	Fair	Fair	Good	Good
Non Time-sensitive applications	Good	Good	Fair	Good	Poor

7.2 Calling Loss

The access performance is another important parameter to evaluate the network performance, in different network modes, the access condition is different, for exemple, in datagram mode, since best-effort service is main stream, only if there is enough resource, the access is admitted. In ATM, since the traffic is transmitted by virtual channels, the explicit calling setup is necessary. Furthermore, in synchronous mode, the more strict access requirement needs to be considered, esbaslishing channel requires not only the quantity of the slots, but the fixed positions, therefore, there exists remarkable calling loss in synchronous mode.

7.2.1 DTM

DTM is a synchronous model, which uses dynamical resource allocation mechanism, in constrast to statical channel allocation technology, DTM partly solves the issue of low utilization, however, it still has some shortcomings, on the one hand, due to all traffic is transmitted by dynamical channels, it causes the calling loss for non time-sensitive applications, on the other hand, for time-sensitive applications, since the temporary idle slot positions can not be used during the session periods, which can not completely remove the issue of lower resource utilization.

In this part, we investigate the calling loss of the sessions of non time-sensitive applications in DTM, assume that the number of arrived sessions at any time satisfies the Poisson arrival process with average number of n_c, thus the probability of k sessions arriving simultaneously is

$$p(k) = e^{-n_c}\frac{k^{n_c}}{k!} \tag{7.1}$$

Then, the parameters are determined as following: firstly, assume that the average size of each session includes 96 slots, then, n_c is calculated by

$$n_c = \frac{B_{non}}{b} \tag{7.2}$$

where B_{non} is the total number of the slots used to transmit non time-sensitive traffic in a frame.

Besides, both the duration time of the sessions,$L_{session}$, and the time interval, $L_{interval}$, also satisfy exponentia distribution with the average value of $\overline{L}_{session}$ and $\overline{L}_{interval}$, respectively. Thus

$$\rho = \frac{\overline{L}_{session}}{\overline{L}_{session} + \overline{L}_{interval}} \tag{7.3}$$

and,

$$\overline{L}_{interval} = \overline{L}_{session}\frac{1-\rho}{\rho} \tag{7.4}$$

where ρ is the utilization of the fiber used to transmit non time-sensitive traffic.

Furthermore, if processing speed of the node is 10G, thus, the length of the frame in DTM is about 19200 slots, among them, half frame is used to carry non time-sensitive traffic, which is constant, therefore, $B_{non} = 9600$ slots, and the size of the packets satisfy the uniform distribution, then the calling loss probability on different utilization can be simulated by varying the average session length and time interval. The results are shown in figure 7.11, the curve illustrate that the calling loss probability remarkably increases with the increasing of the utilization of the resource.

7.2.2 MPsLS

In MPsLS, since the non time-sensitive traffic is transmitted by filler asynchronous slots, the arrived slots can be buffered in queues, it hardly exists the calling loss for non time-sensitive applications if existing residual bandwidth.

7.2.3 Qualitative Evaluation

The qualitative comparison of the calling loss in different modes is shown in figure 7.2.

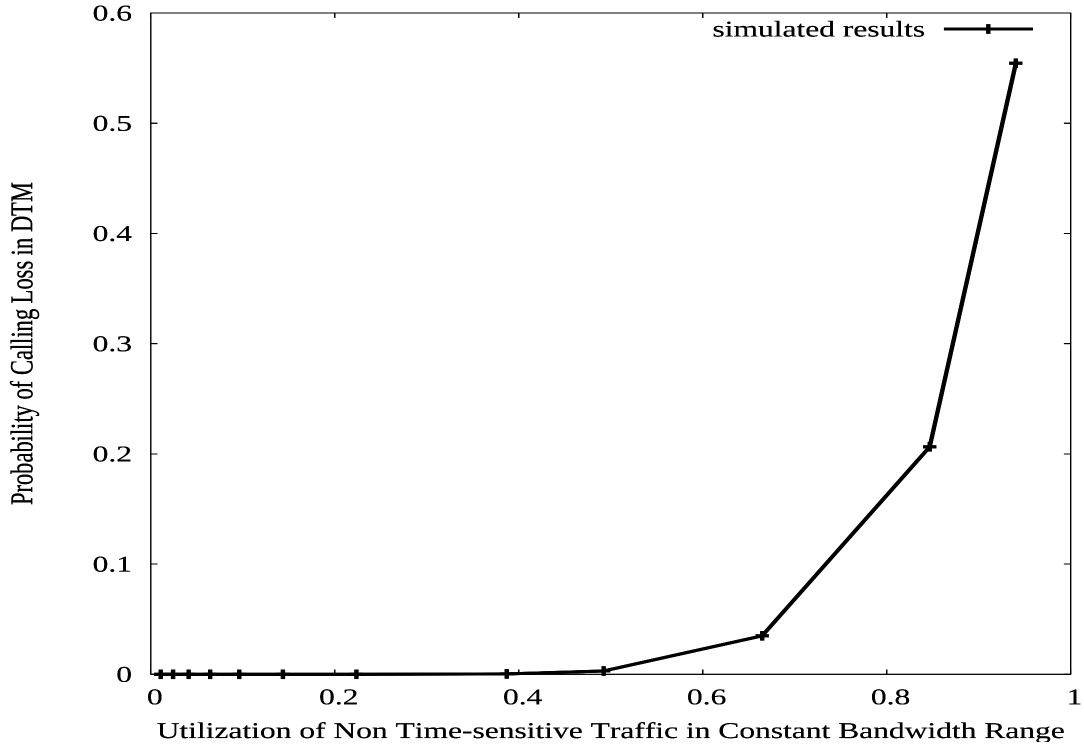

Figure 7.11: Probability of Calling Loss in DTM

Table 7.2: Calling Characteristic in Different QoS Improvement Schemes

Performance Evaluation / Transfer Mode / Type of Application	Asynchronous modes			Synchronous mode	
	DiffServ	MPLS	ATM	MPsLS	DTM
Time-sensitive applications	Fair*	Fair*	Fair**	Fair**	Fair**
Non Time-sensitive applications	Good	Good	Fair**	Good	Poor

Note: Fair* means only resource limit being required

Fair** means explicit calling connection being necessary

7.3 Summary

In this chapter, the simulation results in Sections 7.1 show that in datagram mode, since the size of the packets is variable, there is little larger delay and jitter, which affects the service performance of time-sensitive applications; in ATM model, because using fixed size cells, the bandwidth of CBR service class is reserved at peak bit rate, although the bandwith can be guaranteed, the processing time does not completely match with the arrivaing process, because of the feature of asynchronous transfer, the jitter still exists, the maximum jitter is variable with the scheduling mechanism and the number of CBR flows. Besides, ATM has a remarkable overhead caused by adding ATM header. Therefore, on the asynchronous service model, because of the undertimination of arriving process, the jitter is inevitable.

DTM is a synchronous model, which can remove the jitter, but it remains low resource utilization even if dynamical resource allocation mechanism is used because the idle slot positions can not be used by other flows during the sessions.

In MPsLS, time-sensitive traffic is transmitted synchronously by the slot positions fixed and reserved during the whole sessions, which limits the maximum delay and remove the jitter, while non time-sensitive traffic can use idle appointed slot positions, it remarkably increases the utilization. The qualititive comparison is shown in Table 7.1.

Similarly, Section 7.2 analyzes the calling loss, the qualititive comparison is shown in Table 7.2. In datagram mode, there is hardly strict access limitation, only requiring enough bandwidth, therefore, calling loss is zero; in virtual circuit mode, although it is necessary of explicit calling connection, the virtual circuit is used to improve transmitting speed and efficiency, the establishing of the channel needs only enough resource, the calling loss is almost zero; in synchronous mode, the calling loss can not be ignored, the difference between DTM and MPsLS is that the former uses synchronous channels to transmit non time-sensitive traffic similar to time-sensitive traffic, therefore, there is remarkable calling loss for non time-sensitive applications, the latter uses pesudo-synchronous channel to transmit non time-sensitive traffic similar to vircuit circuit mode, the calling loss for non time-sensitive applications is near zero.

Therefore, MPsLS has potential power to provide better service quality than other switching schemes. But the cost is the increasing of the average delay at low utilization.

Chapter 8

Conclusion

In this paper, we proposed a new data forwarding scheme used to the integrated data network: MPsLS.

In MPsLS network, data are transferred in cyclic mode with a constant period($\tau = 125\mu s$) called a frame. A frame consists of a number of slots with the same size(512 bits). Some slots in the header part of a frame are used as control slots, which dedicates to communicate with neighbouring nodes and to exchange the network control information for routing and setting up connections. The slots in the remaining part of a frame are data slots, which carry the application data and temporary control message. And a slot carries only a segment of a layer-3 packet.

MPsLS uses position-fixed slots to forward synchronously time-sensitive application flows and floating position slots to forward asynchronously non real-time application flows. The channels consisting of position-fixed slots are called appointed channels, those including variable position slots are called filler channels.

Furthermore, appointed channels are divided into the exact synchronization model and the less strict synchronization model. In exact synchronization model the distribution of the positions of all slots on appointed channels is same along the connection path; while in less strict synchronization model the distribution of slot positions may be different along the connection path, although the positions are fixed in each at whole session time.

The appointed channels implement resource reservation at per-flow, each channel only dedicates to one time-sensitive application, the different channels are isolated each other, once an application gets the time slot, which can not been token away by other applications till it is tear down after the session data are entirely finished, thus, the traffic state in one channel does not impact that in other channels, therefore, each time-sensitive application flow can receive expected service quality if the connection channel is established according to the QoS requirements.

In addition, due to tag bits in each slot are introduced to differ the types of data carried by the slots, the idle slots on appointed channels can be used to carry non time-sensitive data temporarily, this mechanism improves the utilization of network resource.

Accordingly, MPsLS combines the advantages of synchronous transfer mode and flexibility of label switching technology. Besides, MPsLS also has following performances:

- **High Access Performance and Ratio of Appointed Slots**

  Analytic formula and calculation results show that MPsLS has good access performance for less strict synchronous model, if allowable offset range is 200, the calling success probability can reach to near 100% for a new calling with 25 slots even if the utilization being 0.9. Furthermore, the ratio of appointed slots to total bandwidth can reach to 90%, even on the worst case, the value also exceeds 80%.

- **Small Delay and Jitter**

  In core nodes, MPsLS transmits all data in slot level synchronous mode, for exact synchronization channels, the delay passing each node is only several slot periods (about one percent of the frame periods), even for less strict synchronization channels with offset range of 200 slots, the value is only 10% of the frame periods, which is obviously less than the delay in DTM model. The delay jitter is 0 at each core nodes.

- **Low Transmission Loss**

  For time-sensitive traffic, at ingress nodes, due to usually equips longer queues for each flow, the loss can drop to very low level, specially, if reserving the channel at peak bit rate, the loss will almost disappear. At core area, since the outgoing slots are reserved, the congestion does not take place, although there exists waiting in less strict synchronous model, all appointed slots can be accommodated in the reorder buffer, therefore, there is no loss. In a ward, MPsLS can control loss according to time-sensitive application QoS requirement during transmission.

  For non time-sensitive traffic, MPsLS does not lead to extra dropping compared with usual IP networks, because MPsLS equips a long queue to store data when the packet burst takes place.

- **Large Bandwidth**

  MPsLS provides large bandwidth to support large bit rate service, if limiting the offset range within 200 slots, the maximum bit rate to per-flow can reach to $4M \times 200 = 800M$, it can satisfy any current applications.

- **Supporting Multicast of Appointed Slots**

  In MPsLS network, appointed channels transmission naturely support multicast, because once the appointed channel is established, the only thing need to be done for implementing multicast is copying the incoming slot to several outgoing slots, while the data carried by the slot need not be changed, including tag bits.

In contrast to other QoS improvement schemes, such as DiffServ and CBR service class of ATM mode, in principle, which are asynchronous transfer based on priority scheduling mechanism, it is inevitable of temporary congestion or contention for resource on the case of time-division multiplexing while the arrival process is stochastic. However, MPsLS adopts synchronous slot transfer and reserves not only bandwidth but also slot positions in the frames for time-sensitive flows, therefore, it avoids the jitter during transmission, the fluctuation of average delay is predictable, it can be controlled according to QoS requirement.

Appendix A: Speed Requirement to MPsLS Core Nodes

MPsLS is a novel switching scheme based on synchronous mode, therefore, it is necessary of the switch having enough bandwidth to support the synchronous slots being forwarded within each slot period.

A.1 Architecture Prototype

The architecture prototype of MPsLS core switch is shown in figure 8.1, the implementation is as following:

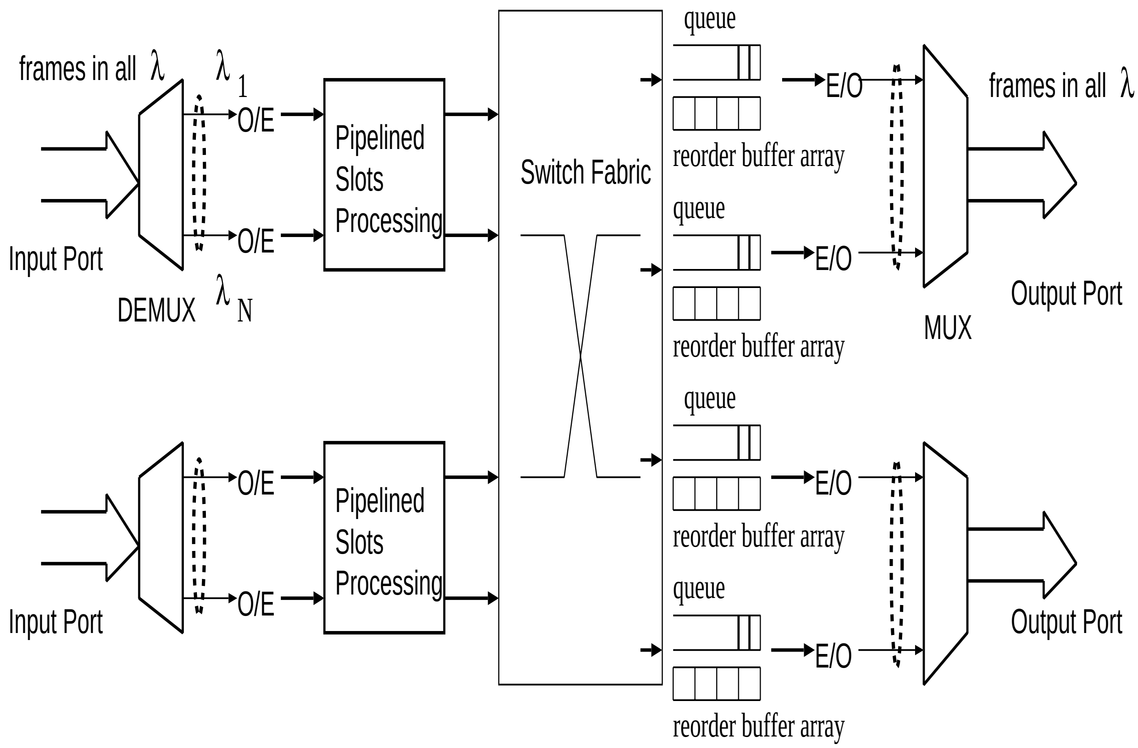

Figure 8.1: Prototype of MPsLS Core Switch

In the input ports of MPsLS core node, firstly, in each slot period, arrived optical signal is demultiplexed according to different λ, then, those optical signals corresponding to different λ are converted into electronical signal synchronously, the electronical signal data (slots) are processed and sent to corresponding output ports across the switch fabric, while at output ports, firstly, the electronical slots storing in queue and reorder buffer array are read out according to the positions in the frames, then they are converted into optical signal with determined λ, at last, multiplexing all the optical signals with different λ into one output optical signal.

A.2 Feasibility Analysis

In the architecture introduced in section A.1, currently, the wavelength division demultiplexing and multiplexing are not bottleneck, the processing also can be performed by multiple pipelines, the possible bottleneck is bandwidth of switch fabric, or width of the data bus, if the fiber is single channel with 10Gb/s, the number of input ports is 20, the required bandwidth is 200Gb/s, the increasing of the switch bandwidth can be realized by increasing the stages of switch matrix [111, 112], such as in Clos-network switch [110], a four-stages network can be scaled up to 320Gb/s; and a five-stages network may be scaled up to 2Tb/s, which can satisfy the requirement of the fiber having 10 channels, if the bandwidth of each channel is 10Gb/s.

A.3 Mechanism of Reading out Data from Reorder Buffer Array

On less strict synchronous channels mode, it is possible that more than one appointed slots arrive in each slot period, which can cause the writing contestion if using only one reorder buffer corresponding to each output port. Accordingly, we suggest using reorder buffer array to replace single reorder buffer in implementation. The reorder buffer array includes several same reorder buffers, the value equals the number of the input ports, it means that each input port utilizes a reorder buffer. At the headers of reorder buffers, the logical "OR" is performed among all slots at bits, then the result is sent to the outgoing frame, because at any slot period corresponding to each slot position in outgoing frame, there is at most one slot being non-zero. This mechanism can avoid the writing contestion in reorder buffer.

Bibliography

[1] "Integrated Services Digital Network (ISDN) Overview", Internetworking Technology Handbook. $http://www.cisco.com/univercd/cc/td/doc/cisintwk/ito\_doc/isdn.htm$.

[2] William A. Flanagan, "ISDN: A Practical Guide To Getting Up and Running", ISBN: 0936648813, Cmp Books, 1995.

[3] Gary C. Kessler and Peter V. Southwick, "ISDN: Concepts, Facilities, and Services", 4th ed. ISBN 0-07-034437-X, McGraw-Hill, 1998.

[4] Anthony S. Acampora, "An Introduction to Broadband Networks: LANs, MANs, ATM, B-ISDN, and Optical Networks for Integrated Multimedia Telecommunication", ISBN: 0306445581, Springer, 1994.

[5] William Stallings, "ISDN and Broadband ISDN with Frame Relay and ATM", ISBN: 0-13-973744-8, Prentice Hall, 1999.

[6] "Broadband ISDN Communications". $http://www.geocities.com/SiliconValley/1047/bisdn.html$.

[7] "Broadband ISDN: Broadband Integrated Services Digital Network (BISDN)". $http://www.javvin.com/protocolBISDN.html$.

[8] Ender Ayanoglu and Nail Akar, "B-ISDN (Broadband Integrated Services Digital Network)". $http://repositories.cdlib.org/cgi/viewcontent.cgi?article=1001\&context=cpcc$.

[9] Koichi Asatani, "Introduction to ATM Networks and B-ISDN", ISBN: 0471967661, John Wiley & Sons; 1st edition, 1997.

[10] "IP-based Networks: Basics", white paper. $http://www.axis.com/documentation/whitepaper/ip\_networks\_basics.htm$.

[11] "Designing Large-Scale IP Internetworks". http://www.cisco.com/univercd/cc/td/doc/cisintwk/idg4/nd2003.htm.

[12] B. Koch, D. Goderis , "QoS: Quality of Service for IP networks". $http://www.ngni.org/qos.htm\sharp\_Toc862363$.

[13] "QoS in IP networks". $http://www.csd.uoc.gr/\ hy536/ip\_qos.pdf$.

[14] Arne Lie, Leif Arne Ronningen, "Distributed Multimedia Plays with QoS guarantees over IP", Third International Conference on WEB Delivering of Music (WEDELMUSIC'03) pp. 12-20.
$http://doi.ieeecomputersociety.org/10.1109/WDM.2003.1233866$.

[15] Xipeng Xiao and Lionel M. Ni, "Internet QoS: A Big Picture".
$http://www.cs.virginia.edu/\ cs757/papers/xiao99internet.pdf$.

[16] Zheng Wang, "Internet QoS: Architectures and Mechanisms for Quality of Service", ISBN: 1558606084, Morgan Kaufmann, 1st edition, 2001.

[17] Dimitrios Miras, "A Survey on Network QoS Needs of Advanced Internet Applications", QoS working group, 2002.
$http://qos.internet2.edu/wg/apps/fellowship/Docs/Internet2AppsQoSNeeds.html$.

[18] Cristina Aurrecoechea, Andrew T. Campbell, Linda Hauw, "A Survey of QoS Architectures", 4th IFIP International Conference on Quality of Service, Paris, France, March 1996. $http://citeseer.ist.psu.edu/article/aurrecoechea96survey.html$.

[19] Daniel Reid and Michael Katchabaw, "Internet QoS: Past, Present and Future", Technical Report, 2004.
$http://www.csd.uwo.ca/\ katchab/pubs/tr\_internetqos.pdf$.

[20] "Specification of the Controlled-Load Network Element Service", RFC 2211.
ftp://ftp.isi.edu/in-notes/rfc2211.txt.

[21] "Specification of Guaranteed Quality of Service", RFC 2212.
ftp://ftp.isi.edu/in-notes/rfc2212.txt.

[22] "General Characterization Parameters for Integrated Service Network Elements", RFC 2215. ftp://ftp.isi.edu/in-notes/rfc2215.txt

[23] Paul P. White, "RSVP and Integrated Services in the Internet: A Tutorial".
http://www.cs.tcd.ie/ htewari/papers/white97rsvp.pdf.

[24] Jussi Laukkanen, "Integrated Services".
$http://www.cs.helsinki.fi/u/kraatika/Courses/QoS00a/jussilaukkanen.pdf$.

[25] "Resource ReSerVation Protocol (RSVP) - Version 1 Functional Specification", RFC 2205. ftp://ftp.isi.edu/in-notes/rfc2205.txt.

[26] "Resource ReSerVation Protocol (RSVP) - Version 1 Applicability Statement Some Guidelines on Deployment", RFC 2208. ftp://ftp.isi.edu/in-notes/rfc2208.txt.

[27] "Resource ReSerVation Protocol (RSVP) - Version 1 Message Processing Rules", RFC 2209. ftp://ftp.isi.edu/in-notes/rfc2209.txt.

[28] "The Use of RSVP with IETF Integrated Services",RFC 2210.
ftp://ftp.isi.edu/in-notes/rfc2210.txt.

[29] K. Nichols and S. Blake, "Definition of the Differentiated Services Field (DS Field) in the IPv4 and IPv6 Headers", RFC 2474, 1998.

[30] S. Blake and D. Black, "An Architecture for Differentiated Services", RFC 2475, 1998.

[31] Chris Metz, "Differentiated Services," IEEE MultiMedia, vol. 07, no. 3, pp. 84-90, Jul-Sept, 2000.

[32] V. Jacobson and K. Nichols, "An Expedited Forwarding PHB", RFC 2598, 1999.

[33] Nandy B., Seddigh N., Pieda P.,"Diffserv's Assured Forwarding PHB: What Assurance does the Customer Have?" NOSSDAV' 99.

[34] J. Heinanen and F. Baker, "Assured Forwarding PHB Group", RFC 2597, 1999.

[35] R. Callon et al., "A Framework for Multiprotocol Label Switching," September 1999.

[36] B. Davie and Y. Rekhter, "MPLS: Technology and Applications," Morgan Kufmann, 2000.

[37] E. Rosen et al., "MPLS Label Stack Encoding," RFC-3032, January 2001.

[38] E. Rosen, A. Viswanathan and R. Callon, "Multiprotocol Label Switching Architecture," RFC-3031, January 2001.

[39] G. Swallow, "MPLS Advantages for Traffic Engineering," IEEE Communication, December 1999.

[40] D. Awduche, "MPLS and Traffic Engineering in IP Networks," IEEE Communication, December 1999.

[41] B. Jamoussi et al., "Constraint-Based LSP Setup using LDP," work in progress, draft-ietf-mpls-cr-ldp-05.txt, February 2001.

[42] D. Awduche et al., "Requirements for Traffic Engineering over MPLS," RFC 2702, September 1999.

[43] D. Awduche et al., "RSVP-TE: Extensions to RSVP for LSP Tunnels," draft-ietf-mpls-rsvp-lsp-tunnel-08.txt, February 2001.

[44] Le Faucheur et al., "MPLS Support of Differentiated Services," work in progress, draft-ietf-mpls-diff-ext-08.txt, February 2001.

[45] Eric Horlait, Nicolas Rouhana, "Differentiated Services and Integrated Services Use of MPLS," iscc, p. 194, Fifth IEEE Symposium on Computers and Communications (ISCC 2000), 2000.

[46] M. A. Bauer, H. A. Akhand, " Managing Quality of Service in Internet Applications Using Differentiated Services, " JNSM: Vol. 10, No. 1, 2002.

[47] V. Fineberg, " QoS Support in MPLS Networks, "May 2003.

[48] K. Siu and R. Jain, "A Brief Overview of ATM: Protocol Layers, LAN Emulation, and Traffic Management", Computer Communications Review (ACM SIGCOMM), Vol. 25, No. 2, April 1995.

[49] A. Alles, "ATM Internetworking", Cisco Systems Inc. White Paper, May 1995. *http://www.cisco.com/warp/public/614/12.html*.

[50] A. Tanenbaum, "Computer Networks", Prentice Hall, 3rd Ed., 1996.

[51] T.M. Chen, S.S. Liu, "ATM Switching Systems", ISBN 0-89006-682-5, Artech House, Inc. 1995.

[52] G. Kessler, "An Overview of ATM Technology", January 1995. *http://www.hill.com/personnel/gck/atm_overview.html*.

[53] ATM Forum, "ATM User Network Interface (UNI) Specification Version 3.1", ISBN 0-13-393828-X, Prentice Hall, Englewood Cliffs, NJ, June 1995.

[54] U. Black, "ATM: Internetworking with ATM", Prentice Hall, Inc. 1997.

[55] M. de Prycker, R. Peschi and T. Van Landegem, "B-ISDN and the OSI Protocol Reference Model", IEEE Network, March, pp 10-18, 1993.

[56] D. McDysan, and D. Sophn, " ATM Theory and Applications ", McGraw Hill, 1999, pp. 585-660.

[57] Sonia Fahmy, Raj Jain, Rohit Goyal, Bobby Vandalore, Shivkumar Kalyanaraman, Sastri Kota, and Pradeep Samudra, " Feedback Consolidation Algorithms for ABR Point-to-Multipoint Connections in ATM Networks ", Infocom 98, San Francisco, March 1998, vol. 3 PP. 1004-1013.

[58] Sonia Fahmy, " Traffic Management for Point to Point and Multipoint Available Bit Rate (ABR) Service is asynchronous Transfer Mode (ATM) Networks ", PHD Dissertation, 1999.

[59] The ATM Forum, "Traffic Management Specification Version 4.1", March 1999.

[60] Network Equipment Technologies, Inc., "Introduction to ATM Traffic Management", 1998.

[61] R. Jain, "Congestion Control and Traffic Management in ATM Networks: Recent Advances and A Survey", Computer Networks and ISDN Systems, Vol. 28, No. 13, October 1996.

[62] S. Kalyanaraman, "Traffic Management for the Available Bit Rate (ABR) Service in Asynchronous Transfer Mode (ATM) networks", Ph.D. Dissertation, Dept. of Computer and Information Sciences, The Ohio State University, August 1997.

[63] "An Introduction to Circuit Emulation Services", White Paper, Cisco Systems.

[64] ATM white papers, "Speaking Clearly with ATM - A practical guide to carrying voice over ATM". *http://www.mfaforum.org/education/2.shtml*.

[65] Raj Jain, "Congestion Control and Traffic Management in ATM Networks: Recent Advances and A Survey", ACM Computer Networks and ISDN Systems, Vol. 28, No. 13, pp. 1723-1738, October 1996.
http://www.cse.wustl.edu/ jain/papers/cnis.htm
or *ftp://netlab.ohio-state.edu/pub/jain/papers/cnis.zip*.

[66] J. Murphy, "Resource Allocation In ATM Networks", Dublin City University School of Electronic Engineering, March 1996.
http://www.eeng.dcu.ie/murphyj/the/the/the.html.

[67] Network Equipment Technologies, "Advanced Traffic Management for Multiservice ATM Networks: An NET White Paper", 1998.
http://internet.net.com/repository/white_papers/adtmatm_wp.

[68] J. K. Ng, S. Song, C. Li, and W. Zhao, "A new method for integrated end-to-end delay analysis in ATM networks", Journal of Communications and Networks.
http://citeseer.ist.psu.edu/294539.html.

[69] J. Ng, S. Song, C. Li and W. Zhao, "Integrated End-to-end Delay Analysis in ATM Networks", Technical Report, Dept. of Computer Science, Hong Kong Baptist University, July 1998. (http://www.comp.hkbu.edu.hk/~jng/Tech-Rpt).
http://citeseer.ist.psu.edu/ng98integrated.html.

[70] Bobby Vandalore, Sonia Fahmy, Raj Jain, Rohit Goyal and Mukul Goyal, "QoS and Multipoint support for Multimedia Applications over ATM ABR service," IEEE Communications Magazine, Vol. 37, No. 1, pp. 53-57, January 1999.
http://www.cse.wustl.edu/~jain/papers/multabr.htm.

[71] Bobby Vandalore, Raj Jain, Rohit Goyal, Sonia Fahmy, "Dynamic Queue Control Functions for ATM ABR Switch Schemes: Design and Analysis," Computer Networks, August 1999, Vol. 31, Issue 18, pp. 1935-1949.
http://www.cse.wustl.edu/~jain/papers/cnis_qctrl.htm.

[72] Bobby Vandalore, "Traffic Management to Enhance Quality of Service (QoS) of Multimedia over Available Bit Rate (ABR) Service in Asynchronous Transfer Mode (ATM) Networks," PhD Dissertation, The Ohio State University, June 2000.
http://www.cse.wustl.edu/~jain/theses/bobby.htm.

[73] G. Babic, R. Jain, A. Durresi, "ATM Performance Testing and Quality of Service Management," in F. Golshani and F. Groom, Ed., The ATM Handbook Published by International Engineering Consortium, Chicago, IL, 2000, ISBN:0-933217-63-3.

[74] Sonia Fahmy, Raj Jain, Sameh Rabie, Rohit Goyal and Bobby Vandalore, "Quality of Service for Internet Traffic over ATM Service Categories, " Computer Communications, 15 September 1999, Vol. 22, Issue 14, pp. 1307-1320.
http://www.cse.wustl.edu/~jain/papers/enterprs.htm.

[75] R. Ballart and Y. C. Ching, "SONET: Now It's the Standard Optical Network", IEEE Commun. Mag, Vol. 29, No. 3, March 1989.

[76] M. J. Karol et al. AT & T Bell Laboratories, "Hierarchical Gigabit ATM Switching: Applications and Performance", Proc. of XIV Int. Switching Symposium 1992.

[77] A. Hac, H.B. Mutlu, "Synchronous Optical Network and Broadband ISDN Protocols", IEEE Computer, Nov. pp26-34, 1993.

[78] G. Sjodin, "An algorithm for handling slot allocation in DTM", Proc. of the 4th-MultiG Workshop, Stockholm, May 12, 1992.

[79] P. Lindgren, "Host Interfacing and Connection Management in the DTM Gigabit Network", Licentiate Thesis, Royal Institute of Technology, March 1994.

[80] P. Lindgren and C. Bohm, "Fast connection establishment in the DTM gigabit network, " in Proc. 5th High Performance Networking (HPN), (Grenoble, France), IFIP, June 1994, pp. 283–294. $http://citeseer.ist.psu.edu/lindgren94fast.html$.

[81] C. Bohm, P. Lindgren, L. Ramfelt P. Sjdin, "Resource Reservation in DTM", 1st IEEE Symposium on Global Data Networking, Cairo, Dec. 1993. $http://citeseer.ist.psu.edu/bohm93resource.html$.

[82] C. Bohm, "The DTM Protocol—Design and Implementation", Licentiate Thesis, Royal Institute of Technology, ISRN KTH/IT/R–94/05–SE, Stockholm, Sweden, Feb. 1994. $http://citeseer.ist.psu.edu/article/bohm94dtm.html$.

[83] C. Bohm, P. Lindgren, L. Ramfelt, and P. Sjodin, "The DTM Gigabit Network," Journal of High Speed Networks, vol. 3, no. 3, pp. 109–126, 1994. $http://citeseer.ist.psu.edu/bohm94dtm.html$

[84] L. Ramfelt, "Performance Analysis of Slot Management in DTM Networks", Technical Report TRITA-IT R 95:23, Dept. of Teleinformatics, KTH, Stockholm, January 1996. $http://citeseer.ist.psu.edu/ramfelt96performance.html$.

[85] C. Bohm, P. Lindgren, M. Hidell, L. Ramfelt, and P. Sjodin, "Fast circuit switching for the next generation of high performance networks," To appear in IEEE Journal on Selected Areas in Communications, vol. 14, Feb. 1996. $http://citeseer.ist.psu.edu/bohm96fast.html$.

[86] "Next Generation Networks in Europe-From ACTS to IST". $http://www.dit.upm.es/infowin/ngn/ngnbook.pdf$

[87] C.-C. Han, C.-J. Hou, and K. G. Shin, "On slot allocation for time-constrained messages in DQDB networks," in Proc. of INFOCOM'95, April 1995. $http://citeseer.ist.psu.edu/han95slot.html$.

[88] Csaba Antal, Jozsef Molner, Sandor Molner, and Gabor Szab. Performance study of distributed channel allocation techniques for a fast circuit switched network. 21(17):1597–1609, November 1998. $http://citeseer.ist.psu.edu/antal98performance.html$.

[89] Csaba Antal, Jozsef Biro, Tamas Henk, Gergely Matefi, "Performance Evaluation of a Time Division Multiplexing Method Applicable for Dynamic Transfer Mode Networks," iscc, pp. 34-50 Fifth IEEE Symposium on Computers and Communications (ISCC 2000), 2000.

[90] Csaba Antal and Sandor Molner, "Fairness of a DTM Dual-bus with Large Inter-node Distances", 8th International Conference on Telecommunication Systems, Modelling and Analysis, March 9-12, 2000, Nashville, Tennessee, USA.

[91] Csaba Antal and Sandor Molner, "Smoothing Algorithms for DTM Networks", Proceedings of the IFIP TC6 WG6.7 Sixth International Conference on Intelligence in Networks: Telecommunication Network Intelligence, Vol. 178, pp. 163-180, 2000.

[92] Chih-Jen Chang and Arne A. Nilsson, "Performance Evaluation of DTM Access Nodes". *http : //www.ece.ncsu.edu/cacc/download_techreport.php?id* = 41.

[93] Chih-Jen Chang and Arne A. Nilsson, "Fair Efficient Call Admission Control Policies for Heterogeneous Traffic Streams in a Packet Routing Server Using the DTM Technology", Proceedings of the IFIP-TC6 / European Commission International Conference on Broadband Communications, High Performance Networking, and Performance of Communication Networks, Vol. 1815, pp. 620 - 631, 2000.

[94] ATM Forum, "Circuit Emulation Service Interoperability Specification Version 2.0", January 1997. *ftp : //ftp.atmforum.com/pub/approved − specs/af − vtoa −* 0078.000.*pdf*

[95] ATM Forum, "Specifications of (DBCES) Dynamic Bandwidth Utilization - In 64Kbps Time Slot Trunking Over ATM - Using CES" July 1997. *ftp : //ftp.atmforum.com/pub/approved − specs/af − vtoa −* 0085.000.*pdf*

[96] ATM Forum, "ATM Trunking using AAL2 for Narrowband Services", February 1999. *ftp : //ftp.atmforum.com/pub/approved − specs/af − vtoa −* 0113.000.*pdf*

[97] J. Leguay, M. Latapy, T. Friedman and K. Salamatian, Describing and Simulating Internet Routes, Preprint cs.NI/0411051, 2004.

[98] J. Yang and Y. Hibino, *Analysis of Delay Characteristics in MPsLS Forwarding Scheme*, IEICE Trans. on Communications, Vol. E89-B, No. 6, pp. 1738-1746, June 2006.

[99] N.T.J.Bailey,*On queueing processes with bulk service*, J. R. Statist. Soc. B, 16, pp. 80-87, 1953.

[100] Mutlu Yaglioglu and Rudi Distler,*Simulation of Hierarchical Scheduler for ATM Switches*, 8th IEEE Symposium on Computer and Comunications, iscc, pp. 535-640, 2003.

[101] K. Murakami and M. Maruyama, *MAPOS - Multiple Access Protocol over SONET/SDH Version 1*, RFC 2171, June 1997.

[102] M. Maruyama and K. Murakami, *MAPOS Version 1 Assigned Numbers*, RFC 2172, June 1997.

[103] K. Murakami and M. Maruyama, *A MAPOS version 1 Extension - Node Switch Protocol*, RFC 2173, June 1997.

[104] K. Murakami and M. Maruyama, *A MAPOS version 1 Extension - Switch-Switch Protocol*, RFC 2174, June 1997.

[105] K. Murakami and M. Maruyama, *MAPOS 16 - Multiple Access Protocol over SONET/SDH with 16 Bit Addressing*, RFC 2175, June 1997.

[106] K. Murakami and M. Maruyama, *IPv4 over MAPOS Version 1*, RFC 2176, June 1997.

[107] S. Shimizu, T. Kawano, K. Murakami and E. Beier, *MAPOS/PPP Tunneling mode*, RFC 3186, December 2001.

[108] O. Okamoto, M. Maruyama and T. Sajima, *Forwarding Media Access Control (MAC) Frames over Multiple Access Protocol over Synchronous Optical Network/Synchronous Digital Hierarchy (MAPOS)*, RFC 3422, November 2002.

[109] T. Ogura, M. Maruyama and T. Yoshida, *Internet Protocol Version 6 over MAPOS (Multiple Access Protocol Over SONET/SDH)*, RFC 3572, July 2003.

[110] E. Oki, Z. Jing, R. Rojas and H. J. Chao, *Concurrent Round Robin Dispatching Scheme for a Clos-Network Switches*, Prof. IEEE ICC 2001, Helsiki, Finland, June 2001.

[111] H. J. Chao, K. L. Deng and Z. Jing, *A Petabit Photonic Packet Switch*, $http://www.ieee-inforcom.org/2003/papers/19\_03.PDF$.

[112] S. Blau and J. Rooth, *AXD 301- A new generation ATM switching system*, Ericsson Review No. 1, 1998.

[113] J. R. Smith, *MPEG-21 Digital Item Adaptation: Enabling Universal Multimedia Access*, Multimedia, IEEE, Vol. 11, Issue 1, pp. 84-87, 2004.

Publications

[1] J. Yang and Y. Hibino: "Analysis of Delay Characteristics in MPsLS Forwarding Scheme", IEICE Trans. on Communications, Vol. E89-B, No. 6, pp. 1738-1746, June 2006.

[2] Yang, Jun and Satoshi Mizuta: "Detailed Analysis of Uphill Moves in Temperature Parallel Simulated Annealing and Enhancement of Exchange Probabilities", Complex Systems, Vol.15, pp. 349-358, Oct. 2005.

[3] J. Yang and Y. Hibino: "MPsLS: A Forwarding Scheme of Guaranteeing QoS in Integrated Services Networks", Technical Report of IEICE, CS2004-100, Vol. 104, No. 493, pp. 21-26, Dec. 2004.

[4] J. Yang and Y. Hibino, *A Novel Switching Scheme Improving Per-flow QoS for Time-sensitive Applications*, in preparing.