UNIVERSITI PUTRA MALAYSIA

TRAFFIC CONTROL IN A SYNCHRONOUS TRANSFER MODE NETWORKS

IBRAHIM KHALIL

TRAFFIC CONTROL IN ASYNCHRONOUS TRANSFER MODE NETWORKS

By

IBRAHIM KHALIL

Thesis Submitted in Fulfillment of the Requirements for
the Degree of Master of Science in the
Faculty of Engineering
Universiti Pertanian Malaysia

September 1995

ACKNOWLEDGEMENTS

First and foremost, I would like to express my earnest gratitude and heart-felt thanks to the chairman of my supervisory committee, Associate Professor Dr. Borhanuddin Mohd Ali, for his guidance and support throughout my research. After drafting me into the MS programme, he helped me find prestigious graduate assistantship for more than a year. He introduced me to the field of high speed networks and laid the foundations of ATM network, upon which this thesis is based. Time and again, his insightful comments helped me to refine ideas, and his impromptu suggestions opened new and exciting avenues for further research. Dr. Borhan was encouraging and sympathetic even during times when I was confused and unsure about my research.

It is also a great honour and pleasure to acknowledge Dr. Abdul Rahman Bidin, another member of the supervisory committee, for his technical support, helpful suggestions and insights. Dr. Abdul Rahman suggested that I apply the concepts of Neural networks to ATM call admission control, which led my research effort in this area. His easy going manner and unflagging team spirit made him a pleasure to work with.

Most of conference and journal articles relating to the research reported in this dissertation carry a third or fourth name on the byline. I was really lucky to have Dr. Malay Raj Mukerjee, an associate professor at UPM, as a collaborator, and sometimes as a criticizer. His helpful comments on bandwidth allocation was simply great.

I had help from many other people over internet. Those making helpful comments are literally too numerous to mention. However, I would like to acknowledge particularly the advice and support of Dr. D. C. Verma, formerly a Univ. of California student and now working at IBM. His Ph.D. dissertation, traffic generator and regular critical comments through e-mails helped me a lot to improve my work on characterizing variation of traffic parameters. Thanks also go to Ramesh Nagarajan, for his Ph.D. dissertation and explanation on Quality-Of-Service; Hiroshi Esaki for papers on Call Admission Control (CAC) and pointing out to me the real research issues on this subject; J. Neves, Hiromatsu, Nordstorm for sending their works on CAC using neural network. I received a few other research related Ph.D. dissertations from Andreas Pitsilledis of Swinburne Univ. of Tech., Paul Skelly of Columbia university, Hui Zhang of University of California at Berkeley. All these greatly influenced my research work.

Thanks are also due to many individuals who helped me in many ways. I am grateful to Prof. Tatsuya Suda and Prof. Aurel Lazar for their generous help while they visited UPM during my early period of research. Prof. Fred Halsall spent considerable amount of time advising me on the organization of the dissertation.

Last but not least, I must mention the tremendous support I received from my parents. I would not have finished this thesis without their support.

TABLE OF CONTENTS

CHAPTER

LIST OF TABLES

Table Page

LIST OF FIGURES

LIST OF ABBREVIATIONS

ABR - Available-Bit-Rate

ATM - Asynchronous Transfer Mode

B-ISDN - Broadband Integrated Services Digital Network

BPNN - Back-Propagation Neural Network

CAC - Call (Connection) Admission Control

CBR - Constant-Bit-Rate

CLP - Cell Loss Probability

CLR - Cell Loss Ratio

DT - Delayed Transmission

ECN - Explicit Congestion Notification

EWMA - Exponentially Weighted Moving Average

FCI - Forward Congestion Indicator

FRP - Fast Reservation Protocol

HRR - Hierarchical Round Robin

IP - Internet Protocol

IPP - Interrupted Poisson Process

IVCLP - Individual Virtual Cell Loss Probability

IT - Immediate Transmission

JW - Jumping Window

LB - Leaky Bucket

LAN	-	Local area Network
MLT	-	Maximum Laxity Threshold
MMPP	-	Markov Modulated Poisson Process
MV	-	Moving Window
NN	-	Neural Network
NNTEM	-	Neural Network Traffic Enforcement Mechanism
PBR	-	Peak- Bit- Rate
PBS	-	Partial Buffer Sharing
QLT	-	Queue Length Threshold
QOS	-	Quality of service
SE	-	Switching Element
SQ	-	Source Quench
STM	-	Synchronous Transfer Mode
TDM	-	Time Division Multiplexing
TCP	-	Transmission Control Protocol
TJW	-	Triggered Jumping Window
UPC	-	Usage Parameter Control
VBR	-	Variable Bit Rate
VC	-	Virtual Channel
VCI	-	Virtual Channel Identifier
VCLP	-	Virtual Cell Loss Probability
VP	-	Virtual Path

VPI - Virtual Path Identifier

Abstract of the thesis submitted to the Senate of Universiti Pertanian Malaysia in fulfillment of the requirements for the degree of Master of Science.

TRAFFIC CONTROL IN ASYNCHRONOUS TRANSFER MODE NETWORKS

By

Ibrahim Khalil

September 1995

Chairman: Associate Professor Dr. Borhanuddin Mohd Ali
Faculty : Engineering

In the 90s, there is an increasing demand for new telecommunication services such as video conferencing, videophone, broadcast television, image transfer and bulk file transfer etc. At the same time, transmission systems at bit rates of 2.5 Gb/s are now being installed, and the expected next generation of 10 Gb/s systems is emerging from the research laboratories. Coupled with that the development and deployment of new technologies systems such as fiber optics and intelligent high-speed switches have made it possible to provide these services in future high-speed integrated services networks like Asynchronous Transfer Mode (ATM). However, because of their new characteristics, these new services pose great challenges not previously encountered in traditional circuit-switched or packet switched networks. For example, features such as large propagation delay as compared to transmission delay, diverse application demands, constraints on call processing capacity, and Quality-Of-Service (QOS) support for different applications all present new challenges arising from the new technology and new applications. Thus, much research is needed not just to improve existing technologies, but to seek a fundamentally different approach toward network architectures and protocols. In particular, new

bandwidth allocation and call admission control algorithms need to be studied to meet these new challenges.

A VP bandwidth allocation problem is studied for services which requires guaranteed connection for a fixed duration of time leading to extensive use of facilities like reservations of transmission capacity in advance. In such a case, the network may offer discounts for users reserving capacities in advance due to the advantage of working with predetermined traffic loads. Similarly, charges may differ for customers wanting to book capacity for a specified time interval. Based on this scenario, various *charge classes* and *booking policies* are introduced. An effective bandwidth allocation scheme is proposed at the VP level with *multiple nested charge classes* where these various classes are allocated bandwidth optimally through some booking policies. The scheme is also shown to be effective in maximizing network revenue. The best tradeoff between revenue gained through greater demand for discount bandwidth units against revenue lost when full-charge bookings request must be turned away because of prior bookings of discount bandwidth units is also sought for.

To increase the average utilization of the network by real-time traffic, new Neural Network (NN) based Call Admission Control (CAC) schemes of various capabilities are introduced. These schemes are able to adapt gracefully to the dynamic behaviour of traffic and time-varying nature of network conditions, especially when new services are being continually introduced after network design and installation. Since traffic parameters in an ATM network may change over time from the values declared at the connection setup, a Neural Network approach is proposed to effectively characterize the changes in traffic parameters declared during call set-up and establish a simple input-output relation from which values of output traffic parameters can be produced. These new parameters can be used for bandwidth allocation in the node downstream.

Abstrak tesis yang dikemukakan kepada Senat Universiti Pertanian Malaysia untuk memenuhi keperluan Ijazah Master Sains

KAWALAN TRAFIK DALAM RANGKAIAN MOD PEMINDAHAN TAKSEGERAK

Oleh

Ibrahim Khalil

September 1995

Pengerusi: Prof. Madya Dr. Borhanuddin Mohd. Ali
Fakulti : Kejuruteraan

Dalam dekad ke '90an, terdapat peningkatan permintaan untuk perkhidmatan telekomunikasi baru seperti persidangan video, telefon video, televisyen penyiaran, pemindahan imej dan pemindahan besar fail. Pada masa yang sama sistem penghantaran dengan kadar bit 2-5 Gb/s sedang dipasangkan manakala sistem generasi "masa depan" 10 Gb/s mula muncul dari makmal penyelidikan. Di samping itu kemunculan suis pantas pintar telah memungkinkan penawaran perkhidmatan tersebut dalam rangkaian pantas terkamir masa depan seperti Mod Pemindahan Taksegerak (Asynchronous Transfer Mode, ATM). Walau bagaimanapun, oleh kerana cirinya yang baru, perkhidmatan ini menimbulkan cabaran yang belum pernah dihadapi dalam rangkaian tradisional pengsuisan litar atau pengsuisan paket, seperti kelengahan perambatan yang tinggi berbanding dengan kelengahan penghantaran, permintaan aplikasi yang berbagai, pengekangan pada keupayaan pemprosesan panggilan dan sokongan kepada kualiti perkhidmatan untuk aplikasi yang berbagal. Oleh itu, tesis ini mengkaji algoritma penguntukan lebarjalur dan kawalan kebenaran masuk panggilan untuk menyahut cabaran ini

Masalah pengagihan Laluan Maya, (Virtual Path,VP) adalah dikaji untuk perkhidmatan yang memerlukan sambungan terjamin untuk jangka masa yang tetap menjurus kepada penggunaan kemudahan yang meluas seperti penempahan keupayaan awal. Dalam kes ini, rangkaian tersebut boleh menawarkan diskaun untuk pengguna yang menempah kapasiti terlebih dahulu kerana manafaat berurusan dengan beban lalu-lintas yang ditentukan terlebih dahulu. Begitu juga, bayaran boleh berbeza untuk pelanggan yang ingin menempah kapasiti untuk satu jangka masa yang tertentu. Berdasarkan kepada senario ini, beberapa kelas pembayaran dan polisi tempahan diperkenalkan. Suatu skim penguntukan lebarjalur di peringkat VP dengan kelas bayaran bergelung berbilang, melalui polisi penempahan yang ditunjukkan sebagai optimum di kalangan kelas bayaran tersebut dan berkesan dalam memaksimumkan hasil rangkaian, juga diperihalkan. Penggantian yang paling baik juga dicari di antara hasil pungutan yang diterima melalui permintaan yang meningkat untuk lebarjalur diskaun ke atas hasil pungutan yang luput apabila penempahan bayaran penuh terpaksa ditolak kerana terdapat penempahan lebarjalur diskaun terlebih awal.

Untuk meningkatkan penggunaan purata rangkaian oleh lalu-lintas masa-nyata, suatu skim pembenaran-masuk panggilan Rangkaian Neural telah diperkenalkan. Skim ini mampu menyesuaikan dengan berkesan kelakuan lalu-lintas yang dinamik dan keadaan sebenar rangkaian yang berubah mengikut masa, terutama sekali apabila perkhidmatan baru diperkenalkan setelah rekabentuk rangkaian dan pemasangannya selesai dibuat. Oleh kerana parameter lalu-lintas dalam rangkaian ATM mungkin berubah daripada nilai yang dimaklumkan pada peringkat penyambungan panggilan suatu pendekatan Rangkaian Neural telah dicadangkan untuk mencirikan perubahan parameter lalu-lintas tersebut yang dideklarasikan pada peringkat penyambungan panggilan tadi, dan seterusnya menetapkan suatu hubungan masukan/keluaran mudah yang mana nilai parameter lalu-lintas keluaran dapat diperolehi darinya. Parameter baru ini boleh digunakan untuk peruntukan lebarjalur di nod seterusnya.

CHAPTER I

INTRODUCTION

Communications systems have been revolutionized by technological advances in the last decade. The speed and capacity of various components in a communication system, such as transmission media, switches, memory, processors, have all followed technological curves that have grown either linearly or exponentially over the last ten years (Fraser, 1991). At the periphery of the network, driven by the same underlying technology- microelectronics, the capability of computers has been drastically increased while the cost has been significantly reduced. To take advantage of the technological trends, and to satisfy the growing need for people and computer to communicate, new integrated services networks are being designed. Unlike the current communications networks which are designed to offer a special type of service, the integrated-services network like Asynchronous Transfer Mode (ATM) will offer multiple services that include data, voice, video, and others. The integration of multiple services into one network will offer a number of advantages, including vast economies of scale, ubiquity of access, and improved statistical multiplexing. However, this integration also presents new challenges for network designers.

Traditional communication networks offer a single type of service that supports one type of applications. For example, since the telephone network has been designed to support interactive voice, which requires a low delay, low rate, fixed bandwidth, two-way, jitter-free service. Alternatively, the cable TV network has been designed to support broadcasting of analog video, which requires a high rate, fixed bandwidth, one-way, jitter-free service. Finally, the data network has been designed to support communication between computers. Although Quality-Of-Service (QOS) has been considered, the current data network only offers a best effort service - there are no guarantees on performance parameters such as delay or throughput. The specialization of each of these networks has allowed the network design to be optimized for that specific type of service.

ATM will have to support applications with diverse traffic characteristics and performance objectives. There have been many applications proposed, and there is no doubt that many more will emerge once the networks are in place. These applications include digital television, digital audio and facsimile transmission, and in general multimedia applications. The QOS expected from the network by these applications varies over a wide range: some are sensitive to delays experienced in the communication network, others are sensitive to loss rates, while yet others are sensitive to delay variations. An ATM network, which aims to support all these services, must attempt to meet the needs of all these applications.

ATM Network Environment

In an ATM network, the transfer of information is made in the form of cells of fixed length, regardless of the information conveyed. Operation is *asynchronous*, in the sense that cells do not have an identified position in time unlike the bytes of a synchronous link that occupy a fixed position within a frame. All communication in an ATM network is *connection-oriented,* i.e., a connection needs to be established before data transmission can begin. The asynchronous transfer mode is a more flexible alternative to the *synchronous* transfer mode (STM), in which cells belonging to a connection could only occupy certain predetermined time-slots on a transmission link. The difference between the two modes is illustrated by Figure 1, which shows some possible patterns of slot usage in the two modes. Four connections are assumed to exist.

In STM, time is divided into frames (marked by the shaded cells in the figure), and each connection gets one or more slots in a frame to transfer its cells. Thus connection 2 can only transfer cells in the first slot of a frame, connection 3 in the second slot, connection 4 in the 3rd slot and connection 1 in the fourth slot. In the central frame shown at the top of the figure, the slot for connection 1 is wasted because there is no cell to use it. In ATM, cells are allowed to use the slots as soon as they are available, and so the cell belonging to connection 2 can go ahead in the empty slot.

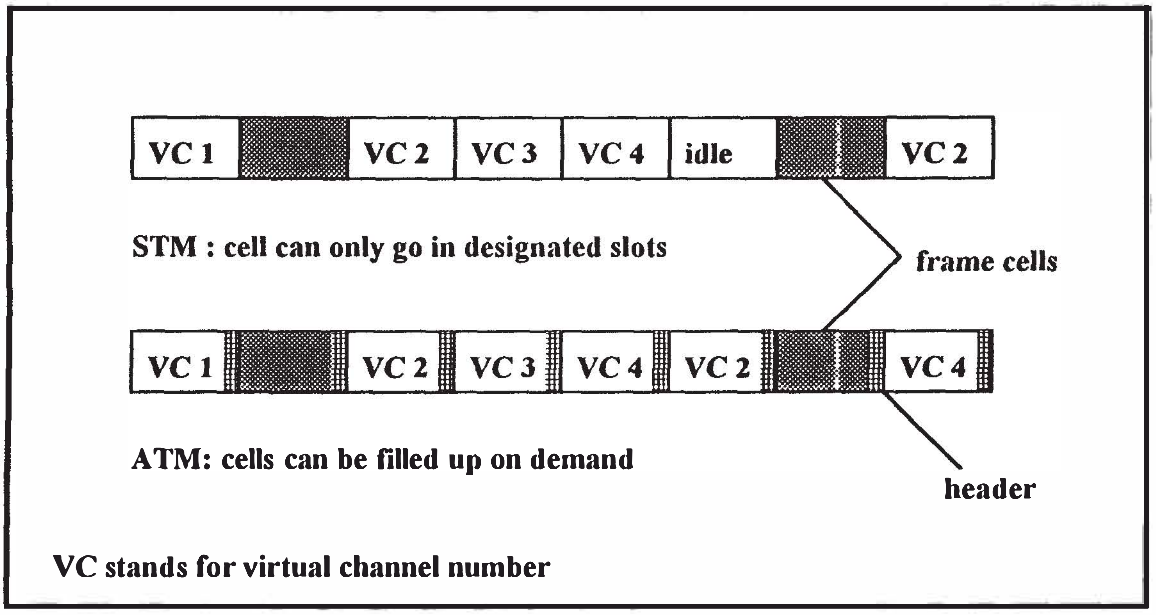

Figure 1: STM and ATM Principles

An ATM logical channel is identified by means of two hierarchical entities: the virtual path (VP) and the virtual channel (VC), each having a subfield in the header of every ATM cell, the virtual path identifier (VPI) and the virtual channel identifier (VCI). The two identifiers are used to establish a virtual link across the interfaces of an ATM connection. The virtual paths allow several virtual channels to be multiplexed or switched in a single block. In an ATM crossconnect network, the crossconnects (or

VP-switches) only handle the virtual path identifiers. Several VPs may be multiplexed on the same physical link. Figure 2 shows the concept of virtual paths. There are three virtual paths, VP1, VP2 and VP3, shown in the figure. For example, VP2 is a virtual path between end nodes N1 and N5. N3 and N4 are transit nodes of virtual path VP2. N3 is a crossconnect.

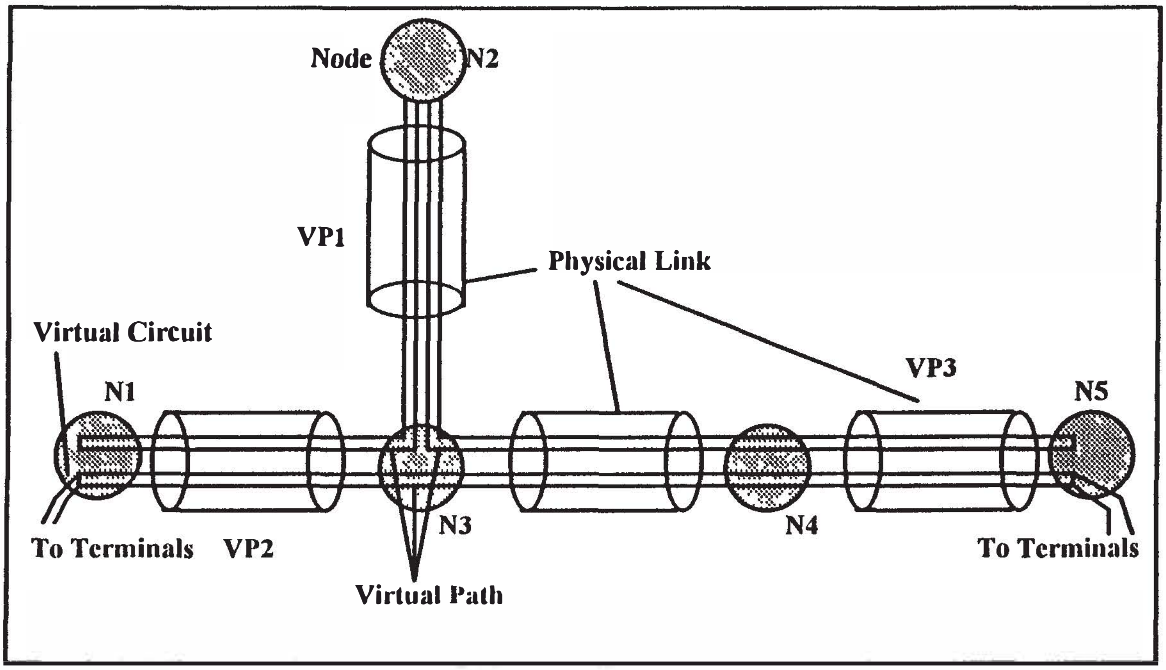

Figure 2 : Schematic Illustration of Virtual Paths

An ATM network consists of a number of ATM *switches* connected in an arbitrary mesh. It is assumed that each switch is accompanied by a *switch manager,* which is responsible for accepting or rejecting channels through the switch. The switch manager should be running the algorithms required to establish connections and perform any admission control tests required to ensure the quality of service needed by individual connections. When a channel is established, the switching fabric maps cells arriving along that channel from the input link to the appropriate output link. The exact details of this mapping mechanism is not a concern here. Any of the the common switching techniques may be used for the mapping operation. These techniques include shared-memory switching (Condruse, 1987), shared medium switching (Gopal, 1987) or space division switching fabrics (like banyan switches) (Hui, 1987). A survey of switching schemes can be found in some recent literature (Ahmadi, 1989; Tobagi, 1990). In this thesis, no assumptions are made regarding the switching scheme used, as long as the switch model proposed below is valid.

An ATM switch can be modeled as a set of input queues terminating the incoming links, connected through an interconnection network to a set of output queues (outgoing links). Usually the only purpose of the input queues is cell/bit synchronization and clock recovery; so input queue size is small. Nevertheless, it should be noted that statistical switching techniques will necessitate large input queue sizes in some cases. The output queues are needed since cells from different incoming links that are to be routed towards the same outgoing link may arrive concurrently. Each switch can be represented as in Figure 3.

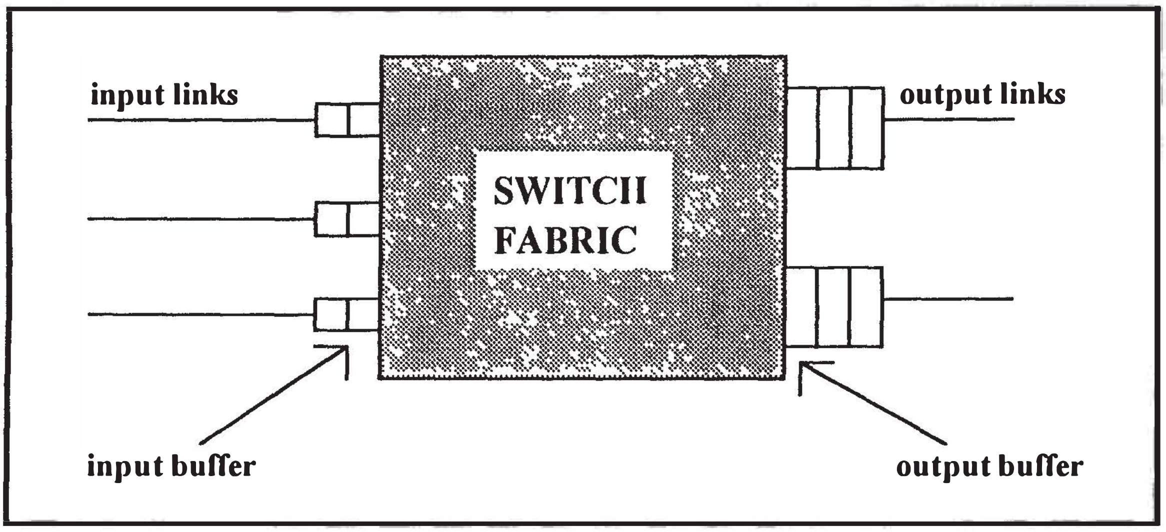

Figure 3: Structure of a Typical ATM Switch