UNIVERSITI PUTRA MALAYSIA

DYNAMIC BANDWIDTH ALLOCATION IN ATM NETWORKS

LOH CHEE HEOK

FSKTM 2001 16

DYNAMIC BANDWIDTH ALLOCATION IN ATM NETWORKS

By

LOH CHEE HEOK

Thesis Submitted in Fulfilment of the Requirement for the Degree of Master of Science in the Faculty of Computer Science and Information Technology Universiti Putra Malaysia

January 2001

With appreciation and respect,

this thesis is dedicated

To my dearest mom, brothers and sister,
who inspired me with confidence and ambitions.

To all my supervisors,
here and abroad,
who ensured it all worthwhile.

Abstract of thesis presented to the Senate of the Universiti Putra Malaysia in fulfilment of the requirement for the degree of Master of Science

DYNAMIC BANDWIDTH ALLOCATION IN ATM NETWORKS

By

LOH CHEE HEOK

January 2001

Chairman : **Dr. Mohamed Othman**

Faculty : **Faculty of Computer Science and Information Technology**

Today's new applications such as World Wide Web, video conferencing and multimedia have introduced a large amount of traffic into the network. Additionally new applications are also heading towards real time process. Instant access to the network, greater level of performances and higher degree of satisfaction has become the main concerns of users using these new applications. Although current transmission mediums have advanced in capacity through means such as optical fiber and Gigabit Ethernet, future and unknown new services tend to consume up the available bandwidth. ATM network is the new technology used to support a wide variety of services including data, voice, video and most possibly other future applications. Its flexibility, efficiency and high throughput have gained popularity but with greater complexity due to different approaches in handling different type of services.

A high-speed network such as ATM networks must have an effective traffic management scheme in order to gain high data throughput with the least cost of operation. Thus, simulation and modeling are the effective methods used to design the trade-off between network parameters and their performances. Effective sharing of network resources such as bandwidth and buffer are studied through the dynamic allocation method. Static allocation scheme has been proven inefficient to provide high resources utilization as can be seen in STM networks compared to ATM networks. However, ATM networks should provide different dynamic allocation methods according to its different services and traffic characteristics. Four dynamic allocation strategies have been designed, evaluated and compared for their performances. They are called Static Bandwidth Allocation, Bandwidth Allocated Proportional to Expected Queue Length, Bandwidth Allocated Proportional to Expected Queue Length with Threshold Value and Bandwidth Allocated with Threshold Interrupt. Bandwidth Allocated with Threshold Interrupt is proven to be the most effective strategy as it could response to congestion immediately.

Abstrak tesis yang dikemukakan kepada Senat Universiti Putra Malaysia sebagai memenuhi keperluan untuk ijazah Master Sains

PERUNTUKAN DINAMIK JALUR-LEBAR RANGKAIAN ATM

Oleh

LOH CHEE HEOK

Januari 2001

Pengerusi : Dr. Mohamed Othman

Fakulti : Fakulti Sains Komputer dan Teknologi Maklumat

Aplikasi terkini seperti Internet, persidangan video dan multimedia merupakan penyumbang utama terhadap kesesakan rangkaian komunikasi. Tambahan pula, aplikasi baru sebegini lebih menuju ke arah masa pemprosesan yang singkat. Faktor-faktor penggalak bagi para pengguna menggunakan aplikasi ini termasuk penimbangan ke atas kecepatan aplikasi tersebut mencapai rangkaian, kualiti persembahan dan tahap kepuasan yang tinggi. Walaupun medium perantaraan seperti fiber optik dan "Gigabit Ethenet" telah mempertingkatkan kapasiti penghantaran tetapi aplikasi baru termasuk yang akan diterokai pada masa hadapan sudah pasti akan menghabiskan jalurlebar yang sedia ada. Dengan itu, rangkaian ATM merupakan teknologi baru yang akan menjadi tulang belakang kepada perkhidmatan baru dan masa depan yang merangkumi data, visual dan suara. Walaupun rangkaian baru ini agak rumit kerana ia menggunakan pendekatan yang berlainan untuk perkhidmatan yang berbeza tetapi ia semakin mendapat

sambutan hangat, hasil dari keupayaannya yang fleksibel, efektif dan tinggi kadar pemulangannya.

Rangkaian yang laju seperti rangkaian ATM memerlukan pengendalian trafik yang berkesan untuk menambah penghantaran data dengan kos operasi yang terendah. Oleh yang demikian, simulasi dan pemodelan merupakan kaedah yang berkesan untuk mendapatkan keseimbangan terbaik antara parameter rangkaian untuk mendapatkan persembahan optimum. Kaedah peruntukan dinamik sumber rangkaian seperti jalur-lebar dan daftar akan dikaji agar memberi tahap perkongsian yang lebih efektif. Membandingkan rangkaian STM dengan rangkaian ATM, skema peruntukan statik dibukti tidak memberi tahap pemulangan penggunaan sumber rangkaian dengan berkesan. Disebabkan perkhidmatan yang berlainan tidak semuanya mempunyai kelakuan trafik yang sama, rangkaian ATM wajar menyediakan strategi peruntukan dinamik yang bersesuaian. Perbandingan tahap pencapaian dilaksanakan dengan mereka empat strategi peruntukan. Strategi-strategi tersebut dinamakan sebagai "Static Bandwidth Allocation", "Bandwidth Allocated Proportional to Expected Queue Length", "Bandwidth Allocated Proportional to Expected Queue Length with Threshold Value" dan "Bandwidth Allocated with Threshold Interrupt". Disebabkan strategi "Bandwidth Allocated with Threshold Interrupt" bertindak dengan segera ke atas kesesakan rangkaian, ia dibuktikan sebagai strategi yang paling efektif.

ACKNOWLEDGEMENTS

It is my great pleasure to express my special thanks for the guidance of supervisor Dr. Mohamed Othman, head of Communication Technology and Network Department, Faculty of Computer Science and Information Technology, University Putra Malaysia. I would like to thanks also to the committee members, Madam Lili Nurliyana, Miss Shyamala Doraisamy and Prof. Madya Dr. Ashwani Kumar Ramani. I benefited greatly from their experiences, expertise and knowledge. With their continuous advice, I built my confidence; reform my attitude and moving me towards the completeness of my work.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

AAL	ATM Adaptation Layer
ABR	Available Bit Rate
ATM	Asynchronous Transfer Mode
BER	Bit Error Ratio
B-ISDN	Broadband Integrated Services Digital Networks
CAC	Call Admission Control
CBR	Constant Bit Rates
CIR	Cell Insertion Ratio
CLP	Cell Loss Priority
CLR	Cell Loss Ratio
DTS	Dynamic Time Slice
FEC	Forward Error Correction
FIFO	First-In-First-Out
GFR	Guaranteed Frame Rate
GMDP	Generally Modulated Deterministic Process
IP	Internet Protocol
ITU-T	International Telecommunications Union-Telecommunications
MBS	Maximum Burst Size
Mbps	Mega bits per second
MCR	Minimum Cell Rate
MMPP	Markov Modulated Poisson Process
NNI	Network-Network Interface
nrt-VBR	non-real-time-VBR
PCR	Peak Cell Rate
QoS	Quality of Services
RT-HBS	Real-Time High Bandwidth Services
rt-VBR	real-time VBR
SAR	Segmentation and Reassembly
SCR	Sustainable Cell Rate
STM	Synchronous Transfer Mode
UBR	Unspecified Bit Rate
UNI	User Network Interface
VC	Virtual Channel

VP	Virtual Path
VCI	Virtual Channel Identifier
VPI	Virtual Path Identifier
VBR	Variable Bit Rate

CHAPTER I

INTRODUCTION

Asynchronous Transfer Mode (ATM) is emerging as the primary networking technology for the next-generation. Its spurred progress is the result of high transmission speed and bandwidth requirement in today's new applications such as multi-media communications, internets, desktop conferencing, video on demand and other networking applications. The ATM protocols are designed so that it can handle isochronous (time critical) data such as telephony (audio) and video, apart from more conventional inter-computer data communications, meaning today's and future applications will be integrated in one service, core transmission medium that is ATM.

ATM enables the rate of transmission of cells within the virtual channel to be variable, reflecting the source activity and resource availability (E. D. Sykas *et al.*, 1991). It adopts packet-oriented switching and multiplexing techniques, but differs from conventional packet switching. Since users first have to establish a virtual connection between the corresponding endpoints prior to the actual information transfer, and all packet (cells) of the same call follow the same route, therefore ATM is also defined as connection-oriented switching method. Through statistical multiplexing technique, ATM is expected to provide high bandwidth utilization and satisfactory quality of service for all connections.

The main argument of this work here is that several individual sources share a high speed transmission link, where its capacity is less than the sum of peak arrival rate of all the sources. Since the network resources are highly competent, they need to be allocated efficiently, such as maximizing bandwidth usage. The network performances would then be compared under several bandwidth allocation strategies. The following subsection will cover traffic management issues on ATM networks.

Traffic Management

Traffic management includes the actions of routing and resource allocation, necessary for setting up virtual connections, as well as protective measures required to maintain throughput in the event of overloading (J. W. Roberts, 1993). The ATM networks have to exercise flexible resource allocation and congestion control to utilize as much as possible the potential increase in network efficiency resulting from the use of statistical multiplexing (E. D. Sykas *et al.*, 1991). Resource allocation can be defined as the network resource reservation strategy used to achieve a balance between two conflicting performance criteria: network utilization and user satisfaction. The network operator seeks to utilize the available resources as much as possible whereas the users strive to obtain the best quality out of his usage of resources. Nevertheless the desired balance problem is difficult to achieve during periods of heavy load particularly if the traffic demands cannot be predicted in advance (R. Jain, 1996).

As known, many ATM sources will simultaneously trying to access the sharing resources such as buffer. At any one time, the network can only transmit one cell, and the rest of the contention cells are queued in the buffer. Buffer space is required to temporarily store these contention cells before transmitting them to other intermediate nodes. How much buffer space is needed to allocate to a queue is the question. Unlimited buffer space would cause cell delay performance to suffer, while inadequate buffer space for a queue would cause serious cell loss due to overflow of buffer. Thus, there is a need to find out the minimum buffer size in order to balance the network performances between cell delay and cell loss. Buffer is also used for traffic shaping where it will reshape the cell flow to conform to the traffic descriptor (N. Giroux and S. Ganti, 1998). Another benefit of buffering ATM cells is to achieve statistical multiplexing gains where bandwidth utilization can be increased especially for bursty traffic such as Variable Bit Rate (VBR) sources (E. D. Sykas *et al.*, 1991).

However, to ease the buffer management of the network, a switching node can be implemented with one or more queuing structures. Each queue may receive incoming cells from one type of service category only. The service categories consist of Constant Bit Rates (CBR), VBR, Available Bit Rate (ABR), Unspecified Bit Rate (UBR) and Guaranteed Frame Rate (GFR). These categories may also be divided to its sub-categories. A common serving strategy for queue management is First-In-First-Out (FIFO) basic. Since there are many queues in the switching node, to fairly allocate network resources to each queue is important in order to meet the Quality of Services (QoS) of each service category. The bandwidth needs to be allocated more flexibly and accurately. Therefore a scheduling mechanism is implemented at each queuing structure to appropriately select the order in which cells should be served (N. Giroux and S. Ganti, 1998). In other words, the queue is prioritized to gain network resources. Without priority handling of traffic within the network, the resources may become underutilized (J. R. Fernandez and M. W. Mutka, 1999).

On top of congestion control discussed above, the ATM networks have to find solutions for bandwidth allocation (J. L. Boudec, 1992). The function here is to fairly distribute bandwidth to all the existing connections and the ones to be set up. Unlike Synchronous Transfer Mode (STM), the ATM networks will allocate bandwidth to connections dynamically and any unused bandwidth will become available to other connections. Amount of bandwidth allocated to a connection is pre-determined before transmission begins. The re-allocation of bandwidth is done over certain periods of times. In order to increase network throughput, the bandwidth should be allocated effectively and wisely. The static bandwidth allocation and dynamic bandwidth allocation strategies are evaluated in (M. Othman *et al.*, 1999) and is concluded that the dynamic allocation strategy is more effective.

Improper traffic management will violate the fairness operation of the network in spite of constantly varying demand. This is true where ATM networks integrate different traffic classes with diversified QoS requirements. This is why resource allocation, although only a part of the traffic management issues, is essential and very important.

Source Model

Understanding the models of computer network's traffic will provide insight to design better protocols, better network topologies, better routing and switching hardware and lastly better services to users. Similarly, implementing or selecting the correct and accurate traffic model will closely contribute to the success design and efficient performance of the ATM networks to be built (G. D. Stamoulis *et al.*, 1994). Without the correct source model, any performance evaluation work is a waste of effort. Although how the data communication traffic behaves is still not fully understood and the traffic models used recently is not necessarily accurate, research will still need to be continued (G. Babic *et al.*, 1998).

Source characterization is necessary to precisely identify the behaviour of each particular source. It provides network management with the ability to manipulate flexibly the various services in terms of connection acceptance, negotiation of the QoS, congestion control, traffic enforcement and resource allocation. The term "Model" for a traffic source refers to an algorithm giving the generation time X_i of the ith cell, for i = 1, 2,…, the X_i's will be taken as random variables. In ATM networks, there is a general trend to visualize cells generation as a succession of active and silent (also called idle) periods (G. D. Stamoulis *et al.*, 1994). Cell generation occurs only during active periods and a group of successive cells that are not interrupted by an idle period is called a burst. The most prominent paradigm of source model exhibiting this behaviour is the ON-OFF model, which will be discussed in chapter two.