UNIVERSITI PUTRA MALAYSIA

DYNAMIC TRAFFIC SCHEDULING AND RESOURCE RESERVATION ALGORITHMS FOR OUTPUT-BUFFERED SWITCHES

SHAMALA SUBRAMANIAM

FSKTM 2002 6

DYNAMIC TRAFFIC SCHEDULING AND RESOURCE RESERVATION ALGORITHMS FOR OUTPUT-BUFFERED SWITCHES

SHAMALA SUBRAMANIAM

DOCTOR OF PHILOSOPHY
UNIVERSITI PUTRA MALAYSIA

2002

DYNAMIC TRAFFIC SCHEDULING AND RESOURCE RESERVATION ALGORITHMS FOR OUTPUT-BUFFERED SWITCHES

By

SHAMALA SUBRAMANIAM

Thesis Submitted to the School of Graduate Studies, Universiti Putra Malaysia, in Fulfilment of the Requirements for the Degree of Doctor of Philosophy

December 2002

Abstract of thesis presented to the Senate of Universiti Putra Malaysia in fulfilment of the requirement of the degree of Doctor of Philosophy

DYNAMIC TRAFFIC SCHEDULING AND RESOURCE RESERVATION ALGORITHMS FOR OUTPUT-BUFFERED SWITCHES

By

SHAMALA SUBRAMANIAM

December 2002

Chairman: Associate Professor Mohamed Othman, Ph.D.

Faculty: Computer Science and Information Technology

Scheduling algorithms implemented in Internet switches have been dominated by the best-effort and guaranteed service models. Each of these models encompasses the extreme ends of the correlation spectrum between service guarantees and resource utilisation. Recent advancements in adaptive applications have motivated active research in predictive service models and dynamic resource reservation algorithms. The OCcuPancy_Adjusting (OCP_A) is a scheduling algorithm focused on the design of the above-mentioned research areas. Previously, this algorithm has been analysed for a unified resource reservation and scheduling algorithm while implementing a tail discarding strategy. However, the differentiated services provided by the OCP_A algorithm can be further enhanced. In this dissertation, four new algorithms are proposed. Three are extensions of the OCP_A. The fourth algorithm is an enhanced version of the Virtual Clock (VC) algorithm, denoted as ACcelErated (ACE) scheduler. The first algorithm is a priority scheduling algorithm (i.e. known as the M-Tier algorithm) incorporated with a multi-tier dynamic resource reservation algorithm. Periodical resource reallocations are implemented. Thus, enabling each tier's resource utilisation to converge to its

desired Quality of Service (QoS) operating point. In addition, the algorithm integrates a cross-sharing concept of unused resources between the various hierarchical levels to exemplify the respective QoS sensitivity. In the second algorithm, a control parameter is integrated into the M-Tier algorithm to ensure reduction of delay segregation effects towards packet loss sensitive traffic. The third algorithm, introduces a delay approximation algorithm to justify packet admission. The fourth algorithm enhances the VC scheduling algorithm. This is performed via the incorporation of dynamic features in the computation of the VC scheduling tag. Subsequently, the delay bound limitation of the parameter is eliminated.

The proposed models are investigated for performance through analytical modelling and discrete-event simulation. The Semi-Markov Process (SMP) modelling approach is extended for the analyses of these cases. The analytical results are compared with the simulation results. The deviations of the results are within the acceptable limits, indicating the applicability of the model in the predictive service models and dynamic resource reservation algorithms. The results obtained have shown that the proposed four algorithms have significantly improved the four performance parameters analysed for delay sensitive traffic. However, the performances of the packet loss sensitive traffic are affected by the mechanisms of the M-Tier model. These consequences are eliminated with the integration of the threshold parameter in the M-Tier algorithm. The regulation of the threshold value has to correlate with the mechanism of the dynamic resource reservations with precision. The ACE scheduler significantly improved the performance of the VC algorithm. It is observed that the ACE scheduler outperforms the other algorithms.

Abtrak thesis yang dikemukakan kepada Senat Universiti Putra Malaysia
sebagai memenuhi keperluan untuk ijazah Doktor Falsafah

ALGORITMA TRANSMISI TRAFIK DINAMIK DAN PENEMPAHAN SUMBER UNTUK SUIS STORAN-LUARAN

Oleh

SHAMALA SUBRAMANIAM

Disember 2002

Pengerusi: Profesor Madya Mohamed Othman, Ph.D.

Fakulti: Sains Komputer dan Teknologi Maklumat

Algoritma transmisi yang dilaksanakan dalam suis Internet telah lama didominasikan oleh model servis 'cubaan-terbaik' dan 'guaranteed'. Setiap model servis berikut merangkumi pelusuk ektrim dalam spektrum korelasi di antara pengesyoran servis dan pengunaan sumber. Perkembangan terkini dalam rekabentuk aplikasi adaptif, telah memberi motivasi untuk penyelidikan dalam model servis penganggaran serta algoritma pengagihan sumber dinamik. OCcuPancy_Adjusting (OCP_A) adalah algoritma transmisi yang berorientasikan faktor-faktor rekabentuk yang tercatat di atas. Algoritma OCP_A telah dikaji untuk sistem pengagihan sumber yang mempunyai storan yang tidak membezakan kelas trafik serta strategi pengguguran paket berdasarkan kebarangkalian hujung. Perkhidmatan yang membezakan ciri-ciri khas setiap kelas trafik serta strategi pengguguran paket yang diimplementasikan oleh algoritma OCP_A boleh dimantapkan. Disertasi ini mencadangkan empat algoritma. Tiga algoritma merupakan lanjutan dari algoritma OCP_A. Algoritma keempat merupakan lanjutan dari algoritma transmisi Virtual Clock (VC) yang dinamakan algoritma ACelErated (ACE). Algoritma pertama adalah berdasarkan algoritma prioriti yang diintegrasikan dalam sistem pengagihan

sumber dinamik (yang dinamakan sebagai M-Tier). Pengagihan sumber berfasa dilaksanakan berdasarkan pemantauan status sistem. Justeru, membolehkan setiap kelas trafik untuk mencapai khidmat kualiti servis yang ideal. Algoritma tersebut turut menekankan kepentingan perkongsian-sumber yang tidak digunakan secara berhiraki. Algoritma kedua mengintegrasikan parameter kawalan dalam algoritma M-Tier. Ini untuk mengurangkan kesan kepentingan yang diberi terhadap trafik berorientasikan kelambatan. Algoritma ketiga memperkenalkan algoritma penganggaran kelambatan untuk memastikan kriteria kemasukan paket ke dalam sistem adalah sah. Algoritma ACE memperbaiki tahap pencapaian VC dengan mengintegrasikan ciri-ciri dinamik dalam pengiraan tag VC.

Prestasi pencapaian algoritma-algoritma telah dianalisa dengan menggunakan kaedah pemodelan analitik dan simulasi diskrit. Dalam perbandingan keputusan antara kaedah analitik dan simulasi, keputusan adalah dalam batasan had penerimaan. Justeru, membuktikan keberkesanaan penggunaan model-model yang dibina. Keputusan hasil kajian telah membuktikan rekabentuk empat algoritma baru telah memperbaiki prestasi empat parameter pencapaian untuk trafik yang sensitif terhadap kelambatan. Namun, pencapaian trafik yang sensitif terhadap pengguguran paket terjejas oleh model M-Tier. Kelemahan ini dapat diatasi oleh parameter kawalan yang disertakan dalam algorithma kedua. Algoritma ACE telah memperbaiki prestasi pencapaian algoritma transmisi VC. Hasil kajian telah membuktikan bahawa algoritma ACE telah mencapai prestasi terbaik. Kesimpulannya, penyelidikan ini dapat meningkatkan serta memberi pesepsi baru terhadap rekabentuk algoritma-algoritma transmisi yang diimplimentasikan dalam suis berhalaju tinggi.

ACKNOWLEDGEMENTS

Zig Zaglar has advocated that it seems universally true that people who have direction in their lives go farther and faster and get more done in all areas of their life. I am privileged to be bestowed with a guide, mentor, researcher and supervisor of pure dedication, Associate Professor Dr. Mohamed Othman. He has indeed devoted relentless effort in ensuring that the essence of my research is truly appreciated. My earnest gratitude to him for guiding the research with true wisdom and intellectual rigor.

I gratefully acknowledge Associate Professor Dr. Yazid Saman and Dr. Rozita Johari, my co-supervisors for their constant encouragements, guidance, advice and constructive discussions. My gratitude to Associate Professor Dr. Abd. Azim Abdul Ghani, Dean, Faculty of Computer Science and Information Technology for providing a stimulating research environment and guidance.

I am forever indebted to my mother, Krish C. Rayan for her endless love and utmost devotion to my education. She has indeed had infinite patience, implicit faith in my capabilities and I could not ask for more inspiration from her. My thanks to my sisters and brothers-in-law, Shanthi, Suganthi, Aru and Kannan for their unceasing cheer and steadfast optimism. Special mention goes to Raja Noora Ashikin who has constantly been a source of inspiration to me.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

ACE	ACcelErated
auxVC	Auxiliary Virtual Clock
CPU	Central Processing Unit
CSE	Current Service Environment
DDP	Delay Differentiation Parameter
Delay-EDD	Delay Earliest-Due-Date
DS	Differentiated Services
EDD	Early-Due-Date
EDF	Earliest Deadline First
FCFS	First Come First Serve
FFQ	Frame-based Fair Queuing
FIFO	First In First Out
FIFO+	First In First Out+
FQ	Fair Queuing
FTP	File Transfer Protocol
GPS	Generalized Processor Sharing
HoL	Head-of-Line
HoL-PJ	Head-of-Line with Priority Jumping
HRR	Hierarchical Round Robin
ID	Identification
IP	Internet Protocol
IQ	Input Queue
ISN	Integrated Services Networks
ISP	Internet Service Providers
LCFS	Last Come First Serve
LDF	Longest Delay First
LR	Latency Rate
LTO	Least Time to Overflow
LTO_LB	Least Time to Overflow with Leaky Bucket
MLT	Minimum Laxity Threshold
M-Tier	Multi-Tier
M-Tier(D)	Multi-Tier with Delay Approximation
M-Tier(T)	Multi Tier with Threshold Derivations
OCP_A	OCcuPancy_Adjusting
OQ	Output-Queued
PFQ	Packet Fair Queuing

PGPS	Packet-by-Packet Generalized Processor Sharing
QoS	Quality of Service
RCSP	Rate Controlled Static Priority
RPQ	Rotating Priority Queue
RPS	Rate Proportional Service
RTT	Round Trip Time
SCFQ	Self-Clocked Fair Queuing
SFI	Service Fairness Index
SMP	Semi-Markov Process
SPFQ	Starting Potential-Based Fair Queuing
STE	Shortest Time to Extinction
TCP	Transmission Control Protocol
TCP/IP	Transmission Control Protocol / Internet Protocol
TDM	Time Division Multiplexing
VC	Virtual Circuit
VLSI	Very Large System Integrated
VoIP	Voice-over-Internet Protocol
WAN	Wide Area Network
WDM	Wave Division Multiplexing
WF2Q	Worst-case Weighted Fair Queuing
WFI	Worst-Case Fairness Index
WFQ	Weighted Fair Queuing
WRR	Weighted Round Robin
WWW	World Wide Web

CHAPTER 1

INTRODUCTION

Communication technologies have transcended a wide spectrum of boundaries and challenges, to enable the globalisation of human interactions. Evolving over many decades, the insinuation of the various generations of computer networks has made a significant impact in our daily lives. Among the most notable technological success innovation is the Internet [25,44,110].

The Internet has grown both in terms of size as well as community penetration, with the advent of Voice-over-Internet-Protocol (VoIP), streaming video and the World Wide Web (WWW) [16,26,44,110]. The IP platform has often been viewed as an hourglass [25]. On the wide top are the applications; on the bottom are all the alternative physical transmission technologies. The narrow waist is the IP. That narrow waist isolates the myriad complexities of the underlying world from the equally daunting complexities of the upper applications. Thus, in an IP platform, the user is empowered to build applications on a minimally defined standard. The transformation of the Internet into an important and ubiquitous commercial infrastructure has not only created rapidly rising bandwidth demand but also significantly changed consumer expectation in terms of performance, security and services [17,72]. Even though the Internet is still extremely small compared to the telephone and cable television networks in terms of the number of users and the quantity of capital invested. it has clearly joined them as a significant aspect of our telecommunications infrastructure [110]. Among the major performance issues of the Internet is the nature of transmission service it supports, which is generally called Quality of Service (QoS) [6,7]. The majority of current Internet traffic is contained in Transmission Control Protocol (TCP) connections generated by applications requiring the transfer of some kind of digital document [73]. The TCP/IP protocol suite used in the Internet has been designed to provide QoS qualitatively on an aggregate network-wide basis [6,26]. This network technology is elastic, routinely losses information, experiences variable and unpredictable delay in data delivery, makes no distinctions between applications with different communication requirements and serves packets in output queues in a simple first-come-first serve (FCFS) order, with the packet at the front of the queue transmitted first [18,26,65]. The network QoS experienced by the TCP connections are mainly through the variable throughput achieved by the congestion control algorithms. An evident feature inherited by a network service model of this nature, is the occurrence of congestion [65]. The phenomenon of congestion has evidently been a critical problem in network management and design, attributed by the exponential increase in the network load. A QoS sensitive flow cannot readily tolerate the effect of packet loss, delay (and delay variation, or jitter), and fluctuations in network throughput and has the potential to adversely impact the performance of the network [44]. Attempting to integrate QoS sensitive flows into conventional IP infrastructures is a rather controversial endeavour [26] and raises several issues. Among which is the wide spectrum of correlation probabilities present as a trade-off between resource reservation and QoS guarantees. Numerous combinations exist as users wish to access a whole plethora of services via the Internet in an Integrated Services Network (ISN) [17]. A network derives an ideal correlation primarily by appropriately scheduling its resources [6,16,68,136]. This dissertation presents approaches for dynamic resource reservation strategies implemented in wide-area networks (WAN) via scheduling algorithms. The derivations of these algorithms are to ensure QoS guarantees whilst achieving high resource utilization.

This chapter is laid out as follows. Section 1.1 presents the environment of discourse. Section 1.2 and 1.3 discusses the switch evolution and its significance to congestion control algorithms. Section 1.4 discusses the issues pertaining to QoS guarantees and scheduling algorithms. Section 1.5 defines the fundamental requirements of scheduling algorithms. Section 1.6 presents the problem statement, section 1.7 discusses the research objectives, section 1.8 presents the research scope and the dissertation organisation is presented in section 1.9.

1.1 Environment of Discourse

The pre-dominant designs of computer networks were circuit-switched telephone networks. These networks carry traffic of a single type, and the traffic behaviour is well-known. Thus, congestion avoidance is simple by reserving enough resources at the start of each call. The trade-off in limiting the number of users is the guarantee of possessing enough resources for a call to achieve its performance target. However, resources can be severely under-utilised, since the resources are blocked by a call, even if idle, are not available to other calls [67].

Early research in computer data networking led to the development of reservationless store-and-forward data networks. These networks are prone to congestion since neither the number of users nor their workload are regulated. The trade-off from the flexibility gained by the statistical multiplexing of network resources is the possibility of congestion or marginal performance guarantees. This problem was recognized quite early and a number of congestion control schemes were proposed, references [26,136] provide a detailed review of these. The nature of rising communication systems, the constantly evolving and growing demands of the diligent 'on-line' population coupled with the stringent requirements of multimedia applications are causing a pressing need for existing algorithms, protocols and architectures to be revamped and redesigned [26,45]. Packet-switching networks have been long dominated by the features of data applications. These applications advocate robust, scalable and reliable end-to-end data transfer from the underlying Internet protocol architecture. Algorithms and protocols that were developed to realize these pre-requisites employed the best-effort service model (also known as the so-called *send and pray* model) [72]. This service model provided no service guarantees to the clients, allowed drastic service degradation when networks were overloaded, required no resource reservations and employed the FCFS algorithms. The requirements of low speed data applications such as telnet and file transfer protocol (FTP) were catered for efficiently via these best-effort services. However, the birth of multimedia technology created a new era of ISN and differentiated services (DS or diff-serv) architecture [98,128]. A packet in a diff-serv domain is classified into a class of service according to the stipulated contract profile and is treated differently by its class.

The incorporation of multimedia technology used in computing and communication system, offers a wide spectrum of opportunity whilst challenging. These challenges are attributed to the following key parameters depicting the characteristics and nature of multimedia traffic [7,26,68]: composed of a variety of traffic patterns. imposes significant requirements on network resources, posses unpredictable as well as highly variable bit rate requirements on multiple time-scales, requires differentiated communication modes and needs integrated services in a common network. Efforts to integrate multimedia applications into traditional data architectures proved to be unsuccessful [26]. This was partially caused by the fact that traditional architectures provided single level best-effort service. The best-effort delivery mechanism of standard computer networks does not lend itself to guaranteeing timely delivery and predictability for real-time data streams. Thus, causing the packets to experience variable delay and performance complications. To resolve this problem, QoS sensitive flows must be given priority handling in routers but still maintaining fairness, fast packet forwarding engines should be built and mechanisms for service differentiation should be defined [72]. This is the scheduling policy. Multimedia applications impose an obligatory fulfilment of resource reservation and traffic dependent services.

One of the core components of a QoS network is the packet scheduling algorithm which determines the transmission order of packets at the output buffers of switches [99]. In recent years, numerous QoS based packet schedulers were derived to enable conducive platforms for the deployments of networked multimedia system. In 1992, Clark et al. [26] described a method of evolving the original Internet architecture to an integrated services network that could support traditional applications as well as emerging real-time applications. Four architectural components were designed: a service level. a service interface, an admission control mechanism and scheduling mechanisms. The following is a simple description of the interactions between the components:

- A service level is defined. This includes all the service semantics: descriptions how packets should be treated within the network, how the application should inject traffic into the network, and how the service should be policed. Knowledge of the service semantics must be available within routers and applications.

- An application invokes a service using the service interface and a signalling protocol. The invocation includes specific information about the traffic characteristics required for the flow, such as the data rate. The network indicates if the service invocation was successful and might also inform the application of any service violation, either by the application's use of the service or from a network failure.

- Admission control uses the information in the service invocation, plus knowledge about other service requests it is currently supporting, to determine if it can accept the new request. Admission control, typically implemented in the routers, policies service use to ensure that applications do not use more resources than they have requested.

- Once a service invocation has been accepted, the network employs router mechanisms for scheduling and queue management to ensure that the packets within the flow receive the requested service.

Subsequently schemes designs revolved around the high weight-age placed on timing requirements of real-time applications [136]. These schemes were