QoS Analysis of Traffic Between an ISP and Future Home Area Network

Eugene Ng
August, 2006

Department of Electrical and Computer Engineering
University of Manitoba
Winnipeg, Manitoba

Sponsored by TRLab, Winnipeg

Abstract

Today's home network usually involves connecting multiple PCs and peripheral devices, such as printers and scanners, together in a network. This provides the benefit of allowing the PCs in the network to share Internet access and other resources. However, it is expected in the future, the home area network (HAN) will grow and extend to other home devices such as home entertainment systems (including digital TV, hi-fi stereo, etc.), appliances, webcam, security alarm system, etc. Connecting other home devices to a HAN provides users with many benefits not available in today's home networks. For example, home devices capable of connecting to the future HAN are able to share the content downloaded from broadband access anywhere in the home. Users can also have remote access and control of their home devices. To extend the home area network to all these different home devices, however, means that the traffic between the ISP and future HAN will be very different from the traffic generated by today's home network. In today's home network, which consists mainly of multiple PCs, a best-effort approach is able to satisfy the need, since most of the traffic generated by PCs is not real-time in nature. However, in future HANs, it is anticipated that traffic generated from home devices requiring real-time applications such as multimedia entertainment systems, teleconferencing, etc. will occupy a large proportion of the traffic between the ISP and future HANs. In addition, given the variety of home devices that could potentially be added to future HANs, the amount and variety of traffic between the ISP and a future HAN will certainly be very different from today's home network that is dominated by Internet/data traffic. To allow HAN users of these real-time applications and various types of home devices to continue enjoying seamless experiences in using their home devices without noticing significant delays or unnecessary interruptions, it is important for the ISP to be able to effectively manage the channel to the home so that it can provide sufficient bandwidth to ensure high QoS for home applications. The aim of this thesis is to understand the types of traffic that will be expected and to develop an analytical model that will represent the traffic behaviour between the ISP and future HANs to understand how to manage the channel to provide high QoS.

In this thesis, we use the continuous-time PH/M/n/m preemptive priority queue to model the traffic behaviour between the ISP and a future HAN. Three classes of traffic are defined in this model: real-time, interactive, and unclassified. Each of these three traffic classes receives a unique priority level. From the model one can approximate the amount of bandwidth required to be allocated for each traffic class for each household so that the total bandwidth required is minimized while the QoS requirements (delay and blocking probability) of the traffic generated by the home devices are met. Thus this model could potentially be used as a network planning tool for ISPs to estimate how much bandwidth they need to provide per household for homes that use home area network. Alternatively, it could also be used to estimate what quality of service (e.g. what is the mean delay and blocking probability expected) given a certain amount of bandwidth per household.

Table of Content

Figures and Tables

List of Figures

List of Tables

Acronyms

ADSL	asymmetrical digital subscriber line
CSMA/CA	carrier sense multiple access with collision avoidance
CTMC	continuous-time Markov chain
DQPSK	differential quadratic phase shift keying
DSSS	direct sequence spread spectrum
FCFS	first-come-first-serve
FIFO	first-in-first-out
HAN	home area network
IFFT	inverse fast Fourier transform
ISP	Internet service provider
LAN	local area network
OFDM	orthogonal frequency division multiplexing
PC	personal computer
POTS	plain old telephone system
RG	residential gateway
TDMA	time division multiple access
QoS	quality of service
UPnP	Universal Plug and Play
WAN	wide area network

Chapter 1

Introduction

1.1 Today's Home Network vs. Future Home Area Network

Today's home network usually involves connecting multiple personal computers and peripheral devices together in a network. This provides the benefit of allowing the PCs in the network to share Internet access and other resources (such as printers, scanners, etc.). However, there are two major drawbacks in today's home networking. First, it is not easy to configure; the user is required to have a certain level of networking knowledge in order to successfully configure a home network. Second, most of the home devices, such as appliances, security alarm system, and digital TV, are excluded from today's home network. As a result, sharing of the material downloaded from the Internet is restricted to the few PCs connected to the home network. One is not able to, for example, download an MP3 music file from the Internet and play it on the hi-fi stereo at home, or download a movie and play it on the TV.

It is envisioned that in the future, the home area network (HAN) will support "the connection of a number of devices and terminals in the home on to one or more networks which are themselves connected in such a way that digital information and content can be passed between devices and any access 'pipe' to the home" [28]. A HAN is a network that allows home users to have remote access and remote control of the home devices and their content or services. In its simplest form, it is the interconnection of multiple PCs as in today's home network, but the future HAN will also allow any home device capable of interfacing with the HAN to be interconnected together, and the home area network itself will connect to the outside world (i.e. the Internet) via the residential gateway (similar to how a gateway connects the LAN to the WAN). Figure 1.1 depicts how HAN might look in the future:

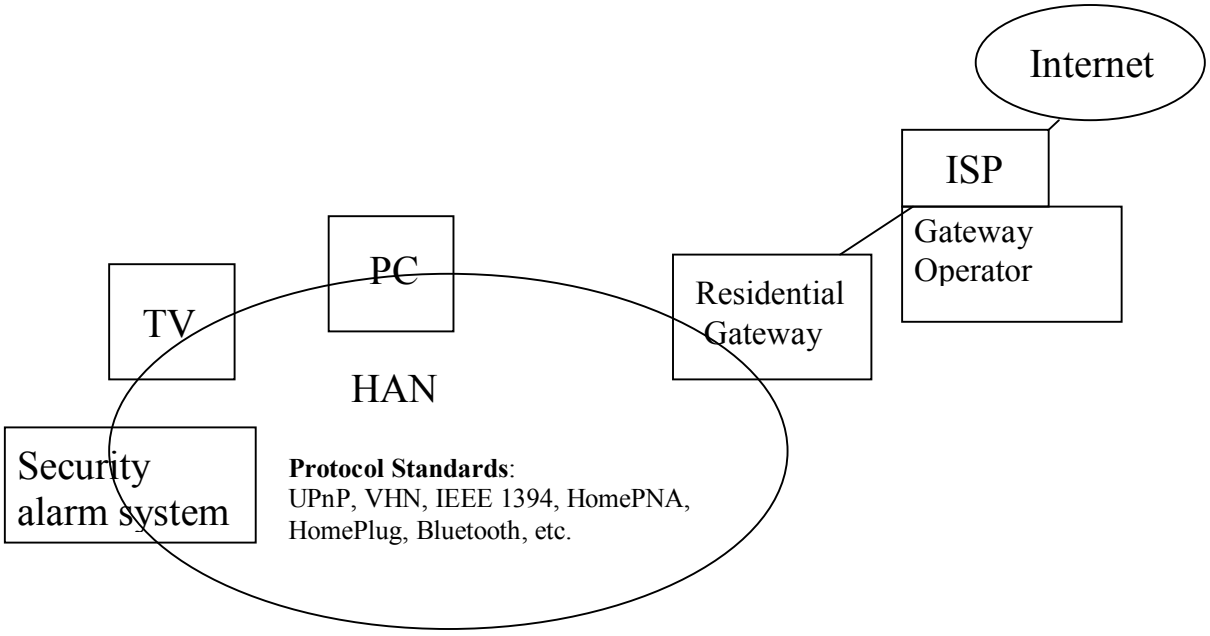


Figure 1.1: Connection between HAN, residential gateway, and the ISP.

As can be seen from Figure 1.1, besides connecting multiple PCs together, the HAN also connects the multimedia entertainment systems (such as digital TV, hi-fi stereo, CD/DVD/MP3 players), appliances, webcams, security alarm system, and almost any home device imaginable.

What are the benefits of interconnecting home devices together? One obvious benefit is that material downloaded using broadband access can be shared by any home device capable of using it, instead of being shared by only the few PCs in today's home network. Another benefit is that it allows the storing of content and making it accessible anywhere in the home. It is perhaps best to illustrate the benefits of HAN through some of the examples listed in [28].

The first example involves a digital TV and a web pad. In a digital TV, the main components involved in displaying information include a display, an MPEG decoder, and audio amplifier. In a web pad, the display is used to display visual information and the headphone is used for audio information. In a HAN, because home devices are connected to each other and thus can potentially interface with each other, the web pad application can use the MPEG decoder in the TV to deliver video to the web pad display and audio to the Web pad headphone [20].

Another example that illustrates the benefit of having a home area network can be in the area of telecare for seniors and those recently out of hospital. With a webcam connected to a home area network, a family member can monitor seniors at home (provided that the senior is willing to be monitored by his/her relative) to ensure their safety no matter where that family member is (for instance, to ensure that a senior at home did not fall down from stairs or become sick and not be able to get out of bed while the other members of the family are away on business or on vacation somewhere else in the world) [20].

1.2 Goal of this Thesis

From the discussion above, it is obvious that the capability of a HAN to remotely control home devices and remotely access their contents offers benefits to home users. However, due to the introduction of more home devices of varying types into the HAN, the traffic between the ISP and a future HAN is anticipated to be very different from the traffic between the ISP and current homes today. Up to now, traffic between an ISP and a home is mainly generated by PCs. This content mainly comes from web browsing, email, and some occasional video and audio streaming. Because most of this traffic is not real-time in nature, the best-effort approach to data transmission used in today's Internet is sufficient to meet the traffic demand. However, the anticipated traffic between an ISP and a future home area network will be drastically different. For instance, multimedia home devices like digital TV generate a lot of high bandwidth real-time traffic that needs to meet certain QoS requirements so that users will be satisfied with the service. Thus, to make the deployment of HANs successful, it is essential that we understand the behavior of the traffic that will be generated between an ISP and future home area networks and also how to guarantee that QoS requirements of the home devices will be met.

The goal of this thesis is to estimate the bandwidth required for each traffic class based on the mean delay and blocking probability requirements. More specifically, this thesis aims to develop an analytical model to represent the behavior of the different types of traffic between an ISP and a HAN; from the model, one can then approximate the bandwidth capacity required for each type of traffic given that certain QoS requirements (mean delay and blocking probability) of the traffic generated by the home devices are met.

This traffic model will be represented as a PH/M/n/m preemptive priority queue and will be analyzed using matrix-geometric method.

1.3 Thesis Organization

This thesis is organized into three parts. Part I gives a literature review of the current state of the art of HAN technology, as well as discussing the traffic types and QoS requirements that are expected in the future home area network. Part II presents the traffic model used to represent the behaviour of traffic between the ISP and future HAN. In part III, I present a performance analysis of the expected traffic between the ISP and future HAN using the model I develop in part II, and from there draw the conclusions and make suggestions for future study.

1.3.1 Synopsis of Part I: Literature Review

Part I is organized into 2 chapters. In Chapter 2, I present a literature review of the current state of the art of HAN technology. In Chapter 3, I review the expected traffic types in future HANs suggested by previous researchers and how priority levels can be assigned to each type of traffic. I will also discuss the QoS constraints that are important for our traffic model and list the QoS parameters of some key home applications.

Current state of the art of HAN technology (Chapter 2)

Chapter 2 provides a brief summary of the current status of home area network technology. The survey will be discussed in two areas: HAN interconnection architectures and HAN technology at the physical and link layers. In terms of HAN interconnection architectures, the two architectures that receive the most attention today, UPnP (Universal Plug and Play) and CableHome, will be discussed. At the physical and link layer, we will give a summary of some of the more promising current technologies, including HomePNA, HomePlug, HomeRF, and IEEE 802.11.

Traffic between ISP and future HAN & QoS Requirements (Chapter 3)

In this chapter, I review previous work in the classification of the expected traffic types in future HANs and the way priority levels were assigned to each type of traffic. I also describe my own strategy for how traffic between an ISP and future HANs will be classified in our traffic model. The logic behind the assignment of priority levels to the traffic classes using a preemptive priority scheme is discussed as well. The QoS requirements (bandwidth, delay, and blocking probability requirements) of some key home applications are also be presented in this chapter.

1.3.2 Synopsis of Part II: Description of Traffic Model

PH/M/n/m Preemptive Priority Traffic Model (Chapter 4)

Part II is organized as a single one chapter, Chapter 4, in which I discuss the logic behind choosing the PH/M/n/m preemptive priority queue as my traffic model to represent the traffic between an ISP and future HANs. I also describe the model in a high-level conceptual overview before presenting the mathematical

detail later in the chapter. Then I describe in detail the arrival and service processes, the behavior of the service discipline, preemptive priority FIFO (first-in-first-out) with delay loss, that is used in our traffic model, as well as the algorithm for constructing the generator matrix Q for our traffic model.

1.3.3 Synopsis of Part III: Performance Analysis

Part III is organized into two chapters, Chapter 5 and Chapter 6.

Performance Analysis (Chapter 5)

In Chapter 5, I first discuss how to obtain the stationary distribution of our preemptive priority queue. Once the stationary distribution is obtained, I discuss the equations for obtaining the performance metrics of interest, which include average queue length, blocking probability, and mean waiting time.

Results and Discussions (Chapter 6)

In Chapter 6, I illustrate the use of our traffic model by implementing an example system with 2-phase Coxian arrival process, 7 servers, and 27 buffer slots, i.e. a PH/M/7/(9+9+9) preemptive priority queue. Afterward, the results of the performance analysis on this example system are discussed.

Part I: Literature Review

Chapter 2

Current State of the Art of Home Area Network Technology

2.1 Standards in HAN Technology

To discuss the emerging standards in HAN technology, it is convenient to divide the discussion into two parts: the HAN interconnection architectures and interconnection technology at the physical and link layer. The following section discusses two emerging HAN interconnection architectures, and in Section 2.3 I discuss the interconnection technologies at the physical and link layer. Since this thesis deals with the QoS analysis of traffic, more emphasis will be placed on describing the QoS requirements specified by these HAN technologies.

2.2 HAN Interconnection Architectures

Two major emerging interconnection architectures contending for the future HAN are UPnP (Universal Plug and Play) and CableHome. UPnP [27, 29] is an interconnection architecture that aims at connecting together all types of personal computers, intelligent appliances, and wireless devices. It defines the Device Control Protocols (DCPs), which allow interactions between home devices. DCPs utilize standard device descriptions written in XML that describe the standard methods for device interaction.

The main features of UPnP are plug and play, automatic device and service discovery**.** Plug and play in UPnP is similar to the concept of plug and play in computers and allows UPnP devices to be "plugged" in to become part of the home network. In automatic discovery, a home device advertises its services (i.e. the actions that it is able to perform) when requested, and in service discovery, devices are able to discover the presence and services of other devices within the same home network. Using UPnP, any home device that is compatible with UPnP can simply plug into the home area network. The UPnP architecture allows these devices to advertise their presence and services to the home area network so that other home devices already in the HAN can automatically discover the existence and services offered by that newly added device. Once the automatic discovery is done, the other home devices are then able to make use of the device's advertised services.

In terms of QoS, UPnP supports both prioritized and parameterized QoS [29]. In prioritized QoS, each packet is assigned a priority number called a *TrafficImportanceNumber*, which is assigned by the application that originally generated the packet. Packets access the shared media by order of their priority. In parameterized QoS, parameters can be used to define the requirements of a traffic stream.

CableHome [5] is a set of specifications specifying the requirements for a residential gateway and to standardize Quality of Service (QoS) and LAN messaging within future home area networks. To guarantee Quality of Service, CableHome uses a type of priority queuing that is referred to in their specification to as "First in, First Out with Priorities, and Highest Priority Queue First" [5]. According to the specification, packets in each outgoing interface in home devices are polled according to their priorities. The polling begins with the highest priority packet that arrives at the queue first, and that packet is extracted out of the

queue and transmitted onto the shared media. If no highest priority packet is found, then the next highest priority packet is polled, and so on. This process is essentially the same behavior as a first-come-first-serve (FCFS) preemptive priority queue, where a packet with the highest priority is allowed to receive service immediately even if another packet with lower priority is already in service when the higher priority packet arrives.

Note that currently, CableHome supports priority-based QoS only. This is different from UPnP, which supports both prioritized and parameterized QoS. In this thesis, we will assume prioritized QoS, since it is supported in both CableHome and UPnP.

2.3 HAN Technologies at the Physical and Link Layer

Emerging HAN technologies at the physical and link layer exist in both wired and wireless networks. Section 2.3.1 discusses the wired HAN technologies that are currently under development. Section 2.3.2 describes the developing wireless HAN technologies.

2.3.1 Wired Connection

HomePNA

The major emerging HAN technologies that use wired interconnection include HomePNA and HomePlug. HomePNA [2, 7, 15] is a physical and link layer HAN technology initiated by the Home Phone Line Networking Alliance. It proposes to use phone lines already available in homes to carry data for the future home area network. First, because phone lines are readily available in today's homes, this technology has the advantage that "no new wire" is necessary to implement the future HAN. Furthermore, using phone lines as communication medium for the future home area network means it is easy for the future HAN to co-exist with existing POTS/ADSL technologies. In addition, phone line is an inherently more secure environment when compared to other medium like wireless and power lines.

HomePNA is designed to co-exist with other phone line services like POTS and xDSL. It uses frequency division multiplexing (FDM) to split the bandwidth of a phone line into three separate frequency spectrums, with each spectrum allocated a different function. The 0 to 4 kHz spectrum is allocated POTS service, the 26 kHz to 1.1 MHz spectrum is reserved for xDSL services, and the 4 to 10 MHz spectrum is used for HomePNA to transmit packets for a home area network.

Currently in its third version (HPNA 3.0), HomePNA can theoretically transmit within the homes at 240 Mbps. Further enhancement of data rates of up to 320 Mbps will be expected in HPNA 3.1, which the specification is expected to be completed in the summer of 2006. HomePNA allows up to 50 devices to be connected the home area network simultaneously.

At the media access control layer, HomePNA uses a modified version of IEEE 802.3 framing and Ethernet CSMA/CD MAC behaviour. To provide QoS, it proposes to use 8 priority levels for packet transmission. As explained in [15], the QoS mechanism in the HomePNA MAC protocol can be viewed as being divided into two parts. In the first part, packets with lower priority levels must wait for all higher priority level packets to finish transmission over the shared media before the lower priority packets can begin

transmission. The second part deals with the mechanism of collision detection and retransmission, with collision possible only for packets with the same priority level.

The "no-new-wire" advantage of HomePNA means that setting up a home area network using HomePNA incur no additional cost in terms of wiring. Moreover, the cost of initial installation is estimated to be less than $100US [7], which is relatively inexpensive. However, in countries where houses do not have many telephone sockets, such as the UK, the cost of using HomePNA as a viable HAN technology might be relatively expensive compared to that of HomePlug and the wireless solutions. Moreover, the use of phone lines as communication channel introduces some problems that need to be overcome. For example, as pointed out in [9], POTS (plain old telephone system) signaling and ringing can produce significant transients. Also, the coupling of AC line-noise from the power lines to the phone lines introduces impulse noise on many phone lines. Another problem with phone line is that telephone instruments on the same wiring present a wide range of frequency-dependent impedances. Fortunately, most of these noise-related problems can be solved by placing low-pass filter at appropriate points in the home area network to reduce the undesirable effects of the noise and low impedances coupled to the phone lines, as discussed in the report of the field tests for HPNA 1.0 in [9].

As mentioned before, HomePNA's uses a modified version of IEEE 802.3 framing and Ethernet CSMA/CA MAC behaviour. IEEE 802.3 is a Layer 2 protocol that has been extensively tested for many years in real use already, so HomePNA, being 802.3 compatible, gives a fairly good confidence that it is a protocol that can support other networking standards and in particular the IP suite. Also, HomePNA is supported by many vendors and has been standardized by ITU-T as Recommendation G.989.1, so it is an accepted standard that can guarantee interoperability between equipments from different manufacturers.

HomePlug

HomePlug [2, 8] is a HAN technology at the physical and MAC layers that utilizes power lines already installed in homes as communication channel for the future HAN. As in the case of HomePNA, HomePlug also has a "no-new-wire" advantage of using existing infrastructure. HomePlug enables any home device with a power plug to connect to the home area network. Since it is common to have many power outlets within a house, with at least one power outlet in each room usually, HomePlug has the advantage that the entire house can be serviced without the need of additional wiring. This is a distinctive advantage of HomePlug over other HAN technologies. For instance, a HAN serviced by HomePNA might encounter the problem of having not enough phone sockets to service all the home devices in a home. Also, full service coverage of a home might sometimes not be possible using HomePlug alone because some areas of a house might not have a phone socket (e.g. not all houses have phone sockets in every room). As for wireless HAN technologies, certain areas of a home might be out of serviceable range due to obstacles that limit wireless transmission within a home, such as a thick wall. Also, wireless transmission over the unlicensed 2.4GHz spectrum often encounters interference due to other devices that share the same frequency spectrum, such as microwave and amateur radio frequencies. Hence, PowerPlug has the advantage over HomePNA and wireless HAN technologies in providing a full service coverage of a home.

Despite the advantages, using power lines as communication medium have some major disadvantages. One major problem is that power lines have an inherently noisy environment for the purpose of data transmission. The attenuation due to AC cycles produce interference on the packets transmitted on a power line. To solve this problem, HomePlug uses orthogonal frequency division multiplexing (ODFM) to split a signal up over 84 available narrowband sub-carriers. The data bit-streams are then modulated onto the sub-carriers using

differential quadratic phase shift keying (DQPSK) and inverse fast Fourier transform (IFFT). This solution helps a wire line channel to deliver a large quantity of data over a relatively short period of time, thus minimizing the effect of the noisy environment of the power lines on data transmission.

Another problem with power lines is privacy issue. Unlike phone lines, which are relatively secure medium, it is possible to have leakage of data to networks within other buildings using power lines. Therefore, it is important to have a security feature in HomePlug to render the leaked data to neighboring buildings as unintelligible. To address this issue, HomePlug uses a 128-bit encryption to encrypt the data transmitted over power lines.

Regulatory issue is yet another obstacle with using power lines. In Europe, the regulations on electromagnetic (EMC) emissions is stricter than in North America, so it is uncertain whether HomePlug is able to deliver the promised data rate under the more stringent European EMC emissions regulation. If the data rate needs to be lowered to reduce emissions, this will lead to lower performance compared to other HAN technologies and an increased risk of the HAN services being unavailable at some power outlets [2].

HomePlug operates in the 4.5 to 21 MHz spectrum and transmits at theoretical data rates of up to 200 Mbps. To ensure QoS, it provides both time division multiple access (TDMA) and CDMA/CA (Collision Sense Multiple Access/Collision Avoidance) access at the MAC layer. TDMA provides contention-free access to the shared media, thus supporting home applications that use parameterized QoS. CDMA, on the other hand, is a contention-based access mechanism. In the HomePlug version of CDMA access, packets are classified into 4 priority classes. At the beginning of a contention window, there is a brief Priority Resolution phase during which pending traffic with lower priority are eliminated from the contention. After the brief Priority Resolution phase, CSMA/CA is then applied to the pending traffic with the highest priority level. In this way, CDMA/CA is used in HomePlug for supporting home applications using prioritized QoS.

In terms of interoperability between equipments from different manufacturers, many vendors are already supporting HomePlug.

2.3.2 Wireless Connection

HomeRF

In the area of wireless HAN technology, HomeRF and the IEEE 802.11 standards appear to be the most promising ones at present. HomeRF [2, 21] is a HAN technology at the physical and MAC layer that addresses issues related to wireless voice and data transmission within home area networks. With 8 simultaneous voice lines specified in version 2 of the standard, HomeRF provides voice transmission at near-wire-line quality. HomeRF operates at the 2.4 GHz ISM band, and home devices can be connected either ad-hoc or as a managed network. It currently supports data rates of up to 10 Mbps, but it is expected to support up to 25 Mbps in the future. HomeRF is capable of supporting up to 127 nodes.

Since HomeRF is concerned with wireless transmission, it addresses the issue of interference by using FHSS with frequency hopping of up to 50 hops/sec with 22 different hop patterns [2]. To address the security issue that is inherent in wireless transmission, HomeRF uses a 128-bit encryption to encrypt the data transmitted. Furthermore, it requires a 24-bit network ID from home devices before a device can be connected to a home area network.

To ensure that the QoS for wireless voice transmission is met, HomeRF provides different QoS mechanism for voice and data. In HomeRF, data traffic is based on a 20ms frame structure (with one hop per frame). For voice traffic, however, however, HomeRF moves to a 10ms subframe substructure (with one hope per subframe). The shorter subframe substructure provides decreased latency and decreased interference. Furthermore, HomeRF uses TDMA access for voice on the MAC layer. The shorter subframe structure and the TDMA access help ensure that the QoS requirements of voice traffic are met.

For data transmission, HomeRF uses a 20 ms frame structure. Data traffic access the shared media using a contention-based access mechanism (CDMA). Furthermore, in HomeRF, streaming multimedia traffic has priority over other data traffic.

IEEE 802.11

IEEE 802.11 was originally a standard for wireless LANs based on the well-established 802.3 standard. There are several variants of this standard. 802.11a and 802.11b are standards drafted to improve upon the original 802.11. However, because these two variants are incompatible with each other, vendors and customers alike were confused at which standard they should use in their wireless products. 802.11g was drafted later in an effort to combine the advantages of the two standards while maintaining backward compatibility with both standards. 802.11e is an extension to provide QoS, which is not addressed in any of the aforementioned variants.

IEEE 802.11a and 802.11b

IEEE 802.11b [11] supports data rates of 5.5 Mbps and 11 Mbps. It operates at the 2.4GHz ISM band and uses a spread spectrum modulation technique called direct sequence spread spectrum (DSSS) at the physical layer. At the MAC layer, 802.11b uses CSMA/CA to detect if any of the RF channels are usable. If a channel is busy, a nodes wanting to send a packet will back off for a random time period.

802.11a [10], on the other hand, supports a much higher data rate of 54 Mbps. This is due to the use of a multi-carrier modulation technique called OFDM (orthogonal frequency division multiplexing), whereas the DSSS used by 802.11b is a single-carrier system.

The disadvantage of 802.11a is that it operates at 5.2 GHz, a frequency often used in military applications. As such, this frequency spectrum is regulated against commercial use in many countries. Both 802.11a and b standards use CDMA/CA at the MAC layer, and they support both ad-hoc and managed network infrastructure.

IEEE 802.11g

802.11g [2, 6, 12] was developed to incorporate the advantages of both the 802.11a and b variants. It supports the same data rate as 802.11a (54 Mbps) while operating at the frequency of 802.11b (2.4 GHz). A data rate of 54 Mbps is made possible in 802.11g by using the same OFDM modulation technique as in 802.11a. At the same time, the 802.11g standard requires mandatory implementation of 802.11b modes. Basically, 802.11g uses DSSS as in 802.11b to achieve data rates of up to 20 Mbps and then further increase

the data rates of up to 54 Mbps using ODFM. In this way, IEEE 802.g is able to support the same data rate as 802.11a while operating at 2.4 GHz, overcoming the frequency spectrum issue of 802.11a.

IEEE 802.11e

802.11e [2] addresses QoS issues, which are not addressed by the 802.11a, b, and g variants. In this extension, traffic is divided into 8 priority levels. In the CSMA/CA protocol that is used by any variant of 802.11, after detecting that the channel is idle, nodes need to wait for a period of time before they are allowed to transmit a packet. In 802.11e, the length of this waiting period depends on the priority level. Traffic with higher priority levels has a shorter waiting time than that of lower priority traffic. This way, 802.11e provides a certain level of QoS, since the shorter waiting period of higher priority traffic results in a higher probability that more higher priority packets are transmitted than lower priority ones.

One advantage of using any of the variants of 802.11 as a HAN technology is that it has been widely adopted by many manufacturers already, thus interoperability of equipments among different vendors should not be an issue. Also, it covers a fairly long range of 25 to 500 meters, whereas HomeRF covers only up to about 50m. However, compared to HomePNA, HomePlug, and HomeRF, 802.11 variants are not easy to configure by general home users. Whereas HomePNA, HomePlug, and HomeRF supports UPnP, 802.11 requires a certain level networking knowledge for a home user to be able to configure it properly. Also, 802.11's operation in the 2.4GHz ISM band means that it shares the same interference problem as HomeRF from many other home devices that also use this frequency spectrum.

2.4 Which HAN Technology Will Be "The Standard"?

As one can see from the above discussion, there are many standards being developed for home networking. As of today, there is not a single "standard" that has been adopted by the home networking industry yet. Since each HAN technologies has its own advantages and disadvantages, it appears that various home networking technologies will coexist in the future, each covering a certain aspect of the requirements of a future home area network.

Chapter 3

Traffic Between ISP and Future HAN & QoS Requirements

3.1 Expected Types of Traffic in Home Area Network

Lei et al [18] define seven classes of traffic in their QoS framework for a residential gateway. These seven classes are security, multimedia, device control, FTP, web surfing, interactive, and unclassified. Each of these seven classes is assigned a unique priority level from 1 to 7, with priority 1 being the lowest priority and priority 7 being the traffic class with the highest priority. Details of how the traffic classes are assigned their priority in Lei et al's QoS framework can be found in [18]. Other standards and research papers have categorized traffic in HAN according to priority classes as well. For example, in the IEEE 802.1D standard, a layer 2 protocol for local and metropolitan area networks that includes a prioritization mechanism, 8 priority levels are defined [26]. From highest to lowest priority, they are network control traffic, voice traffic, video traffic, controlled load traffic, excellent effort traffic, reserved traffic, background traffic, and best effort traffic. A description of these 802.1D priorities can be found in [1].

Although different research papers and standards assign different numbers of priority classes, these classes can be aggregated into three broad classes: real-time, interactive, and unclassified. In this thesis, the real-time, interactive, and unclassified classes are priority level 1, 2, and 3 respectively. Priority 1 (real-time) class has the highest priority, and priority 3 (unclassified) class has the lowest priority. The real-time traffic class is composed of real-time video, real-time audio, telephony, and control signal traffic. Real-time video traffic is traffic generated by real-time video applications such as digital TV, teleconferencing, distance-learning applications (e.g. remote classroom) and streaming video applications. Real-time audio traffic includes traffic generated by real-time audio applications like hi-fi stereo system and other audio streaming applications. Telephony traffic refers to traffic related to the use of telephones. Control signal traffic is composed of both time-critical and safety-critical control signals needed to control the network. The interactive traffic class includes applications that do not require real-time traffic but interact with users, such as web browsing. Finally, the unclassified traffic class includes traffic that is neither real-time nor interactive in nature, such as FTP. These three traffic classes will be used in the traffic model in this thesis, each assigned a unique priority level. Furthermore, preemptive priority is used in our traffic model. A more formal definition of these three priority classes is given in Section 4.3. The priority assignment and preemptive priority will be further discussed in Chapter 4 as well.

As discussed in Chapter 2, a lot of the HAN technologies at the physical and link layer provide their own QoS mechanisms, and each of these technologies divides traffic into different number of priority classes. For example, HomeRF divides traffic into isochronous and asynchronous data, HomePlug differentiates traffic into 4 priority classes, and HomePNA and IEEE 802.11e support 8 priority classes. Although all the HAN technologies discussed in Chapter 2 support priority-based QoS, some of them are not strictly preemptive. However, these different QoS mechanisms within home area networks are not the main concerns of this thesis. This thesis is mainly concerned with providing QoS for the traffic between an Internet service provider and the residential gateway (RG). Once traffic arrives at the residential gateway, which is located within each home, it is the responsibility of the residential gateway to further refine the QoS necessary for

the home devices within the HAN itself. Thus, the traffic model in this thesis tries to handle the QoS issue for the traffic *between* the ISP and RG by differentiating traffic into three priority classes. Once traffic arrives at the RG, the RG will then handle the QoS mechanism *within* the home area network.

3.2 Bandwidth Requirements of Key Applications and Their QoS Parameters

To measure the performance of a network, it is common to use metrics such as throughput, delay, and reliability. Thus, to develop a traffic model to ensure that home devices in HAN receive reasonable quality of service, the model must be capable of providing the throughput, delay, and reliability requirements for the applications in each traffic class. In [13], the bandwidth requirements for key applications corresponding to five different types of traffic (telephony, audio, video, internet/data, and control) between ISP and home area network are identified and summarized in table form. Table 3.1 is mainly taken from Table 3 (page 14) in [13] and from [26].

Table 3.1: Bandwidth Requirements of Key Home Applications (from [13] and [26])

Table 3.2 lists the delay and reliability requirements for different traffic classes. The values in Table 3.2 are from the research done in [25]. See pages 18-31 in [25] for a detailed discussion of their research on the perceived QoS for WWW and streaming video.

Service	Mean Delay (ms)	Jitter (ms)	PER (Packet Error Rate)
High Quality Voice	10	±5	10^-3
Streaming Video	125-150		0.05
HDTV	90	±10	10^-5
Video Conference	10	±5	10^-5
CD Quality Audio	100	±10	10^-5
WWW	1.0s	-	-

Table 3.2: Delay and Reliability Requirements

Part III: Description of Traffic Model

Chapter 4

Traffic Model

4.1 How the Channel Between the ISP and Future HAN is Modeled

The channel between the ISP and the HAN is modeled as a CTMC (continuous-time Markov chain) multi-server preemptive priority queue of type PH/M/n/m (where PH stands for phase-type arrival process, M stands for exponential service process, n specifies the number of servers, and m specifies the number of buffer slots). The channel is modeled as depicted in Figure 4.1.

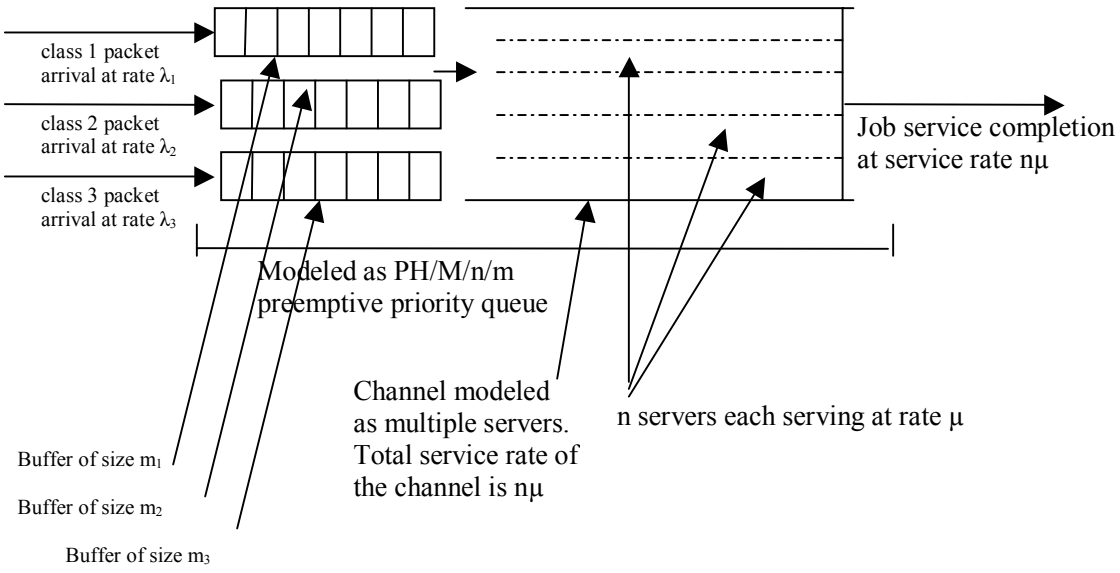


Figure 4.1: Diagram of how the channel between an ISP and a HAN is modeled

As depicted in Figure 4.1, our traffic model differentiates the traffic between an ISP and a HAN into three priority classes. We reserve the highest priority class, class 1, for real-time traffic. The second priority class, class 2, is used for interactive traffic. Traffic that is not real-time or interactive belongs to the lowest priority class, class 3. We divide traffic into priority classes because real-time applications must meet certain QoS requirements to deliver their services to users in a satisfactory manner. Hence we need to differentiate traffic into real-time and non-realtime traffic and assign a higher priority to real-time traffic.

Differentiating traffic into only two priority traffic classes is not sufficient, because a substantial amount of traffic generated by home users is interactive traffic. Interactive traffic has less stringent QoS requirements than real-time traffic, as shown in Table 3.1 and Table 3.2, so grouping interactive traffic with real-time traffic will result in a waste of bandwidth. However, we should not group interactive traffic with the lower priority class either, because interactive traffic has higher QoS requirements than non-realtime traffic that are

not interactive, as will be shown in Table 4.2. Therefore, in this thesis, we differentiate traffic into three priority classes, with real-time traffic occupying the highest priority, interactive traffic having the second highest priority, and the other unclassified traffic having lowest priority.

Our traffic model, depicted in Figure 4.1, has the following properties:

- There are three classes of packets. Each packet is assigned a priority level i = {1, 2, 3}. A packet with priority 1 has the highest priority (i.e. has preemptive priority over priority 2 and 3 packets). Priority 3 is of the lowest priority.
- Packets arrive according to phase-type distribution represented by (α, T_i) of dimension d_i with mean rate λ_i. The arrival process will be discussed in more detail in Section 4.4.
- The server process is represented by n identical servers, each serving at rate μ.
- The service times of packets are iid exponentially distributed random variables, and they are assumed to be independent of the arrival process.
- There are 3 buffers m_i for i = {1, 2, 3}. Each buffer m_i is exclusively for class i packets. For example, m_1 represents the number of waiting slots for class 1 packets, m_2 represents the number of waiting slots for class 2 packets, and so on.
- The system capacity for class i traffic consists of the number of servers plus the buffer space for that class i traffic. In other words, system capacity for class i traffic = $n + m_i$.
- The service discipline is FIFO preemptive priority with delay loss. Upon the arrival of a packet, there are four possible scenarios regarding whether the newly arriving packet will receive service or not:
    1. (number of packets in system < n) If there are less than n packets in the system, the arriving packet randomly chooses a server and enters into service immediately.
    2. (number of packets in system >= n) If there are n or more than n packets in the system already, and if one of the n packets already receiving service is of a lower priority than the new packet, then the new packet of higher priority randomly chooses a server among the pool of servers that are serving a lower priority packet and receives service immediately. The lower priority packet that has been receiving service from the chosen server originally is being preempted and placed back into the appropriate buffer.
    3. (number of packets in system >= n and < $n+m_i$) If there are n or more than n packets in the system already and if all of the n packets already receiving service are of the same or higher priority than the new packet, then the new packet joins the buffer if the buffer for that traffic class is not full yet.
    4. (number of packets in system >= n and > $n+m_i$) If there are n or more than n packets in the system already, and if all of the n packets already receiving service are of the same or higher priority than the new packet, and if the buffer for that traffic class is already full, then the new packet is lost.

4.2 Why Use Preemptive Priority Queue?

To understand our logic behind choosing a preemptive priority access scheme for our traffic model, we should first understand the two approaches generally used to ensure QoS. One approach uses *prioritized (differentiated) QoS* and the other uses *parameterized (scheduled) QoS*. In parameterized QoS, it is the application that triggers the QoS mechanism. When an application needs to transmit data over a network, it first sends a control signal to request a certain amount of resources (e.g. bandwidth, amount of time, etc.) from the network. This is called resource reservation in parameterized QoS. If the network has enough

capacity left to satisfy the requested resources, the reservation will then be granted and the application can begin transmission.

In prioritized QoS, instead of each application triggering its own QoS reservation, applications are aggregated into different service classes. Applications within the same service class have the same priority in accessing the network, and each service class has its priority level assigned to it. In the prioritized approach, jobs that belong to a higher priority service class are selected for service ahead of those with lower priorities, independent of their time of arrival into the system.

As mentioned in Chapter 2, 802.11b/g/a, HomePNA, HomePlug, CableHome, and UPnP are the HAN technologies that are anticipated to be the most widely used in future HANs. These technologies have, or will shortly have, support for priority-based QoS. In these technologies, the way these media-access technologies handle their priority-based QoS scheme is to divide home devices into different service classes with each service class having its own priority level. Devices belonging to the highest priority class are allowed to transmit packets over the network first, and the leftover network capacity is then used for the home devices with lower priority. This behaviour is very similar to the behaviour of a preemptive priority queue, where jobs with highest priority are selected for service ahead of those with lower priorities, independent of their time of arrival into the system. Hence our decision to use a priority queue as our traffic model.

Preemptive priority means that a job with higher priority is allowed to enter service immediately even if another with lower priority is already in service when the higher priority job arrives. We choose to use preemptive priority because in future HANs, the applications requiring real-time transmission are usually of higher priority, thus preemptive priority access can guarantee that the QoS requirements of such real-time applications will be satisfied, even if there are many frequently arriving non real-time requests.

4.3 Priority Level Assignment and Packet Size

The three traffic classes used in our model are real-time traffic, interactive traffic, and unclassified traffic. Real-time traffic includes real-time video and audio traffic, as well as telephony traffic and control signals. Interactive traffic class includes traffic that are interactive in nature, such as web browsing. Unclassified traffic is traffic that is neither real-time nor interactive in nature. Real-time traffic class is assigned the highest priority level (priority 1), because these traffic must be ensured certain QoS requirements to satisfy the real-time traffic demand. Interactive class is assigned priority level 2, because although these applications are not real-time in nature, they nevertheless should have a response time below a certain level to satisfy user expectations. Unclassified class is assigned priority 3, the lowest priority. Table 4.1 lists the priority levels of the three traffic classes.

Priority	Traffic Type
1	Real-time (real-time video, real-time audio, telephony, and control signal)
2	Interactive
3	Unclassified

Table 4.1: Priority Level Assignment

Each of the three priority classes has its own QoS requirements that need to be satisfied. Recall that in Section 3.2.2, we discussed about the QoS requirements of some traffic types, including the delay and reliability requirements of high quality voice, streaming video, HDTV, world wide web, etc. These requirements are listed in Table 3.2. For the purpose of our 3-class traffic model, we choose the QoS requirements for each of the three traffic classes by selecting the values that represent the strictest requirement among the traffic types that belong to the same traffic class. For example, in Table 3.2, high quality voice, streaming video, HDTV, video conference, and CD quality audio all belong to the real-time traffic class in our traffic model. Each of these five traffic types has its own delay and blocking probability requirements that need to be satisfied. For instance, the delay requirement for high quality voice is 10ms (i.e. mean delay for high quality voice traffic must be less than 10ms), the delay requirement for streaming video is 125-150 ms, that for HDTV is 90ms, video conference is 10ms, and CD quality is 100 ms. We notice that 10 ms mean delay is the strictest delay requirement among the traffic types in real-time traffic class, thus we choose the delay requirement for class 1 traffic to be < 10 ms. Following the same logic, the QoS requirements for the three traffic classes in our traffic model are listed below in Table 4.2.

Traffic Class	Delay (ms)	Jitter (ms)	Blocking Probability
Real-time (priority 1)	10	±5	10^-5
Interactive (priority 2)	1000	N/A	0.05
Unclassified (priority 3)	N/A	N/A	0.05

Table 4.2: QoS Requirements for the Three Traffic Classes

It is assumed that the traffic between the ISP and future home area networks will be packet-oriented, since it is the most popular mean of traffic transportation today. Furthermore, in our traffic model, to simplify calculation, it is assumed that all packets of all traffic classes are equal in size and 2500 bytes per packet. This packet size is chosen because it is within the commonly seen packet size, and video and audio traffic can be broken down into packets of such length. In our traffic model, the state space will be based on the number of packets in the system, as will be further discussed in Section 4.6.1.

4.4 EC Phase-Type Arrival Process

In our model, I decide to use the EC (Erlang-Coxian) distribution to represent the arrival process. The EC distribution was discussed in Osogami's PhD thesis [24] in detail. It is a mixture of (n-2) phase Erlang distribution and 2-phase Coxian distribution. The reason I choose the EC distribution to represent the arrival process is because, as of now there is no future home area network existing yet, so it is impossible to know what the actual arrival rate distribution of the different types of traffic will be. Since the EC distribution is able to represent most types of distribution, I chose it to represent the arrival process so that in the future, when we can accurately characterize the distribution of the different types of traffic, we can then use moment matching algorithm to a map a general distribution into an n-phase EC distribution. The details of how to map a general distribution into an n-phase EC distribution are discussed in [24]. Figure 4.2 depicts an n-phase EC distribution, which is taken from [24].

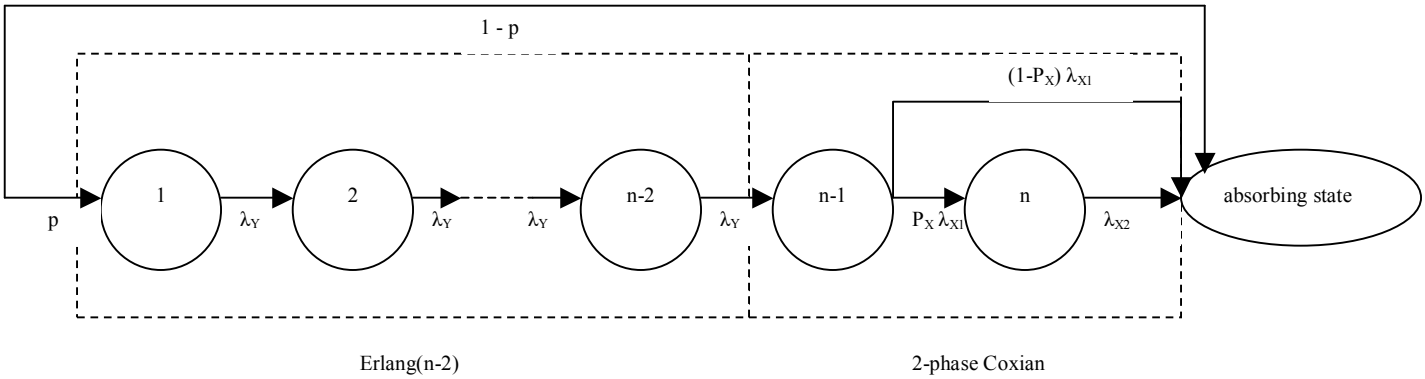


Figure 4.2: n-phase EC (Erlang-Coxian) distribution (from [25])

As described in [24], an n-phase EC distribution is a combination of an (n-2)-phase Erlang distribution and a 2-phase Coxian distribution. When a job arrives at a process that is characterized by an n-phase EC distribution, it either enters the first phase of the (n-2)-phase Erlang distribution with probability p or enters absorption state with probability 1 – p. If the job enters the first phase of the (n-2)-phase Erlang distribution, it then enters phase 2 of the (n-2)-phase Erlang distribution with transition rate λ_Y, and then enters phase 3 of the (n-2)-phase Erlang distribution with transition rate λ_Y, and so on, until it enters the $(n-2)^{th}$ phase of the (n-2)-phase Erlang distribution. After that the job enters the first phase of the 2-phase Coxian distribution (i.e. the $(n-1)^{th}$ phase in the process) with transition rate λ_Y. It then either enters the second phase of the 2-phase Coxian distribution with transition rate $p_x \lambda_{X1}$ and enters absorption state with transition rate $(1 - p_x) \lambda_{X1}$, where p_x represents the probability of a job transiting from phase n-1 to phase n in the process. If a job is in phase n, it then enters absorption state with transition rate λ_{X2}.

An n-phase EC distribution is particularly useful in approximating a general distribution for the purpose of our traffic model because of the properties of Erlang and Coxian distributions. A 2-phase Coxian distribution can approximate any distribution with high second and third moments very well [4, 23]. However, when a general distribution has low second and third moment, a 2-phase Coxian distribution requires many phases to approximate it. By contrast, an n-phase Erlang distribution has the least variability among all n-phase PH distributions [3], but it is limited in the set of distributions it can represent accurately [24]. Thus, as explained in [24], intuitively it makes sense to combine the Erlang and Coxian distributions into an n-phase EC distribution to approximate a general distribution, and this is the reason why behind choosing an n-phase EC distribution to represent our arrival process.

In the sample calculations, we only use the 2-phase Coxian distribution represented by (α, T_i) of dimension 2 to represent the arrival process, since it is simpler for the purpose of numerical calculation. The arrival rate λ_i, where $i = \{1, 2, 3\}$ follows a 2-phase PH distribution with Coxian representation, as depicted in Figure 4.3 below.

Figure 4.3: A 2-phase PH distribution arrival rate with Coxian representation.

According to this representation, a job first enters phase 1. It then either moves to phase 2 with probability p_i or to job arrival (i.e. absorption) with probability $q_i = 1 - p_i$. In matrix form, this is represented by

$$\alpha = [\,1\ 0\,] \qquad T_i^{o} = \begin{bmatrix} \lambda_i^{(1)} q_i \\ \lambda_i^{(2)} \end{bmatrix}$$

$$T_i = \begin{bmatrix} -\lambda_i^{(1)} & \lambda_i^{(1)} p_i \\ 0 & -\lambda_i^{(2)} \end{bmatrix}$$

For each traffic class i, the mean arrival rate $\overline{\lambda_i}$ can be represented by the equation

$$\overline{\lambda_i} = (-\alpha T_i^{-1} e)^{-1} \qquad (4.1)$$

where

e = column vector of 1 of dimension 1 by 2

The packet arrival rate into the first phase of the arrival process, $\lambda_1^{(1)}$, in Equation (4.1), is determined by three variables: the basic bandwidth rate of class i traffic (r_i), the average number of class i applications being used by one household (n_i), and the fraction of households served by the ISP actively using class i applications (δ_i). Multiplying the basic bandwidth of class i traffic r_i by the average number of class i applications being used by one household n_i gives the total bandwidth used by one household for class i traffic. However, at any particular instant, it is likely that only a certain fraction of households served by an ISP would be actively using class i applications. Thus we further multiply the result of $r_i \times n_i$ by δ_i, the fraction of households served by the ISP that are actively using class i applications. Thus, the packet arrival rate into the first phase of the arrival process is represented by Equation (4.2) below.

$$\lambda_i^{(1)} = \delta_i\, r_i\, n_i \qquad (4.2)$$

where

δ_i = fraction of households served by an ISP that is actively using class i applications
r_i = basic bandwidth of class i traffic
n_i = average number of class i applications (jobs) being used by one household

For priority class 1 traffic in particular, since it is composed of real-time video, real-time audio, telephony, and control signal traffic, the packet arrival rate into the first phase of the arrival process for class 1 traffic can be further expanded into equation (4.3) as follow:

$$\lambda_1^{(1)} = \delta_{video}\, r_{video}\, n_{video} + \delta_{audio}\, r_{audio}\, n_{audio} + \delta_{tel}\, r_{tel}\, n_{tel} + \delta_{control}\, r_{control}\, n_{control} \qquad (4.3)$$

where

δ_{video} = fraction of households served by the ISP that are actively using real-time video traffic
δ_{audio} = fraction of households served by the ISP that are actively using real-time audio traffic
δ_{tel} = fraction of households served by the ISP that are actively using telephony traffic
$\delta_{control}$ = fraction of households served by the ISP that are actively using control signal traffic

r_{video} = basic bandwidth of video traffic
n_{video} = average number of video applications being used by one household

r_{audio} = basic bandwidth of audio traffic
n_{audio} = average number of audio applications being used by one household

r_{tel} = basic bandwidth of telephone traffic
n_{tel} = average number of telephone applications being used by one household

$r_{control}$ = basic bandwidth of control signal traffic
$n_{control}$ = average number of applications using control signal traffic in one household

Using a diagram, the arrival process of the traffic model is depicted in Figure 4.4 below.

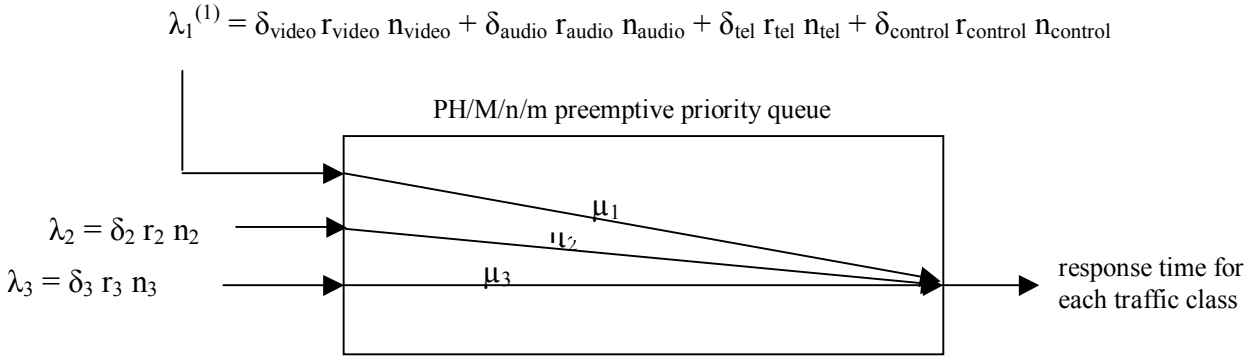


Figure 4.4: High-level diagram of ISP-HAN traffic model

4.5 Exponential Service Process

In our traffic model, it is assumed that the service time of a class i packet follows an exponential distribution with mean service rate μ_i. As mentioned before, the service time follows a service discipline of preemptive priority FIFO with delay-loss. Upon the arrival of a packet, there are four possible scenarios related to whether the newly arriving packet will receive service or not:

1. (number of packets in system < n) If there are less than n packets in the system, the arriving packet randomly chooses a server and enters into service immediately.
2. (number of packets in system ≥ n) If there are n or more than n packets in the system already, and if one of the n packets already receiving service is of a lower priority than the new packet, then the new packet of higher priority randomly chooses a server among the pool of servers that are serving a lower priority packet and receives service immediately. The lower priority packet that has been receiving service from the chosen server is preempted and placed back into the appropriate buffer.
3. (number of packets in system ≥ n and < $n+m_i$) If there are already n or more than n packets in the system and if all of the n packets already receiving service are of the same or higher priority than the new packet, then the new packet joins the buffer if the buffer for that traffic class is not full yet.
4. (number of packets in system ≥ n and > $n+m_i$) If there are n or more than n packets in the system already, and if all of the n packets already receiving service are of the same or higher priority than the new packet, and if the buffer for that traffic class is already full, then the new packet is lost.

Thus, following the above service discipline scheme, the following is what will happen to a newly arrived priority i packet:

Let x_1 = number of priority 1 packets already in system
x_2 = number of priority 2 packets already in system
x_3 = number of priority 3 packets already in system

- Upon the arrival of a priority 1 packet:
  - If $x_1 < n$: priority 1 packet enters a free server at random and receives service from it for a random time period
  - If $n < x_1 \leq n+m_1$: priority 1 packet queues at buffer 1
  - If $x_1 > n+m_1$: packet loss
- Upon the arrival of a priority 2 packet:
  - If $x_2 < n - x_1$: priority 2 packet enters a free server at random and receives service from it for a random time period
  - If $n- x_1 < x_2 \leq n+m_2- x_1$: priority 2 packet queues at buffer 2
  - If $x_2 > n+m_2- x_1$: packet loss
- Upon the arrival of a priority 3 packet:
  - If $x_3 < n- x_1- x_2$: priority 3 customer enters a free server at random and receives service from it for a random time period
  - If $n- x_1- x_2 < z \leq n+m_3- x_1- x_2$: priority 3 customer queues at buffer 3
  - If $x_3 > n+m_3- x_1- x_2$: packet loss

4.6 Construction of the Generator Matrix Q

4.6.1 State Space

In this thesis, the system depicted in Figure 4.1 is modeled as a continuous-time stochastic process $\{X(t), t \geq 0\}$ with state space $\Delta = \{S_1, S_2, S_3\}$. S_1, S_2, and S_3 represent the state spaces of class 1, 2, and 3 traffic respectively The definitions of S_1, S_2, and S_3 will be given shortly. More specifically, the stochastic process X(t) in this thesis is a continuous-time Markov chain where the conditional probability distribution of the future state depends only on the present state and is independent of the past states. In other words,

$$P\{X(t+s) = j \mid X(s) = i, X(u) = x(u), 0 \leq u < s\}$$
$$= P\{X(t+s) = j \mid X(s) = i\}$$

for all s, $t \geq 0$ and nonnegative integers i, j, x(u), $0 \leq u < s$.

This Markov chain is captured in the generator matrix Q given in this thesis. The generator matrix Q has 4 levels. The first three levels represent the class 1, 2, and 3 traffic class respectively. The state space of priority class 1, S_1, is $(n+m_1+1)S_2$, where n = number of servers, m_1 = number of waiting spaces for class 1 traffic, and S_2 = state space of class 2 traffic. This means that the state space of priority class 1 consists of the sum of the total number of servers plus the number of waiting spaces in the buffer for class 1 packets in the system.

The state space of priority class 2, S_2, is $\max((n+m_2+1-i)S_3, (m_2+1)S_3)$, where n = number of servers, m_2 = number of buffer slots for class 2 packets, i = number of class 1 packets already in the system and S_3 = state space of class 3 traffic. $n+m_2+1$ gives the state space of priority class 2 when there is no class 1 packet in the system already. Since the service discipline is preemptive priority, this means if there is any class 1 packet in the system, it will receive service first and preempts any class 2 packet already receiving service. Thus the number of class 2 packets in the system is equal to $n+m_2+1$ minus the number of class 1 packets in the system. This gives the term $(n+m_2+1-i)S_3$.

$\max((n+m_2+1-i)S_3, (m_2+1)S_3)$ means that the state space is the maximum of $(n+m_2+1-i)S_3$ or $(m_2+1)S_3$. The necessity of this max operation can be explained by the following example. Let's say the number of class 1 packets already in the system, i, is greater than the number of servers in the system, thus i > n. According to the equation that makes the state space of priority class 2 $S_2 = (n+m_2-i+1)S_3$, this gives $S_2 < m_2+1$. However, this is not correct, since even if all the servers have been occupied by class 1 packets, the state space of class 2 packets should still be m_2+1, since the buffer space of class 2 packets is exclusively for class 2 packets and is not affected by packets of the other classes. Thus, the equation for the state space of class 2 packets $S_2 = \max((n+m_2+1-i)S_3, (m_2+1)S_3)$. Note that the number of class 3 packets already in system has no consequence on the state space of priority class 2, since class 2 packets have preemptive priority over class 3 packets.

The state space of priority class 3 $S_3 = (\max(n+m_3-i-j+1, m_3+1)) \times (pn_1 \times pn_2 \times pn_3)$, where n = number of servers, m_3 = number of buffer slots for class 3 packets, i = number of class 1 packets already in the system, j = the number of class 2 packets already in the system, and pn_i = umber of phases in the arrival processk for class i traffic where i = {1, 2, 3}. Similar to the logic used in obtaining the state space of

priority class 2, $n+m_3+1-i-j$ gives the state space of priority class 3 given i priority 1 packets and j number of priority 2 packets already in the system. $(\max(n+m_3-i-j+1), m_3+1)\times(pn_1 \times pn_2 \times pn_3)$ means that the state space is the maximum of $(n+m_2-i+1)\times(pn_1 \times pn_2 \times pn_3)$ or $(m_3+1)\times(pn_1 \times pn_2 \times pn_3)$. As in the case of priority class 2 state space calculation, this max operation ensures that the state space of priority class 3 has a minimum of m_3+1, since the buffer for priority class 3 is unaffected by the number of class 1 and 2 packets already in the system.

Note that in the calculation of S_3, the term $\max(n+m_3-i-j+1, m_3+1)$ is multiplied $(pn_1 \times pn_2 \times pn_3)$. This is because the generator matrix Q needs to keep track of which phase the n-phase arrival process is currently in. For example, if the arrival process is a 2-phase arrival process for all three traffic classes, then $(pn_1 \times pn_2 \times pn_3)$ = 2 x 2 x 2 = 8, or if the arrival process is a 3-phase arrival process with 3 traffic classes, then $(pn_1 \times pn_2 \times pn_3)$ = 3 x 3 x 3 = 27.

In summary, the size of the state space of the generator matrix Q, S, and priority class i, S_i, is given by the following equations:

State space $S_1 = (n+m_1+1)S_2$ (4.4)

State space $S_2 = \max((n+m_2-i+1)S_3, (m_2+1)S_3)$ (4.5)

State space $S_3 = (\max(n+m_3-i-j+1, m_3+1))\times(pn_1 \times pn_2 \times pn_3)$ (4.6)

$\Delta = \{(S_1, S_2, S_3) \mid 0 \le S_1 \le (n+m_1+1)S_2,\ 0 \le S_2 \le \max((n+m_2+1-i)S_3, (m_2+1)S_3),\ 0 \le S_3 \le (\max(n+m_3+1-i-j, m_3+1)\times(pn_1 \times pn_2 \times pn_3)\}$

where

S = state space of generator matrix Q
S_1 = state space of priority class 1
S_2 = state space of priority class 2
S_3 = state space of priority class 3
n = number of servers
m_1 = number of waiting spaces in the buffer for class 1 packets
m_2 = number of waiting spaces in the buffer for class 2 packets
m_3 = number of waiting spaces in the buffer for class 3 packets
i = number of class 1 packets already in the system
j = number of class 2 packets already in the system
pn_i = number of phases in the n-phase arrival process for class i traffic, where i = {1, 2, 3}

The states in the generator matrix Q represent the number of packets and the arrival phase in the system. There are three possible events that would cause a change of state of the generator matrix Q:

1. *Packet arrival*. A packet arrives when the buffer is not full. This is equivalent to the case when the arrival process enters absorption state, resulting in an increase in the number of packets in the system by one.

2. *Service completion*. A packet has finished receiving service from the system, resulting in a decrease in the number of packets in the system by one.
3. *Phase change in arrival process without packet arrival*. This occurs when a packet arrives at the first phase of the arrival process or changes from one phase to another without entering the absorption state of the arrival process. There is no change in the number of packets in the system.

4.6.2 The Generator Matrix Q

The construction algorithm for the generator matrix is presented below, followed by the construction algorithm in matrix forms.

Construction Algorithm for Q Matrix

The generator matrix, Q, is a matrix with dimension $(n+m_1+1)$ by $(n+m_1+1)$.
The mathematical form of the construction algorithm for the generator matrix Q can be represented as:

$$Q = \begin{bmatrix} Q_{i,j} & Q_{i,j} & & & & & \\ & Q_{i,j} & Q_{i,j} & Q_{i,j} & & & \\ & & \cdot & \cdot & \cdot & & \\ & & & \cdot & \cdot & \cdot & \\ & & & & Q_{i,j} & Q_{i,j} & Q_{i,j} \\ & & & & & Q_{i,j} & Q_{i,j} \end{bmatrix}$$

In mathematical form, the Q matrix can be represented as follow:

$$Q_{i,j} = \begin{cases} A_{0,0} & \forall i = 0, j = 0 \\ A_{0,1} & \forall i = 0, j = 1 \\ A_{i,j} & \forall 1 \le i \le n, i-1 \le j \le i+1 \\ A_2 & \forall n+1 \le i \le n+m_1, j = i-1 \\ A_1 & \forall n+1 \le i \le n+m_1-1, j = i \\ A_0 & \forall n+1 \le i \le n+m_1-1, j = i+1 \\ A_1 + A_0 & \forall i = n+m_1, j = i \\ 0 & \forall otherwise \end{cases}$$

The above equation can be expressed in matrix form as follow:

$$
Q = \begin{bmatrix}
A_{0,0} & A_{0,1} & & & & & & & & & \\
A_{1,0} & A_{1,1} & A_{1,2} & & & & & & & & \\
 & A_{2,1} & A_{2,2} & A_{2,3} & & & & & & & \\
 & & \ddots & \ddots & \ddots & & & & & & \\
 & & & \ddots & \ddots & \ddots & & & & & \\
 & & & & \ddots & \ddots & \ddots & & & & \\
 & & & & & \ddots & \ddots & \ddots & & & \\
 & & & & & & A_{n,n-1} & A_{n,n} & A_{n,n+1} & & \\
 & & & & & & & A_2 & A_1 & A_0 & \\
 & & & & & & & & \ddots & \ddots & \ddots \\
 & & & & & & & & & A_1 & A_0 \\
 & & & & & & & & & A_2 & A_1 + A_0
\end{bmatrix} \qquad (4.7)
$$

Details about the construction algorithm for the sub-matrices $A_{i,j}$ in the generator matrix Q are described in Appendix A.

4.7 Example of Generator Matrix Q

To illustrate the above construction algorithm, we show the generator matrix Q of a small system where the number of servers n = 8 and the size of the system m = 4. Following the construction algorithm for the generator matrix Q in Section 4.6.2 and Appendix A, the generator matrix Q of such a system is

$$
Q = \begin{bmatrix}
A_{0,0} & A_{0,1} & & & & & & & & & & & \\
A_{1,0} & A_{1,1} & A_{1,2} & & & & & & & & & & \\
 & A_{2,1} & A_{2,2} & A_{2,3} & & & & & & & & & \\
 & & A_{3,2} & A_{3,3} & A_{3,4} & & & & & & & & \\
 & & & A_{4,3} & A_{4,4} & A_{4,5} & & & & & & & \\
 & & & & A_{5,4} & A_{5,5} & A_{5,6} & & & & & & \\
 & & & & & A_{6,5} & A_{6,6} & A_{6,7} & & & & & \\
 & & & & & & A_{7,6} & A_{7,7} & A_{7,8} & & & & \\
 & & & & & & & A_{8,7} & A_{8,8} & A_{8,9} & & & \\
 & & & & & & & & A_2 & A_1 & A_0 & & \\
 & & & & & & & & & A_2 & A_1 & A_0 & \\
 & & & & & & & & & & A_2 & A_1 & A_0 \\
 & & & & & & & & & & & A_2 & A_1 + A_0
\end{bmatrix}
$$

where

$A_{0,0}$ (a 13 by 13 matrix) =

$$\begin{bmatrix}
A_{0,0,0,0} & (I_2 \otimes T_2^o \alpha \otimes I_2) I_{13,12} & & & & & & & & & & & \\
\mu_2 I_{8\times(12,13)} & A_{0,0,1,1} & (I_2 \otimes T_2^o \alpha \otimes I_2) I_{12,11} & & & & & & & & & & \\
 & 2\mu_2 I_{8\times(11,12)} & . & . & & & & & & & & & \\
 & & . & . & . & & & & & & & & \\
 & & & . & . & . & & & & & & & \\
 & & & & . & . & (I_2 \otimes T_2^o \alpha \otimes I_2) I_{6,5} & & & & & & \\
 & & & & 8\mu_2 I_{8\times(5,6)} & . & & (I_2 \otimes T_2^o \alpha \otimes I_2) I_{5,5} & & & & & \\
 & & & & & 8\mu_2 I_{8\times(5,5)} & . & & . & & & & \\
 & & & & & & . & . & & & & & \\
 & & & & & & & . & A_{0,0,8,8} & (I_2 \otimes T_2^o \alpha \otimes I_2) I_{5,5} & & & \\
 & & & & & & & & 8\mu_2 I_{8\times(5,5)} & A_{0,0,n+m_2-i,n+m_2-i} & & &
\end{bmatrix}$$

(Note 1:

$$(I_2 \otimes T_2{}^o \alpha \otimes I_2) I_{6,5} = \begin{bmatrix}
I_2 \otimes T_2{}^o \alpha \otimes I_2 & & & & & & \\
 & I_2 \otimes T_2{}^o \alpha \otimes I_2 & & & & & \\
 & & I_2 \otimes T_2{}^o \alpha \otimes I_2 & & & & \\
 & & & I_2 \otimes T_2{}^o \alpha \otimes I_2 & & & \\
 & & & & I_2 \otimes T_2{}^o \alpha \otimes I_2 & & \\
 & & & & & 0_{8,8}
\end{bmatrix}$$

$0_{8,8}$ = a zero matrix of size 8 by 8

$$(I_2 \otimes T_2{}^o \alpha \otimes I_2) I_{5,5} = \begin{bmatrix}
I_2 \otimes T_2{}^o \alpha \otimes I_2 & & & & & \\
 & I_2 \otimes T_2{}^o \alpha \otimes I_2 & & & & \\
 & & I_2 \otimes T_2{}^o \alpha \otimes I_2 & & & \\
 & & & I_2 \otimes T_2{}^o \alpha \otimes I_2 & & \\
 & & & & I_2 \otimes T_2{}^o \alpha \otimes I_2
\end{bmatrix}$$

and so on. This notation will be encountered throughout this section and in Appendix A).

(Note 2: $I_{8x(11,12)} = I_{88,96}$, $I_{8x(5,6)} = I_{40,48}$, and so on. This notation will also be encountered throughout this section and in Appendix A.)

$A_{1,1}$ (a 12 by 12 matrix) =

$$\begin{bmatrix}
A_{1,1,0,0} & (I_2 \otimes T_2^o \alpha \otimes I_2) I_{12,11} & & & & & & & & & & \\
\mu_2 I_{8\times(11,12)} & A_{1,1,1,1} & (I_2 \otimes T_2^o \alpha \otimes I_2) I_{11,10} & & & & & & & & & \\
 & 2\mu_2 I_{8\times(10,11)} & . & . & & & & & & & & \\
 & & . & . & . & & & & & & & \\
 & & & . & A_{1,1,6,6} & (I_2 \otimes T_2^o \alpha \otimes I_2) I_{6,5} & & & & & & \\
 & & & & 7\mu_2 I_{8\times(5,6)} & A_{1,1,7,7} & (I_2 \otimes T_2^o \alpha \otimes I_2) I_{5,5} & & & & & \\
 & & & & & 7\mu_2 I_{8\times(5,5)} & . & . & & & & \\
 & & & & & & . & . & . & & & \\
 & & & & & & & . & A_{1,1,7,7} & (I_2 \otimes T_2^o \alpha \otimes I_2) I_{5,5} & & \\
 & & & & & & & & 7\mu_2 I_{8\times(5,5)} & A_{1,1,11,11} & &
\end{bmatrix}$$

$A_{2,2}$ (11 by 11 matrix) =

$$\begin{bmatrix}
A_{2,2,0,0} & (I_2 \otimes T_2^o \alpha \otimes I_2) I_{11,10} & & & & & & & & & \\
\mu_2 I_{8\times(10,11)} & A_{2,2,1,1} & (I_2 \otimes T_2^o \alpha \otimes I_2) I_{10,9} & & & & & & & & \\
 & 2\mu_2 I_{8\times(9,10)} & . & . & & & & & & & \\
 & & . & . & . & & & & & & \\
 & & & . & . & (I_2 \otimes T_2^o \alpha \otimes I_2) I_{6,5} & & & & & \\
 & & & & 6\mu_2 I_{8\times(5,6)} & A_{2,2,6,6} & (I_2 \otimes T_2^o \alpha \otimes I_2) I_{5,5} & & & & \\
 & & & & & 6\mu_2 I_{8\times(5,5)} & . & . & & & \\
 & & & & & & . & . & . & & \\
 & & & & & & & . & A_{2,2,6,6} & (I_2 \otimes T_2^o \alpha \otimes I_2) I_{5,5} & \\
 & & & & & & & & 6\mu_2 I_{8\times(5,5)} & A_{2,2,10,10} &
\end{bmatrix}$$

$A_{7,7}$ (6 by 6 matrix) =

$$\begin{bmatrix}
A_{7,7,0,0} & (I_2 \otimes T_2{}^o \alpha \otimes I_2) I_{6,5} & & & & \\
\mu_2 I_{8\times(5,6)} & A_{7,7,1,1} & (I_2 \otimes T_2{}^o \alpha \otimes I_2) I_{5,5} & & & \\
 & \mu_2 I_{8\times(5,5)} & . & . & & \\
 & & . & . & . & \\
 & & & . & A_{7,7,1,1} & (I_2 \otimes T_2{}^o \alpha \otimes I_2) I_{5,5} \\
 & & & & \mu_2 I_{8\times(5,5)} & A_{7,7,5,5}
\end{bmatrix}$$

A_1 (5 by 5 matrix) = $\begin{bmatrix}
A_{8,8,0,0} & (I_2 \otimes T_2{}^o \alpha \otimes I_2) I_5 & & & \\
0 & . & . & & \\
 & . & . & . & \\
 & & . & A_{8,8,0,0} & (I_2 \otimes T_2{}^o \alpha \otimes I_2) I_5 \\
 & & & 0 & A_{8,8,4,4}
\end{bmatrix}$

$A_{0,0,0,0}$ (13 by 13 matrix) =

$$\begin{bmatrix}
T_1\oplus T_2\oplus T_3 & I_2\otimes I_2\otimes T_3^o\alpha & & & & & & & & & & & \\
\mu_3 I_8 & (T_1\oplus T_2\oplus T_3)-\mu_3 I_8 & . & & & & & & & & & & \\
 & 2\mu_3 I_8 & (T_1\oplus T_2\oplus T_3)-2\mu_3 I_3 & . & & & & & & & & & \\
 & & . & . & . & & & & & & & & \\
 & & & . & . & . & & & & & & & \\
 & & & & & 8\mu_3 I_8 & (T_1\oplus T_2\oplus T_3)-8\mu_3 I_8 & . & & & & & \\
 & & & & & & . & . & . & & & & \\
 & & & & & & & . & (T_1\oplus T_2\oplus T_3)-8\mu_3 I_8 & I_2\otimes I_2\otimes T_3^o\alpha & & & \\
 & & & & & & & & 8\mu_3 I_8 & (T_1\oplus T_2)-8\mu_3 I_8 & & &
\end{bmatrix}$$

$A_{0,0,1,1}$ (12 by 12 matrix) =

$$\begin{bmatrix}
(T_1\oplus T_2\oplus T_3)-\mu_2 I_8 & I_2\otimes I_2\otimes T_3^o\alpha & & & & & & & & & & \\
\mu_3 I_8 & (T_1\oplus T_2\oplus T_3)-\mu_2 I_8-\mu_3 I_8 & . & & & & & & & & & \\
 & 2\mu_3 I_8 & (T_1\oplus T_2\oplus T_3)-\mu_2 I_8-2\mu_3 I_8 & . & & & & & & & & \\
 & & . & . & . & & & & & & & \\
 & & & & 7\mu_3 I_8 & (T_1\oplus T_2\oplus T_3)-\mu_2 I_8-7\mu_3 I_8 & . & & & & & \\
 & & & & & . & . & . & & & & \\
 & & & & & & . & (T_1\oplus T_2\oplus T_3)-\mu_2 I_8-7\mu_3 I_8 & I_2\otimes I_2\otimes T_3^o\alpha & & \\
 & & & & & & & 8\mu_3 I_8 & (T_1\oplus T_2)-\mu_2 I_8-7\mu_3 I_8 & &
\end{bmatrix}$$

$A_{0,0,2,2}$ (11 by 11 matrix) =

$$\begin{bmatrix}
(T_1\oplus T_2\oplus T_3)-2\mu_2 I_8 & I_2\otimes I_2\otimes T_3^o\alpha & & & & & & & & & \\
\mu_3 I_8 & (T_1\oplus T_2\oplus T_3)-2\mu_2 I_8-\mu_3 I_8 & . & & & & & & & & \\
 & . & . & . & & & & & & & \\
 & & 6\mu_3 I_8 & (T_1\oplus T_2\oplus T_3)-2\mu_2 I_8-6\mu_3 I_8 & . & & & & & & \\
 & & & . & . & . & & & & \\
 & & & & 6\mu_3 I_8 & (T_1\oplus T_2\oplus T_3)-2\mu_2 I_8-6\mu_3 I_8 & I_2\otimes I_2\otimes T_3^o\alpha & & & & \\
 & & & & & 6\mu_3 I_8 & (T_1\oplus T_2)-2\mu_2 I_8-6\mu_3 I_8 & & & &
\end{bmatrix}$$

$A_{0,0,7,7}$ (6 by 6 matrix) =

$$\begin{bmatrix} (T_1 \oplus T_2 \oplus T_3) - 7\mu_2 I_8 & I_2 \otimes I_2 \otimes T_3^o \alpha & & & & \\ \mu_3 I_8 & (T_1 \oplus T_2 \oplus T_3) - 7\mu_2 I_8 - \mu_3 I_8 & . & & & \\ & . & . & . & & \\ & & . & . & . & \\ & & & . & (T_1 \oplus T_2 \oplus T_3) - 7\mu_2 I_8 - \mu_3 I_8 & I_2 \otimes I_2 \otimes T_3^o \alpha \\ & & & & \mu_3 I_8 & (T_1 \oplus T_2) - 7\mu_2 I_8 - \mu_3 I_8 \end{bmatrix}$$

$A_{0,0,8,8}$ (5 by 5 matrix) =

$$\begin{bmatrix} (T_1 \oplus T_2 \oplus T_3) - 8\mu_2 I_8 & I_2 \otimes I_2 \otimes T_3^o \alpha & & & \\ 0 & . & . & & \\ & . & . & . & \\ & & . & (T_1 \oplus T_2 \oplus T_3) - 8\mu_2 I_8 & I_2 \otimes I_2 \otimes T_3^o \alpha \\ & & & 0 & (T_1 \oplus T_2) - 8\mu_2 I_8 \end{bmatrix}$$

$A_{0,0,12,12}$ (5 by 5 matrix) =

$$\begin{bmatrix} T_3 - 8\mu_2 I_8 & I_2 \otimes I_2 \otimes T_3^o \alpha & & & \\ 0 & . & . & & \\ & . & . & . & \\ & & . & T_3 - 8\mu_2 I_8 & I_2 \otimes I_2 \otimes T_3^o \alpha \\ & & & 0 & -8\mu_2 I_8 \end{bmatrix}$$

$A_{1,1,0,0}$ (12 by 12 matrix) =

$$\begin{bmatrix} (T_1 \oplus T_2 \oplus T_3) - \mu_1 I_8 & I_2 \otimes I_2 \otimes T_3^o \alpha & & & & & & & & & & \\ \mu_3 I_8 & (T_1 \oplus T_2 \oplus T_3) - \mu_1 I_8 - \mu_3 I_8 & . & & & & & & & & & \\ & . & . & . & & & & & & & & \\ & & 7\mu_3 I_8 & (T_1 \oplus T_2 \oplus T_3) - \mu_1 I_8 - 7\mu_3 I_8 & . & & & & & & & \\ & & & . & . & . & & & & & \\ & & & & . & (T_1 \oplus T_2 \oplus T_3) - \mu_1 I_8 - 7\mu_3 I_8 & I_2 \otimes I_2 \otimes T_3^o \alpha & & & & \\ & & & & & 7\mu_3 I_8 & (T_1 \oplus T_2) - \mu_1 I_8 - 7\mu_3 I_8 & & & & \end{bmatrix}$$

$A_{1,1,1,1}$ (11 by 11 matrix) =

$$\begin{bmatrix}
(T_1\oplus T_2\oplus T_3)-\mu_1 I_8-\mu_2 I_8 & I_2\otimes I_2\otimes T_3^o\alpha & & & & & & & & & \\
\mu_3 I_8 & (T_1\oplus T_2\oplus T_3)-\mu_1 I_8-\mu_2 I_8-\mu_3 I_8 & . & & & & & & & & \\
 & . & . & . & & & & & & & \\
 & & 6\mu_3 I_8 & (T_1\oplus T_2\oplus T_3)-\mu_1 I_8-\mu_2 I_8-6\mu_3 I_8 & . & & & & & & \\
 & & & . & . & . & & & & \\
 & & & & . & (T_1\oplus T_2\oplus T_3)-\mu_1 I_8-\mu_2 I_8-6\mu_3 I_8 & I_2\otimes I_2\otimes T_3^o\alpha & & & \\
 & & & & & 6\mu_3 I_8 & (T_1\oplus T_2)-\mu_1 I_8-\mu_2 I_8-6\mu_3 I_8 & & &
\end{bmatrix}$$

$A_{1,1,6,6}$ (6 by 6 matrix) =

$$\begin{bmatrix}
(T_1\oplus T_2\oplus T_3)-\mu_1 I_8-6\mu_2 I_8 & I_2\otimes I_2\otimes T_3^o\alpha & & & & \\
\mu_3 I_8 & (T_1\oplus T_2\oplus T_3)-\mu_1 I_8-6\mu_2 I_8-\mu_3 I_8 & . & & & \\
 & . & . & . & & \\
 & & . & . & . & \\
 & & & . & (T_1\oplus T_2\oplus T_3)-\mu_1 I_8-6\mu_2 I_8-\mu_3 I_8 & I_2\otimes I_2\otimes T_3^o\alpha \\
 & & & & \mu_3 I_8 & (T_1\oplus T_2)-\mu_1 I_8-6\mu_2 I_8-\mu_3 I_8
\end{bmatrix}$$

$A_{1,1,7,7}$ (5 by 5 matrix) =

$$\begin{bmatrix}
(T_1\oplus T_2\oplus T_3)-\mu_1 I_8-7\mu_2 I_8 & I_2\otimes I_2\otimes T_3^o\alpha & & & \\
0 & (T_1\oplus T_2\oplus T_3)-\mu_1 I_8-7\mu_2 I_8 & . & & \\
 & . & . & . & \\
 & & . & (T_1\oplus T_2\oplus T_3)-\mu_1 I_8-7\mu_2 I_8 & I_2\otimes I_2\otimes T_3^o\alpha \\
 & & & 0 & (T_1\oplus T_2)-\mu_1 I_8-7\mu_2 I_8
\end{bmatrix}$$

$A_{1,1,11,11}$ (5 by 5 matrix) =

$$\begin{bmatrix}
T_3-\mu_1 I_8-7\mu_2 I_8 & I_2\otimes I_2\otimes T_3{}^o\alpha & & & \\
0 & . & . & & \\
 & . & . & . & \\
 & & . & T_3-\mu_1 I_8-7\mu_2 I_8 & I_2\otimes I_2\otimes T_3{}^o\alpha \\
 & & & 0 & -\mu_1 I_8-7\mu_2 I_8
\end{bmatrix}$$

$A_{2,2,0,0}$ (11 by 11 matrix) =

$$\begin{bmatrix}
(T_1\oplus T_2\oplus T_3)-2\mu_1 I_8 & I_2\otimes I_2\otimes T_3^o\alpha & & & & & & & & & \\
\mu_3 I_8 & (T_1\oplus T_2\oplus T_3)-2\mu_1 I_8-\mu_3 I_8 & . & & & & & & & & \\
 & . & . & . & & & & & & & \\
 & & 6\mu_3 I_8 & (T_1\oplus T_2\oplus T_3)-2\mu_1 I_8-6\mu_3 I_8 & . & & & & & & \\
 & & & . & . & . & & & & \\
 & & & & . & (T_1\oplus T_2\oplus T_3)-2\mu_1 I_8-6\mu_3 I_8 & I_2\otimes I_2\otimes T_3^o\alpha & & & & \\
 & & & & & 6\mu_3 I_8 & (T_1\oplus T_2)-2\mu_1 I_8-6\mu_3 I_8 & & & &
\end{bmatrix}$$

$A_{2,2,1,1}$ (10 by 10 matrix) =

$$\begin{bmatrix}
(T_1\oplus T_2\oplus T_3)-2\mu_1 I_8-\mu_2 I_8 & I_2\otimes I_2\otimes T_3^o\alpha & & & & & & & & \\
\mu_3 I_8 & (T_1\oplus T_2\oplus T_3)-2\mu_1 I_8-\mu_2 I_8-\mu_3 I_8 & . & & & & & & & \\
 & . & . & . & & & & & & \\
 & & 5\mu_3 I_8 & (T_1\oplus T_2\oplus T_3)-2\mu_1 I_8-\mu_2 I_8-5\mu_3 I_8 & . & & & & & \\
 & & & . & . & . & & & \\
 & & & & 5\mu_3 I_8 & (T_1\oplus T_2)-2\mu_1 I_8-\mu_2 I_8-5\mu_3 I_8 & & & &
\end{bmatrix}$$

$A_{2,2,2,2}$ (9 by 9 matrix) =

$$\begin{bmatrix}
(T_1\oplus T_2\oplus T_3)-2\mu_1 I_8-2\mu_2 I_8 & I_2\otimes I_2\otimes T_3^o\alpha & & & & & & & \\
\mu_3 I_8 & (T_1\oplus T_2\oplus T_3)-2\mu_1 I_8-2\mu_2 I_8-\mu_3 I_8 & . & & & & & & \\
 & . & . & . & & & & & \\
 & & . & . & . & & & & \\
 & & & . & (T_1\oplus T_2\oplus T_3)-2\mu_1 I_8-2\mu_2 I_8-4\mu_3 I_8 & I_2\otimes I_2\otimes T_3^o\alpha & & & \\
 & & & & 4\mu_3 I_8 & (T_1\oplus T_2)-2\mu_1 I_8-2\mu_2 I_8-4\mu_3 I_8 & & &
\end{bmatrix}$$

$A_{2,2,5,5}$ (6 by 6 matrix) =

$$\begin{bmatrix}
(T_1\oplus T_2\oplus T_3)-2\mu_1 I_8-5\mu_2 I_8 & I_2\otimes I_2\otimes T_3^o\alpha & & & & \\
\mu_3 I_8 & (T_1\oplus T_2\oplus T_3)-2\mu_1 I_8-5\mu_2 I_8-\mu_3 I_8 & . & & & \\
 & . & . & . & & \\
 & & . & (T_1\oplus T_2\oplus T_3)-2\mu_1 I_8-5\mu_2 I_8-\mu_3 I_8 & I_2\otimes I_2\otimes T_3^o\alpha & \\
 & & & \mu_3 I_8 & (T_1\oplus T_2)-2\mu_1 I_8-5\mu_2 I_8-\mu_3 I_8 &
\end{bmatrix}$$

$A_{2,2,6,6}$ (5 by 5 matrix) =

$$\begin{bmatrix} (T_1 \oplus T_2 \oplus T_3) - 2\mu_1 I_8 - 6\mu_2 I_8 & I_2 \otimes I_2 \otimes T_3^{\,o}\alpha & & & \\ 0 & . & . & & \\ & . & . & . & \\ & & . & (T_1 \oplus T_2 \oplus T_3) - 2\mu_1 I_8 - 6\mu_2 I_8 & I_2 \otimes I_2 \otimes T_3^{\,o}\alpha \\ & & & 0 & (T_1 \oplus T_2) - 2\mu_1 I_8 - 6\mu_2 I_8 \end{bmatrix}$$

$A_{2,2,10,10}$ (5 by 5 matrix) =

$$\begin{bmatrix} T_3 - 2\mu_1 I_8 - 6\mu_2 I_8 & I_2 \otimes I_2 \otimes T_3^{\,o}\alpha & & & \\ 0 & . & . & & \\ & . & . & . & \\ & & . & T_3 - 2\mu_1 I_8 - 6\mu_2 I_8 & I_2 \otimes I_2 \otimes T_3^{\,o}\alpha \\ & & & 0 & -2\mu_1 I_8 - 6\mu_2 I_8 \end{bmatrix}$$

$A_{7,7,0,0}$ (6 by 6 matrix) =

$$\begin{bmatrix} (T_1 \oplus T_2 \oplus T_3) - 7\mu_1 I_8 & I_2 \otimes I_2 \otimes T_3^{\,o}\alpha & & & & \\ \mu_3 I_8 & (T_1 \oplus T_2 \oplus T_3) - 7\mu_1 I_8 - \mu_3 I_8 & . & & & \\ & . & . & . & & \\ & & . & (T_1 \oplus T_2 \oplus T_3) - 7\mu_1 I_8 - \mu_3 I_8 & I_2 \otimes I_2 \otimes T_3^{\,o}\alpha \\ & & & \mu_3 I_8 & (T_1 \oplus T_2) - 7\mu_1 I_8 - \mu_3 I_8 \end{bmatrix}$$

$A_{7,7,1,1}$ (5 by 5 matrix) =

$$\begin{bmatrix} (T_1 \oplus T_2 \oplus T_3) - 7\mu_1 I_8 - \mu_2 I_8 & I_2 \otimes I_2 \otimes T_3^{\,o}\alpha & & & \\ 0 & . & . & & \\ & . & . & . & \\ & & . & . & I_2 \otimes I_2 \otimes T_3^{\,o}\alpha \\ & & & 0 & (T_1 \oplus T_2) - 7\mu_1 I_8 - \mu_2 I_8 \end{bmatrix}$$

$A_{7,7,5,5}$ (5 by 5 matrix) =

$$\begin{bmatrix} T_3 - 7\mu_1 I_8 - \mu_2 I_8 & I_2 \otimes I_2 \otimes T_3{}^o\alpha & & & \\ 0 & \ddots & \ddots & & \\ & \ddots & \ddots & \ddots & \\ & & \ddots & \ddots & I_2 \otimes I_2 \otimes T_3{}^o\alpha \\ & & & 0 & -7\mu_1 I_8 - \mu_2 I_8 \end{bmatrix}$$

$A_{8,8,0,0}$ (5 by 5 matrix) =

$$\begin{bmatrix} (T_1 \oplus T_2 \oplus T_3) - 8\mu_1 I_8 & I_2 \otimes I_2 \otimes T_3{}^o\alpha & & & \\ 0 & \ddots & \ddots & & \\ & \ddots & \ddots & \ddots & \\ & & \ddots & (T_1 \oplus T_2 \oplus T_3) - 8\mu_1 I_8 & I_2 \otimes I_2 \otimes T_3{}^o\alpha \\ & & & 0 & (T_1 \oplus T_2) - 8\mu_1 I_8 \end{bmatrix}$$

$A_{8,8,4,4}$ (5 by 5 matrix) =

$$\begin{bmatrix} (T_1 \oplus T_2 \oplus T_3) - 8\mu_1 I_8 & I_2 \otimes I_2 \otimes T_3{}^o\alpha & & & \\ 0 & \ddots & \ddots & & \\ & \ddots & \ddots & \ddots & \\ & & \ddots & (T_1 \oplus T_2 \oplus T_3) - 8\mu_1 I_8 & I_2 \otimes I_2 \otimes T_3{}^o\alpha \\ & & & 0 & (T_1 \oplus T_2) - 8\mu_1 I_8 \end{bmatrix}$$

$A_{0,1}$ (12 by 13 matrix) =

$$\begin{bmatrix} (T_1{}^o\alpha \otimes I_2 \otimes I_2) I_{13,12} & & & & & & & & & & & \\ & (T_1{}^o\alpha \otimes I_2 \otimes I_2) I_{12,11} & & & & & & & & & & \\ & & \ddots & & & & & & & & & \\ & & & \ddots & & & & & & & & \\ & & & & (T_1{}^o\alpha \otimes I_2 \otimes I_2) I_{6,5} & & & & & & & \\ & & & & & (T_1{}^o\alpha \otimes I_2 \otimes I_2) I_{5,5} & & & & & & \\ & & & & & & \ddots & & & & & \\ & & & & & & & (T_1{}^o\alpha \otimes I_2 \otimes I_2) I_{5,5} & & & & \\ & & & & & & & & 0 & & & \end{bmatrix}$$

$A_{1,2}$ (11 by 12 matrix) =

$$\begin{bmatrix} (T_1^o\alpha\otimes I_2\otimes I_2)I_{12,11} & & & & & & & & & & \\ & (T_1^o\alpha\otimes I_2\otimes I_2)I_{11,10} & & & & & & & & & \\ & & \cdot & & & & & & & & \\ & & & \cdot & & & & & & & \\ & & & & (T_1^o\alpha\otimes I_2\otimes I_2)I_{6,5} & & & & & & \\ & & & & & (T_1^o\alpha\otimes I_2\otimes I_2)I_{5,5} & & & & & \\ & & & & & & \cdot & & & & \\ & & & & & & & (T_1^o\alpha\otimes I_2\otimes I_2)I_{5,5} & & \\ & & & & & & & & 0 & & \end{bmatrix}$$

$A_{2,3}$ (10 by 11 matrix) =

$$\begin{bmatrix} (T_1^o\alpha\otimes I_2\otimes I_2)I_{11,10} & & & & & & & & & & \\ & (T_1^o\alpha\otimes I_2\otimes I_2)I_{10,9} & & & & & & & & & \\ & & \cdot & & & & & & & & \\ & & & \cdot & & & & & & & \\ & & & & (T_1^o\alpha\otimes I_2\otimes I_2)I_{6,5} & & & & & & \\ & & & & & (T_1^o\alpha\otimes I_2\otimes I_2)I_{5,5} & & & & & \\ & & & & & & \cdot & & & & \\ & & & & & & & (T_1^o\alpha\otimes I_2\otimes I_2)I_{5,5} & & \\ & & & & & & & & 0 & & \end{bmatrix}$$

$A_{6,7}$ (6 by 7 matrix) =

$$\begin{bmatrix} (T_1^{\,o}\alpha\otimes I_2\otimes I_2)I_{7,6} & & & & & & \\ & (T_1^{\,o}\alpha\otimes I_2\otimes I_2)I_{6,5} & & & & & \\ & & (T_1^{\,o}\alpha\otimes I_2\otimes I_2)I_{5,5} & & & & \\ & & & \cdot & & & \\ & & & & \cdot & & \\ & & & & & (T_1^{\,o}\alpha\otimes I_2\otimes I_2)I_{5,5} & \\ & & & & & 0 & \end{bmatrix}$$

$A_{7,8}$ (5 by 6 matrix) = $\begin{bmatrix} (T_1^o \alpha \otimes I_2 \otimes I_2) I_{6,5} & & & & & \\ & (T_1^o \alpha \otimes I_2 \otimes I_2) I_{5,5} & & & & \\ & & \ddots & & & \\ & & & \ddots & & \\ & & & & (T_1^o \alpha \otimes I_2 \otimes I_2) I_{5,5} & \\ & & & & & 0 \end{bmatrix}$

A_0 (5 by 5 matrix) = $\begin{bmatrix} (T_1^o \alpha \otimes I_2 \otimes I_2) I_5 & & & & \\ & \ddots & & & \\ & & \ddots & & \\ & & & \ddots & \\ & & & & (T_1^o \alpha \otimes I_2 \otimes I_2) I_5 \end{bmatrix}$

$A_{1,0}$ (13 by 12 matrix) = $\begin{bmatrix} \mu_1 I_{8\times(12,13)} & & & & & & & & & & & \\ & \mu_1 I_{8\times(11,12)} & & & & & & & & & & \\ & & \ddots & & & & & & & & & \\ & & & \ddots & & & & & & & & \\ & & & & \mu_1 I_{8\times(5,6)} & & & & & & & \\ & & & & & \mu_1 I_{8\times(5,5)} & & & & & & \\ & & & & & & \ddots & & & & & \\ & & & & & & & \mu_1 I_{8\times(5,5)} & & & & 0 \end{bmatrix}$

$A_{2,1}$ (12 by 11 matrix) =

$\begin{bmatrix} 2\mu_1 I_{8\times(11,12)} & & & & & & & & & & \\ & 2\mu_1 I_{8\times(10,11)} & & & & & & & & & \\ & & \ddots & & & & & & & & \\ & & & \ddots & & & & & & & \\ & & & & 2\mu_1 I_{8\times(5,6)} & & & & & & \\ & & & & & 2\mu_1 I_{8\times(5,5)} & & & & & \\ & & & & & & \ddots & & & & \\ & & & & & & & 2\mu_1 I_{8\times(5,5)} & & & 0 \end{bmatrix}$

$A_{3,2}$ (11 by 10 matrix) = $\begin{bmatrix} 3\mu_1 I_{8\times(10,11)} & & & & & & & & & \\ & 3\mu_1 I_{8\times(9,10)} & & & & & & & & \\ & & \ddots & & & & & & & \\ & & & \ddots & & & & & & \\ & & & & 3\mu_1 I_{8\times(5,6)} & & & & & \\ & & & & & 3\mu_1 I_{8\times(5,5)} & & & & \\ & & & & & & \ddots & & & \\ & & & & & & & 3\mu_1 I_{8\times(5,5)} & 0 \end{bmatrix}$

$A_{7,6}$ (7 by 6 matrix) = $\begin{bmatrix} 7\mu_1 I_{8\times(6,7)} & & & & & \\ & 7\mu_1 I_{8\times(5,6)} & & & & \\ & & 7\mu_1 I_{8\times(5,5)} & & & \\ & & & \ddots & & \\ & & & & \ddots & \\ & & & & & 7\mu_1 I_{8\times(5,5)} \quad 0 \end{bmatrix}$

$A_{8,7}$ (6 by 5 matrix) = $\begin{bmatrix} 8\mu_1 I_{8\times(5,6)} & & & & \\ & 8\mu_1 I_{8\times(5,5)} & & & \\ & & \ddots & & \\ & & & \ddots & \\ & & & & 8\mu_1 I_{8\times(5,5)} \quad 0 \end{bmatrix}$

A_2 (5 by 5 matrix) = $\begin{bmatrix} 8\mu_1 I_{5\times8} & & & & \\ & \ddots & & & \\ & & \ddots & & \\ & & & \ddots & \\ & & & & 8\mu_1 I_{5\times8} \end{bmatrix}$

$\alpha = [\ 1\ 0\ ]$ $\qquad T_1^{\,o} = \begin{bmatrix} \lambda_1^{(1)} q_1 \\ \lambda_1^{(2)} \end{bmatrix}$ $\qquad T_1 = \begin{bmatrix} -\lambda_1^{(1)} & \lambda_1^{(1)} p_1 \\ 0 & -\lambda_1^{(2)} \end{bmatrix}$

$T_2^{\,o} = \begin{bmatrix} \lambda_2^{(1)} q_2 \\ \lambda_2^{(2)} \end{bmatrix}$ $\quad T_2 = \begin{bmatrix} -\lambda_2^{(1)} & \lambda_2^{(1)} p_2 \\ 0 & -\lambda_2^{(2)} \end{bmatrix}$ $\qquad T_3^{\,o} = \begin{bmatrix} \lambda_3^{(1)} q_3 \\ \lambda_3^{(2)} \end{bmatrix}$ $\quad T_3 = \begin{bmatrix} -\lambda_3^{(1)} & \lambda_3^{(1)} p_3 \\ 0 & -\lambda_3^{(2)} \end{bmatrix}$

Chapter 5

Performance Analysis

5.1 Steady-State Probability

Once the traffic is modeled as a CTMC preemptive priority queue, one can then use matrix-geometric method [22] to analyze the performance metrics. To obtain the performance metrics of each traffic class, we first need to calculate the invariant vector from the generator matrix Q. Let π denote the invariant vector $\pi = [\pi_0\ \pi_1\ \pi_2\ ...\ \pi_{m1}]$ of the generator matrix Q such that $\pi Q = 0$ and $\pi \mathbf{e} = 1$, where **e** is a column vector of 1's. We use the Gauss-Seidel method to calculate the invariant vector π.

To illustrate the use of the Gauss-Seidel method to calculate the invariant vector π, we use a system with n = 2 and $m_1 = m_2 = m_3 = 3$ for illustrative purpose. Let the generator matrix Q be

$$Q = \begin{bmatrix} A_{0,0} & A_0 & & \\ A_{1,0} & A_{1,1} & A_0 & \\ & A_2 & A_1 & A_0 \\ & & A_2 & A_1+A_0 \end{bmatrix}$$

This gives the following set of 5 linear equations:

(1) $0 = \pi_0 A_{0,0} + \pi_1 A_{1,0}$
(2) $0 = \pi_0 A_0 + \pi_1 A_{1,1} + \pi_2 A_2$
(3) $0 = \pi_1 A_0 + \pi_2 A_1 + \pi_3 A_2$
(4) $0 = \pi_2 A_0 + \pi 3(A_1 + A_0)$
(5) $\pi_0 + \pi_1 + \pi_2 + \pi_3 = 1$

Manipulating the above set of equations gives

(1) $\pi_0 = (-\pi_1 A_{1,0}) A_{0,0}^{-1}$
(2) $\pi_1 = -(\pi_0 A_0 + \pi_2 A_2) A_{1,1}^{-1}$
(3) $\pi_2 = -(\pi_1 A_0 + \pi_3 A_2) A_1^{-1}$
(4) $\pi_3 = (-\pi_2 A_0)(A_1 + A_0)^{-1}$
(5) $\pi_0 + \pi_1 + \pi_2 + \pi_3 = 1$

Then, to calculate the invariant vector π, we iterate through the set of equations as follows:

(1) $\pi_0(n+1) := (-\pi_1(n) A_{1,0}) A_{0,0}^{-1}$
(2) $\pi_1(n+1) := -(\pi_0(n+1) A_0 + \pi_2(n) A_2) A_{1,1}^{-1}$
(3) $\pi_2(n+1) := -(\pi_1(n+1) A_0 + \pi_3(n) A_2) A_1^{-1}$
(4) $\pi_3(n+1) := (-\pi_2(n+1) A_0)(A_1 + A_0)^{-1}$

with the initial vector $n = [\ n_0\ n_1\ n_2\ n_3\ ] = [\ 0\ 1\ 0\ 0\ ]$ and the condition $\| \pi_i(n+1) - \pi_i(n) \| < \varepsilon$, where $\varepsilon = 10^{-9}$.

5.2 Queue Length

5.2.1 Joint Stationary Distribution

Once the stationary vector π is obtained, we then have the joint stationary distribution. The joint stationary distribution in our traffic model is the probability that there are i class 1, j class 2, and k class 3 packets in the system. In this thesis, we use P(i,j,k) to represent $\pi_{i,j,k}$ in the performance analysis in the subsequent sections.

In summary, the joint distribution stationary distribution is

$P(i,j,k) = \pi_{i,j,k}$ (5.1)

= probability that there are i class 1, j class 2, and k class 3 packets in the system

5.2.2 Marginal Distribution

Class 1

The marginal distribution of class 1 traffic is the probability that there are i class 1 packets in the system. It is given by equation 5.2.

$$P(i,.,.) = \sum_{j=0}^{\max(n-i,0)+m_2} \sum_{k=0}^{\max(n-i-j,0)+m_3} \pi_{i,j,k} \quad (5.2)$$

where

n = number of servers
m_2 = number of waiting spaces in the buffer for class 2 traffic
m_3 = number of waiting spaces in the buffer for class 3 traffic
i = number of class 1 packets in system
j = number of class 2 packets in system
k = number of class 3 packets in system

Equation 5.2 gives the probability that there are i class 1 packets in the system and is calculated by summing up all $\pi_{i,j,k}$ where i = n and where n is any non-negative integer. The upper bound of the first summation sign is $\max(n-i,0)+m_2$ because the queue in our traffic model has a preemptive priority service discipline. Recall that in our traffic model, there are n servers and m_2 buffer slots in the system. So the maximum possible number of class 2 packets in the system at any single instance is $n+m_2$. However, because the system has a preemptive priority service discipline, the maximum possible number of class 2 packets in the system at any single instance becomes $n - i + m_2$. (The minus i in $n-i+m_2$ is because class 1 traffic has preemptive priority over class 2 traffic, so the maximum number of class 2 packets that can be in service given that there are i class 1 packets already receiving service is n-i). Note that it is possible for i to be greater than n (i.e. $i > n$) when there are class 1 packets in the buffer. To ensure that $n-i \geq 0$, we use max(n-i,0) to guarantee this condition. Thus, the upper bound for the first summation term is $\max(n-i,0) + m_2$.

The upper bound $\max(n-i-j,0) + m_3$ for the second summation term follows by the same logic. Because of the preemptive priority service discipline, the maximum number of class 3 packets in the n-server channel is

n-i-j. Adding together the maximum number of class 3 packets in the buffer, m_3, gives the upper bound $\max(n-i-j,0) + m_3$.

Class 2

Similarly, the marginal distribution of class 2 traffic is the probability that there are i class 1 packets and j class 2 packets in the system and is given by equation 5.3:

$$P(i,j,.) = \sum_{k=0}^{\max(n-i-j,0)+m_3} \pi_{i,j,k} \qquad (5.3)$$

5.2.3 First Moment (Average Queue Length)

Once the marginal distribution is known, the average queue length can be calculated easily. The average queue length for class i traffic is the average number of class i packets in the system (i.e. in the servers and buffer). It is well known that if X is a discrete random variable, then E[X], the expected value, or mean, of X, is $E[X] = \sum_k x_k P_X(x_k)$, where $P_X(x_k)$ denotes the probability that the random variable X has a value of x_k [19, p. 127]. Since calculating the average queue length means finding the first moment of a random variable, the average queue length of class 1 traffic, which we denote as $(E[X])_1$, can be calculated by equation (5.4).

$$(E[X])_1 = \sum_{i=1}^{n+m_1} iP(i,.,.) \qquad (5.4)$$

$$= P(1,.,.) + 2P(2,.,.) + 3P(3,.,.) + \ldots + (n+m_1-1)P(n+m_1-1,.,.) + (n+m_1)P(n+m_1,.,.)$$

For class 2, the average queue length $(E[X])_2$ is as follows:

$$(E[X])_2 = \sum_{i=0}^{n+m_1} \sum_{j=1}^{\max(n-i,0)+m_2} jP(i,j,.) \qquad (5.5)$$

$$= P(0,1,.) + 2P(0,2,.) + 3P(0,3,.) + \ldots + (n+m_2-1)P(0,n+m_2-1,.) + (n+m_2)P(0,n+m_2,.) +$$
$$P(1,1,.) + 2P(1,2,.) + \ldots + (n+m_2-2)P(1,n+m_2-2,.) + (n+m_2-1)P(1,n+m_2-1,.) +$$
$$P(2,1,.) + 2P(2,2,.) + \ldots + (n+m_2-3)P(2,n+m_2-3,.) + (n+m_2-2)P(2,n+m_2-2,.) + \ldots +$$
$$P(n+m_1-1,1,.) + 2P(n+m_1-1,2,.) + \ldots + (m_2-1)P(n+m_1-1,m_2-1,.) + m_2P(n+m_1-1,m_2,.) +$$
$$P(n+m_1,1,.) + 2P(n+m_1,2,.) + \ldots + (m_2-1)P(n+m_1,m_2-1,.) + m_2P(n+m_1,m_2,.)$$

The average queue length for class 3 traffic, $(E[X])_3$, is:

$$(\mathrm{E}[X])_3 = \sum_{i=0}^{n+m_1} \sum_{j=0}^{\max(n-i,0)+m_2} \sum_{k=1}^{\max(n-i-j,0)+m_3} kP(i,j,k) \qquad (5.6)$$

5.2.4 Second Moment (Variance of Queue Length)

Once the first moment (average queue length) is known, its variance can easily be calculated. We know that the variance of a random variable X, VAR[X], is defined as $\mathrm{VAR}[X] = \mathrm{E}[(X - \mathrm{E}[X])^2] = \mathrm{E}[X^2] - \mathrm{E}^2[X]$ [19, p.133]. In the context of our traffic model, $(\mathrm{VAR}[X])_i$, where i = {1, 2, 3}, represents the variance of queue length of class i traffic, and since the term $\mathrm{E}[X^2]$ is calculated by:

$$\mathrm{E}[X^2] = \sum_k x_k^{\,2} P_X(x_k) \qquad (5.7)$$

The equation for the variance of the queue length for class 1 traffic, $(\mathrm{VAR}[X])_1$, is:

$$(\mathrm{VAR}[X])_1 = (\mathrm{E}[X^2])_1 - (\mathrm{E}^2[X])_1$$
$$= \sum_{i=1}^{n+m_1} i^2 P(i,.,.) \; - \; [\sum_{i=1}^{n+m_1} iP(i,.,.)]^2 \qquad (5.8)$$

Similarly, the variance of the queue length for class 2, $(\mathrm{VAR}[X])_2$, is:

$$(\mathrm{VAR}[X])_2 = (\mathrm{E}[X^2])_2 - (\mathrm{E}^2[X])_2$$
$$= \sum_{i=0}^{n+m_1} \sum_{j=1}^{\max(n-i,0)+m_2} j^2 P(i,j,.) \; - \; [\sum_{i=0}^{n+m_1} \sum_{j=1}^{\max(n-i,0)+m_2} jP(i,j,.)]^2 \qquad (5.9)$$

And the variance of the queue length for class 3, $(\mathrm{VAR}[X])_3$, is:

$$(\mathrm{VAR}[X])_3 = (\mathrm{E}[X^2])_3 - (\mathrm{E}^2[X])_3$$
$$= \sum_{i=0}^{n+m_1} \sum_{j=0}^{\max(n-i,0)+m_2} \sum_{k=1}^{\max(n-i-j,0)+m_3} kP(i,j,k) \; - \; [\sum_{i=0}^{n+m_1} \sum_{j=0}^{\max(n-i,0)+m_2} \sum_{k=1}^{\max(n-i-j,0)+m_3} kP(i,j,k)]^2$$

(5.10)

5.3 Blocking Probability

The blocking probability is the probability that, upon the arrival of a class i packet, the queue is already full (i.e. all the n servers are occupied and no lower priority class packet can be preempted, and the m_i buffer slots are all occupied) so the newly arrived packet must be dropped. For traffic class 1, the blocking probability Pb_1 is the probability that upon the arrival of a class 1 packet, there are n class 1 packets already

in service and all the m_1 buffer slots are full. This is equivalent to the probability that there are $n+m_1+1$ class 1 packets in the system. Thus the equation for Pb_1 is:

$$Pb_1 = \pi_{n+m_1+1} \qquad (5.11)$$

For traffic class 2, the blocking probability Pb_2 is the probability that, upon the arrival of a class 2 packet, the n servers are all occupied by either class 1 and class 2 packets, and all the m_2 buffer slots are filled. This is equivalent to the probability that there are already i class 1 packets and n-i class 2 packets receiving service and m_2+1 class 2 packets in the buffer when a new class 2 packet arrives. Thus the blocking probability Pb_2 is

$$Pb_2 = \sum_{i=0}^{n+m_1} P(i, \max(n-i,0) + m_2 + 1,.) \qquad (5.12)$$

Similarly, for traffic class 3, the blocking probability Pb_3 is the probability that, upon the arrival of a class 3 packet, the n servers are all occupied, and all the m_3 buffer slots are filled. This is equivalent to the probability that there are already i class 1 packets, j class 2 packets, and n-i-j class 3 packets receiving service, and m_3+1 class 3 packets in the buffer when a new class 3 packet arrives. Thus the blocking probability Pb_3 is

$$Pb_3 = \sum_{i=0}^{n+m_1} \sum_{j=0}^{\max(n-i,0)+m_2} P(i, j, \max(n-i-j,0) + m_3 + 1) \qquad (5.13)$$

5.4 Delay

5.4.1 First Moment (Mean Delay)

The mean delay is the average amount of time a packet needs to wait in a queue before being finished receiving service and leaving the system. It is easy to calculate the mean delay and delay variance once the average queue length $(E[X])_i$ and blocking probability Pb_1 are known. According to Little's Law, the average number of customers (N) in a stable system equals the average arrival rate of customers ($\overline{\lambda}$) multiplied by the average waiting time of a customer ($\overline{W}$). In equation form, this is

$$N = \overline{\lambda}\overline{W} \qquad (5.14)$$

Rearranging the equation, it gives

$$\overline{W} = \frac{N}{\overline{\lambda}} \qquad (5.15)$$

In our traffic model, the average number of packets N is almost equivalent to the average queue length E[X] described in the previous section. However, in our traffic model, a newly arrived packet will find the queue to be full and be dropped from the system with blocking probability Pb. Thus we need to compensate for the effect of blocking probability on the average queue length. Therefore, in our traffic model, N is equivalent to E[X] divided by (1 – Pb); this division by (1 – Pb) compensates for the packet loss due to the blocking probability. The bottom term in equation (5.15), $\overline{\lambda}$, is the average customer arrival rate. Since the arrival rate in our traffic model is represented by phase-type distribution with representation (α, T) with dimension d, the average arrival rate is then $\overline{\lambda} = (-\alpha T^{-1}\mathbf{e})^{-1}$, where **e** is a column vector of 1. Thus, the mean delay of class i traffic is

$$\overline{W_i} = \frac{N_i}{\overline{\lambda_i}} = \frac{(E[X])_i}{-\alpha T_i^{-1}\mathbf{e}} \qquad (5.16)$$

5.4.2 Second Moment (Jitter)

The delay variance, also called jitter, follows from the second moment of Little's Law [14]. Recall in the previous section that $\overline{W_i} = \frac{N_i}{\overline{\lambda_i}}$. Thus the delay variance, VAR[W_i] , can be calculated by VAR[W_i] = VAR[N_i] / $\overline{\lambda_i^2}$. However, this VAR[W_i] is actually the variance around zero, not the variance around the mean delay. To obtain the variance around mean delay, recall that $(VAR[X])_i = (E[X^2])_i - (E^2[X])_I$ [19]. Thus, if we denote $VAR[\overline{W_i}]$ as the variance around mean delay and VAR[W_i] as the variance around zero, then

$$VAR[\overline{W_i}] = \mathrm{VAR}[W_i] - \overline{W_i^2} = \frac{VAR[N_i]}{\overline{\lambda_i}} - \overline{W_i^2} = \frac{(VAR[X])_i}{\overline{\lambda_i}} - \overline{W_i^2} \qquad (5.17)\ \text{[from [14]]}$$

where

$(VAR[X])_i$ is described in Section 5.3

$\overline{\lambda_i} = (-\alpha T^{-1}\mathbf{e})^{-1}$

$\overline{W_i}$ is described in Section 5.5.1

i = {1, 2, 3}

Chapter 6

Results and Discussions

6.1 Implementation of a PH/M/7/(9+9+9) Example System

We implemented the analytical model described in Chapter 5 with a PH/M/7/(9+9+9) preemptive priority queue (i.e. number of servers n = 7 and buffer space $m_1 = m_2 = m_3 = 9$) with 2-phase Coxian arrival process. The input parameters of the traffic model for the example system are described in Table 6.1.

Input Parameters	Values	Meaning
packet_size	2500 bytes per packet	The size of packet, assumed to be 2500 bytes per packet.
δ_{video}, δ_{audio}	0.8	Fraction of households served by an ISP that are actively using real-time video and audio applications
δ_{tel}	0.8	Fraction of households served by an ISP that are actively using telephony applications
$\delta_{control}$	0.5	Fraction of households served by an ISP that are actively using control traffic
δ_2	0.8	Fraction of households served by an ISP that are actively using class 2 (interactive) applications
δ_3	0.5	Fraction of households served by an ISP that are actively using class 3 (unclassified) applications
λ_{video}	20 Mbits / sec	Arrival rate of real-time video traffic
λ_{audio}	0.256 Mbits / sec	Arrival rate of real-time audio traffic
λ_{tel}	0.128 Mbits / sec	Arrival rate of telephony traffic
$\lambda_{control}$	0.1 Mbits / sec	Arrival rate of control traffic
$\lambda_1^{(1)}$	831.95 packets/sec	Arrival rate of class 1 traffic at phase 1 of arrival process
$\lambda_1^{(2)}$	831.95 packets/sec	Arrival rate of class 1 traffic at phase 2 of arrival process
$\lambda_2^{(1)}$	24 packets / sec	Arrival rate of class 2 traffic at phase 1 of arrival process
$\lambda_2^{(2)}$	24 packets / sec	Arrival rate of class 2 traffic at phase 2 of arrival process
$\lambda_3^{(1)}$	12.5 packets / sec	Arrival rate of class 3 traffic at phase 1 of arrival process
$\lambda_3^{(2)}$	12.5 packets / sec	Arrival rate of class 3 traffic at phase 2 of arrival process
q_1	0.9	Probability of a class 1 packet entering absorption phase in arrival process without passing through phase 2
q_2	0.9	Probability of a class 2 packet entering absorption phase in arrival process without passing through phase 2
q_3	0.9	Probability of a class 3 packet entering absorption phase in arrival process without passing through phase 2

Table 6.1: Input parameters of traffic model

Class 1 Arrival Rate

In our implementation, we assume that all packets are of size 2500 bytes, or 0.02 Mbits (2500 bytes x 8 bits/byte = 20000 kbits = 0.02 Mbits, per packet). Furthermore, we assume that an average of 80% of all the households served by a particular ISP actively use real-time video and audio applications during peak hours, thus $\delta_{video\_audio} = 0.8$. For real-time multimedia traffic, we assume that an average household uses the bandwidth equivalent of 1 HDTV, 2 streaming videos and 2 MP3 applications. The calculation for λ_{video}, the arrival rate of real-time video traffic, is as follows:

λ_{video}
= 1 HDTV application x basic bandwidth for 1 HDTV application (18 Mbits per application) + 2 streaming video applications x bandwidth requirement of 1 streaming video application (1 Mbits per application)
= 1 x r_{HDTV} + 2 x $r_{streaming\_video}$
= 1 x 18 Mbits/sec + 2 x 1 Mbits/sec
= 20 Mbits/sec

Similarly, the arrival rate of real-time audio traffic, λ_{audio}, is:

λ_{audio} = basic bandwidth for 1 MP3 application x 2 MP3 applications
= 0.128 Mbits/sec x 2
= 0.256 Mbits/sec

The arrival rate of telephony traffic, λ_{tel}, is:

λ_{tel} = basic bandwidth for telephony application x 2 telephony applications
= 0.064 Mbits/sec x 2
= 0.128 Mbits/sec

The arrival rate of control signal traffic, $\lambda_{control}$, is:

$\lambda_{control}$ = basic bandwidth for control traffic x 2 control signal applications
= 0.05 Mbits/sec * 2
= 0.1 Mbits/sec

Recall that the arrival rate of class 1 traffic is calculated by equation (4.2). Thus the arrival rate of class 1 traffic, λ_1, is:

$\lambda_1 = \delta_{video}\lambda_{video} + \delta_{audio}\lambda_{audio} + \delta_{tel}\lambda_{tel} + \delta_{control}\lambda_{control}$
= 0.8(20 Mbits/sec + 0.256 Mbits/sec) + 0.3(0.128 Mbits/sec) + 0.5(0.1 Mbits/sec)
= 16.639 Mbits/sec * (1 byte / 8 bits) * (1 packet / 2500 bytes)
= 831.95 packets / sec

Class 2 Arrival Rate

For class 2 traffic, the arrival rate is:

$\lambda_2 = \delta_2 r_2 n_2$
= (0.8)(0.3 Mbits/sec)(2)
= 0.48 Mbits/sec * (1 byte / 8 bits) * (1 packet / 2500 bytes)
= 24 packets/sec

Class 3 Arrival Rate

For class 3 traffic, the arrival rate is:

$\lambda_3 = \delta_3 r_3 n_3$
= (0.5)(0.05 Mbits/sec)(10)
= 0.25 Mbits/sec * (1 byte / 8 bits) * (1 packet / 2500 bytes)
= 12.5 packets/sec

The parameters discussed in Table 6.1 are only estimated parameters that we provide for the traffic model as an example. When actually using the traffic model for bandwidth estimation, the values for the input parameters in Table 6.1 should be estimated according to survey of average household usage. For example, one needs to collect data on how many class i applications are used for each of the three traffic classes in an average household. Also, the δ_i should be estimated according to data collected on the proportion of households that actively use class i applications at a particular time. We do not have such data available, because at present, to the best of our knowledge, there is no actual home area network existing yet and so we are not able to obtain such data. Therefore we could only estimate the values of the input parameters. However, in the future, when test homes with home area networks equipped become available, one should first collect the necessary data to estimate the input parameters for the traffic model.

2-phase Coxian Arrival Process

Once again, because there is no actual future HAN existing yet, a 2-phase Coxian arrival process is used in the example system as a simple example for the purpose of illustrating the possible use of the traffic model to help estimate the optimal bandwidth allocation for network planning purpose. However, when HAN becomes available for testing, one should collect data on each of the three types of traffic, determine the general distribution of the packet arrival rate for each traffic type, and then map the general distributions into n-phase EC distributions to approximate the arrival process for each traffic class, as described in Section 4.4.

6.2 Bandwidth Optimization Procedure

Since the goal of the traffic model is to determine the bandwidth allocation scheme that utilizes the minimum total bandwidth while satisfying the QoS constraints (mean delay and blocking probability) of each traffic class i, the following steps are used to run the traffic model to achieve the goal of determining the best bandwidth allocation scheme.

Step 1: Run the traffic model with various $n\mu_1$ (i.e. varying the amount of bandwidth allocated to class 1 traffic) and record the mean delay (W_1) and blocking probability (Pb_1) of class 1 vs. a particular $n\mu_1$ for each run.

Step 2: Plot the results from step 1 in two plots. One as the plot of class 1 bandwidth ($n\mu_1$) vs. class 1 mean delay (W_1). (n is the number of servers.) The other as the plot of class 1 bandwidth vs. class 1 blocking probability (Pb_1).

Step 3: In the plot of class 1 bandwidth ($n\mu_1$) vs. class 1 mean delay (W_1), use polynomial interpolation or other regression analysis methods to interpolate a function f_1 where $n\mu_1 = f_1(W_1)$ for the $n\mu_1$ vs. W_1 data points observed from running the traffic model. Similarly, in the plot of class 1 bandwidth vs. class 1 blocking probability (Pb_1), use polynomial interpolation or other regression analysis methods to interpolate a function g_1 where $n\mu_1 = g_1(Pb_1)$ for the μ_1 vs. Pb_1 data points observed from running the traffic model.

Step 4a: Determine the minimal value of the function f_1 (i.e. the minimum μ_1) which meets the constraints $W_1 < 10$ms (see Table 4.2 for the QoS constraints) and stability constraint $n\mu_1 > 16.7$ Mbps.
Step 4b: Determine the minimal value of the function g_1 (i.e. the minimum μ_1) which meets the constraint $Pb_1 < 10\text{^}-5$ (see Table 4.2 for the QoS constraints) and stability constraint $n\mu_1 > 16.7$ Mbps.

Step 5: Compare the two minimum $n\mu_1$ found in steps 4a and 4b. Select the one with the higher value of μ_1 and discard the one with the lower value.

Step 6: Run the traffic model with various $n\mu_2$ (i.e. varying the amount of bandwidth allocated to class 2 traffic) while keeping the value of $n\mu_1$ fixed at the value selected in step 5. Record the mean delay (W_2) and blocking probability (Pb_2) of class 2 vs. a particular $n\mu_2$ for each run.

Step 7: Plot the results from step 6 into two plots. One as the plot of class 2 bandwidth ($n\mu_2$) vs. class 2 mean delay (W_2). The other as the plot of class 2 bandwidth vs. class 2 blocking probability (Pb_2).

Step 8: In the plot of class 2 bandwidth ($n\mu_2$) vs. class 2 mean delay (W_2), use polynomial interpolation or other regression analysis methods to interpolate a function f_2 where $n\mu_2 = f_2(W_2)$ for the $n\mu_2$ vs. W_2 data points observed from running the traffic model. Similarly, in the plot of class 2 bandwidth vs. class 2 blocking probability (Pb_2), use polynomial interpolation or other regression analysis methods to interpolate a function g_2 where $n\mu_2 = g_2(Pb_2)$ for the $n\mu_2$ vs. Pb_2 data points observed from running the traffic model.

Step 9a: Determine the minimal value of the function f_2 (i.e. the minimum $n\mu_2$) which meets the constraint $W_2 < 1.0$s (see Table 4.2 for the QoS constraints) and stability constraint $n\mu_2 > 0.48$ Mbps.
Step 9b: Determine the minimal value of the function g_2 (i.e. the minimum μ_1) which meets the constraint $Pb_2 < 0.05$ (see Table 4.2 for the QoS constraints) and stability constraint $n\mu_2 > 0.48$ Mbps.

Step 10: Compare the two minimum $n\mu_2$ found in steps 9a and 9b. Select the one with the higher value of μ_2 and discard the one with the lower value.

Step 11: Run the traffic model with various $n\mu_3$ (i.e. varying the amount of bandwidth allocated to class 3 traffic) while keeping the value of $n\mu_1$ and $n\mu_2$ fixed at the values selected in step 5 and step 10. Record the mean delay (W_3) and blocking probability (Pb_3) of class 3 vs. a particular $n\mu_3$ for each run.

Step 12: Plot the results from step 11 into two plots. One as the plot of class 3 bandwidth ($n\mu_3$) vs. class 3 mean delay (W_3). The other as the plot of class 3 bandwidth vs. class 3 blocking probability (Pb_3).

Step 13: In the plot of class 3 bandwidth ($n\mu_3$) vs. class 3 mean delay (W_3), use polynomial interpolation or other regression analysis methods to interpolate a function f_3 where $n\mu_3 = f_3(W_3)$ for the $n\mu_3$ vs. W_3 data points observed from running the traffic model. Similarly, in the plot of class 3 bandwidth vs. class 3 blocking probability (Pb_3), use polynomial interpolation or other regression analysis methods to interpolate a function g_3 where $\mu_3 = g_3(Pb_3)$ for the $n\mu_3$ vs. Pb_3 data points observed from running the traffic model.

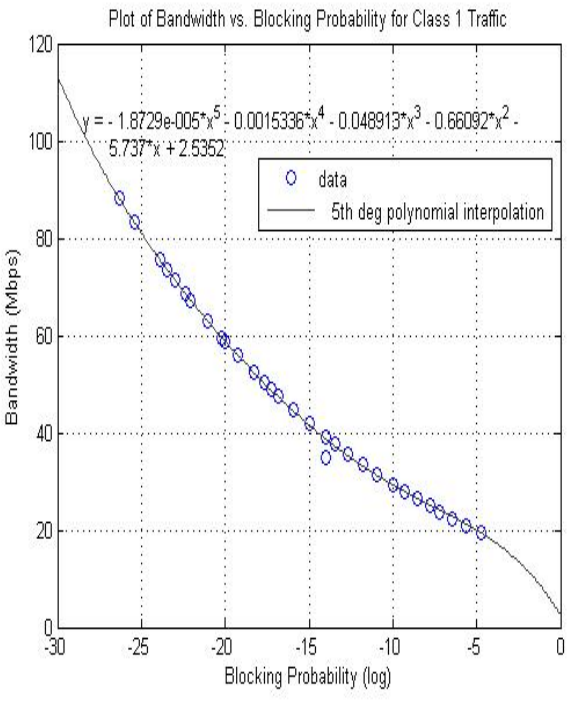
Step 14a: Determine the minimal value of the function f_3 (i.e. the minimum $n\mu_3$) with stability constraint $n\mu_3 > 0.25$ Mbps.
Step 14b: Determine the minimal value of the function g_3 (i.e. the minimum $n\mu_3$) while meeting the constraint $Pb_3 < 0.05$ (see Table 4.2 for the QoS constraints) and stability constraint $n\mu_3 > 0.25$ Mbps.

Step 15: Compare the two minimum $n\mu_3$ found in steps 14a and 14b. Select the one with the higher value and discard the one with the lower value.

After running the above procedure, we will have obtained the bandwidth allocation scheme that minimizes total bandwidth usage while satisfying the QoS constraints for all 3 traffic classes listed in Table 4.2.

Class 1 Traffic

The plot of bandwidth vs. mean delay for class 1 traffic in Figure 6.1(a) and the plot of bandwidth vs. blocking probability for class 1 traffic in Figure 6.1(b) show the result of applying the above procedure for the parameters shown in Table 6.1.



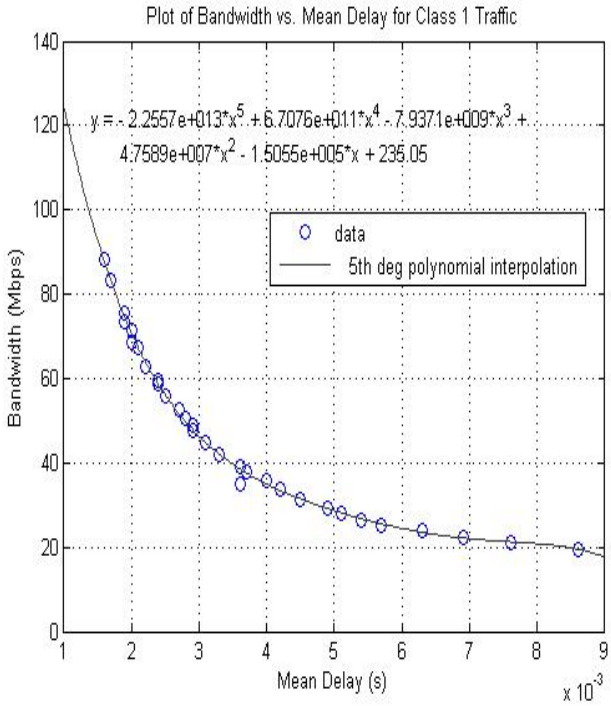


Figure 6.1(a): Plot of Bandwidth Required vs. Mean Delay for Class 1 Traffic. Figure 6.1(b): Plot of Bandwidth Required vs. Blocking Probability for Class 1 Traffic

The above two plots show the amount of bandwidth that should be allocated to class 1 traffic given a mean delay and blocking probability requirement for class 1 traffic. From the above two sets of data, one can use data fitting to generate two polynomial equations that approximate the behavior of bandwidth vs. mean delay and bandwidth vs. blocking probability, respectively. For example, using the implementation of the PH/M/7/(9+9+9) example system described in Section 6.1, the polynomial equation for describing the behavior of bandwidth required vs. mean delay for class 1 traffic is:

$$n\mu_1 = f_1(W_1) = (-2.2557e{+}013)\, W_1^5 + (6.7076e{+}011)\, W_1^4 + (-7.9371e{+}009)\, W_1^3 + (4.7589e{+}007)\, W_1^2 + (-1.5055e{+}005) W_1 + 235.05$$

where W_1 represents the mean delay of class 1 traffic and $f_1(W_1)$, the value of the function f_1 at W_1, represents the bandwidth allocated for class 1 traffic given that the mean delay for class 1 traffic is W_1.

To determine the minimum bandwidth needed for class 1 traffic to satisfy the mean delay requirement of 10ms (from Table 3.2), a local search optimization method called golden section search method [30] was used on the polynomial function $f_1(W_1)$ on the range $0 \leq W_1 \leq 0.01$ to determine the local minimum within the range $0 \leq W_1 \leq 0.01$s. The reason why it is decided to use a local search method like golden section search to find the minimum is because usually, from the plot, we have a fairly good initial guess of where the minimum bandwidth will be. If getting trapped in a local minimum is an issue, one can use a global search method like simulated annealing or genetic algorithm to search for the global minimum. In this thesis, the resultant minimum bandwidth obtained using the golden section search method was 16.8397 Mbps for each household.

Similarly, to determine the minimum class 1 bandwidth to satisfy the QoS requirement of < 10e-5 blocking probability for class 1 traffic, we first used the polynomial interpolation to approximate the relationship between bandwidth required versus blocking probability for class 1 traffic. The resulting polynomial equation for the example system in Section 6.1 was determined to be

$$n\mu_1 = g_1(Pb_1) = -0.18756\, Pb_1^5 - 0.1103\, Pb_1^4 + 0.30508\, Pb_1^3 + 3.9773\, Pb_1^2 - 19.511\, Pb_1^2 + 43.471$$

where $g_1(Pb_1)$ represents the amount of bandwidth that needs to be allocated to class 1 traffic to produce a blocking probability of Pb_1 for class 1 traffic. Again, using golden section search on the polynomial equation $g_1(Pb_1)$, the minimum class 1 bandwidth required to satisfy the QoS requirement of blocking probability of less than 10e-5 for class 1 traffic was determined to be 21.4382 Mbps.

From the above discussion, it is obvious that to meet the QoS requirements of *both* mean delay (< 10ms) and blocking probability (< 10e-5) for class 1 traffic, the ISP needs to allocate a minimum of 21.44 Mbps of bandwidth for class 1 traffic for each household. Note that 21.44 Mbps also satisfies the stability constraint for class 1 traffic, $n\mu_1 > 16.7$ Mbps. Also, from the data points sampled, the highest variance was found to be 1.9930×10^{-5} s^2 and the lowest variance was 2.1083×10^{-6} s^2.

Recall that in Table 4.2, class 1 traffic also has a jitter requirement of 5ms. This has been satisfied, since the highest jitter measured from running the traffic model was $\sqrt{1.9930 \times 10^{-5}\, s^2} = 4.46$ ms, which is less than the 5ms jitter requirement for class 1 traffic.

Traffic class 2 bandwidth required

From the above discussion, it has been determined that an ISP needs to allocate a minimum of 21.44 Mbps of bandwidth for class 1 traffic for each household. The amount of traffic class 2 bandwidth required is determined by fixing the class 1 bandwidth at 21.44 Mbps. The following plots illustrate the relationship between class 2 bandwidth required vs. mean delay and blocking probability for class 2 traffic given that class 1 traffic is allocated a bandwidth of 21.44 Mbps.

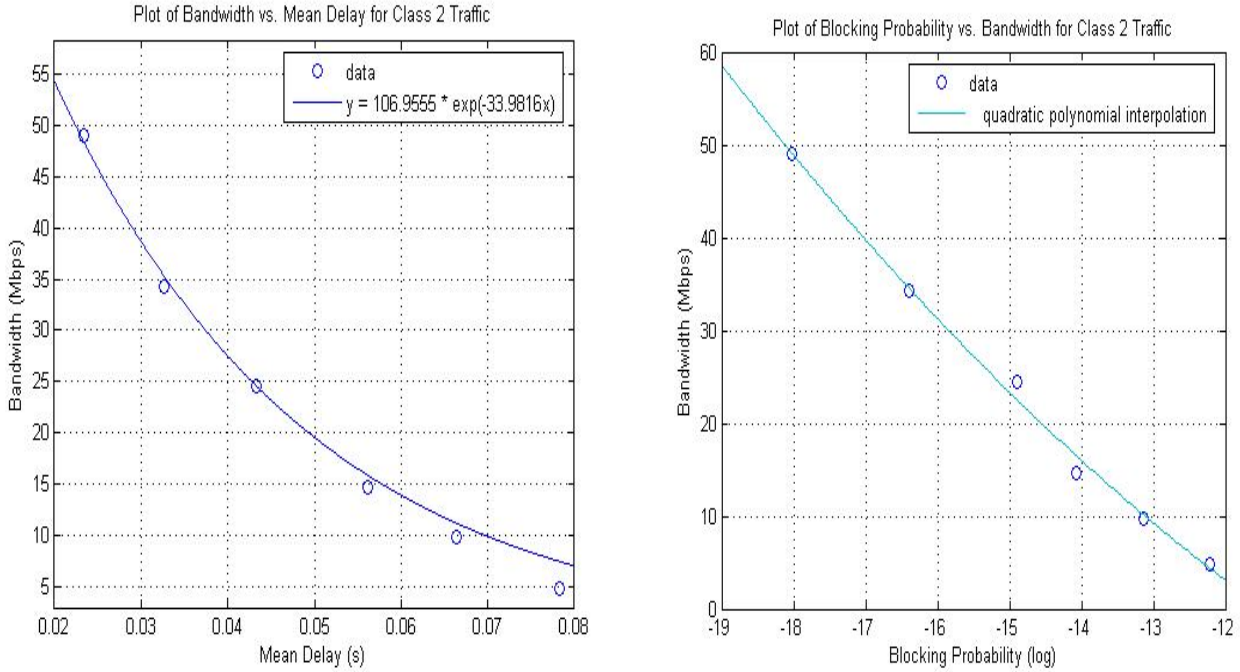


Figure 6.2(a): Plot of Bandwidth Required vs. Mean Delay for Class 2 Traffic, and (b): Plot of Bandwidth Required vs. Blocking Probability for Class 2 Traffic.

Figure 6.2(a) shows the amount of bandwidth that should be allocated to class 2 traffic given various mean delay requirements for class 2 traffic at $n\mu_1$ = 21.44 Mbps, and Figure 6.2(b) shows the amount of bandwidth that should be allocated to class 2 traffic given various blocking probability requirements for class 2 traffic at $n\mu_1$ = 21.44 Mbps . From the above two plots, one can use data fitting to generate two equations that approximate the behavior of bandwidth vs. mean delay and bandwidth vs. blocking probability, respectively. For example, using the implementation of the PH/M/7/(9+9+9) example system described in Section 6.1 and following the logic outlined for class 1 traffic, the equation for describing the behavior of bandwidth required vs. mean delay for class 2 traffic is:

$$n\mu_2 = f_2(W_2) = 106.9555e^{-33.9816W_2} \qquad (6.6)$$

where W_2 represents the mean delay of class 2 traffic and $f_2(W_2)$, the value of the function f_2 at W_2, represents the bandwidth allocated for class 2 traffic given that the mean delay for class 2 traffic is W_2.

If we let $y = f_2(W_2)$ and $x = W_2$, equation (6.6) can be written as:

$$y = 106.9555e^{-33.9816x} \qquad (6.7)$$

The equation $f_2(W_2)$ is determined by fitting the data in Figure 6.2(a) into an exponential equation of the form:

$$y = pe^{-qx} \qquad (6.8)$$

where p and q are real-number coefficients. The above exponential form is, again, used because the data points observed in Figure 6.2(a) appear to exhibit exponential decay. The coefficients p and q, in turn, are obtained by observing that rearranging all the terms in Equation (6.8) onto the left-hand side gives

$$y = pe^{-qx}$$
$$\ln(y) - \ln(p)e^{-qx} = 0 \qquad (6.9)$$

Thus to find p and q, we try to find the values of p and q that minimize:

$$\sum_i (\tilde{y}_i - (\tilde{p}_i - qx_i))^2 \qquad (6.10)$$

where $\tilde{y}_i = \ln(y_i)$ and $\tilde{p}_i = \ln(p_i)$. The values of p and q obtained for this particular example system are p = 106.9555 and q = 33.9816.

To determine the minimum bandwidth needed for class 2 traffic to satisfy the mean delay requirement of 1.0s (from Table 3.2), the golden section search method was used on the function $f_2(W_2)$ on the range $0 \le W_2 \le$ 0.0875s to determine the local minimum within the range $0 \le W_2 \le 0.0875$s. The resulting minimum bandwidth obtained was 0.4809 Mbps for each household. Also, from the data points sampled, the highest variance was found to be 0.0037 s^2 and the lowest variance was 0.0011 s^2. There is no jitter requirement for class 2 traffic.

From Figure 6.2(b), it is obvious that at an allocation of 0.4809 Mbps of bandwidth for class 2 traffic will satisfy the QoS requirement of < 0.05 blocking probability for class 2 traffic, Thus allocating 0.4809 Mbps of bandwidth for class 2 traffic for each household should be sufficient to satisfy the mean delay and blocking probability requirements of class 2 traffic.

Traffic class 3 bandwidth required

From the above section, it has been determined that ISP needs to allocate a minimum of 21.44 Mbps of bandwidth for class 1 traffic and a minimum of 0.49 Mbps of bandwidth for class 2 traffic for each household. To determine the amount of bandwidth that needs to be allocated for class 3 traffic, we run the traffic model at $n\mu_1$ = 21.44 Mbps and at $n\mu_2$ = 0.49 Mbps. The plots in Figure 6.3 illustrate the relationship between the class 3 bandwidth required versus mean delay and the class 3 bandwidth required versus blocking probability for class 3 traffic given that class 1 traffic is allocated a bandwidth of 21.44 Mbps and class 2 traffic is allocated a bandwidth of 0.49 Mbps.

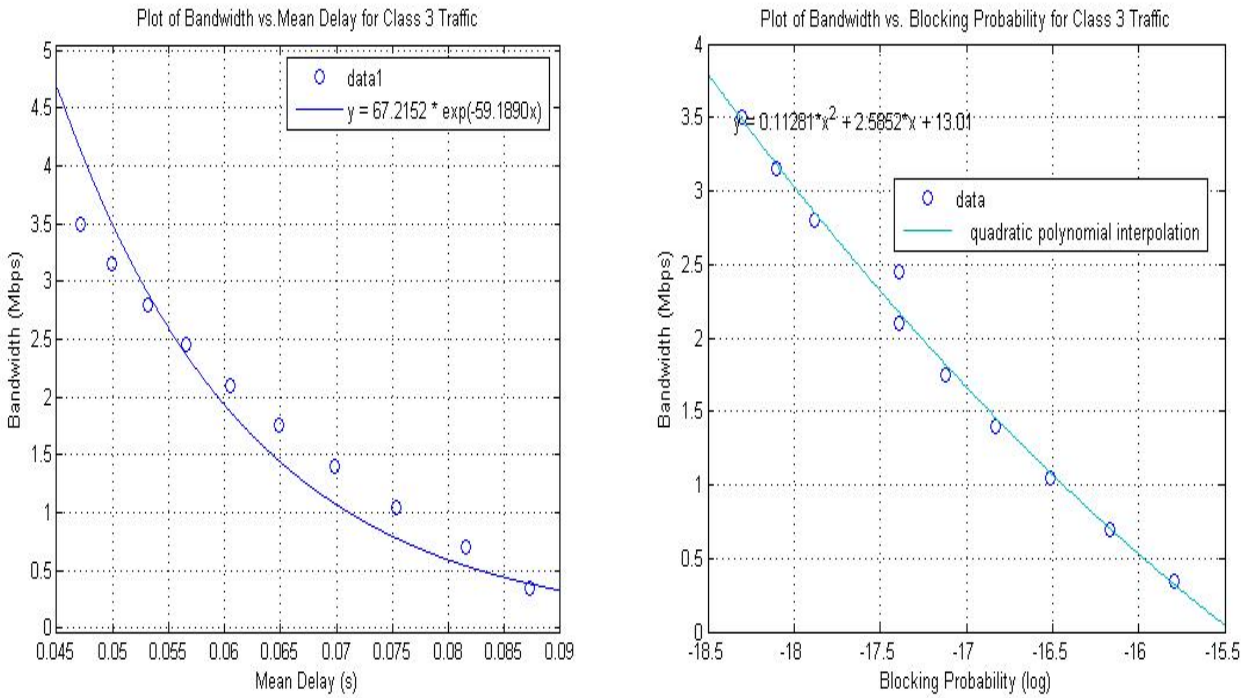


Figure 6.3(a): Plot of Bandwidth Required vs. Mean Delay for Class 3Traffic , and (b) Plot of Bandwidth Required vs. Blocking Probability for Class 3 Traffic.

Figure 6.3(a) above shows the bandwidth that should be allocated to class 3 traffic given various mean delay requirements for class 3 traffic with $n\mu_1 = 21.44$ Mbps and $n\mu_2 = 0.4809$ Mbps. Figure 6.3(b) shows the amount of bandwidth that should be allocated to class 3 traffic given various blocking probability requirements for class 3 traffic with $n\mu_1 = 21.44$ Mbps and $n\mu_2 = 0.4809$ Mbps. There is no QoS requirement for mean delay for class 3 (unclassified) traffic. The blocking probability should be less than 0.05 for class 3 traffic. To satisfy this QoS requirement, along with meeting the stability constraint $n\mu_3 > 0.25$ Mbps, 0.26 Mbps of bandwidth should be allocated for class 3 traffic for each household.

6.3 Conclusions and Discussions

From the results above, one can draw the conclusion that if we model the channel between the ISP and home area network using our example traffic model of PH/M/7/(9+9+9) preemptive priority queue with input parameters listed in Table 6.1, a bandwidth allocation scheme allocating 21.44 Mbps for real-time traffic, 0.49 Mbps for interactive traffic, and 0.26 Mbps for unclassified traffic for each household will minimize the total bandwidth required to satisfy the QoS constraints of all three traffic classes.

Needless to say, the example system serves as an illustrative purpose only, demonstrating the possibility of using a PH/M/n/m preemptive priority queue as an analytical model to help estimate the amount of bandwidth required to be allocated to each traffic class while satisfying the QoS constraints of each traffic class. As mentioned before, all the input parameters are estimates only; in the future, when actual home area networks come into existence, these input parameters should be based on data collected from pilot homes that use home area networks on a normal day-to-day basis. For example, prior to using the traffic model,

sufficient data should be collected to determine the statistical distribution of the arrival process, and then use the moment-generating algorithm described in [24] to approximate the actual distribution with an n-phase EC distribution. Also, data regarding the average number of class i applications that a typical household uses at different times of the day should be collected as well to provide values for the input parameters in Table 6.1. In summary, the purpose of this thesis is to demonstrate the possible use of the traffic model in bandwidth estimation for the channel between ISP and future home area network.

Although this thesis is mainly concerned with the use of the traffic model in estimating network capacity planning for the future HANs, it is possible for ISPs to put the model into immediate usage. One possible use is as follow. Let's say if an ISP owns a certain network infrastructure so that it has the capacity to provide x_1 amount of class 1 bandwidth, x_2 class 2 bandwidth, and x_3 class 3 bandwidth currently. If the ISP is interested in finding out what level of QoS its current network infrastructure can support, it can collect data on current home device usage to be used as information for calculating the packet arrival rate for the three traffic classes in the traffic model. Meanwhile, since the service rates μ_1, μ_2, and μ_3 represents the bandwidth required for class 1, 2, and 3 traffic respectively, the x_1, x_2, and x_3 bandwidth for class 1, 2, and 3 traffic that the ISP currently is capable of providing can be used as the rates μ_1, μ_2, and μ_3 in the traffic model. Given the packet arrival rates (λ_1, λ_2, λ_3) and service rates (μ_1, μ_2, and μ_3) for all the three traffic classes in the traffic model, the mean delay, delay variance, and blocking probability of all three traffic classes can then be determined. This gives a general idea of the level of QoS the ISP is capable of supporting given its current network infrastructure.

Another possibility of putting the traffic model into immediate usage is to use it to study the relationship between packet arrival rates and bandwidth required. An ISP interested in understanding the effects of varying the input parameters in Table 6.1 on bandwidth required and QoS for each traffic class can run the traffic model by varying the input parameters. The table of results for various input parameters can then be used to plot surface graphs to study the relationship between input parameters, bandwidth required, and QoS provided for each traffic class.

In summary, although the emphasis of this thesis is on demonstrating the possible use of the traffic model to optimize the bandwidth required for each traffic class while satisfying the QoS constraints, the traffic model can be put into more immediate use in estimating the relationship between packet arrival rates, bandwidth required, and QoS.

6.4 Future Work

As mentioned before, since there are no future home area networks existing yet, our input parameters are all mostly based on "best-guess" only. Therefore, once pilot homes become available for data collection, one should collect data regarding the actual behavior of home users in using the devices in HAN. Also, the packet arrival distribution of class i traffic should be determined from data collected from actual HAN and then approximated into n-phase EC-distribution using moment generation algorithm.

Appendix A

This section describes the construction algorithm for constructing the submatrices $A_{i,j}$ in the generator matrix Q described by Equation (4.7) on page 28.

Construction Algorithm for $A_{i,i-1}$ Matrix

The matrix $A_{i,i-1}$ is a max(n+m_2-i, m_2+1) x max(n+m_2-i+1, m_2+1) matrix. The mathematical form of the construction algorithm for the matrix $A_{i,i-1}$ can be represented as:

$$(A_{i,i-1})_{row,col} = \begin{cases} i\mu_i I_{(d1*d2*d3)*(\max(n+m_2+1-i-row,m_2+1),\max(n+m_2+1-i-row+1,m_2+1))} & \forall 0<=row<=n+m_2-i, col=row \\ 0 & \forall otherwise \end{cases}$$

(4.9)

The above equation can be expressed in matrix form as follow.

$A_{i,i-1}$ (n+m_2-i+1 by n+m_2-i) =

$$\begin{bmatrix} i\mu_1 I_{(d_q\times d_2\times d_3)\times(\max(n+m_2+1-i-row,m_2+1),\max(n+m_2+1-i-row+1,m_2+1))} & & \\ & \ddots & i\mu_1 I_{(d_1\times d_2\times d_3)\times(\max(n+m_2+1-i-row,m_2+1),\max(n+m_2+1-i-row+1,m_2+1))} \\ & & \ddots \\ & i\mu_1 I_{(d_1\times d_2\times d_3)\times(\max(n+m_2+1-i-row,m_2+1),\max(n+m_2+1-i-row+1,m_2+1))} & 0 \end{bmatrix}$$

where row = row number of the matrix, with 0 as initial row number

(4.10)

Construction Algorithm for $A_{i,i+1}$ Matrix

The matrix $A_{i,i+1}$ is a max(n+m_2-i, m_2+1) x max(n+m_2-i-1, m_2+1) matrix. The mathematical form of the construction algorithm for the generator matrix Q can be represented as:

$$(A_{i,i+1})_{row,col} = \begin{cases} (T_1^o \alpha \otimes I_{d2} \otimes I_{d3}) I_{\max(n+m_2+1-i-row,m_2+1),\max(n+m_2+1-1-i-row,m_2+1)} & \forall 0<=row<=n+m_2-i-1, col=row \\ 0 & \forall otherwise \end{cases}$$

(4.11)

The above equation can be expressed in matrix form as follow.

$A_{i,i+1}$ ($n+m_2$-i-1 by $n+m_2$-i matrix) =

$$\begin{bmatrix} (T_1{}^{o}\alpha \otimes I_d \otimes I_d) I_{\max(n+m_2+1-i-row,m_2+1),\max(n+m_2+1-1-i-row,m_2+1)} & & \\ & (T_1{}^{o}\alpha \otimes I_d \otimes I_d) I_{\max(n+m_2+1-i-row,m_2+1),\max(n+m_2+1-1-i-row,m_2+1)} & \\ & & \ddots \\ & & (T_1{}^{o}\alpha \otimes I_d \otimes I_d) I_{\max(n+m_2+1-i-row,m_2+1),\max(n+m_2+1-1-i-row,m_2+1)} \\ 0 & & \end{bmatrix}$$

(4.12)

Note that in the matrix in (4.12), $(T_1{}^{o}\alpha \otimes I_d \otimes I_d) I_{\max(n+m_2+1-i-row,m_2+1),\max(n+m_2+1-1-i-row,m_2+1)}$ represents a diagonal matrix of size max($n+m_2$+1-i-row, m_2+1) by max($n+m_2$+1-1-i-row, m_2+1) as follow:

$$(T_1{}^{o}\alpha \otimes I_d \otimes I_d) I_{\max(n+m_2+1-i-row,m_2+1),\max(n+m_2+1-1-i-row,m_2+1)}$$

$$= \begin{bmatrix} (T_1{}^{o}\alpha \otimes I_d \otimes I_d) & & & \\ & (T_1{}^{o}\alpha \otimes I_d \otimes I_d) & & \\ & & \ddots & \\ & & & (T_1{}^{o}\alpha \otimes I_d \otimes I_d) \end{bmatrix}$$

This notation also applies to the subsequent matrix elements describing the generator matrix Q.

Construction Algorithm for $A_{i,i}$ Matrix

The matrix $A_{i,i}$ is a max($n+m_2$-I, m_2+1) x max($n+m_2$-i, m_2+1) matrix. The mathematical form of the construction algorithm for matrix $A_{i,i}$ can be represented as:

$$(A_{i,i})_{row,col} = \begin{cases} A_{i,i,\min(row,n-i),\min(row,n-i)} \\ A_{i,i,row,row} \\ \min(row,n-i) * \mu_2 I_{(d1*d2*d3)*(\max(n+m_2+1-i-row,m_2+1),\max(n+m_2+1-i-row+1,m_2+1))} \\ (I_{d1} \otimes T_2^{\,o} \alpha \otimes I_{d3}) I_{\max(n+m_2+1-i-row,m_2+1),\max(n+m_2+1-i-row-1,m_2+1)} \end{cases}$$

$$\begin{gathered} \forall 0 <= row <= n+m_2-i-1, col = row \\ \forall row = n+m_2-i, col = row \\ \forall 1 <= row <= n+m_2-i, col = row-1 \\ \forall 0 <= row <= n+m_2-i-1, col = row+1 \end{gathered}$$

(4.13)

The above equation can be expressed in matrix form as follow.

$A_{i,i}$ ($n+m_2+1-i$ by $n+m_2+1-i$ matrix) =

$$\begin{bmatrix} A_{i,i,\min(row,n-i),\min(row,n-i)} & (I_{d_1} \times T_2^{\,o} \alpha \times I_{d_3}) I_{\max(n+m_2+1-i-row,m_2+1),\max(n+m_2+1-i-row-1,m_2+1)} & & \\ \min(row,n-i) \times \mu_2 I_{(d_1 \times d_2 \times d_3) \times (\max(n+m_2-i-row,m_2+1),\max(n+m_2-i-row+1,m_2+1))} & A_{i,i,\min(row,n-i),\min(row,n-i)} & (I_{d_1} \times T_2^{\,o} \alpha \times I_{d_3}) I_{\max(n+m_2+1-i-row,m_2+1),\max(n+m_2+1-i-row-1,m_2+1)} & \\ & \min(row,n-i) \times \mu_2 I_{(d_1 \times d_2 \times d_3) \times (\max(n+m_2-i-row,m_2+1),\max(n+m_2-i-row+1,m_2+1))} & \ddots & \ddots \\ \ddots & \ddots & A_{i,i,\min(row,n-i),\min(row,n-i)} & (I_{d_1} \times T_2^{\,o} \alpha \times I_{d_3}) I_{\max(n+m_2+1-i-row,m_2+1),\max(n+m_2+1-i-row-1,m_2+1)} \\ & & \min(row,n-i) \times \mu_2 I_{(d_1 \times d_2 \times d_3) \times (\max(n+m_2-i-row,m_2+1),\max(n+m_2-i-row+1,m_2+1))} & A_{i,i,row,row} \end{bmatrix}$$

(4.14)

<u>Construction Algorithm for $A_{i,i,j,j}$ Matrix Except for $j=n+m_2-i$</u>

The matrix $A_{i,i,j,j}$ (except for A matrix where $j = n + m_2 - i$) is a $\max(n+m_2-i, m_2+1)$ x $\max(n+m_2-i, m_2+1)$ matrix. The mathematical form of the construction algorithm for matrix $A_{i,i,j,j}$ can be represented as:

$$(A_{i,i,j,j})_{row,col} = \begin{cases} (T_1 \oplus T_2 \oplus T_3) - i\mu_1 I_{d1*d2*d3} - j\mu_2 I_{d1*d2*d3} - \min(row, n-i-j) * \mu_3 I_{d1*d2*d3} \\ (T_1 \oplus T_2) - i\mu_1 I_{d1*d2*d3} - j\mu_2 I_{d1*d2*d3} - \min(row, n-i-j) * \mu_3 I_{d1*d2*d3} \\ \min(row, n-i-j) * \mu_3 I_{d1*d2*d3} \\ I_{d1} \otimes I_{d2} \otimes T_3^{\,o} \alpha \\ 0 \end{cases}$$

$$\forall 0 <= row <= n + m_3 - i - j - 1, col = row$$
$$\forall row = n + m_3 - i - j, col = row$$
$$\forall 1 <= row <= n + m_3 - i - j, col = row - 1$$
$$\forall 0 <= row <= n + m_3 - i - j - 1, col = row + 1$$
$$\forall otherwise$$

(4.15)

The above equation can be expressed in matrix form as follow.

$A_{i,i,j,j}$ =

$$\begin{bmatrix} (T_1 \oplus T_2 \oplus T_3) - i\mu_1 I_{d_1 \times d_2 \times d_3} - j\mu_2 I_{d_1 \times d_2 \times d_3} - \min(row, n-i-j) \times \mu_3 I_{d_1 \times d_2 \times d_3} & I_{d_1} \times I_{d_2} \times T_3^o \alpha & & \\ \min(row, n-i-j) \times \mu_3 I_{d_1 \times d_2 \times d_3} & (T_1 \oplus T_2 \oplus T_3) - i\mu_1 I_{d_1 \times d_2 \times d_3} - j\mu_2 I_{d_1 \times d_2 \times d_3} - \min(row, n-i-j) \times \mu_3 I_{d_1 \times d_2 \times d_3} & . & \\ & \min(row, n-i-j) \times \mu_3 I_{d_1 \times d_2 \times d_3} & . & . \\ . & . & . & \\ . & (T_1 \oplus T_2 \oplus T_3) - i\mu_1 I_{d_1 \times d_2 \times d_3} - j\mu_2 I_{d_1 \times d_2 \times d_3} - \min(row, n-i-j) \times \mu_3 I_{d_1 \times d_2 \times d_3} & I_{d_1} \times I_{d_2} \times T_3^o \alpha & \\ & \min(row, n-i-j) \times \mu_3 I_{d_1 \times d_2 \times d_3} & (T_1 \oplus T_2) - i\mu_1 I_{d_1 \times d_2 \times d_3} - j\mu_2 I_{d_1 \times d_2 \times d_3} - \min(row, n-i-j) \times \mu_3 I_{d_1 \times d_2 \times d_3} & \end{bmatrix}$$

(4.16)

Construction Algorithm for $A_{i,i,j,j}$ Matrix where $j=n+m_2-i$

The matrix $A_{i,i,j,j}$ where $j=n+m_2-i$ is a (m_2+1) x (m_2+1) matrix. The mathematical form of the construction algorithm for matrix $A_{i,i,j,j}$ can be represented as:

$$(A_{i,i,j,j})_{row,col} = \begin{cases} T_3 - i\mu_1 I_{d1*d2*d3} - (n-i)\mu_2 I_{d1*d2*d3} & \forall 0 <= row <= m_3 - 1, col = row \\ -i\mu_1 I_{d1*d2*d3} - (n-i)\mu_2 I_{d1*d2*d3} & \forall row = m_3, col = row \\ I_{d1} \otimes I_{d2} \otimes T_3^o \alpha & \forall 0 <= row <= m_3 - 1, col = row + 1 \\ 0 & \forall otherwise \end{cases}$$

(4.17)

The above equation can be expressed in a (m_2+1) x (m_2+1) matrix form as follow.

$A_{i,i,j,j} =$

$$\begin{bmatrix} T_3 - i\mu_1 I_{(d_1\times d_2\times d_3)} - (n-i)\mu_2 I_{(d_1\times d_2\times d_3)} & I_{d_1}\otimes I_{d_2}\otimes T_3^{o}\alpha & & \\ 0 & \ddots & \ddots & \\ & \ddots & \ddots & \ddots \\ & & \ddots & T_3 - i\mu_1 I_{(d_1\times d_2 x d_3)} - (n-i)\mu_2 I_{(d_1\times d_2\times d_3)} \quad I_{d_1}\otimes I_{d_2}\otimes T_3^{o}\alpha \\ & & & 0 \quad -i\mu_1 I_{(d_1\times d_2\times d_3)} - (n-i)\mu_2 I_{(d_1\times d_2\times d_3)} \end{bmatrix}$$

(4.18)

Construction Algorithm for A_2 Matrix

The matrix A_2 is a (m_2+1) x (m_2+1) matrix. The mathematical form of the construction algorithm for matrix A_2 can be represented as:

$$(A_2)_{row,col} = \begin{cases} n\mu_1 I_{(m_2+1)*(d1*d2*d3)} & \forall 0 <= row <= m_2, rol = col \\ 0 & \forall otherwise \end{cases} \quad (4.19)$$

The above equation can be expressed in a (m_2+1) x (m_2+1) matrix form as follow.

$$A_2 = \begin{bmatrix} n\mu_1 I_{(m_2+1)\times(d_1\times d_2\times d_3)} & & \\ & \ddots & \\ & & \ddots \\ & & n\mu_1 I_{(m_2+1)\times(d_1\times d_2\times d_3)} \end{bmatrix}$$

(4.20)

Construction Algorithm for A_1 Matrix

The matrix A_1 is a (m_2+1) x (m_2+1) matrix. The mathematical form of the construction algorithm for matrix A_1 can be represented as:

$$(A_1)_{row,col} = \begin{cases} A_{n,n,0,0} & \forall 0 <= row <= m_2 - 1, col = row \\ A_{n,n,m_2,m_2} & \forall row = m_2, col = row \\ (I_{d1}\otimes T_2^{\,o}\alpha\otimes I_{d3})I_{m_2+1} & \forall 0 <= row <= m_2 - 1, col = row + 1 \\ 0 & \forall otherwise \end{cases}$$

(4.21)

The above equation can be expressed in a a (m_2+1) x (m_2+1) matrix form as follow.

$$A_1 = \begin{bmatrix} A_{n,n,0,0} & (I_{d_1} \otimes T_2{}^{o}\alpha \otimes I_{d_3})I_{m_2+1} & & \\ 0 & . & . & \\ & . & . & . \\ & & . & A_{n,n,0,0} \quad (I_{d_1} \otimes T_2{}^{o}\alpha \otimes I_{d_3})I_{m_2+1} \\ & & & 0 \qquad\qquad A_{n,n,m_2,m_2} \end{bmatrix}$$

(4.22)

Construction Algorithm for A_0 Matrix

The matrix A_0 is a (m_2+1) x (m_2+1) matrix. The mathematical form of the construction algorithm for matrix A_0 can be represented as:

$$(A_0)_{row,col} = \begin{cases} (T_1{}^{o}\alpha \otimes I_{d2} \otimes I_{d3})I_{m_2+1} & \forall 0 <= row <= m_2, col = row \\ 0 & \forall otherwise \end{cases}$$

(4.23)

The above equation can be expressed in a a (m_2+1) x (m_2+1) matrix form as follow.

$$A_0 = \begin{bmatrix} (T_1{}^{o}\alpha \otimes I_{d_2} \otimes I_{d_3})I_{m_2+1} & & \\ & . & \\ & & . \\ & & . \\ & & (T_1{}^{o}\alpha \otimes I_{d_2} \otimes I_{d_3})I_{m_2+1} \end{bmatrix} \qquad (4.24)$$

References

[1] 802.1.D™ IEEE Standard for Local and Metropolitan Area Networks: Media Access Control (MAC) Bridge. IEEE Computer Society, 2004; available online at www.ieee.org or http://standards.ieee.org/getieee802.

[2] Adams, C.E. Home Area Network Technologies, *BT Technology Journal*, 20(2): 53-72, April 2002.

[3] D. Aldous and L. Shepp. The least variable phase type distribution is Erlang. Communications in Statistics, *Stochastic Models*, 3(3):467-473, 1987.

[4] T. Altiok. On the phase-type approximations of general distributions. *IIE Transactions*, 17(2):110-116, 1985.

[5] CableHome 1.1 Specification. August 12, 2005; available at http://www.cablelabs.com/projects/cablehome/specifications/.

[6] Carney, W. *IEEE 802.11g: New Draft Standard Clarifies Future of Wireless LAN*, white paper. Texas Instruments, 2002.

[7] E. Frank and J. Holloway, Connecting the Home With a Phone Line Network Chip Set, *IEEE Micro*, March 2000; available online at http://www.homepna.org/docs/wp1.asp

[8] S. Gardner et al, HomePlug Standard Brings Networking to the Home, November 2, 2003; available online at http://www.commsdesign.com/main/2000/12/0012feat5.htm

[9] HomePNA Specification 1.0 Field Tests Status. HomePNA Inc., 1999; available online at http://www.homepna.org

[10] IEEE Standards 802.11™a-1999(R2003). Institute of Electrical and Electronics Engineers Inc, 2003. Available online at http://standards.ieee.org/getieee802/.

[11] IEEE Standards 802.11™b-1999(R2003). Institute of Electrical and Electronics Engineers Inc, 2003. Available online at http://standards.ieee.org/getieee802/.

[12] IEEE Standards 802.11™g-2003. Institute of Electrical and Electronics Engineers Inc, 2003; available online at http://standards.ieee.org/getieee802/.

[13] Cecile de Jong et al, RGE-D.2.1 Access in the Netherlands in 2002. 2002; available online at http://www.rge.brabantbreedband.nl/.

[14] Gross, D. and Harris, C.M. *Fundamentals of Queueing Theory*. Chapter5, 3rd ed. John Wiley & Sons, Inc., New York, 1998.

[15] Kangude, S., Copeland J., and Sherman M., An Analysis of the Home PAN Collision Resolution Mechanism, *28th Annual IEEE International Conference on Local Computer Networks (LCN'03)* pp. 250-259.

[16] Kwaaitaal et al, RGE-D3.1 Home Networking Technologies Overview and Analysis. Residential Gateway Environment Project. December 2003. Available online at http://www.rge.brabantbreedband.nl/.

[17] Kwaaitaal et al, RGE-D3.2 Device and Service Discovery in Home Networks: Overview and Analysis. Residential Gateway Environment Project. May 2004. Available online at http://www.rge.brabantbreedband.nl/.

[18] Lei, B., Ananda, A.L., and Teck, T.S., QoS-aware Residential Gateway, *Proceedings of the 27th Annual IEEE Conference on Local Computer Networks*, 2002. pp. 518-524.

[19] Leon-Garcia, A., *Probability and Random Processes for Electrical Engineering*, Chapter 3, 2nd ed. Addison-Wesley Publishing Company, Inc., New York, 1994.

[20] Miller, B.A., Nixon, T., Tai, C., and Wood, M.D. Home Networking with Universal Plug and Play. *IEEE Communications Magazine*, 39(12):104-109.

[21] Myers, E. HomeRF Overview and Market Positioning. Available online at http://www.palowireless.com/homerf/homerf.asp

[22] Neuts, Marcel F., *Matrix-Geometric Solutions in Stochastic Models: An Algorithmic Approach*. Dover Publications Inc., 1994.

[23] C. A. O'Cinneide. Phase-type Distributions and Majorization. *Annals of Applied Probability*, 1(3):219-227, 1991.

[24] Osogami, T., *Analysis of Multi-server Systems via Dimensionality Reduction of Markov Chains*. PhD thesis, Carnegie Mellon University, Pittsburgh, USA. June 2005.

[25] Pluijmaekers et al. RGE-D.4.3. Modelling of the performance of a hierarchical caching architecture. May 2004. Available online at http://www.rge.brabantbreedband.nl/.

[26] *Quality of Service in the Home Networking Model*, white paper. HomeRF Working Group, 2001; available online at www.homerf.org.

[27] Rose, B. Home Networks: A Standards Perspective. *IEEE Communications Magazine*, 39(12): 78-85, December 2001.

[28] Turnbull, J. G. Introducing Home Area Networks. *BT Technology Journal*, 20(2): 30-38, April 2002.

[29] UPnP QoS Architecture: 1.0. March 10, 2005; available at http://www.upnp.org

[30] Venkataraman, P. *Applied Optimization with MATLAB Programming*. pp. 214 - 217. John Wiley & Sons, Inc., 2002.