# Merging of Bursty Traffic in Weakly Stable Markovian Networks

by

Syed Irfan Chanth Basha

B.S. (C.S), University of Madras, Chennai, 2002

Submitted to the Department of Electrical Engineering and Computer Science and the Faculty of the Graduate School of the University of Kansas in partial fulfillment of the requirements for the degree of Master of Science in Computer Science

---

Dr.Victor Wallace, Chair

---

Dr.Man Kong, Chair

---

Dr.Victor Frost, Member

---

Date Thesis Accepted

*Dedicated to my parents.*

# Acknowledgments

All thanks due to Allah.

If it was not for my parents then I would not be here. If it was not for my advisors then I would not have left here successfully without achieving what I have achieved. My sincere and humble thanks to Dr.Victor L.Wallace for being there for me when I needed him the most from the very beginning of my life in KU. My utmost gratitude for Dr.Man Kong for providing me with invaluable advice both in the academic and non-academic aspects of life in the US. My work in this field could not be complete if it was not for the support provided by Dr.Appie van de Liefvoort. I am in debt to him for supporting me in my thesis and research and for his valuable time spent on me. My thanks to Dr.Victor Frost for providing the necessary help when requested in both my research and thesis.

I would also like to thank Mr.Ramakrishnan Krishnaswamy, fellow Masters student and fellow research assistant for listening patiently to the ideas "bounced" off him by me and thanks to Mr.Jayesh Kumaran for providing timely help when requested. I would definitely like to thank Mr.William J.Routt, my manager in Sprint, for understanding the importance of my work as a graduate student and constantly providing encouragement to complete my thesis while working with Sprint. Finally, to all of my friends and family who help me complete my Masters, phew ...

**Abstract**

The relationship between burstiness (a form of multi time-scale auto-dependence) in ethernet traffic and its effect on backbone network performance has been explored using new tools for the construction and analysis of dependent matrix-exponential (MED) queueing network models. These tools are tailored to treat models with pseudo heavy tailed distributions and auto-covariances which are significant for lags extending over many orders of magnitude. They also introduce a new flexibility into the construction and solution of many station network models involving such processes.

An algebraic approach extending the art of MED queueing to more complex and extended networks of queues is outlined and used in the generation of an access/backbone mode. It employs the tools of Kronecker products, hat spaces, and nearly completely decomposable (NCD) operators to model event matrices modularly and hierarchically in time, function and network scale.

The tools also improve algebraic intuition and computational effectiveness. The models examined allow multiple access-networks with realistically bursty, correlated traffic to be included in a network with a backbone network model. The models employ the concept of weak stability and multi-modal MED traffic to explore the effects of the multi-time-scale properties found in real traffic. Recent evidence confirms that these are plausible models that can faithfully represent the pathologies of published trace data.

The models have been explored to demonstrate the development of insight into the effect of access network traffic and design on call losses on the backbone. Interesting in-

sights have been presented, and offer guidance towards future, deeper explorations of other properties of such models.

# Contents

# List of Tables

# List of Figures

# Chapter 1

# Introduction

It is well known that traffic in local area networks is often severely "bursty" in character, giving rise to persistent problems in guaranteeing an acceptable quality of service for network clients. Burstiness of traffic can result in poor performance due to limited buffering available in the network.

With bursty traffic, each time a burst occurs there would be a rapid filling of buffer space on the access network and the backbone network, followed by a slower draining as traffic calms. This results in excessive packet loss in both the access and backbone networks, and consequently excessive delivery delays due to retransmissions and other effects. The intermittent bursts also introduce large queueing delays due to the very full buffers, and the time it takes for them to empty.

Design of network control under these circumstances requires understanding of the

mechanism by which the phenomenon works, and how performance is affected by changes in traffic or design. Quantitative models of performance are needed which are predictive of the effect of changes to parameters of design upon the desired measures of performance.

Much effort has been expended to determine an analytical explanation of the causes of this random burstiness and its effects upon the buffering medium [1][7][13][14][15][21]. The most intriguing of the ideas to come out of this work is the notion of weak stability [5] in a queue. Under this view, the input stream is made up of sub-streams of traffic, each with different average traffic rates. These are switched among by a much slower process. We refer to these sub-streams as modes. Jelenkovic [5] calls modes whose average rate of arrival is not less than the average service rate *unstable* modes, and otherwise *stable*. If the average arrival rate of the combined process is less than the average service rate and there is at least one unstable sub-stream, then the queue is termed *weakly-stable*, signifying that, though the whole process is stable, it becomes unstable for intervals of time, from time to time.

Both intuitively and by design, it can be seen that if the average service rate of the server is less than the fastest rate mode, then the queue will build rapidly until the mode is once again stable[5][6]. So the weak-stability explains the observation that for such traffic, there could be buffer overflows even with very large buffers. If the service rate of the server is increased then, to compensate for the buffer overflow, the utilization of the server would conversely have to be extremely low. This process of striking a balance between the utilization and buffer capacity has sparked the need for various solution models.

Though the effect of burstiness on a single LAN is now reasonably well understood

[8][6], very little work has been done to analyze how this burstiness, filtered through the buffering medium, affects the backbone network performance. It is our purpose to examine the effect of merging the output of several access networks with bursty traffic onto a backbone network. We propose to extend this understanding to the backbone.

We propose to do this analytically, using some dependent matrix-exponential (MED) queueing models which have recently been developed[6][8]. These models exploit the concept of nearly completely decomposable (NCD) matrix exponential (ME) modes to represent the traffic, and exploit the concept of weakly stable MED queues to represent the effect of the traffic on performance [19]. By merging output traffic from weakly stable queues, we believe we will have a simple credible model for approaching the behavior and stochastic structure of a backbone network being fed from multiple bursty access networks. The exploration of the consequences of merging these models is expected to yield insight into what the merging of bursty streams of traffic does to the performance of backbone queues.

To better assist us in the development of such complex models involving many dimensions of state space, we utilize the concepts of Kronecker operations and Hat Space. These tools aid in the formation of event matrices hierarchically in either time, function or network scale. They also improve algebraic intuition and effectiveness when addressing a variety of design issues.

As stated earlier, the nearly completely decomposable (NCD), matrix exponential traffic modes represent the traffic entering the model. When one or more NCD classes of states have an arrival rate markedly faster than the average arrival rate then we call them

burst modes [6][8]. Transition between these burst modes and states with much slower rates of arrival (base modes) are relatively less frequent than arrival events. Also, it is understood that the time spent in these burst modes must be a small fraction of the time spent in base modes. Krishnaswamy [8] has demonstrated that the Bellcore traffic traces exhibit the modal behavior used here.

The transition from one mode to another is instantaneous. A transition from a burst mode almost always results in a return to a non-bursty base mode [8]. Such traffic characteristics, keeping in memory the non-bursty mode from which the transition to bursty mode took place, can pose an interesting challenge, since they clearly violate the memoryless property assumed in conventional analytical performance studies.

When bursty traffic is filtered through the access network the output of the access network will still be bursty, though usually to a lesser degree because of the packet losses incurred, as well as the spreading of inter-arrival times caused by the access server. Thus, the backbone network is fed by traffic which is spread and clipped, tough yet bursty. Our model will allow analysis of the effect of this action.

The flexibility of applying the algebraic concepts for the analysis of the ethernet/internet traffic is that we can trace the effects of events when the merging of this filtered traffic in the backbone from multiple access networks occurs. The insight drawn in this research is to explore the behavior in the backbone. This work will take the simplest possible network models exhibiting the merging behavior and to explore what insights develop. The issues to be dealt with are to predict the forms of burstiness in a backbone from known burstiness in

the access networks. Success is determined by the quality of insights drawn and explanations of the phenomena already published.

The mathematical tools to be used are discussed with appropriate examples and the analysis of the network with focus on the call-loss probability is completed and documented towards the later half of this thesis. The organization of the thesis is given below.

Chapter two describes mathematically the tools that are available and used. A brief explanation about the matrix exponential representations of traffic (event sequences) and systems, and the familiar formulae in use for this kind of problem is given. It also includes moment matching technologies [17] to reduce state space in modelling. An introduction and analysis of the Kronecker products and hats are given, which in combination is a tool for representing the complex matrices needed and finally the modal ME representation for dependent traffic in local networks.

In chapter three, an ME network theory is described along with its representations of network components, their modal dependencies and the equations to be solved. The general techniques used for the construction of modal queues in meaningful ways are illustrated with simple examples without losing its focus to clarify the generality of the technique.

Having laid the foundation for understanding the network, chapter four deals with the development of the pieces of the general backbone model and the equations that put it together, step by step. It includes the general formulation and then illustrates it with the simplified model that will be used for insightful results in the following chapter, chapter five. It explains the need for the insight generating model and justifies the simplification done

which are acceptable.

The last chapter, chapter five, describes the measures and strategy involved with the execution of test cases. This is where the discoveries and confirmations of the findings are explained and reasoned. It also includes the explanation and reasoning behind the choices done in selecting the BBTR (burst to base traffic ratio) as one of the critical parameters. This chapter uses the results obtained by the trace driven simulation done by Krishnaswamy [8] for Bellcore (October) traffic as its bench mark to maintain the burst stability in the network. The final chapter also includes the potential limitations faced by this model and some suggested ways it can be further extended in its purpose and application.

# Chapter 2

# Background

This chapter deals with the mathematical foundation of this thesis. A brief walk through on the concepts of linear algebraic queueing theory including the matrix exponential representation of traffic is given here. An introduction to the Kronecker operators is dealt here along with the modal ME representation for dependent traffic.

## 2.1 Matrix exponential distribution

A matrix exponential distribution is defined as a probability distribution with representation $(\mathbf{p}, \mathbf{B}, \boldsymbol{\varepsilon}')$ [9] and [11]

$$F(t) = 1 - \mathbf{p}exp(-\mathbf{B}t)\boldsymbol{\varepsilon}', t \geq 0, \tag{2.1}$$

where $\mathbf{p}$ is the starting vector for the process, $\mathbf{B}$ is the process rate operator which must be non-singular and $\boldsymbol{\varepsilon}'$ (the transpose of $\boldsymbol{\varepsilon}$) is the summing operator.

The order of the representation is indicated by the dimension of the $\mathbf{B}$ matrix and the degree of the distribution $F(t)$ is the minimal order of all its representations. Its probability density function is defined as [12][17]

$$f(t) = \frac{dF(t)}{dt} = \mathbf{p}exp(-\mathbf{B}t)\mathbf{B}\boldsymbol{\varepsilon}' \tag{2.2}$$

If $T_1, T_2, T_3, \ldots$ is a sequence of ME random variables then the joint probability density function over any finite sequence inter-event times is given by

$$f_{T_1,T_2,\ldots,T_n}(t_1, t_2, \ldots, t_n) = \pi(0)exp(-\mathbf{B}t_1)\mathbf{L}\ldots exp(-\mathbf{B}t_n)\mathbf{L}\boldsymbol{\varepsilon}' \tag{2.3}$$

where $\pi(t)$ is a vector representing the internal state of the process at time t and $\mathbf{L}$ is the event rate matrix. If the process is renewal then $\mathbf{L} = \mathbf{B}\boldsymbol{\varepsilon}'\mathbf{p}$ where $\mathbf{p}$ is the starting vector for the process, the rank of $\mathbf{L}$ being 1. The $n^{th}$ moments satisfy the following

$$E[X^n] = \int_0^\infty x^n f(t)dt = n!\mathbf{p}\mathbf{V}^n\boldsymbol{\varepsilon}' \tag{2.4}$$

where $\mathbf{V} = \mathbf{B}^{-1}$. The Laplace-Stieltjes Transform of $f(t)$ is given by

$$\mathbf{B}^*(s) = \int_0^\infty exp^{st} f(t)dt = \mathbf{p}(\mathbf{I} + s\mathbf{V})^{-1}\boldsymbol{\varepsilon}' = \mathbf{p}(\mathbf{B} + s\mathbf{I})^{-1}\mathbf{B}\boldsymbol{\varepsilon}' \tag{2.5}$$

Matrix exponential distributions have rational Laplace-Stieltjes transform and are more general than the phase-type (PH) distributions as defined by Neuts [12]. ME distributions place fewer constraints on its representation. Although the class of second degree Matrix exponential distributions is equivalent to the physically based phase-type distributions higher degree representation may not have physical representation. Phase type distributions are, in fact, a strict subset of matrix exponential distributions.

According to Neuts [12], the class of distributions with rational Laplace-Stieltjes transforms is dense in the set of all distributions, which means that any density function can be approximated arbitrarily closely by a density function with a rational transform. Some of such distributions are exponential, Erlangian, Coxian, hypoexponential, hyperexponential, Marie and mixtures of convolutions of these distributions. Distributions belonging to the above class have probabilistically interpretable components and have close relationship with Markov chains. An advantage of using matrix exponential distributions is that higher order moments can be matched using Van de Liefvoort's algorithm [17] and they can be represented in different canonical forms through the use of similarity transforms. The LAQT solution method does not depend on the process representation and hence any representation can be chosen while modelling.

## 2.2 Moment matching

Van de Liefvoort's algorithm [17] can be used to match the moments of the distribution. From the set of power moments $E[X^n]$ of a continuous distribution $F(t)$ an ME distribution

$(\mathbf{p}, \mathbf{B}, \epsilon)$ can be generated. Let,

$$r_n = \frac{E[X^n]}{n!} \tag{2.6}$$

be the set of normalised or reduced moments of the distribution. Applying the algorithm the following was generated by Van de Liefvoort [17],

$$\mathbf{p} = \begin{bmatrix} 1 & 0 \end{bmatrix} \tag{2.7}$$

$$\mathbf{V} = \begin{bmatrix} r_1 & r_1 \\ (r_2 - r_1^2)/(r_1) & (r_3 - 2r_1r_2 + r_1^3)/(r_2 - r_1^2) \end{bmatrix} \tag{2.8}$$

$$\boldsymbol{\varepsilon}' = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \tag{2.9}$$

using the first three moments. This representation has the prescribed moments and a rational Laplace-Stieltjes transform. The boundary conditions of the power moments have been discussed in [12]. For a third order representation, the first five moments are mapped into $(\mathbf{p}, \mathbf{B}, \boldsymbol{\varepsilon}')$.

$$\mathbf{p} = \begin{bmatrix} 1 & 0 & 0 \end{bmatrix} \tag{2.10}$$

$$\mathbf{V} = \begin{bmatrix} r_1 & r_1 & 0 \\ \frac{r_2 - r_1^2}{r_1} & \frac{r_3 - 2r_1r_2 + r_1^3}{r_2 - r_1^2} & r_1 \\ 0 & -\frac{r_n^- 2r_1r_2r_3 + r_3^2 + r_4r_1^2 - r_4r_2}{(r_1^2 - r_2)^2 r_1} & \frac{\beta}{\gamma} \end{bmatrix}$$

$$\boldsymbol{\varepsilon}' = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$$

where

$$\beta = -r_2^4 r_1 + 3r_2^2 r_1^2 r_3 - 2r_1 r_2 r_3^2 - 2r_2 r_1^3 r_4 + 2r_2^2 r_1 r_4 - r_3^2 r_1^3 + r_3^3 + 2r_3 r_4 r_1^2$$
$$- 2r_3 r_2 r_4 + r_5 r_1^4 - 2r_5 r_1^2 r_2 + r_5 r_2^2,$$
$$\gamma = (r_2^3 - 2r_1 r_2 r_3 + r_3^2 + r_4 r_1^2 - r_4 r_2)(r_1^2 - r_2)$$

## 2.3 Introducing correlations

Auto-correlations can be arbitrarily introduced into a renewal process preserving the marginal distribution using Mitchell's method [11]. The progress rate matrix of the arrival process ($\mathbf{B}$) describes what happens between the arrival events and the event rate matrix describes ($\mathbf{L}$) what happens at the time of an arrival. $\mathbf{L}$ can be altered from representing a renewal process to represent a semi-Markov arrival process. The starting vector $\mathbf{p}$ and the progress rate matrix $\mathbf{B}$ are kept unchanged from the renewal process representation. An $\mathbf{L}$ can be chosen such that $\mathbf{p}$ and $\mathbf{B}$ remain unchanged,

$$\mathbf{L} = \beta(\mathbf{B}\boldsymbol{\varepsilon}'\mathbf{p} - \mathbf{B}) + \mathbf{B} \tag{2.11}$$

$$\mathbf{L} = (1 - \gamma)(\mathbf{B}\boldsymbol{\varepsilon}'\mathbf{p} - \mathbf{B}) + \mathbf{B} \tag{2.12}$$

where from Mitchell's et.al [11] $\gamma$ is the equilibrium solution for the embedded chain on $\mathbf{B}^{-1}\mathbf{L}$ and $\beta = 1 - \gamma$ is called the persistence of the process, and as $\beta \to 0$, the autocorrelation exhibited by the point process increases. This method can be used to match the decay of the desired autocorrelation structure rather than matching the entire autocorrelation structure of the process, which Mitchell has found out to be more effective. The covariance of a sequence of MEs (assuming stationarity) is given by,

$$cov[X_n, X_{n+k}] = \mathbf{pV}(\mathbf{Y})^k\mathbf{V}\boldsymbol{\varepsilon}' - (\mathbf{pV}\boldsymbol{\varepsilon}')^2 \tag{2.13}$$

and the variance,

$$var[X_0] = 2\mathbf{pV}^2\boldsymbol{\varepsilon}' - (\mathbf{pV}\boldsymbol{\varepsilon}')^2 \tag{2.14}$$

where $\mathbf{V} = \mathbf{B}^{-1}$ and $\mathbf{Y} = \mathbf{VL}$ and where $\mathbf{p}$ and $\boldsymbol{\varepsilon}'$ are chosen so that $\mathbf{Y}\boldsymbol{\varepsilon}' = \boldsymbol{\varepsilon}'$ and $\mathbf{pY} = \mathbf{p}$. Thus $\boldsymbol{\varepsilon}'$ is the right eigenvector of $\mathbf{Y}$ with eigenvalue 1 and $\mathbf{p}$ is the left eigenvector of $\mathbf{Y}$ with eigenvalue 1. The value of $\mathbf{p}$ is assumed to be unique and its existence is guaranteed if 1 is the largest eigenvector of $\mathbf{Y}$. The autocorrelation is obtained by dividing the covariance by the variance,

$$autocorr[X_n, X_{n+k}] = \frac{\mathbf{pVY}^k\mathbf{V}\boldsymbol{\varepsilon}' - (\mathbf{pV}\boldsymbol{\varepsilon}')^2}{2\mathbf{pV}^2\boldsymbol{\varepsilon}' - (\mathbf{pV}\boldsymbol{\varepsilon}')^2} \tag{2.15}$$

If $\mathbf{L} = \mathbf{B}\boldsymbol{\varepsilon}'\mathbf{p}$ then $\mathbf{Y}$ is of rank 1, the process becomes renewal, and hence $cov[X_n, X_{n+k}] = 0$. The matrix $\mathbf{B}$ in the above equations is the autogenous event matrix and the matrix $\mathbf{L}$ is the endogenous event matrix for the MED arrival module. Since there is no input port and only one output port for this module, it is the complete description.

## 2.4 Kronecker operations and hat spaces

The modules we are using are complex involving many dimensions in the state spaces. Establishing a common plane for computation and algebraic manipulation is often frustrating since disjoint operator spaces by nature are disjoint. This has been the driving force to understanding the concepts of Kronecker operations.

Kronecker products are used to combine processes operating in different spaces. It is a way to preserve independence among process in different spaces, whereby disjoint operator spaces are embedded into the direct product space. For example in a ME/ME/1/N queueing system the arrival process $< \mathbf{B}_a, \mathbf{L}_a >$ and the service process $< \mathbf{B}_s, \mathbf{L}_s >$ operate in two disjoint spaces namely the arrival space and the service space. To algebraically represent their behavior in the combined state space, we find it useful to employ Kronecker products.

The Kronecker product is represented by the symbol $\otimes$. According to [10], in order to preserve the independence of each space, the operators in the disjoint spaces are combined in the system space using the Kronecker product.

The Kronecker product of two matrices $\mathbf{K_1}$ (operating on objects in space 1) and $\mathbf{K_2}$ (operating on objects in space 2) is defined by,

$$\mathbf{K} = \mathbf{K_1} \otimes \mathbf{K_2} = \begin{bmatrix} (\mathbf{K}_1)_{11}\mathbf{K}_2 & \dots & (\mathbf{K}_1)_{1n_2}\mathbf{K}_2 \\ \vdots & \ddots & \vdots \\ (\mathbf{K}_1)_{n_1 1}\mathbf{K}_2 & \dots & (\mathbf{K}_1)_{n_1 n_2}\mathbf{K}_2 \end{bmatrix} \tag{2.16}$$

where $\mathbf{K}_1$ is of size $n_1 \times n_2$ and the $\mathbf{K}_2$ is of size $m_1 \times m_2$.

Some properties of Kronecker product [3]:

1. The Kronecker product is a bi-linear operator. Given $\alpha$ real,

$$\mathbf{A} \otimes (\alpha\mathbf{B}) = \alpha(\mathbf{A} \otimes \mathbf{B}) \tag{2.17}$$

$$(\alpha\mathbf{A}) \otimes \mathbf{B} = \alpha(\mathbf{A} \otimes \mathbf{B}) \tag{2.18}$$

2. The Kronecker product distributes over addition:

$$(\mathbf{A} + \mathbf{B}) \otimes \mathbf{C} = (\mathbf{A} \otimes \mathbf{C}) + (\mathbf{B} \otimes \mathbf{C}) \tag{2.19}$$

$$\mathbf{A} \otimes (\mathbf{B} + \mathbf{C}) = (\mathbf{A} \otimes \mathbf{B}) + (\mathbf{A} \otimes \mathbf{C}) \tag{2.20}$$

3. The Kronecker product is associative:

$$(\mathbf{A} \otimes \mathbf{B}) \otimes \mathbf{C} = \mathbf{A} \otimes (\mathbf{B} \otimes \mathbf{C}) \tag{2.21}$$

4. The Kronecker product is not commutative in general. I.e, if $\mathbf{A} \neq \mathbf{B}$, then

$$(\mathbf{A} \otimes \mathbf{B}) \neq (\mathbf{B} \otimes \mathbf{A}) \tag{2.22}$$

5. Matrix multiplication, when dimensions are compatible,

$$(\mathbf{A} \otimes \mathbf{B})(\mathbf{C} \otimes D) = (\mathbf{AC} \otimes \mathbf{BD}) \tag{2.23}$$

The interpretation of ordering in Kronecker products is better visualized when applied to a transition matrix. When applied to this space, the Kronecker product shows the combined effect of simultaneous transitions in two spaces. For example, $\mathbf{K}_1 \otimes \mathbf{K}_2$ shows simultaneous transition in two spaces, whereas $\mathbf{K}_1 \otimes \mathbf{I}_2$ shows an operator with (potentially) change occurring in space 1, but no change in space 2. We refer to $\mathbf{K}_1 \otimes \mathbf{I}_2$ as an *extension* of the operator $\mathbf{K}_1$, which is an operator from space 1 to space 1, into an operator from space (1,2) to space (1,2) combined space. The expression $\mathbf{K}_1 \otimes \mathbf{I}_2$ shows the extension of transitions in one space into two spaces, the second of which is unaffected.

As can be observed, as the number of dimensions used in computation increase, the Kronecker representation gets more complicated. As shall be discussed later, the ordering needs to be taken into special consideration as well. In order for a more intuitive representation of the combination of the two dimensions and spaces, the *hats* are introduced.

Mitchel, et.al; [10] explain clearly the usage of hats. An operator that operates on a space only (say arrival space) is embedded in the system space by taking the Kronecker product of it with the identity operator from the remaining spaces. To preserve the ordering, hat spaces are introduced.

For example if there are N spaces, then a matrix in one space is extended into the

larger space by "hatting" it. So, in natural space ordering $< K_1, K_2, \ldots K_n >$

$$\begin{aligned}
\widehat{\mathbf{K}}_1 &= \mathbf{K_1} \otimes \mathbf{I_2} \otimes \mathbf{I_3} \ldots \otimes \mathbf{I_N} \\
\widehat{\mathbf{K}}_2 &= \mathbf{I_1} \otimes \mathbf{K_2} \otimes \mathbf{I_3} \ldots \otimes \mathbf{I_N} \\
\widehat{\mathbf{K}}_N &= \mathbf{I_1} \otimes \mathbf{I_2} \otimes \mathbf{I_3} \ldots \otimes \mathbf{K_N}
\end{aligned} \tag{2.24}$$

where $\mathbf{I_i}$ is the identity matrix of dimensions $m_i \times m_i$ and $\mathbf{K}_i$ is the matrix of that dimension. The above construction of hats is applicable to square matrices. Each process embedded into system space is differentiated from the non-embedded process by the use of hats, thus the product space is sometimes referred to as *hat space*.

Notice that the hatted symbol identifies in its subscript the space it operates in, and doesn't by itself identify the ordering of a Kronecker representation, which is arbitrary. When rendering a hat equation in its Kronecker form, consistency of order of terms is essential even though the actual order used can be otherwise arbitrarily chosen. This is where the non-commutative nature of the Kronecker product operation plays a role.

Although not commutative, there is a sense in which the ordering of the spaces is arbitrary. Provided the ordering is changed consistently within an equation, the import of the equation does not change. $\mathbf{A} \otimes \mathbf{B}$ differs from $\mathbf{B} \otimes \mathbf{A}$ only in a permutation of rows and columns. For example, equations 2.16 - 2.22 are equally correct if the matrices were rendered in the order **B**, **C**, **A**. All high level matrices in the expressions are permuted in the same way in the result.

Nevertheless, in various manipulations for insight or computational advantage, the ordering is decisively important. $\mathbf{A} \otimes \mathbf{B}$ keeps $\mathbf{B}$ intact in sub-matrices of the product, whereas $\mathbf{B} \otimes \mathbf{A}$ keeps $\mathbf{A}$ intact in the sub-matrices. Some Kronecker orderings will often be more revealing than others for a given purpose and situation, so there is an advantage associated with that choice. Its choice can be deferred in general equations by using the hats. Hence the more compact hat notation is often used rather than explicitly using Kronecker products in the equations.

The property, under space-order $< 1, 2 >$

$$\widehat{\mathbf{K}}_1 \widehat{\mathbf{K}}_2 = \mathbf{K}_1 \otimes \mathbf{K}_2 = \widehat{\mathbf{K}}_2 \widehat{\mathbf{K}}_1 \tag{2.25}$$

satisfies the independence of spaces condition. The above relationship can be proved as given below while taken into consideration the space-order $< 1, 2 >$

$$\widehat{\mathbf{K}}_1 \widehat{\mathbf{K}}_2 = (\mathbf{K}_1 \otimes \mathbf{I}_2)(\mathbf{I}_1 \otimes \mathbf{K}_2) \tag{2.26}$$

$$= (\mathbf{K}_1 \mathbf{I}_1) \otimes (\mathbf{I}_2 \mathbf{K}_2) \tag{2.27}$$

$$= (\mathbf{K}_1 \otimes \mathbf{K}_2)$$

Now taking $\widehat{\mathbf{K}}_2\widehat{\mathbf{K}}_1$ and applying the same ordering $< 1, 2 >$

$$\widehat{\mathbf{K}}_2\widehat{\mathbf{K}}_1 = (\mathbf{I}_1 \otimes \mathbf{K}_2)(\mathbf{K}_1 \otimes \mathbf{I}_2) \tag{2.28}$$

$$= (\mathbf{I}_1\mathbf{K}_1) \otimes (\mathbf{K}_2\mathbf{I}_2) \tag{2.29}$$

$$= \mathbf{K}_1 \otimes \mathbf{K}_2$$

Thus,

$$\widehat{\mathbf{K}}_1\widehat{\mathbf{K}}_2 = \mathbf{K}_1 \otimes \mathbf{K}_2 = \widehat{\mathbf{K}}_2\widehat{\mathbf{K}}_1$$

showing the hatted matrices over different spaces commute under matrix multiplication even though Kronecker products do not. Notice that if space order is changed, so are the Kronecker products.

Taking $\widehat{\mathbf{K}}_2\widehat{\mathbf{K}}_1$ with ordering $< 2, 1 >$ gives us

$$\widehat{\mathbf{K}}_2\widehat{\mathbf{K}}_1 = (\mathbf{K}_2 \otimes \mathbf{I}_1)(\mathbf{I}_2 \otimes \mathbf{K}_1) \tag{2.30}$$

$$= (\mathbf{K}_2\mathbf{I}_2) \otimes (\mathbf{I}_1\mathbf{K}_1) \tag{2.31}$$

$$= \mathbf{K}_2 \otimes \mathbf{K}_1$$

and similarly for $\widehat{\mathbf{K}}_1\widehat{\mathbf{K}}_2$, so that

$$\widehat{\mathbf{K}}_2\widehat{\mathbf{K}}_1 = \mathbf{K}_2 \otimes \mathbf{K}_1 = \widehat{\mathbf{K}}_1\widehat{\mathbf{K}}_2$$

But clearly $\mathbf{K}_1 \otimes \mathbf{K}_2 \neq \mathbf{K}_2 \otimes \mathbf{K}_1$, in general. So we must state an ordering for any equation

involving both hats and $\otimes$, but not if only hats are used or only $\otimes$ is used.

Similarly, the Kronecker product of two row-vectors $\mathbf{v}_1$ and $\mathbf{v}_2$ over spaces 1 and 2 respectively is defined as a vector

$$\mathbf{v}_1 \otimes \mathbf{v}_2 = [(\mathbf{v}_1)_1\mathbf{v}_2 \ldots (\mathbf{v}_1)_{n_1}\mathbf{v}_2]$$

where $\mathbf{v}_1$ is of dimension $n_1$ and $\mathbf{v}_2$ is of dimension $n_2$.

We define $\widehat{\mathbf{v}}_1$ under space order $< 1, 2 >$ as

$$\widehat{\mathbf{v}}_1 = \mathbf{v}_1 \otimes \mathbf{I}_2 = [(\mathbf{v}_1)_1\mathbf{I}_2 \ldots (\mathbf{v}_1)_{n_1}\mathbf{I}_2] \tag{2.32}$$

where $\mathbf{I}_2$ is an $n_2 \times n_2$ identity matrix and $\widehat{\mathbf{v}}_1$ is an $n_2 \times n_1 n_2$ matrix. Similarly, for $\widehat{\mathbf{v}}_2$

$$\begin{aligned} \widehat{\mathbf{v}}_2 &= \mathbf{I}_1 \otimes \mathbf{v}_2 \\ &= \begin{bmatrix} \mathbf{v}_2 & & & \\ & \mathbf{v}_2 & & \\ & & \ddots & \\ & & & \mathbf{v}_2 \end{bmatrix} \end{aligned} \tag{2.33}$$

an $n_1 \times n_1 n_2$ matrix.

Then in space-order $< 1, 2 >$,

$$\begin{aligned} \mathbf{v}_1\widehat{\mathbf{v}}_2 &= \mathbf{v}_1(\mathbf{I}_1 \otimes \mathbf{v}_2) \\ &= (\mathbf{v}_1\mathbf{I}_1 \otimes \mathbf{v}_2) \\ &= \mathbf{v}_1 \otimes \mathbf{v}_2 \end{aligned} \tag{2.34}$$

and

$$\begin{aligned} \mathbf{v}_2\widehat{\mathbf{v}}_1 &= \mathbf{v}_2(\mathbf{v}_1 \otimes \mathbf{I}_2) \\ &= \mathbf{v}_1 \otimes \mathbf{v}_2 \end{aligned} \tag{2.35}$$

so that $\mathbf{v}_1\widehat{\mathbf{v}}_2 = \mathbf{v}_2\widehat{\mathbf{v}}_1 = \mathbf{v}_1 \otimes \mathbf{v}_2$.

In space-order $< 2, 1 >$, $\mathbf{v}_1\widehat{\mathbf{v}}_2 = \mathbf{v}_2\widehat{\mathbf{v}}_1 = \mathbf{v}_2 \otimes \mathbf{v}_1$. For N spaces in natural space-order, from the above equations $\mathbf{v}_1 \otimes \mathbf{v}_2 \otimes \ldots \otimes \mathbf{v}_n$=$\mathbf{v}_1\widehat{\mathbf{v}}_2 \ldots \widehat{\mathbf{v}}_n$ by associativity of Kronecker products and inductive application using the results of the above equations.

For the dependent matrix exponential description $\widehat{\mathbf{B}}$ and $\widehat{\mathbf{L}}$ operating in N independent spaces in natural order, the resultant matrices are given by the following,

$$\begin{aligned} \widehat{\mathbf{B}}_1 &= \mathbf{B_1} \otimes \mathbf{I_2} \otimes \mathbf{I_3} \ldots \otimes \mathbf{I_N} \\ \widehat{\mathbf{B}}_2 &= \mathbf{I_1} \otimes \mathbf{B_2} \otimes \mathbf{I_3} \ldots \otimes \mathbf{I_N} \\ \widehat{\mathbf{B}}_N &= \mathbf{I_1} \otimes \mathbf{I_2} \otimes \mathbf{I_3} \ldots \otimes \mathbf{B_N} \end{aligned} \tag{2.36}$$

and

$$\begin{aligned}
\widehat{\mathbf{L}}_1 &= \mathbf{L_1} \otimes \mathbf{I_2} \otimes \mathbf{I_3} \ldots \otimes \mathbf{I_N} \\
\widehat{\mathbf{L}}_2 &= \mathbf{I_1} \otimes \mathbf{L_2} \otimes \mathbf{I_3} \ldots \otimes \mathbf{I_N} \\
\widehat{\mathbf{L}}_N &= \mathbf{I_1} \otimes \mathbf{I_2} \otimes \mathbf{I_3} \ldots \otimes \mathbf{L}_N
\end{aligned} \tag{2.37}$$

Starting (row) vector $\mathbf{p}$:

$$\mathbf{p} = \mathbf{p}_1 \otimes \mathbf{p}_2 \ldots \otimes \mathbf{p}_n \tag{2.38}$$

$$= \mathbf{p}_1 \widehat{\mathbf{p}}_2 \ldots \widehat{\mathbf{p}}_n \tag{2.39}$$

$$\mathbf{p} = \mathbf{p}_1 \otimes \mathbf{p}_2 \tag{2.40}$$

$$= \mathbf{p}_1 \widehat{\mathbf{p}}_1 \tag{2.41}$$

Summing vector $\boldsymbol{\varepsilon}'$

$$\boldsymbol{\varepsilon}' = \boldsymbol{\varepsilon}'_1 \otimes \boldsymbol{\varepsilon}'_2 = \widehat{\boldsymbol{\varepsilon}'}_2 \boldsymbol{\varepsilon}'_1 = \widehat{\boldsymbol{\varepsilon}'}_3 \widehat{\boldsymbol{\varepsilon}'}_2 \boldsymbol{\varepsilon}'_1 \tag{2.42}$$

$$\boldsymbol{\varepsilon}' = \widehat{\boldsymbol{\varepsilon}'}_n \widehat{\boldsymbol{\varepsilon}'}_{n-1} \ldots \widehat{\boldsymbol{\varepsilon}'}_2 \boldsymbol{\varepsilon}'_1 \tag{2.43}$$

## 2.5 Modes and weak stability

Before we proceed any further we need to understand the difference between the weakly-stable and the unstable network. The following is an excerpt from Krishnaswamy's thesis

[8] which provides us with a clear idea about what modes are and its different types. The ethernet traffic can be adequately represented by a continuous time Markov chain (cpMc) model that captures the dependencies and burstiness observed in the traffic, and that this traffic model can be used to predict the queueing behavior of networking systems that handle this traffic.

Burstiness in traffic can be characterized by a modal cpMc. A modal cpMc can be defined as a cpMc whose state transitions are nearly completely decomposable, with decomposition classes each defining a particular arrival rate, inter-arrival time distribution and/or dependency. These classes are called modes because to a general observer the traffic pattern appears different whenever the nearly-decomposable process jumps from class to class. To appear bursty one or more modes must have a mean arrival rate which is significantly greater than the mean arrival rate taken over all modes.

The concept of weak-stability, discussed by Jelenkovich [5], suggests that with respect to traffic, a stable queueing system can be either weakly stable or strictly stable. When it is weakly stable the arrival rate of one or more modes exceed the service capacity, though the overall service rate is greater than the overall arrival rate. As the modal traffic moves from mode to mode, the system moves from stable to unstable state, even though the overall system is stable. On the other hand, when the service capacity is greater than the arrival rate of the fastest mode, the system is strictly stable. The Figure 2.1 explains the concept of weak stability. The modes with mean arrival rate higher than the mean service rate are unstable and the rest of the modes in the traffic are stable. As the service rate is adjusted upward, unstable modes can become stable.

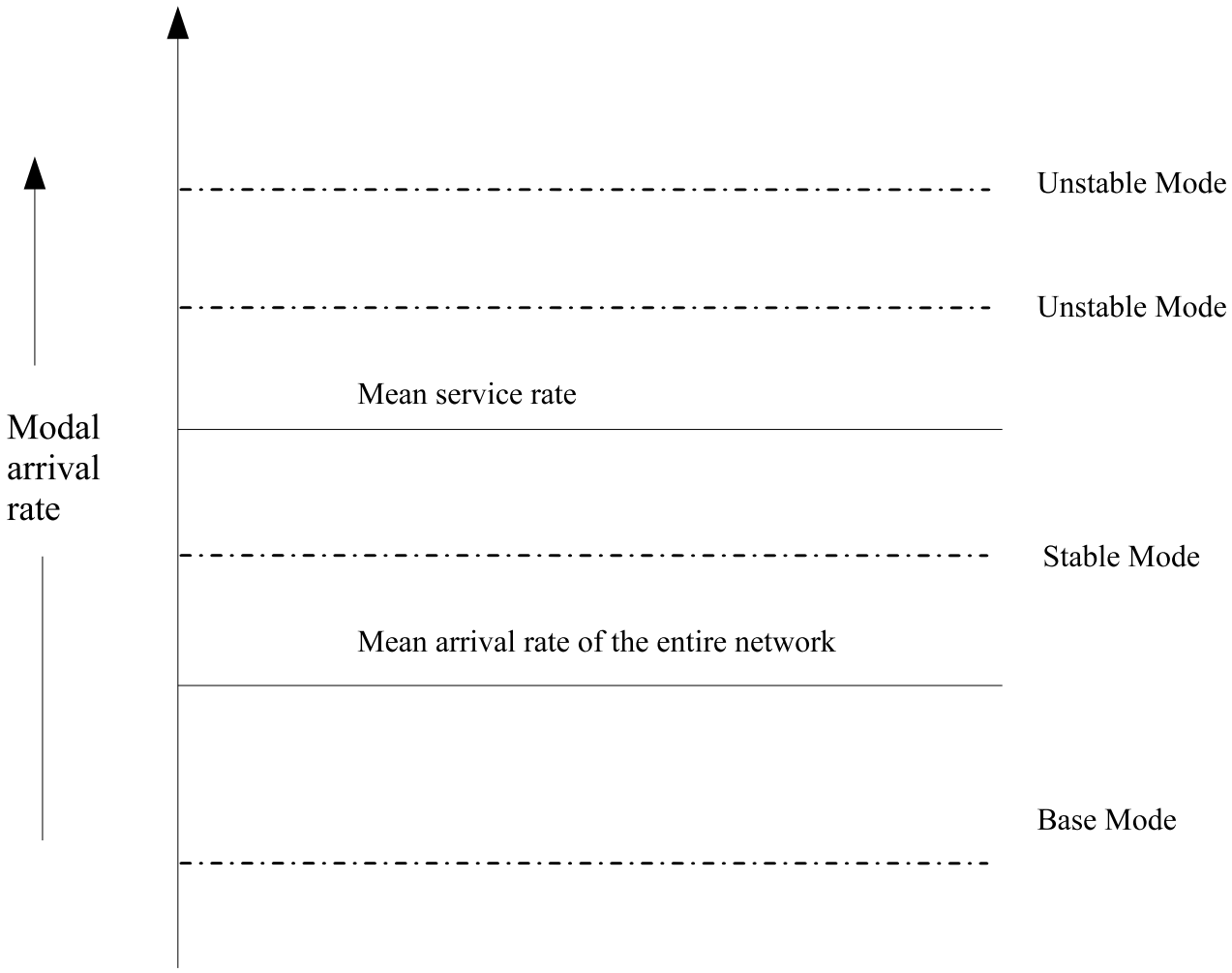



Figure 2.1: Modes in traffic

# Chapter 3

# Modelling Techniques

## 3.1 Introduction

In the earlier chapter we had discussed the Kronecker operations. We have discussed about the problems and the new solutions. The concept of an algebraic ME network is a new theoretical approach by Wallace which is a modernisation of a theory developed in [19][20] making a new use of the ME formulation of van-der Liefvoort [17]. In this chapter we will discuss the organization and setting up of the network. We need a tool that lets us look directly at the components of a general Markovian and ME network and how their behaviors combine to produce system behavior.

This section of this chapter gives a better understanding of the spaces involved with the construction of the network model. Following the conclusions of Krishnaswamy [8], each

access queue is, in general, represented by a multi-dimensional stochastic process consisting of five state spaces with varying degrees of dependence. These represent:

1. Mode Space - identifies modes the process is in.
2. Arrival Space - generates MED inter-arrival times within current mode.
3. Burst Duration Space - generates MED duration of mode before change.
4. Queue Space - identifies the number of packets in buffer queue.
5. Service Space - generates MED service time distribution.

These spaces undergo transformation and merging when the arrivals pass the access network and reach the backbone network. Our model builds on the queueing model employed by Krishnaswamy, extending it to a network model that allowed exploration of the interplay between a pair of access subnetworks and a backbone subnet. The earlier solution model followed by Krishnaswamy[8] could not incorporate the functionality of the merging traffic. This is because the mathematical procedure used to calculate the steady state vector did not account for the possibility of the merging of two traffic streams. A new solution was proposed by Dr.Victor Wallace and this thesis utilizes this new approach to effectively compute the steady-state vector values and the call loss probabilities.

Based on the properties of modes observed, the inter-arrival times during mode m, can be modelled using a suitable ME distribution. Even if the individual arrivals were exponentially distributed, when the arrival stream is considered with the different modes, it is autocorrelated over several time-scales.

The duration space represents the burst durations in a mode. The mode duration space is represented by the $< \mathbf{B}_d(m), \mathbf{M}_d(m), \mathbf{L}_d(m) >$ combination. $\mathbf{L}_d(m)$ represents the event transition matrix for the transition from mode m to base mode and $\mathbf{M}_d(m)$ represents the event transition matrix for the transition from base mode to mode m. The matrices become scalars if exponentially distributed durations are assumed.

The process in mode space is considered independent from the arrival space and the arrival mechanism is a function of the mode. This discovery following the findings of Krishnaswamy [8] were vital in the computation of the steady state vector $\pi$ values. This also greatly reduces the state space for computational purposes. The concept of independent modal arrival was incorporated from the very first step of construction of the matrices. This tool was effective in increasing the computation speed and reduces the amount of memory used in Maple during mathematical computations.

## 3.2 Network description

The **B** and **L** are generalized uses of the process rate matrix and event transition matrix. These matrices were introduced in chapter two. The **B** represents the changes in state which

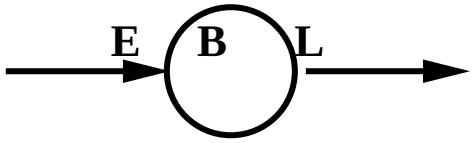



Figure 3.1: Generic representation

do not result in generation of an output event. The effect is *autogenous*,i.e. within the module. Its usually called the progress rate matrix.

The **L** represents changes in state which result simultaneously, in generation of an output event. The effect is *endogenous*, specific to the module affecting the world outside the module. Its usually called the event rate matrix. If there are multiple ports, there will be one **L** per output port.

The **E** represents changes in state resulting from input events as a result of the events from outside the module. The effect is *exogenous*. It is a new matrix introduced and is called the event transition matrix, denoted by a probability matrix. For every input port into a module, there will be one **E**.

We illustrate the use of these matrices by a very simple network, represented in Figure 3.2. We construct the network with the arrival and departure process modelled outside our

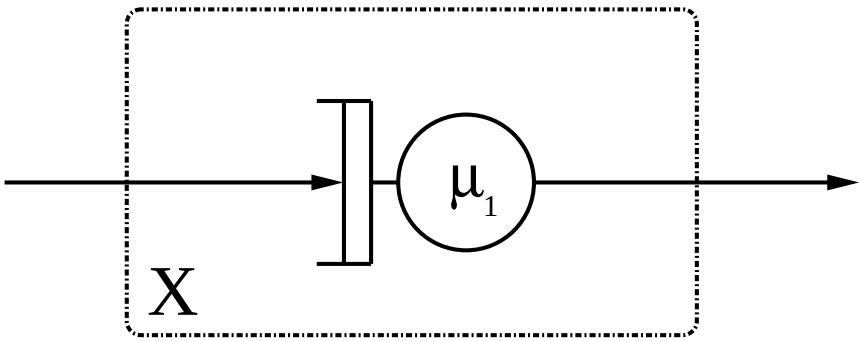



Figure 3.2: Simple Model

model. We assume that the queue length is infinite. The corresponding **B**, **L** and **E** matrices are generated below. $\lambda$ is the rate of arrivals into the system and $\mu_1$ is the service rate. We

get;

$$\mathbf{B}_X = \begin{bmatrix} 0 & & & \\ 0 & \mu_1 & & \\ & \ddots & \ddots & \end{bmatrix}$$

The departures are modelled as;

$$\mathbf{L}_X = \begin{bmatrix} 0 & & & \\ \mu_1 & 0 & & \\ & \ddots & \ddots & \end{bmatrix}$$

Notice that the effect of a departure event is to reduce the state. Therefore the $\mu$ appears in the lower diagonal.

The effect of the change in the system state on the adjoining state is shown as;

$$\mathbf{E}_X = \begin{bmatrix} 0 & 1 & & \\ 0 & 0 & 1 & \\ & \ddots & \ddots & \ddots \end{bmatrix}$$

Notice that the effect of an arrival event into the module is to increase the state, which it does with probability 1. Therefore the probability values are place in the upper diagonal.

Note that if the service intervals were ME distributed with $(\mathbf{B}_s, \mathbf{L}_s)$, the state space would be two dimensional and the $\mu$ elements would become matrices.

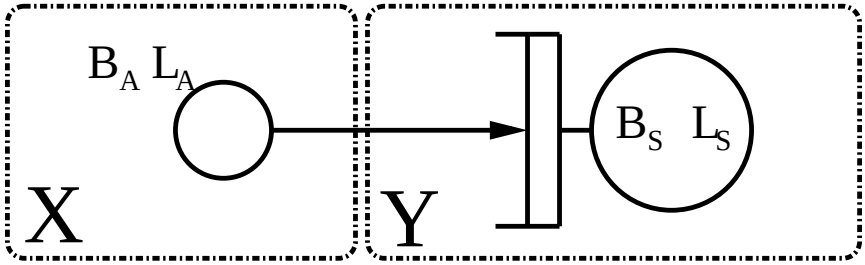



Figure 3.3: MEd arrivals and service

$$\mathbf{B} = \begin{bmatrix} 0 & 0 & & \\ 0 & \mathbf{B}_S & 0 & \\ 0 & & \mathbf{B}_S & \\ 0 & & \ddots & \ddots \end{bmatrix} \tag{3.1}$$

$$\mathbf{L} = \begin{bmatrix} 0 & & & \\ \mathbf{L}e_S' & 0 & & \\ & \mathbf{L}_S & & \\ & & \mathbf{L}_S & \end{bmatrix} \tag{3.2}$$

The exogenous matrix is defined as follows

$$\mathbf{E} = \begin{bmatrix} 0 & p & & & \\ 0 & 0 & \mathbf{I}_S & & \\ 0 & 0 & 0 & \mathbf{I}_S & \\ & & & & \ddots \end{bmatrix} \tag{3.3}$$

What we have seen so far is the step-by-step construction of the network module which is essential for network modelling. It can be seen that the fragments of the network are also

algebraic ME networks. Each fragment of the network has a state space $S$ described by $\mathbf{B}$, $\mathbf{L}$ and $\mathbf{E}$. These fragments of the network are also known as modules. Over the course of this thesis, the term modules would be used interchangeably with the fragment of the network, either the access network or the backbone.

If the queue were finite (lossy), the same algebra would apply except that

$$\mathbf{B}_q = \begin{bmatrix} 0 & & & \\ 0 & \mu & & \\ 0 & 0 & \mu & \\ 0 & 0 & 0 & \mu \end{bmatrix}, \mathbf{L}_q = \begin{bmatrix} 0 & & & \\ \mu & 0 & & \\ 0 & \mu & 0 & \\ 0 & 0 & \mu & 0 \end{bmatrix}, \mathbf{E}_q = \begin{bmatrix} 0 & 1 & 0 & \\ 0 & 0 & 1 & \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \end{bmatrix} \tag{3.4}$$

The $(n, n)^{th}$ element in the $\mathbf{E}$ matrix is 1.

$$\mathbf{B}_{aq} = \mathbf{B}_a \otimes \mathbf{I}_q + \mathbf{I}_a \otimes \mathbf{B}_q - \mathbf{L}_a \otimes \mathbf{E}_q = \begin{bmatrix} \lambda & -\lambda & & \\ 0 & \lambda+\mu & -\lambda & \\ 0 & 0 & \lambda+\mu & -\lambda \\ 0 & 0 & 0 & \mu \end{bmatrix}$$

and

$$\mathbf{L}_{aq} = \begin{bmatrix} 0 & & & \\ \mu & 0 & & \\ 0 & \mu & 0 & \\ 0 & 0 & \mu & 0 \end{bmatrix} \tag{3.5}$$

$\Rightarrow \mathbf{Q}_{aq} = \mathbf{L}_{aq} - \mathbf{B}_{aq}$.

## 3.3 Construction of a network

The following section aids in the description of the connection between the different modules to describe a larger module. A new joint space is created. The figure 3.4 is combination of

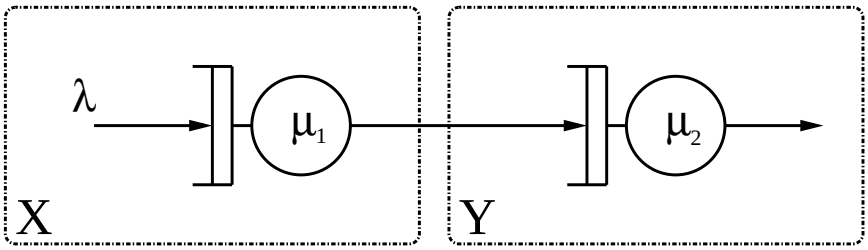



Figure 3.4: Simple Cascaded Network

two modules. The effect of the module X is seen on the module Y due to the link between them. The autogenous events in each of the modules extend into the joint space. Hence the *hatting* of each $\mathbf{B}$ matrix is done, $\widehat{\mathbf{B}}$. The endogenous transitions join with simultaneous exogenous events of connected module $\widehat{\mathbf{L}}$ and $\widehat{\mathbf{E}}$.

$$\mathbf{B}_{YX} = -(\widehat{\mathbf{B}}_Y + \widehat{\mathbf{B}}_X + \widehat{\mathbf{E}}_Y \times \widehat{\mathbf{L}}_X) \tag{3.6}$$

The above equation can be represented in Kronecker notation with the state space ordering of $< Y, X >$.

$$\mathbf{B}_{YX} = (\mathbf{B}_Y \otimes \mathbf{I}_X + \mathbf{I}_Y \otimes \mathbf{B}_X - \mathbf{E}_Y \otimes \mathbf{L}_X) \tag{3.7}$$

When the above equation is expanded to show the matrices obtained we get the following forms of matrices whose properties have already been discussed in chapter 2. While expanding the matrices, the ordering was taken into account for consistency.

$$\mathbf{I}_Y \otimes \mathbf{B}_X = \begin{bmatrix} \mathbf{B}_X & & \\ 0 & \mathbf{B}_X & \\ & \ddots & \ddots \end{bmatrix}$$

$$\mathbf{B}_Y \otimes \mathbf{I}_X = \begin{bmatrix} 0 & & \\ -\mu_2 I_x & \mu_2 I_x & \\ & \ddots & \ddots \end{bmatrix} \tag{3.8}$$

$$\mathbf{E}_Y \otimes \mathbf{L}_X = \begin{bmatrix} 0 & \mathbf{L}_X & \\ & 0 & \mathbf{L}_X \\ & & \ddots & \ddots \end{bmatrix}$$

What we have seen so far is the basic combination of two modules to better highlight the relationship between the autogenous, endogenous and exogenous event rate transition matrices.

We have a simple connection described in the figure 3.5. It is comprised of two modules $X$ and $Y$ making up the network module $XY$. Each module $X$ and $Y$ have the event transitions occurring between themselves and $XY$ module has external events associated with it. An input to the $XY$ module is also an input event to the $X$ module. The $XY$ module has an external exogenous event entering it which causes transitions in the state of

module $X$ described by exo-event matrix, $E_X$; and an external event exiting the system is represented by the departure event matrix, the endo-event matrix, $L_Y$ as seen in figure 3.5. The event matrices are calculated as follows The autogenous matrix of the $XY$ module, $\mathbf{B}_{XY}$

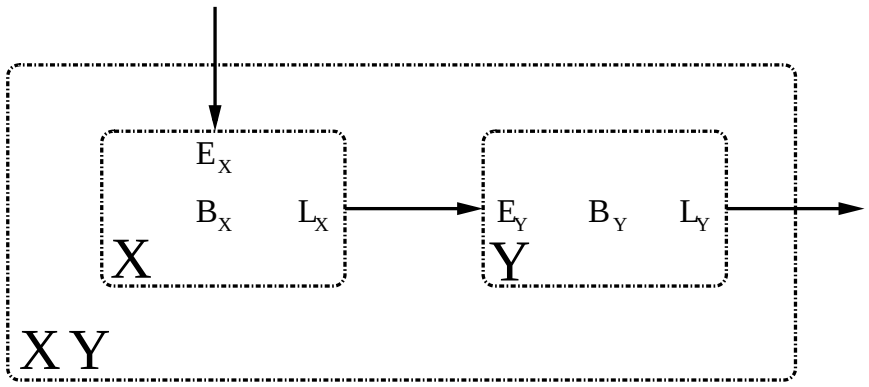



Figure 3.5: A simple connection

is obtained by performing a Kronecker sum "$\oplus$" between the two autogenous matrices of the $X$ and $Y$ modules, $\mathbf{B}_X$ and $\mathbf{B}_Y$. The $\mathbf{B}_{XY}$ matrix also takes into account the interaction between the two modules . The Kronecker product between $\mathbf{L}_X$ and $\mathbf{E}_Y$ is a result of the interaction.

$$\mathbf{B}_{XY} = (\mathbf{B}_X \otimes \mathbf{I}_Y) + (\mathbf{I}_X \otimes \mathbf{B}_Y) - (\mathbf{L}_X \otimes \mathbf{E}_Y)$$

The exogenous event matrix $\mathbf{E}_{XY}$ is the Kronecker product of the $\mathbf{E}_X$ and $\mathbf{I}_Y$. This is the external event exiting the system with-respect-to the $X$ module and entering the $Y$ module.

$$\mathbf{E}_{XY} = \mathbf{E}_X \otimes \mathbf{I}_Y$$

The network module as such would have an overall event that produces a change in the neighboring module (if there was one ) namely the endogenous event matrix. It can be

either absorbed into the module or shown to exit the system as the $\mathbf{L}_{XY}$ matrix.

$$\mathbf{L}_{XY} = \mathbf{I}_X \otimes \mathbf{L}_Y$$

The network of modules discussed above could be considered the platform over which various complex networks can be constructed. The combination of modules can be further extended as follows. The figure 3.6 is a simple branched network. The output from module X gets

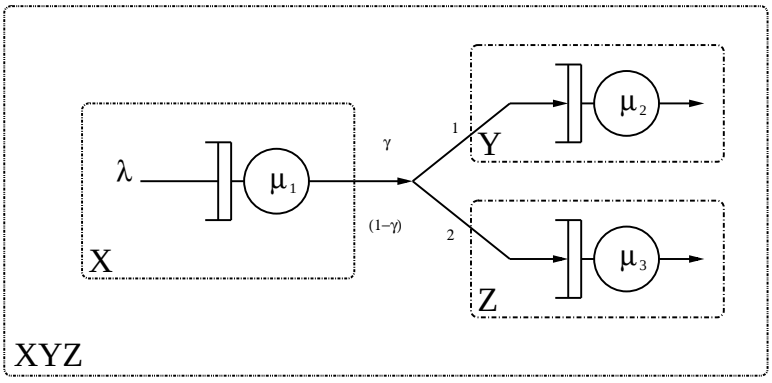



Figure 3.6: Simple Branched Network

distributed with a certain probability amongst the Y and Z modules. If the arrival and service rates are as indicated in the figure, we can write the autogenous, exogenous and endogenous matrices.

$$\mathbf{B}_X = \begin{bmatrix} \lambda & -\lambda & & \\ 0 & \lambda+\mu_1 & -\lambda & \\ & \ddots & \ddots & \ddots \end{bmatrix} \tag{3.9}$$

$$\mathbf{B}_Y = \begin{bmatrix} 0 & & \\ -\mu_2 & \mu_2 & \\ & \ddots & \ddots \end{bmatrix} \tag{3.10}$$

$$\mathbf{B}_Z = \begin{bmatrix} 0 & & \\ -\mu_3 & \mu_3 & \\ & \ddots & \ddots \end{bmatrix} \tag{3.11}$$

The main purpose of introducing this combination of modules is to understand the behavior of the **L** and the **E** modules. It can be seen there, since there is one output port from module X we have just one **L** but for every input port into the module Y and Z there is a **E**.

$$\mathbf{L}_X = \begin{bmatrix} 0 & & \\ \mu_1 & 0 & \\ & \ddots & \ddots \end{bmatrix} \tag{3.12}$$

$$\mathbf{E}_Y = \begin{bmatrix} 0 & 1 & & \\ 0 & 0 & 1 & \\ & & \ddots & \ddots \end{bmatrix} \tag{3.13}$$

$$\mathbf{E}_Z = \begin{bmatrix} 0 & 1 & & \\ 0 & 0 & 1 & \\ & & \ddots & \ddots \end{bmatrix} \tag{3.14}$$

Combining the event rate matrices we get the **B** for the entire network.

$$\mathbf{B}_{XYZ} = \mathbf{B}_X \oplus \mathbf{B}_Y \oplus \mathbf{B}_Z - [\gamma \mathbf{L}_X \otimes \mathbf{E}_Y \otimes \mathbf{I}_Z + (1-\gamma)\mathbf{L}_X \otimes \mathbf{I}_Y \otimes \mathbf{E}_Z]$$

It can also be represented in Hat Space notation.

$$\mathbf{B}_{XYZ} \;=\; \widehat{\mathbf{B}}_X + \widehat{\mathbf{B}}_Y + \widehat{\mathbf{B}}_Z - [\gamma \widehat{\mathbf{L}}_X \widehat{\mathbf{E}}_Y + (1-\gamma)\widehat{\mathbf{L}}_X \widehat{\mathbf{E}}_Z]$$

The figure 3.7 describes the conceptual framework of the modules to form a network. The basic methodology followed is highlighted below.

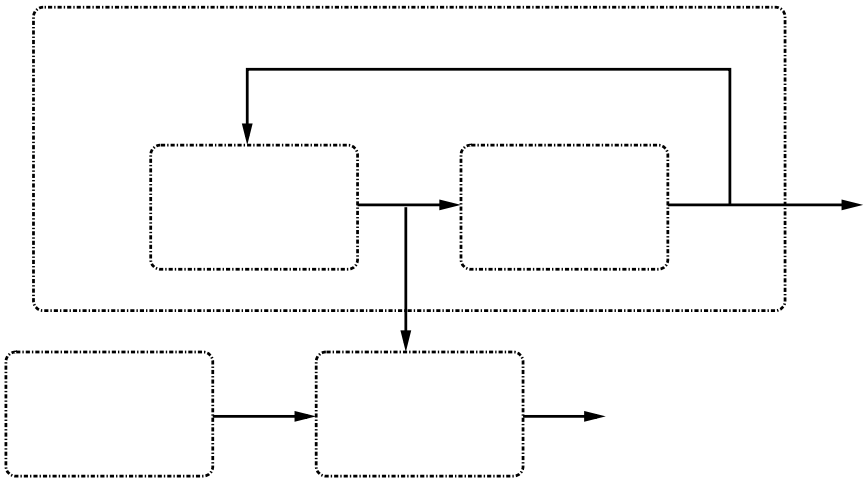


Figure 3.7: Network Framework

- Modules + Directed Connections = network.
- A new module is formed by rolling together the old modules and the connections between them.
- Repeat until there is only one module.
- The resulting unique module describes the infinitesimal generator of the network's Markovian or pseudo-Markovian model.

Similarly, the arrival process at any directed connection can be determined by combining modules to produce a source module for the connection. The combination of modules is a strictly algebraic manipulation of module-descriptive relations.

## 3.4 Solving MED networks

Once the components of the ME network have been developed, the next step is to solve the algebraic ME networks. The computation of the infinitesimal generator for Markov or ME chain is denoted by $\mathbf{Q}$. From the above figure 3.4 we get $\mathbf{Q}$ matrix as

$$\mathbf{Q} = \mathbf{L} - \mathbf{B}$$

$$\mathbf{Q} = \begin{bmatrix} -\mathbf{B}_X & \mathbf{L}_X & & & \\ \mu_2\mathbf{I}_X & -(\mathbf{B}_X + \mu_2\mathbf{I}_X) & \mathbf{L}_X & & \\ 0 & \mu_2\mathbf{I}_X & -(\mathbf{B}_X + \mu_2\mathbf{I}_X) & \mathbf{L}_X & \\ & \ddots & \ddots & \ddots & \ddots \end{bmatrix} \tag{3.15}$$

We expand the Q matrix to get

$$\mathbf{Q} = \begin{bmatrix} -\lambda & \lambda & 0 & \cdots & 0 & & & & & \\ 0 & -[\lambda+\mu_1] & \lambda & \cdots & \mu_1 & 0 & & & & \\ 0 & 0 & -[\lambda+\mu_1] & \ddots & 0 & \mu_1 & 0 & & & \\ 0 & 0 & 0 & \ddots & 0 & 0 & \ddots & \ddots & & \\ \mu_2 & 0 & 0 & 0 & -[\lambda+\mu_2] & \lambda & 0 & \cdots & 0 & \\ 0 & \mu_2 & 0 & 0 & 0 & -[\lambda+\mu_1+\mu_2] & \lambda & \cdots & \mu_1 & 0 \end{bmatrix}$$

Examples from the connections described above show that $\pi_m \mathbf{Q}_m = 0$.

## 3.5 Network Model

In the previous sections we have discussed how to construct the basic building blocks of the network. We have also discussed the mathematics involved in explaining the basic models. Now lets see how the final model looks. This leads into the modeling of backbone behavior. The same rules for combining modules are applied to the backbone. The traffic entering the backbone is a combined result of the merging of the traffic exiting out of the access networks.

We have, as we can see, two access networks (for simplicity and insight purposes) getting two independent streams of traffic merging into a backbone network. Each module has a buffer with a predetermined queue length and a server with calculated service rate values. Independent of the arrival modules, we have the 'modal' module that has the distribution of

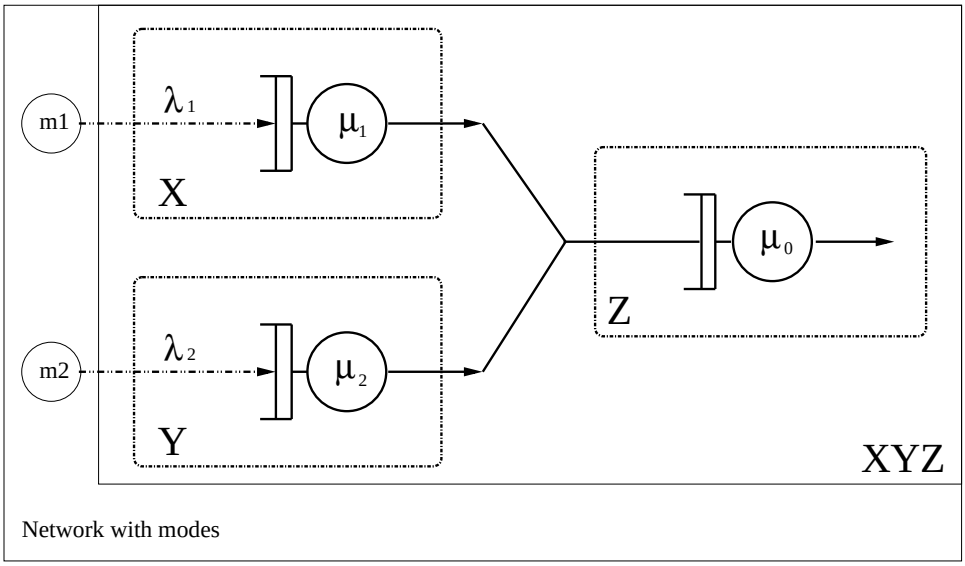



Figure 3.8: The network diagram with 2 modules, each with modal traffic, merging into a backbone.

the modes for the different arrivals that enter each node of the network (access networks).

The arrival mechanism is a function of the mode.

# Chapter 4

# Modeling Backbone Traffic

## 4.1 Introduction

Using our algebraic approach, this chapter deals with the construction of a matrix exponential multi-modal queueing model to calculate the merging onto the backbone of different arrival streams of dependent traffic. The buffer and service rates can be varied sufficiently so that a relationship between the call loss probability and a measure of burstiness called the burst-to-base ratio.

To gain insight into the way in which bursty arrivals affect an entire network, we need to reduce the number of parameters, while retaining the main elements of dependency. Our model consists of three modal queues, representing two access network queues and a backbone. We view the entire effect of a local network of users as a single queue with bursty

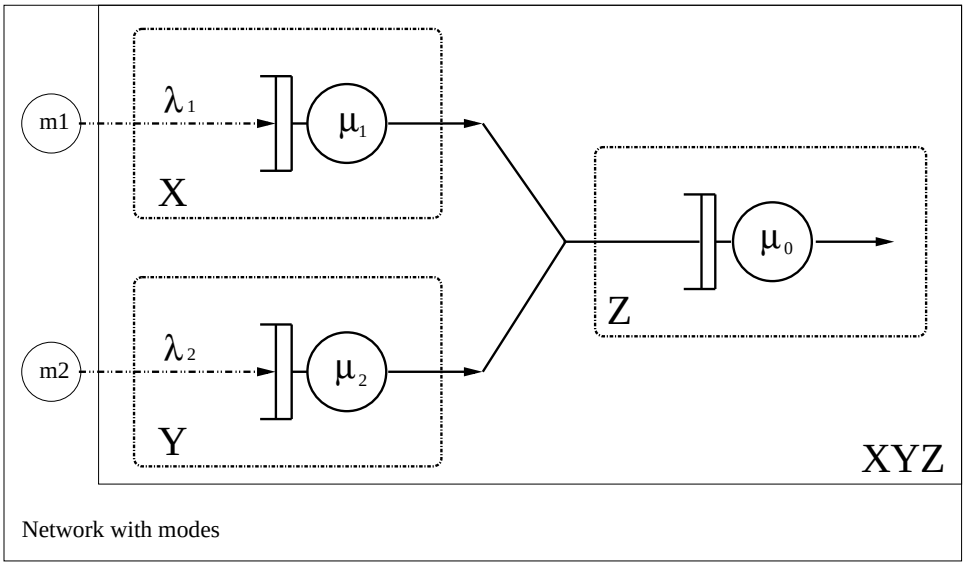



Figure 4.1: The network diagram with 2 modules merging into a backbone.

arrivals, which we call an *access network*. The access network has incoming traffic that is exponential per mode but multi-modal. We view the internetwork joining access networks as a single queue serving a stream of traffic formed by merging the outputs of the access networks. We call this part of the model the *backbone network*. The output traffic from the access networks is merged to become the input traffic to the backbone network.

The merging of the traffic can increase as the square of the number of modes. However, if bursts are rare, as they are in our case, then most combined modes are highly unlikely and growth of the number of modes can be linear, or even fixed. We will analyze this network, shown schematically in Figure 4.1, in order to see how the long-term dependence induced by the bursty access streams affect the performance on the backbone and overall.

It was not part of the goals of this research to study computational efficiencies re-

sulting from the NCD or QBD character of these modes. Such algebraic tools are likely to have an enormous impact on the tractability and richness of more complex models than are treated here. As a result we simplify the state structure by a number of assumptions which enhance analytical tractability and reduce the number of parameters to be considered. We make the queue capacities finite and small, and we reduce the number of modes to the most important ones. We also make the per-mode arrivals exponential, and service times exponential and independent. Nevertheless, the model will be interesting and give a clear idea of the potentials of the methodology.

Making the per-mode inter-arrival times simple and independent still leaves the entire input traffic model multi-time-scale and strongly dependent over an extended time frame. We can see from Figure 4.1 that we have applied the rules discussed in chapter three to construct this network. We can see the effects on the backbone network due to the events that leave the access networks. The merging of the access networks is clearly shown here. This model gives a scaled down insight into what one can expect to see in real time networks.

## 4.2 Mathematical solution

The modes shift from base mode to one of the burst modes, and then back. Let $\overline{\delta_0}$ be the infinitesimal rate of occurrence of the shift from base mode to any of the burst modes, and $\overline{\delta_i}$, $i = 1, 2, 3$ be the rate of occurrence of the shift from base to burst mode $i$. This $\overline{\delta_0} = \overline{\delta_1} + \overline{\delta_2} + \overline{\delta_3}$. Let $\underline{\delta_i}$, $i = 1, 2, 3$ be the rate of occurrence of a shift from mode $i$ to the base mode, so $1/\underline{\delta_i}$ is the mean sojurn time in a burst mode $i$ and $1/\overline{\delta_0}$ is the mean sojurn time

in base mode. The process description of the mode-space module $< \mathbf{B}_m, \mathbf{L}_m >$ is thus given by

$$\mathbf{B}_m = \begin{bmatrix} \overline{\delta_0} & 0 & 0 & 0 \\ 0 & \underline{\delta_1} & 0 & 0 \\ 0 & 0 & \underline{\delta_2} & 0 \\ 0 & 0 & 0 & \underline{\delta_3} \end{bmatrix}$$

$$\mathbf{L}_m = \begin{bmatrix} 0 & \overline{\delta_1} & \overline{\delta_2} & \overline{\delta_3} \\ \underline{\delta_1} & 0 & 0 & 0 \\ \underline{\delta_2} & 0 & 0 & 0 \\ \underline{\delta_3} & 0 & 0 & 0 \end{bmatrix}$$

The arrival stream entering the backbone is in reality the departure streams from the access networks. Hence, the modes presented to the backbone come from the access modes. Since we are predetermining the service rate of the access networks, it is observed that the arrival rates are reduced in the backbone due to buffer overflows in access. The effects of modal traffic has been discussed in both chapter two and three. The effects of modes is critical in our discussion and their importance in controlling the behavior of the network cannot be overstated. The infinitesimal generator $\mathbf{Q}_m$ can be calculated by the formula :

$$\mathbf{Q}_m = \mathbf{L}_m - \mathbf{B}_m$$

Having calculated the $\mathbf{Q}_m$ we can find the steady state probability $\pi_m$ as follows: $\pi_m \mathbf{Q}_m = 0$.

We can also notice that this network fits the different spaces discussed in Chapter 3.

Each access network has the arrival space, duration space and the mode space. We can see that the mode space of the backbone is a combination of the modes of the arrival space. The backbone has the queue space and service space and its arrival space is determined by its inputs. For our model the combination of the entire arrival streams into each of the access networks merging into the backbone signifies the flexibility and power of our modelling. We want to focus our model on the degree of burst traffic, not on its finer characteristics.

The arrival rate and the transition rate matrices for the access networks can be defined as follows. For insight purposes we assign the arrival rate for each of the access networks as $\mathbf{B}_a = \lambda$ and the transition event rate as $\mathbf{L}_a = \lambda$.

The event descriptors for the access network's queues are:

$$\mathbf{B}_q = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & \mu & 0 & 0 \\ 0 & 0 & \mu & 0 \\ 0 & 0 & 0 & \mu \end{bmatrix}, \mathbf{L}_q = \begin{bmatrix} 0 & 0 & 0 & 0 \\ \mu & 0 & 0 & 0 \\ 0 & \mu & 0 & 0 \\ 0 & 0 & \mu & 0 \end{bmatrix}, \mathbf{E}_q = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

For a finite buffer space, notice the $(n, n)^{th}$ element which is 1, as discussed in chapter three about the finite queues with overflows. Utilizing the concepts dealt with in the chapter three

in section 3.2, we can construct the $\mathbf{B}$, $\mathbf{L}$ and $\mathbf{E}$ for each of the modules $X$, $Y$ and $Z$ as

$$\mathbf{B}_X = \begin{bmatrix} 0 & & & \\ 0 & \mu & & \\ 0 & 0 & \mu & \\ 0 & 0 & 0 & \mu \end{bmatrix}, \mathbf{L}_X = \begin{bmatrix} 0 & & & \\ \mu & 0 & & \\ 0 & \mu & 0 & \\ 0 & 0 & \mu & 0 \end{bmatrix}, \mathbf{E}_X = \begin{bmatrix} 0 & 1 & 0 & \\ 0 & 0 & 1 & \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \end{bmatrix} \tag{4.1}$$

$$\mathbf{B}_Y = \begin{bmatrix} 0 & & & \\ 0 & \mu & & \\ 0 & 0 & \mu & \\ 0 & 0 & 0 & \mu \end{bmatrix}, \mathbf{L}_Y = \begin{bmatrix} 0 & & & \\ \mu & 0 & & \\ 0 & \mu & 0 & \\ 0 & 0 & \mu & 0 \end{bmatrix}, \mathbf{E}_Y = \begin{bmatrix} 0 & 1 & 0 & \\ 0 & 0 & 1 & \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \end{bmatrix} \tag{4.2}$$

Applying the concepts of the Kronecker sums and products and the hat space as discussed earlier in chapter two, we can combine the three spaces; (two Access network spaces $X, Y$ and one Backbone $Z$).The combination and the interaction of the spaces forms the network that interacts with the mode space on a different plane.

The overall arrival rate matrix after merging of the two access networks with the backbone is:

$$\mathbf{B}_{XYZ} = \widehat{\mathbf{B}}_X + \widehat{\mathbf{B}}_Y + \widehat{\mathbf{B}}_Z - (\widehat{\mathbf{L}}_X + \widehat{\mathbf{L}}_Y)\widehat{\mathbf{E}}_Z \tag{4.3}$$

The event rate matrix for the entire network is

$$\mathbf{L}_{XYZ} = \widehat{\mathbf{L}}_Z \tag{4.4}$$

Hence the infinitesimal generator matrix

$$\mathbf{Q}_{XYZ} = \mathbf{L}_{XYZ} - \mathbf{B}_{XYZ} \tag{4.5}$$

If there are two modes for each of the access networks, then it's observed that there would be four modes in the joint mode space. The above set of equations needs to be calculated four times for every single combination of free parameters. The $\mathbf{Q}_{XYZ}$ indicated above is the generic representation of the $\mathbf{Q}$ matrix. When the $\mathbf{Q}$ is to be represented as a function of the modes, then it is indicated as $\mathbf{Q}_{XYZ}(m_1, m_2)$ where $m_1$ and $m_2$ are the modes of the respective access networks. Then the steady state probability vector in the space $XYZ$, $\pi_{XYZ}$ can be calculated as the solution vector for $\pi_{XYZ}$

$$\pi_{XYZ}\mathbf{Q}_{XYZ} = 0 \tag{4.6}$$

It has been noted that, though the modes could have been combined with the $X, Y$ and $Z$ space to make $MXYZ$ space, it has been determined that we can separate the modes at the very beginning of the calculation. The steady state vector of the modes is calculated. It is then combined with the steady state probability vector of each of the four combination of modes. This improves the efficiency of the calculation.

Since the arrival rates are a function of the modes, the final state probability, $\pi$, is given by

$$\pi_0 = \pi_m[1]\pi_{XY}(1, 1)$$

where $\pi_{XY}(1,1)$ indicates the $\pi$ for the combination of mode 1 from module $X$ and $Y$. On similar lines we can write

$$\pi_1 = \pi_m[2]\pi_{XY}(1,2)$$

$$\pi_2 = \pi_m[3]\pi_{XY}(2,1) \tag{4.7}$$

$$\pi_3 = \pi_m[4]\pi_{XY}(2,2) \tag{4.8}$$

## 4.3 Calculation of the results

The objective now is to exercise our model. We will critique the performance calculations of the model, comparing the model behavior with intuitive expectations. To perform this comparison we would have to first take a look at the variables that are used for computation, which can be categorized into free variables and dependent variables.

The independent variables are those that would typically be statistically measured as the environment of an operating network – such variables as:

1. Buffer capacity of access networks.
2. Burst arrival rate in the access network.
3. Burst duration in the access network.
4. Base arrival rate in the access network.

5. Base duration in the access network.

6. Backbone/access network service rate

The dependent variables are those that would typically be observed as measurements of performance – such variables are

1. Network call loss probability

2. Backbone call loss probability

3. Throughput

4. output stream distributions

5. output stream auto-covariances

For presentation, the parameters are chosen more for insight and are often ratios rather than absolute value. In our case, we will concentrate on call loss probability as a function of certain derived parameters. We will hold the burst arrival rates and burst durations at observed values derived from the Bellcore October trace data as reported in [8]. We will also hold the buffer capacities fixed at a value that will not cause our state space to become too large. The access traffic intensity – the ratio of mean arrival rate to the mean service rate, whose values are varied from 0.2 to 0.9 signifying the high packet loss in the access networks, and the base arrival rates and mean base duration are chosen as our primary parameters for analysis.

The backbone traffic intensity is calculated based on the burst and base modes seen in the Bellcore traffic. The service rate of the backbone, as a design choice, is chosen to be of value in the range of the individual service rate of the access network and the sum of the service rates of the access networks. The backbone traffic intensity is the ratio of mean service rate of the access network and the mean service rate of the backbone. Once these values are obtained, the mean arrival rate for the access network can be calculated as follows.

$$\lambda_{mean} = \frac{burst\ arrival\ rate * burst\ duration + base\ arrival\ rate * base\ duration}{base\ duration + burst\ duration} \tag{4.9}$$

Once the value of the $\lambda_{mean}$ and $\rho$, the traffic ratio of the access networks, are known, the service rate of the access network, $\mu$, can be computed from the equation.

$$\begin{aligned} \rho &= \frac{\lambda_{mean}}{\mu} \\ \Rightarrow \mu &= \frac{\lambda_{mean}}{\rho} \end{aligned} \tag{4.10}$$

Any change in the arrival or mode duration affects the overall $\mu$, if we want to keep a specific $\rho$ value. The above set of values is associated with the access networks. One of the positive outcomes of this model is the facility to compute the call-loss probability (CLP) with relative ease. The computation of CLP is associated with the modification of the burst to base traffic ratio (BBTR).

The main reason for choosing this derived parameter (BBTR) as a measure is because of the fact that BBTR signifies the bursty nature of the arrivals entering the network. Its a ratio describing the relationship between the bursty traffic and the base traffic. The traffic

characteristics are determined by the amount of traffic entering the network in the given interval of time. Long burst with small number of arrivals during that burst is much more tolerable than short bursts with a large number of arrivals during that burst periods. In this thesis, we have defined the burstiness of traffic as observed by Krishnaswamy [8] in the Bellcore October traffic trace. We are maintaining the burst stability and modifying the BBTR by varying the base arrival rate. By maintaining the burst stability we determine the nature of the burst arrivals in the given interval of time.

Since the call loss probability is to be computed, the variations in burst to base ratio is taken into account. The burst to base traffic ratio ($BBTR$) is defined as follows:

$$BBTR = \frac{burst\ duration * burst\ arrival\ rate}{mean\ arrival\ rate * burst\ duration + base\ duration} \tag{4.11}$$

$$= \frac{burst\ duration * burst\ arrival\ rate}{mean\ arrival\ rate * cycle\ time} \tag{4.12}$$

$$= \frac{\lambda(0)/\overline{\delta_0}}{\lambda_{mean} * cycle\ time} \tag{4.13}$$

The BBTR is the ratio of the packets in burst to the total packets. To increase the burst to base traffic ratio, do one or more of the following

1. decrease $\lambda_{mean}$
2. decrease cycle time
3. increase burst duration or burst arrival

It has been observed that varying the base arrival rate has the most significant effect on the

ratio. As the burst to base ratio increases, it is noted that the overall CLP increases. The overall CLP is calculated from the sum of the individual CLP at the access network and the CLP at the backbone. Approximate value range for the service rate of the backbone server is greater than the service rate of the individual access network and much less than the total service rates. The CLP is calculated by the formula [11]

$$CLP = \frac{\pi_n \mathbf{L}_a \mathbf{e}'}{\sum_i \pi_i \mathbf{L}_a \mathbf{e}'} \tag{4.14}$$

The CLP is the probability of arrival occurring while the buffer is full. It is also referred as the fraction of arrivals that find the buffer full.

# Chapter 5

# Results and insights

## 5.1 Introduction

We have constructed the network consisting of the access networks and the backbone. We have seen the traffic exiting from the individual access networks merging as they enter the backbone. With all the efforts spent in looking at the access networks, the importance of the backbone network in maintaining the overall stability of the network is overlooked. This is what makes it challenging and interesting to look at. The behavior of the backbone is governed by the service rate of the access networks and also by the burst characteristics of the traffic entering the access networks, whose effect is seen in the arrival to the backbone. We can expect to see the backbone behave, to an extent, like the access network when it is fed with bursty traffic filtered through the access networks. From the service rate that is chosen for the backbone network, we can see the weak-stability characteristic appear when

the bursty traffic reaches the backbone. We can also notice the variations in the CLP in the backbone as the service rate of the access and the service rate of the backbone are changed. We can expect the access network to behave more erratically as the burstiness of the traffic into it is increased. This effect coupled with the effects resulting from the changes in the time scale of the burst arrival traffic, the changes in the duration of the bursts and how fast the arrivals occur during a burst, govern the modelling of the backbone. The queue capacity is taken as the point beyond which the arrivals entering the queue are considered to be lost. These contribute to the packet loss of the network.

What we propose to develop is a measure of burstiness for streams to use as a primary parameter for exploration. We shall exercise the model described in the section 4.5 from chapter 4 using Maple. The main purpose is to gather insights of the CLP as function of burstiness, under various parameter values.

## 5.2 Experimental setup

Ours is an insight generating model. From chapter four, we can notice the parameters that are considered independent and those that are dependent. We have restricted ourselves to a fixed queue capacity. Another important decision on the selection of modes was done. We chose to have two modes in each of the access network's arrival stream, which when merged would give a total of four modes entering a backbone. More modes would make it harder to focus on the behavior of bursts. There is one burst mode and one base mode per access network. We have also assigned the buffer capacity to 13 due to memory used by Maple for

computation. The amount of memory used in Maple for computation is dependent on the size of the matrices computed. For instance, in this case, there are three modules with buffer capacity of 13 each and there are two modes in the access networks for each access network. When *hatted*, we would find the memory used for the matrix of size

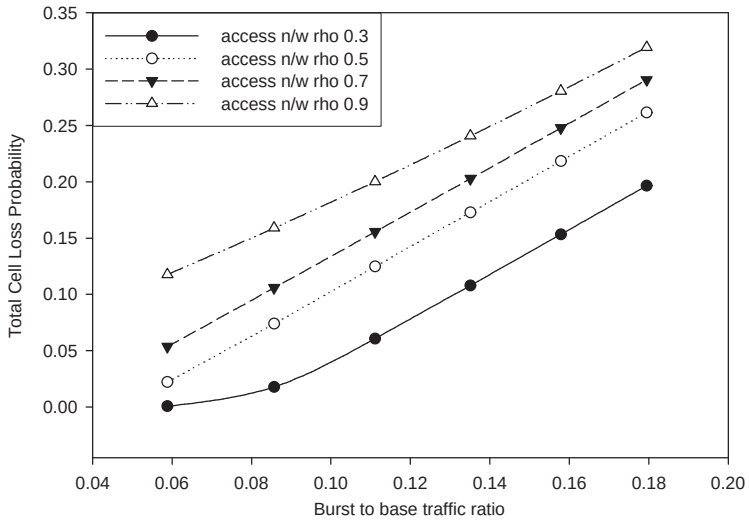
$$Size\ of\ Matrix = (13^3 \times 2^2) \times (13^3 \times 2^2)$$

$$= 8788 \times 8788$$

As can be seen from the above calculation, the total number of elements under consideration for a simple insight generating model is equal to 77228944. Such is the increase in the state space.

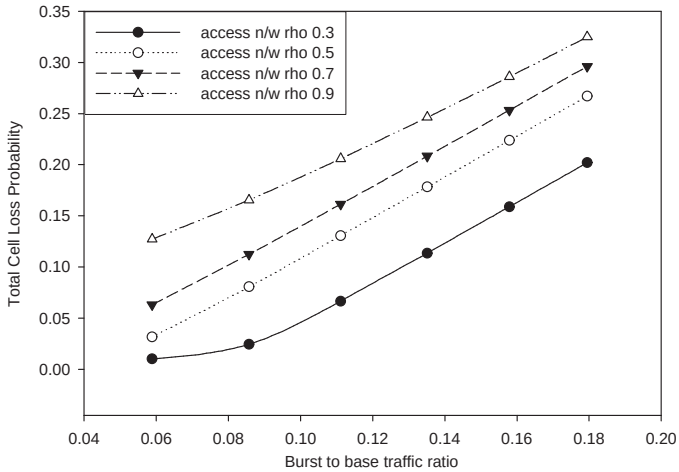
## 5.3 Case 1

We have modelled and analyzed a network topology. We were successful in representing the merging of arrival streams into the backbone, from more than one access network, mathematically. We have plotted eye-catching graphs from the results we got, running the calculations in Maple. But, now the question that bothers us all is, do these values represent anything worthwhile? Our first experiment was to run a sanity check on our model. Intuitively, one would expect the total CLP to increase as the BBTR of the arrivals increases. General knowledge dictates that as the burstiness, as measured by BBTR, increases at the access network, the CLP would increase. BBTR here refers to the ratio of the burst traffic to the




The service rate of the backbone (muz) is fixed to 3333.33 (pkts/sec).
The rho for the access network is varied from 0.3 to 0.9.
Notice the increase in the total CLP with the increase in the BBTR.

Figure 5.1: BB $rho = 0.3$ BBTR vs CLP




The service rate of the backbone (muz) is fixed to 1666.66 (pkts/sec). The rho for the access network is varied from 0.3 to 0.9. Notice the increase in the total CLP with the increase in the BBTR.

Figure 5.2: BB $rho = 0.6$ BBTR vs CLP

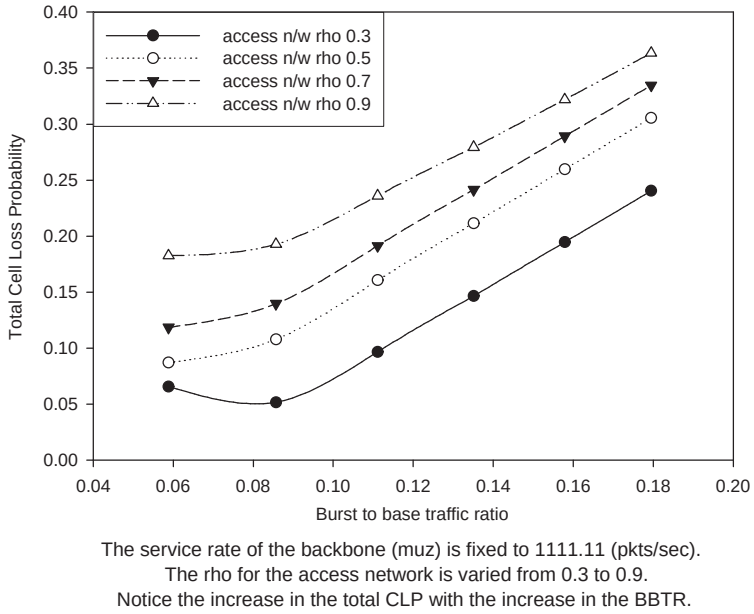



Figure 5.3: BB $rho = 0.9$ BBTR vs CLP

base traffic in an interval of time. In order for us to compute the total CLP, we maintain control over the backbone service rate. Hence, for a given backbone service rate, the CLP of the access network and the backbone is calculated for various traffic intensities of the access networks. The total CLP is taken as the sum of the access network CLP and backbone CLP.

It is observed that, the increase in CLP as the BBTR increases is as expected as seen from the graphs 5.1, 5.2 and 5.3. Also noted, is the overall increase of the CLP in the access network with higher traffic ratio in the access network. Notice the shift in the curves in the different graphs which indicate the increase in the values of the access network CLP as the $\rho$ of the access network increases. These observations are a clear indication that our model is working as expected.

## 5.4 Case 2

Having satisfied the basic requirements for a sanity check, the next step was to take a look at the elements that are the building blocks of this thesis, the access and backbone network CLP. We have confirmed that as the BBTR increases, the total CLP increases but what intrigues us is the relative contribution to this effect by the backbone and access network. This case deals with the effect on the access network and backbone network CLP independently. The insight generated here is to determine which element is more susceptible to changes as the BBTR is varied at the access network.

In this case, the service rate of the backbone network is fixed and the BBTR at the access network is increased. The initial value of the backbone service rate is calculated from the observed access network burst and base values from Bellcore October trace as shown by Krishnaswamy [8]. The $\rho$ for the backbone is fixed, based on the Bellcore trace values and the service rate of the backbone is calculated from the service rate of the access networks that are considered to be the arrival rate into the backbone when the servers are busy. It is true that as the values of the access networks are changed, the $\rho$ of the backbone also changes, but the graphs following, indicate the $\rho$ of the backbone dependent on the group of parameter values of the access networks when the weak stability was first observed. The access network display weak stability when the access network $\rho = 0.5$ and this value is taken for calculating the backbone network service rate. The backbone service rate takes three values which are 1111.11 pkts/second, 1666.666 and 3333.333 pkts/second closely simulating the 0.3, 0.5 and 0.9 traffic intensity in the backbone.

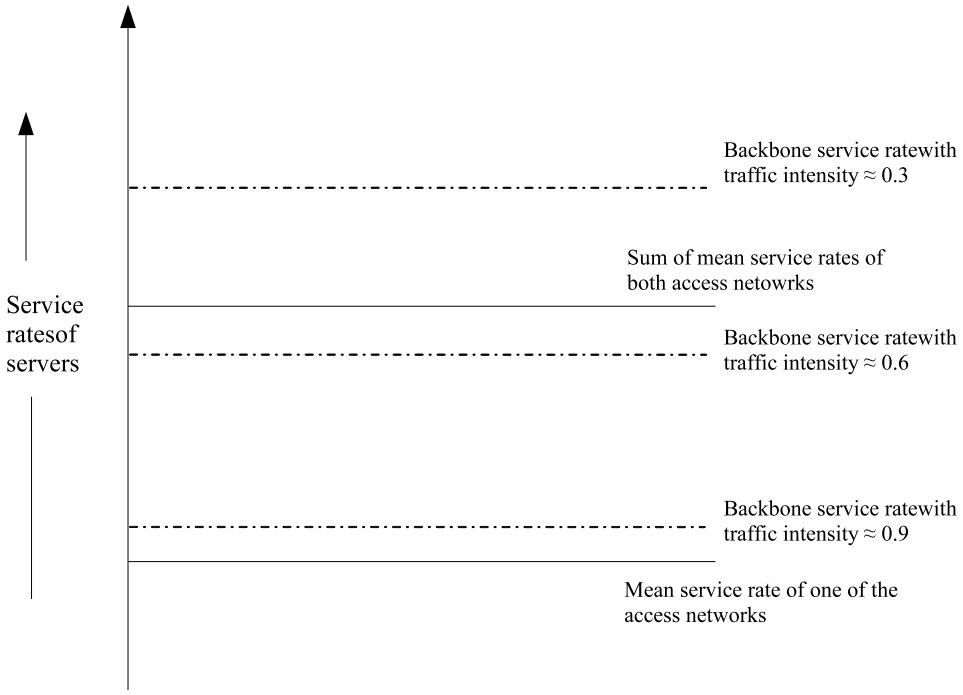
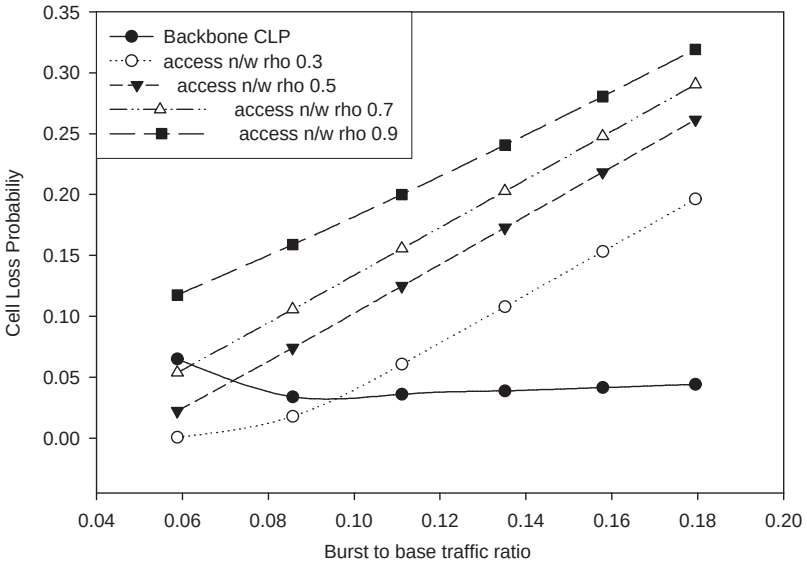



Figure 5.4: Range of backbone service rate relative to access network service rates

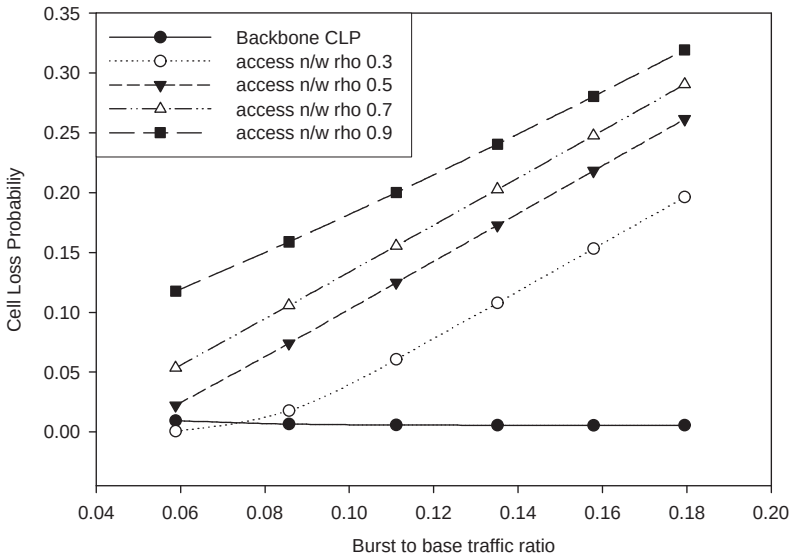
This is done to regulate the traffic entering the backbone. The arrival rate into the backbone starts from being much lower than the service rate of the backbone to being much greater than the service rate of the backbone as BBTR increases. From the network topology, we can say that the mean rate of arrivals to the backbone network is limited by the losses on the access networks, and ultimately by the mean service rate of the access networks. The traffic entering the backbone, starts from being stable and becomes weakly stable mode. This can be seen from the values assigned to the backbone service rate in the figure 5.4. When the service rate is 1111.11 pkts/second, it is noted that the backbone service rate lies from anywhere close to service rate of single access network to much greater than the sum of the service rates of the access networks.

The graphs 5.5, 5.6 and 5.7 show the relationships between the access network



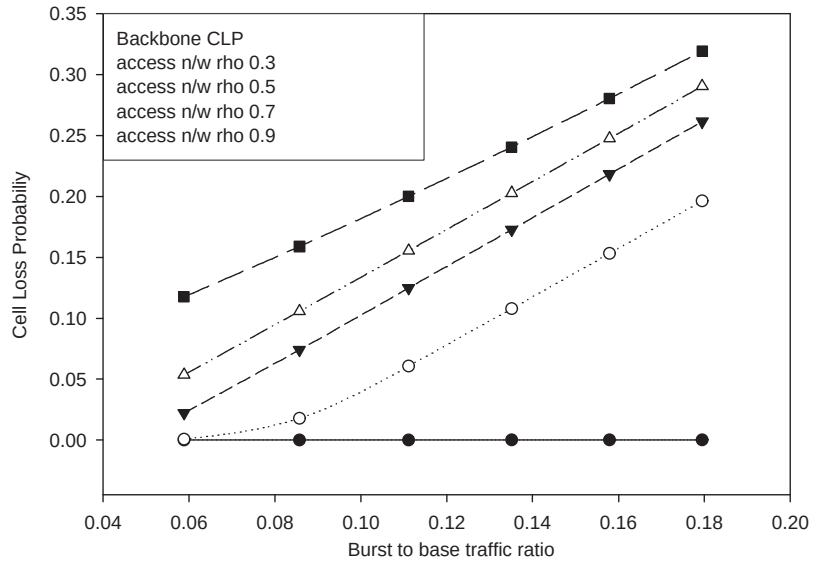

Backbone rho = 0.9. The rho for the access network is varied from 0.3 to 0.9. Access network CLP increases with BBTR. The backbone CLP reaches a constant value.

Figure 5.5: BB $rho$=0.9 BBTR vs CLP - Access network and Backbone CLP




Backbone rho = 0.6. The rho for the access network is varied from 0.3 to 0.9. Access network CLP increases with BBTR. The backbone CLP reaches a constant value.

Figure 5.6: BB $rho$=0.6 BBTR vs CLP - Access network and Backbone CLP




Backbone rho = 0.3. The rho for the access network is varied from 0.3 to 0.9. Access network CLP increases with BBTR. The backbone CLP reaches a constant value.

Figure 5.7: BB $rho$=0.3 BBTR vs CLP - Access network and Backbone CLP

CLP or the backbone CLP with the BBTR. The solid lines shows the dependence of the backbone CLP on the BBTR. The increase in BBTR results in more bursty traffic entering the backbone but since the backbone service rate is held constant, any further increase in the BBTR beyond a certain limit has no significant effect on the CLP of the backbone.

A simple analogy is like a tub with a leak in it being filled with water. The volume of the tub being the capacity of the backbone. Once the tub is filled up to the brim, keeping the water output from the leak constant, any variation in the speed in which the water keeps filling the tub makes no difference, since the water always overflows as long as the leak is not bigger than the arrival. This justifies the call-loss-probability in the backbone reaching a constant for a given service rate of the backbone as the burst to base ratio is increased, which is actually increasing the burstiness of the traffic.

It can be concluded that the increase in the BBTR results in the increase in total CLP as expected but the contribution to the increase in the CLP, predominantly comes from the access network when compared to the backbone.

## 5.5 Case 3

We have found that our model satisfies the sanity check. We were able to identify the level of contribution the access network and the backbone network provided towards the variation in the total CLP by fixing the backbone service rate and varying the BBTR. We found that for a given backbone service rate, the contribution of the access network towards the total CLP was much greater when compared to the effect of the backbone. The next question that follows the above discussion is, what is the significance of the backbone properties. Is there a role that backbone plays in regulating the total CLP? This question can be answered by taking a look at the next scenario.

In the above two cases, the backbone service rate was held constant. Any variations to the parameters was done on the access networks. Now the backbone service rate is varied to observe the effects on the CLP. As explained in case two, we are more focussed on the weak stability of the network. Previously, for a given backbone value, the access network values were changed. But now, for a given set of values for the access networks, the backbone service rate is varied.

From the graphs 5.1, 5.2 and 5.3, we can see that for the same values of arrival rate

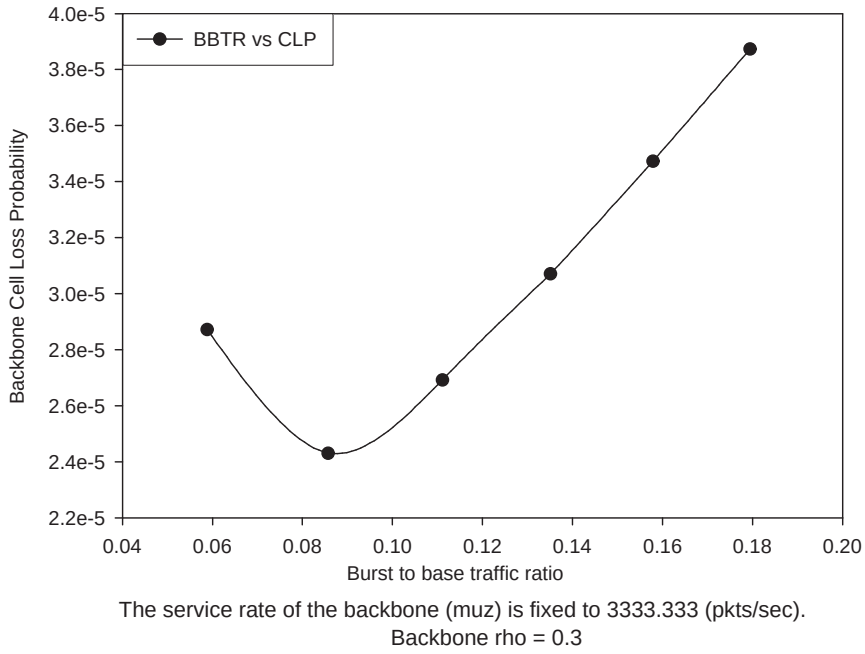



Figure 5.8: BB $rho$=0.3 BBTR vs Backbone CLP

and service rate of the access network, there is a drop in the value of the total CLP. This is shown more clearly by the following figures 5.8, 5.9 and 5.10. The bold lines in figure 5.11 give the backbone CLP and it can be noted that as the service rate of the backbone is increased, the CLP becomes much less, but more stable. Since the CLP is dependent on the BBTR, any variation in the BBTR affects the CLP of a network for a given service rate of the backbone. The graph indicating the various thresholds of stability in chapter two, figure 2.1 provides a visual aid to understand the limit on the increase of the burst-to-base traffic ratio.

As the BBTR increases it is noted that the mean arrival rate of the entire network slowly increases and the burst arrival rate increases beyond the mean service rate. This is when the burst mode becomes an unstable mode. Hence any dramatic increase in the burst-

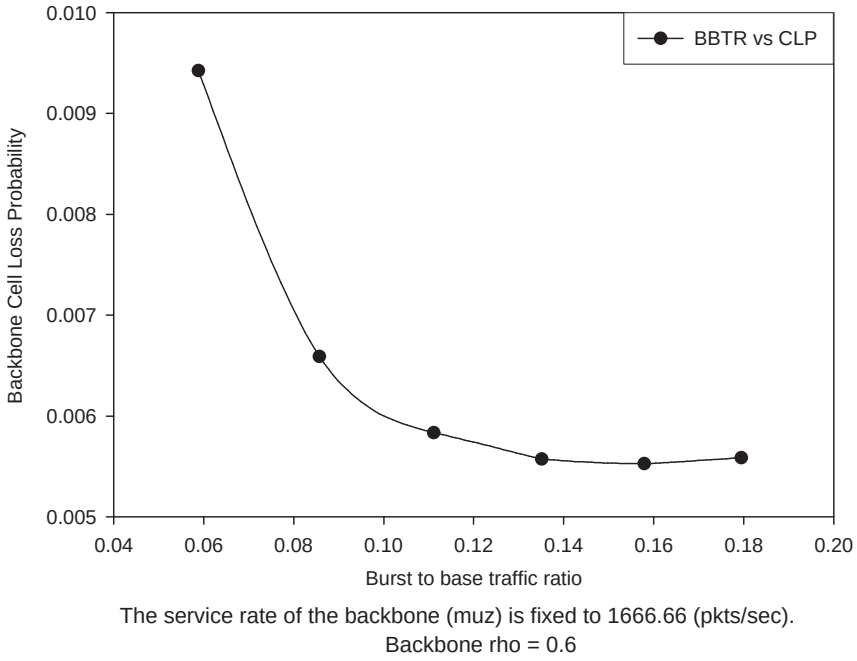



Figure 5.9: BB $rho$=0.6 BBTR vs Backbone CLP

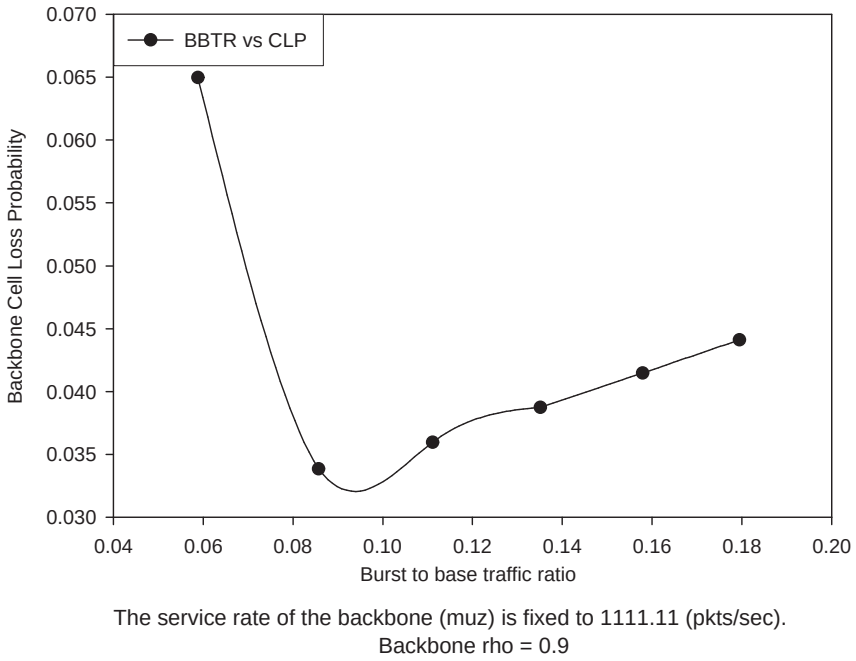



Figure 5.10: BB $rho$=0.9 BBTR vs Backbone CLP

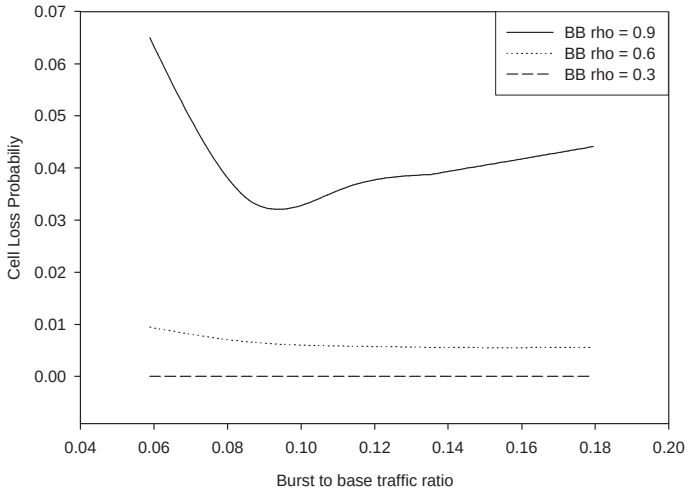



Figure 5.11: Comparison between Backbone CLP at different rho

to-base ratio would result in the entire network to be modelled as an unstable network rather than a weakly stable network. Intuitively, any variation in the backbone service rate would affect the call loss probability at the backbone. The call loss probability of the backbone has an influence on the total call loss probability which is actually the sum of the backbone and access network call loss probabilities. It can be seen that as the service rate of the backbone increases, total call loss probability decreases. This is because there is a decrease in the CLP at the backbone which contributes to the decrease in the total CLP.

From figures 5.8, 5.9 and 5.10, it can be seen that as the burst to base ratio increases, the call loss probability decreases till a certain point and after that starts increasing again in the backbone network. This at first glance seems like an anomaly since our understanding was that as the burst to base ratio increases the CLP should also increase, which happens to be true with the case of the total CLP. A possible conclusion can be drawn for this anomaly

which sounds intuitively plausible. It can be explained as follows. When the burst to base ratio is less, the amount of traffic entering the access network in the base mode is relatively high. Hence the amount of packet loss in the access network is less which causes the backbone network to receive a large volume of traffic and hence a heavy call loss probability in the backbone.

As the burst to base ratio increases, the CLP at the access network increases resulting in the drop in volume of traffic to the backbone, consequently resulting in the drop in the CLP in the backbone. Beyond a certain value, the burst to base ratio increase results in heavy burst traffic coming into the access network and even though there is a high CLP in the access network, there is still a relative increase in the traffic hitting the backbone resulting in increase of the CLP in the backbone.

The sum of the access network and backbone CLP result in the increase of the total network CLP which clearly matches the intuitive thought process. These graphs enable us to design the backbone network better such that there are no steep increase or drop in CLP and at the same time focussing on the near optimal configuration.

## 5.6 Case 4

We have satisfied the sanity check conditions with our model. We have had unexpected results, such as the dip in the CLP values, and have made an attempt to answer them intuitively. The next step is the observation of any behavior which this model would display,

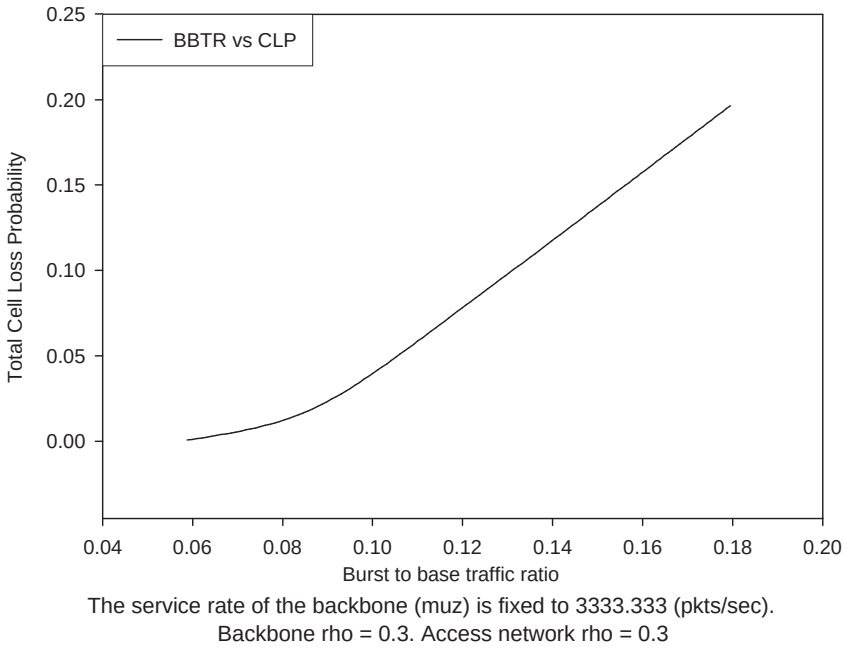



Figure 5.12: BB $rho$=0.3 BBTR vs Total CLP with Access network $rho$=0.3

not part of our original plan.

It has been noted that for a fixed service rate in the backbone, any variation in the traffic intensity, for the incoming access network produces almost identical call loss probability variations in the backbone. For a given backbone service rate, as long as the BBTR is held constant along with the burst capacity, the total CLP remains same for different values of $\rho$ in the access network. For instance, for backbone service rate = $3333.3333 packets/sec$ and for burst to base ratio of 0.05882 the total call loss probability for an access network with $\rho = 0.9$ is 0.0000287 and for access network with $\rho = 0.7$ its again 0.0000287. The figures 5.12, 5.13, 5.14 and 5.15 show the variation of the total CLP with respect to the BBTR when the service rate of the backbone is 3333.333 packets/sec. The $\rho$ of the access network take the values $0.3, 0.5, 0.7$ and 0.9. The figures 5.16, 5.17, 5.18 and 5.19 show the

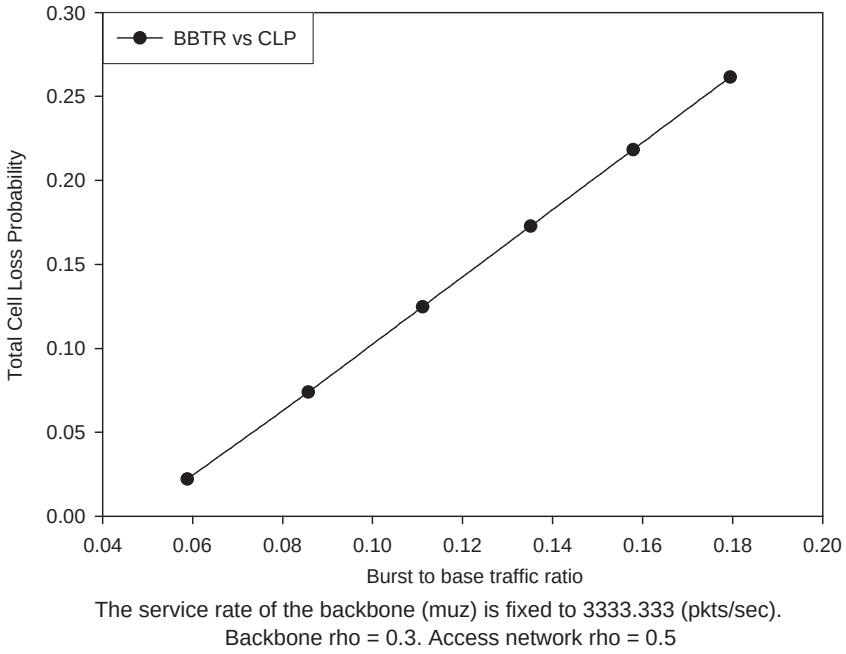



Figure 5.13: BB $rho$=0.3 BBTR vs Total CLP with Access network $rho$=0.5

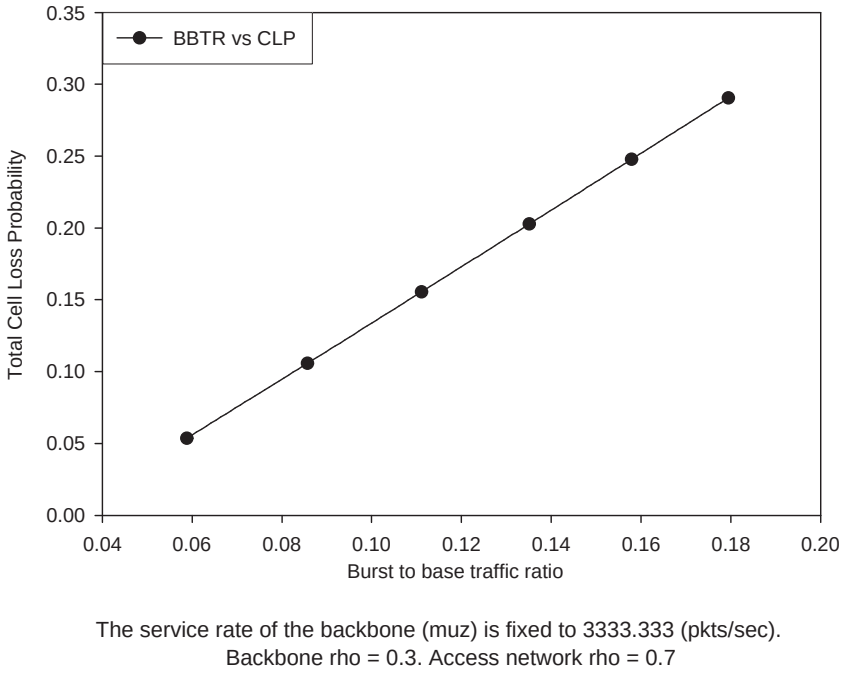



Figure 5.14: BB $rho$=0.3 BBTR vs Total CLP with Access network $rho$=0.7

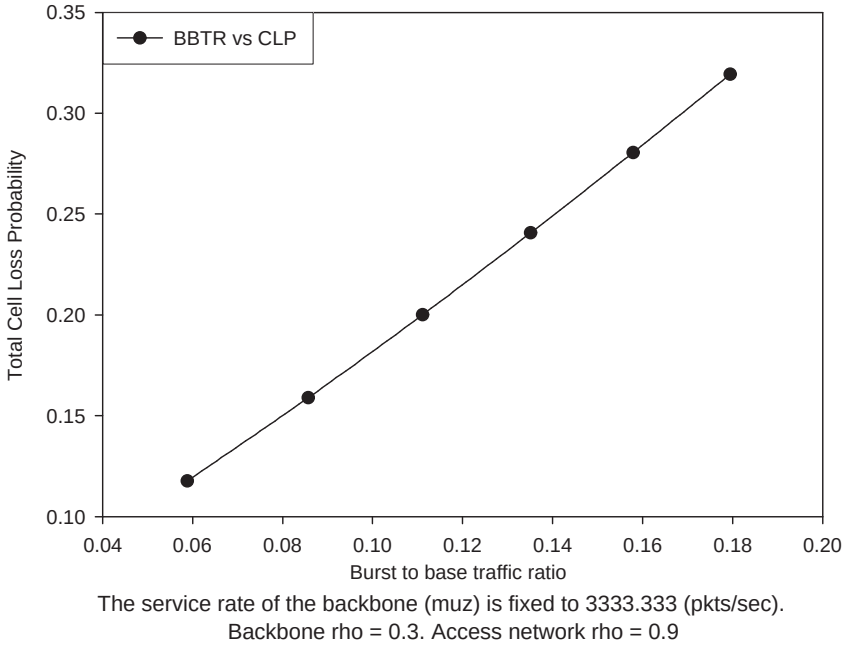



Figure 5.15: BB $rho$=0.3 BBTR vs Total CLP with Access network $rho$=0.9

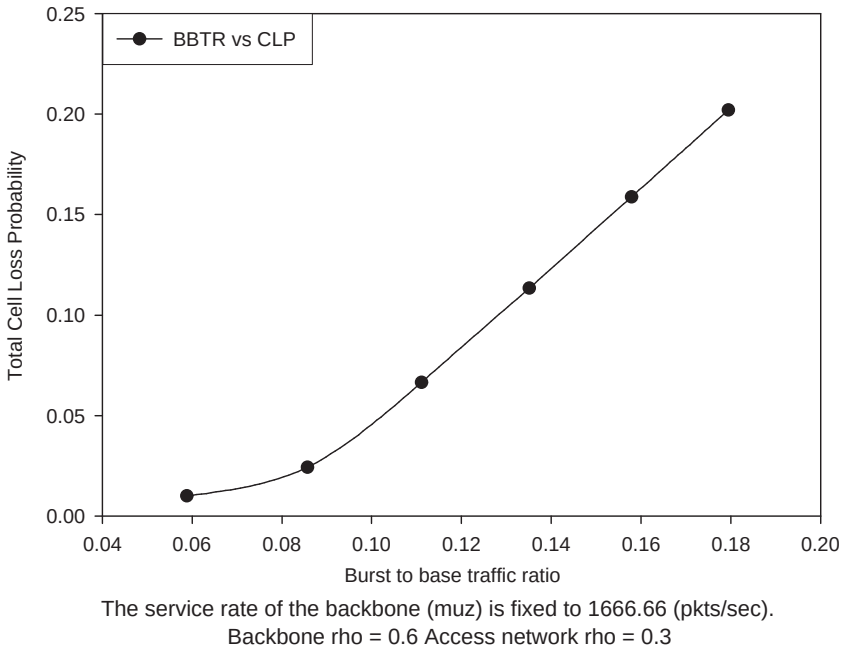



Figure 5.16: BB $rho$=0.6 BBTR vs Total CLP with Access network $rho$=0.3

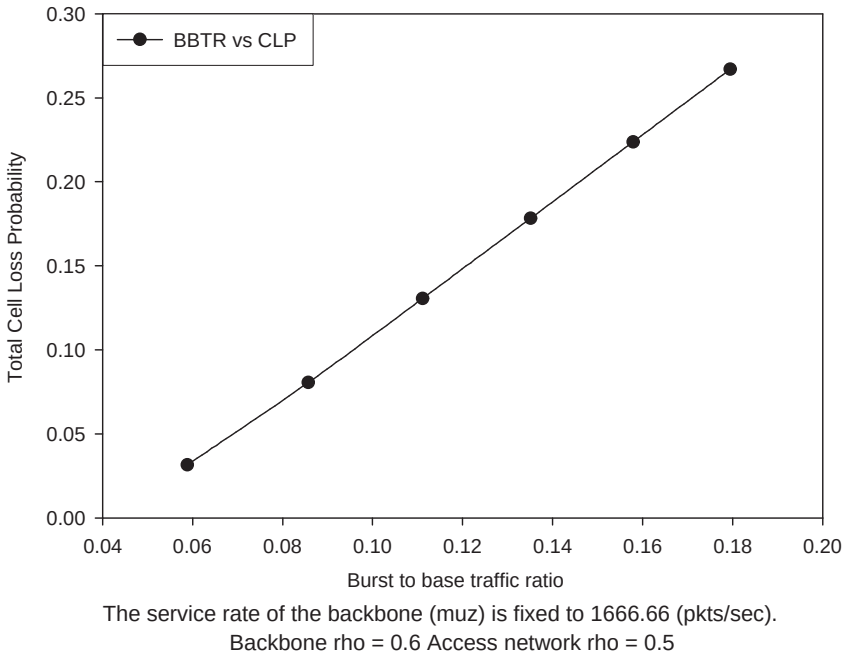



Figure 5.17: BB $rho$=0.6 BBTR vs Total CLP with Access network $rho$=0.5

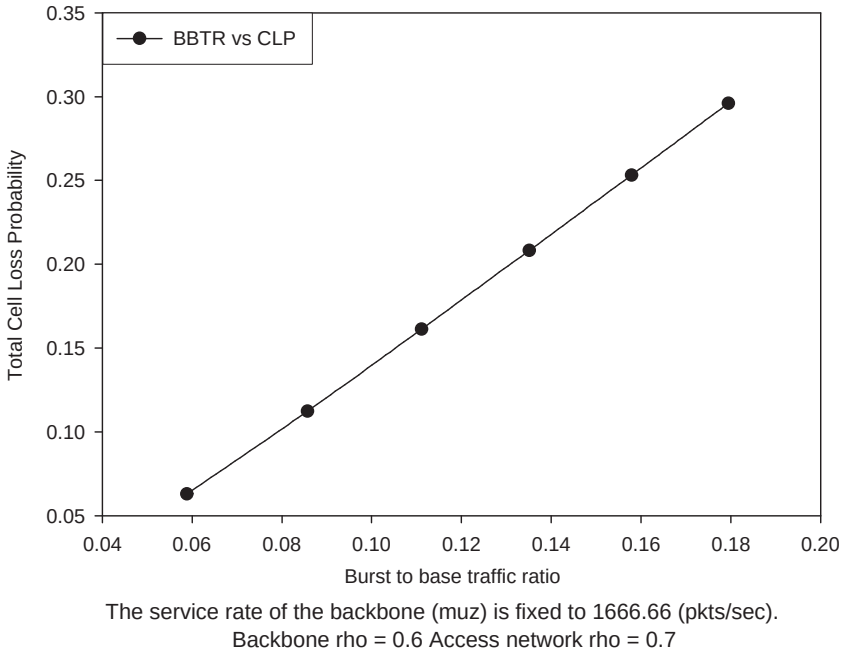



Figure 5.18: BB $rho$=0.6 BBTR vs Total CLP with Access network $rho$=0.7

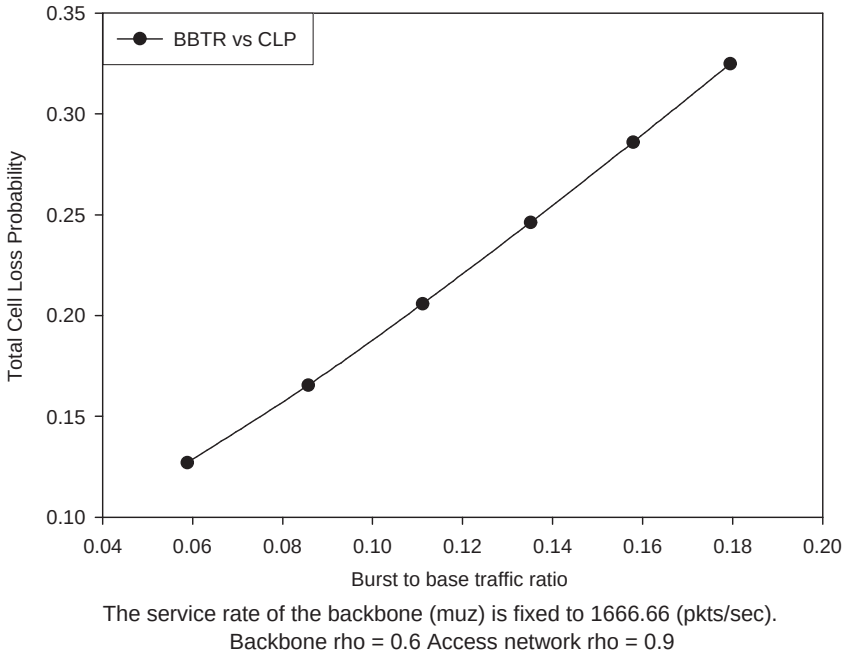



Figure 5.19: BB $rho$=0.6 BBTR vs Total CLP with Access network $rho$=0.9

variation of the total CLP with respect to the BBTR when the service rate of the backbone is 1666.66 packets/sec. The $\rho$ of the access network take the values $0.3, 0.5, 0.7$ and 0.9.

The figures 5.20, 5.21, 5.22 and 5.23 show the variation of the total CLP with respect to the BBTR when the service rate of the backbone is 1111.111 packets/sec. The $\rho$ of the access network take the values $0.3, 0.5, 0.7$ and 0.9.

The explanation for the backbone taking the service rates of 1111.111,1666.666 and 3333.333 is that the backbone is assigned values that closely resemble the current backbone service rates in real time networks. Many networks, including the one in the University of Kansas have the backbone $\rho$ equal to around 0.5. For insight and calculation purposes, we have taken the backbone $\rho$ to be equal to 0.3,0.6 and 0.9 with 0.3 corresponding to 3333.333

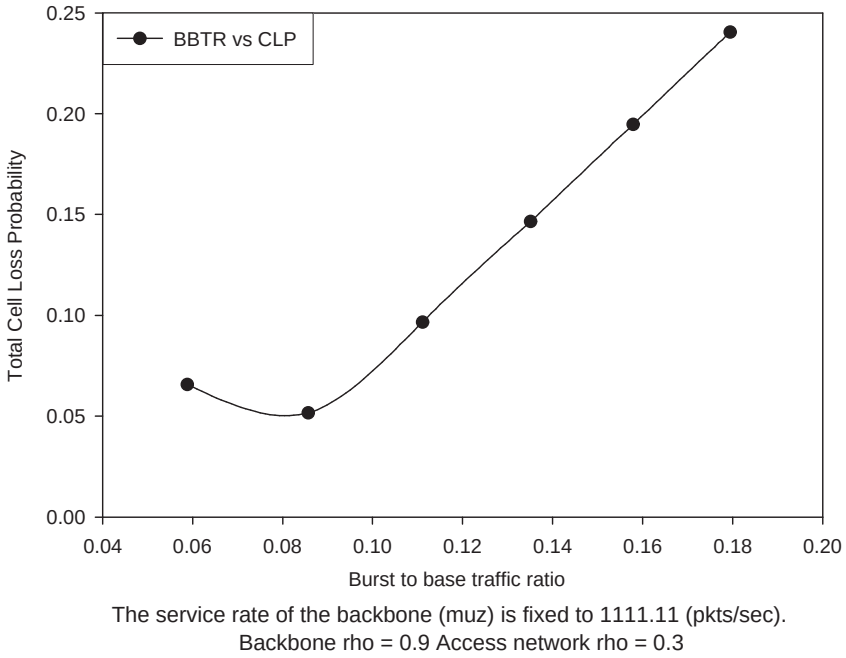



Figure 5.20: BB $rho$=0.9 BBTR vs Total CLP with Access network $rho$=0.3

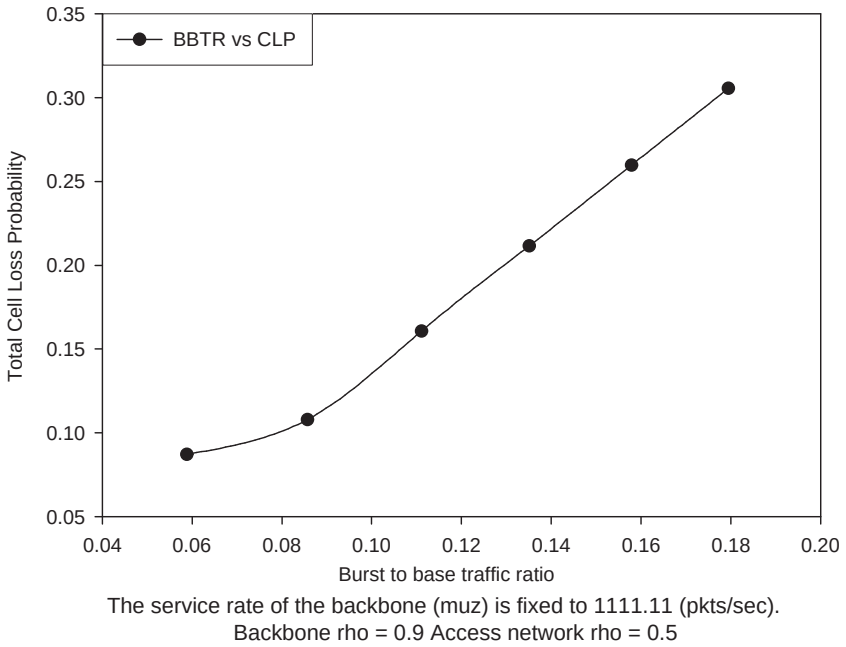



Figure 5.21: BB $rho$=0.9 BBTR vs Total CLP with Access network $rho$=0.5

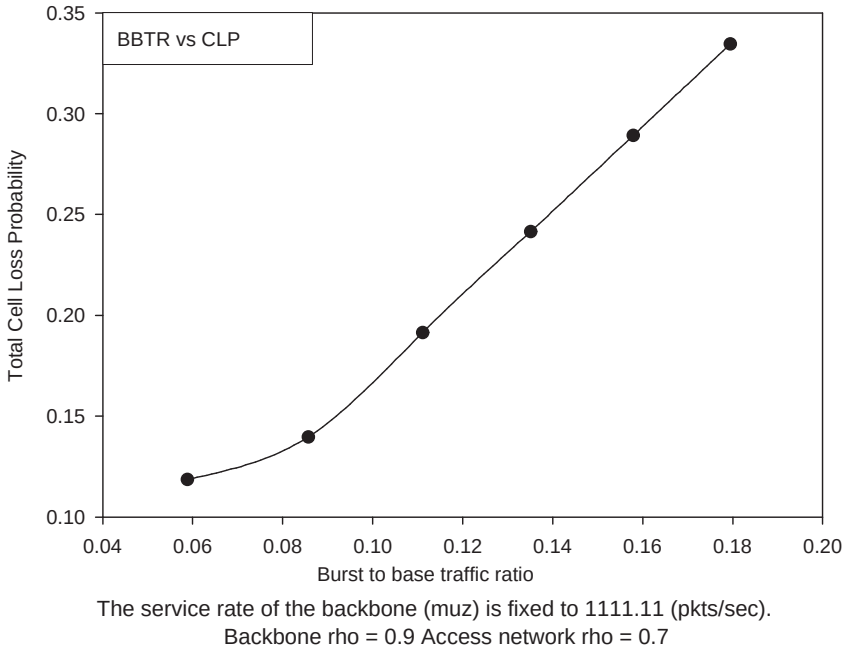



Figure 5.22: BB $rho$=0.9 BBTR vs Total CLP with Access network $rho$=0.7

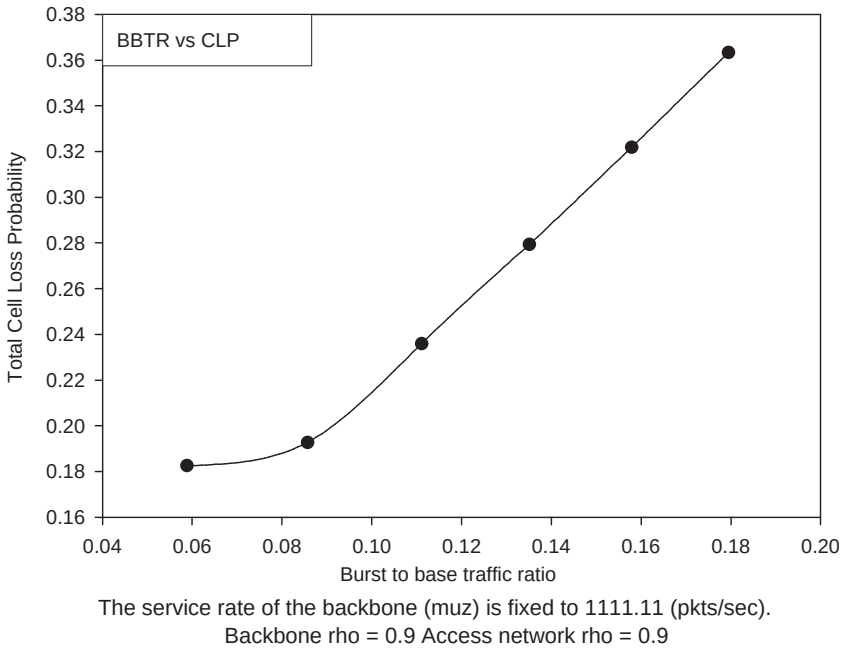



Figure 5.23: BB $rho$=0.9 BBTR vs Total CLP with Access network $rho$=0.9

| BBTR – | Backbone CLP BB $\rho = 0.9$ | Access CLP $\rho = 0.3$ | Access CLP $\rho = 0.5$ | Access CLP $\rho = 0.7$ | Access CLP $\rho = 0.9$ |
|---|---|---|---|---|---|
| 0.058823529 | 0.064981097 | 0.00068001 | 0.022171726 | 0.053544049 | 0.117589631 |
| 0.085714286 | 0.033849736 | 0.017759691 | 0.074025018 | 0.105754713 | 0.158883692 |
| 0.111111111 | 0.035959071 | 0.060666341 | 0.124741382 | 0.15546096 | 0.200021163 |
| 0.135135135 | 0.038735849 | 0.107814659 | 0.172764981 | 0.202728063 | 0.240635935 |
| 0.157894737 | 0.04146626 | 0.153215361 | 0.218268525 | 0.247683371 | 0.280452005 |
| 0.17948718 | 0.044109094 | 0.196379451 | 0.26144247 | 0.290459026 | 0.319270386 |

Table 5.1: The CLP values access network rho 0.9

| BBTR – | Backbone CLP BB $\rho = 0.6$ | Access CLP $\rho = 0.3$ | Access CLP $\rho = 0.5$ | Access CLP $\rho = 0.7$ | Access CLP $\rho = 0.9$ |
|---|---|---|---|---|---|
| 0.058823529 | 0.00942709 | 0.00068001 | 0.022171726 | 0.053544049 | 0.117589631 |
| 0.085714286 | 0.00659004 | 0.017759691 | 0.074025018 | 0.105754713 | 0.158883692 |
| 0.111111111 | 0.005834334 | 0.060666341 | 0.124741382 | 0.15546096 | 0.200021163 |
| 0.135135135 | 0.00557383 | 0.107814659 | 0.172764981 | 0.202728063 | 0.240635935 |
| 0.157894737 | 0.005527581 | 0.153215361 | 0.218268525 | 0.247683371 | 0.280452005 |
| 0.17948718 | 0.005586904 | 0.196379451 | 0.26144247 | 0.290459026 | 0.319270386 |

Table 5.2: The CLP values access network rho 0.6

packets/sec, 0.6 corresponding to 1666.666 packets/sec and finally 0.9 corresponding to 1111.11 packets/sec.

Here are a list of the values that were calculated that further show the variations in the ratios and CLPs. It can be concluded from the tables 5.3, 5.2 and 5.1, that the burst

| BBTR – | Backbone CLP BB $\rho = 0.3$ | Access CLP $\rho = 0.3$ | Access CLP $\rho = 0.5$ | Access CLP $\rho = 0.7$ | Access CLP $\rho = 0.9$ |
|---|---|---|---|---|---|
| 0.058823529 | 2.87161E-05 | 0.00068001 | 0.022171726 | 0.053544049 | 0.117589631 |
| 0.085714286 | 2.42981E-05 | 0.017759691 | 0.074025018 | 0.105754713 | 0.158883692 |
| 0.111111111 | 2.69169E-05 | 0.060666341 | 0.124741382 | 0.15546096 | 0.200021163 |
| 0.135135135 | 3.07029E-05 | 0.107814659 | 0.172764981 | 0.202728063 | 0.240635935 |
| 0.157894737 | 3.47253E-05 | 0.153215361 | 0.218268525 | 0.247683371 | 0.280452005 |
| 0.17948718 | 3.873E-05 | 0.196379451 | 0.26144247 | 0.290459026 | 0.319270386 |

Table 5.3: The CLP values access network rho 0.3

| BBTR | Total CLP $\rho = 0.3$ | Total CLP $\rho = 0.5$ | Total CLP $\rho = 0.7$ | Total CLP $\rho = 0.9$ | BB CLP |
|---|---|---|---|---|---|
| 0.058823529 | 0.065661107 | 0.087152815 | 0.118525135 | 0.182570722 | 0.064981097 |
| 0.085714286 | 0.051609427 | 0.107874747 | 0.139604444 | 0.19273342 | 0.033849736 |
| 0.111111111 | 0.096625411 | 0.160700458 | 0.191420032 | 0.235980237 | 0.035959071 |
| 0.135135135 | 0.146550508 | 0.211500832 | 0.241463915 | 0.279371787 | 0.038735849 |
| 0.157894737 | 0.194681622 | 0.259734787 | 0.289149633 | 0.321918267 | 0.04146626 |
| 0.17948718 | 0.240488544 | 0.305551562 | 0.334568119 | 0.363379477 | 0.044109094 |

Table 5.4: Table 4: Total CLP, BB rho = 0.9 with Access network range

| BBTR | Total CLP $\rho = 0.3$ | Total CLP $\rho = 0.5$ | Total CLP $\rho = 0.7$ | Total CLP $\rho = 0.9$ | BB CLP |
|---|---|---|---|---|---|
| 0.058823529 | 0.0101071 | 0.031598814 | 0.062971137 | 0.127016721 | 0.00942709 |
| 0.085714286 | 0.02434973 | 0.080615054 | 0.11234475 | 0.165473728 | 0.00659004 |
| 0.111111111 | 0.066500674 | 0.130575718 | 0.161295294 | 0.205855499 | 0.005834334 |
| 0.135135135 | 0.113388488 | 0.17833881 | 0.208301893 | 0.246209766 | 0.00557383 |
| 0.157894737 | 0.158742942 | 0.223796106 | 0.253210952 | 0.285979587 | 0.005527581 |
| 0.17948718 | 0.201966354 | 0.267029373 | 0.29604593 | 0.324857289 | 0.005586904 |

Table 5.5: Table 5: Total CLP, BB rho = 0.6 with Access network range

to base ratio plays the most critical role in determining the backbone call loss probability.

The next set of values from tables 5.6, 5.5 and 5.4, actually are those of the total CLP and the backbone CLP .

| BBTR | Total CLP $\rho = 0.3$ | Total CLP $\rho = 0.5$ | Total CLP $\rho = 0.7$ | Total CLP $\rho = 0.9$ | BB CLP |
|---|---|---|---|---|---|
| 0.058823529 | 0.000708726 | 0.022200442 | 0.053572765 | 0.117618347 | 2.87161E-05 |
| 0.085714286 | 0.01778399 | 0.074049316 | 0.105779012 | 0.15890799 | 2.42988E-05 |
| 0.111111111 | 0.060693257 | 0.124768299 | 0.155487876 | 0.20004808 | 2.69164E-05 |
| 0.135135135 | 0.107845362 | 0.172795684 | 0.202758766 | 0.240666638 | 3.07028E-05 |
| 0.157894737 | 0.153250087 | 0.21830325 | 0.247718096 | 0.280486731 | 3.47253E-05 |
| 0.17948718 | 0.196418181 | 0.2614812 | 0.290497756 | 0.319309116 | 3.873E-05 |

Table 5.6: Table 6: Total CLP, BB rho = 0.3 with Access network range

# Chapter 6

# Conclusion and future work

## 6.1 Conclusion and insight

We have paid little attention to exploiting known numerical efficiencies in solving the models, preferring to keep the model as simple as possible, and reducing the scope of the project to manageable size. The insights obtained in chapter five serve as demonstration, we hope, of the possibilities of the new methodology to contribute to general understanding of the dynamics of bursty traffic on larger networks. This thesis also shows how this type of model can be used to gain more detailed insight, even with simplified models. The analysis of the burstiness (a form of multi time-scale auto-dependence) in the access network, filtered through the buffering medium, reaching the backbone was done.

The relationship between burstiness in ethernet traffic and its effect on backbone

network performance has been explored using new tools for the construction and analysis of dependent matrix-exponential (MED) queueing network models. These tools are tailored to treat models with pseudo heavy tailed distributions and auto-covariances which are significant for lags extending over many orders of magnitude. They also introduce a promising new flexibility into the construction and solution of many station network models involving such processes.

An algebraic approach extending the art of MED queueing to more complex and extended networks of queues is outlined and used in the generation of an access/backbone mode. It employs the tools of Kronecker products, hat spaces, and nearly completely decomposable (NCD) operators to model event matrices modularly and hierarchically in time, function, and network scale. The tools also improve algebraic intuition and computational effectiveness.

The models examined allow multiple access-networks with realistically bursty, correlated traffic to be included in a network with a backbone network model. The models employ the concept of weak stability and multi-modal MED traffic to explore the effects of the multi-time-scale properties found in real traffic. Recent evidence confirms that these are plausible models that can faithfully represent the pathologies of published trace data.

The models have been explored to demonstrate the development of insight into the effect of access network traffic and design on call losses on the backbone. The effects of the changes in the access network's traffic intensity on the total CLP is observed. The effects on the backbone as a result of the merging of weakly stable Markovian queues from the access

network was better understood by manipulating the parameters that would yield the best insight into this model. Interesting insights have been presented, and offer guidance towards future, deeper explorations of other properties of such models.

## 6.2 Future work

Future work should address the numerical efficiency issues in solving for state probabilities and other measures of performance. We could also use a simulated validation to confirm numerical work using traces as opposed to analytical models. Further understanding of the modular traffic can be obtained by increasing the number of modes entering the access network. This is generally achieved by modelling the network using modal hyper-exponential arrival streams into the access networks as observed by Krishnaswamy [8]. This would result in the increase in the number of modes entering the backbone. This also would increase the probability of occurrence of weakly stable modes in the backbone providing more insights.

The model in our experiments turned out to be Markov modulated Poisson process (MMPP), but the general model we have described in Chapter 3, is a MED modulated MED process. It would be interesting to explore the effects of the richer model upon the results of our experiments. This MED modulated MED process model could then match the experimental observations more closely. The scope of this project can be extended to meet the requirements of the insights sought for by the researcher.

# Bibliography

[1] A.T. Anderson and B.F. Nielsen. A markovian approach for modelling packet traffic. *IEEE Journal on selected areas in communications*, 16(5):719–732, 1998.

[2] J.P. Buzen. Computational algorithms for closed queueing networks with exponential servers. *Communicaitons of the ACM*, 16(9), 1973.

[3] Alexander Graham. *Kronecker Products and Matrix Calculus : with Applications*. Ellis Horwood Limited, 1981.

[4] J.R. Jackson. Jobshop-like queueing systems. *Management Science*, 10(1), August 21-24 1963.

[5] P.R. Jelenkovic, A.A. Lazar, and Nemo Semret. The effect of multiple time scales and subexponentiality in MPEG video streams on queueing behavior. *IEEE Journal of Selected Areas in Communications*, 15(6), August 1997.

[6] R.E. Juliano. *A linear algebraic based solution method for queing systems with highly correlated arrival process*. PhD thesis, The University of Kansas, January 2000.

[7] Shoji Kasahara and B.F. Nielsen. Internet traffic modeling: Markovian approach to self-similar traffic and prediction of loss probability for finite queues. *IEICE Transactions on Communications*, E84-B(8):2134–2141, August 2001.

[8] R. Krishnaswamy. Capturing modes in bursty network traffic. Master's thesis, The University of Kansas, 2004.

[9] L.R. Lipsky. *Queueing Theory : A linear algebraic approach.* Macmillan Publishing Company, 1992.

[10] K. Mitchell, Khosrow Sohraby, Appie van de Liefvoort, and Jerry Place. Approximation models of wireless cellular networks using moment matching. *IEEE journal on Selected Areas in communications*, 19(11), November 2001.

[11] K. Mitchell and A. van de Liefvoort. Approximation models of feed-forward g/g/1/n queueing networks with correlated arrivals. *Perfomance Evaluation*, 251:137–152, 2003.

[12] M.F. Neuts. *Algorithmic Probability : A Collection of Problems (Stochastic Modeling).* Chapman and Hall, London, 1995.

[13] K. Park and W. Willinger. *Self-similar network traffic and performance evaluation.* Wiley, 2000.

[14] Vern Paxson and Sally Floyd. Wide area traffic: the failure of poisson modeling. *IEEE/ACM Transactions on Networking*, 3(3):226–244, 1995.

[15] P. Salvador, R. Valadas, and A. Pacheco. Multiscale fitting procedure using markov modulated poisson processes. *Telecommunications systems*, 23:123–148, June 2003.

[16] C.L. Samelson. *Product Form Solution For Queueing Networks With Poisson Arrivals And General Service Time Distributions With Finite Mean.* PhD thesis, University of Kansas, 1973.

[17] A. van de Liefvoort. The moment problem for continuous distribution. Technical report, Computer Science Telecommunications Program, University of Missouri - Kansas City, 1990.

[18] V.L. Wallace. On the representation of markovian systems by network models. Technical Report 42, Systems Engineering Laboratory, University of Michigan, Ann Arbor, August 1969.

[19] V.L. Wallace. Toward an algebraic theory for markovian networks. *Proceedings of the Symposium on Telecommunications and Computer Networks, MRI Symposium Proceedings*, 22:397–408, 1972.

[20] V.L. Wallace and K.B. Irani. On network linguistics and the conversational design of queueing networks. *Journal of the ACM*, 18:616–629, October 1971.

[21] T. Yoshihara, S. Kasahara, and Y. Takahashi. Practical time-scale fitting of self-similar traffic with markov modulated poisson process. *Telecommunications systems*, 17:185–211, 2001.