HELSINKI UNIVERSITY OF TECHNOLOGY
Department of Electrical and Communications Engineering
Networking Laboratory

Shuping Liu

Data Traffic Performance Analysis of a Cellular System with Finite User Population

Master's thesis submitted in partial fulfillment of the requirements for the degree of Master of Science in Technology, Espoo, Finland, August 15, 2004

Supervisor: Professor Jorma Virtamo, HUT

Instructor: Professor Jorma Virtamo

HELSINKI UNIVERSITY OF TECHNOLOGY Abstract of the Master's Thesis

Author:	Shuping Liu
Title of the thesis:	Data Traffic Performance Analysis of a Cellular System with Finite User Population
Date:	August 15, 2004
Number of pages:	83

Department:	Department of Electrical and Communications Engineering
Professorship:	S-38 Networking Technology
Major:	Teletraffic Theory
Minor:	Communications Networks and Systems

Supervisor:	Professor Jorma Virtamo
Instructors:	Professor Jorma Virtamo

Cellular systems have experienced a dramatic development over the past fifteen years. However, due to the random nature of traffic and the inherent "elasticity" of data transfers in cellular systems, few models have been explored about the user performance for wireless data channel. Some notable exceptions are the recent works by Borst for CDMA/HDR systems [11], and by Bonald and Proutière for cellular systems with Poissonian arrivals and infinite user population [4, 8].

In practice the user population in the real system cannot be infinite. This thesis develops two models with finite user population: the OCOF model and the OCMF model, in which a customer can transmit one flow or multiple flows at the same time, respectively. Arrival rate in the OCOF model is not Poissonian while the OCMF model allows a Poissonian arrival rate. In fact, the OCMF model is an intermediate model between the OCOF model and the model by Bonald et al. and the model by Bonald et al. can be regarded as a limit case of the OCOF and OCMF models when the number of users tends to infinity.

We identify two limit regimes, quasi-stationary and fluid, where the motion of the customers occurs on an infinitely slow and an infinitely fast time scale, respectively. Using queueing network theory, the average throughputs of the system in the two regimes are evaluated. The differences between the models are compared by a set of numerical examples.

Keywords: Cellular system, Data Traffic Performance, Throughput
Finite User Population

Acknowledgements

This Master's thesis was written in the Networking Laboratory of Helsinki University of Technology for the FIT project during the year 2004.

First of all, I would like to thank the supervisor and instructor of the thesis, *Professor Jorma Virtamo*, for his patient guidance and all the help during the whole process. *Lic. Sc. Esa Hyytiä* and *Ph.D Pasi Lassila* deserve my gratitude for all the help they provided me in learning the secrets of Latex. I also thank *M. Sc. Riikka Susitaival* for kindly offering me the template for this thesis.

I wish also to thank all the people in the laboratory for creating an enjoyable atmosphere to work.

Finally, I would like to express my gratitude to my parents and all my friends, for everything you have provided me so far.

Espoo, Finland, 15th August, 2004

Shuping Liu

Contents

List of abbreviations

QLD	Queue Length Dependent
SSFR	Single Server Fixed Rate
ROP	Random Observer Property
FCFS	First Come First Served
PS	Processor Sharing
IS	Infinite Server
LCFS-FR	Last Come First Served Pre-emptive Resume
i.i.d	independent and identically distributed
NMT	Nordic Mobile Telephony
GSM	Global system for Mobile
NA-TDMA	
CDMA	Code Division Multiple Access
PDC	Personal Digital Cellular
MS	Mobile Station
BS	Base Station
MSC	Mobile Switching Center
CDMA/HDR	Code Division Multiple Access/High Data Rate
QS	Quasi-Stationary regime
FL	Fluid regime
T	sojourn time
S	service time ($S = \frac{\sigma}{R}$)
OCOF	One Customer One Flow
OCMF	One Customer Multiple Flows

Chapter 1

Introduction

1.1 Background

Recently cellular systems have been developing greatly due to the operational limitations of conventional mobile telephone systems [15]. The next generation wireless cellular systems are expected to support a wide variety of high-speed data applications, in addition to conventional voice services and current low-bandwidth data service such as short message. The integration of these heterogeneous applications on a common transmission infrastructure raises similar challenges as in wireline integrated networks. In wireless environments, these issues are further exacerbated by interference problems, the intrinsically limited bandwidth, and the highly variable and unpredictable propagation characteristics. Specially, the channel quality may vary widely among spatially distributed users due to distance-related attenuation. In addition, the channel conditions for a given user may vary dramatically over time because of fading effects.

Very few papers address the issue of user performance for wireless data channel due to these random nature of traffic and the inherent elasticity of data transfers. Some notable exceptions are the recent works by Borst, Bonald and Proutiére .

In the work by Borst, the author evaluated user performance for CDMA/HDR systems. The results show that user performance is insensitive to the flow size distribution in a symmetric scenario.

In the works by Bonald and Proutiére, the authors developed some models for cellular system. However, all these models are based on the assumption that the customer number is infinite and arrival rate to the system is Poissonian. These assumptions are not valid in the real world systems.

In this thesis, based on queueing network theory we develop a new model for cellular system, named OCOF (One Customer One Flow) finite population model. The OCOF model is based on finite user population and does not require the arrival rate to the system necessarily to be Poissonian. Also give another finite population model, named OCMF (One Customer Multiple Flows) finite population model, which is an intermediate model between the OCOF model and the model by Bonald et al.

1.2 Objective

Objectives of this thesis can roughly be divided into two parts.

Firstly, we introduce the queueing network theory and the existing data models for cellular system.

Secondly, we develop and analyze two finite population models for a cellular system.

1.3 Structure of the thesis

The thesis is organized as follows: in Chapter 2 the theoretical background is introduced and finite queueing models are discussed. In Chapter 3 the existing data models for cellular system are presented.

In Chapter 4 and 5 we develop two finite population models for cellular system, named OCOF and OCMF model, respectively. The comparison of different models is discussed in Chapter 6.

Conclusion are drawn and some further work considerations are presented in Chapter 7.

Chapter 2

Theoretical Background

In this chapter the different queueing networks are introduced in general. Their stationary distribution and performance are discussed. Notion of balanced fairness and its main features are described. Two finite customer population queueing models are presented. This chapter is based on the results represented in [5–7, 9, 10, 12, 16, 19–21].

2.1 Queueing networks

A queueing network is a directed graph $G = (V, E)$, where $V = 1, 2, \ldots, M$ is a set of nodes and $E \subseteq V \times V$ is a set of arcs. Nodes represent active resources of some system and arcs define possible paths taken by customers. The workload of the system of resources is characterized by the service time at each node and the routing probabilities between nodes. For example, data packets traverse a network moving from a queue in a router to the queue in another router [12].

2.1.1 Classification of queueing networks

Queueing networks can be classified into three basic types:

Open networks: In open networks there is at least one arc coming from outside and at least one arc going out. The incoming arc comes from some infinite source of customers and the outgoing arc goes to some external sink. For example, Figure 2.1 shows a network with three nodes and two external arcs.

Closed networks: In closed networks there are no external arcs and so the network population is constant and all customers are forever circulating.

For example, Figure 2.1 with external arcs removed from nodes 1 and 2 is a closed network.

Mixed networks: Multi-class networks that are closed with respect to some customer types and open with respect to others are said to be mixed networks.

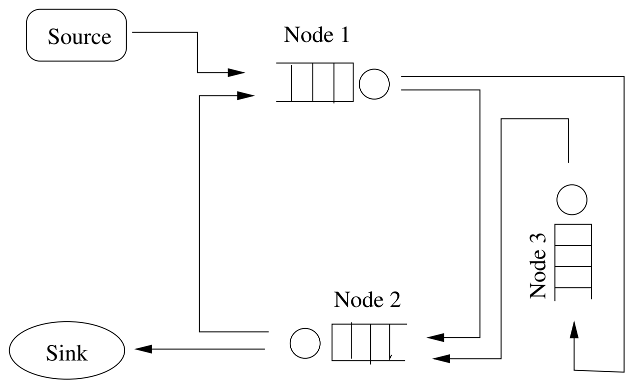

Figure 2.1: A simple queueing network

These three types of queueing networks relate to computer and communication systems in a very simple way. The open network can represent a transaction system with external arrivals and departures. The closed network is typically used for modelling batch type workloads where the multiprogramming level is held constant. The mixed network can represent both transaction and batch type workloads simultaneously in a computer system.

In the following, we'll introduce four kinds of queueing networks.

2.1.2 Jackson network

Definition

Jackson network is an open network. A Jackson network with M labeled nodes is defined as follows:

- Node i with n customers is QLD (Queue Length Dependent) with service rate $\mu_i(n)$.
- After service completion in a node the customer makes a probabilistic choice of either leaving the network or entering another node, independent of past history.
- The network is open and any external arrivals to node i form a Poisson stream.

Let us consider an open network with M nodes. The state space is given by:

$$S = \{\mathbf{n} = (n_1, n_2, \ldots, n_M) : \; n_i \geq 0 \quad \forall i\},$$

where n_i is the queue length at node i.

We define

$$\pi(\mathbf{n}) = \pi(n_1, n_2, \ldots, n_M)$$

as the stationary distribution (if it exists) that the network is in state $\mathbf{n}$.

Traffic equations

The arrivals to node i consist of external arrivals and arrivals from other nodes j. The mean arrival rate is thus

$$\lambda_i = \nu_i + \sum_{j=1}^{M} \lambda_j q_{ji}, \quad i = 1, 2, \ldots, M. \tag{2.1}$$

where

- λ_i is the mean arrival rate to node i in the network,
- ν_i is the mean external arrival to node i in the network,
- q_{ij} is the constant probability that on leaving node i a customer next goes to node j.

In the steady state the following equation holds:

$$\nu = \sum_{i=1}^{M} \lambda_i q_{i0}, \tag{2.2}$$

where

- ν is the mean total external arrivals to the network,
- q_{i0} is the probability of an external departure from node i.

Stationary distribution of the network

The stationary distribution of the network is of the product form,

$$\pi(n_1, n_2, \ldots, n_M) = \prod_{i=1}^{M} \pi_i(n_i), \tag{2.3}$$

where

$$\pi_i(n_i) = \pi_i(0)\left(\frac{\lambda_i^{n_i}}{\prod_{j=1}^{n_i}\mu_i(j)}\right),$$

and $\mu_i(j)$ is the service rate of node i when there are j customer in node i.

In open networks with all SSFR (Single Server Fixed Rate) nodes and fixed service rate μ_i at node i, the stationary distribution becomes:

$$\pi(n_1, n_2, \ldots, n_M) = \prod_{i=1}^{M}(1-\rho_i)\rho_i^{n_i}, \tag{2.4}$$

where $\rho_i = \frac{\lambda_i}{\mu_i}$.

Jackson's theorem

From the above stationary distribution of the network, it is can be seen that for a Jackson network in the steady state with arrival rate λ_i to node i:

1. The number of customers at any node is independent of the number of customers at every other node.
2. Node i behaves stochastically as if it were subjected to Poisson arrivals with rate λ_i.

ROP (Random Observer Property)

For an open Jackson queueing network, suppose $\pi(\mathbf{n})$ is the stationary probability that the network is in state $\mathbf{n}$. Then the probability that the network is in state $\mathbf{n}$ immediately before an arrival to any node is also $\pi(\mathbf{n})$.

Performance measures

Performance measures can be derived directly from Jackson's theorem and Little's result.

1. The mean queue length L_i of node i:

$$L_i = \frac{\rho_i}{1-\rho_i}. \tag{2.5}$$

2. The average number of customers in the network:

$$L = \sum_{i=1}^{M}\frac{\rho_i}{1-\rho_i}. \tag{2.6}$$

3. The average time spent in node i:

$$T_i = \frac{1}{\mu_i(1-\rho_i)}. \tag{2.7}$$

4. The average queueing time:

$$W_i = \frac{\rho_i}{\mu_i(1-\rho_i)}. \tag{2.8}$$

5. The mean time in the network of a customer entering node i satisfies the set of linear equations [21]:

$$T_{id} = T_i + \sum_{j=1}^{M} q_{ij} T_{jd}. \tag{2.9}$$

2.1.3 Gordon-Newell network

Definition

A Gordon-Newell network with M labeled nodes is a closed network defined as follows:

- Node i is QLD (Queue Length Dependent) with service rate $\mu_i(n)$ when it has n customers.
- A customer completing service at a node chooses next node probabilistically, independent of past history.
- The network is closed and has fixed population K.

The state space of the network is finite:

$$S(M,K) = \{\mathbf{n} = (n_1, n_2, \ldots, n_M) : \; n_i \geq 0 \quad \text{and} \quad \sum_{i=1}^{M} n_i = K \quad \forall i\},$$

where n_i is the queue length at node i.

We also define:

$$\pi(\mathbf{n}) = \pi(n_1, n_2, \ldots, n_M)$$

as the stationary distribution (if it exists) that the network is in state $\mathbf{n}$.

Traffic equations

The routing probabilities satisfy:

$$\sum_{j=1}^{M} q_{ji} = 1, \quad i = 1, 2, \dots, M. \tag{2.10}$$

The traffic equations are:

$$\lambda_i = \sum_{j=1}^{M} \lambda_j q_{ji}, \quad i = 1, 2, \dots, M. \tag{2.11}$$

where

- λ_i is the mean arrival rate to node i in the network,
- q_{ij} is the constant probability that on leaving node i a customer next goes to node j.

The traffic equations form a set of homogeneous linear equations that have the form $\boldsymbol{\lambda}(\mathbf{I}-\mathbf{Q}) = \mathbf{0}$, where $\mathbf{I}$ is the identity matrix and $\mathbf{Q}$ is the routing probability matrix. So the number of solutions for the traffic equations is infinite. However, all solutions differ only by a multiplicative factor. Let $(e_1, e_2, \dots, e_M)$ be any non-zero solution. Then e_i is proportional to the arrival rate at node i.

Stationary distribution of the network

The stationary distribution of the network is of the product form,

$$\pi(n_1, n_2, \dots, n_M) = \frac{1}{G} \prod_{i=1}^{M} x_i(n_i), \tag{2.12}$$

where

$$x_i(n_i) = \left(\frac{e_i^{n_i}}{\prod_{j=1}^{n_i} \mu_i(j)} \right), \quad \sum_{i=1}^{M} n_i = K,$$

and G is the normalizing constant defined by:

$$G(M, K) = \sum_{\mathbf{n} \in S(M,K)} \prod_{i=1}^{M} x_i(n_i). \tag{2.13}$$

Arrival theorem

In a closed network with K customers, the state probabilities seen by a customer entering any node are the same as the equilibrium probabilities (stationary distribution) $\pi(K-1,\mathbf{n})$ in a network with $K-1$ customers [21].

Computation of the normalizing constant

1. *The closed network with SSFR nodes only*

   For SSFR node i,

$$x_i(n_i) = x_i^{n_i}, \quad \text{where } x_i = \frac{e_i}{\mu_i}.$$

   For the closed network with SSFR nodes only, we have:

$$G(m,n) = \sum_{\mathbf{n}\in S(m,n)} \prod_{i=1}^{m} x_i^{n_i}.$$

   For $m, n \geq 0$ we split $G(m,n)$ into the cases when $n_m = 0$ and when $n_m > 0$:

$$\begin{aligned} G(m,n) &= \sum_{\substack{\mathbf{n}\in S(m,n)\\ n_m=0}} \prod_{i=1}^{m} x_i^{n_i} + \sum_{\substack{\mathbf{n}\in S(m,n)\\ n_m>0}} \prod_{i=1}^{m} x_i^{n_i} \\ &= \sum_{\mathbf{n}\in S(m-1,n)} \prod_{i=1}^{m-1} x_i^{n_i} + x_m \sum_{\substack{\mathbf{n}\in S(m,n)\\ k_i=n_i(i\neq m)\\ k_m=n_m-1}} \prod_{i=1}^{m} x_i^{k_i} \\ &= G(m-1,n) + x_m G(m,n-1) \end{aligned} \tag{2.14}$$

   The boundary conditions can be calculated from definitions:

$$G(m,0) = 1, \quad m > 0, \tag{2.15}$$
$$G(0,n) = 0, \quad n \geq 0. \tag{2.16}$$

   In this case, the customer rate through node i is given by:

$$\lambda_i = e_i \frac{G(M,K-1)}{G(M,K)}. \tag{2.17}$$

2. *The closed network with QLD nodes*

   We define the main generating function as follows:

$$f_i(z) = \sum_{n=0}^{\infty} x_i(n) z^n, \tag{2.18}$$
$$g_m(z) = \prod_{i=1}^{m} f_i(z). \tag{2.19}$$

So,

$$g_m(z) = \prod_{i=1}^{m} \sum_{n_i=0}^{\infty} x_i(n_i) z^{n_i} = \sum_{k=0}^{\infty} \left[\sum_{\mathbf{n} \in S(m,k)} \prod_{i=1}^{m} x_i(n_i) \right] z^k = \sum_{k=0}^{\infty} G(m,k) z^k.$$

Then,

$$\begin{aligned} g_m(z) &= g_{m-1}(z) f_m(z) \\ &= \sum_{k_1=0}^{\infty} G(m-1, k_1) z_1^k \sum_{k_2=0}^{\infty} x_m(k_2) z^{k_2} \\ &= \sum_{k=0}^{\infty} \left[\sum_{j=0}^{k} G(m-1, j) x_m(k-j) \right] z^k. \end{aligned}$$

From the above two equations, we find that:

$$G(1,k) = x_1(k), \tag{2.20}$$

$$\begin{aligned} G(m,k) &= \sum_{j=0}^{k} G(m-1,j) x_m(k-j) \\ &= G(m-1,k) + \sum_{j=0}^{k-1} G(m-1,j) x_m(k-j). \end{aligned} \tag{2.21}$$

In this case, the customer rate through node i is also given by:

$$\lambda_i = e_i \frac{G(M, K-1)}{G(M,K)}. \tag{2.22}$$

2.1.4 BCMP network

BCMP network can be an open network or a closed network. A BCMP network has the following characteristics:

- Four types of nodes in the network are allowed:

  type 1 FCFS (First Come First Served)
  type 2 PS (Processor Sharing)
  type 3 IS (Infinite Server)
  type 4 LCFS-FR (Last Come First Served Pre-emptive Resume)

- Multiple class of customers.

- With service stages information. Refer to [12] for details.

In a BCMP network, if there is a non-zero probability that a class-r customer can change to class s and a class-s customer can change to class r, then r and s are said to be *interchangeable.* The interchangeable relation is an equivalent relation and so it can partition the set of classes {1,2,...,K} such that two customer classes belong to the same subset P if and only if they are interchangeable.

In particular, we want to consider the nodes visited by any class in a subset P of the class partition. Thus we define a *routing chain* for subset P as a set of pairs (i,r), where node i is reachable by class r customers, $r \in P$. The routing chain contains information about all the nodes that a given customer class can visit, and all the classes that a customer can change into.

BCMP theorem

The BCMP theorem states that queueing networks with nodes of type 1, 2, 3 and 4 and multi-class traffic have a product form solution for the steady state joint probability distribution of the node states.

Notation:

$$\begin{aligned} J_{ir} &= \text{the number of stages for class } r, \\ A_{irs} &= \text{the probability of class } r \text{ customer entering stage } s, \\ \mu_{irs} &= \text{the service rate for class } r \text{ customer at stage } s, \\ \mu_{ir} &= \text{the service rate for class } r \text{ customer}, \\ n_{irs} &= \text{the number of customers of class } r \text{ in stage } s \text{ of service}, \\ n_{ir} &= \text{the number of customers of class } r, \\ n_i &= \text{the number of customers at the node.} \end{aligned}$$

In a general mixed multi-class queueing network the network state is denoted by:

$$\mathbf{n} = (\mathbf{n}_1, \mathbf{n}_2, \ldots, \mathbf{n}_M),$$

where $\mathbf{n}_i$ depends on the type of node.

For type 1 nodes, $\mathbf{n}_i = (r_{i1}, r_{i2}, \ldots, r_{i,n_i})$, where r_{ij} is the class of the jth customer waiting in FCFS order at node i.

For type 2 and type 3 nodes, $\mathbf{n}_i = (\mathbf{n}_{i1}, \mathbf{n}_{i2}, \ldots, \mathbf{n}_{iR})$ and $\mathbf{n}_{ir} = (n_{ir1}, n_{ir2}, \ldots, n_{irJ_r})$, where n_{irs} is the number of customers at node i of class r and in stage s of their service.

For type 4 nodes, $\mathbf{n}_i = ((r_{i1}, s_{i1}), (r_{i2}, s_{i2}), \ldots, (r_{i,n_i}, s_{i,n_i}))$, where r_{ij} is the class of the jth customer waiting in arrival order at node i and s_{ij} is its stage of service.

Then the steady state probability distribution of $\mathbf{n}$ is:

$$\pi(\mathbf{n}_1, \mathbf{n}_2, \ldots, \mathbf{n}_M) = G^{-1} d(K) \pi_1(n_1)) \pi_2(n_2) \ldots \pi_M(n_M), \tag{2.23}$$

where G is the normalizing constant, and

$$d(K) = \begin{cases} \prod_{k=0}^{K-1} \nu(k) & \text{if the network is open and has only 1 chain,} \\ \prod_{c=1}^{m} \prod_{k=0}^{K_c-1} \nu_c(k) & \text{if the network is open and has } m \text{ chains} \\ & \text{(chain } c \text{ has population } K_c\text{),} \\ 1 & \text{if the network is closed.} \end{cases} \quad (2.24)$$

$$\pi_i(\mathbf{n}_i) = \prod_{j=1}^{n_i} \frac{e_{ir_{ij}}}{\mu_i(j)} \qquad \text{if } i \text{ is type 1,} \quad (2.25)$$

$$\pi_i(\mathbf{n}_i) = n_i! \prod_{r=1}^{R} \prod_{s=1}^{J_{ir}} \left[\frac{1}{n_{irs}!} \left(\frac{e_{ir} A_{irs}}{\mu_{irs}} \right)^{n_{irs}} \right] \qquad \text{if } i \text{ is type 2,} \quad (2.26)$$

$$\pi_i(\mathbf{n}_i) = \prod_{r=1}^{R} \prod_{s=1}^{J_{ir}} \left[\frac{1}{n_{irs}}! \left(\frac{e_{ir} A_{irs}}{\mu_{irs}} \right)^{n_{irs}} \right] \qquad \text{if } i \text{ is type 3,} \quad (2.27)$$

$$\pi_i(\mathbf{n}_i) = \prod_{j=1}^{n_i} \left[e_{ir_{ij}} \frac{A_{ir_{ij}s_{ij}}}{\mu_{ir_{ij}s_{ij}}} \right] \qquad \text{if } i \text{ is type 4.} \quad (2.28)$$

Properties of BCMP networks

- All nodes in BCMP network preserve the Markov property.
- When a node is subjected to Poisson arrivals then the departure process is also Poisson and independent of the arrival process. This is called the $M \Rightarrow M$ property and holds for all four types of nodes in BCMP networks.
- If a Markovian queueing network has the local balance property then the $M \Rightarrow M$ property holds for all nodes in the network.
- The arrival theorem for single class networks extends to BCMP networks.

2.1.5 Whittle network

In a Whittle network, the service rate at each node is a function of the number of units at all the nodes, in contrast to a Jackson network whose service rate at each node depends only on the number of units at that node [19]. In particular, the Whittle network has the following characteristics:

- Nodes with PS service discipline.
- Capacities are balanced.

Model description

Consider an open queueing network of M nodes. Then,

- The exogenous arrivals to node i is Poisson process with intensity ν_i, $i = 1, 2, \ldots, M$.
- Customer requirement is i.i.d exponentially distributed with mean $1/\mu_i$.
- Denote by $\rho_i = \lambda_i/\mu_i$ the traffic intensity at node i.
- Denote by ϕ_i the capacity of node i, which depends on the state of the system $\mathbf{X} = (X_1, X_2, \ldots, X_M)$, where X_i is the random variable of the number of customers at node i. $\mathbf{x} = (x_1, x_2, \ldots, x_M)$ is a value of the random variable $\mathbf{X}$, where x_i is the number of customers at node i. Note that to be consistent with other part in the thesis, here we use vector $\mathbf{x} = (x_1, x_2, \ldots, x_M)$ to present the network state. In [6], the author uses x instead.
- The service discipline at each node is processor-sharing.
- After service completion at node i, a customer is routed to node j with probability p_{ij} and leaves the network with probability $p_i \equiv 1 - \sum_j p_{ij}$.
- The effective arrival rate λ_i at node i is uniquely defined by the traffic equations:

$$\lambda_i = \nu_i + \sum_j \lambda_j p_{ji}, \quad i = 1, 2, \ldots, M. \tag{2.29}$$

Balance property

Let us define vector $\mathbf{e_i}$ as unit vector with value 1 in the ith component and 0 elsewhere, $i = 1, 2, \ldots, M$. Then the balance property is defined as follows:

Balance property: Capacity allocation $\phi = (\phi_1, \phi_2, \ldots, \phi_M)$ is balanced if:

$$\phi_i(\mathbf{x})\phi_j(\mathbf{x} - \mathbf{e}_i) = \phi_j(\mathbf{x})\phi_i(\mathbf{x} - \mathbf{e}_j), \quad \forall i, j \ \forall \mathbf{x} : x_i \geq 0, x_j \geq 0. \tag{2.30}$$

Let $\langle \mathbf{x}, \mathbf{x} - \mathbf{e}_{i_1}, \ldots, \mathbf{x} - \mathbf{e}_{i_1} - \ldots - \mathbf{e}_{i_{n-1}}, 0 \rangle$ be a direct path from state $\mathbf{x}$ to state 0. This path has length $n \equiv |x|$, that is, the length is equal to the number of flows in state $\mathbf{x}$. The balance property 2.30 implies that the expression

$$\Phi(\mathbf{x}) = \frac{1}{\phi_{i_1}(\mathbf{x})\phi_{i_2}(\mathbf{x} - \mathbf{e}_{i_1}) \ldots \phi_{i_n}(\mathbf{x} - \mathbf{e}_{i_1} - \ldots - \mathbf{e_{i_{n-1}}})}, \tag{2.31}$$

is independent of the considered direct path [6]. A positive function Φ: $\mathbb{Z}_+^K \mapsto \mathbb{R}_+$, referred to as the balance function, defines the capacities unambiguously:

$$\phi_i(\mathbf{x}) = \frac{\Phi(\mathbf{x}-\mathbf{e}_i)}{\Phi(\mathbf{x})}, \quad \forall \mathbf{x} : x_i > 0. \tag{2.32}$$

Conversely, if there exists a positive function Φ such that the capacities satisfy 2.32, these capacities are balanced.

The balance property 2.30 implies that:

$$\frac{\phi_i(\mathbf{x}-\mathbf{e}_j)}{\phi_i(\mathbf{x})} = \frac{\phi_j(\mathbf{x}-\mathbf{e}_i)}{\phi_j(\mathbf{x})}, \quad \forall \mathbf{x} : x_i > 0, \tag{2.33}$$

which is an alternative way of writing the balance property. This means that the experienced relative change in allocation of one flow class caused by the removal of a flow of the other class, is equal for all flow class pairs.

Theorem

In [6], the following theorems are given:

Theorem 2.1.1 *For a Whittle network, the stationary distribution of $\boldsymbol{X}$ is:*

$$\pi(\mathbf{x}) = \Phi(\mathbf{x}) \prod_{i=1}^{M} \rho_i^{x_i}. \tag{2.34}$$

Theorem 2.1.2 *A processor-sharing network is a Whittle network if and only if the stationary distribution of* $\mathbf{X}$ *remain unchanged when at any node i and for any α_i, $0 \le \alpha_i \le 1$, customers require an exponentially distributed service of mean $1/\alpha_i \times 1/\mu_i$ with probability α_i, a null service with probability $1-\alpha_i$.*

The previous results could be extended to closed queueing network. First consider the case where routing is irreducible in the sense that each node is visited by the N customers, and

- Fixed number of customers N.
- $\nu_i = 0$ for all i.
- A particular node is referred to as the source, say node 0, whose capacity is a function ψ of the number of customers at this node only. We have $\phi_0(\mathbf{x}) = \psi_0(N - |\mathbf{x}|)$, where $\mathbf{x} = (x_1, x_2, \ldots, x_M)$ denotes the number of customers at any other node, characterize the system state and $|\mathbf{x}| = \sum_{i=1}^{M} x_i$.

- The arrival rate λ_i at node i is uniquely defined by the equations:

$$\lambda_i = \sum_j \lambda_j p_{ji}, \quad i = 1, 2, \ldots, M. \tag{2.35}$$

- $\rho_i = \lambda_i / \mu_i, \quad i = 1, 2, \ldots, M.$

Theorem 2.1.3 *For a closed irreducible Whittle network, the stationary distribution of* $\mathbf{X}$ *is:*

$$\pi(\mathbf{x}) = \Phi(\mathbf{x}) \prod_{i=1}^{M} \rho_i^{x_i} \prod_{n=1}^{N-|\mathbf{x}|} \frac{\rho_0}{\psi_0(n)}. \tag{2.36}$$

Theorem 2.1.4 *A closed irreducible processor-sharing network is a Whittle network if and only if the stationary distribution of* $\mathbf{X}$ *remains unchanged when at any node* $i \neq 0$ *and for any* α_i, $0 \leq \alpha_i \leq 1$, *customers require an exponentially distributed service of mean* $1/\alpha_i \times 1/\mu_i$ *with probability* α_i, *a null service with probability* $1 - \alpha_i$.

Similar results hold when routing is reducible. In this case, we

- Denote by $C_1, C_2, \ldots, C_K$ the K subsets of nodes such that routing is irreducible on each of these subsets.
- Each set C_k is visited by a fixed number of customers M_k.
- The arrival rate λ_i at any node i of C_k is uniquely defined by the equations:

$$\lambda_i = \sum_{j \in C_k} \lambda_j p_{ji}, \quad i \in C_k. \tag{2.37}$$

- Define $\varrho_k = \lambda_{i_k} / \mu_{i_k}$ and $\rho_i = \lambda_i / \mu_i$ for $i \in C_k / i_k$.
- For each k, the capacity of a particular node $i_k \in C_k$ is a function ψ_k of the number of customers at this node only. This node is referred to as the source of set C_k.
- Denote by M the number of nodes other than the sources, i.e., nodes in the set $\bigcup_k \{C_k / i_k\}$.
- The state space is $\mathbf{x} = (x_1, x_2, \ldots, x_M)$, where x_i is the customer number in the nodes other than the sources.

Theorem 2.1.5 *If the capacities* $\phi_1, \phi_2, \ldots, \phi_M$ *are balanced by a function* Φ, *the stationary distribution* π *of* $\mathbf{X}$ *is:*

$$\pi(\mathbf{x}) = \Phi(\mathbf{x}) \prod_{i=1}^{M} \rho_i^{x_i} \prod_{k=1}^{K} \prod_{n_k=1}^{M_k - \sum_{i \in C_k / i_k} x_i} \frac{\varrho_k}{\psi_k(n_k)} \tag{2.38}$$

Insensitivity

The results in this section are based on [6].

The network is said to be insensitive if the stationary distribution of customer number in each node is insensitive to any traffic characteristics except the traffic intensities $\rho_i, i = 1, 2, \ldots, M$. Insensitivity is a key property of Whittle network, which is proved as follows:

Consider the case where a node i is split into two separate nodes, i_1 and i_2, such that:

- Total capacity of two nodes depends on the total number of customers in these two nodes only, and
- Capacity is equally shared between customers.

i.e.,

$$\phi_{i_1}(\mathbf{x}) = \frac{x_{i_1}}{x_{i_1} + x_{i_2}} \phi_i(\mathbf{x}) \qquad \phi_{i_2}(\mathbf{x}) = \frac{x_{i_2}}{x_{i_1} + x_{i_2}} \phi_i(\mathbf{x}). \tag{2.39}$$

Then $\phi_1, \ldots, \phi_{i_1}, \phi_{i_2}, \ldots, \phi_M$ are balanced by Φ if and only if $\phi_1, \ldots, \phi_i, \ldots, \phi_M$ are balanced by $\widetilde{\Phi}$, with

$$\Phi(\mathbf{x}) = \binom{x_{i_1} + x_{i_2}}{x_{i_1}} \widetilde{\Phi}(x_1, \ldots, x_i, \ldots, x_M). \tag{2.40}$$

Proof of equation (2.40) is in the following:

$$\begin{aligned} \phi_{i_1}(\mathbf{x}) &= \frac{\Phi(\mathbf{x} - \mathbf{e}_{i_1})}{\Phi(\mathbf{x})} \\ &= \frac{\binom{x_{i_1} - 1 + x_{i_2}}{x_{i_1} - 1} \widetilde{\Phi}(x_1, \ldots, x_i - 1, \ldots, x_M)}{\binom{x_{i_1} + x_{i_2}}{x_{i_1}} \widetilde{\Phi}(x_1, \ldots, x_i, \ldots, x_M)} \\ &= \frac{x_{i_1}}{x_{i_1} + x_{i_2}} \phi_i(\mathbf{x}). \end{aligned}$$

The same applies for $\phi_{i_2}(\mathbf{x})$. This is to say that equation (2.40) is correct.

Based on theorem 1, the stationary distribution for the network with node i split is:

$$\pi(\mathbf{x}) = \binom{x_{i_1} + x_{i_2}}{x_{i_1}} \widetilde{\Phi}(x_1, \ldots, x_i, \ldots, x_M) \prod_{j=1}^{M} \rho_j^{x_j}. \tag{2.41}$$

If we remove the information of nodes i_1 and i_2 by summing over these two nodes, we get:

$$\begin{aligned}\pi(\mathbf{x}) &= \sum_{x_{i_1}+x_{i_2}=x_i} \binom{x_{i_1}+x_{i_2}}{x_{i_1}} \widetilde{\Phi}(x_1,\ldots,x_i,\ldots,x_M) \prod_{j=1}^{M} \rho_j^{x_j} \\ &= \sum_{x_{i_1}+x_{i_2}=x_i} x_i! \prod_{k=i_1,i_2} \frac{\rho_k^{x_k}}{x_k!} \cdot \widetilde{\Phi}(x_1,\ldots,x_i,\ldots,x_M) \prod_{j\neq i_1,i_2} \rho_j^{x_j} \\ &= \widetilde{\Phi}(x_1,\ldots,x_i,\ldots,x_M)(\rho_{i_1}+\rho_{i_2})^{x_i} \prod_{j\neq i_1,i_2} \rho_j^{x_j}. \end{aligned} \tag{2.42}$$

On the other hand, the stationary distribution $\widetilde{\pi}$ of the number of customers at nodes $1,\ldots,i,\ldots,M$ is given by:

$$\widetilde{\pi}(\mathbf{x}) = \widetilde{\Phi}(x_1,\ldots,x_i,\ldots,x_M)(\rho_{i_1}+\rho_{i_2})^{x_i} \prod_{j\neq i_1,i_2} \rho_j^{x_j}. \tag{2.43}$$

From equation (2.42) and (2.43), we can see that the aggregated stationary distribution for the network with the split node is the same as that for the network with an integrated node.

Now if a node i is split into P phases instead of just two, we could have Cox distribution service time for node i. Because Cox distribution approximates infinitely closely any distribution, we can conclude that the Whittle network is insensitive to the distribution of service times.

Performance measures

In [6], the authors give the following performance measures for Whittle network:

Denote by T_i the sojourn time in node i of a customer arriving at node i. Then according to Little's formula:

$$\mathrm{E}[x_i] = \lambda_i \mathrm{E}[T_i]. \tag{2.44}$$

The throughput of node i is defined as:

$$\gamma_i = \frac{1/\mu_i}{\mathrm{E}[T_i]} = \frac{\rho_i}{\mathrm{E}[x_i]}. \tag{2.45}$$

Instead of one node, consider a non-empty set of nodes $I = \{1,2,\ldots,M\}$. Then again Little's formula gives:

$$\sum_{i\in I} \mathrm{E}[x_i] = \lambda_I \mathrm{E}_I[T_I], \tag{2.46}$$

where T_I is sojourn time of first customer arriving at any node in I and λ_I is the arrival rate at nodes I:

$$\lambda_I = \sum_{i \in I} \left[\nu_i + \sum_{j \neq I} \lambda_j p_{ji} \right]. \tag{2.47}$$

Denote by $1/\mu_I$ the mean required quantity of service at nodes I, given by $\lambda_I/\mu_I = \sum_{i \in I} \rho_i$. Similarly we can define the throughput for nodes I:

$$\gamma_I = \frac{1/\mu_I}{\mathrm{E}_I[T_I]} = \frac{\sum_{i \in I} \rho_i}{\sum_{i \in I} \mathrm{E}[x_i]}. \tag{2.48}$$

Consider a Whittle network such that the total capacity of nodes I depends on the number of customers present at these nodes through their sum only and is equally shared between these customers, that is:

$$\frac{\sum_{i \in I} \phi_i(\mathbf{x})}{\sum_{i \in I} x_i} = \frac{\phi_i(\mathbf{x})}{x_i}, \quad i \in I, \quad x_i > 0, \tag{2.49}$$

Then $\gamma_I = \gamma_i$ for all $i \in I$.

2.2 Balanced fairness

Balanced fairness has the following characteristics:

1. Balanced fairness is the most efficient insensitive allocation.
2. In balanced fairness, a network link is saturated or a flow rate limit constraint is attained in any state. It is the only insensitive allocation with this kind of characteristics.
3. For balanced fairness, the data network is empty with the highest probability.

In [7, 9], Bonald and Proutièrethe introduce the balanced fairness in detail. We take some main results from [7, 9] as follows:

2.2.1 Network model

The network consists of a set of links $\mathcal{L} = \{1, 2, \ldots, L\}$, where link l has a finite capacity $C_l > 0$. A random number of flows compete for the access to these links.

There are K flow classes, each class k characterized by a route r_k. Each route r_k is a non-empty subset of the set $\mathcal{L}$. Let us denote $l \in r$ when link l belongs to the set of links used by route r.

A single flow is also characterized by the volume of information to be transferred, i.e. the flow size, on the flow's route. The duration of this transfer depends on the flow rate that varies in time. Relation between the flow size s and the time t when flow is transferred on route r is as follows:

$$s = \int_{t_{\text{start}}}^{t_{\text{end}}} c(t)dt,$$

where

- t_{start} is the arrival time of the flow,
  t_{end} is the end time of the flow,
  $c(t)$ denotes the flow rate at time t, i.e. the capacity allocated to this flow on each link $l \in r$ at time t, $t \in [t_{\text{start}},\ t_{\text{end}}]$. The allocated capacity $c(t)$ is additionally limited by condition $c(t) \leq C_l$ for all $l \in r$.

The state of the network is denoted by $\mathbf{x} = (x_1, \ldots, x_K)$, where x_k is the number of active flows of class k. The aggregated capacity ϕ_k is the capacity allocated for all flows of class k. Within a class k the capacity ϕ_k is shared equally between flows, i.e. each flow of class k is given the capacity of ϕ_k / x_k. The capacity allocation $\phi = (\phi_1, \ldots, \phi_K)$ is considered feasible when the following conditions hold:

$$\sum_{k:l \in r_k} \phi_k(\mathbf{x}) \leq C_l \quad \forall l \in \mathcal{L}. \tag{2.50}$$

That is, the aggregated capacity of flow classes traversing through link l may not exceed the link capacity C_l.

The traffic intensity of class k is denoted by ρ_k, which corresponds to mean volume of information offered by flows of class k per unit of time. Denote by $A_l = \sum_{k:l \in r_k} \rho_k(\mathbf{x})$.

The traffic conditions are given by the inequalities:

$$A_l \leq C_l, \quad l \in \mathcal{L}. \tag{2.51}$$

The above data network could be represented by the PS queueing network considered in previous Whittle network. Each customer in Whittle network corresponds to an ongoing flow in case of exponential flow size distribution, or to a phase of an ongoing flow in case of phase-type size distribution, i.e. class k here corresponds to the node i in the model of Whittle network.

2.2.2 Insensitive allocation

Sufficient condition for insensitivity

Based on the *balance property* in Whittle network, we have:

Capacity allocation $\phi = (\phi_1, \phi_2, \ldots, \phi_K)$ is balanced if:

$$\phi_k(\mathbf{x})\phi'_k(\mathbf{x} - \mathbf{e}_k) = \phi'_k(\mathbf{x})\phi_k(\mathbf{x} - \mathbf{e}'_k), \quad \forall k, k' \, \forall \mathbf{x} : x_k \geq 0, x'_k \geq 0. \tag{2.52}$$

and the capacities is defined by balance function Φ:

$$\phi_k(\mathbf{x}) = \frac{\Phi(\mathbf{x} - \mathbf{e}_k)}{\Phi(\mathbf{x})}, \quad \forall \mathbf{x} : x_k \geq 0. \tag{2.53}$$

The insensitivity of an allocation is implied by the balance property (2.52). The balance property is a sufficient condition for the insensitivity.

Necessary conditions for insensitivity

The converse is also true, i.e. an allocation is balanced if the invariant measure of the number of flows of each class is insensitive to any traffic character excluding the traffic intensities ρ. The balance property is implied by each of the following forms of insensitivity [7]:

1. *Insensitivity to the flow size distribution*: Assuming Poisson flow arrivals and identically and independently distributed (i.i.d.) flow sizes, the Poisson flow arrival process of any flow class can be changed to any phase-type distribution with the same expected value, and the invariant measures of the process of the number of flows remains unchanged.

2. *Insensitivity to the flow arrival process*: Assuming i.i.d. flow sizes, the Poisson flow arrivals process of any flow class can be changed to Poisson session arrivals, and the invariant measures of the process of the number of flows remains unchanged.

3. *Time-scale insensitivity*: Assuming Poisson flow arrivals and exponential i.i.d. flow sizes, the flow inter-arrival times and flow sizes of any flow class can be multiplied by the same constant, and the invariant measures of the process of the number of flows remains unchanged.

Any allocation satisfying one of the previous properties is balanced.

Property of insensitive allocation

1. Given balance function Φ, the stability condition for the corresponding allocation is:

$$\sum_{x} \Phi(\mathbf{x}) \prod_{k=1}^{K} \rho_k^{x_k} \leq \infty. \tag{2.54}$$

2. For any class k, the mean durationof a class-k of size σ satisfies:

$$T_k(\sigma) \geq \frac{\sigma}{a_k} \quad \text{and} \quad T_k(\sigma) \geq \frac{\sigma}{C_l - A_l} \qquad \forall l \in r_k \tag{2.55}$$

2.2.3 Balanced fairness

Based on equation (2.53), there are infinitely many insensitive allocations. Every of these allocations has to be feasible. From equations (2.50), and (2.53), one could get:

$$\sum_{k:l \in r_k} \frac{\Phi(\mathbf{x} - \mathbf{e}_k)}{\Phi(\mathbf{x})} = \frac{1}{\Phi(\mathbf{x})} \sum_{k:l \in r_k} \Phi(\mathbf{x} - \mathbf{e}_k) \leq C_l \quad \forall \mathbf{x},\ \forall l \in \mathcal{L},$$

Then,

$$\Phi(\mathbf{x}) \geq \frac{1}{C_l} \sum_{k:l \in r_k} \Phi(\mathbf{x} - \mathbf{e}_k), \quad \forall x,\ \forall l \in \mathcal{L},$$

Thus an unambiguous balanced allocation, denoted as *balanced fairness*, exists as follows:

For $\forall x \neq 0$,

$$\Phi^{\mathrm{BF}}(\mathbf{x}) = \max_{l \in \mathcal{L}} \left\{ \frac{1}{C_l} \sum_{k:l \in r_k} \Phi^{\mathrm{BF}}(\mathbf{x} - \mathbf{e}_k) \right\}. \tag{2.56}$$

The above equation is recursive with the initial assumption $\Phi^{\mathrm{BF}}(0) = 1$. Any link realizing the maximum in (2.56) is called saturated in state $\mathbf{x}$.

2.3 Finite customer population model

In the following we will introduce two kinds of finite customer population queue models, Binomial model and Engset model [18, 21]:

2.3.1 Binomial model

Binomial model refers to the following teletraffic model [18]:

- Finite number of independent customers ($n < \infty$). The customers are on-off type.
- Idle time are i.i.d and exponentially distributed with mean $1/\nu$.
- As many servers as customers ($s = n$).
- Service times are i.i.d and exponentially distributed with mean $1/\mu$.
- No waiting places ($m = 0$).

This is an $M/M/n/n/n$ queue by Kendall's notation. The queue is lossless.

On-off type customer

The customers are on-off type between idleness and activity. Take a call traffic as an example, when on a call is going on and when off a call is idle, as shown in Figure 2.2:

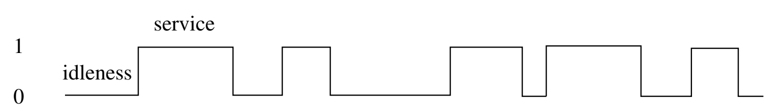

Figure 2.2: On-off type customer in binomial model

Stationary distribution

Denote by $X(t)$ the number of active customers. It is obvious that process $X(t)$ is a Markov process, whose state transition diagram is shown in Figure 2.3:

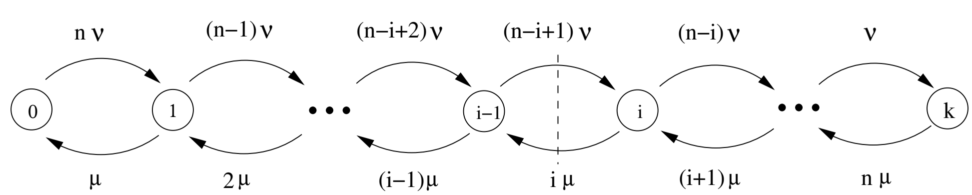

Figure 2.3: The state transition diagram of Binomial queue model

Let π_i the stationary distribution of state i. Then LBE (Local Balance Equation) is:

$$\begin{aligned}
\pi_{i-1}(n-i+1)\nu &= \pi_i \cdot i\,\mu \Rightarrow \\
\pi_i &= \frac{(n-i+1)\nu}{i\mu}\pi_{i-1} \Rightarrow \\
\pi_i &= \frac{n!}{i!(n-i)!}\left(\frac{\nu}{\mu}\right)^i \pi_0 \Rightarrow \\
\pi_i &= \binom{n}{i}\left(\frac{\nu}{\mu}\right)^i \pi_0, \quad i = 0, 1, \ldots, n.
\end{aligned} \tag{2.57}$$

With normalizing condition,

$$\sum_{i=0}^{n} \pi_i = \pi_0 \sum_{i=0}^{n} \binom{n}{i}\left(\frac{\nu}{\mu}\right)^i = 1. \tag{2.58}$$

We get that:

$$\pi_0 = \left[\sum_{i=0}^{n} \binom{n}{i}\left(\frac{\nu}{\mu}\right)^i\right]^{-1} = \left(\frac{\mu}{\nu+\mu}\right)^n, \tag{2.59}$$

and

$$\pi_i = \binom{n}{i}\left(\frac{\nu}{\nu+\mu}\right)^i \left(\frac{\mu}{\nu+\mu}\right)^{n-i}, \quad i = 0, 1, \ldots, n. \tag{2.60}$$

i.e. each source is on with probability $\frac{\nu}{\nu+\mu}$ and off with probability $\frac{\mu}{\nu+\mu}$, independent of the others.

If we denote $\widehat{a} = \frac{\nu}{\mu}$, then the stationary distribution of state i can be written as follows:

$$\begin{aligned}
p\{X = i\} &= \pi_i \\
&= \frac{\binom{n}{i}\left(\frac{\nu}{\mu}\right)^i}{\left(1+\frac{\nu}{\mu}\right)^n} \\
&= \frac{\binom{n}{i}\left(\frac{\nu}{\mu}\right)^i}{\sum_{j=0}^{n}\binom{n}{j}\left(\frac{\nu}{\mu}\right)^j} \\
&= \frac{\binom{n}{i}\widehat{a}^i}{\sum_{j=0}^{n}\binom{n}{j}\widehat{a}^j}, \quad i = 0, 1, \ldots, n.
\end{aligned} \tag{2.61}$$

2.3.2 Engset model

The definition of Engset model is as follows [18]:

- Finite number of independent customers ($n < \infty$). The customers are on-off type.
- Idle times are i.i.d and exponentially distributed with mean $1/\nu$.
- Less servers than customers ($s < n$).
- Service times are i.i.d and exponentially distributed with mean $1/\mu$.
- No waiting places ($m = 0$).

This is an $M/M/s/s/n$ queue by Kendall's notation. The queue is lossy.

On-off type customer

The customer is on-off type between idleness and activity. If the system is full when an idle customer tries to become an active customer, a new idle period starts, as shown in Figure 2.4:

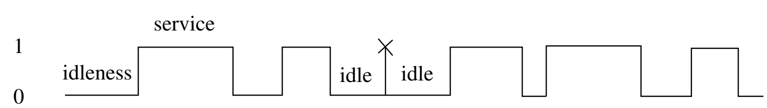

Figure 2.4: On-off type customer in Engset model

Stationary distribution

Let $X(t)$ be the number of active customers. The process $X(t)$ is a Markov process with state transition diagram as shown in Figure **??**:

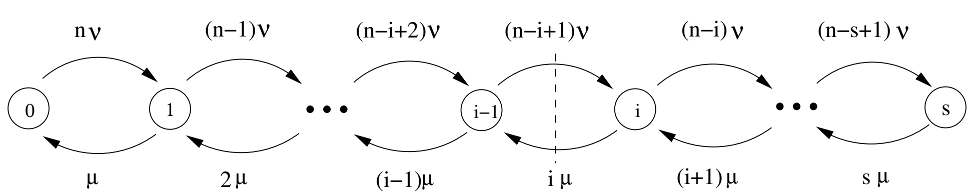

Figure 2.5: The state transition diagram of Engset queue model

Denote by π_i the stationary distribution of state i. Then LBE (Local Balance Equation) is:

$$\begin{aligned}\pi_{i-1}(n-i+1)\nu &= \pi_i \cdot i\,\mu \Rightarrow \\ \pi_i &= \frac{(n-i+1)\nu}{i\mu}\,\pi_{i-1} \Rightarrow \\ \pi_i &= \frac{n!}{i!(n-i)!}\left(\frac{\nu}{\mu}\right)^i \pi_0 \Rightarrow \\ \pi_i &= \binom{n}{i}\left(\frac{\nu}{\mu}\right)^i \pi_0, \quad i = 0, 1, \ldots, s. \end{aligned} \tag{2.62}$$

With normalizing condition,

$$\sum_{i=0}^{s} \pi_i = \pi_0 \sum_{i=0}^{s} \binom{n}{i}\left(\frac{\nu}{\mu}\right)^i = 1. \tag{2.63}$$

We get that:

$$\pi_0 = \left[\sum_{i=0}^{s} \binom{n}{i}\left(\frac{\nu}{\mu}\right)^i\right]^{-1}. \tag{2.64}$$

Thus the stationary distribution is a truncated binomial distribution:

$$P\{X=i\} = \pi_i = \frac{\binom{n}{i}\left(\frac{\nu}{\mu}\right)^i}{\sum_{j=0}^{s}\binom{n}{j}\left(\frac{\nu}{\mu}\right)^j}, \quad i = 0, 1, \ldots, s. \tag{2.65}$$

Similarly, denote $\widehat{a} = \frac{\nu}{\mu}$, then the stationary distribution of state i is:

$$P\{X=i\} = \frac{\binom{n}{i}\widehat{a}^i}{\sum_{j=0}^{s}\binom{n}{j}\widehat{a}^j}, \quad i = 0, 1, \ldots, s. \tag{2.66}$$

Blocking probability

Blocking probability in Engset model could be *Time Blocking*, denoted by B_t, or *Call Blocking*, denoted by B_c [18].

B_t is the probability that all s servers are occupied at an arbitrary time, which is equivalent with the fraction of time that all s servers are occupied. For a stationary Markov process, it equals the stationary distribution when the state is in s, i.e.:

$$B_t = P\{X=s\} = \frac{\binom{n}{s}\widehat{a}^s}{\sum_{j=0}^{s}\binom{n}{j}\widehat{a}^j}. \tag{2.67}$$

B_c is the probability that an arriving customer finds all s servers occupied, which is equivalent with the fraction of arriving customers that are lost. The state distribution seen by an arriving customer is the same as the stationary distribution in a system with one less customer [21]. Thus,

$$B_c = \pi_i(n-1) = \frac{\binom{n-1}{s}\widehat{a}^s}{\sum_{j=0}^{s}\binom{n-1}{j}\widehat{a}^j}. \tag{2.68}$$

The load of the system

We will discuss two kinds of loads in the following, Realized load, ρ_{real}, and Intended load, ρ_{int}, which is based on [21].

For the intended load, the sojourn time of each customer in BS tends to 0, as shown in Figure 2.6:

Figure 2.6: The intended load model in Engset queue

For the realized load, the sojourn time of each customer in BS is larger than 0, as shown in Figure 2.4. Let's consider one cycle which consists of one idle period and one service period. The service period is a real active period with the probability of $1 - B_c$, and is shrunk to blocking event with the probability of B_c which is shown in Figure 2.7:

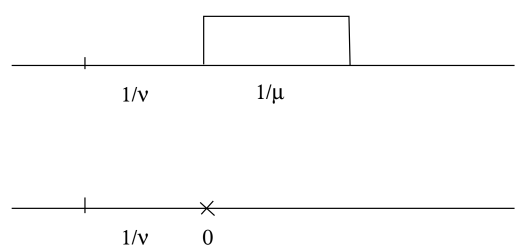

Figure 2.7: The cycle period which consists of idle period and service period

Thus the mean sojourn of a cycle is as follows:

$$\frac{1}{\nu} + (1 - B_c) \cdot \frac{1}{\mu} + B_c \cdot 0 = \frac{1}{\nu} + (1 - B_c) \cdot \frac{1}{\mu}. \tag{2.69}$$

The probability that the call is in the idle phase is:

$$\frac{1/\nu}{1/\nu + (1 - B_c) \cdot 1/\mu} = \frac{1}{1 + (1 - B_c)\widehat{a}}. \tag{2.70}$$

So the offered realized load, denoted by a, is:

$$a = \frac{\widehat{a}}{1 + (1 - B_c)\widehat{a}} n. \tag{2.71}$$

and the carried realized load, denoted by a_c, is:

$$a_c = (1 - B_c)a. \tag{2.72}$$

We have the following relationship between realized load and intended load:

$$\rho_{real} < \rho_{int}. \tag{2.73}$$

Chapter 3

Modelling data communication in cellular systems

Cellular systems have experienced a dramatic development over the past twenty years due to the operational limitations of conventional mobile telephone systems [15].

- NMT (Nordic Mobile Telephony) was opened in 1981. NMT was the world's first multinational cellular network [17].
- In 1992 the first digital cellular system, GSM (Special Mobile Group) was deployed in Germany. GSM is a European standard system.
- In the United States, an NA-TDMA system (IS-54) and a CDMA system (IS-95) have been developed. NA-TDMA was deployed in 1993 and CDMA is planned for deployment in 1995.
- A Japanese system, PDC (Personal Digital Cellular), was deployed in Osaka in June 1994.

A basic cellular system consists of three parts:

- *MS (Mobile Station)*: A mobile telephone unit contains a control unit, a transceiver, and an antenna system.
- *BS (Base Station) and MSC (Mobile Switching Center)*: BS and MSC are the central elements for all cellular system. It interfaces with telephone company and other networks, controls call and data processing, and handles billing activities. We usually call this part only BS.
- *The interconnection network*: Network relative to the cellular system.

As the radio wave propagational factors limit the coverage area, there must be a large number of BSs in a large service area. The resulting figure is similar to a section of the cells in a beehive, hence the term cellular [13].

Due to the random nature of traffic and the inherent "elasticity" of data transfers in cellular systems, very few models have been explored about user performance for wireless data channel. Most existing models indeed represent data transfer as circuit services [1]. Some notable exceptions are the recent works by Borst for CDMA/HDR systems [11], and by Bonald and Proutière for cellular systems with Poissonian arrival and infinite user population [4, 8].

In the work by Borst, user performance is explicitly evaluated and shown to be insensitive to the flow size distribution in a symmetric scenario where all users experience the same fast fading and the resource allocation is that realized by the PF scheduler. The radio channel is modelled at flow level by a processor-sharing queue which is known to have the insensitivity property [19].

In the works by Bonald and Proutière, two kinds of models are evaluated, with the basic assumption that the customer number is infinite. Here we denote these two models Static Model and Dynamical Model, where the customers in different classes *don't and do move in the cellular system*, respectively. We will introduce them in the following.

3.1 Static Model

The static model is based on [8] by Bonald and Proutière.

3.1.1 Radio resource

In the model by Bonald and Proutière, radio resource is assumed to be time-shared between active users and the fraction of time base station b transmits to user u is denoted by ϕ_u^b, with $\sum_u \phi_u^b = 1$. Then the data rate of user u is:

$$R_u = R \times \phi_u^b, \tag{3.1}$$

where R is the peak data rate, obtained in the absence of any other users in the cell, i.e., for $\phi_u^b = 1$.

In the model, it is also assumed that the peak data rate R is approximately constant during data transfer and only depends on the distance r from BS b to user u. The customers in each annular ring between $r_k - dr_k$ and r_k are grouped into a single class k. The corresponding cell is illustrated in Figure 3.1. Note that the class k can be continuous or discrete.

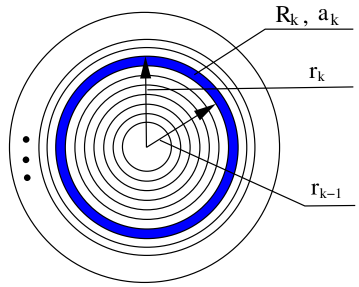

Figure 3.1: Ring shapes in the cell system

Denote R_1 the maximum peak rate and by r_1 the maximum distance at which this maximum peak rate is achieved:

$$R_k = R_1 \quad \text{for all}\, r_k \leq r_1. \tag{3.2}$$

The power P_u received by user u is equal to $P \times \Gamma_u$, where P is the transmission power of the BS and Γ_u denotes the path loss as follows:

$$\Gamma_u = \begin{cases} 1 & \text{if}\, r \leq \varepsilon, \\ \left(\frac{r_1}{r}\right)^\alpha & \text{otherwise,} \end{cases}$$

where r_1 denotes the maximum distance at which the full power P is received and α is the path loss exponent which characterizes the radio environment (typical values of α are between 2 and 5). Assuming that the maximum peak rate R_1 can be achieved, then the peak rate function is:

$$R_k = R_1 \times \begin{cases} 1 & \text{if}\, r_k \leq r_1, \\ \left(\frac{r_1}{r_k}\right)^\alpha & \text{otherwise,} \end{cases} \tag{3.3}$$

Equation (3.3) corresponds to an ideal case where the set of achievable peak rates is continuous. In practice, sometimes only a discrete set of peak rates is available, $R_1 > R_2 > \ldots > R_K$, i.e. K classes. In this case these rates define a set of concentric rings of external radius, $r_1 < r_2 < \ldots < r_K$, corresponding to regions where these rates are achievable.

3.1.2 Traffic characteristics

Assumptions about traffic characteristics in the model by Bonald et al.:

- Data flows arrive as a Poisson process with intensity $\lambda \times ds$ in any area of surface ds;
- The traffic demand is uniformly distributed;

- Flow sizes are independent and identically distributed with mean σ.

In the case of continuous peak rates, the traffic intensity generated by class-k customers is:

$$d\bar{\rho}_k = \lambda\,\sigma \times 2\pi r_k dr_k. \tag{3.4}$$

The load generated by class-k customers is the ratio of their traffic intensity to their peak rate, i.e.,

$$\rho_k = \frac{d\bar{\rho}_k}{R_k}. \tag{3.5}$$

Thus the load of a cell of radius r is:

$$\rho = \int_0^r \rho_k = \int_0^r \frac{d\bar{\rho}_k}{R_k}. \tag{3.6}$$

Similarly, in the case of discrete peak rates, the traffic intensity generated by class-k customers is:

$$\bar{\rho}_k = \lambda\,\sigma\pi(r_k^2 - r_{k-1}^2) \qquad \text{with} \quad r_0 = 0. \tag{3.7}$$

The load generated by class-k customers is:

$$\rho_k = \frac{\bar{\rho}_k}{R_k}. \tag{3.8}$$

The cell load is:

$$\rho = \sum_k \rho_k, \tag{3.9}$$

3.1.3 Throughput performance

We define the throughput γ_k of class-k customers whose distance from the BS is r_k, as the ratio of the mean flow size σ to the mean sojourn time T_k, i.e.

$$\gamma_k = \frac{\sigma}{T_k}. \tag{3.10}$$

In [3], Bonald et al. have deduced that in the case of single cell, the following result holds:

$$T_k = \frac{\sigma}{(1-\rho)R_k}. \tag{3.11}$$

Thus, we have:

$$\gamma_k = (1-\rho)R_k. \tag{3.12}$$

Note that class k can be continuous and discrete.

From equation (3.12), it can seen that the throughput decreases linearly with the cell load.

3.2 Dynamical Model

The dynamical model by Bonald et al. evaluated the impact of mobility on the flow-level performance of wireless data systems. The results show that for a broad class of fading processes, performance varies monotonically with the speed of the rate variations [4], improving with higher mobility.

In [4] Bonald et al. considered two limit regimes, termed *fluid* and *quasi-stationary*, obtained when rate variations have the same instantaneous statistics, but occur on an infinitely fast and an infinitely slow time scale, respectively. It is easy to analyze these regimes since their performance is insensitive. However, in practice, the mobility-induced fading evolves on an intermediate time scale, whose performance depends in some complicated fashion on the evolution of the queue itself. Such a system is not an ordinary PS queue and the insensitivity property of the PS queue can not be applied. Fortunately, the two limit regimes Bonald et al. considered provide simple performance bounds that only depend on easily calculated load factors and are remarkably tight in certain usual cases.

3.2.1 Model description

In the model, each class is characterized by the flow size (in bits) and its feasible transmission rate.

Assumptions:

- K classes, each class corresponding to given statistical flow size and rate variation characteristics;
- Class-k flows arrive as a Poisson process with rate λ_k;
- The size of class-k flow is represented as σ_k, and the size of the i-th arriving class-k flow is denoted by σ_{ki};
- The feasible rate at time t of class-k flow is represented as $R_k(t)$, and the feasible rate at time t of the i-th arriving class-k flow is denoted by $R_{ki}(t)$;

Here class k can be continuous and discrete.

If $R_k = \mathrm{E}[R_k(0)]$ is the time-average feasible rate of a class-k flow, then one has the following definition:

- The load of class-k flow: $\rho_k = \lambda_k \mathrm{E}[F_k]/R_k$.
- The total load: $\rho = \sum_{k=1}^{K} \rho_k$.

Assuming packet scheduling results in fair sharing at flow level, the actual transmission rate of the i-th arriving class-k flow, if present at time t, is [2]:

$$R_{ki}(t)\frac{1}{n}, \tag{3.13}$$

where n denotes the total number of flows present at time t.

The flow-level model defined by equation (3.13) corresponds to a Processor-Sharing type queue where the service rate of each customer is modulated by an independent stochastic process.

3.2.2 Throughput performance

Definition of fluid and quasi-stationary regimes

Consider the generic rate process for class-k flow,

$$R_k^{(s)}(t) \equiv R_k(st),$$

where s represents the speed of the rate process. In case $R_k(t)$ is a Markov process, the process $R_k^{(s)}(t)$ may be obtained by scaling the transition rate with s.

1. When $s \to \infty$, the variations completely vanish, and the rate process reduces to a constant. This corresponds to *fluid* regime, with

   $$R_k^{\text{fl}}(t) \equiv R_k^{(\infty)}(t) = R_k.$$

   - The load of class-k flow: $\rho_k^{\text{fl}} = \lambda_k \text{E}[\sigma_k]/R_k$.
   - The total load: $\rho \equiv \rho^{\text{fl}} = \sum_k \rho_k^{\text{fl}}$.

2. When $s \to 0$, the changes completely disappear, and the rate process freezes in some initial state. This corresponds to *quasi-stationary* regime, with

   $$R_k^{\text{qs}}(t) \equiv R_k^{(0)}(t) = R_k(0).$$

   - The load of class-k flow: $\rho_k^{\text{qs}} = \lambda_k \text{E}[\sigma_k/R_k(0)] = \lambda_k \text{E}[\sigma_k]/R_k^{(\text{qs})}$, where $R_k^{\text{qs}} = \text{E}[1/R_k(0)]^{-1}$.
   - The total load: $\rho^{\text{qs}} = \sum_k \rho_k^{\text{qs}}$.

By Jensen's inequality, we have:

$$\rho_k^{\text{fl}} \leq \rho_k^{\text{qs}}. \tag{3.14}$$

Throughput performance

In case of stability, the stationary distributions in the fluid and quasi-stationary regimes are represented as $\pi^{\text{fl}}(x_1, x_2, \ldots, x_K)$ and $\pi^{\text{qs}}(x_1, x_2, \ldots, x_K)$, where n_k is the number of on-going flows of class-k. They are given by:

$$\pi^{\text{fl}}(x_1, x_2, \ldots, x_K) = \pi^{\text{fl}}(0)\frac{x!}{\prod_{i=1}^{x} G(i)} \prod_{k=1}^{K} \frac{(\rho_k^{\text{fl}})^{x_k}}{x_k!}, \tag{3.15}$$

where $\pi^{\text{fl}}(0)$ is the normalization constant, and $x = \sum_k x_k$.

$$\pi^{\text{qs}}(x_1, x_2, \ldots, x_K) = \pi^{\text{qs}}(0)\frac{x!}{\prod_{i=1}^{x} G(i)} \prod_{k=1}^{K} \frac{(\rho_k^{\text{qs}})^{x_k}}{x_k!}, \tag{3.16}$$

where $\pi^{\text{qs}}(0)$ is the normalization constant, and $x = \sum_k x_k$.

In view of equations (3.15) and (3.16), the mean number of on-going flows of class-k is:

$$\mathrm{E}^{\text{fl}}[x_k] = \frac{\rho_k^{\text{fl}}}{1 - \rho^{\text{fl}}}. \tag{3.17}$$

$$\mathrm{E}^{\text{qs}}[x_k] = \frac{\rho_k^{\text{qs}}}{1 - \rho^{\text{qs}}}. \tag{3.18}$$

By Little's formula, one obtains the mean sojourn time of class-k flows:

$$\mathrm{E}[T_k] = \frac{\mathrm{E}[x_k]}{\lambda_k}. \tag{3.19}$$

Then the throughput of class-k flow is as follows:

$$\gamma_k \equiv \frac{\mathrm{E}[\sigma_k]}{\mathrm{E}[T_k]} = \frac{\rho_k \, R_k}{\mathrm{E}[x_k]}. \tag{3.20}$$

Thus one gets:

$$\gamma_k^{\text{fl}} = R_k(1 - \rho^{\text{fl}}), \tag{3.21}$$

$$\gamma_k^{\text{qs}} = R_k^{\text{qs}}(1 - \rho^{\text{qs}}). \tag{3.22}$$

The quasi-stationary and fluid regimes provide explicit performance estimates, which are provably optimistic and conservative, respectively. The results in [4] also shows that the performance in fluid regime is better than that in quasi-stationary regime.

Chapter 4

OCOF finite population model

Because of the random nature of cellular system very few models have been developed for wireless data channel. In recent years, Bonald et al. established some useful models in this field. However, the model by Bonald et al. is based on the following assumptions:

- Data flows arrive at BS as Poissonian process;
- The flow that arrives at BS selects the user in the cell with a uniform probability;
- The selected user moves in such a way that it occupies all points in the cell uniformly;
- The user population in the cell system is infinite. [1]

So in the model by Bonald et al., the movement doesn't change the user constellation.

Of course in the practice the user number in the cell system can not be infinite. In the thesis, we will develop another kind of model with finite user population. As in Figure 4.1, there are 8 users all together in the cell. At one instant of time, users make one constellation. User performance in this constellation could be calculated based on queueing network theory. At another instant of time, user positions make another constellation after moving. Therefore, we could get the final user performance by externally averaging over different user performances in different constellations.

In this model, we assume that only one flow is going for one customer at the same time. That is to say, a customer could receive or generate new flow only after the old flow ends. The model is so called OCOF(One Customer One Flow) finite population model.

[1] In this thesis we consider the mobile cell system with moblity and so we only discuss the dynamical model by Bonald et al.

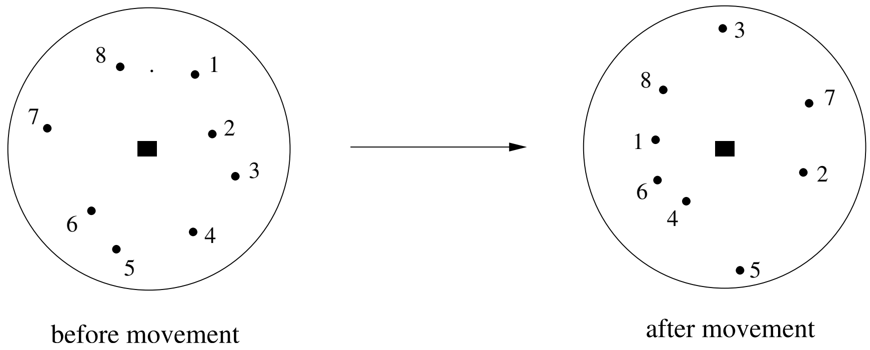

Figure 4.1: The user position constellation before and after movement

In the following session, we will model and calculate the throughput in a single cell mobile system:

Object: A single cell mobile system with M customers.

Model: A two-node closed network. Node 1 represents the BS and is modelled as a PS queue, formed by all active customers. Node 0 is modelled as a IS queue, formed by all thinking customers. In node 0, all the customers have the same service time distribution with mean $1/\nu$. The closed network modelled is shown in Figure 4.2.

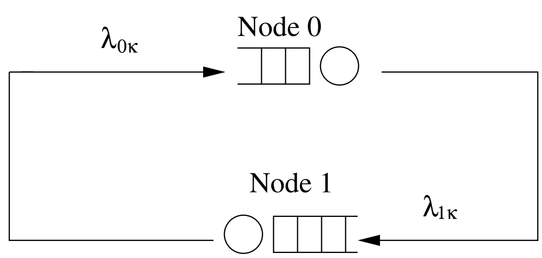

Figure 4.2: The closed network modelling the cell system

The customers in each annular ring are grouped into a single class. There are K classes in the system, with M_k maximum customer in each class k, $k = 1, 2, \ldots, K$. Denote $\mathbf{M} = (M_1, M_2, \ldots, M_K)$, and $\sum_{k=1}^{K} M_k = M$.

When there is only one class-k customer in the system, the capacity of node 1 (BS) is [4]:

$$\mu_k = \frac{R_k}{\sigma}, \tag{4.1}$$

where

- $R_k = R_0 \times \begin{cases} 1, & r_k \leq r_0 \\ \left(\frac{r_0}{r_k}\right)^{\alpha}, & \text{otherwise} \end{cases}$;
  R_0 is the maximum peak rate;

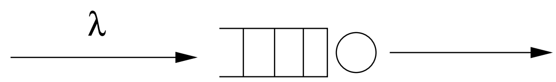

Figure 4.3: The queue model in the model by Bonald et al.

r_0 is the maximum distance at which R_0 is achieved;
σ is the mean flow size.

In the model by Bonald et al., only node 1 is considered, as shown in Figure 4.3:

State Description: For node 1,

$$\mathbf{x} = (x_1, x_2, \ldots, x_K),$$

where x_k is the customer number of class k in node 1.

For node 0,

$$\mathbf{y} = (y_1, y_2, \ldots, y_K),$$

where y_k is the customer number of class k in node 0.

We denote the state space of the network by $\varsigma(K)$. So $(\mathbf{x}, \mathbf{y}) \in \varsigma(K)$.

Note that:

- $M_k = x_k + y_k$;
- $\mathbf{M} = \mathbf{x} + \mathbf{y}$.

In the following sections we also denote:

- $x = |\mathbf{x}| = \sum_{k=1}^{K} x_k$;
- $y = |\mathbf{y}| = \sum_{k=1}^{K} y_k$.

4.1 Stationary distribution of the network state

There are two basic ways to get the state stationary distribution of the network, which will be discussed in the following:

4.1.1 BCMP theorem

As described in chapter 2, the state space is usually denoted with the detail information about service stages in the BCMP queueing networks, i.e.,

$$\mathbf{x} = (\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_K),$$

$$\mathbf{y} = (\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_K),$$

where $\mathbf{x}_k = (x_{k1}, x_{k2}, \ldots, x_{kJ_k})$ and $\mathbf{y}_k = (y_{k1}, y_{k2}, \ldots, y_{kJ_k})$, x_{ks} and y_{ks} are the customer number of class k and in stage s of their service at node 1 and node 0 respectively.

By the BCMP theorem [12],

$$\pi(\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_K) = x! \prod_{k=1}^{K} \prod_{s=1}^{J_k} \left[\frac{1}{x_{ks}!} \left(\frac{e_{1k} A_{1ks}}{\mu_{1ks}} \right)^{x_{ks}} \right],$$

$$\pi(\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_K) = \prod_{k=1}^{K} \prod_{s=1}^{J_k} \left[\frac{1}{y_{ks}!} \left(\frac{e_{0k} A_{0ks}}{\mu_{0ks}} \right)^{y_{ks}} \right].$$

Here, we do not require such a detailed solution. The information about stages could be removed by summing the above equations over all stages and using the multinomial theorem [12]. From the traffic equations, we can get $e_{0i} = e_{1i}$, where e_{jk} is the relative arrival rate of class k customer in node j, $k = 1, 2, \ldots, K, j = 0, 1$. Without loss of generality, we take $e_{0k} = e_{1k} = 1$. Thus we have:

For node 1,

$$\begin{aligned} \pi(\mathbf{x}) &= \pi(x_1, x_2, \ldots, x_K) \\ &= x! \prod_{k=1}^{K} \frac{1}{x_k!} \left(\frac{e_{1k}}{\mu_{1k}}\right)^{x_k} \\ &= x! \prod_{k=1}^{K} \frac{1}{x_k!} \rho_{1k}^{x_k}, \end{aligned}$$

where

- $\rho_{1k} = 1/\mu_{1k} = \sigma/R_k$;
- μ_{1k} is the full service rate available by a class-k customer in node 1.

For node 0,

$$\begin{aligned}\pi(\mathbf{y}) &= \pi(y_1, y_2, \ldots, y_K) \\ &= \prod_{k=1}^{K} \frac{1}{y_k!} \left(\frac{e_{0k}}{\mu_{0k}}\right)^{y_k} \\ &= \prod_{k=1}^{K} \frac{1}{y_k!} \rho_{0k}^{y_k},\end{aligned}$$

where

- $\rho_{0k} = 1/\nu$;
- μ_{0k} is the service rate available by a class-k customer in node 0.

Then stationary distribution of the whole network is:

$$\begin{aligned}\pi(\mathbf{x}, \mathbf{y}) &= \widehat{G}^{-1} \cdot x! \prod_{k=1}^{K} \frac{1}{x_k!} \rho_{1k}^{x_k} \cdot \prod_{k=1}^{K} \frac{1}{y_k!} \rho_{0k}^{y_k} \\ &= \widehat{G}^{-1} x! \prod_{k=1}^{K} \frac{1}{x_k!} \left(\frac{\sigma}{R_k}\right)^{x_k} \prod_{k=1}^{K} \frac{1}{(M_k - x_k)!} \left(\frac{1}{\nu}\right)^{M_k - x_k} \\ &= \widehat{G}^{-1} x! \prod_{k=1}^{K} \frac{1}{x_k!(M_k - x_k)!} \left(\frac{\nu\sigma}{R_k}\right)^{x_k} \\ &= \widehat{G}^{-1} x! \prod_{k=1}^{K} \frac{1}{x_k!(M_k - x_k)!} \rho_k^{x_k}, \end{aligned} \tag{4.2}$$

where

- $\rho_k = \frac{\nu\sigma}{R_k}$;
- $\widehat{G}^{-1} = \sum_{(\mathbf{x},\mathbf{y}) \in \varsigma(K)} x! \prod_{k=1}^{K} \frac{1}{x_k!(M_k - x_k)!} \rho_k^{x_k}$.

Alternatively, by absorbing a factor $1/\prod_{k=1}^{K} M_k!$ in the normalization constant we could write the network stationary distribution in the following form:

$$\pi(\mathbf{x}, \mathbf{y}) = G^{-1} x! \prod_{k=1}^{K} \binom{M_k}{x_k} \rho_k^{x_k}, \tag{4.3}$$

where

- $\rho_k = \frac{\nu\sigma}{R_k}$;
- $G^{-1} = \sum_{(\mathbf{x},\mathbf{y})\in\varsigma(K)} x! \prod_{k=1}^{K} \binom{M_k}{x_k} \rho_k^{x_k}$.

We denote the unnormalized state probability $\chi(\mathbf{x}) = x! \prod_{k=1}^{K} \binom{M_k}{x_k} \rho_k^{x_k}$.

We will use the later definition for the stationary distribution of the network in the following chapter.

Note: when vector $\mathbf{x}$ is given, vector $\mathbf{y}$ is determined accordingly by $\mathbf{y} = \mathbf{M} - \mathbf{x}$. So in the later chapter we denote the stationary distribution of the whole network by $\pi(\mathbf{x})$, which is not the same as one only for node 1 earlier. $\mathbf{x} \in \varsigma(K)$.

4.1.2 Whittle network theorem

Equivalently, the cell system could also be modelled as a Whittle network, where the capacity of each node depends on the global state of the network. In our case, the routing in the Whittle network is reducible and there are K irreducible closed subnetworks, $C_1, C_2, \ldots, C_K$. Each subnetwork C_k is formed by the class k and has the maximum customer number M_k. Each subnetwork consists of two nodes, node 1 and node 0. Node 0 is regarded as the source node, whose capacity is a function ψ_k of the number of customer at this node only, i.e., $\psi_k = y_k\,\nu$, for $k = 1, 2, \ldots, K$. The arrival rate satisfies that $e_{0k} = e_{1k}$, where e_{jk} is the arrival rate of class k customer at node j, $k = 1, 2, \ldots, K, j = 0, 1$. Without loss of generality, here we take it as 1. The state is described as $\mathbf{x} = (x_1, x_2, \ldots, x_K)$, where x_k is the customer number in node 1 of network C_k. The whole Whittle network is depicted in Figure 4.4.

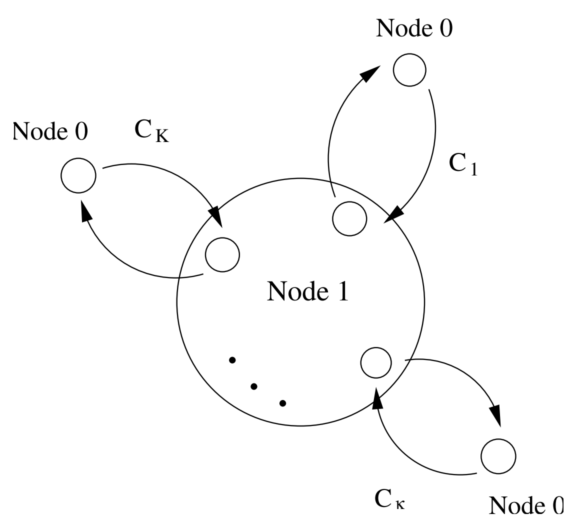

Figure 4.4: Modeling the cell system as a whittle network with K irreducible closed subnetworks

If the capacities of node 1s are balanced by a function Φ, based on Theorem 5 in Chapter 2, we could get the stationary distribution of the whole network as follows [6]:

$$\begin{aligned}\pi(\mathbf{x}) &= \Phi(\mathbf{x}) \prod_{k=1}^{K} \widetilde{\rho}_k^{x_k} \prod_{k=1}^{K} \prod_{n_k=1}^{y_k} \frac{\varrho_k}{\psi_k(n_k)} \\ &= \Phi(\mathbf{x}) \prod_{k=1}^{K} \left[\rho_k^{x_k} \prod_{n_k=1}^{y_k} \frac{\varrho_k}{\psi_k(n_k)} \right],\end{aligned}$$

where

- $\Phi(\mathbf{x}) = \frac{x!}{x_1!x_2!\cdots x_K!} \cdot \frac{1}{\left(\frac{R_1}{\sigma}\right)^{x_1}\left(\frac{R_2}{\sigma}\right)^{x_2}\cdots\left(\frac{R_K}{\sigma}\right)^{x_K}}$

  $\widetilde{\rho}_k = e_{1k}\,\sigma = \sigma$

  $\varrho_k = e_{0k}\,\sigma = \sigma$

  $\prod_{n_k=1}^{y_k} \frac{\varrho_k}{\psi_k(n_k)} = \frac{\sigma^{y_k}}{\nu^{y_k}\, y_k!} = \frac{\sigma^{y_k}}{\nu^{M_k - x_k}(M_k - x_k)!}$

Proof of $\Phi(\mathbf{x})$:

$$\frac{\Phi(\mathbf{x}-\mathbf{e}_k)}{\Phi(\mathbf{x})} = \frac{\frac{(x-1)!}{x_1!\cdots(x_i-1)!\cdots(x_K)!} \cdot \frac{1}{\left(\frac{R_1}{\sigma}\right)^{x_1}\cdots\left(\frac{R_k}{\sigma}\right)^{x_k-1}\cdots\left(\frac{R_K}{\sigma}\right)^{x_K}}}{\frac{x!}{x_1!\cdots x_i!\cdots(x_K)!} \cdot \frac{1}{\left(\frac{R_1}{\sigma}\right)^{x_1}\cdots\left(\frac{R_k}{\sigma}\right)^{x_k}\cdots\left(\frac{R_K}{\sigma}\right)^{x_K}}} = \frac{x_k}{x}\frac{R_k}{\sigma} = \phi_k(\mathbf{x}).$$

So,

$$\begin{aligned}\pi(\mathbf{x}) &= \frac{x!}{x_1!x_2!\cdots x_K!} \frac{1}{\left(\frac{R_1}{\sigma}\right)^{x_1}\left(\frac{R_2}{\sigma}\right)^{x_2}\cdots\left(\frac{R_K}{\sigma}\right)^{x_K}} \prod_{k=1}^{K} \sigma^{x_k} \frac{\sigma^{y_k}}{\nu^{M_k-x_k}(M_k-x_k)!} \\ &= x! \prod_{k=1}^{K} \frac{1}{x_k!(M_k-x_k)!} \left(\frac{\nu\sigma}{R_k}\right)^{x_k} \cdot \frac{\sigma^M}{\nu^{M_k}}. \qquad (4.4)\end{aligned}$$

After normalization, we get the following stationary distribution of the whole network:

$$\begin{aligned}\pi(\mathbf{x}) &= \widehat{G}^{-1} x! \prod_{k=1}^{K} \frac{1}{x_k!(M_k-x_k)!} \left(\frac{\nu\sigma}{R_k}\right)^{x_k} \\ &= \widehat{G}^{-1} x! \prod_{k=1}^{K} \frac{1}{x_k!(M_k-x_k)!}\, \rho_k^{x_k}, \qquad (4.5)\end{aligned}$$

where

- $\rho_k = \frac{\nu\sigma}{R_k}$;
- $\widehat{G}^{-1} = \sum_{(\mathbf{x},\mathbf{y})\in\varsigma(K)} x! \prod_{k=1}^{K} \frac{1}{x_k!(M_k-x_k)!} \rho_k^{x_k}$.

As in BCMP theorem above, we can also write the network stationary distribution in the following form:

$$\pi(\mathbf{x}) = G^{-1} x! \prod_{k=1}^{K} \binom{M_k}{x_k} \rho_k^{x_k}, \tag{4.6}$$

where

- $\rho_k = \frac{\nu\sigma}{R_k}$;
- $G^{-1} = \sum_{(\mathbf{x},\mathbf{y})\in\varsigma(K)} x! \prod_{k=1}^{K} \binom{M_k}{x_k} \rho_k^{x_k}$.

4.2 Two special cases

Case 1 All the customers in node 1 have the same service time distribution with mean $\mu = \frac{R}{\sigma}$. So there is only one class in the system.

From equations (4.3) or (4.6), we could get the customer number distribution of the network,

$$\chi(x) = x! \frac{1}{x!(M-x)!} \left(\frac{\nu\sigma}{R}\right)^x = \frac{1}{(M-x)!} \rho^x, \tag{4.7}$$

where $\rho = \frac{\nu\sigma}{R}$.

For this simple case, we can also analyze it from the point of a simple Markov chain as follows, assuming that both thinking time and flow size have exponential distribution,

For node 1,

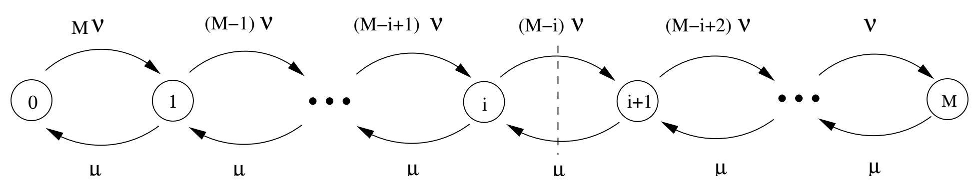

Figure 4.5: the Markov chain for node 1

$$\pi_i \cdot \mu = \pi_{i-1} \cdot (M-i+1)\nu \Rightarrow \pi_i = \frac{1}{(M-i)!} \rho_i \cdot M! \pi_0,$$

where $\rho = \frac{\nu\sigma}{R}$.

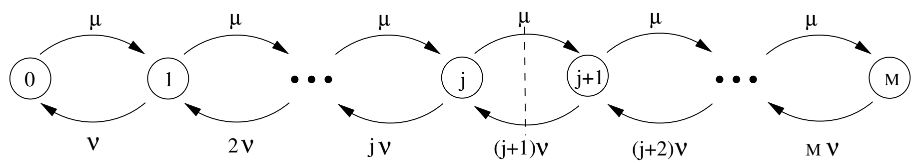

Figure 4.6: the Markov chain for node 0

For node 0,

$$\pi_j \cdot j\nu = \pi_{j-1} \cdot \mu \Rightarrow \pi_j = \frac{1}{j!}(\frac{1}{\rho})^j \cdot \pi_0 = \frac{1}{(M-i)!}\rho^i \cdot \frac{1}{\rho^M}\pi_0,$$

where $\rho = \frac{\nu\sigma}{R}$.

It can be seen that the results from the Markov chain analysis are the same as what we obtained (see 4.7) from the earlier formula.

Case 2 Each class in the system has only one customer, $M_k = 1$, and M = 3. From the equation (4.3) or (4.6), it can be concluded that,

$$\chi(\mathbf{x}) = x! \prod_{k=1}^{K} \rho_k^{x_k}.$$

$$\begin{array}{ll} \chi(0,0,0) = 1 & \chi(1,0,0) = \rho_1 \\ \chi(0,1,0) = \rho_2 & \chi(0,0,1) = \rho_3 \\ \chi(1,1,0) = 2\rho_1\rho_2 & \chi(1,0,1) = 2\rho_1\rho_3 \\ \chi(0,1,1) = 2\rho_2\rho_3 & \chi(1,1,1) = 6\rho_1\rho_2\rho_3 \end{array}$$

The behavior of on-off type customer i is depicted in Figure 4.7. From Figure 4.7, we find that:

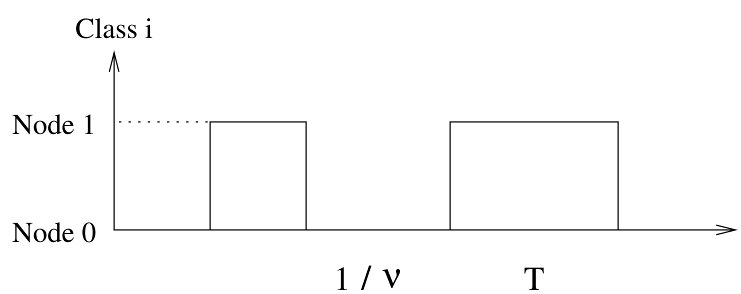

Figure 4.7: On-off type customer i

$$\frac{T_k}{T_k + 1/\nu} = p_{1k} = \mathrm{E}[X_k] = p_{0k} \cdot 0 + p_{1k} \cdot 1, \tag{4.8}$$

where T_k is the sojourn time of class k customer in node 1. p_{1k} is the probability that customer k is in node 1. X_k is random variable presenting the customer number of class k in node 1. x_k is its possible value and $\mathrm{E}[X_k]$ is its mean value.

From equation (4.8), we get,

$$T_k = \frac{p_{1k}}{1-p_{1k}} \cdot \frac{1}{\nu},$$
$$\gamma_k = \frac{1-p_{1k}}{p_{1k}} \cdot \nu\sigma. \tag{4.9}$$

Take class 1 as an example,

$$\begin{aligned} T_1 &= \frac{p_{11}}{1-p_{11}} \cdot \frac{1}{\nu} \\ &= \frac{1+2\rho_2+2\rho_3+6\rho_2\rho_3}{1+\rho_2+\rho_3+2\rho_2\rho_3} \cdot \frac{\rho_1}{\nu}. \end{aligned}$$

$$\begin{aligned} \gamma_1 &= \frac{\sigma}{T_1} \\ &= \frac{1+\rho_2+\rho_3+2\rho_2\rho_3}{1+2\rho_2+2\rho_3+6\rho_2\rho_3} \cdot \frac{\nu\sigma}{\rho_1} \\ &= \frac{1+\rho_2+\rho_3+2\rho_2\rho_3}{1+2\rho_2+2\rho_3+6\rho_2\rho_3} \cdot R_1 \\ &= \frac{G[0,1,1]}{\frac{\partial}{\partial\rho_1}G[1,1,1]} \cdot R_1, \end{aligned}$$

where $G[M_1, M_2, M_3]$ is the normalizing constant with the maximum customer number M_k in each class k, $k = 1, 2, 3$.

If the loads of class 2 and 3 in the system are very low, the throughput of class 1 is $\gamma_1 = R_1$;

If the loads of class 2 and 3 in the system are very high, the throughput of class 1 is $\gamma_1 = \frac{1}{3}R_1$, which is consistent with the PS principle of node 1.

4.3 The average throughput of class k

Definition 1 The throughput (γ_k) of class k in the cell system is the average bandwidth available for a class k flow when it is being sent.

With little formula [14], we have:

$$\mathrm{E}[X_k] = \lambda_{1k}\mathrm{E}[T_{1k}], \tag{4.10}$$

where λ_{1k} is the arrival rate of class k customer to node 1. $E[T_{1k}]$ is the average sojourn time spent by class k customer in node 1.

$$\gamma_k = \frac{\sigma}{\mathrm{E}[T_{1k}]} = \frac{\lambda_{1k}\,\sigma}{\lambda_{1k}\mathrm{E}[T_{1k}]} = \frac{\lambda_{1k}\,\sigma}{\mathrm{E}[X_k]}, \tag{4.11}$$

To calculate the throughput of class k customer, we need to know how to get λ_{1k}. It is easy to find that:

$$\lambda_{1k} = M_k \frac{G[\mathbf{M}-\mathbf{e}_k]}{G[\mathbf{M}]}\,\nu.^2 \tag{4.12}$$

The proof of equation (4.12) is in the following,

Considering the queue in node 1, we calculate the average rate class-k customers are leaving this node,

$$\begin{aligned}
\lambda_{1k} &= \sum_{\mathbf{x}:x_k>0} \pi(\mathbf{x}) \cdot \frac{x_k}{x}\frac{R_k}{\sigma} \\
&= \frac{1}{G[\mathbf{M}]} \sum_{\mathbf{x}:x_k>0} x! \prod_{k=1}^{K} \binom{M_k}{x_k} \rho_k^{x_k} \cdot \frac{x_k}{x}\frac{R_k}{\sigma} \\
&= \frac{1}{G[\mathbf{M}]} \sum_{\mathbf{x}:x_k>0} \prod_{j\neq k} \frac{M_j!}{x_j!(M_j-x_j)!}\,\rho_j^{x_j} \cdot x! \frac{M_k!}{x_k!(M_k-x_k)!}\,\rho_k^{x_k} \cdot \frac{x_k}{x}\frac{R_k}{\sigma} \\
&= \frac{1}{G[\mathbf{M}]} \sum_{\mathbf{x}:x_k>0} \prod_{j\neq k} \frac{M_j!}{x_j!(M_j-x_j)!}\,\rho_j^{x_j} \cdot (x-1)! \\
&\quad \frac{M_k!}{(x_k-1)!((M_k-1)-(x_k-1))!}\,\rho_k^{x_k-1} \cdot \rho_k \frac{R_k}{\sigma} \\
&= M_k \cdot \frac{1}{G[\mathbf{M}]} \sum_{\mathbf{x}:x_k>0} (x-1)! \prod_{j\neq k} \frac{M_j!}{x_j!(M_j-x_j)!}\,\rho_j^{x_j} \\
&\quad \cdot \frac{(M_k-1)!}{(x_k-1)!((M_k-1)-(x_k-1))!} \cdot \rho_k^{x_k-1} \cdot \nu \\
&= M_k \frac{G[\mathbf{M}-\mathbf{e}_k]}{G[\mathbf{M}]}\,\nu.
\end{aligned}$$

Alternatively we can also prove it based on the queue in node 0. First we rewrite the stationary distribution of the whole network as follows:

$$\pi(\mathbf{x}) = \frac{1}{G[\mathbf{M}]}\,(M-y)! \prod_{k=1}^{K} \binom{M_k}{y_k} \rho_k^{M_k-y_k} \qquad (x_k = M_k - y_k),$$

[2] If $\widehat{G}[\mathbf{M}]$ had been adopted as the definition of normalization constant (see section 4.1), the factor M_k would not appear here.

and calculate the average rate class-k customers are leaving node 0,

$$\begin{aligned}
\lambda_{1k} &= \lambda_{0k} \\
&= \sum_{\mathbf{x}} \pi(\mathbf{x}) \cdot y_k \, \nu \\
&= \frac{1}{G[\mathbf{M}]} \sum_{\mathbf{y}: y_k > 0} (M-y)! \prod_{k=1}^{K} \binom{M_k}{y_k} \rho_k^{M_k - y_k} \cdot y_k \, \nu \\
&= \frac{1}{G[\mathbf{M}]} \sum_{\mathbf{y}: y_k > 0} (M-y)! \prod_{j \neq k} \frac{M_j!}{y_j!(M_j - y_j)!} \rho_j^{M_j - y_j} \cdot \frac{M_k!}{y_k!(M_k - y_k)!} \rho_k^{M_k - y_k} \cdot y_k \, \nu \\
&= M_k \cdot \frac{1}{G[\mathbf{M}]} \sum_{\mathbf{y}: y_k > 0} ((M-1)-(y-1))! \prod_{j \neq k} \frac{M_j!}{y_j!(M_j - y_j)!} \rho_j^{M_j - y_j} \\
& \quad \cdot \frac{(M_k - 1)!}{(y_k - 1)!((M_k - 1) - (y_k - 1))!} \rho_k^{(M_k - 1)-(y_k - 1)} \cdot \nu \\
&= M_k \frac{G[\mathbf{M} - \mathbf{e}_k]}{G[\mathbf{M}]} \nu.
\end{aligned}$$

We could also understand equation (4.12) in two other ways, see Appendix A.

Theorem 4.3.1 *The throughput of class k in general case is,*

$$\gamma_k = M_k \frac{G[\boldsymbol{M} - \boldsymbol{e}_k]}{\frac{\partial}{\partial \rho_k} G[\boldsymbol{M}]} R_k, \tag{4.13}$$

where $G[M_1, \cdots, M_K]$ is the normalizing constant with M_k customers in class k, $k = 1, 2, \ldots, K$.

The proof of the above theorem is as follows,

$$G[\mathbf{M}] = \sum_{\mathbf{x}} x! \prod_{k=1}^{K} \frac{M_k!}{x_k!(M_k - x_k)!} \rho_k^{x_k} \Rightarrow$$

$$\frac{\partial}{\partial \rho_k} G[\mathbf{M}] = \sum_{\substack{\mathbf{x} \\ x_k \geq 1}} x! \cdot \frac{M_k!}{x_k!(M_k - x_k)!} x_k \, \rho_k^{x_k - 1} \prod_{j \neq k} \frac{M_j!}{x_j!(M_j - x_j)!} \rho_j^{x_j} \Rightarrow$$

$$\rho_k \frac{\partial}{\partial \rho_k} G[\mathbf{M}] = \sum_{\mathbf{x}} x_k \cdot x! \prod_{j} \frac{M_j!}{x_j!(M_j - x_j)!} \rho_j^{x_j} \Rightarrow$$

$$\mathrm{E}[X_k] = \sum_{\mathbf{x}} x_k \cdot G[\mathbf{M}]^{-1} x! \prod_{j} \frac{M_j!}{x_j!(M_j - x_j)!} \rho_j^{x_j} = \frac{\frac{\partial}{\partial \rho_k} G[\mathbf{M}]}{G[\mathbf{M}]} \rho_k.$$

With the help of equations (4.11) and (4.12), we can get,

$$\gamma_k = \frac{\lambda_{1k} \, \sigma}{\mathrm{E}[X_k]} = M_k \frac{G[\mathbf{M} - \mathbf{e}_k]}{\frac{\partial}{\partial \rho_k} G[\mathbf{M}]} \cdot \frac{\nu \sigma}{\rho_k} = M_k \frac{G[\mathbf{M} - \mathbf{e}_k]}{\frac{\partial}{\partial \rho_k} G[\mathbf{M}]} R_k.$$

Note that γ_k here relates to the vector $\mathbf{M}$. So in the following section we will denote the throughput as $\gamma_k[\mathbf{M}]$.

Now let us calculate the throughput $\gamma_k, k = 1, 2$ in the following case:

Case 3 where

$n = 4$	$K = 2$
$r = 10\,\text{km}$	$r_1 = 4\,\text{km}$
$R_1 = 40\,\text{Mbits/s}$	$R_2 = 20\,\text{Mbits/s}$
$\nu = 0.5\,/\text{s}$	$\sigma = 10\,\text{Mbits}$

By calculating, we can get the following results,

$\rho_1 = 0.1250$			$\rho_2 = 0.2500$
M_1	M_2	γ_1^{qs}	γ_2^{qs}
0	4	0.0000	11.0938
1	3	22.1875	12.0588
2	2	24.1176	13.1098
3	1	26.2195	14.2481
4	0	28.4962	0.0000

Table 4.1: The results of the throughput for each class in Case 3

4.4 The average throughput of the cell system in the quasi-stationary regime

Now let us consider the average throughput in the whole cell system in case of quasi-stationary regime, where the rate variations occur on an infinitely slow time scale. Assume a discrete set of peak rates is available. The cell with the area A and the radius r is divided into K rings. Each ring $k, k = 1, 2, \ldots, K$, has the area of a_k and the radius between r_{k-1} and r_k with $r_0 = 0$. The customer obtains the same peak rate R_k in each ring k. So we group the traffic in ring k into the class k. The cell is illustrated in Figure 3.1. The customers are also assumed to be distributed uniformly in the cell.

Then the probability that a customer is in class k is,

$$p_k = \frac{a_k}{A} = \frac{\pi(r_k^2 - r_{k-1}^2)}{\pi r^2} = \frac{r_k^2 - r_{k-1}^2}{r^2}, \quad \text{with} \quad r_0 = 0. \tag{4.14}$$

Denote vector $\mathbf{M} = (M_1, M_2, \ldots, M_K)$, where M_k indicates the customer number in class k. There are altogether n customers in the system, i.e. $\sum_{k=1}^{K} M_k = n$.

Now the M_k are random variables. Let $\mathbf{m} = (m_1, m_2, \ldots, m_K)$ be a possible value of $\mathbf{M}$.

The random vector $\mathbf{M}$ follows the multinomial distribution [12],

$$\begin{aligned} \mathbb{P}[\mathbf{M} = \mathbf{m}] &= p_r\{\mathbf{m}\} \\ &= p_r\{m_1, m_2, \ldots, m_K\} \\ &= \binom{n}{m_1, m_2, \ldots, m_K} p_1^{m_1} p_2^{m_2} \cdots p_K^{m_K} \\ &= \frac{n!}{m_1! m_2! \cdots m_K!} p_1^{m_1} p_2^{m_2} \cdots p_K^{m_K} \quad \text{for } |\mathbf{m}| = n. \end{aligned} \tag{4.15}$$

The average time spent by a class-k customer in node 1, conditioned on $\mathbf{M} = \mathbf{m}$, is:

$$\bar{T}_k[\mathbf{m}] = \mathrm{E}[T_k[\mathbf{m}]] = \frac{\sigma}{\gamma_k[\mathbf{m}]}. \tag{4.16}$$

The average time spent by an arbitrary customer in node 1, conditioned on $\mathbf{M} = \mathbf{m}$, is:

$$\bar{T}[\mathbf{m}] = \mathrm{E}[T[\mathbf{m}]] = \frac{1}{n} \sum_k m_k \bar{T}_k[\mathbf{m}]. \tag{4.17}$$

Then the average time spent by a customer in node 1 is:

$$\begin{aligned} \mathrm{E}[T] &= \mathrm{E}[\mathrm{E}\,[\mathrm{T} \,|\, \mathbf{M} = \mathbf{m}]] \\ &= \sum_{\mathbf{m}} p_r\{\mathbf{m}\} \mathrm{E}\,[\mathrm{T} \,|\, \mathbf{M} = \mathbf{m}] \\ &= \sum_{\mathbf{m}} p_r\{\mathbf{m}\} \mathrm{E}[T[\mathbf{m}]] \\ &= \sum_{\mathbf{m}} \frac{n!}{m_1! m_2! \cdots m_K!} p_1^{m_1} p_2^{m_2} \cdots p_K^{m_K} \cdot \frac{1}{n} \sum_k m_k \bar{T}_k[\mathbf{m}] \\ &= (n-1)! \sum_{\mathbf{m}} \left[\frac{1}{m_1! m_2! \cdots m_K!} p_1^{m_1} p_2^{m_2} \cdots p_K^{m_K} \cdot \sum_k m_k \bar{T}_k[\mathbf{m}] \right]. \end{aligned} \tag{4.18}$$

The average throughput of the whole cell system is:

$$\mathrm{E}[\gamma^{\mathrm{qs}}] = \frac{\sigma}{\mathrm{E}[T]}. \tag{4.19}$$

Consider again the situation in Case 3:

Case 4 where,

$n = 4$ $K = 2$
$r = 10\,\text{km}$ $r_1 = 4\,\text{km}$
$R_1 = 40\,\text{Mbits/s}$ $R_2 = 20\,\text{Mbits/s}$
$\nu = 0.5\,/\text{s}$ $\sigma = 10\,\text{Mbits}$

		$p_1 = 0.1600$			$p_2 = 0.8400$			
M_1	M_2	γ_1^{qs}	T_1	γ_2^{qs}	T_2	$\bar{T}[\mathbf{m}]$	$\text{E}[T]$	$\text{E}[\gamma^{\text{qs}}]$
0	4	0.0000	0.0000	11.0938	0.9014	0.9014		
1	3	22.1875	0.4507	12.0588	0.8293	0.7346		
2	2	24.1176	0.4146	13.1098	0.7628	0.5887	0.7978	12.5339
3	1	26.2195	0.3814	14.2481	0.7018	0.4615		
4	0	28.4962	0.3509	0.0000	0.0000	0.3509		

Table 4.2: The average throughput in case 4

By calculation, we can get the results, shown in Table 4.2.

Could we obtain a recursive equation for the throughput? To do this, what we should do is to calculate the normalization constant $G[\mathbf{m}]$ and its derivative $\frac{\partial}{\partial \rho_i} G[\mathbf{m}]$ for γ_k, (this section needs the further work) :-)

One way is to calculate the normalization constant $\widehat{G}[\mathbf{m}]$ (see section 4.1),

$$\widehat{G}[\mathbf{m}] = \widehat{G}[m_1, m_2, \ldots, m_K] = \sum_{x_1=0}^{m_1} \sum_{x_2=0}^{m_2} \cdots \sum_{x_K=0}^{m_K} x! \prod_{k=1}^{K} \frac{1}{x_k!(m_k - x_k)!} \rho_k^{x_k}.$$

Denote $f(z) = f_1(z) f_2(z) \cdots f_K(z) = a_0 + a_1 z^1 + a_2 z^2 + \cdots + a_m^m$, where

$$f_k(z) = \sum_{x_k=0}^{m_k} \frac{1}{x_k!(m_k - x_k)!} \rho_k^{x_k} z^{x_k}.$$

Then,

$$\widehat{G}[\mathbf{m}] = \sum_{k=1}^{m} k! a_k. \tag{4.20}$$

Maybe it is necessary to use the following generating function of $\widehat{G}[\mathbf{m}]$ to get the final solution:

$$\widetilde{G}[\mathbf{z}] = \sum_{\mathbf{m}=\mathbf{0}}^{\infty} G[\mathbf{m}] z_1^{m_1} z_2^{m_2} \cdots z_K^{m_K} = \frac{e^{z_1 + z_2 + \cdots + z_K}}{1 - (\rho_1 z_1 + \rho_2 z_2 + \cdots + \rho_K z_K)}. \tag{4.21}$$

The proof of equation (4.21),

$$
\begin{aligned}
\widetilde{G}[\mathbf{z}] &= \sum_{\mathbf{m}=\mathbf{0}}^{\infty} \sum_{\mathbf{x}=\mathbf{0}}^{\infty} x! \prod_{k=1}^{K} \frac{1}{x_k!(m_k - x_k)!} \rho_k^{x_k} \cdot z_k^{m_k} \\
&= \sum_{\mathbf{m}=\mathbf{0}}^{\infty} \sum_{\mathbf{x}=\mathbf{0}}^{\infty} x! \prod_{k=1}^{K} \frac{1}{x_k!(m_k - x_k)!} \rho_k^{x_k} \cdot z_k^{x_k} z_k^{m_k - x_k} \\
&= \sum_{\mathbf{m}=\mathbf{0}}^{\infty} \sum_{\mathbf{x}=\mathbf{0}}^{\infty} x! \prod_{k=1}^{K} \frac{(\rho_k\, z_k)^{x_k}\, z_k^{m_k - x_k}}{x_k!(m_k - x_k)!} \\
&= \sum_{\mathbf{x}=\mathbf{0}}^{\infty} \sum_{\mathbf{m}=\mathbf{x}}^{\infty} x! \prod_{k=1}^{K} \frac{(\rho_k\, z_k)^{x_k}}{x_k!} \cdot \frac{z_k^{m_k - x_k}}{(m_k - x_k)!} \\
&= \sum_{\mathbf{x}=\mathbf{0}}^{\infty} \sum_{\mathbf{m}'=\mathbf{0}}^{\infty} x! \prod_{k=1}^{K} \frac{(\rho_k\, z_k)^{x_k}}{x_k!} \cdot \frac{z_k^{m'_k}}{(m'_k)!} \quad (\text{let } m'_i = m_i - x_i) \\
&= \sum_{\mathbf{x}=\mathbf{0}}^{\infty} x! \prod_{k=1}^{K} \frac{(\rho_k\, z_k)^{x_k}}{x_k!} \cdot \sum_{\mathbf{m}'=\mathbf{0}}^{\infty} \frac{z_k^{m'_k}}{(m'_k)!} \\
&= e^{z_1 + z_2 + \cdots + z_K} \sum_{\mathbf{x}=\mathbf{0}}^{\infty} x! \prod_{k=1}^{K} \frac{(\rho_k\, z_k)^{x_k}}{x_k!} \\
&= e^{z_1 + z_2 + \cdots + z_K} \sum_{x=0}^{\infty} \sum_{\mathbf{x}:|\mathbf{x}|=x} \frac{x!}{x_1! x_2! \cdots x_K!} (\rho_1\, z_1)^{x_1} (\rho_2\, z_2)^{x_2} \cdots (\rho_K\, z_K)^{x_K} \\
&= e^{z_1 + z_2 + \cdots + z_K} \sum_{x=0}^{\infty} (\rho_1\, z_1 + \rho_2\, z_2 + \cdots + \rho_K\, z_K)^{x} \\
&= \frac{e^{z_1 + z_2 + \cdots + z_K}}{1 - (\rho_1 z_1 + \rho_2 z_2 + \cdots + \rho_K z_K)}.
\end{aligned}
$$

Another way is to calculate the normalizaiton constant $G[\mathbf{m}]$ (see section 4.1),

$$\begin{aligned}
G[\mathbf{m}] &= G[m_1, m_2, \cdots, m_K] \\
&= \sum_{x=0}^{n} x! \sum_{\sum_i x_i = x} \prod_{k=1}^{K} \frac{m_k!}{x_k!(m_k - x_k)!} \rho_k^{x_k} \\
&= 1 + \sum_{x=1}^{n-1} x! \sum_{\sum_i x_i = x} \prod_{k=1}^{K} \frac{m_k!}{x_k!(m_k - x_k)!} \rho_k^{x_k} + n! \prod_{k=1}^{K} \rho_k^{m_k} \qquad (4.22) \\
&\quad \text{(difficult to get recursive relationship from here)} \\
&= \sum_{x=0}^{n-m_k} x! \cdot \sum_{\sum_i x_i = x} \prod_{j \neq k} \frac{m_j!}{x_j!(m_j - x_j)!} \rho_j^{x_j} + \\
&\quad \sum_{x=0}^{n-m_k} (x+1)! \binom{m_k}{1} \rho_k \cdot \sum_{\sum_i x_i = x} \prod_{j \neq k} \frac{m_j!}{x_j!(m_j - x_j)!} \rho_j^{x_j} + \\
&\quad \sum_{x=0}^{n-m_k} (x+2)! \binom{m_k}{2} \rho_k^2 \cdot \sum_{\sum_i x_i = x} \prod_{j \neq k} \frac{m_j!}{x_j!(m_j - x_j)!} \rho_j^{x_j} + \cdots + \\
&\quad \sum_{x=0}^{n-m_k} (x+m_k)! \binom{m_k}{m_k} \rho_k^{m_k} \cdot \sum_{\sum_i x_i = x} \prod_{j \neq k} \frac{m_j!}{x_j!(m_j - x_j)!} \rho_j^{x_j} \\
&= \sum_{x=0}^{n-m_k} x! \left[\binom{m_k}{0} \rho_k^0 + (x+1)! \binom{m_k}{1} \rho_k^1 + \cdots + (x+m_k)! \binom{m_k}{m_k} \rho_k^{m_k} \right] \\
&\quad \sum_{\sum_i x_i = x} \prod_{j \neq k} \frac{m_j!}{x_j!(m_j - x_j)!} \rho_j^{x_j} \\
&= \sum_{x=0}^{n-m_k} \sum_{l=0}^{m_k} (x+l)! \binom{m_k}{l} \rho_k^l \sum_{\sum_i x_i = x} \prod_{j \neq k} \frac{m_j!}{x_j!(m_j - x_j)!} \rho_j^{x_j} \\
&= \sum_{x=0}^{n-m_k} \sum_{l=0}^{m_k} \frac{(x+l)!}{x!} \binom{m_k}{l} \rho_k^l \cdot x! \sum_{\sum_i x_i = x} \prod_{j \neq k} \frac{m_j!}{x_j!(m_j - x_j)!} \rho_j^{x_j}. \qquad (4.23)
\end{aligned}$$

In the sum of the above equation, the second term is G[$\mathbf{m}$] with class k less. As for the first term, if we let

$$f(x, m_k) = \sum_{l=0}^{m_k} \frac{(x+l)!}{x!} \binom{m_k}{l} \rho_k^l.$$

then it is easy to get the recursive eqution for $f(x, m_k)$ as follows:

$$f(x+1, m_k) = f(x, m_k) + m_k \rho_k f(x+1, m_k - 1). \qquad (4.24)$$

Now it is possible for us to get the recursive equation about $G[\mathbf{m}]$.

$$
\begin{aligned}
\frac{\partial}{\partial \rho_k} G[\mathbf{m}] &= \sum_{\substack{\mathbf{x} \\ x_k \geq 1}} x! \cdot \frac{m_k!}{x_k!(m_k - x_k)!} x_k \, \rho_k^{x_k - 1} \cdot \prod_{j \neq k} \frac{m_j!}{x_j!(m_j - x_j)!} \rho_j^{x_j} \\
&= \sum_{\substack{\mathbf{x} \\ x_k \geq 1}} x! \, \frac{m_k(m_k-1)!}{(x_k-1)!((m_k-1)-(x_k-1))!} \rho_k^{x_k-1} \cdot \prod_{j \neq k} \frac{m_j!}{x_j!(m_j - x_j)!} \rho_j^{x_j} \\
&\overset{x_k - 1 \to x'_k}{=} \sum_{\substack{\mathbf{x}-\mathbf{e}_k \\ x_k \geq 1}} (x'+1)! \, \frac{m_k m'_k!}{x'_k!(m'_k - x'_k)!} \rho_k^{x'_k} \cdot \prod_{j \neq k} \frac{m_j!}{x_j!(m_j - x_j)!} \rho_j^{x_j} \\
& \quad \text{where } \mathbf{m}' = \mathbf{m} - \mathbf{e}_k \\
&= m_k \sum_{\substack{\mathbf{x}-\mathbf{e}_k \\ x_k \geq 1}} (x'+1)x'! \prod_{j} \frac{m_j!}{x_j!(m_j - x_j)!} \rho_j^{x_j} \\
&= m_k G[\mathbf{m} - \mathbf{e}_k] + \sum_{\mathbf{x}' = \mathbf{x} - \mathbf{e}_k} x' \pi(\mathbf{x}') \\
&= m_k G[\mathbf{m} - \mathbf{e}_k] + \sum_{i=1}^{m-1} \left[i \cdot G^{-1} \frac{\rho^i}{(m-1-i)!} \right],
\end{aligned} \tag{4.25}
$$

where $\rho = \frac{\nu\sigma}{R}$, $\quad G^{-1} = \left(\sum_{i=0}^{m-1} \frac{\rho^i}{(m-1-i)!} \right)^{-1}$

4.5 The average throughput of the cell system in the fluid regime

In the fluid regime, the user moves on an infinitely fast time scale. The rate variations completely vanish and the rate process reduces to a constant, denoted by R. Similarly as in the quasi-stationary regime, we assume that there are K classes and n customers in the system. Then,

$$
R = \sum_{k=1}^{K} p_k R_k, \tag{4.26}
$$

where p_k is calculated as equation (4.14).

In the case of the fluid regime, all customers fall into one class. The queueing network model is shown in Figure 4.8.

From equation (4.11), we get:

$$
\mathrm{E}[\gamma^{\mathrm{fl}}] = \frac{\lambda_1 \sigma}{\mathrm{E}[X]}, \tag{4.27}
$$

where,

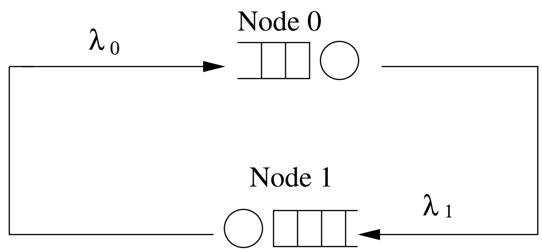

Figure 4.8: The closed network mode of the cell system in the fluid regime

- $\lambda_1 = \sum_{\mathbf{x}} (n-x)\pi(\mathbf{x})\,\nu$ or $\lambda_1 = n\,\frac{G[n-1]}{G[n]}\,\nu$;
- $\mathrm{E}[\mathbf{x}] = \sum_{\mathbf{x}} x\pi(\mathbf{x})$;
- $\pi(\mathbf{x}) = G^{-1}x!\binom{n}{x}\rho^x = G^{-1}\frac{n!}{(n-x)!}\,\rho^x$;
- $\rho = \frac{\nu\sigma}{R}$.

Alternatively we could also get the throughput in the fluid regime from equation (4.13),

$$\mathrm{E}[\gamma^{\text{fl}}] = n\,\frac{G[n-1]}{\frac{d}{d\rho}G[n]}\,R, \tag{4.28}$$

where $\rho = \frac{\nu\sigma}{R}$.

With the same condition in Case 3, we could calculate the throughput in the fluid regime as follows:

Case 5 where

$n = 4$	$K = 2$
$r = 10\,\text{km}$	$r_1 = 4\,\text{km}$
$R_1 = 40\,\text{Mbits/s}$	$R_2 = 20\,\text{Mbits/s}$
$\nu = 0.5\,/\text{s}$	$\sigma = 10\,\text{Mbits}$

By calculating, we can get the following results,

$p_1 = 0.1600$	$p_2 = 0.8400$
R	$p_1R_1 + p_2R_2 = 23.2000$
ρ	0.2155
λ_1	1.4644
$\mathrm{E}[\mathbf{x}]$	1.0713
$\mathrm{E}[\gamma^{\text{fl}}]$	13.6697

Table 4.3: The results of the throughput in the fluid regime in Case 5

Compared the results in Case 4 and Case 5, it can be seen that $\gamma^{\mathrm{fl}} > \gamma^{\mathrm{qs}}$. In the finite user population model, we get the same result in the model by Bonald et al., i.e. the throughput increases with the mobility, although this increase in throughput is not so big.

It is easy to calculate the throughput in two regimes by computer. Assume that the number of the classes is 2, it only take several seconds to calculate them. Of course, the size of the system that could be calculated and the calculation speed depend on the specific computer.

Up to now, both the model by Bonald et al. and the OCOF finite user population model assume that users occupy the positions uniformly. It is easy to extend this kind of uniform distribution to any other distributions.

Chapter 5

OCMF finite population model

In the previous chapter we evaluated the throughput for the OCOF finite population model, where there is only one flow for one customer at the same time. However, sometimes one customer could receive or generate the new flow during the duration of the old flow. For example, one could receive short message while talking with the other. In this case, there are several flows for one customer at the same time. This corresponds another model, so called OCMF(One Customer Multiple Flows) finite population model. We will consider a single cell mobile system in the following as in the OCOF finite population model.

Assumptions:

- n customers in the system;
- K classes in the system. For customers in class k the full service rate is $R_k, k = 1, 2, \ldots, K$; The number of customers in class k is m_k.
- Data flows from a customer in class k arrive at BS as a Poissonian process with rate of $\lambda^k, k = 1, 2, \ldots, K$;
- The total arrival rate to BS is $\lambda = \sum_{k=1}^{K} m_k \lambda^k$;
- Each flow with the mean size of σ;

In the new model, the flows for a customer are assumed to arrive as a Poisson process while the customer number is finite. In this sense the OCMF finite population model is an intermediate model between the model by Bonald et al. and the OCOF finite population model.

5.1 The average throughput of the cell system in the quasi-stationary regime

From equation (3.11), we obtain that the sojourn time of a class-k customer in the system, conditioned on $\mathbf{M} = \mathbf{m}$, is:

$$T_k[\mathbf{m}] = \frac{\sigma}{(1-\rho[\mathbf{m}])R_k} = \frac{S_k}{1-\rho[\mathbf{m}]}, \tag{5.1}$$

where

- R_k is the full service rate available by a customer in class k;
- $S_k = \frac{\sigma}{R_k}$, which is the average service time of a class-k customer with full service rate;
- $\rho[\mathbf{m}] = \sum_{i=1}^{n} \rho_i = \sum_{k=1}^{K} m_k \lambda^k S_k$.

So the average sojourn time of a customer in the system, conditioned on $\mathbf{M} = \mathbf{m}$, is:

$$\begin{aligned} \bar{T}[\mathbf{m}] &= \frac{1}{n}\sum_{k=1}^{K} m_k T_k[\mathbf{m}] \\ &= \frac{1}{n}\sum_{k=1}^{K} m_k \frac{S_k}{1-\rho[\mathbf{m}]} \\ &= \frac{1}{1-\rho[\mathbf{m}]} \cdot \frac{\sigma}{n}\sum_{k=1}^{K} \frac{m_k}{R_k}. \end{aligned} \tag{5.2}$$

Alternatively we may calculate $\bar{T}[\mathbf{m}]$ as follows:

The average service time of a customer conditioned on $\mathbf{M} = \mathbf{m}$ is:

$$\begin{aligned} \bar{S} &= \frac{1}{n}\sum_{k=1}^{K} m_k S_k \\ &= \frac{\sigma}{n}\sum_{k=1}^{K} \frac{m_k}{R_k}. \end{aligned} \tag{5.3}$$

So the the average sojourn time of a customer in the system, conditioned on $\mathbf{M} = \mathbf{m}$, is:

$$\begin{aligned} \bar{T}[\mathbf{m}] &= \frac{\bar{S}}{1-\rho[\mathbf{m}]} \\ &= \frac{1}{1-\rho[\mathbf{m}]} \cdot \frac{\sigma}{n}\sum_{k=1}^{K} \frac{m_k}{R_k}. \end{aligned} \tag{5.4}$$

From the above we can see that the results are the same in both ways.

Then, the average sojourn time of a customer in the system is:

$$\mathrm{E}[T] = \sum_{\mathbf{m}} p_r\{\mathbf{m}\}\bar{T}[\mathbf{m}], \tag{5.5}$$

where $p_r\{\mathbf{m}\}$ is calculated as equation (4.15).

The average throughput of the whole cell system is:

$$\mathrm{E}[\gamma^{\mathrm{qs}}] = \frac{\sigma}{\mathrm{E}[T]} = \frac{\sigma}{\sum_{\mathbf{m}} p_r\{\mathbf{m}\}\bar{T}[\mathbf{m}]}. \tag{5.6}$$

Because data flows arrive at BS as a Poisson process, it is necessary to consider the condition for the stability of the system:

$$\rho[\mathbf{m}] = \sum_{k=1}^{K} m_k \lambda^k S_k = \sum_{k=1}^{K} m_k \frac{\lambda^k \sigma}{R_k} \leq 1, \ \forall \mathbf{m}. \tag{5.7}$$

We can also consider the stability condition in another way. When the parameters of the system are known, it is possible to find a class, whose load for a customer is maximum among all K classes. Then if all the customers reside in this class, the total load of the system reaches the highest value. The system will be stable as long as this worst case is stable, i.e.,

$$\rho_{\mathrm{worst}} = n \frac{\lambda^{k_0}\sigma}{R_{k_0}} \leq 1, \tag{5.8}$$

where $\lambda^{k_0}/R_{k_0} = \max_{k=1}^{K} \lambda^k/R_k$.

If we take the same arrival rate for each class, i.e. $\lambda^k = \frac{1}{n}\lambda$, the total arrival rate is λ and the evaluation will be simple. In this case,

$$\rho[\mathbf{m}] = \sum_{k=1}^{K} m_k \lambda^k S_k = \lambda \cdot \frac{1}{n} \sum_{k=1}^{K} m_k S_k = \lambda \bar{S}, \tag{5.9}$$

and the stability condition is:

$$\rho_{\mathrm{worst}} = n \cdot \max_{k=1}^{K} \frac{\lambda\sigma}{nR_k} = \frac{\lambda\sigma}{R_K} \leq 1, \tag{5.10}$$

where R_K is the outermost class in the cell.

Taking the conditions in Case 3 as an example, we could calculate the throughput in the OCMF finite population model as follows:

Case 6 where,

By calculating, we can get the results shown in Table 5.1:

$n = 4$ $K = 2$
$r = 10\,\text{km}$ $r_1 = 4\,\text{km}$
$R_1 = 40\,\text{Mbits/s}$ $R_2 = 20\,\text{Mbits/s}$
$\lambda^1 = 0.4\,/\text{s}$ $\lambda^2 = 0.4\,/\text{s}$
$\sigma = 10\,\text{Mbits}$

$p_1 = 0.1600$				$p_2 = 0.8400$
M_1	M_2	$\bar{T}[\mathbf{m}]$	E[T]	E[γ^{qs}]
0	4	2.5000		
1	3	1.4583		
2	2	0.9375	1.9083	5.2401
3	1	0.6250		
4	0	0.4167		

Table 5.1: The throughput for the OCMF finite population model in Case 6

5.2 The average throughput of the cell system in the fluid regime

In the fluid regime, all the customers are in the same class with the full service rate as follows:

$$R = \sum_{k=1}^{K} p_k R_k, \tag{5.11}$$

where p_k is calculated as equation (4.14).

So the average sojourn time of a customer in the system is:

$$\mathrm{E}[T] = \frac{\sigma}{(1-\rho)R} = \frac{\bar{S}}{1-\rho}, \tag{5.12}$$

where

- $\bar{S} = \frac{\sigma}{R}$, which is the full service time of customer i;
- $\rho = \sum_{i=1}^{n} \rho_i = \bar{S} \sum_{i=1}^{n} \lambda^i = \frac{\lambda\sigma}{R}$.

The average throughput of the whole cell system is:

$$\mathrm{E}[\gamma^{\text{fl}}] = \frac{\sigma}{\mathrm{E}[T]} = \frac{\sigma}{\bar{S}}(1-\rho) = R(1-\rho), \tag{5.13}$$

where

- $\bar{S} = \frac{\sigma}{R}$, which is the full service time of customer i;

- $\rho = \sum_{i=1}^{n} \rho_i = \bar{S} \sum_{i=1}^{n} \lambda^i = \frac{\lambda\sigma}{R}$.

Similarly in the quasi-stationary regime, we can get the stability condition for the system as follows:

$$\rho = \frac{\lambda\sigma}{R} \leq 1. \tag{5.14}$$

An example is given in Case 7 in the following:

Case 7 where

$n = 4$	$K = 2$
$r = 10\,\text{km}$	$r_1 = 4\,\text{km}$
$R_1 = 40\,\text{Mbits/s}$	$R_2 = 20\,\text{Mbits/s}$
$\lambda^1 = 0.4\,/\text{s}$	$\lambda^2 = 0.4\,/\text{s}$
$\sigma = 10\,\text{Mbits}$	

By calculating, we can get the following results:

$p_1 = 0.1600$	$p_2 = 0.8400$
R	$p_1 R_1 + p_2 R_2 = 23.2000$
ρ	0.1724
$\text{E}[\gamma^{\text{fl}}]$	19.2000

Table 5.2: The results of the throughput in the fluid regime in Case 7

From the results in Case 6 and Case 7, it can be seen that $\gamma^{\text{fl}} > \gamma^{\text{qs}}$ in the OCMF finite population model, which is consistent with the results in both the model by Bonald et al. and the OCOF finite population model. That is to say, the throughput increases with the mobility.

Chapter 6

Numerical comparison of the results

Up to now we have three models for evaluating the performance of wireless data channel in a single cell mobile system. In this chapter we will compare numerically their performance in the case of a finite user population, based on the fixed load and arrival rate. Three kinds of loads are choosed as follows:

- $\rho \to 0$;
- $\rho = 0.4$;
- $\rho = 0.8$.

In the OCOF finite population model, there are two limits for the load as follows:

- When $\nu \to 0$, the service time of a customer in node 0 tends to infinity. For each class $k, k = 1, 2, \ldots, K$, all the m_k customers reside in node 0. The load of the system tends to 0;
- When $\nu \to \infty$, the service time of a customer in node 0 tends to 0. For each class $k, k = 1, 2, \ldots, K$, all the m_k customers reside in node 1. The load of the system is equal to 1;

It is possible to get any value of the load (ρ_O) between 0 and 1 by adjusting ν. In order to get the correct value of ν for a given ρ_O, it is necessary to know how to calculate ρ_O. Please see Appendix B.

In the OCMF finite population model, assume that $\lambda^k = \frac{1}{n}\lambda_M, k = 1, 2, \ldots, K$. The total arrival rate to node 1 is $\sum_{k=1}^{K} m_k \lambda^k = \lambda^k \sum_{k=1}^{K} m_k = \frac{\lambda_M}{n} n = \lambda_M$. Then the load can be calculated in the following ways:

- For the QS regime:

$$\begin{aligned}\rho_M^{\text{qs}} &= \sum_{\mathbf{m}} p_r\{\mathbf{m}\}\rho[\mathbf{m}] \\ &= \sum_{\mathbf{m}} p_r\{\mathbf{m}\}\lambda_M \bar{S} \\ &= \sum_{\mathbf{m}} p_r\{\mathbf{m}\}\lambda_M \cdot \frac{1}{n}\sum_{k=1}^{K} m_k \frac{\sigma}{R_k} \\ &= \frac{\lambda_M \sigma}{n} \sum_{\mathbf{m}} p_r\{\mathbf{m}\} \sum_{k=1}^{K} \frac{m_k}{R_k}. \end{aligned} \tag{6.1}$$

  where R_k is the full service rate available for class-k customers.

  From equation (5.10), the stability condition is:

$$\frac{\lambda_M \sigma}{R_K} \leq 1 \Rightarrow \lambda_M \leq \frac{R_K}{\sigma}, \tag{6.2}$$

  where R_K is the outer class in the cell.

- For the FL regime:

$$\rho_M^{\text{fl}} = \frac{\lambda_M \sigma}{R_M^{\text{fl}}} = \frac{\lambda_M \sigma}{R}, \tag{6.3}$$

  where $R = \sum_{k=1}^{K} p_k R_k$. Note that R_k here is the full service rate available for class-k customer.

  The stability condition is:

$$\frac{\lambda_M \sigma}{R} \leq 1 \Rightarrow \lambda_M \leq \frac{R}{\sigma}. \tag{6.4}$$

So we can change the value of λ_M to get the above three typical loads.

In the model by Bonald et al., assume that the arrival rate to node 1 is λ_B. Based on chapter 3, we can calculate the load as follows:

- For the QS regime:

$$\rho_B^{\text{qs}} = \frac{\lambda_B \sigma}{R_B^{\text{qs}}}, \tag{6.5}$$

  where $R_B^{\text{qs}} = \mathrm{E}[\frac{1}{R_k}]^{-1} = \left(\sum_{k=1}^{K} p_k \frac{1}{R_k}\right)$;

  The stability condition is:

$$\frac{\lambda_B \sigma}{R_B^{\text{qs}}} \leq 1 \Rightarrow \lambda_B \leq \frac{R_B^{\text{qs}}}{\sigma}. \tag{6.6}$$

- For the FL regime:

$$\rho_B^{\text{fl}} = \frac{\lambda_B \sigma}{R_B^{\text{fl}}} = \frac{\lambda_B \sigma}{R}, \tag{6.7}$$

  where $R = \sum_{k=1}^{K} p_k R_k$.

  The stability condition is:

$$\frac{\lambda_B \sigma}{R} \leq 1 \Rightarrow \lambda_B \leq \frac{R}{\sigma}. \tag{6.8}$$

We can get the three typical loads by adjusting the parameter λ_B.

From equations (6.1) and (6.5), (6.3) and (6.7), it is seen that the loads for the OCMF model both in the QS regime and in the FL regime are the same as those for the model by Bonald et al., respectively. That is to say, if $\lambda_M = \lambda_B$, then $\rho_M^{\text{qs}} = \rho_B^{\text{qs}}$ and $\rho_M^{\text{fl}} = \rho_B^{\text{fl}}$.

Before comparing the results for the three models at the different loads, we introduce a general result as follows:

General result 1 Based on equations (3.21) and (5.13), it is easy to see that the throughputs for the OCMF model and the model by Bonald et al. in the FL regime are the same while the results are different in the QS regime.

Let us consider the case where there are 10 classes and a finite number of users in the system, where

$n = 5$	$K = 10$	$r = 15\,\text{km}$
$r_1 = 6\,\text{km}$	$r_2 = 7\,\text{km}$	$r_3 = 8\,\text{km}$
$r_4 = 9\,\text{km}$	$r_5 = 10\,\text{km}$	$r_6 = 11\,\text{km}$
$r_7 = 12\,\text{km}$	$r_8 = 13\,\text{km}$	$r_9 = 14\,\text{km}$
$r_{10} = 15\,\text{km}$	$R_1 = 80\,\text{Mbits/s}$	$R_2 = 75\,\text{Mbits/s}$
$R_3 = 70\,\text{Mbits/s}$	$R_4 = 65\,\text{Mbits/s}$	$R_5 = 60\,\text{Mbits/s}$
$R_6 = 55\,\text{Mbits/s}$	$R_7 = 50\,\text{Mbits/s}$	$R_8 = 40\,\text{Mbits/s}$
$R_9 = 30\,\text{Mbits/s}$	$R_{10} = 20\,\text{Mbits/s}$	$\sigma = 10\,\text{Mbits}$

From equations (6.2),(6.4),(6.6) and (6.8), we can get all the critical values of the arrival rate for the OCMF model and the model by Bonald et al. The results are given in Table 6.1. Accordingly the critical loads are shown in Table 6.2.

That is to say, the OCMF model is valid only when the intervals of the arrival rate (/s) in the QS regime and the FL regime are [0,2.000] and [0,5.2644], respectively. Similarly, the model by Bonald et al. is valid only when the intervals of the arrival rate (/s) in the QS regime and the FL regime are [0, 4.2999] and [0, 5.2644], respectively.

model	the critical value of arrival rate in QS regime	the critical value of arrival rate in FL regime
OCMF model	2.0000 /s	5.2644 /s
Bonald et al. model	4.2999 /s	5.2644 /s

Table 6.1: The critical values of arrival rate for the OCMF model and the model by Bonald et al. in Case 8

model	the critical value of the load in QS regime	the critical value of the load in FL regime
OCMF model	0.4651	1.0000
Bonald et al. model	1.0000	1.0000

Table 6.2: The critical values of the load for the OCMF model and the model by Bonald et al. in Case 8

In contrast, the OCOF model is stable for all reasonable value of the thinking time. That is to say, in the OCOF model the thinking time can vary from 0 to ∞.

6.1 Numerical comparison when $\rho \to 0$

Case 8 where $\rho \to 0$. The calculation results are shown in Table 6.3 and 6.4.

For QS regime

model	arrival rate	the load	throughput
OCOF model	$\nu = 10^{-7}$	$\rho_O^{\text{qs}} = 1.1628 \times 10^{-7}$	$\gamma_O^{\text{qs}} = 42.9993$
OCMF model	$\lambda_M = 5.0000 \times 10^{-7}$	$\rho_M^{\text{qs}} = 1.1628 \times 10^{-7}$	$\gamma_M^{\text{qs}} = 42.9993$
Bonald et al. model	$\lambda_B = 5.0000 \times 10^{-7}$	$\rho_B^{\text{qs}} = 1.1628 \times 10^{-7}$	$\gamma_B^{\text{qs}} = 42.9993$

Table 6.3: The throughput comparison for the different models in the QS regime when $\rho \to 0$

From the results in Table 6.3 and 6.4, it can be seen that, when ρ tends to 0, the throughputs for the three models are the same both in the QS regime and in the FL regime. This is reasonable. In the OCOF finite population model, if we let ν tend to zero, the load (ρ_O) tends to 0. In this case, for each class k, all the m_k customers always reside in node 0. Then the data flow at node 1 could be regarded as infinite flow and the arrival rate is $m_k \nu$. This is the same as the traffic characteristic assumption in the OCMF finite population model and the model by Bonald et al. So when $\rho \to 0$, the three models are essentially identical.

Also we could calculate the throughput for the OCMF model and the model by Bonald et al. when the arrival rate is 0. The results are given in Table 6.5. It can

For FL regime

model	arrival rate	the load	throughput
OCOF model	$\nu = 2.0000 \times 10^{-7}$	$\rho_O^{\text{fl}} = 1.8995 \times 10^{-7}$	$\gamma_O^{\text{fl}} = 52.6444$
OCMF model	$\lambda_M = 10^{-6}$	$\rho_M^{\text{fl}} = 1.8995 \times 10^{-7}$	$\gamma_M^{\text{fl}} = 52.6444$
Bonald et al. model	$\lambda_B = 10^{-6}$	$\rho_B^{\text{fl}} = 1.8995 \times 10^{-7}$	$\gamma_B^{\text{fl}} = 52.6444$

Table 6.4: The throughput comparison for the different models in the FL regime when $\rho \to 0$

be seen from this table that the results are the same as those when $\rho \to 0$.

For FL regime

model	arrival rate	the load	throughput
For QS regime			
OCMF model	$\lambda_M = 0$	$\rho_M^{\text{qs}} = 0$	$\gamma_M^{\text{qs}} = 42.9993$
Bonald et al. model	$\lambda_B = 0$	$\rho_B^{\text{qs}} = 0$	$\gamma_B^{\text{qs}} = 42.9993$
For FL regime			
OCMF model	$\lambda_M = 0$	$\rho_M^{\text{fl}} = 0$	$\gamma_M^{\text{fl}} = 52.6444$
Bonald et al. model	$\lambda_B = 0$	$\rho_B^{\text{fl}} = 0$	$\gamma_B^{\text{fl}} = 52.6444$

Table 6.5: The throughput of the OCMF model and the model by Bonald et al. when $\rho = 0$

6.2 Numerical comparison when ρ is fixed

Case 9 where $\rho = 0.4$. The calculation results are given in Tables 6.6 and 6.7.

For QS regime

model	arrival rate	the load	throughput
OCOF model	$\lambda_1 = 0.1920$ ($\nu = 0.3997$)	$\rho_O^{\text{qs}} = 0.4$	$\gamma_O^{\text{qs}} = 30.3333$
OCMF model	$\lambda_M = 1.7200$	$\rho_M^{\text{qs}} = 0.4$	$\gamma_M^{\text{qs}} = 24.0662$
Bonald et al. model	$\lambda_B = 1.7200$	$\rho_B^{\text{qs}} = 0.4$	$\gamma_B^{\text{qs}} = 25.7993$

Table 6.6: The throughput comparison for different models in the QS regime when $\rho = 0.4$

In case 9, the system is finite. So the OCOF model is appropriate. From the results in Tables 6.6 and 6.7, we see that when $\rho = 0.4$ the throughputs calculated with the OCMF model and the model by Bonald et al. are lower than those with the OCOF model. In the QS regime, the throughput of the OCMF model deviates from that of the OCOF model by 20.7% and the throughput of the model by Bonald et al. deviates from that of the OCOF model by 15.0%. It can also be seen from the tables that in the FL regime, the throughput calculated with

For FL regime

model	arrival rate	the load	throughput
OCOF model	$\lambda_1 = 2.1059$ ($\nu = 0.4750$)	$\rho_O^{\text{fl}} = 0.4$	$\gamma_O^{\text{fl}} = 37.1645$
OCMF model	$\lambda_M = 2.1060$	$\rho_M^{\text{fl}} = 0.4$	$\gamma_M^{\text{fl}} = 31.5844$
Bonald et al. model	$\lambda_B = 1.1060$	$\rho_B^{\text{fl}} = 0.4$	$\gamma_B^{\text{fl}} = 31.5844$

Table 6.7: The throughput comparison for the different models in the FL regime when $\rho = 0.4$

the OCMF model is the same as that with the model by Bonald et al., which is consistent with the above general result. In the FL regime, the throughputs of the OCMF model and the model by Bonald et al. deviate from that of the OCOF model by 15.0%.

Case 10 where $\rho = 0.8$. In this case, the load is greater than the critical load (0.4651) of the OCMF model in the QS regime. So in the QS regime, we just compare the numerical results for the OCOF model and the model by Bonald et al. The calculation results are shown in Tables 6.8 and 6.9.

For QS regime

model	arrival rate	the load	throughput
OCOF model	$\lambda_1 = 0.4043$ ($\nu = 1.1695$)	$\rho_O^{\text{qs}} = 0.8$	$\gamma_O^{\text{qs}} = 19.1014$
Bonald et al. model	$\lambda_B = 3.4400$	$\rho_B^{\text{qs}} = 0.8$	$\gamma_B^{\text{qs}} = 8.5993$

Table 6.8: The throughput comparison for the OCOF model and the model by Bonald et al. in the QS regime when $\rho = 0.8$

For FL regime

model	arrival rate	the load	throughput
OCOF model	$\lambda_1 = 4.2117$ ($\nu = 1.3128$)	$\rho_O^{\text{fl}} = 0.8$	$\gamma_O^{\text{fl}} = 23.5055$
OCMF model	$\lambda_M = 4.2115$	$\rho_M^{\text{fl}} = 0.8$	$\gamma_M^{\text{fl}} = 10.5294$
Bonald et al. model	$\lambda_B = 4.2115$	$\rho_B^{\text{fl}} = 0.8$	$\gamma_B^{\text{fl}} = 10.5294$

Table 6.9: The throughput comparison for the different models in the FL regime when $\rho = 0.8$

From the results in Table 6.8 , it can be seen that in the QS regime, when $\rho = 0.8$, the throughput of the model by Bonald et al. is lower than that of the OCOF model by 55.0%.

From the results in Table 6.9, we can see that in the FL regime, when $\rho = 0.8$, the throughputs of the OCMF model and the model by Bonald et al. are lower than that of the OCOF model by 55.2%.

Based on Case 9 and Case 10, it can be seen that when the load of the system increases the arrival rate will increase accordingly. And with increasing arrival rate and load of the system, the deviation of the other two models from the OCOF model increases.

Further calculation shows that the OCOF model is stable for all values of the thinking time (from 0 to ∞) while the OCMF model and the model by Bonald et al. are not (for them the arrival rate is lower than the critical value). When the arrival rate is higher than the critical value, the throughput will be negative, which is not reasonable.

6.3 Numerical comparison when the arrival rate is fixed

Case 11 where the arrival rate is equal to 0.2. The calculation results are given in Tables 6.10 and 6.11.

For QS regime

model	arrival rate	the load	throughput
OCOF model	$\lambda_1 = 0.2000$ ($\nu = 0.4195$)	$\rho_O^{\text{qs}} = 0.4162$	$\gamma_O^{\text{qs}} = 29.8654$
OCMF model	$\lambda_M = 0.2000$	$\rho_M^{\text{qs}} = 0.0465$	$\gamma_M^{\text{qs}} = 40.8855$
Bonald et al. model	$\lambda_B = 0.2000$	$\rho_B^{\text{qs}} = 0.0465$	$\gamma_B^{\text{qs}} = 40.9993$

Table 6.10: The throughput comparison for the different models in the QS regime when the arrival rate is 0.2

For FL regime

model	arrival rate	the load	throughput
OCOF model	$\lambda_1 = 0.2000$ ($\nu = 0.04032$)	$\rho_O^{\text{fl}} = 0.0380$	$\gamma_O^{\text{fl}} = 51.0565$
OCMF model	$\lambda_M = 0.2000$	$\rho_M^{\text{fl}} = 0.0380$	$\gamma_M^{\text{fl}} = 50.6444$
Bonald et al. model	$\lambda_B = 0.2000$	$\rho_B^{\text{fl}} = 0.0380$	$\gamma_B^{\text{fl}} = 50.6444$

Table 6.11: The throughput comparison for the different models in the FL regime when the arrival rate is 0.2

Case 12 where the arrival rate is equal to 0.4. The calculation results are given in Tables 6.12 and 6.13.

From Tables 6.10 and 6.12, it can be seen that in the QS regime, the throughput of the OCOF model is lower than that of the OCMF model and the model by Bonald et al. when the arrival rate is given. When the arrival rate is 0.2, it is lower than that of the OCMF model and the model by Bonald et al. by 27.0%

For QS regime

model	arrival rate	the load	throughput
OCOF model	$\lambda_1 = 0.4000$ ($\nu = 1.1440$)	$\rho_O^{\text{qs}} = 0.7926$	$\gamma_O^{\text{qs}} = 19.3185$
OCMF model	$\lambda_M = 0.4000$	$\rho_M^{\text{qs}} = 0.0930$	$\gamma_M^{\text{qs}} = 38.7589$
Bonald et al. model	$\lambda_B = 0.4000$	$\rho_B^{\text{qs}} = 0.0930$	$\gamma_B^{\text{qs}} = 38.9993$

Table 6.12: The throughput comparison for the different models in the QS regime when the arrival rate is 0.4

For FL regime

model	arrival rate	the load	throughput
OCOF model	$\lambda_1 = 0.4000$ ($\nu = 0.08132$)	$\rho_O^{\text{fl}} = 0.0760$	$\gamma_O^{\text{fl}} = 49.4935$
OCMF model	$\lambda_M = 0.4000$	$\rho_M^{\text{fl}} = 0.0760$	$\gamma_M^{\text{fl}} = 48.6444$
Bonald et al. model	$\lambda_B = 0.4000$	$\rho_B^{\text{fl}} = 0.0760$	$\gamma_B^{\text{fl}} = 48.6444$

Table 6.13: The throughput comparison for the different models in the FL regime when the arrival rate is 0.4

and 27.2%, respectively. When the arrival rate is 0.4, it is lower by 50.2% and 50.5%, respectively. The reason is that the load of the OCOF model is far higher than that of the OCMF model and the model by Bonald et al. while the arrival rates are the same. It is obvious that the difference of the performance between the OCOF model and the other two models increases with the arrival rate.

It can also be seen that in the QS regime, when the arrival rates are the same, $\gamma_O^{\text{qs}} < \gamma_M^{\text{qs}} < \gamma_B^{\text{qs}}$. In this sense, we can say that the OCMF model is an intermediate model between the OCOF model and the model by Bonald et al. In fact, it is a good way to compare different models based on the arrival rate because the arrival rate defines the carried traffic in the system while the load is an internal parameter. It is difficult for us to know in advance the load of the OCOF system. The above results show that the performance worsens when the user number goes from infinite to finite, given the arrival rate to the system is kept fixed.

From Tables 6.11 and 6.13, we can see that in the FL regime, the throughput of the three models are almost the same when the arrival rate is given. And when the arrival rate is 0.2, the load is so small (0.0380) that the throughputs for the three models approach to the maximum throughput of the system (52.6444 when $\rho \to 0$). For the OCMF model and the model by Bonald et al., they are 50.6444. On the contrary, from Table 6.9 we see that the load is so high that all 5 customers almost always reside in node 1 and share the capacity of the system. The throughput in this case is 10.5294, which is approximately equal 1/5 of 50.3344.

In Table 6.11, we also see that the thinking time for the customers is very small (0.04032). So all 5 customers reside in node 0 most of their time and the arrival rate should be about 5 times the thinking time. The given arrival rate is 0.2000, which is approximately 5 times the thinking time. It is understandable.

Based on all above results, we can see that the throughput of the OCMF model is always lower than that of the model by Bonald et al. In the OCMF model, it is possible that the load is the highest in some user position constellation, which contributes more negatively to the throughput. So it makes the above resuls understandable.

6.4 The relationship among the arrival rate, the load and the throughput in the three models

Now let us look in more detail at the relationship between the arrival rate, the load and the throughput of the OCOF,OCMF models and the model by Bonald et al., based on the conditions in Case 8.

The results are given in Table 6.14, 6.15 and 6.16.

	QS regime		FL regime	
ν	ρ_O^{qs}	γ_O^{qs}	ρ_O^{fl}	γ_O^{fl}
10^{-6}	1.1628×10^{-6}	42.9993	9.4977×10^{-7}	52.6444
10^{-4}	1.1628×10^{-4}	42.9953	9.4975×10^{-5}	52.6404
0.01	0.0116	42.6014	0.0095	52.2460
0.1	0.1127	39.2123	0.0931	48.7988
1	0.7451	20.6704	0.6941	27.1420
10	0.9998	9.5882	0.9998	11.7653
100	1.0000	8.6912	1.0000	10.6409
10^4	1.0000	8.6008	1.0000	10.5300
10^6	1.0000	8.5999	1.0000	10.5289

Table 6.14: The relationship among the arrival rate, the load and the throughput in the OCOF model

From Table 6.14, it is easy to see that in the OCOF model the thinking time can vary from 0 to ∞ and the load can have any value between 0 and 1, which is consistent with the former conclusion.

Based on Table 6.15 and 6.16, we know that in the OCMF model and the model by Bonald et al. the value of the arrival rate is limited. Above some value of the arrival rate, the system will not be stable again. The load begins to greater than 1 and the throughput starts to be negative, which are not reasonable. The exact critical values of the arrival rate for the OCMF model and the model by Bonald et al. are shown in Table 6.1.

	QS regime		FL regime	
λ_M	ρ_M^{qs}	γ_M^{qs}	ρ_M^{fl}	γ_M^{fl}
10^{-6}	2.3256×10^{-7}	42.9993	1.8995×10^{-7}	52.6444
10^{-4}	2.3256×10^{-5}	42.9983	1.8995×10^{-5}	52.6434
0.01	0.0023	42.8939	0.0019	52.5444
0.1	0.0233	41.9439	0.0190	51.6444
1	0.2326	32.2716	0.1900	42.6444
10	2.3256	-52.3000	1.8995	-47.3556

Table 6.15: The relationship among the arrival rate, the load and the throughput in OCMF model

	QS regime		FL regime	
λ_B	ρ_B^{qs}	γ_B^{qs}	ρ_B^{fl}	γ_B^{fl}
10^{-6}	2.3256×10^{-7}	42.9993	1.8995×10^{-7}	52.6444
10^{-4}	2.3256×10^{-5}	42.9983	1.8995×10^{-5}	52.6434
0.01	0.0023	42.8993	0.0019	52.5444
0.1	0.0233	41.9993	0.0190	51.6444
1	0.2326	32.9993	0.1900	42.6444
10	2.3256	-57.0007	1.8995	-47.3556

Table 6.16: The relationship among the arrival rate, the load and the throughput in the model by Bonald et al.

From the results in Table 6.14, 6.15 and 6.16, it can also be seen that for the given conditions the throughputs with the OCMF model and the model by Bonald et al. begin to deviate from the throughput of the OCOF model obviously when the arrival rate reaches 0.1 /s. The deviation increases with the arrival rate and the load.

Chapter 7

Conclusions

7.1 Summary

We have explored a finite user population model for cellular system, the OCOF model. In the OCOF model only one flow is allowed for each customer simultaneously. The arrival rate to the system is not Poissonian and the user population is finite. Two regimes are identified, the QS (Quasi-stationary) regime and the FL (Fluid) regime, where the rate variations occur on an infinitely slow and an infinitely fast time scale, respectively. The QS regime and FL regime provide the conservative and optimistic performance estimates respectively. The results show that the performance of the limit regimes is insensitive, and only depends on an appropriately defined load factor.

Another finite user population model is also developed in the thesis, the OCMF model. In the OCMF model multiple flows are allowed for one customer at the same time. Arrival rate to the system could be Poissonian and the user population is finite. In this sense, the OCMF model is an intermediate model between the OCOF model and the model by Bonald et al. The numerical results also support this kind of conclusion. In fact, the model by Bonald et al. is a limit case of the OCOF and OCMF models.

The performance evaluation results are compared in a finite system for the OCOF model, the OCMF model and the model by Bonald et at. The OCOF model is stable for all values of the thinking time. In the OCOF model the thinking time of a customer can vary from 0 to ∞. On the contrary, in the OCMF model and the model by Bonald et al. the values of arrival rate are limited by the stability condition.

The results also show that in a finite populaiton system, when the loads are the same, the throughput of the OCOF model is higher than those with the OCMF model and the model by Bonald et al.

A good way to compare the different models is to keep the arrival rate fixed. The numerical results show that when the arrival rates are the same, $\gamma_O^{\mathrm{qs}} < \gamma_M^{\mathrm{qs}} < \gamma_B^{\mathrm{qs}}$ in the QS regime. In this sense, we can say that the OCMF model is an intermediate model between the OCOF model and the model by Bonald et al. The difference between the OCOF model and the other two models increases with the load.

The numerical results show that in the QS regime the throughput of the OCMF model is always lower than that of the model by Bonald et al. However, in FL regime the throughputs of the OCMF model and the model by Bonald et al. are the same.

7.2 Further work

With the increase of the number of customers and classes, the state space will grow exponentially. The calculation of normalization constant depends on the specific computer capacity. So the size of the system that could be calculated and the calculation speed are limited on the specific computer. Deriving an efficient, possibly recursive way of calculating the normalization constant would therefore have a big practical value.

Another future challenge is to extend the finite user population models to several cells.

Bibliography

[1] E. Altman. Capacity of multi-service cellular networks with transmission-rate control: A queueing analysis. In *ACM Mobicom*, 2002.

[2] T. Bonald. A score-based opportunistic scheduler for fading radio channels. In *European Wireless*, 2004.

[3] T. Bonald, S. Borst, N. Hegde, and A. Proutière. Wireless data performance in multi-cell scenarios. In *ACM SIGMETRICS*, 6 2004.

[4] T. Bonald, S. C. Borst, and A. Proutière. How mobility impacts the flow-level performance of wireless data systems. In *IEEE InfoCom*, 2004.

[5] T. Bonald and A. Proutière. Insensitive bandwidth sharing. In *Proceedings of the IEEE Globecom 2002*, pages 82–91, 2002.

[6] T. Bonald and A. Proutière. Insensitivity in processor-sharing networks. *Performance Evaluation*, 49:193 – 209, 2002.

[7] T. Bonald and A. Proutière. Insensitive bandwidth sharing in data networks. *Queueing Systems*, 44:69 – 100, 2003.

[8] T. Bonald and A. Proutière. Wireless downlink data channels: User performance and cell dimensioning. In *MobiCom'03*, pages 14–19, Sep. 2003.

[9] T. Bonald and A. Proutière. On performance bounds for balanced fairness. *Performance Evaluation*, 55, 2004.

[10] T. Bonald, A. Proutière, J. Roberts, and J. Virtamo. Computational aspects of balanced fairness. In *Proceedings of the 18th International Teletraffic Congress (ITC-18)*, pages 801 – 810, September 2003.

[11] S. Borst. User-level performance of channel-aware scheduling algorithms in wireless data networks. In *IEEE Infocom*, 2003.

[12] P. G. Harrison and N. M. Patel. *Performance Modelling of Communication Networks and Computer Architectures*. Wokingham Addison-Wesley, 1994.

[13] S. G. Hg̈gman. Mobile communication system and service. Technical report, Helsinki University of Technology, 4 2003.

[14] L. Kleinrock. *Queueing System.* Wiley and Sons, 1976.

[15] W. C. Y. Lee. *Mobile Cellular Telecommunications–Analog and Digital System.* McGraw-Hill, New York, 1995.

[16] R. Nelson. *Probability, Stochasitic Processes, and Queueing Theory–The Mathematics of Computer Performance Modelling.* Springer-Verlag, New York, 1995.

[17] Focusing on Telecommunications. The Journey into Telecommunications, 2004. http://www.nokiainfo.net/nokiahistory.php/.

[18] A. Penttinen and E. Nyberg. Introduction to teletraffic theory. Lecture notes, Helsinki University of Technology, 2 2003.

[19] R. F. Serfozo. *Introduction to Stochastic Networks.* Springer Verlag, 1999.

[20] V. Timonen. Simulation studies on performance of balanced fairness. Master's thesis, Helsinki University of Technology, Oct. 2003.

[21] J. Virtamo. Queueing theory. Lecture notes, Helsinki University of Technology, 9 2004.

Appendices

Appendix A

Understanding of the expression for arrival rate at node 1

We could also understand the following equation (4.12) in two other ways,

$$\lambda_{1k} = M_k \frac{G[\mathbf{M} - \mathbf{e}_k]}{G[\mathbf{M}]} \nu.$$

1. Consider the two limits,

   - When $\nu \to 0$, for each class k, all the M_k customers reside in node 0 and the arrival rate to node 1 is $\lambda_{1k} = M_k \nu$;
   - When $\nu \to \infty$, for each class k, all the M_k customers reside in node 1 and the arrival rate to node 1 is $\lambda_{1k} = \frac{M_k}{M} \cdot \frac{R_k}{\sigma}$.

   In the following example (example 1), there are altogether 4 customers ($n = 4$) and 2 classes ($K = 2$) in the cell system with the radius of $r = 10\,\text{km}$. For class 1, $R_1 = 40\,\text{Mbits/s}$, $r_1 = 4\,\text{km}$; For class 2, $R_2 = 20\,\text{Mbits/s}$. The flow size $\sigma = 10\,\text{Mbits}$. In this example,

   - $\frac{M_1}{M} \cdot \frac{R_k}{\sigma} = M_1$, and
   - $\frac{M_2}{M} \cdot \frac{R_k}{\sigma} = \frac{M_2}{2}$.

   With equation (4.12), we could calculate λ_{1k} explicitly by summing over the states in the above two limits. The results are shown in Table 7.1,

   From Table 7.1, it is seen that the results calculated by equation (4.12) are the same as what they should be in the two limits.

	$\nu = 10^5$	$\nu = 10^{-10}$
$M_1 = 0$ $M_2 = 4$	$\lambda_{11} = 0$ $\lambda_{12} = 2$	$\lambda_{11} = 0$ $\lambda_{12} = 4 \times 10^{-10}$
$M_1 = 1$ $M_2 = 3$	$\lambda_{11} = 1$ $\lambda_{12} = 1.5$	$\lambda_{11} = 1 \times 10^{-10}$ $\lambda_{12} = 3 \times 10^{-10}$
$M_1 = 2$ $M_2 = 2$	$\lambda_{11} = 2$ $\lambda_{12} = 1$	$\lambda_{11} = 2 \times 10^{-10}$ $\lambda_{12} = 2 \times 10^{-10}$
$M_1 = 3$ $M_2 = 1$	$\lambda_{11} = 3$ $\lambda_{12} = 0.5$	$\lambda_{11} = 3 \times 10^{-10}$ $\lambda_{12} = 1 \times 10^{-10}$
$M_1 = 4$ $M_2 = 0$	$\lambda_{11} = 4$ $\lambda_{12} = 0$	$\lambda_{11} = 4 \times 10^{-10}$ $\lambda_{12} = 0$

Table 7.1: The arrival rate to node 1 with different thinking time in different constellations

2. Calculating the throughput of class k by the method in case 2. The unnormalized stationary distribution of the network is,

$$\chi(\mathbf{x}) = x! \prod_{k=1}^{K} \frac{M_k!}{x_k!(M_k - x_k)!} \rho_k^{x_k}.$$

From equation (4.9),

$$\begin{aligned}
\gamma_k &= \frac{1 - p_{1k}}{p_{1k}} \cdot \nu\sigma \\
&= \frac{\sum_{j=0}^{M_k} \frac{M_k - j}{M_k} \sum_{\mathbf{x}:x_k=j} \chi(\mathbf{x})}{\sum_{j=0}^{M_k} \frac{j}{M_k} \sum_{\mathbf{x}:x_k=j} \chi(\mathbf{x})} \cdot \nu\sigma \\
&= \frac{\sum_{j=0}^{M_k-1} \frac{M_k - j}{M_k} \sum_{\mathbf{x}:x_k=j} \chi(\mathbf{x})}{\sum_{j=1}^{M_k} \frac{j}{M_k} \sum_{\mathbf{x}:x_k=j} \chi(\mathbf{x})} \cdot \nu\sigma \\
&= \frac{\sum_{j=0}^{M_k-1} \frac{M_k - j}{M_k} \cdot \frac{M_k!}{j!(M_k-j)!} \rho_k^j \cdot \sum_{\mathbf{x}:x_k=0} \chi(\mathbf{x})}{\sum_{j=1}^{M_k} \frac{j}{M_k} \cdot \frac{M_k!}{j!(M_k-j)!} \rho_k^j \cdot \sum_{\mathbf{x}:x_k=0} \chi(\mathbf{x})} \cdot \nu\sigma \\
&= \frac{\sum_{j=0}^{M_k-1} \frac{(M_k-1)!}{j!(M_k-1-j)!} \rho_k^j \cdot \sum_{\mathbf{x}:x_k=0} \chi(\mathbf{x})}{\sum_{j=1}^{M_k} \frac{j}{M_k} \cdot \frac{M_k!}{x_k!(M_k-x_k)!} \rho_k^{x_k} \cdot \sum_{\mathbf{x}:x_k=0} \chi(\mathbf{x})} \cdot \nu\sigma \\
&= \frac{\sum_{j=0}^{M_k-1} \frac{(M_k-1)!}{j!(M_k-1-j)!} \rho_k^j \cdot \sum_{\mathbf{x}:x_k=0} \chi(\mathbf{x})}{\frac{1}{M_k} \rho_k \sum_{j=1}^{M_k} j \rho_k^{j-1} \frac{M_k!}{j!(M_k-j)!} \cdot \sum_{\mathbf{x}:x_k=0} \chi(\mathbf{x})} \cdot \nu\sigma \\
&= M_k \frac{G[\mathbf{M} - \mathbf{e}_k]}{G[\mathbf{M}]} \frac{\nu\sigma}{\rho_k} \\
&= M_k \frac{G[\mathbf{M} - \mathbf{e}_k]}{G[\mathbf{M}]} R_k.
\end{aligned}$$

Take the same example as above and consider class 1,

$$\chi(\mathbf{x}) = x! \cdot \frac{2!}{x_1!(2-x_1)!}\,\rho_1^{x_1} \cdot \frac{2!}{x_2!(2-x_2)!}\,\rho_2^{x_2},$$

$$\begin{array}{lll}
\chi(0,0) = 1 & \chi(1,0) = 2\rho_1 & \chi(2,0) = 2\rho_1^2 \\
\chi(0,1) = 2\rho_2 & \chi(1,1) = 8\rho_1\rho_2 & \chi(2,1) = 12\rho_1^2\rho_2 \\
\chi(0,2) = 2\rho_2^2 & \chi(1,2) = 12\rho_1\rho_2^2 & \chi(2,2) = 24\rho_1^2\rho_2^2
\end{array}$$

While $\mathbf{M} \to \mathbf{M} - \mathbf{e}_1$, we get similarly,

$$\chi(\mathbf{x}) = x! \cdot \frac{1!}{x_1!(1-x_1)!}\,\rho_1^{x_1} \cdot \frac{2!}{x_2!(2-x_2)!}\,\rho_2^{x_2},$$

$$\begin{array}{ll}
\chi(0,0) = 1 & \chi(1,0) = \rho_1 \\
\chi(0,1) = 2\rho_2 & \chi(1,1) = 4\rho_1\rho_2 \\
\chi(0,2) = 2\rho_2^2 & \chi(1,2) = 6\rho_1\rho_2^2
\end{array}$$

From the above results, we get,

$$\begin{aligned}
G[\mathbf{M}] &= 1 + 2\rho_2 + 2\rho_2^2 + 2\rho_1 + 8\rho_1\rho_2 + 12\rho_1\rho_2^2 + 2\rho_1^2 + 12\rho_1^2\rho_2 + 24\rho_1^2\rho_2^2 \\
\frac{\partial}{\partial\rho_1}[\mathbf{M}] &= 2 + 8\rho_2 + 12\rho_2^2 + 4\rho_1 + 24\rho_1\rho_2 + 48\rho_1\rho_2^2 \\
G[\mathbf{M}-\mathbf{e}_i] &= 1 + 2\rho_2 + 2\rho_2^2 + \rho_1 + 4\rho_1\rho_2 + 6\rho_1\rho_2^2.
\end{aligned}$$

With equation (4.9),

$$\begin{aligned}
\gamma_1 &= \frac{1-p_{11}}{p_{11}}\,\nu\sigma \\
&= \frac{\frac{2}{2}\cdot(1+2\rho_2+2\rho_2^2) + \frac{1}{2}\cdot(2\rho_1+8\rho_1\rho_2+12\rho_1\rho_2^2)}{\frac{1}{2}\cdot(2\rho_1+8\rho_1\rho_2+12\rho_1\rho_2^2) + \frac{2}{2}\cdot(2\rho_1^2+12\rho_1^2\rho_2+24\rho_1^2\rho_2^2)} \\
&= 2\cdot\frac{1+2\rho_2+2\rho_2^2+\rho_1+4\rho_1\rho_2+6\rho_1\rho_2^2}{2+8\rho_2+12\rho_2^2+4\rho_1+24\rho_1\rho_2+48\rho_1\rho_2^2}\cdot\frac{\nu\sigma}{\rho_1} \\
&= 2\cdot\frac{G[\mathbf{M}-\mathbf{e}_i]}{G[\mathbf{M}]}\cdot R_1 \\
&= M_1\,\frac{G[\mathbf{M}-\mathbf{e}_i]}{\frac{\partial}{\partial\rho_1}[\mathbf{M}]}\,R_1.
\end{aligned}$$

Appendix B

How to calculate the load (ρ_O) of the OCOF model

The intended load in the QS regime

In each user position constellation $\mathbf{m}$, with the probability $P_r[\mathbf{m}]$ the intended load in the QS regime is:

$$\rho[\mathbf{m}] = \sum_{k=1}^{K} m_k \rho_k = \sum_{k=1}^{K} m_k \cdot \frac{\nu\sigma}{R_k}, \tag{7.1}$$

where $\sum_{k=1}^{K} m_k = n$.

Then the average intended load in the QS regime is:

$$\rho_{int}^{\text{qs}} = \sum_{\mathbf{m}} P_r[\mathbf{m}]\rho[\mathbf{m}]. \tag{7.2}$$

The real load in the QS regime

In each user position constellation $\mathbf{m}$, with the probability $P_r[\mathbf{m}]$ the real load in the QS regime is:

$$\rho[\mathbf{m}] = \sum_{k=1}^{K} \sum_{j=0}^{m_k} \pi(y_k = j) \cdot j \cdot \frac{\nu\sigma}{R_k}. \tag{7.3}$$

So the average real load in the QS regime is:

$$\rho_{real}^{\text{qs}} = \sum_{\mathbf{m}} P_r[\mathbf{m}]\rho[\mathbf{m}]. \tag{7.4}$$

Alternatively we can calculate the real load in another way as follows:

In each user position constellation $\mathbf{m}$, with the probability $P_r[\mathbf{m}]$ the real load in the QS regime is:

$$\rho[\mathbf{m}] = \sum_{k=1}^{K} \rho_k = \sum_{k=1}^{K} \frac{\lambda_k \sigma}{R_k}, \tag{7.5}$$

where,

$$\lambda_k = m_k \frac{G[\mathbf{m} - \mathbf{e}_k]}{G[\mathbf{m}]} \nu. \tag{7.6}$$

Then the average real load in the QS regime is calculated as in equation (7.4).

The intended load in the FL regime

In the FL regime, all the customers have the following full service rate:

$$R = \sum_{k=1}^{K} p_k R_k \tag{7.7}$$

Then the intended load in the FL regime is:

$$\rho_{int}^{\text{fl}} = n \cdot \frac{\nu\sigma}{R}. \tag{7.8}$$

The real load in the FL regime

The real load in the FL regime is:

$$\rho_{real}^{\text{fl}} = \sum_{j=0}^{n} \pi(y = j) \cdot j \cdot \frac{\nu\sigma}{R}. \tag{7.9}$$

We can also calculate it as follows:

$$\rho_{real}^{\text{fl}} = \frac{\bar{\lambda}\sigma}{R}, \tag{7.10}$$

where,

$$\bar{\lambda} = n \frac{G[n-1]}{G[n]} \nu. \tag{7.11}$$