A Special $D^{[X]}/D/1/m$ Queueing System: A Correlated Arrival Process

Thesis

Written by:
Árpád Szlávik

Supervisor:
László Szeidl
associate professor

Eötvös Loránd University of Budapest
Faculty of Science

Department of Probability Theory and Statistics

1998.

Contents

List of Figures

Chapter 1

Introduction: The general description

In the beginning the need for a new mathematical apparatus for describing telecommunication systems helped the birth of the queueing theory. Yet the development and spread of the theory of queues was due to the high utilization capability of its models and methods everywhere where queues appeared. Meanwhile the theory of mass services remained to be a strong machinery for analysing the performance of telecommunication networks and its evolution is still highly encouraged by the appearance of new networks and new ways of communication.

In the time of computers and Internet, fiber optics, mobile phone systems and video phones telecommunication is not only telephony any more and as a result using different ways of communication at the same time became necessary. As the communication forms a traffic in the networks, having various ways of communication at the same time is referred to as having different sort of traffic flows in the same network. Two classes of telecommunication networks are traditionally distinguished in digital communication:
circuit-switched and **packet-switched network**s.

There are two key-words describing circuit-switched networks:

– **connection**, that is a path in the network established between the source and the destination before the start of the communication but used for the whole time of the communication,

– **fixed resource allocation**, that means a dedicated fixed capacity (**bandwidth**, **transmission rate**) between the source and the destination for the established connection and for the whole communication. The deterministic resource allocation means that the circuit-switching technique provides a very simple connection (in technological terms) and good quality of communication (the **P**ublic **S**witched **T**elephone **N**etworks – **PSTN**, the traditional telephone network – is a well-known circuit-switched type network). On the other hand this technique wastes resources because a fixed bandwidth is allocated irrespectively to the utilization (if one picks up the phone and gets a connection but remains silent, the same amount of resources are allocated as if she/he would speak).

In packet-switched networks **packet**s are carrying the user information. The packets can be routed individually from the source to the destination – there is no dedicated path for the data transfer, while the original sequence of the packets is restored at the receiving terminal. In packet-switched networks packet loss or long delays can occur but the packet-switching technique is "efficient" since the network is only loaded when there is user data to be transmitted at the sending terminal (**IP** – **I**nternet **P**rotocol – network is an example for packed-switched networks).

A probable solution for networks that are capable of providing a wide variety of communication services with different traffic characteristics (bandwidth, quality) may be half-way between circuit-switched and packet-switched networks like in **ATM network**s (ATM stands for **A**synchronous **T**ransfer **M**ode). ATM sets up connections but does not allocate peak bandwidth (efficiency) and for this uses some "good" features of both the traditional networks:

– Data is transmitted in small packets so the transmission rate has fine granularity (large and fine scale of bandwidth can be achieved).

– Connection (**virtual connection**) is built up on the basis of the so-called **traffic contract**, that is a conditional contract between a user and the network coordinator: the user demanding a certain **q**uality **o**f **s**ervice – **QoS** (small packet loss probability, small delays); the coordinator demanding the traffic characteristics of the user's communication (traffic type, transmission rate), and as a result a guaranteed service quality can be achieved.

– The packets of the ATM are fixed size packets called **cell**s – the size of a cell is 53 octets (octet=8 bits, where octet is the used terminology for byte in ATM) with a 5 octet **header** and a 48 octet **payload**. Data packets are segmented and packed into ATM cells. After that streams of ATM cells are entering the ATM network and are sent by the server through **ATM link**s (their capacity, the transmission rate, shortly the **rate** is measured for example in **bit/secundum** or **cell/secundum**) to the destination equipment for depaketization. The traffic data is packed into the payload part only but the header consists of the so-called: virtual path identifier (VPI), virtual channel identifier (VCI), payload type (PT), reserved field (Res), cell loss priority (CLP) and header error check (HEC) – if there would not be the header part we could think of ATM cells as "bigger bits". There are **ATM multiplexer**s in ATM networks as well, where queues can appear in the so-called **buffer** part of it and these multiplexers serve for handling (multiplexing) a number of possibly different traffic flows at the same time.

ATM is a technology which allows for the transition of data, voice and video traffic simultaneously in the same transition environment (**Figure 1.1**) though in first approach it seems to be efficient for the so-called **broadband** (large bandwidth) services only. It is because the 48 octet payload is not small enough for the **narrowband** services like compressed voice of telephones as for the mobile phones (if we want to fill the whole cell) since large packetization delays can appear but that is not acceptable for the voice traffic (more delays can occur in the air interface that we can not influence so we have to manage well the delays we can have influence on).

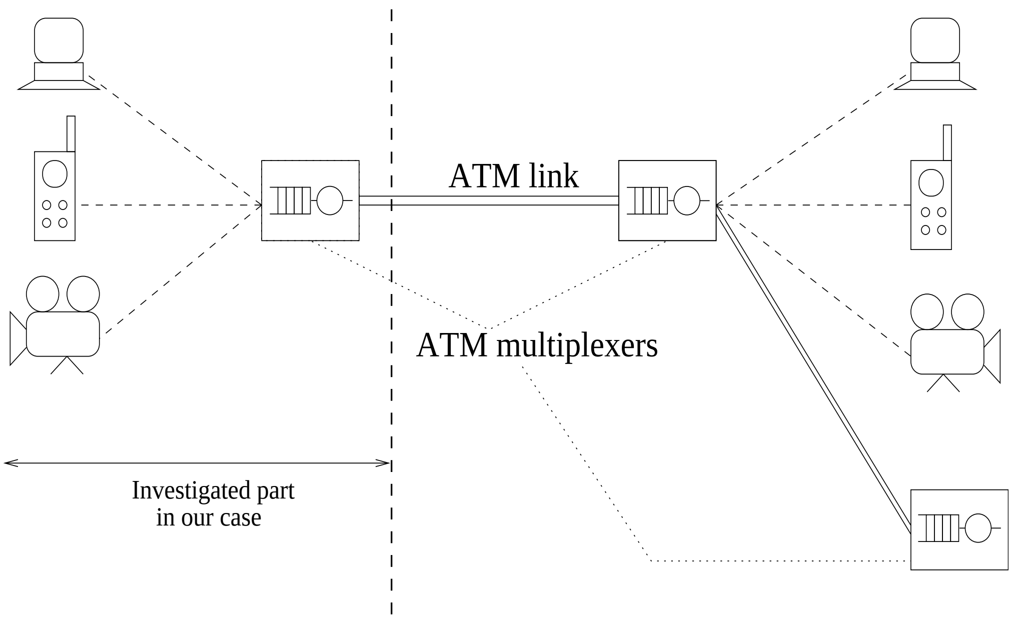

Figure 1.1: ATM

An **A**TM **A**daptation **L**ayer called **AAL2** (there exists AAL1, AAL3/4 and AAL5 as well) is standardized recently that provides the bandwidth-efficient transition of low bit-rate, delay-sensitive applications:

– transition packets are the **minicell**s smaller than ATM cells (size of at most 24 octets of payload and 3 octets of header) that gives even finer granularity for the bandwidth and enables low packetization delay,

– because of the variable size of the minicells the little packets generated by the **voice coders** (coding the voice into digital data), producing variable amount of data in a time unit, can be transmitted efficiently.
([McDS94])

AAL2 will be used in the access network of the **3-rd generation mobile phone system**s (the so-called analogous NMT mobile system being an example for the 1-st and the **GSM** – **G**lobal **S**ystem for **M**obile Communications being an example for the 2-nd one). In the recently standardized AAL2 based mobile system voice packets are generated periodically over the radio interface from mobile phones to base stations. Therefore the investigation of queueing systems with periodical (deterministic time) batch (group) arrivals and with deterministic service time, for the sake of simplicity, is justified. These kind of queueing systems will be referred to as $D^{[X]}/D/1$ queueing models using the notation of [BLT96], where the unusual $[X]$ superscript emphasises the fact of group arrivals of the customers (it is a sequence of non-negative random variables in the discrete-time model, hence the model is stochastic, nevertheless the classical $D/D/1$ notation of *Kendall* ([Kle75]) would stand for deterministic models).

Due to the voice coding in some types of mobile terminals (GSM for example) voice can be modelled as an **on/off type traffic** – data packets generated by the source during the on-periods, no packets generated during the off-periods, and the traffic stream consists of alternating on and off-periods (**Figure 1.2**).

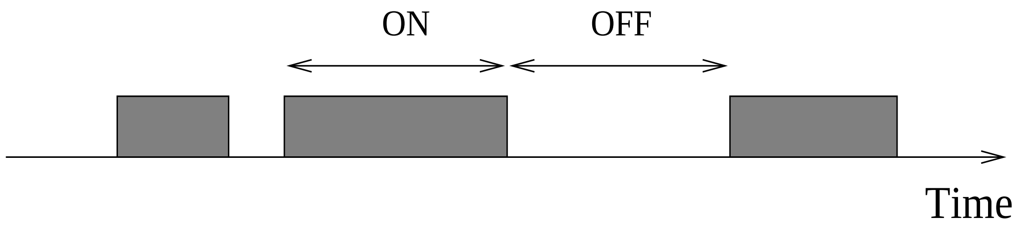

Figure 1.2: On/off type traffic

Voice traffic is **bursty** (having a two party conversation after speaking the probability of speaking for a while is greater than becoming silent – from a chosen party's point of view), in other words there is correlation inherent in the flow of packets issued by an individual voice coder. This results in a correlation between the number of successively generated voice packets when modelling it by $D^{[X]}/D/1$ queues, thus we can not assume **i**ndependent, **i**dentically **d**istributed (**i. i. d.**) sequence of batches (random variables) arriving to the queue.

In this thesis we are investigating a special $D^{[X]}/D/1/m$ queueing model with $m \leq \infty$ storage capacity available in the system and a special non i. i. d. batch arrival (the superposition of some simple on/off type traffic arriving to the ATM multiplexer). The precise model is given in *Section 2.1*.

In *Chapter 2* we are going to investigate the **conditions of steadiness** for our queueing system for both the $m = \infty$ and $m < \infty$ cases. We are interested in the **number of customers in the system** or in other words the **queue length**. This being a random variable we seek its probabilistic description (the mean value, the average behaviour of its steady-state distribution for example). In the case $m < \infty$ we also need the **probability of rejection** for a customer (the probability of having no space for waiting left more in the system, namely the probability that a customer finds the buffer full).

In ATM's transport mechanism and especially for packet-based voice transmission the quality of service to the users is quite often quantified by the probability of dropping packets due to buffer saturation so the analysis of the probability of rejection in the $m < \infty$ case provides guidelines for the right buffer sizing of ATM multiplexers. The tail distribution of the queue length is generally considered to be one of the most important performance measures of a queueing system of infinite storage capacity. Althought in practice ATM multiplexers have naturally finite storage capacity (finite buffer) the investigation of the infinite case is justified because proper dimensioning of buffers can be evaluated from the tail distribution.

The results can be used for uderstanding the behaviour of mobile generated voice traffic in AAL2/ATM from mobile phones to base stations.

In *Chapter 3* the precise model descriptions and the summarized results of some papers are given. These papers study ATM multiplexers and on/off sources but they are discussed in the *Appendix* because of their weak connection to the thesis.

In [BLT96] a $D^{[X]}/D/1/m$, $m \leq \infty$ queueing system is considered with an i. i. d. sequence of batch arrivals. In [Bru88] and [Vit86] a $D^{[X]}/D/1/\infty$ queue is examined, a $D^{[X]}/D/K/m$, $K \geq 1$, $m < \infty$ queue is considered in [JPN80] – all of these queueing systems are treating batch arrivals arising from the superposition of some on/off sources. In [WB97] a $D^{[X]}/D/1/\infty$ system is investigated with group arrivals originating from some generalized on/off sources. A $G^{[X]}/D/1/m$, $m < \infty$ queue with two different blocking policies is studied in [TGA88], some $G^{[X]}/D/1/\infty$ queues are described in [RV91]. Finally in [WB94] a thorough description is given of a special $D^{[X]}/D/1/m$, $m \leq \infty$ with a single on/off source generating the batch arrival (not only the models differ from each other in these papers but the model conditions as well).

Acknowledgements

I would like to thank *László Szeidl* for the useful help in preparing the thesis. I would also like to thank *András Gergely Valkó*[1] for representing the model for the thesis and *Szabolcs Malomsoky*[1] for the aid in searching some related works and for the help in investigating the practical use of the mathematical model. I am also grateful for the assistance of *Ferenc Máthé*[1] in describing the technical environment of the thesis.

The thesis was supported by *Ericsson Telecommunications Ltd., Traffic Analysis and Network Performance Laboratory.*

[1]Research engineers, Ericsson Telecommunications Ltd.

Chapter 2

The investigated model and the results

2.1 A special $D^{[X]}/D/1/m$ queue

The following $D^{[X]}/D/1/m$ queueing system is considered in the thesis:

The arrival process

• The **interarrival time** of the customers is constant Δ time units (the arrival process is deterministic), the n-th arrival occurs at time

$$t_n \stackrel{def}{=} t_0 + n\Delta,\ n \geq 1$$

(t_0 is the **initial time** of the queueing system, the time when we first look at the system).

• Denote the number of customers arriving in t_n by $Z_n,\ n \geq 1$.
Z_n originates from $N \geq 1$ integer number of customer generating units,

$$Z_n \stackrel{def}{=} X_1(t_n) + X_2(t_n) + \ldots + X_N(t_n),\ \ n \geq 1,$$

where the i-th generating unit, $i = 1, \ldots, N$ generates $X_i(t_n)$ number of customers in the time $t_n,\ n \geq 1$. $Z_n,\ n \geq 0$ and $X_i(t_n),\ i = 1, \ldots, N,\ n \geq 0$, are random variables on the $(\Omega, \mathcal{A}, Pr)$ (Kolmogorov-type) probability space and all the random variables, stochastic processes and sequences will be interpreted in the future on this same probability space.

Let us denote

$$r_{a,b}^{n,n+1}(i) \stackrel{def}{=} Pr\{X_i(t_{n+1}) = b|X_i(t_n) = a\},$$

$$n \geq 0,\ a, b \in \{0, 1\},\ i = 1, \ldots, N,$$

where $X_i(t_0)$ is the **initial state** of the customer generating unit

($Z_0 \stackrel{def}{=} X_1(t_0) + X_2(t_0) + \ldots + X_N(t_0)$ is the initial state of the arrival process, the number of customers arriving in the system in time t_0). We have the model conditions for these probabilities

$$r_{a,b}^{n,n+1}(i) = r_{a,b}^{0,1}(i) = r_{a,b}^{0,1}(1) \stackrel{def}{=} r_{a,b},$$

$$\text{for all } n \geq 1,\ a, b \in \{0, 1\},\ i = 1, \ldots, N,$$

which means the following two conditions:

1. $\{X_i(t_n),\ n \geq 0\}$ for all $i = 1, \ldots, N$ are **homogeneous discrete-time Markov chain**s by definition, i. e.

   $$Pr\{X_i(t_{n+1}) = a_{n+1}|X_i(t_0) = a_0, \ldots, X_i(t_n) = a_n\} =$$

   $$= Pr\{X_i(t_{n+1}) = a_{n+1}|X_i(t_n) = a_n\} = r_{a_n,a_{n+1}}^{n,n+1}(i),$$

   for all $n \in \tau$, for the τ **time space** of the Markov chain ($\tau \stackrel{def}{=} \{0, 1, \ldots\}$) and $a_i \in S_X$ ($0 \leq i \leq n + 1$), for the S_X **state space** of the Markov chain ($S_X \stackrel{def}{=} \{0, 1\}$), and the homogenity means that these probabilities does not depend on n

   ($r_{a,b}^{n,n+1}(i) = r_{a,b}(i)$, for all $n \geq 0,\ i = 1, \ldots, N,\ a, b \in \{0, 1\}$).

2. The $\{X_i(t_n),\ n \geq 0\}$, $i = 1, \ldots, N$ Markov chains are equally distributed

   ($r_{a,b}(i) = r_{a,b},\ i = 1, \ldots, N,\ a, b \in \{0, 1\}$).

The given **one-step transition probabilities** of the Markov chains are

$$R = \begin{pmatrix} r_{0,0} & r_{0,1} \\ r_{1,0} & r_{1,1} \end{pmatrix} \stackrel{def}{=} \begin{pmatrix} 1 - p_1 & p_1 \\ p_0 & 1 - p_0 \end{pmatrix}$$

for fixed $0 < p_0, p_1 < 1$ (**Figure 2.1** on the next page).

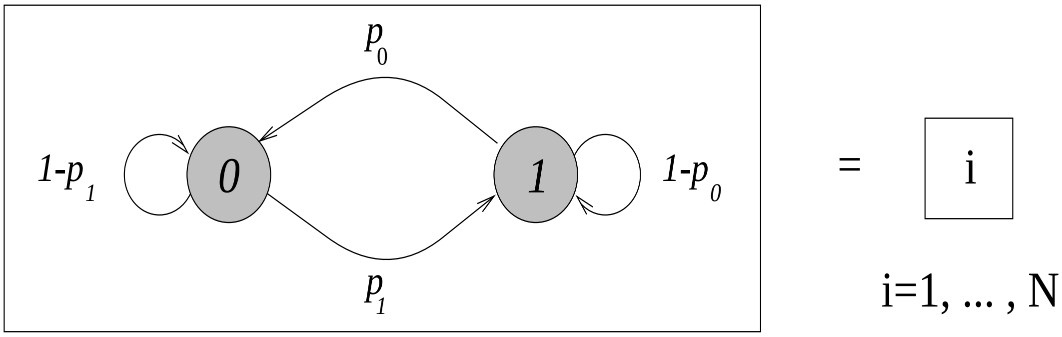

Figure 2.1: A customer generating unit

• The generation of the customers by a generating unit only depends on the generating state the unit is in and does not depend on the states of the other generating units (independence of the generating units).

The structure of the service facility

• Customers arrive in a single-server queue (**number of service stations** equals 1) with m **storage capacity** available, where $m \leq \infty$.

• The **service time** of a customer is constant δ time units.

• The server can serve up to K customers within the period Δ, where $K \geq 1$ is a fixed integer (hence $0 \leq K\delta \leq \Delta$ must hold true).

The discipline of the service facility

• Provided that there is storage capacity available, the customers of a group are accepted in the queue in random order until either the customers in the group are exhausted or the queue fills to its capacity.

• Customers that are accepted by the queue and belong to different arrival groups are scheduled to be served in **f**irst-**c**ome-**f**irst-**s**erved (**FCFS**) order, while customers within the same group are served in order of acceptance.

The scheme of the queue is depicted in **Figure 2.2** with the units of **Figure 2.1** (previous page).

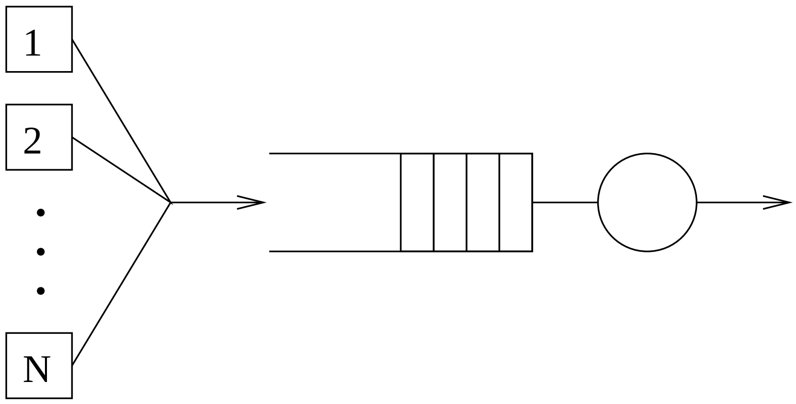

Figure 2.2: A $D^{[X]}/D/1/m$ queue

REMARK: Notice that the customer generating units can be defined in the next equivalent manner as well:
Each source ($i = 1, \dots, N$) generates customers in the following same way:
There is X and Y which are two independent geometrically distributed random variables with the corresponding probabilities

$a(n) = Pr\{X = n\} = p_0(1 - p_0)^{n-1}$,
$b(n) = Pr\{Y = n\} = p_1(1 - p_1)^{n-1}$, for all $n \geq 1$.

We produce two sequence of i. i. d. random variables $X_i^{(1)}$, $X_i^{(2)}$, $\dots$ and $Y_i^{(1)}$, $Y_i^{(2)}$, $\dots$ equally distributed to X and Y respectively and these construct the length of the on-periods and the off-periods of a the i-th generating unit. An $n \geq 1$ long on-period means that the generating unit generates exactly one customer per Δ time units for n consecutive time slots and an n-long off-period means that the source does not generate customer for n consecutive time units. Thus the production of customers by the i-th so called **on/off source** is described by the

$$X_i^{(1)},\ Y_i^{(1)},\ X_i^{(2)},\ Y_i^{(2)},\ \dots$$

sequence of random variables (alternating on/off periods) and the arrival process is determined from the superposition of this N number of infinite $0, 1$ sequences each started with a chosen t_0 initial time, hence Z_n equals the number of sources being in on-phase in time t_n (see **Figure 2.3** on the top of the next page) like in [Bru88] and [WB97].

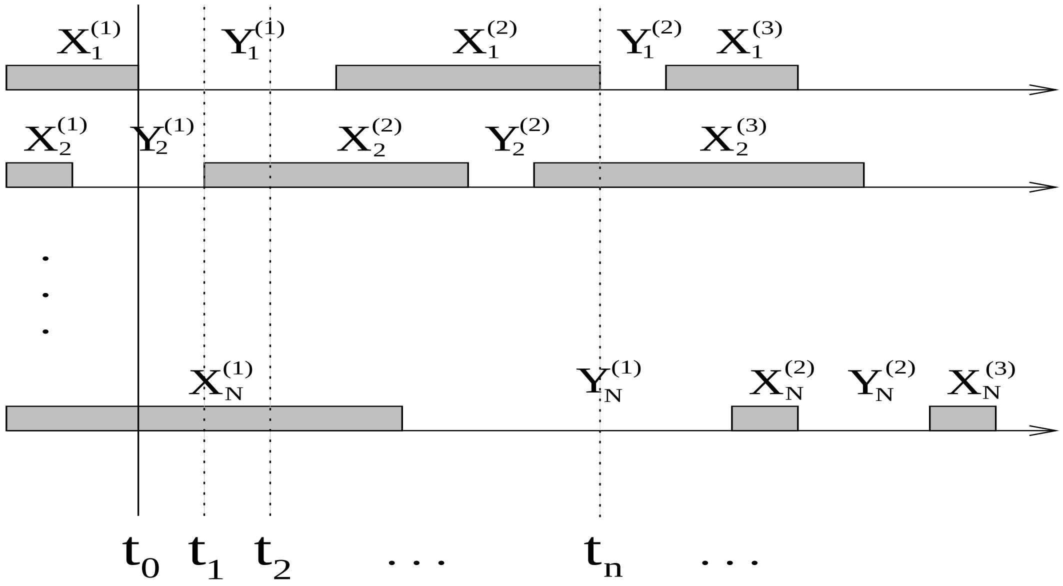

Figure 2.3: Superposition of on/off type sources

REMARK: Notice that the server of capacity: $K \geq 1$ customers of δ service time served within the Δ time period is equivalent to K number of parallel servers with 1 customer of $K\delta$ service time to be served within the same Δ period of time in this discrete-time model (**Figure 2.4**) like in [JPN80].

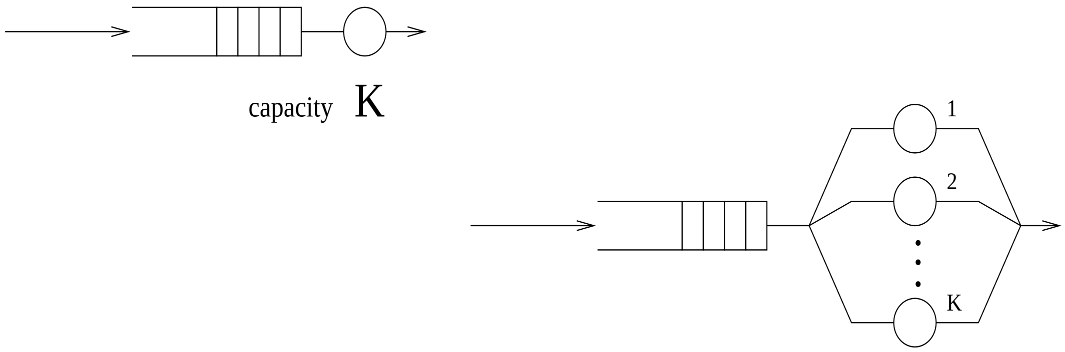

Figure 2.4: Equivalent queues

2.2 The arrival process

Consider the arrival process described in the previous section, namely the $\{Z_n = X_1(t_n) + \ldots + X_N(t_n),\ n \geq 1\}$ sequence of random variables with an arbitrarily fixed $\mathbf{X^{(0)}} \stackrel{def}{=} (X_1(t_0), \ldots, X_N(t_0))^T$ initial state vector, where T stands for the **vector transpose**, and where $X_i(t_0) \in \{0,1\}$ (thus for $Z_0 = X_1(t_0) + \ldots + X_N(t_0)$ the $Z_0 \in \{0, \ldots, N\}$).

For fixed $\mathbf{X^{(0)}}$ the $X_i(t_n),\ i = 1, \ldots, N$ random variables are trivially independent of each other for all $n \geq 1$ (*independence of the generating units*). Consequently the $\{Z_n = X_1(t_n) + \ldots + X_N(t_n),\ n \geq 0\}$ sequence of random variables constitutes a **homogeneous** discrete-time Markov chain by definition, i. e.

$$Pr\{Z_{n+1} = i_{n+1} | Z_0 = i_0, Z_1 = i_1, \ldots, Z_n = i_n\} =$$

$$= Pr\{Z_{n+1} = i_{n+1} | Z_n = i_n\},$$

$$\text{for all } i_0, i_1, \ldots, i_{n+1} \in S_Z \text{ and}$$

$$n \in \tau,$$

where $\tau = \{0, 1, \ldots\}$ is the time space of the Markov chain and

$$S_Z \stackrel{def}{=} \{0, 1, \ldots, N\}$$

is the state space of the Markov chain (the homogenity means that these probabilities does not depend on n). Therefore the arrival process is fully determined by the Z_0 initial state and the one-step transition probabilities:

$$p_{l,k} \stackrel{def}{=} Pr\{Z_{n+1} = k | Z_n = l\},$$

$$k, l \in S_Z,$$

$$n \in \tau.$$

REMARK: This means that our N identical, independent (for fixed $\mathbf{X^{(0)}}$ initial state vector) generating units can be replaced by a single generating unit of $N+1$ corresponding states with the $Z_0 = \sum_{i=0}^{N} X_i(t_0)$ initial state and with the $p_{l,k}$ transition probabilities (**Figure 2.5** on the next page).

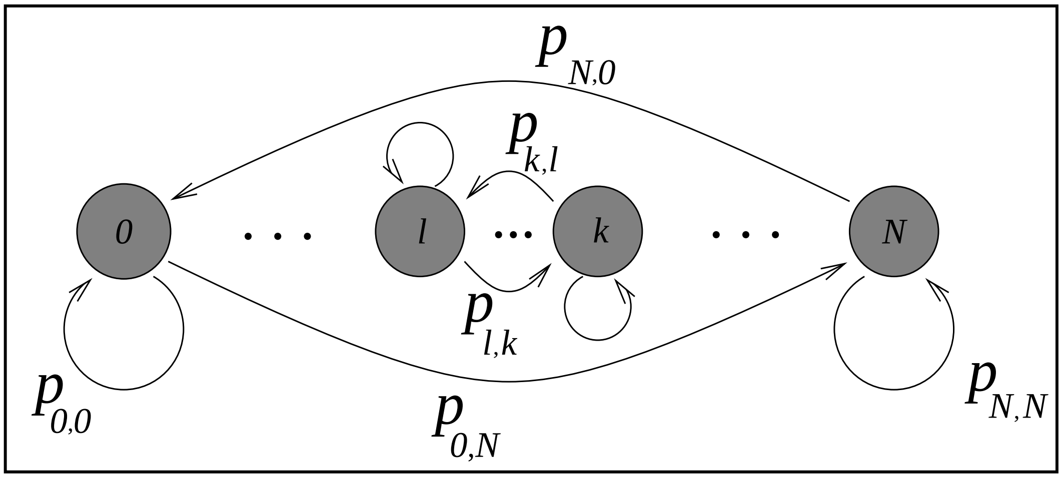

Figure 2.5: Single generating unit

The **one-step transition probability matrix** $P = \{p_{l,k}\}_{l,k=0}^{N}$ of the homogeneous Markov chain evolves as follows:

$$p_{l,k} = \sum_{i=0}^{\min\{l,k\}} \binom{l}{i}(1-p_0)^i p_0^{l-i} \binom{N-l}{k-i} p_1^{k-i}(1-p_1)^{N-l-k+i},$$

$$k, l = 0, 1, \ldots, N \quad (k, l \in S_Z).$$

Control: For fixed $l = 0, 1, \ldots, N$ the sum of the probabilities $p_{l,k}$ have to be 1:

$$\sum_{k=0}^{N} p_{l,k} = \sum_{k=0}^{l}\sum_{i=0}^{k} \binom{l}{i}(1-p_0)^i p_0^{l-i} \binom{N-l}{k-i} p_1^{k-i}(1-p_1)^{N-l-k+i} +$$

$$+ \sum_{k=l+1}^{N}\sum_{i=0}^{l} \binom{l}{i}(1-p_0)^i p_0^{l-i} \binom{N-l}{k-i} p_1^{k-i}(1-p_1)^{N-l-k+i} \stackrel{1}{=}$$

$$\stackrel{1}{=} \sum_{i=0}^{l} \binom{l}{i}(1-p_0)^i p_0^{l-i} \sum_{k=i}^{l} \binom{N-l}{k-i} p_1^{k-i}(1-p_1)^{N-l-k+i} +$$

$$+ \sum_{i=0}^{l} \binom{l}{i}(1-p_0)^i p_0^{l-i} \sum_{k=l+1}^{N} \binom{N-l}{k-i} p_1^{k-i}(1-p_1)^{N-l-k+i} \stackrel{2}{=}$$

$$\stackrel{2}{=} \sum_{i=0}^{l} \binom{l}{i}(1-p_0)^i p_0^{l-i} \sum_{s=0}^{N-i} \binom{N-l}{s} p_1^{s}(1-p_1)^{N-l-s} \stackrel{3}{=}$$

$$\stackrel{3}{=} \sum_{i=0}^{l} \binom{l}{i}(1-p_0)^i p_0^{l-i} \sum_{s=0}^{N-l} \binom{N-l}{s} p_1^{s}(1-p_1)^{N-l-s} \quad \stackrel{4}{=} \quad 1,$$

where the $\overset{1}{=}$ equation stands because of changing the summation, the $\overset{2}{=}$ because introducing new summation variable for $k - i \overset{def}{=} s$, the $\overset{3}{=}$ because for $s > N - l$ the $\binom{N-l}{s}$ is defined to be 0 and the $\overset{4}{=}$ because the second sum is the sum of all the probabilities of the binomial distribution with $(N - l, p_1)$ parameters and then the first sum summarizes the probabilities of the binomial distribution with (l, p_0) parameters. $\square$

We can show the following properties of the Markov chain:

- The $\{Z_n,\ n \geq 0\}$ Markov chain is **irreducible**:

For all $i, j \in S_Z$ there exists $n \in \tau : p_{i,j}^n > 0$. That is because $p_{i,j}^n$ is a **n-step transition probability**, so element of the P^n matrix.

- The $\{Z_n,\ n \geq 0\}$ Markov chain is **aperiodic**:

For all $i, j \in S_Z$ the greatest common divisor of the numbers $\{n \in \tau : p_{i,j}^n > 0\}$ equals unity (it is evident from the matrices P^n).

PROPOSITION 1 *The $\{Z_n = X_1(t_n) + \ldots + X_N(t_n),\ n \geq 0\}$ discrete-time finite-state Markov chain has a unique $\pi = (\pi_0,\ \pi_1,\ \ldots,\ \pi_N)^T$* **ergodic** *distribution, i. e. a π* **stationary distribution***, that is*

$$\begin{aligned} \textit{for all } i, j \in S_Z \textit{ there exists } \pi_j \overset{def}{=} \lim_{n\to\infty} p_{i,j}^n, \\ \textit{for all } i \in S_Z :\ \pi_i \geq 0, \\ \sum_{i\in S_Z} \pi_i = 1, \end{aligned} \tag{2.1}$$

$$\sum_{i\in S_Z} \pi_i p_{i,j}^n = \pi_j,\ \textit{for all } n \geq 1 \textit{ and } j \in S_Z, \tag{2.2}$$

and for this (limiting) stationary distribution $\pi_i > 0$, for all $i \in S_Z$. This π ergodic distribution does not depend on the Z_0 initial state.

Proof: A homogeneous, irreducible, aperiodic, finite-state Markov chain is ergodic (See [KT75] or [Kli86]). $\square$

2.2.1 The ergodic distribution of the arrival process

Consider the $\{Z_n,\ n \geq 0\}$ Markov chain (arrival process) for fixed $N \geq 1$, $0 < p_0, p_1 < 1$ and an arbitrarily fixed Z_0 initial state. The

$\pi = (\pi_0, \ldots, \pi_N)^T$ ergodic distribution of the Markov chain is the unique solution of the equations (2.1) and (2.2), so

$$\sum_{i=0}^{N} \pi_i p_{i,j} = \pi_j, \quad j = 0, \ldots, N \quad (\pi^T P = \pi^T),$$

$$\sum_{i=0}^{N} \pi_i = 1.$$

PROPOSITION 2 *For fixed $N \geq 1,\ 0 < p_0, p_1 < 1$ the $\pi = (\pi_0, \ldots, \pi_N)^T$ unique ergodic distribution of the $\{Z_n,\ n \geq 0\}$ Markov chain is a binomial distribution with $(N, \dfrac{1}{1+q})$ parameters, where*

$$q \stackrel{def}{=} \frac{p_0}{p_1},$$

therefore

$$\pi_i = \binom{N}{i} \Big(\frac{1}{1+q}\Big)^i \Big(1 - \frac{1}{1+q}\Big)^{N-i}, \quad 0 \leq i \leq N.$$

Proof: The ergodic distribution does not depend on the Z_0 initial state so we can assume the $(0, \ldots, 0)^T$ initial state vector for the N generating units ($Z_0 = 0$). Having this assumption the number of customers generated by the units in any time $t_n,\ n \geq 1$ ($X_i(t_n),\ i = 1, \ldots, N$) are independent of each other, the number of customers generated by the generating units are identically distributed in addition, hence it is enough to calcutate the $\Pi \stackrel{def}{=} (\Pi_0, \Pi_1)^T$ ergodic distribution of the $\{X_1(t_n),\ n \geq 0\}$ Markov chain. It is enough because the limiting distribution, too, is the sum of N number of independent, identically distributed random variables, therefore the π ergodic distribution is then the **N-th convolutional power** of Π, i. e.

$$\pi_i = \binom{N}{i} \Pi_1^i \Pi_0^{N-i}, \quad 0 \leq i \leq N.$$

The Π is the solution of the following set of equations ("$N = 1$" for the $P = R$ transition probability matrix and for equations (2.1) and (2.2)):

$$(\Pi_0, \Pi_1) \begin{pmatrix} r_{0,0} & r_{0,1} \\ r_{1,0} & r_{1,1} \end{pmatrix} = (\Pi_0, \Pi_1),$$

$$\Pi_0 + \Pi_1 = 1.$$

Substituting $r_{0,0} = 1 - p_1,\ r_{0,1} = p_1,\ r_{1,0} = p_0,\ r_{1,1} = 1 - p_0$ in the first equation we get the system:

$$\begin{array}{rcrcl} -p_1\Pi_0 & + & p_0\Pi_1 & = & 0, \\ \Pi_0 & + & \Pi_1 & = & 1. \end{array}$$

Then

$$\Pi_0 = \frac{p_0}{p_0 + p_1} = \frac{q}{q+1},$$

$$\Pi_1 = \frac{p_1}{p_0 + p_1} = \frac{1}{q+1},$$

and that completes the proof. $\square$

REMARK: Realize that $p_0 + p_1 = 1$ condition means that the $\{Z_n,\ n \geq 0\}$ Markov chain is an independent, identically distributed sequence of random variables (the number of customers produced in time t_n does not depend on the number generated in time $t_{n-1},\ n \geq 1$).

We can derive the following equation for the number of arrivals:

$$E\{Z\} \stackrel{def}{=} \lim_{n\to\infty} E\{Z_n\} = N\frac{1}{1+q}, \tag{2.3}$$

where $q = \dfrac{p_0}{p_1}$ and for the Z random variable $E\{Z\}$ is the **mean value**, the **first moment** (the mean value of a binomially distributed random variable is the product of its two parameters).

2.3 Distribution of the queue lenght

Let $L(t)$ denote the number of customers in the system (**the queue length**) in the time $t \geq t_0$. Define $L(t)$ **continuous from left**, i. e.

$$L(t) = \lim_{u \to t-0} L(u) \stackrel{def}{=} L(t-0)$$

and define

$$L_n \stackrel{def}{=} L(t_n) = L(t_n - 0), \; n \geq 1,$$

the number of customers in the system right before the n-th arrival ($L_0 \stackrel{def}{=} L(t_0)$ is the initial state of the queue length, the number of customers in the system in the initial time t_0).

2.3.1 The $m = \infty$ case

Consider $m = \infty$ storage capacity available in our system and the stochastic processes $\{Z_n, \; n \geq 0\}$ and $\{L_n, \; n \geq 0\}$ defined previously with a fixed Z_0 and L_0 initial state. The sequence of $\{\mathbf{Y_n} \stackrel{def}{=} (L_n, Z_n), \; n \geq 0\}$ vector-valued random variables constitutes a discrete-time **homogeneous** Markov chain by definition, i. e.

$$\begin{gathered}
Pr\{(L_n, Z_n) = (l_n, s_n) | (L_0, Z_0) = (l_0, s_0), \ldots \\
\ldots, (L_{n-1}, Z_{n-1}) = (l_{n-1}, s_{n-1})\} = \\
= Pr\{(L_n, Z_n) = (l_n, s_n) | (L_{n-1}, Z_{n-1}) = (l_{n-1}, s_{n-1})\} \stackrel{def}{=} \\
\stackrel{def}{=} b^{n-1,n}_{(l_{n-1}, s_{n-1})(l_n, s_n)}, \\
l_i \geq 0, \; s_i = 0, 1, \ldots, N, \; 0 \leq i \leq n, \quad n \geq 1,
\end{gathered}$$

according to the following recursive evaluation of L_n:

$$L_n = (L_{n-1} + Z_{n-1} - K)^+, \quad n \geq 1,$$

where $x^+ \stackrel{def}{=} max\{0, x\}$, and the homogenity means that these probabilities does not depend on n, hence we can denote the one-step transition probabilities of the Markov chain by

$$b_{(l_{n-1}, s_{n-1})(l_n, s_n)} \stackrel{def}{=} b^{n-1,n}_{(l_{n-1}, s_{n-1})(l_n, s_n)}, \; n \geq 1.$$

REMARK: Notice that if we define $L(t)$ continuously from the right the recursion changes to $L_n = (L_{n-1} - K)^+ + Z_n$.

The state space of the Markov chain is a countable set, denote it by S^∞,

$$S^\infty \stackrel{def}{=} \{(l,s),\ l \geq 0,\ 0 \leq s \leq N\},$$

The $B^\infty = \{b_{(l_i,s_i)(l_j,s_j)}\}, \quad (l_i,s_i),(l_j,s_j) \in S^\infty$, one step transition probability matrix of this Markov chain evolves as follows:

$$b_{(l_i,s_i)(l_j,s_j)} = \begin{cases} p_{s_i,s_j} & ,\text{if} \quad l_j = (l_i + s_i - K)^+ \\ 0 & ,otherwise \end{cases}$$

$$l_i, l_j \geq 0, \quad 0 \leq s_i,\ s_j \leq N.$$

REMARK: We may assume that $N > K$ or else the evolution of $\{L_n\}$ is trivial.

Some properties of the $\{\mathbf{Y_n} = (L_n, Z_n),\ n \geq 0\}$ Markov chain can be derived:

- $\{\mathbf{Y_n},\ n \geq 0\}$ Markov chain is **irreducible**:

Due to the $N > K$ assumption (and $0 \leq s_i \leq N$) for all pairs of states there exists an n time that the n-step transition probability is greater than 0 between these states.

This stands between any chosen state and the $(0,0)$ state as well, besides $p_{0,0} > 0$, so

- $\{\mathbf{Y_n},\ n \geq 0\}$ is an **aperiodic** Markov chain:

For all $(l_i,s_i),(l_j,s_j) \in S^\infty$ $\{n \geq 1 :\ b^n_{(l_i,s_i)(l_j,s_j)} > 0\}$ set is not empty because we have an irreducible Markov chain (here $b^n_{(l_i,s_i)(l_j,s_j)}$ are the n-step transition probabilities, so the elements of the $B^{\infty n}$ matrix) and the greatest common divisor of the elements of the set equals unity since there exists an n that $n+1$ is also in the set (there exists an n that there is an n-step walk between the two states going across the $(0,0)$ state).

Therefore for all $(l_i,s_i),(l_j,s_j) \in S^\infty$ there exists $\lim\limits_{n\to\infty} b^n_{(l_i,s_i)(l_j,s_j)}$ moreover the limes is either 0, for all $(l_i,s_i),(l_j,s_j) \in S^\infty$ or there exists a unique

$\beta \stackrel{def}{=} (\beta_{(0,0)}, \ldots, \beta_{(0,N)}, \beta_{(1,0)}, \ldots, \beta_{(1,N)}, \ldots)^T$

stationary distribution being ergodic as well:

$\lim\limits_{n\to\infty} b^n_{(l_i,s_i)(l_j,s_j)} \stackrel{def}{=} \beta_{(l_j,s_j)} > 0, \quad$ for all $(l_i,s_i),(l_j,s_j) \in S^\infty$.

([KT75] or [Kli86], Chapter 3., §3., Theorem 1. and 3.)

A sufficient condition for the ergodicity

THEOREM 3 *If* $N\dfrac{1}{1+q} < K < N$ *then the* $\{\mathbf{Y_n} = (L_n, Z_n),\ n \geq 0\}$ *Markov chain has a unique*

$$\beta = (\beta_{(0,0)}, \ldots, \beta_{(0,N)}, \beta_{(1,0)}, \ldots, \beta_{(1,N)}, \ldots)^T$$

ergodic distribution not depending on the (L_0, Z_0) *initial state.*

Proof: For $N > K$ the $\{\mathbf{Y_n},\ n \geq 0\}$ Markov chain is homogeneous, irreducible and aperiodic thus it has stationary limiting distribution as we saw previously.

An $(l, s) \in S^\infty$ state is **recurrent** if the probability of returning to (l, s) in finite steps is equal to 1, i. e.

$$\sum_{n=1}^{\infty} f^n_{(l,s)(l,s)} = 1,$$

where

$$f^n_{(l,s)(l,s)} \stackrel{def}{=} Pr\{\mathbf{Y_n} = (l, s), \mathbf{Y_i} \neq (l, s), i = 1, \ldots, n-1 | \mathbf{Y_0} = (l, s)\}$$

(for irreducible Markov chains every state is recurrent if one is).

The chain is **positive recurrent** if for its β stationary distribution: $\beta_{(l,s)} > 0,\ \forall (l, s) \in S^\infty$ (for a homogeneous, aperiodic, irreducible Markov chain every state is positive recurrent if one is).

If we find a condition for positive recurrence the chain will have a unique ergodic limiting distribution as well ([KT75], [Kli86]). We will proceed according to the §3. of [Bor76]:

Let us introduce the $\{\xi_n \stackrel{def}{=} Z_n - K,\ n \geq 0\}$ Markov chain that has the same properties as $\{Z_n,\ n \geq 0\}$, and

$$L_n = (L_{n-1} + \xi_{n-1})^+,\ n \geq 1. \tag{2.4}$$

We are investigating the joint probability function of the (L_n, ξ_n) two-dimensional vector-valued random variables for $n \geq 0$. From this we can trivially derive its first marginal distribution, the queue length.

LEMMA 4

$$L_n = U_{n-1} - \min\{-L_0, U_0, \ldots, U_{n-1}\},\ n \geq 1, \tag{2.5}$$

where $U_n \stackrel{def}{=} \xi_0 + \cdots + \xi_n,\ n \geq 0.$

Proof: We will use induction. The equality (2.5) holds for $n = 1$. Suppose that (2.5) holds for $n > 1$:

Two cases are distinguished now:
(a) If

$$U_n = \min\{-L_0, U_0, \ldots, U_n\},$$

then it means

$$U_{n-1} + \xi_n \leq \min\{-L_0, U_0, \ldots, U_{n-1}\},$$

so using the inductional condition

$$L_n + \xi_n \leq 0$$

holds. Now using (2.4)

$$L_{n+1} = (L_n + \xi_n)^+ = 0 = U_n - \min\{-L_0, U_0, \ldots, U_n\}.$$

(b) If

$$U_n = U_{n-1} + \xi_n > \min\{-L_0, U_0, \ldots, U_n\},$$

then using (2.5)

$$L_n + \xi_n > 0.$$

By (2.4) we get then

$$L_{n+1} = L_n + \xi_n = U_n - \min\{-L_0, U_0, \ldots, U_{n-1}\} =$$

$$= U_n - \min\{-L_0, U_0, \ldots, U_n\}. \quad \square$$

The sequence $\{\xi_n^*,\ n \geq 0\}$ is **strictly stationary** if the joint distribution of the random variables $(\xi_{k+n_1}^*, \xi_{k+n_2}^*, \ldots, \xi_{k+n_j}^*)$ does not depend on k for arbitrary j and $n_1, \ldots, n_j$ $(k + n_i \geq 0,\ i = 1, \ldots, j)$.

Let the $\{\xi_n^*,\ n \geq 0\}$ sequence be a Markov chain with the same transition probabilities as the $\{\xi_n,\ n \geq 0\}$ Markov chain, and let the ξ_0^* be equally distributed to the unique limiting ergodic distribution of the $\{\xi_n,\ n \geq 0\}$ Markov chain, hence the sequence $\{\xi_n^*,\ n \geq 0\}$ is strictly stationary.

(The finite dimensional distributions of the sequence

$$\{\xi_{l,k} \stackrel{def}{=} \xi_{l+k},\ k \geq 0\}$$

converge as $l \longrightarrow \infty$ to the corresponding finite-dimensional distributions of the strictly stationary sequence $\{\xi_n^*,\ n \geq 0\}$, where this kind of convergence will be referred to as **weak convergence**.)

We can extend the $\{\xi_n^*,\ n \geq 0\}$ sequence to be strictly stationary on the whole $\{-\infty < n < \infty\}$ time-axis (we can proceed this way without loss of generality since by *Kolmogorov's theorem on the extension of the compatible distributions* we can construct the required distribution of $\{\xi_n^*,\ -\infty < n < \infty\}$ using the compatible finite-dimensional distributions of the original sequence: [Rén68]). The $\{L_n^*,\ n \geq 0\}$ sequence will be given by the initial condition L_0^* and the strictly stationary sequence $\{\xi_n^*,\ -\infty < n < \infty\}$ with the recursion (2.4) – (2.5), the *LEMMA 4*, holds for the extended versions of the $(.)^*$ sequences as well.

Let us define

$$V_n^* \stackrel{def}{=} \xi_{-1}^* + \cdots + \xi_{-n}^*,\ n \geq 1, \quad V_0^* \stackrel{def}{=} 0,$$

$$\overline{V}_n^* \stackrel{def}{=} \max\{V_0^*, \ldots, V_n^*\},$$

$$V^* \stackrel{def}{=} \overline{V}_\infty^* = \sup_{k \geq 0} V_k^*.$$

The $\{\xi_n^*,\ -\infty < n < \infty\}$ strictly stationary sequence is **metrically transitive** if for the W one-to-one measure-preserving shift transformation of sets $A \in \mathcal{M}$ such that

$$W\{\omega \in \Omega : \xi_n^*(\omega) \in B_0, \ldots, \xi_{n+k}^*(\omega) \in B_k\} =$$
$$= \{\omega \in \Omega : \xi_{n+1}^*(\omega) \in B_0, \ldots, \xi_{n+k+1}^*(\omega) \in B_k\},$$

for an arbitrary collection of Borel sets $B_0, \ldots, B_k \in \mathcal{B}(\mathbb{R}^1)$, for $k \geq 0$, where $\mathcal{M} \subset \mathcal{A}$ is the σ-algebra generated by $\{\xi_n^*,\ -\infty < n < \infty\}$ (we are on the $(\Omega, \mathcal{A}, Pr)$ probability space), only the sets of measure 0 and 1 are **invariant** under W, where $A \in \mathcal{M}$ is invariant under W if WA and A coincide up to a set of measure 0.

(The corresponding transformation on $\mathcal{M}$-measurable random variables will be denoted by the letter Q so that

$$\xi_{n+1}^* = Q\xi_n^*.$$

For η $\mathcal{M}$-measurable random variable the sequence $\{Q^n\eta : -\infty < n < \infty\}$ will obviously be strictly stationary.)

LEMMA 5 *For $\{\xi_n^*,\ -\infty < n < \infty\}$ strictly stationary and metrically transitive sequence*

$$\lim_{n \to \infty} Pr\{L_n^* > x\} = Pr\{V^* > x\}, \quad \textit{for all } x \geq 0$$

exists and is proper,

$$Pr\{V^* < \infty\} = 1,\ \textit{if } E\{\xi_0^*\} < 0.$$

Proof: By the finiteness of ξ^*_{-1} we have

$$Pr\{V^* = \infty\} = Pr\{\sup_{n\geq 0}\{V_n^* - \xi^*_{-1}\} = \infty\} = Pr\{W^{-1}\{V^* = \infty\}\},$$

hence the event $\{V^* = \infty\}$ is invariant under W, so $Pr\{V^* = \infty\}$ is 0 or 1 because of the metrical transitivity. By *Doob's theorem* ([Doo53]) for $\{\xi_n^*,\ -\infty < n < \infty\}$ strictly stationary and metrically transitive sequence

$$\frac{1}{k}\sum_{i=n}^{n+k}\xi_i^* \overset{k\to\infty}{\longrightarrow} E\{\xi_0^*\},$$

with probability 1.

$E\{\xi_0^*\} < 0$ means that the sequence $V_1^*, V_2^*, \ldots$ contains with probability 1 only a finite number of positive terms, so $V^* = \sup_{k\geq 0} V_k^*$ is finite with probability 1, so we have proved the second assertion. $E\{\xi_0^*\} < 0$ also means that $U_n^* \overset{n\to\infty}{\longrightarrow} -\infty$ with probability 1 (we will use this to prove the first assertion). It is easy to see that $U_n \overset{n\to\infty}{\longrightarrow} -\infty$ with probability 1 likewise in the $(.)^*$ case.

Let denote

$$v_{x,n} \overset{def}{=} Pr\{\overline{V}^*_{n-1} \leq x, V_n^* > x\},$$

hence

$$Pr\{V^* > x\} = \sum_{n=1}^{\infty} v_{x,n},$$

and the series on the right converges. Moreover, since the (2.5) recursion stands for L^*_{n+1}, we can write

$$Pr\{L^*_{n+1} > x\} = Pr\{\max\{U_n^* - U_0^*, \ldots, U_n^* - U_n^*\} > x\} +$$
$$+ Pr\{\max\{U_n^* - U_0^*, \ldots, U_n^* - U_n^*\} \leq x, U_n^* + L_0^* > x\} =$$
$$= Pr\{\overline{V}^*_n > x\} + Pr\{\max\{U_n^* - U_0^*, \ldots, U_n^* - U_n^*\} \leq x, U_n^* + L_0^* > x\}.$$

Then

$$Pr\{\overline{V}^*_n > x\} = Pr\{V^* > x\} - \sum_{k>n} v_{x,k}$$

and

$$Pr\{\max\{U_n^* - U_0^*, \ldots, U_n^* - U_n^*\} \leq x, U_n^* + L_0^* > x\} \leq$$
$$\leq Pr\{x - L_0^* < U_n^* \leq x\},$$

what means that this converges to 0 as $n \to \infty$ for all fixed L_0^* initial state. Therefore the whole assertion is proved. □

LEMMA 6 *If* $E\{\xi_0^*\} < 0$ *then*

$$\lim_{n\to\infty} Pr\{L_n < x\} = Pr\{\sup_{n\geq 0} V_n^* < x\} = Pr\{V^* < x\}.$$

Proof: For arbitrary m and n, $m \leq n$ let denote

$$F_{n,m} \stackrel{def}{=} \max\{U_n + L_0, U_n - U_0, \ldots, U_n - U_{m-1}\},$$

$$G_{n,m} \stackrel{def}{=} \max\{U_n - U_m, \ldots, U_n - U_n\}.$$

With this notation used for the $(.)^*$ sequence as well

$$|Pr\{L_{n+1} < x\} - Pr\{V^* < x\}| \stackrel{(2.5)}{=}$$

$$= |Pr\{\max\{U_n + L_0, U_n - U_0, \ldots, U_n - U_n\} < x\} - Pr\{V^* < x\}| =$$
$$= |Pr\{F_{n,m} < x, G_{n,m} < x\} - Pr\{G_{n,m} < x\} +$$
$$+ Pr\{G_{n,m} < x\} - Pr\{G_{n,m}^* < x\} +$$
$$+ Pr\{G_{n,m}^* < x\} - Pr\{V^* < x\}| \leq$$
$$\leq |Pr\{F_{n,m} < x, G_{n,m} < x\} - Pr\{G_{n,m} < x\}| +$$
$$+ |Pr\{G_{n,m} < x\} - Pr\{G_{n,m}^* < x\}| +$$
$$+ |Pr\{G_{n,m}^* < x\} - Pr\{V^* < x\}| \stackrel{def}{=}$$
$$\stackrel{def}{=} A_1 + A_2 + A_3.$$

Let us fix an arbitrary $x \geq 0$.

For $n \geq m$

$$A_1 = Pr\{F_{n,m} \geq x, G_{n,m} < x\} \leq Pr\{F_{n,m} \geq x\} =$$
$$= Pr\{\max\{U_n + L_0, U_n - U_0, \ldots, U_n - U_n\} \geq x\} \leq$$
$$\leq Pr\{U_n + L_0 \geq x\} + Pr\{\max\{U_n - U_0, \ldots, U_n - U_n\} \geq x\} =$$
$$= Pr\{U_n + L_0 \geq x\} +$$
$$+ Pr\{\max\{\xi_1 + \ldots + \xi_{m-1} + \xi_m + \ldots + \xi_n, \ldots, \xi_m + \ldots + \xi_n\} \geq x\} \leq$$
$$\leq Pr\{U_n + L_0 \geq x\} + Pr\{(m-1)(N-K) + \xi_m + \ldots + \xi_n \geq x\},$$

therefore for every $\varepsilon > 0$, $m > 0$ and $L_0 > 0$ there exists an $n_1 \stackrel{def}{=} n_1(\varepsilon, m) > 0$ such that

$$A_1 < \frac{\varepsilon}{3}, \quad \text{for } n \geq n_1,$$

because for every fixed $m > 0$: $\xi_m + \ldots + \xi_n \stackrel{n\to\infty}{\longrightarrow} -\infty$ with probability 1 ($E\{\xi_0^*\} < 0$).

For every $\varepsilon > 0$ and $m > 0$ there exists an $n_2 \stackrel{def}{=} n_2(\varepsilon, m) > 0$ such that

$$A_2 < \frac{\varepsilon}{3}, \quad n \geq n_2,$$

because the $\{\xi_{l,k} = \xi_{l+k}, k \geq 0\}$ sequence weakly converges as $l \longrightarrow \infty$ to the $\{\xi_n^*, n \geq 0\}$ sequence (by definition).

For every $\varepsilon > 0$ and $m > 0$ there exists an $n_3 \stackrel{def}{=} n_3(\varepsilon, m) > 0$ such that

$$A_3 < \frac{\varepsilon}{3}, \quad \text{for } n \geq n_3,$$

as we saw in *LEMMA 5.*

Collecting these we get:

For all $\varepsilon > 0$, $m > 0$ and $x \geq 0$ there exists an $0 < n_0 \stackrel{def}{=} \max\{n_1, n_2, n_3\}$ such that

$$|Pr\{L_{n+1} < x\} - Pr\{V^* < x\}| < \varepsilon, \quad \text{for } n \geq n_0,$$

and that is what we wanted to show. □

The $0 > E\{\xi_0^*\} = \lim_{n\to\infty} E\{Z_n - K\} = E\{Z\} - K \stackrel{(2.3)}{=} N\frac{1}{1+q} - K$ condition thus means that the stationary distribution of the original Markov chain is proper but this can stand only if the Markov chain is positive recurrent. This completes the proof of the theorem. □ □

REMARK: Notice that the $N\frac{1}{1+q} < K$ condition is in a certain sense minimal, because the queue length will grow very fast if the mean of the arrivals is greater than the service capacity.

The ergodic distribution of the queue length

Assume that $N\frac{1}{1+q} < K < N$ stands for the $\{\mathbf{Y_n} = (L_n, Z_n),\ n \geq 0\}$ Markov chain so there is a unique β ergodic distribution for it. Denote

$$\rho_l \stackrel{def}{=} \lim_{n\to\infty} Pr\{L_n = l\} = Pr\{L = l\}, \quad l \geq 0,$$

the first marginal distribution of β and

$$\pi_s = \lim_{n \to \infty} Pr\{Z_n = s\} = Pr\{Z = s\}, \quad 0 \le s \le N,$$

the second marginal distribution of β that is already determined in *PROPOSITION 2.*

$$\Big(\rho_l = \sum_{s=0}^{N} \beta_{(l,s)}, \ \pi_s = \sum_{l=0}^{\infty} \beta_{(l,s)}\Big).$$

For all $l \ge 0, \ 0 \le s \le N \ (n \ge 1)$

$$Pr\{L_n = l, Z_n = s\} =$$

$$\sum_{k=0}^{\infty} \sum_{m=0}^{N} Pr\{L_n = l, Z_n = s | L_{n-1} = k, Z_{n-1} = m\} Pr\{L_{n-1} = k, Z_{n-1} = m\} =$$

$$= \sum_{k=0}^{\infty} \sum_{m=0}^{N} \chi_{\{(k+m-K)^+ = l\}} p_{m,s} Pr\{L_{n-1} = k, Z_{n-1} = m\},$$

where $\chi_{\{expression\}} \stackrel{def}{=} \begin{cases} 1 & \text{, if the } expression \text{ is true} \\ 0 & \text{, otherwise} \end{cases}$.

For $n \longrightarrow \infty$ we get the following recursion for β:

$$\beta_{(l,s)} = \sum_{k=0}^{\infty} \sum_{m=0}^{N} \chi_{\{(k+m-K)^+ = l\}} p_{m,s} \beta_{(k,m)} \ , \tag{2.6}$$

$$\sum_{l=0}^{\infty} \sum_{s=0}^{N} \beta_{(l,s)} = 1, \tag{2.7}$$

where $\beta_{(l,s)} = \lim_{n \to \infty} Pr\{L_n = l, Z_n = s\}, \ p_{m,s} = Pr\{Z_n = s | Z_{n-1} = m\}$.

REMARKS:

1. Notice that the recursive equations (2.6) and (2.7) can be generalized for arbitrary $\{Z_n, \ n \ge 0\}$ Markov chains.

2. Presuming some stability conditions we could deal with the recursion numerically by supposing $\beta_{(l,s)} = 0$ for $l > l_0$ (l_0 "great enough") because then we would have finite number of unknown probabilities.

3. Another possible way of examining the recursion is by means of simulation using the regenerative quality of the queue length ($L = 0$).

SPECIAL (INDEPENDENT) CASE:

From the recursions for the general Markov chain $\{Z_n,\ n \geq 0\}$ and for that special case when it is an i. i. d. (independent, identically distributed with the Z random variable) sequence of random variables we can derive the probability generating function of the queue length ([BLT96] with the ergodicity condition $\mu_Z \stackrel{def}{=} E\{Z\} < K$):

For the independent case we can place

$$p_{m,s} = \pi_s,$$

$$\beta_{(l,s)} = \rho_l \pi_s,$$

into the generalized recursions of (2.6) and (2.7), where $\pi_s = Pr\{Z = s\}$, so

$$\rho_0 = \sum_{k=0}^{K} \rho_k \sum_{s=0}^{K-k} \pi_s,$$

$$\rho_l = \sum_{m=0}^{K+l} \rho_{K+l-m} \pi_m, \quad l \geq 1.$$

Let $Q(z) \stackrel{def}{=} \sum_{l=0}^{\infty} \rho_l z^l,\ |z| \leq 1\ (z \in \mathbb{C})$ be the **p**robability **g**enerating **f**unction (**p. g. f.**) of the random variable L, $P(z) \stackrel{def}{=} \sum_{s=0}^{\infty} \pi_s z^s,\ |z| \leq 1$ the p. g. f. of Z. From the derived recursions:

$$z^K Q(z) = z^K \rho_0 + \sum_{l=1}^{\infty} \sum_{s=0}^{K+l} \rho_{K+l-s} z^{K+l-s} \pi_s z^s =$$

$$= \rho_0 z^K + Q(z)P(z) - \rho_0(\pi_0 + \cdots + \pi_K z^K) - \rho_1 z(\pi_0 + \cdots + \pi_{K-1} z^{K-1}) - \cdots$$
$$\cdots - \rho_K z^K(\pi_0),$$

hence we can get for the unknown $Q(z),\ |z| \leq 1$ p. g. f.:

$$Q(z) = \frac{\rho_0 z^K - \sum_{l=0}^{K} \rho_l z^l \Pi_{K-l}(z)}{z^K - P(z)},$$

where $\Pi_j(z) \stackrel{def}{=} \sum_{s=0}^{j} \pi_s z^s,\ 0 \leq j \leq K$.

For calculating the $\rho_l,\ 0 \le l \le K$ unknown probabilities we have two equalities already:

$$\rho_0 = \sum_{l=0}^{K} \rho_l \Pi_{K-l}(1), \tag{2.8}$$

$$1 = Q(1) = \frac{K\rho_0 - \rho_0 \Pi'_K(1) - \sum_{l=1}^{K} [\rho_l \Pi'_{K-l}(1) + l\rho_l \Pi_{K-l}(1)]}{K - \mu_Z}, \tag{2.9}$$

where the equation (2.8) comes from the recursion for ρ_0 and the equation (2.9) from using a *L'Hospital rule* for calculating $Q(1)$.

$Q(z)$ is bounded in $|z| \le 1$ thus all the roots in $|z| \le 1$ of the denominator of $Q(z)$ must be the roots of the nominator as well (for $z = 1$, too). Consequently the roots of $z^K - P(z)$ can be used for determining further equations on the unknown probabilities (for example with *Rouché's theorem* we can see that there are exactly K roots of the denominator on $|z| \le 1$ if the ergodicity condition holds, [BLT96]). Therefore we are going to have just enough equations to determine the $K + 1$ unknown probabilities (the nominator is a polinomial of degree K consequently the K roots and the multiplicative constant can be derived as well).

REMARK: From the probability generating function we can derive any moment of the stationary queue length by calculating the derivatives of the p. g. f. in $z = 1$ using the *L'Hospital rule* (for example $E\{L\} = \lim_{z\to 1} Q'(z)$).

2.3.2 The $m < \infty$ case

REMARK: It would be also interesting to examine the system with the acceptance rule: if there is not enough storage capacity available in the queue for the whole group arriving at the same time they are all rejected ([TGA88]).

Consider $m < \infty$ storage capacity available in the system, the $\{Z_n,\ n \ge 0\}$ arrival process and the $\{L_n^{(m)},\ n \ge 0\}$ queue length defined previously with the initial states Z_0 and $L_0^{(m)}$. The $\{\mathbf{Y_n^{(m)}} \stackrel{def}{=} (L_n^{(m)}, Z_n),\ n \ge 0\}$ sequence of vector-valued random variables forms a homogeneous discrete-time Markov chain by definition again with the $\tau = \{0, 1, \ldots\}$ time space and a finite state space S according to the following recursion:

$$L_n^{(m)} = (\min\{L_{n-1}^{(m)} + Z_{n-1}, m\} - K)^+, \quad n \ge 1,$$

where

$$S \stackrel{def}{=} \{(l,s):\ l = 0,\dots,(m-K)^+,\ s = 0,1,\dots,N\}.$$

If $m \le K$, then the Markov chain is the same as the Markov chain of the arrival process, so we can assume that $m > K$.

REMARK: If we assume the queue length process continuous from the right the recursion would change to $L_n^{(m)} = \min\{(L_{n-1}^{(m)} - K)^+ + Z_n,\ m\},\quad n \ge 1$ and the S state space would change to $S = \{(l,s):\ 0 \le l \le m,\ 0 \le s \le N\}$.

The elements of the $B = \{b_{(l_i,s_i)(l_j,s_j)}\}$,$(l_i,s_i),(l_j,s_j) \in S$ one-step transition probability matrix of this Markov chain are:

$$b_{(l_i,s_i)(l_j,s_j)} = \begin{cases} p_{s_i,s_j} & ,\text{if}\quad l_j = (\min\{l_i + s_i, m\} - K)^+ \\ 0 & ,otherwise \end{cases}$$

$$0 \le l_i, l_j \le m - K, \quad 0 \le s_i, s_j \le N.$$

Some properties of this Markov chain:

- $\{\mathbf{Y_n^{(m)}} = (L_n^{(m)}, Z_n),\ n \ge 0\}$ Markov chain is **irreducible** if $N > K$:

$0 \le L_{n-1}^{(m)} + Z_{n-1} \le m - K + N$, for all $n \ge 1$,
so in case of
$N \le K : L_{n+1}^{(m)} \le L_n^{(m)}$, for all $n \ge 1$ (for example we can not get from $(0,0)$ state to $(1,0)$ state in n steps with positive probability for any $n \ge 1$).

From the one-step transition probability matrix it can be easily seen that for $N > K$ the Markov chain is irreducible.

Specially, assuming $N > K$, the $(0,0)$ state can be reached from any state, $p_{0,0} > 0$ on the other hand, that also means that:

- The $\{\mathbf{Y_n^{(m)}},\ n \ge 0\}$ Markov chain is **aperiodic** for $N > K$ as well:

For all $(l_i,s_i),(l_j,s_j) \in S$ the greatest common divisor of the elements of the $\{n \ge 1:\ b^n_{(l_i,s_i)(l_j,s_j)} > 0\}$ set equals unity because the chain is irreducible and $p_{0,0} > 0$.

Gathering it all we can formulate the following theorem:

THEOREM 7

1. *If $m \leq K$ then the $\{\mathbf{Y_n^{(m)}} = (L_n^{(m)}, Z_n),\ n \geq 0\}$ Markov chain is the same as the $\{Z_n,\ n \geq 0\}$ Markov chain thus has the same unique ergodic distribution not depending on the $(L_0^{(m)}, Z_0)$ initial state.*

2. *If $m > K$ and $N > K$ then the $\{\mathbf{Y_n^{(m)}},\ n \geq 0\}$ Markov chain is homogeneous, irreducible, aperiodic and so has a unique*

$$\beta^{(m)} \stackrel{def}{=} (\beta^{(m)}_{(0,0)}, \ldots, \beta^{(m)}_{(0,N)}, \ldots, \beta^{(m)}_{(m-K,0)}, \ldots, \beta^{(m)}_{(m-K,N)})^T$$

   ergodic distribution that does not depend on the $(L_0^{(m)}, Z_0)$ initial state.
   □

([KT75] or [Kli86], Chapter 3., §3., Theorem 0. and 3.)

The ergodic distribution

Assume $m > K$ and $N > K$ for the $\{\mathbf{Y_n^{(m)}} = (L_n^{(m)},\ Z_n),\ n \geq 0\}$ Markov chain with the B one-step transition probability matrix. Therefore there exists a unique $\beta^{(m)}$ ergodic distribution that is the unique solution of the following set of equations:

$$(\beta^{(m)}_{(0,0)}, \ldots, \beta^{(m)}_{(0,N)}, \ldots, \beta^{(m)}_{(m-K,0)}, \ldots, \beta^{(m)}_{(m-K,N)})B = {\beta^{(m)}}^T, \quad (2.10)$$

$$\sum_{(l,s)\in S} \beta^{(m)}_{(l,s)} = 1. \quad (2.11)$$

Solving equations (2.10) and (2.11) we get the joint ergodic distribution of $(L_n^{(m)}, Z_n)$:

$$\beta^{(m)}_{(l,s)} = \lim_{n\to\infty} Pr\{L_n^{(m)} = l, Z_n = s\} \stackrel{def}{=} Pr\{L^{(m)} = l, Z = s\}, \quad (l,s) \in S,$$

and the first marginal distribution, the queue length probabilities,

$$\rho_l^{(m)} = \lim_{n\to\infty} Pr\{L_n^{(m)} = l\} \stackrel{def}{=} Pr\{L^{(m)} = l\},$$

can be derived from $\beta^{(m)}$ as follows:

$$\rho_l^{(m)} = \sum_{s=0}^{N} \beta^{(m)}_{(l,s)} = \sum_{s=0}^{N} Pr\{L^{(m)} = l, Z = s\},$$

$$0 \leq l \leq m-K.$$

$$\left(\pi_s = Pr\{Z = s\} = \sum_{l=0}^{m-K} \beta_{(l,s)}^{(m)}\right)$$

Some numerical examples are calculated in *MATLAB*:

$$K = 3,\ N = 4\ (m = 4):$$

$$q = \tfrac{1}{4}: \quad \rho^{(4)} = (0.3061,\ 0.6939)^T,$$
$$q = \tfrac{2}{3}: \quad \rho^{(4)} = (0.8233,\ 0.1767)^T;$$

$$K = 3,\ N = 5\ (m = 5):$$

$$q = \tfrac{3}{2}: \quad \rho^{(5)} = (0.8999,\ 0.0873,\ 0.0128)^T,$$
$$q = \tfrac{4}{1}: \quad \rho^{(5)} = (0.9928,\ 0.0068,\ 0.0004)^T.$$

The mean queue length

Having determined the $\rho_l^{(m)},\ 0 \leq l \leq m-K$ probabilities we can get an important performance measure of the queue, the **stationary mean queue length**:

$$E\{L^{(m)}\} = \sum_{l=0}^{m-K} l\rho_l^{(m)},$$

moreover we can get any moment of $L^{(m)}$ because we know the distribution.

Some numerical examples for the mean queue length are depicted in the **Figures 2.6** and **2.7** on **Page 35** and in the **Figures 2.8** and **2.9** on **Page 36**.

The probability of rejection

Another important performance measure of the queueing system in the $m < \infty$ case is the $p^{[R]}$ **probability of overflow**, the probability of customer-rejection in the stationary state, i. e.

$$p^{[R]} \stackrel{def}{=} Pr\{R > 0\} = Pr\{L^{(m)} + Z > m\} = \sum_{(l,s)\in H} \beta_{(l,s)}^{(m)},$$

where R is a random variable counting the number of customers being rejected in the stationary state and

$$H \stackrel{def}{=} \{(l,s) \in S:\ l + s > m\}$$

is a part of the S state space.

Some figures are depicted for the probability of customer-rejection in the **Figures 2.10** and **2.11** on **Page 37** and in the **Figures 2.12** and **2.13** on **Page 38**.

We can also derive the $E\{R\}$ **mean number of customers being rejected** in the system:

$$E\{R\} = \sum_{(l,s)\in H} (l+s-m)\beta_{(l,s)}^{(m)},$$

and then the p_{LOSS} **expected fraction of the total number of customers being rejected** by the system:

$$p_{LOSS} \stackrel{def}{=} \frac{E\{R\}}{E\{Z\}}.$$

REMARK: Notice that these results can be derived for an arbitrary arrival process that forms an ergodic finite-state discrete-time Markov chain.

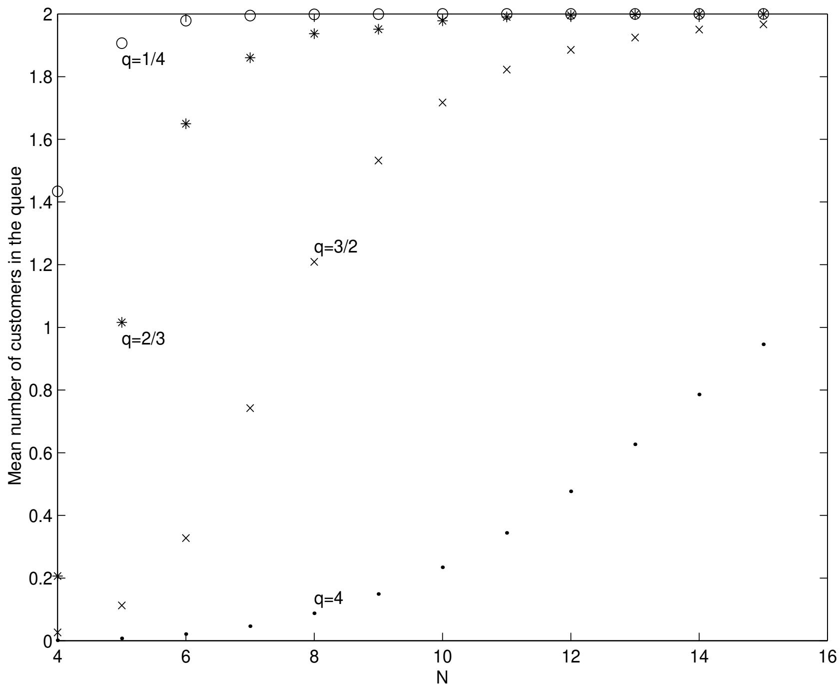

Figure 2.6: $E\{L^{(5)}\}$ for $K = 3$

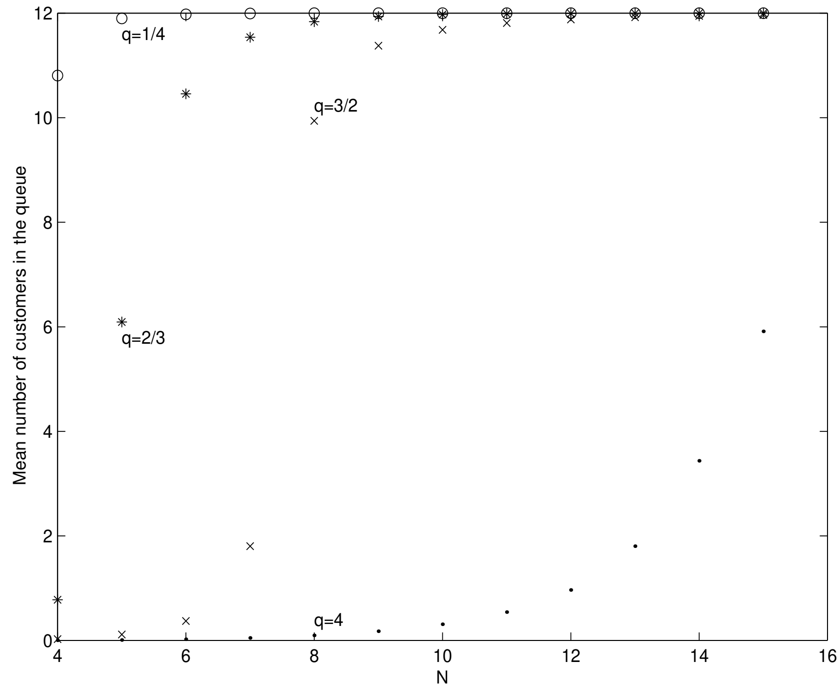

Figure 2.7: $E\{L^{(15)}\}$ for $K = 3$

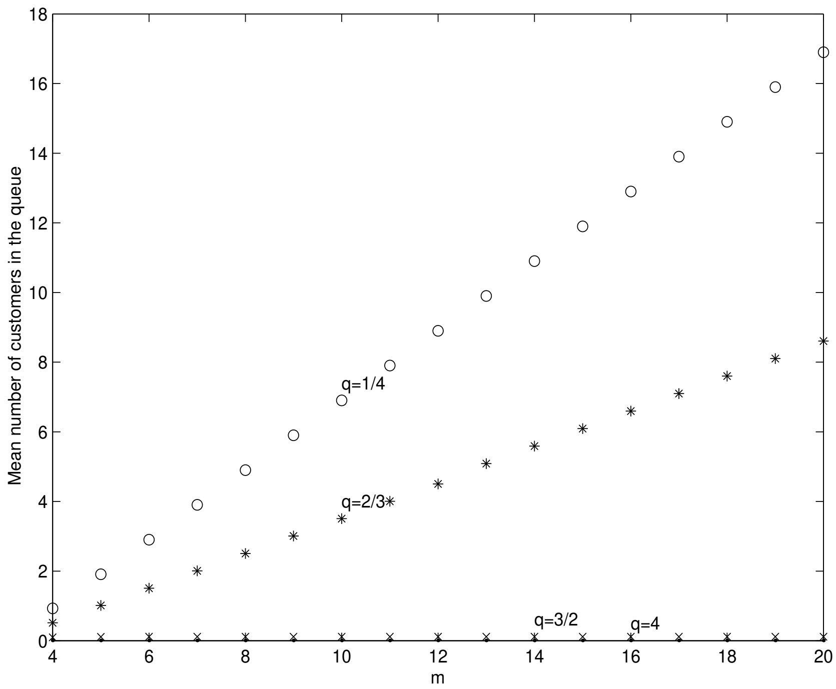

Figure 2.8: $E\{L^{(m)}\}$ for $N = 5$ and $K = 3$

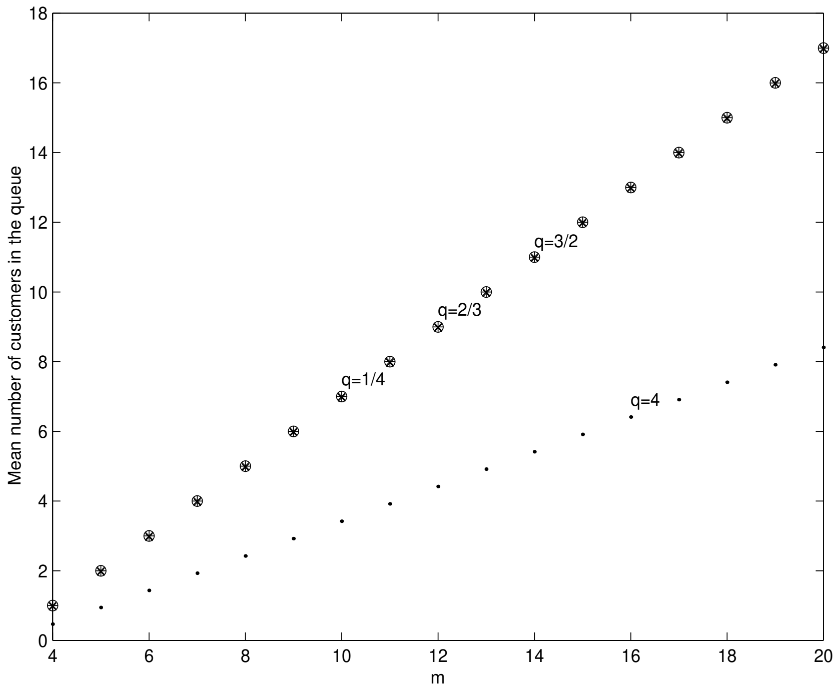

Figure 2.9: $E\{L^{(m)}\}$ for $N = 15$ and $K = 3$

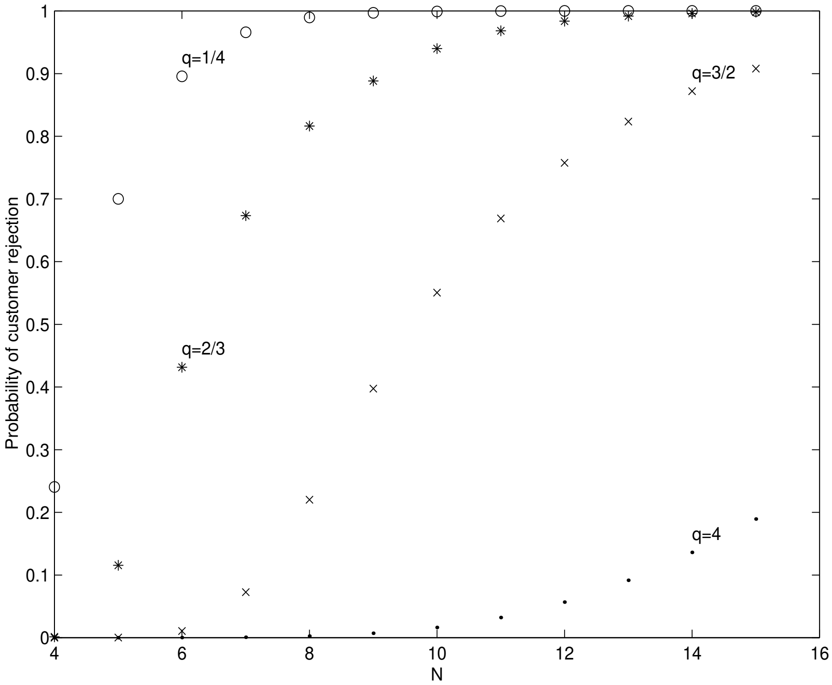

Figure 2.10: $p^{[R]}$ for $m = 5$ and $K = 3$

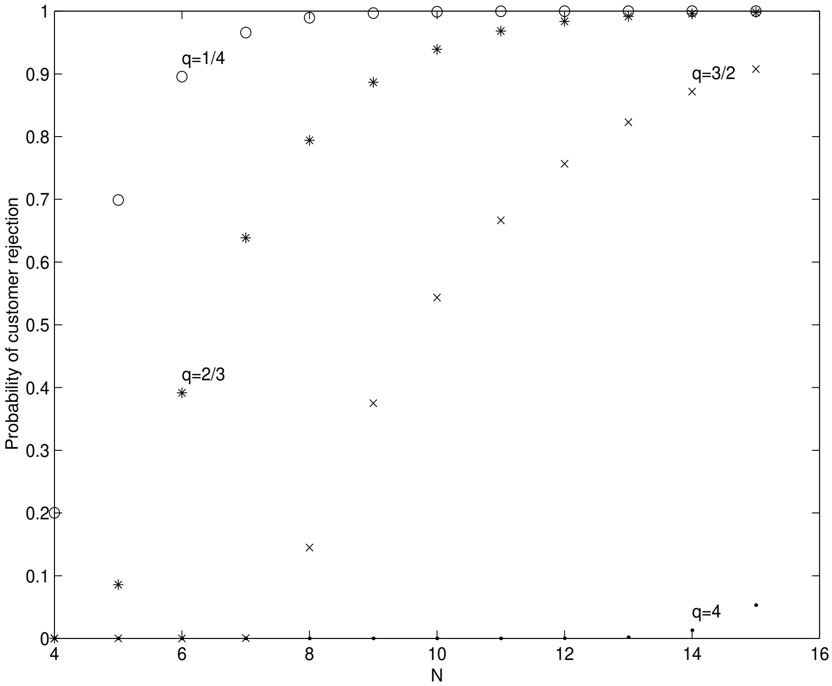

Figure 2.11: $p^{[R]}$ for $m = 15$ and $K = 3$

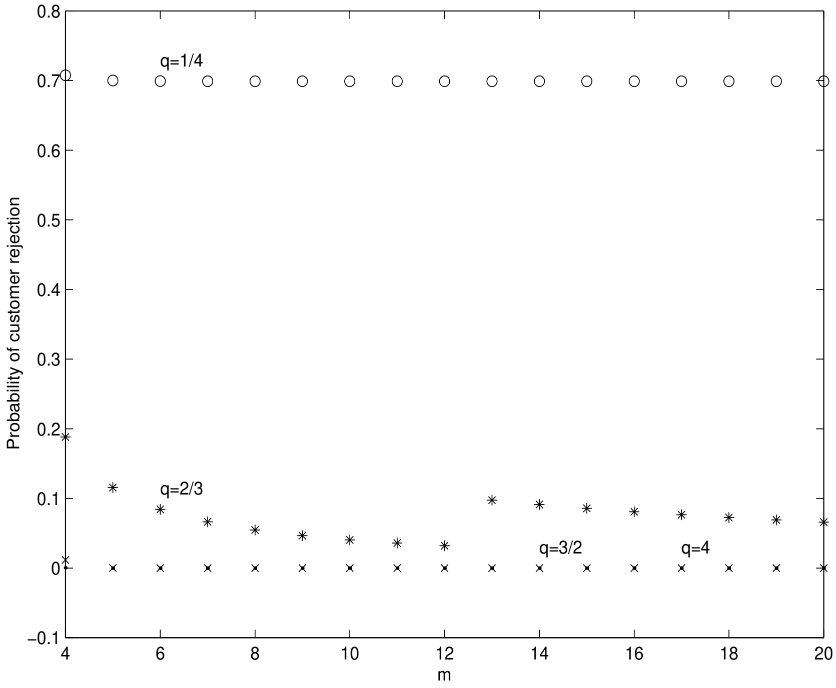

Figure 2.12: $p^{[R]}$ for $N = 5$ and $K = 3$

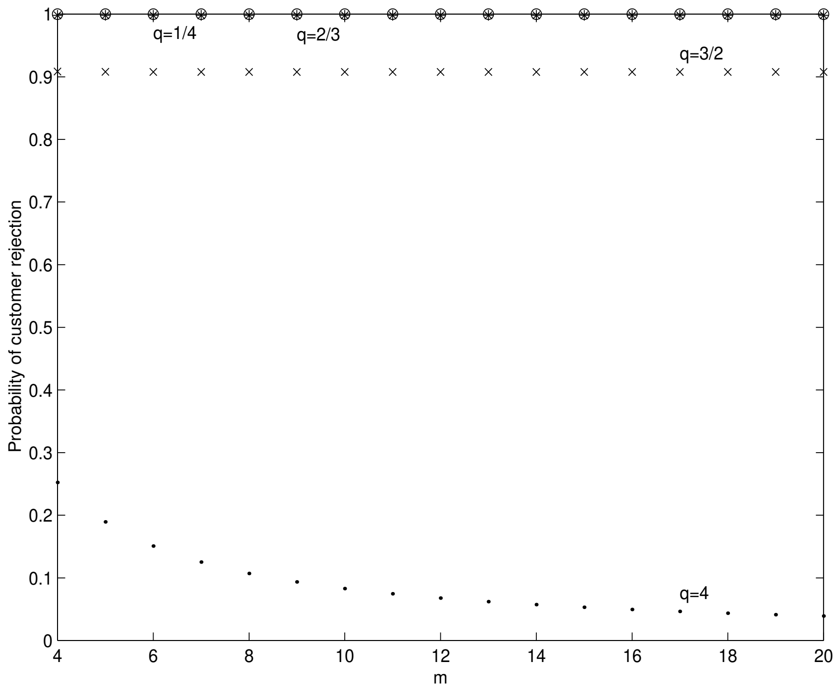

Figure 2.13: $p^{[R]}$ for $N = 15$ and $K = 3$

Further planned works

- Discussion of other performance measures of queueing systems like the **waiting time** of a customer, the current **work backlog**, the **length of a busy** or **an idle period** ([Kle75]) for our system is planned.

- The investigation of $D^{[X]}/D/1$ queues with more general non i. i. d. flows of batch arrivals, too, can be done in the future.

- Finding he connection of the results of discrete-time models to the results of the continuous-time queueing models treating similar problems can be a task as well.

Chapter 3

Appendix

[BLT96]

Model description

A generalized $D^{[X]}/D/1/m$ $(m \leq \infty)$ queueing system is studied in the paper:

The arrival process

• The interarrival time of the customers is constant Δ time units (the arrival process is deterministic), the n-th arrival occurs at time $t_n = t_0 + n\Delta,\ n \geq 1$.

• Denote the number of customers arriving in t_n by $X_n,\ n \geq 1$. The successive sizes of customer groups $X_1, X_2, \ldots$ constitute an i. i. d. sequence of random variables. Let $b_i = Pr\{X_1 = i\},\ i \geq 0$ denote the common **p**robability **m**ass **f**unction (**p. m. f.**) of the random variables of the sequence. Denote the **z-transform**, the probability generating function (p. g. f.) of X_1 by $B(z) \stackrel{def}{=} E\{z^{X_1}\},\ |z| \leq 1$. It is assumed that $\mu_X \stackrel{def}{=} E\{X_1\} < \infty$.

The structure of the service facility

• Customers arrive in a single-server queue (number of service stations = 1) with m storage capacity available, where $m \leq \infty$.

• The service time of a customer is constant δ time units.

• The server can serve up to K customers within the period Δ, where $K \geq 1$ is a fixed integer (hence $0 \leq K\delta \leq \Delta$ must hold true).

The discipline of the service facility

• Provided that there is storage capacity available, the customers of a group are accepted in the queue in random order until either the customers in the group are exhausted or the queue fills to its capacity.

• Customers that are accepted by the queue and belong to different arrival groups are scheduled to be served in FCFS order, while customers within the same group are served in order of acceptance.

Results

Conditions of steadiness ([Pak69]), ergodicity conditions are examined, the probability generating function of the number of customers in the queue and also its first two moments, the Laplace-Stieltjes transform of the waiting time distribution (in $m < \infty$ case the waiting time distribution and the probability of rejection as well) are derived in the paper. A root-finding algorithm is also presented for approximating the case of Poisson distributed group arrivals.

[Bru88]

Model description

A special $D^{[X]}/D/1/\infty$ queue with group arrivals is considered in the paper:

The arrival process

• The interarrival time of the customer groups is a fixed constant (time unit).

• There are $N \geq 1$ number of identical customer generating sources independent of each other. Each source generates customers in the following same way:

There is X and Y which are two independent geometrically distributed random variables with the corresponding probabilities

$a(n) = Pr\{X = n\} = p_0(1 - p_0)^{n-1}$ and

$b(n) = Pr\{Y = n\} = p_1(1 - p_1)^{n-1}$,

for all $n \geq 1$ and for fixed $0 < p_0, p_1 < 1$. We produce two sequence of i. i. d. random variables $X^{(1)}$, $X^{(2)}$, … and $Y^{(1)}$, $Y^{(2)}$, … equally distributed to X and Y respectively and these construct the length of the on-periods and the off-periods of a generating unit. An $n \geq 1$ long on-period means that the generating unit generates exactly one customer per time slot for n consecutive time units and an n-long off-period means that the source does not generate customer for n consecutive time units. Thus the production of customers by the i-th so called on/off source is described by the $X^{(1)}$, $Y^{(1)}$, $X^{(2)}$, $Y^{(2)}$, $\ldots$ sequence of random variables (alternating on/off periods).

The structure of the service facility

• There is a single service station in the system with infinite storage capacity.

• The service time of a customer is a time unit.

• The server can serve one customer per a time unit.

The discipline of the service facility

• Only the number of customers in the system is investigated so there is no need for determing the discipline of the service facility.

Results

The joint steady-state probability generating function (p. g. f.) of the number of customers in the system and the number of active sources is investigated and a functional equation is obtained. An explicit expression is then derived for the p. g. f. of the number of active sources (number of arrivals) and some expressions for the queue length distribution is calculated as well (solving the functional equation in a special case the mean queue length and the second moment of the queue length is got by derivation). It is also shown using some numeric examples that althought the number of arrivals in the steady-state depends solely on

$$\alpha \stackrel{def}{=} \frac{1}{1 + \dfrac{p_0}{p_1}},$$

the behaviour of the queue length is not fully determined by α.

[JPN80]

Model description

Some special $D^{[X]}/D/1/m$ $(m < \infty)$ queue with batch arrivals are considered in the paper:

The arrival process

• The interarrival time of the customer groups is a time unit.

• There are $N \geq 1$ customer generating units working according **Figure 3.1** where the state "n" means:

1. The unit generates exactly $n = 1$ customer per time unit.
2. The unit generates exactly $n > 1$ customers per time unit.

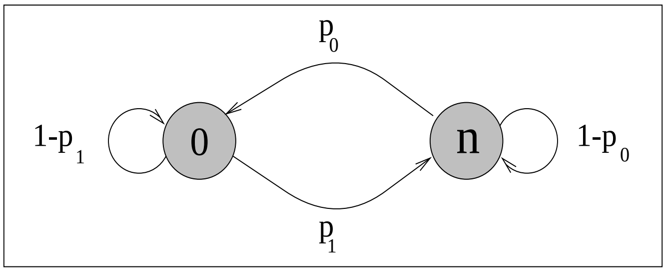

Figure 3.1: A generating unit in [JPN80]

The structure of the service facility

• There are $K \geq 1$ servers in the queue with $m < \infty$ storage capacity available.

• The service time of a customer is a time unit.

• The servers can serve exactly one customer per a time unit.

The discipline of the service facility

• Only the number of customers in the system and the expected fraction of the total number of customers rejected are investigated so there is no need for determing the discipline of the service facility.

Results

In case 1. the ergodic distribution of the number of arrivals is derived. In both the cases a finite system of equation is obtained for the steady-state queue length distribution and also the expected fraction of the customers rejected (probability of customer rejection) is expressed. After the system of equation is numerically solved for a few examples through *Gauss-Seidel iteration*.

[Vit86]

Model description

A lot of complex networks are examined numerically in the paper validated by the theoretical results of investigating the following $D^{[X]}/D/1/\infty$ queue:

The arrival process

• The interarrival time of the customers is a fixed constant (a time unit).

• Customers are generated by $N \geq 1$ independent, different generating units described in **Figure 3.2** ((p_{0^i}, p_{1^i}) are the parameters of the i-th generating unit for $i = 1, \ldots, N$).

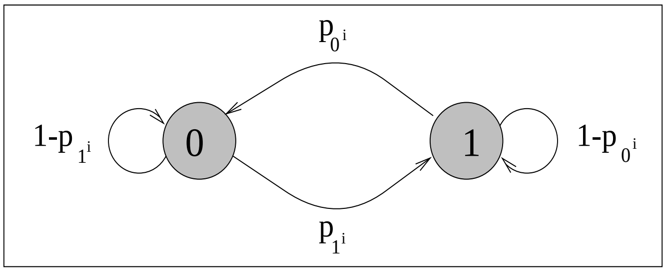

Figure 3.2: The i-th customer generating unit in [Vit86]

The structure of the service facility

• There is one server in the system with ∞ storage capacity.

• The service time is a time unit and the server can serve exactly one customer per a time unit.

The discipline of the service facility

- The customers are accepted in the queue in random order.
- The different groups of customers are served in FCFS order.

Results

Matrix-functional equation is derived for the moment generating function of the distribution of the number of customers in the queue and using the properties of *Kronecker products* of matrices and the properties of their eigenvalues the average number of customers in the queue is obtained. The mean waiting time is thus calculated by *Little's result*.

[WB97]

Model description

A special $D^{[X]}/D/1/\infty$ queue with group arrivals arising from some generalized on/off sources is considered in the paper:

The arrival process

• The interarrival time of the customer groups is a fixed constant - the time unit called **time slot**.

• There are $N_i,\ 1 \leq i \leq I$ number of customer generating sources of type i. Each source generates customers in the following way:
For all $1 \leq i \leq I$ there is X_i and Y_i which are two independent generally distributed positive integer valued random variables with the corresponding probability generating function for $|z| \leq 1$

$A_i(z) = \sum_{n=1}^{\infty} a_i(n) z^n$ and

$B_i(z) = \sum_{n=1}^{\infty} b_i(n) z^n$, where $a_i(n) \stackrel{def}{=} Pr\{X_i = n\}$ and $b_i(n) \stackrel{def}{=} Pr\{Y_i = n\}$, for all $n \geq 1$. We produce two sequence of i. i. d. random variables $X_i^{(1)},\ X_i^{(2)},\ \ldots$ and $Y_i^{(1)},\ Y_i^{(2)},\ \ldots$ equally distributed to X_i and Y_i respectively and these construct the length of the on-periods and the off-periods of the i-th generating unit. An $n \geq 1$ long on-period means that the generating unit generates exactly one customer per time slot for n consecutive time slots and an n-long off-period means that the source does not generate customer for n consecutive time slots. Thus the production of customers by the i-th so called on/off source is described by the $X_i^{(1)},\ Y_i^{(1)},\ X_i^{(2)},\ Y_i^{(2)},\ \ldots$ sequence of random variables (alternating on/off periods).

The structure of the service facility

• There is a single service station in the system with infinite storage capacity.

• The service time of a customer is a time slot.

• The server can serve one customer per a time slot.

The discipline of the service facility

• Only the number of customers in the system is investigated so there is no need for determing the discipline of the service facility.

Results

A Markovian state description is given for the system and a functional equation for the joint steady-state probability generating function of some infinite number of random variables (containing the number of customers in the system) is determined. It is futhermore supposed that in many cases the tail distribution can be approximated by a geometric distribution and its parameters are approximated using the obtained theoretical result and also a lot of numerical examples are shown for the approximation.

[TGA88]

Model description

Analysis of $G^{[X]}/D/1/m$ $(m < \infty)$ queueing system with group arrivals is investigated in the paper.

The arrival process

• A_n denotes the random variable which describes the time interval between the n-th and $(n+1)$-th group arrival (the n-th interarrival time for $n \geq 1$).

• X_n random variable describes the size of the n-th group of customers $(n \geq 1)$.

The structure of the service facility

• Customers arrive in a single server queue with $m < \infty$ storage capacity available.

• The service time of a customer is exactly one time unit.

• The server can serve one customer per a time unit.

The discipline of the service facility

• The customers of a group are accepted in the queue in random order if there is storage capacity available until either the customers in the group are all accepted or the queue is filled to its capacity (the remaining customers are rejected - **b**locking **p**olicy **1**, **BP1**).

The customers of the same group are accepted in the queue in random order if there is enough room for all of them else they are all rejected - **b**locking **p**olicy **2** (**BP2**).

(Both the blocking policies are studied.)

• The different customer groups are served in FCFS order while the customers of the group are served in order of acceptance.

Results

Assuming i. i. d. interarrival times and time-independent group sizes distribution an iterative algorithm is delivered to determine the equilibrium probabilities of the number of customers in the system and the probabilities of rejection are expressed as well for both the blocking policies BP1 and BP2. Numerical examples are calculated and the application of the model for some practical problems are presented, too.

[RV91]

Model description

$G^{[X]}/D/1/\infty$ queues for which the customer arrival process produces a **"stable queue"** are studied in the paper. The "stable queue" is defined as follows using the assumption of the customer service time of the server being one time unit:

Let $N(u,t)$ be the number of customer arrivals in the time interval (u,t). The arrival process is said to produce a "stable queue" if for all t time there exists an $u \geq 1$ integer such that $N(t-s,t) < s$, for all $s > u$ (for a sufficiently long interval the number of arrivals is less than the maximum number of customers which can be served). Note that the service facility is also used for determining the stability.

Furthermore two arrival processes producing "stable queues" are investigated in the paper:

"$N * D_0/D/1$":
The arrival process is generated by N sources, each transmitting one customer per every D_0 time units. The N customer arrival epochs in a time interval $[t-D_0, t)$ are independently and uniformly distributed in this interval.

"$N_1 * D_1 + \ldots + N_r * D_r/D/1$":
$r > 1$ types of sources are considered and there are N_i sources of type i generating a customer every D_i time units likewise the previous process for $i = 1, \ldots, r$.

Results

For the "$N * D_0/D/1$" case the exact distribution of the number of customers present in the system is specified, for the "$N_1 * D_1 + \ldots + N_r * D_r/D/1$" case lower and upper bounds are given for the same distribution function and a lot of numerical results are examined as well.

[WB94]

Model description

A $D^{[X]}/D/1/m$ ($m \leq \infty$) queueing system is studied in the paper:

The arrival process

- The interarrival time of the customers is a fixed constant (a time unit).

- Customers are generated by a single unit described in **Figure 3.3** on the next page ($N = 1$).

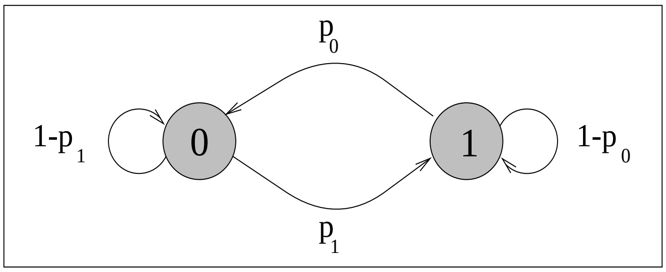

Figure 3.3: The generating unit in [WB94]

The structure of the service facility

• Number of service stations equals 1 that has m storage capacity available including the customer under service ($m \leq \infty$).

• The service time is constant $\delta > 1$ time units for each customer.

• The server can serve exactly one customer per a time unit (every customer need δ time units to be served).

The discipline of the service facility

• If there is storage capacity available, the customers are accepted in the queue, else they are rejected.

• Customers are served in FCFS order.

Results

Closed-form expressions of the probability generating function (p. g. f.) of the number of customers in the system, the probability of rejection, the p. g. f. of the work backlog, the p. g. f. of the waiting time of an arbitrary part of the customers, the joint p. g. f. of two consecutive interdeparture times are derived in the paper and also some numerical examples are presented using the obtained results.

Bibliography

[BLT96] C. Bisdikian, J. S. Lew, A. N. Tantawi. The generalized $D^{[X]}/D/1$ queue: A flexible computer communications model. *Telecommunication Systems*, Vol. 6. No. 2. pp. 127-146, 1996.

[Bor76] A. A. Borovkov. *Stochastic Processes in Queueing Theory.* Springer-Verlag, New York, 1976.

[Bru88] H. Bruneel. Queueing behavior of statistical multiplexers with correlated inputs. *IEEE Transactions on Communications,* Vol. 36. No. 12. pp. 1339-1341, 1988.

[Doo53] J. L. Doob. *Stochastic processes.* Wiley, New York, 1953.

[JPN80] N. Janakiraman, B. Pagurek, L. E. Neilson. Multiplexing Low-Speed Buffered Data Terminals. *IEEE Transactions on Communications*, Vol. COM-28. pp. 1838-1843, 1980.

[Kle75] L. Kleinrock. *Queueing systems.* Wiley-Interscience, New York, 1975.

[Kli86] G. Klimov. *Probability Theory and Mathematical Statistics.* Mir Publishers, Moskow, 1986.

[KT75] S. Karlin, H. M. Taylor. *A First Course in Stochastic Processes.* Academic Press, London, 1975.

[McDS94] D. E. McDysan, D. L. Spohn. *ATM: theory and application.* McGraw-Hill series on computer communications, New York, 1995.

[Pak69] A. G. Pakes. Some conditions for ergodicity and recurrence of Markov chains. *Oper. Res.*, Vol. 17. pp. 1058-1061, 1969.

[Rén68] A. Rényi. *Valószínűségszámítás.* Tankönyvkiadó, Budapest, 1968.

[RV91] J. W. Roberts, J. T. Virtamo. The Superposition of Periodic Cell Arrival Streams in an ATM Multiplexer. *IEEE Transactions on Communications*, Vol. 39. No. 2. pp. 298-303, February 1991.

[TGA88] P. Tran-Ghia, H. Ahmadi. Analysis of a Discrete-time $G^{[X]}/D/1-S$ Queueing System with Applications in Packet-switching Systems. *IEEE INFOCOM'88*, pp. 0861-0870, 1988.

[Vit86] A. M. Viterbi. Approximate Analysis of Time-Synchronous Packet Network. *IEEE Journal on Selected Areas in Communications*, Vol. SAC-4. pp. 879-890, 1986.

[WB94] S. Wittevrongel, H. Bruneel. Performance analysis of ATM queues with constant service times and nonindependent arrivals. *(Unknown source)*, at least 1994.

[WB97] S. Wittevrongel, H. Bruneel. Deriving the tail distribution of the buffer contennts in a statistical multiplexer with general heterogeneous on/off sources. *Proceedings: International Conference on the Performance and Management of Complex Communication Networks (PMCCN'97)*, pp. 37-56, 1997.