Cyclic-Waiting and Vacational Queuing Systems

Theses of the PhD dissertation

Kárász Péter
ELTE IK Doktori Iskola
(vezető: Dr. Demetrovics János)
Numerikus és szimbolikus számítások doktori program
(programvezető: Dr. Járai Antal)

Témavezető: Dr. Lakatos László
a matematikai tudományok kandidátusa
egyetemi docens

2008
1 Introduction

Cyclic-waiting queuing systems are characterized by the feature that a request for service can be repeated only after a constant period of time. A typical example can be an automatic redialling device, which again and again attempts to dial the called number after a deterministic time if the line is engaged. This also occurs at airports where planes can start landing upon arrival or have to start a circular manoeuvre, when the runway is used or there are other planes waiting, and they can only put their next request to land after they have completed a full cycle. Another application in digital technology is the use of an optical buffer, which is a device that is capable of temporarily storing light (or rather, data in the form of light). As light cannot be frozen, a typical optical buffer is realized by a single loop, in which data circulate a variable number of times.

The dissertation describes and further generalizes the results of Lakatos on cyclic-waiting systems. It investigates the so-called relative priority systems, which serves two types of customers of different priority. In these systems only one customer of first type can be present, it can only be accepted for service in the case of a free system, whereas in all other cases the requests of such customers are turned down. There is no such restriction on customers of second type; they are serviced immediately or join a queue in case of a busy server. This model can be applied for systems accepting impatient customers who need urgent service; if they cannot get serviced, they leave the system and find another server which is free. I examined this type of system with different continuous distributions in [4, 5, 6, 13], with discrete ones in [9, 10, 12]; simulation results were also included in [13, 1].

2 Continuous cyclic waiting retrial systems

Chapters 2 and 3 summarize my results on Lakatos-type queuing systems with two different priority customers and various continuous and discrete service time distributions, respectively.

In Chapter 2 an embedded Markov-chain is defined, whose states are identified with the number of customers in the system at moments just before the
service of a new customer begins. For this chain the following probabilities are introduced:

- \( a_{ji} \): the probability of appearance of \( i \) customers of second type at the service of a \( j^{th} \)-type customer \((j = 1, 2)\) if at the beginning there is only one customer in the system;

- \( b_i \): the probability of appearance of \( i \) customers of second type at the service of a second-type customer, if at the beginning of service there are at least two customers in the system;

- \( c_i \): the probability of appearance of \( i \) customers of second type after free state.

These transition probabilities, as well as the equilibrium probabilities are given through their generating functions.

**Theorem 2.1.1** Consider a Lakatos-type queuing system with two types of customers forming Poisson-processes with parameters \( \lambda_1 \) and \( \lambda_2 \). Service time distributions of customers of either type may be exponential with parameters \( \mu_i \) or uniform in the intervals \([\alpha_j, \beta_j]\) \((\alpha_j \) and \( \beta_j \) are multiples of \( T \)); thus four different cases are to be considered \((i, j = 1, 2)\).

The matrix of transition probabilities of the defined chain has the form:

\[
\begin{pmatrix}
  a_{00} & a_{01} & a_{02} & c_0 & c_1 & c_2 & c_3 & \ldots \\
  a_{20} & a_{21} & a_{22} & a_{23} & \ldots \\
  0 & b_0 & b_1 & b_2 & \ldots \\
  0 & 0 & b_0 & b_1 & \ldots \\
  \vdots & \vdots & \vdots & \vdots & \ddots
\end{pmatrix}.
\]

The elements of the matrix are determined by their generating functions below. The type of service time distribution is indicated in the upper index: \( A_j^{(exp,uni)}(z) \) indicates the type of service time distribution of \( j^{th} \)-type customers, and \( B_j^{(exp,uni)}(z) \) indicates the type of service time distribution of second-type customers.
They are connected with the relation

\[ p \]

where \( A \) and \( B \) are the first two probabilities of the equilibrium distribution.

**Theorem 2.1.2** The generating function of the equilibrium distribution of this chain is:

\[
P(z) = \sum_{i=0}^{\infty} p_i z^i = \frac{p_0 (z C(z) - B(z)) + p_1 (A_2(z) - B(z))}{z - B(z)},
\]

where \( p_0 \) and \( p_1 \) are the first two probabilities of the equilibrium distribution. They are connected with the relation \( p_1 = \frac{1 - \alpha}{\alpha_0} p_0 \), and

\[
p_0 = \frac{1 - B'(1)}{1 - B'(1) + C'(1) + \frac{1 - \alpha}{\alpha_0} (A_2'(1) - B'(1))},
\]
where
\[
A_j' \exp(1) = \frac{\lambda_2}{\lambda_2 + \mu_j} \left( 1 + \frac{\lambda_2 T}{1 - e^{-\mu_j T}} \right),
\]
\[
A_j' \uni(1) = \frac{\lambda_2}{\lambda_2 (\beta_j - \alpha_j) (1 - e^{\lambda_2 T})} + \frac{\lambda_2 (\alpha_j + \beta_j + T)}{2},
\]
\[
B' \exp(1) = 1 - \frac{\lambda_2 T}{1 - e^{-\lambda_2 T}} \left( e^{-\lambda_2 T} - \frac{\lambda_2}{\lambda_2 + \mu_2} \cdot \frac{1 - e^{-(\lambda_2 + \mu_2) T}}{1 - e^{-\mu_2 T}} \right),
\]
\[
B' \uni(1) = \frac{\lambda_2 (\alpha_2 + \beta_2 + T)}{2},
\]
\[
C'(1) = \frac{\lambda_1}{\lambda_1 + \lambda_2} A_1'(1) + \frac{\lambda_2}{\lambda_1 + \lambda_2} A_2'(1).
\]

It is also shown that the ergodicity of the process does not depend on customers of first type; its necessary and sufficient condition is stated in the next theorem.

**Theorem 2.1.4** The condition of existence of the ergodic distribution is the fulfilment of one of the following inequalities.

If the service time of second-type customers is exponentially distributed (regardless of the distribution of service time of first-type customers):
\[
\frac{\lambda_2}{\mu_2} < e^{-\lambda_2 T} \frac{1 - e^{-\mu_2 T}}{1 - e^{-\lambda_2 T}}.
\]

If the service time of second-type customers is uniformly distributed (regardless of the distribution of service time of first-type customers):
\[
\frac{\lambda_2 (\alpha_2 + \beta_2 + T)}{2} < 1.
\]

The dependence on input parameters of condition (2) is simple, but it is more complex in (1). For the latter case it is more convenient if parameters that guarantee ergodicity are represented in a graph. The shaded areas (the areas under the curves, excluding the curves themselves) in Figure 2.3 show pairs \((\mu_2, \lambda_2)\) suitable for ergodicity at some arbitrary values of the cycle-time (at \(T = 0.2, 0.4, 0.6, 0.8, \) and \(1.0\)). The graph also displays the limit case \(T \to 0\).
It can be clearly seen that at fixed capacity of the server (at constant $\mu_2$) the longer the cycle-time is, the lower input intensity the system can tolerate. Additionally, at fixed intensity of the input flow, with longer cycle-time the server has to be of higher performance.

The influence of idle periods — while the system is waiting for the next entity to reach its starting position to be able to start its service — becomes less and less while $T \to 0$. Within this transition conditions of ergodicity (1) and (2) turn into classical conditions $\frac{1}{\mu_2} < \frac{1}{\lambda_2}$ (see Figure 2.3) and $\frac{a_2 + b_2}{2} < \frac{1}{\lambda_2}$, i.e. the expectation of the service time must be less than the expectation of the inter-arrival time. In the light of this, (2) is even more straightforward: the mean value of the service time increased by the average idle time (on average $\frac{T}{2}$ time is needed for the next customer in the queue to reach the starting position) has to be less than the average inter-arrival time. Unfortunately, (1) cannot be interpreted in such a clear probabilistic way.

As a corollary to the theorems, the limit distributions while $T \to 0$ are also given for all combinations of the service time distributions. In a separate section some generalizations are carried out, with no restrictions on the boundaries of the uniform distributions. To validate results, equilibrium probabilities of states 0 and 1 are taken a closer look, examining their depen-
dence on the input parameters. To underpin my theory, numerical computer experiments are carried out, which complete the investigation of the continuous case.

3 Discrete cyclic-waiting retrial systems

In the light of technical applications it is important to consider discrete models. Chapter 3 is devoted to discrete, relative-priority Lakatos-type systems. The method in this case follows a similar routine as in Chapter 2. Cycle-time $T$ is divided into $n$ equal time-slices. The probability of appearance of a $j$th-type customer during a certain time-slice is $r_j$, i.e. inter-arrival times are geometrically distributed with parameters $r_j$ $(j = 1, 2)$. However, unlike with continuously distributed service times, in discrete systems different types of customers do appear during the same time-slice with non-zero probability. There are several ways to deal with this phenomenon, called collision; we opt for three methods to treat it: both of them are refused; first-type customers are accepted, but second-type ones are refused; both of them are accepted, but first-type ones are served first.

The same embedded Markov-chain is defined as in the continuous case, but when applying the third collision treatment method, in addition to previously defined transition probabilities new ones have to be introduced. Let $a_{12i}$ denote the probability of appearance of $i$ customers of second-type at the service of a first-type customer, if the service process started with the simultaneous appearance of customers of different types. Two service time distributions are considered; for one possibility service times are geometrically distributed with parameters $q_j$, i.e. the service of a $j$th-type customer continues during a time unit with probability $q_j$, and terminates with probability $1 - q_j$. The other examined alternative is when service time distributions are uniform in the intervals $[\gamma_j, \delta_j]$, where $\gamma_j$ and $\delta_j$ are multiples of $T$, i.e. the probability that the service of a $j$th-type customer is any time units in this interval is $q_j = \frac{T}{n(\delta_j - \gamma_j)}$. All three collision disciplines are investigated with each system. The matrix of transition probabilities is the same as in the continuous case; their generating functions, as well as that of the equilibrium
probabilities are determined.

**Theorem 3.1.1** Consider a Lakatos-type queuing system serving two types of customers in which inter-arrival time distributions are geometric with parameters \( r_j \) \((j = 1, 2)\). Service time distributions of customers of either type may be geometric with parameters \( q_i \) or uniform in the intervals \([\gamma_j, \delta_j]\); thus (considering the three collision disciplines) 12 different cases are to be examined \((i, j = 1, 2)\). The matrix of transition probabilities is identical with the one in the continuous case; their generating functions are given below. The type of service time distribution is indicated in the upper index: \( A_j^{\{\text{geo}, \text{uni}\}}(z) \) indicates the type of service time distribution of \( j^{\text{th}} \)-type customers, \( A_{12}^{\{\text{geo}, \text{uni}\}}(z) \) indicates the type of service time distribution of first-type customers, and \( B_j^{\{\text{geo}, \text{uni}\}}(z) \) indicates the type of service time distribution of second-type customers.

\[
A_j^{\text{geo}}(z) = \frac{(1 - r_2)(1 - q_j)}{1 - q_j (1 - r_2)} + z \frac{r_2 (1 - q_j)}{1 - q_j (1 - r_2)} + \frac{r_2 q_j (1 - q_j^2)}{(1 - q_j (1 - r_2)) (1 - q_j^n (1 - r_2 + r_2 z)^n)},
\]

\[
A_j^{\text{uni}}(z) = \frac{q_j}{r_2} \left( (1 - r_2)^\frac{\gamma_j}{r_2} - (1 - r_2)^\frac{\delta_j}{r_2} \right) + z q_j \left( (1 - r_2)^\frac{\gamma_j}{r_2} - (1 - r_2)^\frac{\delta_j}{r_2} \right) \times
\]

\[
\times \frac{1 - (1 - r_2)^n}{r_2 (1 - r_2)^n} \frac{(1 - r_2 + r_2 z)^\frac{\gamma_j}{r_2}}{1 - \frac{1 - r_2 + r_2 z}{1 - r_2}^n} + z q_j (1 - r_2 + r_2 z) \left[ n \frac{(1 - r_2 + r_2 z)^\frac{\gamma_j}{r_2} - (1 - r_2 + r_2 z)^\frac{\delta_j}{r_2}}{1 - (1 - r_2 + r_2 z)^n} - (1 - r_2)^\frac{\delta_j}{r_2} \frac{\gamma_j}{r_2} \right],
\]
Theorem 3.1.2  

The generating function of the equilibrium distribution of the defined chain is:

\[ P(z) = \sum_{i=0}^{\infty} p_i z^i = \frac{p_0 (z C(z) - B(z)) + p_1 z (A_2(z) - B(z))}{z - B(z)}, \]

and \( C(z) \) depends on collision policies:

I. \[ C(z) = \frac{r_1(1-r_2)}{r_1+r_2-r_1r_2} A_1(z) + \frac{r_2(1-r_1)}{r_1+r_2-r_1r_2} A_2(z) + \frac{r_1r_2}{r_1+r_2-r_1r_2}, \]

II. \[ C(z) = \frac{r_1}{r_1+r_2-r_1r_2} A_1(z) + \frac{r_2(1-r_1)}{r_1+r_2-r_1r_2} A_2(z), \]

III. \[ C(z) = \frac{r_1(1-r_2)}{r_1+r_2-r_1r_2} A_1(z) + \frac{r_2(1-r_1)}{r_1+r_2-r_1r_2} A_2(z) + \frac{r_1r_2}{r_1+r_2-r_1r_2} A_{12}(z). \]
where \( p_0 \) and \( p_1 \) are the first two probabilities of the equilibrium distribution. They are connected with the relation
\[
p_0 = \frac{1 - B'(1)}{1 - B'(1) + C'(1) + \frac{1 - c_{a_2}}{a_{20}} (A'_2(1) - B'(1))},
\]
where
\[
A'_j^{\text{geo}}(1) = \frac{r_j}{1 - q_j} \left( 1 + \frac{nr_j q_j}{1 - q_j} \right),
\]
\[
A'_j^{\text{uni}}(1) = -a_{j0} + \frac{T}{\delta_j - \gamma_j} \left[ (1 - r_j)^2 - (1 - r_j)^2 \frac{\delta_j}{1 - (1 - r_j)^n} \right] \frac{(1 - r_j)^n}{1 - (1 - r_j)^n} + \frac{nr_j \gamma_j + \delta_j + T}{2},
\]
\[
A'_{12}^{\text{geo}}(1) = 1 + \frac{nr_2}{1 - q_1^n},
\]
\[
A'_{12}^{\text{uni}}(1) = 1 + \frac{nr_2 \gamma_1 + \delta_1 + T}{2},
\]
\[
B'^{\text{geo}}(1) = 1 - \frac{nr_2}{1 - (1 - r_2)^n} \left[ (1 - r_2)^n - \frac{r_2 q_2}{1 - q_2} (1 - r_2)^n \right],
\]
\[
B'^{\text{uni}}(1) = \frac{nr_2 \gamma_2 + \delta_2 + T}{2},
\]
and \( C'(1) \) depends on collision policies:
\[
I. \quad C'(1) = \frac{r_1(1 - r_2)}{r_1 + r_2 - r_1 r_2} A'_1(1) + \frac{r_2 (1 - r_1)}{r_1 + r_2 - r_1 r_2} A'_2(1),
\]
\[
II. \quad C'(1) = \frac{r_1}{r_1 + r_2 - r_1 r_2} A'_1(1) + \frac{r_2 (1 - r_1)}{r_1 + r_2 - r_1 r_2} A'_2(1),
\]
\[
III. \quad C'(1) = \frac{r_1(1 - r_2)}{r_1 + r_2 - r_1 r_2} A'_1(1) + \frac{r_2 (1 - r_1)}{r_1 + r_2 - r_1 r_2} A'_2(1) + \frac{r_1 r_2}{r_1 + r_2 - r_1 r_2} A'_{12}(1).
\]
It is also shown that ergodicity of the process depends neither on first-type customers nor on collision policies, and its necessary and sufficient condition is stated in the next theorem.

**Theorem 3.1.4** The condition of existence of the equilibrium distribution is the fulfillment of one of the following inequalities.

If the service time of second-type customers is geometrically distributed (regardless of the distribution of service time of first-type customers and the applied collision discipline):
\[
\frac{r_2 q_2}{1 - q_2} \frac{1 - q_2^n (1 - r_2)^n}{1 - q_2 (1 - r_2)} < (1 - r_2)^n.
\]
If the service time of second-type customers is uniformly distributed (regardless of the distribution of service time of first-type customers and the applied collision discipline):

\[
\frac{nw_2 \gamma_2 + \delta_2 + T}{T} < 1.
\]  

(4)

The dependence on input parameters of condition (4) is simple and can be interpreted easily: considering that \(\frac{T_{\frac{1}{n}r_2}}{n}r_2\) is the average inter-arrival time, the condition rewritten in the form \(\frac{\gamma_2 + \delta_2}{2} + \frac{T}{2} < \frac{T}{n}r_2\) necessitates that the average service time increased by the average idle time (on average \(\frac{T}{2}\) time is needed for the next customer in the queue to reach the starting position) be less than the average inter-arrival time.

For the more complex geometric relation (3) it is more convenient if parameters that guarantee ergodicity are represented in a graph. Numerical examination of inequality (3) reveals that for fixed values of \(n\) and \(q_2\), the process is ergodic if \(r_2 \in (0, r_2^{\text{max}})\). The shaded areas (the areas under the curves, excluding the curves themselves) in Figure 3.1 show pairs \((q_2, r_2)\) suitable for ergodicity at some arbitrary values of \(n\) (at \(n = 1, 2, 4, 6, 8, \) and 10).

It can be seen that in the (somewhat degenerate) case \(n = 1\), the condition of ergodicity is \(r_2 < 1 - q_2\), which has a clear probabilistic interpretation; namely, that the probability of appearance of a new customer of second-type during a time slice (which is \(T\) in this case) has to be less than the probability of completing a service during the same amount of time.

Like in the continuous case, equilibrium probabilities of states 0 and 1 are also investigated. Additionally, the expected value of the queue-length in the examined type of queuing systems is also given.

**Theorem 3.2.1** The expected value of the length of the queues in cyclic-waiting systems serving two types of customers is given by

\[
P'(1) = 1 - p_0 + \frac{(1 - p_0) B''(1) + p_0 \left( C''(1) + \frac{1 - \alpha_0}{\alpha_2} (A''_2(1) - B''(1)) \right)}{2 (1 - B'(1))}.
\]
Queue-length is explicitly determined in the case of geometric service time distributions, and its dependence on input parameters is analyzed graphically.

Both in the continuous and the discrete case all four possibilities (considering the service time distributions of the two types of customers) have been investigated separately, and presented at different forums [1, 2, 3, 4, 5, 6, 9, 10, 12, 13, 14]. In both chapters the unified theory is given, i.e. generating functions can be combined together as necessary, in accordance with the type of service time distributions of first- and second-type customers. Similarly, the generating function of the equilibrium distribution, as well as the conditions of ergodicity are valid for all cases.

4 Queuing systems with vacation

A queuing system with vacation is a queuing system in which the server intermittently spends time away from the queue, perhaps because of a breakdown and repair or because it is serving other jobs. Some examples are
token-passing schemes in local area networks with distributed channel access control, and single-server multi-queue models.

The last chapter of the dissertation is devoted to the investigation of a queuing system which accepts bulk-arrivals, and where the first appearing group of requests initiates a vacation during which the server is prepared for the service, and actual service can only begin after this period of time. Classical methods provide the full description of the system, but usually result in very complicated expressions, especially at higher states, which makes the applicability of the methods rather difficult.

Lakatos applied a different method: the functioning of the $M/G/1$ system can be described as a regenerative process, in which the end-points of the busy periods are the regeneration points. This way the equilibrium probabilities can be determined with the help of the expectation of times spent at certain levels during a busy period and the average length of the busy period.

The novelty is to calculate times spent above level $i - 1$ and above level $i$, based on the idea that they have the same structure; and thus receive a recurrence relation on the expected value of time spent on level $i$ as a difference of the previous ones. In Chapter 4 the results of the previously mentioned systems are generalized, and recursive formulae are given on the equilibrium probabilities of the $M/G/1$ system with vacation at the beginning of the busy period, and which accepts bulk arrivals [7, 8, 11].

**Theorem 4.2.1** In the bulk-arrival $M/G/1$ system with vacation at the beginning of the busy period the ergodic distribution exists if $q < 1$; and equilibrium probabilities are determined by

$$p_i = \frac{\xi'_i}{\zeta'_i}, \quad i = 0, 1, 2, \ldots,$$

where $\xi'_i$ and $\zeta'$ are the mean values of time spent on the $i^{th}$ level for a busy period, and the duration of the busy period in the system with vacation, respectively.

The average length of the busy period proves to be $\eta\alpha\tau_1 \frac{1}{1-\eta}$, where $\eta$ is the mean length of the vacation, $\alpha$ is the average number of customers arriving
in a group, \( \tau \) is the mean length of a service of a customer, and \( \varrho \) is the utilization factor of the system. The mean times spent on certain levels are determined by applying the above method, and are given recursively.

**Theorem 4.2.4** The mean value of time spent on the \( i \)th level for a busy period satisfies the recurrence relations

\[
\begin{align*}
\xi'_0 &= \tau, \\
\xi'_1 &= \frac{\tau}{c_0} + f_1 (\eta - \tau), \\
\xi'_2 &= \xi_2 + (2 - f_1) \left( \frac{\tau}{1 - \varrho} - \zeta_1 \right) - \frac{\tau}{c_0} + f_2 (\eta - \tau), \\
&\vdots \\
\xi'_i &= \xi_i + \sum_{k=2}^{i-1} (1 - f_1 - \ldots - f_{i-k}) \xi_k + \\
&\quad + (1 - f_1 - \ldots - f_{i-1}) \left( \frac{\tau}{1 - \varrho} - \zeta_1 \right) + f_i (\eta - \tau), \quad (i \geq 3),
\end{align*}
\]

where \( \xi_i \) and \( \zeta_i \) are the mean values of time spent on the \( i \)th level, and above it for a busy period in the ordinary (non-vacational) system, respectively; \( c_k \) is the probability of appearance of \( k \) new groups of customers during the service of another customer; \( f_k \) is the probability of \( k \) requests being present at the end of vacation.

**References**


