Wissenschaftliche Berichte des
Institutes für Fördertechnik und Logistiksysteme
der Universität Karlsruhe (TH) Band 70

Prof. Dr.-Ing. Kai Furmans (Hrsg.)

Marc Schleyer

# Discrete Time Analysis of Batch Processes in Material Flow Systems

universitätsverlag karlsruhe

Marc Schleyer

# Discrete Time Analysis of Batch Processes in Material Flow Systems

Wissenschaftliche Berichte des
Institutes für Fördertechnik und Logistiksysteme
der Universität Karlsruhe (TH)
Band 70

# Discrete Time Analysis of Batch Processes in Material Flow Systems

von
Marc Schleyer

universitätsverlag karlsruhe

Dissertation, Universität Karlsruhe (TH)
Fakultät für Maschinenbau, 2007
Referenten: Prof. Dr.-Ing. K. Furmans, Prof. Dr.-Ing. R. Bruns, Prof. Dr. K.-H. Waldmann

**Impressum**

Universitätsverlag Karlsruhe
c/o Universitätsbibliothek
Straße am Forum 2
D-76131 Karlsruhe
www.uvka.de



Universitätsverlag Karlsruhe 2007
Print on Demand

ISSN: 0171-2772
ISBN: 978-3-86644-167-5

# Vorwort

*I can no other answer make, but, thanks, and thanks.*

William Shakespeare

Die vorliegende Arbeit entstand während meiner Tätigkeit als wissenschaftlicher Mitarbeiter am Institut für Fördertechnik und Logistiksysteme der Universität Karlsruhe (TH). Das Vorwort möchte ich nutzen, um einigen Leuten, die zum Gelingen der vorliegenden Dissertationsschrift beigetragen haben, zu danken.

Meinem akademischen Lehrer Herrn Prof. Dr.-Ing. Kai Furmans, Inhaber des Lehrstuhls Logistik und Leiter des Instituts für Fördertechnik und Logistiksysteme, gilt mein Dank für die Übernahme des Referats und besonders für die Unterstützung und Förderung meiner Arbeit als wissenschaftlicher Mitarbeiter. Durch seine eigenen Forschungsarbeiten hat er mir die Vorteile der zeitdiskreten bedientheoretischen Methoden für die Modellierung von Materialflusssystemen vor Augen geführt und mich somit für diese Arbeit motiviert. Ich genoss eine von Freiheit geprägte Arbeitsatmosphäre, was maßgeblich zum Gelingen meiner Promotion beigetragen hat.

Den Korreferenten Prof. Dr.-Ing. Rainer Bruns und Prof. Dr. Karl-Heinz Waldmann danke ich für die Übernahme des Korreferats, sowie Prof. Dr.-Ing. Christoph Stiller für die Übernahme des Vorsitzes der mündlichen Prüfung.

Ich möchte mich bei Prof. Dr.-Ing. Dr. h.c. Dieter Arnold für sein Vertrauen in meine Person bedanken, weil er mir durch die Anstellung als wissenschaftlicher Mitarbeiter an seinem Institut das Verfassen dieser Arbeit überhaupt erst ermöglichte.

Ebenfalls bedanke ich mich bei Maria di Mascolo, in deren Forschergruppe ich Dank eines Stipendiums für die Dauer eines halben Jahres wertvolle Erfahrungen sammeln durfte. In dieser Zeit hatte ich perfekte Bedingungen zum Forschen und so ist es mir gelungen, die analytisch schwierigsten Fragestellungen meiner Arbeit zu beantworten.

Mein Dank gilt meinen Kollegen, die mich stets durch inspirierende Diskussionen tatkräftig unterstützt haben. Die sehr angenehme Arbeitsatmosphäre und die zahlreichen gemeinsamen Unternehmungen werde ich stets in guter Erinnerung behalten. Hervorheben möchte ich Christian Lippolt, der als Betreuer meiner Diplomarbeit mein Interesse am wissenschaftlichen Arbeiten geweckt hat und mich dazu bewegt hat, das Verfassen einer Dissertationsschrift anzustreben. Ich möchte mich bei Steffen Haak bedanken, der mich tatkräftig bei der Programierung einer Anwender-Oberfläche unterstützt hat, die es ermöglicht, die Algorithmen dieser Arbeit für die Praxis auf anschauliche Art und Weise nutzbar zu machen.

Besonders möchte ich mich bei meinen Freunden bedanken, bei denen ich stets ein sehr angenehmes und hervorragend funktionierendes soziales Umfeld vorfinden konnte. Besonders derer Dank gilt Clemens van Dinther, der in zahlreichen Diskussionen mit wertvollen Anregungen zum Gelingen dieser Arbeit beitragen hat.

Mein tiefster persönlicher Dank gilt meinen Eltern Maria und Karl Schleyer, die stets dafür sorgten, dass ich alle Freiheiten und Möglichkeiten hatte, mich in meiner Persönlichkeit zu entwickeln. Sie waren immer am Fortschritt der Arbeit interessiert und haben an mich geglaubt. Das Gelingen dieser Arbeit habe ich im allerhöchsten Maße ihrer Erziehung und Zuneigung zu verdanken, weshalb Ihnen diese Arbeit gewidmet ist.

Karlsruhe, April 2007 Marc Schleyer

# Kurzfassung

*Marc Schleyer*

## Zeitdiskrete Modellierung von Batch-Prozessen in Materialflusssystemen

Diese Arbeit beschäftigt sich mit der Entwicklung von analytischen Verfahren für die Leistungsbewertung von Batch-Prozessen in Materialflusssystemen. Die praktische Anwendung der entwickelten zeitdiskreten Modelle ist vor allem in der Grobplanungsphase von Logistiksystemen zu sehen. Für diese Planungsaufgabe weisen sie einen ausreichenden Detailierungsgrad auf und es können in kurzer Zeit verschiedene Planungsszenarien quantitativ untersucht und bewertet werden. Die für die verschiedenen Planungsszenarien erforderlichen Systemkapazitäten können berechnet werden.

Im Gegensatz zu einer herkömmlichen zeitkontinuierlichen Modellierung eröffnet die Diskretisierung der Zeit neue Möglichkeiten für die Modellierung und Analyse von stochastischen Systemen. Anstatt mit Mittelwerten und Varianzen können nun Leistungskenngrößen (z.B. Wartezeiten, Durchlaufzeiten, Bestände etc.) mit Wahrscheinlichkeitsverteilungen beschrieben werden, wodurch ihre Aussagekraft erhöht wird. Dies eröffnet dem Planer die Möglichkeit Materialflusssysteme so auszulegen, dass Kundenaufträge in einer vorgegebenen Zeit und mit einer vorgegebenen Wahrscheinlichkeit, die in der industriellen Praxis gewöhnlich zwischen 95% und 99% liegt, erfüllt werden können.

Es besteht Forschungsbedarf geeignete zeitdiskrete analytische Methoden für die Analyse von Materialflusssystemen unter generellen stochastischen Verteilungsannahmen zu entwickeln. Vor allem in Materialflusssystemen kommt es bedingt durch die Zusammenfassung von Aufträgen, durch Transporte und durch Sortiervorgänge zur Bildung von Batches. Daher werden in der vorliegenden Arbeit verschiedene analytische Modelle entwickelt, die Batch-Prozesse in der Evaluierung von Materialflusssystemen mit berücksichtigen.

Zunächst werden drei verschiedene Prozesse der Batch-Bildung im Detail untersucht. Es können zwei grundlegende Batch-Bildungs Prozesse, nämlich zum einen die Batch-Bildung nach Menge und zum anderen nach Zeit, identifiziert werden. Diese beiden Prozesse können in verschiedenster Art modifiziert werden. So wird zusätzlich die Batchbildung nach Mindestmenge betrachtet, bei der innerhalb einer vorgegebenen Zeit das Batch gebildet wird, aber der Prozess erst abgeschlossen ist, bis eine Mindestmenge erreicht ist. Für die genannten Batch-Bildungsprozesse werden unter generellen stochastischen Verteilungsannahmen die Warte- und Zwischenabgangszeitverteilung exakt bestimmt.

Des Weiteren werden Modelle zur Leistungsbewertung des G/G/1-Bediensystems mit

Batch-Ankünften, des $G/G^{[K,K]}/1$-Bediensystems, des $G/G^{[L,K]}/1$-Bediensystems und der Sortierung von Batches eingeführt. Verfahren für die Berechnung der Warte- und Zwischenabgangszeitverteilung werden vorgestellt. Der Systemzustand des G/G/1-Bediensystems mit Batch-Ankünften wird zum Ankunftszeitpunkt eines Kundenauftrags und der Systemzustand des $G/G^{[L,K]}/1$-Bediensystems zum Abgangszeitpunkt eines bearbeiteten Kundenauftrags analysiert.

Zusätzlich zu der analytischen Modellbeschreibung wird das Systemverhalten jedes einzelnen Modelles untersucht. Sowohl mathematisch als auch in einer Reihe von nummerischen Beispielen werden die Zusammenhänge zwischen Eingangs- und Ergebnissgrößen aufgezeigt. Es ist ersichtlich, dass die Wahrscheinlichkeit einen Auftrag rechtzeitig zu erfüllen sensitiver auf Parameteränderungen reagiert als die mittlere Systemverweilzeit. Es wird gezeigt, dass eine Analyse, die auf die Berechnung von Mittelwerten beruht, die Konsequenzen eines instabilen Prozessverhaltens auf die Wahrscheinlichkeit der rechtzeitigen Auftragerfüllung unterschätzt. Dies motiviert den Einsatz der vorgestellten zeitdiskreten analytischen Verfahren für die praktische Anwendungen.

# Abstract

*Marc Schleyer*

## Discrete Time Analysis of Batch Processes in Material Flow Systems

Scope of this doctoral thesis is the development of appropriate models for the evaluation of batch processes in material flow systems. The presented analytical methods support the long range planning in an early planning stage, in which capacities are determined to minimize the facility costs under the condition of cycle time targets. In this planning stage a rough and extensive "what-if" analysis is required in order to find a competitive solution. We choose an analytical approach for a performance evaluation of material flow systems since it is more time efficient and allows deep insights into the general system's behavior.

Performance measures based on system averages are not sufficient to verify whether the requested shipping times can be met with an acceptable probability, which usually lies between 95% and 99%, possibly depending on order types. Therefore, for the evaluation of design alternatives in respect to their ability to reach the requested sojourn time from order entry to exit, discrete time queueing models are proposed. These models enable an analysis of general distributed process on the basis of distributions.

Since there is still a lack of appropriate discrete time models for the analysis of material flow processes, we are motivated to find new solutions for problems in this field. Especially, models for the description of batch processes are missing. Due to efficiency reasons batch processes are very common in material flow systems. Therefore, we develop a variety of batch queueing models in the discrete time domain.

At first, we study different batch building processes in detail. Two basic batch building modes, the capacity and the timeout rule, and additionally a possible modification called the minimum batch size rule, are analyzed. For the named batch building modes we derive both interdeparture and waiting time distribution.

Next, we introduce analytical models for an evaluation of the G/G/1-queue with batch arrivals, $G/G^{[K,K]}/1$-queue, $G/G^{[L,K]}/1$-queue, and batch split operation. Again, each model provides methods for a detailed analysis of the waiting and departure process on the basis of discrete distributions. Furthermore, the system's state of the G/G/1-queue with batch arrivals is investigated at the arrival instant and the system's state of the $G/G^{[L,K]}/1$-queue at the departure instant.

In addition to the analytical descriptions we point out the system's behavior of each analytical model. We show both, mathematically and numerically, the dependence of performance measures on the input parameters. Various numerical experiments explain that the probability of an on-time order fulfillment react more sensitively on parameter changes than mean values. We present that an analysis focusing on mean values underestimates the consequences of an instable process behavior on the on-time order fulfillment.

# Contents

# Glossary of Notation

| | |
|---|---|
| $[\tau_{n+k+1} - \tau_n](l)$ | time interval between $\tau_{n+k+1}$ and $\tau_n$ depending on the position $l$ of customer $n$ |
| $\alpha_i^n(k,l)$ | distribution which describes the probability that the time interval between $\tau_{n+k+1}$ and $\tau_n$ is $i$ time units depending on the position $l$ of customer $n$ |
| $\beta_i$ | idle time distribution |
| $\Delta_m(f_x)$ | operator which shift the elements of a the distribution $f_x$ down by $m$ units |
| $\delta_n$ | departure instant of the $n$th customer |
| $\eta_i$ | distribution of the number of customers at the arrival instant |
| $\lambda$ | arrival rate |
| $\lambda_{batch}$ | batch arrival rate |
| $\nu_i$ | distribution of the number of customers at the departure instant |
| $\otimes$ | convolution operator |
| $\overline{a}_i$ | probability that the random variable $A$ is at least $i$ |
| $\Phi_{k,m}^n(l)$ | distribution that the time interval $[\tau_{n+k+1} - \tau_n](l)$ is composed of $m$ interarrival time intervals |
| $\pi_m(f_x)$ | operator which sums up all values of the distribution $f_x$ with $x < m$ and adds their sum to $\pi_m$ |
| $\sigma_u$ | $u$%-quantile $(0 \leq u \leq 1)$ |
| $\tau_n$ | arrival instant of the $n$th customer |
| $\varpi_i$ | distribution of the number of customers in the queue immediately after the service start |
| $\xi$ | probability that the sum of at least one random variable is $j$ |
| $A$ | interarrival time |
| $a_i^{h \to j}$ | interarrival stream from node $h$ to $j$ |

| | |
|---|---|
| $a_i$ | interarrival time distribution |
| $B$ | service time |
| $b_i$ | service time distribution |
| $b_{batch,i}$ | service time distribution for a whole batch ($i = 1, \ldots, b_{batch,max}$) |
| $B_{batch}$ | time to serve a whole batch |
| $c_X^2$ | squared coefficient of variation (scv) of the random variable $X$ |
| $c_{batch,i}$ | working balance distribution |
| $D$ | interdeparture time |
| $d_i$ | interdeparture time distribution |
| $E(X)$ | expectation of the random variable $X$ |
| $E(X^n)$ | $n$th moment of the random variable $X$ |
| $e_i$ | distribution of the number of still missing customers |
| $H$ | number of customers in the queue immediately after the service start |
| $K$ | server capacity |
| $k$ | collecting size |
| $L$ | minimum batch size |
| $N(\tau)$ | number of customers at the arrival instant |
| $N(t)$ | number of customers in the queue at time instant $t$ |
| $N_a$ | number of arrivals during one collecting process |
| $N_c$ | number of missing arrivals at the timeout instant if batch building according to the minimum batch size rule is applied |
| $o_l$ | distribution which describes the probability that an arbitrary chosen customer is located at position $l$ within a batch |
| $P$ | probability measure |
| $P(A)$ | probability of the event $A$ |
| $P(G_c = q)$ | probability that an arbitrary customer is part of a batch building process which requires $q$ additional arrivals if the batch building process according to the minimum batch size rule is applied |
| $p_{hj}$ | probability splitting from node $h$ to $j$ |

| | |
|---|---|
| $p_{ij}$ | transition probabilities |
| $q_l$ | distribution which describes the probability that an arbitrary chosen customer is element of a batch of size $l$ |
| $R$ | residual lifetime of a renewal process |
| $R_a$ | residual interarrival time |
| $R_a^{t_{out}}$ | residual interarrival time at the timeout instant if batch building according to the minimum batch size rule is applied |
| $R_y$ | remaining customers at the end of a collecting process |
| $r_{a,i}$ | residual interarrival time distribution |
| $r_i$ | residual lifetime distribution |
| $r_{y,i}$ | remaining customer distribution |
| $t_{out}$ | collecting time |
| $U$ | age of a renewal process |
| $u_i$ | distribution of the age |
| $V(l)$ | sojourn time of an arbitrary customer depending on position $l$ within his batch |
| $v_i$ | sojourn time distribution |
| $v_i(l)$ | sojourn time distribution of an arbitrary customer depending on position $l$ within his batch |
| $VAR(X)$ | variance of the random variable $X$ |
| $W(l)$ | waiting time of an arbitrary customer depending on position $l$ within his batch |
| $W^{II}$ | waiting time of an individual customer during the service of his batch |
| $w_i^{II}$ | waiting time distribution of an arbitrary individual customer during the service of the his batch |
| $W^{k,l+q}$ | waiting time for the customers who arrive with batch $k$ ($1 \leq k \leq l+q$); $l$ number of arrivals before the timeout ends; $q$ number of additional arrivals in order to complete the collecting process; the batch building process according to the minimum batch size rule is applied |

| | |
|---|---|
| $W^q$ | waiting time of the $q$th arrival encountered after the start of a service period if a batch service according to the minimum batch size rule is applied |
| $w_i(l)$ | waiting time distribution of an arbitrary customer depending on position $l$ within his batch |
| $w^I_{batch,i}$ | batch waiting time distribution |
| $W^I_{batch}$ | batch waiting time |
| $w_i$ | waiting time distribution |
| $w^{II}_i(l)$ | waiting time distribution of an individual customer during the service of his batch depending on position $l$ within his batch |
| $X^n_{batch}$ | working balance of the $n$th batch |
| $Y$ | batch size |
| $Y^{t_{out}}$ | number of customers collected within $t_{out}$ if batch building according to the *minimum batch size* rule is applied |
| $Y_d$ | batch size of the collected batch |
| $y_i$ | batch size distribution |
| $y_{d,i}$ | batch size distribution of the collected batch |
| $Z$ | number of customers who arrive during a service period |
| $z_i$ | distribution of the number of customers who arrive during a service period |
| W | waiting time |

# 1. Introduction

*Tout vient à qui sait attendre.*

Mary M. Curie

Waiting. We are in a permanent condition of waiting. Waiting for a great variety of events, both important and less important events. Waiting entails uncertainty and this uncertainty can often grow and evoke certain emotions within us, be it a flurry of excitement, fear and anxiety, or even pleasure. So, the situation of being caught in a traffic congestion with a seemingly endless queue of cars standing bumper to bumper, compounded by the urgency to reach a particular destination can often transform a usually calm person into a case of rage. In another situation, a hospital patient who waits for a critical surgery suffers fear of not knowing the outcome and will pin all hopes on a successful surgical procedure. In contrast, just like the saying "pleasure caused by anticipation is the best pleasure" suggests, there are so many waiting situations which cause the purest pleasure. Thus, the nine month awaiting the arrival of the newborn is a period of joy and happiness for the expectant parents.

Whether waiting is pleasant or unpleasant depends strongly on the expected outcome. However, an unpleasant waiting situation can be mitigated if we stay patient. Often, our impatience is unfounded. Therefore we refer to the opening quote: "All things come to those who know to wait". Especially, good ideas require patience.

The fact that waiting is omnipresent spurs many philosophers and other intellectuals to create a large amount of literature dealing with the subject of waiting. One such work is the famous drama "Waiting for Godot" by Samuel Beckett, which deals just about a waiting situation. This absurd play confronts the reader with the agony of Vladimir and Estragon awaiting the arrival of Godot. Beckett describes two evenings of paralyzed waiting in an endless sequence of similar evenings. Nothing happens, Godot never arrives.

In the sense of Beckett, waiting is static and is not activity, it is "inactivity". Many people esteem waiting time as lost time. However, when put it in a different light, we can see it inversely: Waiting is gained time, time for doing something which we wished to do for a long time. Therefore, we can use this gained time, for example, to think thoroughly about problems we have in order to make well-grounded decisions.

In material flow systems, we have to recognize unfortunately that material units, customer orders, and information are unable to gain any additional value through waiting. In this context, waiting time is considered as lost time. In addition, inventories and buffers have to be provided in order to bridge the waiting time. The purpose of this work is to introduce analytical models which allow a quantitative analysis of waiting phenomena in material flow systems. The presented insights into the system's behavior can be used to reduce waiting times in real world material flow systems, such that the gained time can be used for more meaningful activities, e.g. thinking and making well-grounded decisions.

## 1.1. Problem Description and Scope of the Book

In the current work we focus on the performance analysis of material flow systems. Let us exemplify the main criteria of such a performance analysis by means of two questions:

*What does a company profit from the manufacture of high quality products if the resulting demand is low due to the prices being too expensive?* Therefore, companies are forced to reduce their costs in order to offer at competitive prices. In this context, companies should focus on minimizing all activities which have no additional value for their products: Inventory, rework due to problems in product quality, scrap, customer claims etc. are all examples for such activities. However, before valueless activities can be minimized, processes which provide no additional value have to be identified and analyzed.

In general, processes which are related to the material flow provide no additional value in contrast to production processes. Thus, customers are not willing to pay more if material units are transported from machine A to B or if finished products are stored in an automated high rack warehouse. However, it is absolutely necessary to transport material units from machine A to B, since the manufacturing process on machine B has to follow the process on machine A. In addition, it is necessary to hold a stock of finished goods as means of reacting quickly to customer orders. It follows that the distance from machine A to B has to be kept to the minimum, that the transportation from machine A to B has to be carried out reliably, and that a sufficient buffer of material units has to be present at each machine in the production process. Equally, given the customer demand the required stock of finished goods has to be computed.

In the recent years companies have realized that their competitiveness depends crucially on the performance of their material flow system. In order to organize the material flow, practitioners and researchers introduced different strategies such as Just-in-Time, Kanban, CONWIP, Control Point Policy, Continuous Flow, Heijunka etc. in order to organize the material flow. The purpose of all these policies is the same, the reduction of valueless activities. In these cases in which analytical methods for a system's evaluation under general assumptions exist, it is addressed by a mean value analysis of the work in progress. In the current work, we propose models for the determination of the mean work in progress as well, however, let us ask a further question.

*What does a company profit from the manufacture of high quality products if they are unable to deliver them on-time?* Therefore, companies have to ensure an on-time delivery of orders to fulfill customer expectations. Material flow systems should be designed in such a way it guarantees the on-time order fulfilment with a given high probability which lies usually between 95% and 99%, depending on order types. Performance measures based on system averages are not sufficient to verify whether the requested shipping times can be met with an acceptable probability. We recognize that there is a lack of appropriate analytical methods to analyze material flows on the basis of general distributions. By means of this work we provide analytical methods on the basis of general distributions which allow the determination of the sojourn time distribution and therefore the calculation of the probability of an on-time order fulfillment.

Furthermore, batch processes are very common in material flow systems. We identify a requirement for research to develop appropriate models for the evaluation of batch processes. The purpose of the current work is to provide a toolkit of analytical models to analyze a variety of batch processes in material flow systems. Different batch building modes, batch arrival processes, batch service processes, and finally sorting of batches have to be investigated. The waiting and the departure process for each material flow element have to be analyzed in detail. We assume that time is discrete since this delivers advantages regarding modeling of material flow systems. These advantages will be pointed out in this work.

By means of our approach we address the following scope of application. The presented analytical methods support the long range planning in an early planning stage, in which capacities are determined to minimize the facility costs under the condition of cycle time targets. In this planning stage a rough and extensive "what-if" analysis is required in order to find a competitive solution. In this context, we have to state that material flow systems can be evaluated by discrete event simulation as well. We choose an analytical approach for a performance evaluation of material flow systems since it is more time efficient and allows deep insights into the general system's behavior.

Recapitulating, the following problems are addressed:

- Performance evaluation of material flow systems with the focus on batch processes
- Design and redesign of material flow systems under the constraint of an on-time order fulfillment
- Long range planning of material flow systems in an early planning stage
- Extensive "what-if" analysis
- Explanatory statements of material flow phenomena in a stochastic environment

Thus, the research purpose is to add new model approaches to the existing tool-kit of analytical models found in the literature and give explanations for the system's behavior in material flow systems not investigated so far.

## 1.2. Organisation of the Book

The overall structure of the current work is depicted in Figure 1.1.

The Chapters from 1 to 3 intend to familiarize the reader with fundamentals and the motivation behind our work. After an introduction, Chapter 2 presents definitions of probability theory focusing on the discrete time domain. In addition, we give a summary of some basic discrete time queueing models which allow the modeling of one-piece flows in material flow systems. We conclude this chapter presenting the advantages of modeling material flow systems using discrete time analysis. These advantages motivate strongly our research. Afterwards, in Chapter 3, we exemplify a variety of reasons why batch processes occur in material flow systems. A literature review relevant to the current work is provided. We distinguish between batch arrival, batch service, batch building, and queueing network analysis with respect to batch processes. The main part

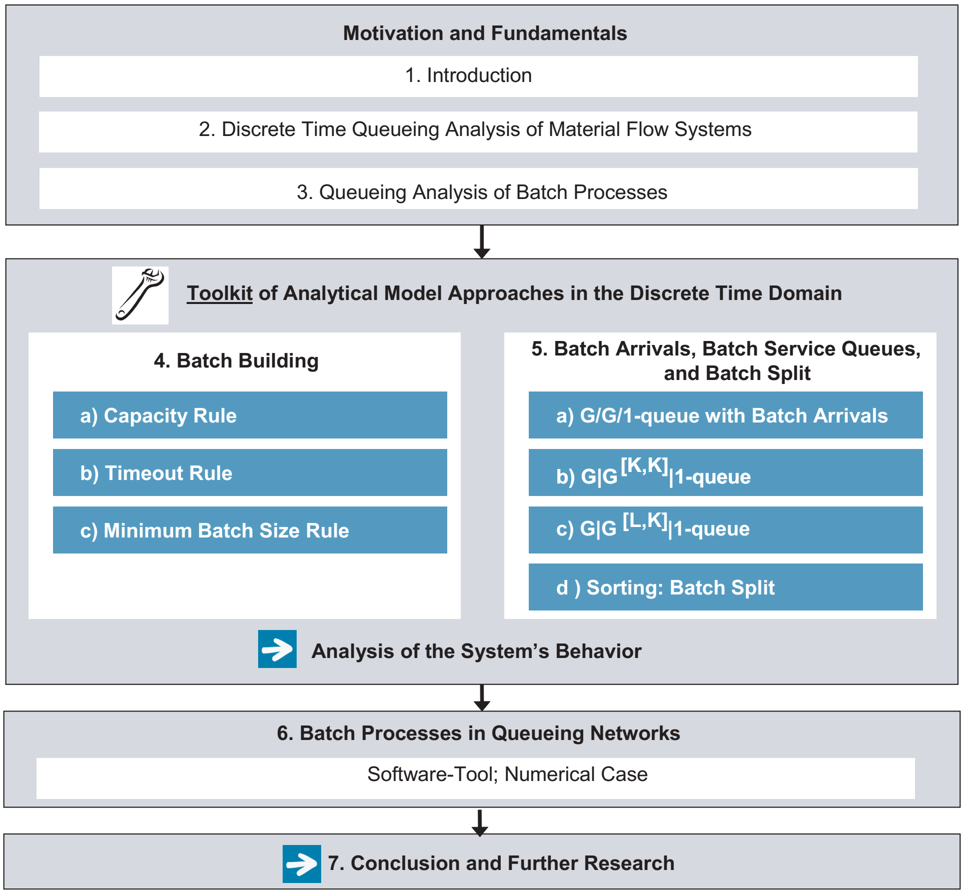


Figure 1.1.: Organisation of the book

of this work are the Chapters 4 and 5. Here, we introduce a comprehensive toolkit of analytical approaches for the analysis of material flow systems in which batch processes are involved. Normally, a toolkit has the property that an user can choose the required tool independent of the other tools which are else provided in the toolkit. Therefore, we construct Chapters 4 and 5 in a way that the sections in which one of the tools is described can be read, understood and used separately from one another. As such, each tool is explained with the given parameters and assumptions, and exemplified briefly with a few real world applications. The proposed tools for the analysis of material flow systems are: Batch building under the capacity rule, batch building under the time-out rule, batch building under the minimum batch size rule, G/G/1-queue with batch arrivals, $G/G^{[K,K]}/1$-queue, $G/G^{[L,K]}/1$-queue, and batch split. Each tool provides methods for a detailed analysis of the waiting and departure process on the basis of discrete distributions. Furthermore, the system's state of the G/G/1-queue with batch arrivals is investigated at the arrival instant and the system's state of the $G/G^{[L,K]}/1$-queue at the departure instant. In addition to the analytical descriptions we point out the system's behavior of each analytical model. We show both, mathematically and numerically, the dependence of performance measures on the input parameters.

Chapter 6 illustrates how the different tools presented in the previous chapters can be taken to model material flow networks. In order to make a network analysis possible we developed a software solution well-suited for the application to practical problems. A tool library supports the modeling of both one-piece and batch flows. Arbitrary material flow networks can be modeled user-friendly via “drag and drop” and can be parameterized via provided windows. This chapter is concluded by presenting a numerical example.

The final chapter of this work provides a concluding discussion of the contribution of this research, summarizes the main results, and also discusses limitations of this approach. Finally, an outlook on further research is given.

# 2. Discrete Time Queueing Analysis of Material Flow Systems

*"What does it profit a man to study such unpleasant phenomena?" The answer, of course, is that through understanding we gain compassion, and it is exactly this which we need since people will be waiting in longer and longer queues as civilization progresses, and we must find ways to tolerate these unpleasant situations.*

Leonard Kleinrock about queueing theory

This chapter familiarizes the reader with fundamentals of discrete time queueing analysis and presents advantages for the evaluation of material flow systems using discrete time queueing analysis.

Section 2.1.1 gives a brief overview of basic definitions of probability theory, which we will use in the approaches presented later.

In Section 2.1.2 the issue of renewal processes in the discrete time domain is addressed. We show how the residual lifetime distribution of a renewal process can be determined, playing an important role in our subsequent analysis.

Section 2.1.3 presents a literature review concerning queueing models in the discrete time domain. We choose models which are elementary for the analysis of material flow systems. Batch processes are not considered. Understanding these models helps the reader with the study of our analysis of batch processes in Chapters 4 and 5.

Finally, in Section 2.2 we discuss the main advantages of discrete time queueing analysis for modeling material flow systems, which serves as the main motivation of our work to find new analytical approaches for a better analysis. Such advantages are exemplified in comparison with queueing methods in the continuous time domain and simulation.

## 2.1. Basic Definitions of Probability Theory in Discrete Time Domain

It is intended to familiarize the reader with the notations used and provide some basic relations of probability theory, focusing on the discrete time domain. However, it is assumed that the reader is familiar with the basic properties and methods of probability and queueing theory. For a detailed study of probability and queueing theory we refer to Feller (1968) and Kleinrock (1975). We recommend the book of Tran-Gia (1996) for an introduction to the analysis of queueing systems in the discrete time domain.

### 2.1.1. Definitions

In our analysis we assume that time is discrete. This means that we observe our system at spaced time instants sequentially numbered by 0,1,2,.... The time period between two subsequent time instants is called the time unit length and is assumed to be constant. Hence, events in a discrete time domain will only be recorded at time instants which are multiples of the time unit length.

In our analysis events are described by a discrete random variable. Given a discrete random variable $X$, we denote its distribution which is also called probability mass function (pmf) by

$$P(X = i) = x_i \quad \forall i = 0,1,...,x_{max}. \tag{2.1}$$

$P$ denotes a probability measure and its possible range of values is from zero to a finite bound $x_{max}$. We assume a finite value range $x_{max}$ since this is in accordance with real applications. The distribution function of $X$, which is called cumulative distribution function abbreviated with CDF, is given by

$$P(X \leq i) = \sum_{j=0}^{i} x_j \quad \forall i = 0,1,...,x_{max}. \tag{2.2}$$

Subsequently, when we use the notion distribution, we refer to the pmf. Several important parameters can be derived from the distribution of a discrete random variable. We get the mean value of $X$ by

$$E(X) = \sum_{i=0}^{x_{max}} i \cdot x_i. \tag{2.3}$$

The $n$th moment of $X$ is defined as

$$E(X^n) = \sum_{i=0}^{x_{max}} i^n \cdot x_i. \tag{2.4}$$

The second central moment of $X$ is referred to as the variance and it is obtained by

$$VAR(X) = E([X - E(X)]^2) = E(X^2) - E(X)^2. \tag{2.5}$$

The squared coefficient of variation (scv) is denoted by $c_X^2$ as a normalized measure of statistical dispersion and is defined as

$$c_X^2 = \frac{VAR(X)}{E(X)^2}. \tag{2.6}$$

We use $c_X^2$ in our analysis to measure the process variability. Processes with low values of $c_X^2$ indicate stable processes. Otherwise, processes with high values of $c_X^2$ indicates unstable processes.

Managers of material flow systems are often interested to know if a process can be performed within a given time period with a given probability. If the distribution of a process is known, this can be indicated by the appropriate quantile. The $u\%$-quantile of a discrete distribution, denoted by $\sigma_u$, gives then the value at which the CDF exceeds $u$. Therefore, we define

$$\sigma_u \Leftrightarrow \sum_{j=0}^{\sigma_u} x_j \geq u \wedge \sum_{j=0}^{\sigma_u - 1} x_j < u. \tag{2.7}$$

The distribution of the sum of two independent nonnegative random variables $X$ and $Y$ is called the convolution of their distributions and can be computed by

$$z_i = \sum_{j=0}^{i} x_j \cdot y_{i-j} = x_i \otimes y_i \qquad i = 0,1,\ldots,z_{max}, \tag{2.8}$$

where $\otimes$ is defined as the convolution operator. Furthermore, we get the difference of two independent nonnegative random variables $X$ and $Y$ by

$$z_i = \sum_{j=0}^{\infty} x_{i+j} \cdot y_j = x_i \otimes -y_i \qquad i = 0,1,\ldots,x_{max}. \tag{2.9}$$

This operation is called the negative convolution.

In Chapters 4 and 5 conditional probabilities are frequently needed for our analysis. This is a powerful tool which eases the description of stochastic processes. Let $P(A)$ denote the probability that the event $A$ occurs, where $P(A)$ is a real number in the range of $0 \leq P(A) \leq 1$. The probability of the event $A$ under the condition that the event $B$ has happened is denoted by $P(A|B)$ and is defined as

$$P(A|B) = \frac{P(A \cap B)}{P(B)}. \tag{2.10}$$

Given is the sequence of events $\{A_i\}$. If $P(A_i)$ and the conditional probabilities $P(B|A_i)$ are known, $P(B)$ can be calculated by the law of total probability:

$$P(B) = \sum_i P(A_i,B) = \sum_i P(B|A_i) \cdot P(A_i). \tag{2.11}$$

Equation (2.10) and (2.11) lead to Bayes theorem written as

$$P(A_i|B) = \frac{P(B|A_i) \cdot P(A_i)}{P(B)} = \frac{P(B|A_i) \cdot P(A_i)}{\sum_i P(B|A_i) \cdot P(A_i)}. \tag{2.12}$$

### 2.1.2. Discrete Time Renewal Process

In the following, we define the renewal process in the discrete time domain. In addition, we derive the residual lifetime distribution which we will need in our further work.

Given is a sequence of events on a discrete time axis. The time interval between event $n$ and $n-1$ is described by the random variable $X^n$ and its distribution is given by $x_i^n$,

$i = 0,1,2,\ldots,x_{max}$. We define this sequence of events as a discrete time renewal process if the length of all intervals are independent from each other and identically distributed (see Tran-Gia (1996)). It follows that

$$x_i^n = x_i \quad \forall\ n = 0,1,2,\ldots \text{ and } \forall\ i = 0,1,2,\ldots. \tag{2.13}$$

Since $X^n$ is independent and identically distributed (iid), each event marks a renewal point. The process is reset at the renewal point and the time interval to the next event is described by $x_i$.

Let us assume that a renewal process is observed at an arbitrary time instant $t^*$. This arbitrary time instant is equally distributed on every possible time instant on the underlying time axis. The time interval from $t^*$ to the succeeding event is defined as the residual life time, denoted by $R$, and the time interval from $t^*$ to the preceding event is defined as the age, denoted by $U$. The residual lifetime and the age of a discrete time renewal process are depicted in Figure 2.1.

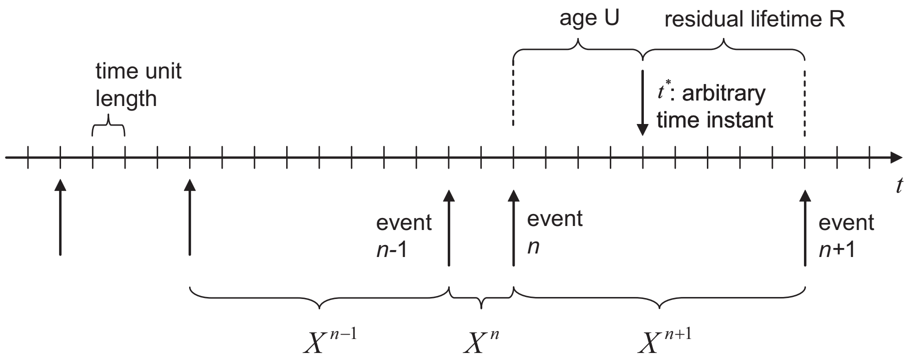


Figure 2.1.: Discrete time renewal process

Since time is assumed to be discrete, we have to distinguish if the arbitrary observation instant $t^*$ lies immediately before or immediately after discrete time instants. First, let us assume that $t^*$ lies immediately before discrete time instants. If an event takes place at the observation instant $t^*$, the occurrence of this event is observed and the residual lifetime is zero. In this case, the age is the time period from the preceding event to $t^*$. Secondly, if we assume that $t^*$ lies immediately after discrete time instants, the occurrence of this event is not observed and the residual lifetime is the time period until the occurrence of the succeeding event. In this case, the age is zero.

If $t^*$ lies immediately before discrete time instants, we conclude that the value range of $R$ is from 0 to $x_{max} - 1$ and that of $U$ is from 1 to $x_{max}$. If $t^*$ lies immediately after discrete time instants, $R$ is defined from 1 to $x_{max}$ and $U$ from 0 to $x_{max} - 1$. In the following, we derive the distribution of the residual lifetime and the age. Since the age has the same statistical properties as the residual lifetime, the age is identical to the residual lifetime except that the value range is shifted by one time unit length. Therefore, only the derivation of the residual lifetime distribution is required.

**Observation Immediately Before Discrete Time Instants**

We assume that $t^*$ lies immediately before discrete time instants. The residual lifetime distribution is denoted by $r_i$ and its derivation can be found in Tran-Gia (1996). It follows that

$$r_i = \sum_{h=i+1}^{x_{max}} P(R = i|Q = h)P(Q = h), \tag{2.14}$$

where $P(Q = h)$ is the probability that $t^*$ lies in an interval of length $h$.

$P(Q = h)$ is proportional to $x_h$. Furthermore, $P(Q = h)$ is proportional to the interval length $h$ because it is more probable to observe a long interval than a short one. Thus we get

$$P(Q = h) = C \cdot h \cdot x_h, \tag{2.15}$$

where $C$ is a constant. Since the sum of all probability has to be one, we can determine $C$:

$$1 = \sum_{h=0}^{x_{max}} P(Q = h) = C \sum_{h=0}^{x_{max}} h \cdot x_h = C \cdot E(X) \Rightarrow C = \frac{1}{E(X)}. \tag{2.16}$$

If the observed interval $X$ is of length $h$, there are $h$ possible observation points which can all be met with the same probability $1/h$. It follows that

$$P(R = i|Q = h) = \frac{1}{h} \quad \forall i = 0,1, \ldots ,k - 1. \tag{2.17}$$

Combining Equations from (2.14) to (2.17) yields

$$r_i = \frac{1}{E(X)} \sum_{h=i+1}^{x_{max}} x_h = \frac{1}{E(X)}(1 - \sum_{h=0}^{i} x_h) \quad \forall i = 0, \ldots ,x_{max} - 1. \tag{2.18}$$

The derivation of the distribution of the age is analogous. The value range is shifted by one time unit length. We get

$$u_i = \frac{1}{E(X)}(1 - \sum_{h=0}^{i-1} x_h) \quad \forall i = 1, \ldots ,x_{max}, \tag{2.19}$$

and we can write

$$r_i = u_{i+1} \quad \forall i = 0, \ldots ,x_{max} - 1. \tag{2.20}$$

**Observation Immediately After Discrete Time Instants**

We assume that $t^*$ lies immediately after discrete time instants. The derivation of $r_i$ and $u_i$ is analogous to the case that $t^*$ lies immediately before discrete time instants. Now, the value range of $R$ is from 1 to $x_{max}$ and that of $U$ is from 0 to $x_{max} - 1$. It yields

$$r_i = \frac{1}{E(X)}(1 - \sum_{h=0}^{i-1} a_h) \quad \forall i = 1, \ldots ,x_{max} \tag{2.21}$$

and

$$r_i = u_{i-1} \quad \forall i = 1, \ldots ,x_{max}. \tag{2.22}$$

### 2.1.3. Material Flow Modeling

In this section we summarize briefly some basic queueing models in the discrete time domain, which allow the modeling of simple problems in material flow systems. These models are restricted to an one-piece flow.

As in the famous Queueing Network Analyzer of Whitt (1983) for the analysis of general open queueing networks in the continuous time domain, we have to model the following basic operations in the discrete time domain as well:

- The service operation by means of the G/G/1-queue[1]; the interarrival and the service time are iid
- The split operation in order to split an incoming stochastic stream into two or more outgoing streams
- The merge of independent stochastic streams

In Figure 2.2 the named basic operations are depicted.

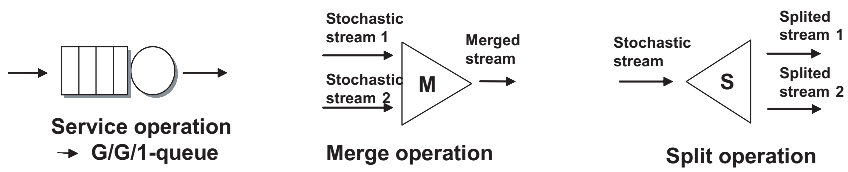


Figure 2.2.: Basic operations for modeling material flow networks in the discrete time domain

As by Whitt it is assumed that the nodes of a material flow network are treated as stochastically independent. In order to connect the nodes, the departure process of each of the named model elements has to be known. Furthermore, for a performance evaluation the waiting times of the G/G/1-queues has to be determined. Thus, Konheim (1975), Ackroyd (1980), Grassmann and Jain (1989), and Tran-Gia (1996) present analytical approaches to estimate the waiting time distribution of the discrete time G/G/1-queue. Grassmann and Jain (1989) give a numerical method which is based on the Wiener-Hopf factorization of the underlying random walk. They compare their method with related methods suggested in the literature of queues and show that their method performs very

[1]The Kendall's notation is widely used to classify elementary queueing systems:

$$A/B/m \text{ - queueing discipline,}$$

where $A$ indicates the interarrival time distribution, $B$ the service time distribution and $m$ the number of servers. E.g. $M$ denotes exponential (Markov process), $D$ deterministic (Dirac), $G$ general, $E_k$ $k$-Erlang and $H$ hyperexponential. Implicit with this notation is the assumption that the interarrival times and the service times are independent and that the random variables of each sequence are identically distributed. The queueing discipline determines which customer is selected from the queue for service when a server becomes idle. Generally, if the queueing discipline is not indicated, it is assumed that the queueing discipline is First-Come-First-Served.

well and is often faster by several orders of magnitude. Grassmann and Jain develop three algorithms for the calculation of the waiting time distribution. Algorithm 3 converges faster than algorithm 1 and 2. However, they have no proof of convergence for algorithm 2 and 3. In contrast, they show that algorithm 1, which converges slower than algorithm 2, always converges.

Once the waiting time distribution is known, the interdeparture time distribution can be determined (see Grassmann and Jain (1988)). It has to be distinguished between two cases. First, if customer $n+1$ arrives at the queueing system and has to wait, the interdeparture time between customer $n$ and $n+1$ will be the service time. However, if customer $n+1$ arrives at an empty system and initiates a busy period, his waiting time is zero and the interdeparture time between customer $n$ and $n+1$ will be the sum of the service and the idle time.

Furmans (2004a) presents a model for the stochastic split of material- and information flows in the discrete time domain, which is based on the ordinary Markovian split in the literature (see Whitt (1983)). In the current work this approach is extended in order to model the split of batch arrivals (see Section 5.5).

There is no exact method to calculate the distribution of a merged stochastic stream. Again, Furmans (2004a) introduces an approximation method which assumes that the merged stream is again a renewal process.

The three briefly presented basic operations (see Figure 2.2) allow a rough and a fast analysis of material flow systems. With the aim to increase the level of detail regarding modeling of material flow processes, this work introduces a variety of new models for the description of batch processes.

## 2.2. Advantages of Modeling Material Flow Systems in Discrete Time

Queueing analysis for general networks in the continuous time scale is well studied. An enormous amount of literature with regard to this topic exists. Some of them which provide a comprehensive insight and an overview about queueing theory, are the works of Kleinrock (1975), Gnedenko and König (1983), Wolff (1989), and Buzacott and Shanthikumar (1993). There are a variety of literature modeling material flow systems by means of queueing systems in the continuous time scale as well, as such Greiling (1997), Rall (1998), and Furmans (2000).

Since the transmission of data packets in communication networks, such as in ATM (Asynchronous Transfer Mode) networks, occurs in slots, scientists in telecommunication science have been investigating discrete time queueing models intensively since the 1980s. Among them are Ackroyd (1980), Bruneel and Kim (1992), and Hübner and Tran-Gia (1995). Discrete time queueing models are well suited to describe material flow systems as well. As a motivation to find appropriate solutions for problems in this field, we shall subsequently discuss advantages of analyzing material flow systems using discrete time queueing models.

**Accuracy**

An analysis of material flow systems by means of general queueing systems in a continuous time domain is based on the description of stochastic processes by the first two moments (see Shanthikumar and Buzacott (1981), Whitt (1983), Whitt (1993), and Hopp et al. (2002)). Concerning the G/G/1-queue, the first two moments of the interarrival and service time are used to calculate the mean waiting time and the first two moments of the interdeparture time using approximations (see Whitt (1993) and Bolch et al. (1998)).

In contrast, the before mentioned approach of Grassmann and Jain (1989) for the determination of the waiting time distribution of the discrete time G/G/1-queue is exact within an $\epsilon$-neighborhood.

Since the queueing systems in our work are analyzed on the basis of discrete distributions, we have the possibility to explain the influence of further standardized moments like skewness, which is a measure for the asymmetry, and kurtosis, which is a measure for the "peakedness". Moments higher than the second influence the accuracy of computation. This fact can be illustrated by the subsequent numerical examples analyzing the G/G/1-queue.

In each of the four examples given in Table 2.1, we varied the service time distribution, but kept its mean value and scv constant. We calculated the mean waiting time using the approach of Grassmann and Jain (1989), which is exact within an $\epsilon$-neighborhood. As shown in Table 2.1, the mean waiting time $E(W)$ is obviously different for each of the five service time distributions. The maximum relative deviation is denoted by $\Delta_{max}$. It yields $\Delta_{max} = 0{,}073$ for the first example. This result demonstrates the fact that more information is required for an accurate determination of $E(W)$ than the first two moments of the interarrival and service time distribution.

However, by an analysis with standard 2-parameters-approximation methods for the G/G/1-queue, in which general processes are described by the first two moments, we get the same result for each of the five service time distributions. Thus, we calculated $E(W)$ by means of the well known G/G/1-approximation formulas of Marchal (1976), Reiser and Kobayashi (1974), Buzacott and Shanthikumar (1993), Krämer and Langenbach-Belz (1976), and Page (1972). The relative deviation of each of these approaches compared to the solution of Grassmann and Jain is illustrated in Table 2.1. The minimum deviation is marked bold.

The accuracy of the 2-parameter approximations is dependent on the scv of the interarrival and service time distribution as well as the utilization (see Shanthikumar and Buzacott (1980)). Therefore, in some cases one approximation works better than another. Shanthikumar and Buzacott (1980) provide a guide to select the appropriate approximation for different ranges of the scv of the interarrival and service time distribution. Just as Shanthikumar and Buzacott, we observe that the approach by Reiser and Kobayashi (1974) seems to give a high percentage of error.

Comparing the 2-parameter approximations with the results obtained by Grassmann and Jain, we observe that the deviations are remarkably high in some cases. In particular, there is a high percentage of error yielded by all approximations in example 2.

Although the number of experiments we performed is not sufficient in order to statistically quantify the deviation of 2-parameter approximations against discrete time approaches, the data in Table 2.1 shows clearly that the deviations can be remarkable. This fact drives us to develop new discrete time queueing models which describe real world material flow processes more accurately.

We conclude that discrete time queueing analysis is a highly accurate analytical tool with respect to interpret stochastic processes.

### Level of Detail

Material flow systems should be designed in a way such that it guarantees the order fulfillment in a predetermined time with a chosen probability (e.g. 95%). Hence, the distribution of the time for which an order remains in a system, defined as the sojourn time, has to be known in order to determine its quantiles. For example, this is very crucial in warehouses and distribution centers in which the requested shipping times should be met with an acceptable probability, which usually lies between 95% and 99%, depending on order types.

Figure 2.3 exemplifies the distribution of the sojourn time in an arbitrary material flow process. By means of this distribution, we can determine the minimum time length so that the concerned process is finished with a chosen probability. This minimum time length which corresponds to the quantile is marked by a bold line in Figure 2.3. The chosen quantile in this case is the 95%-quantile.

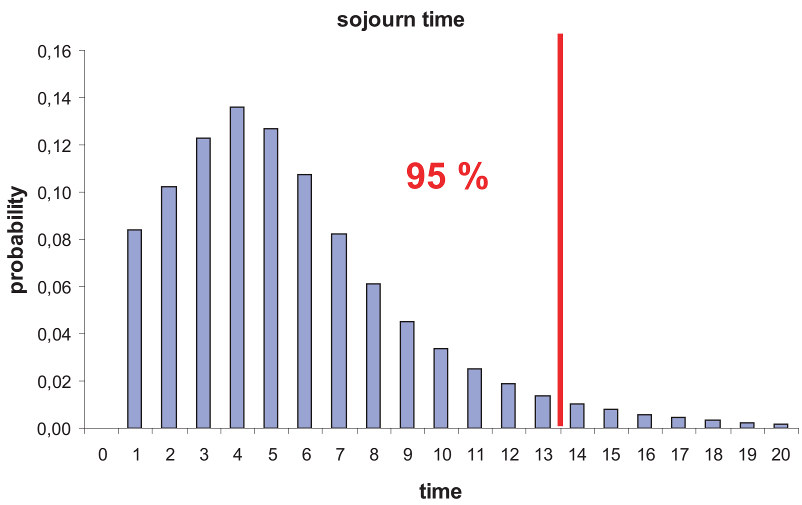


Figure 2.3.: Illustration of the 95%-quantile using an arbitrary sojourn time distribution

Thus, an analysis on the basis of distributions is required. This can be achieved by means of discrete time queueing analysis. An objective of our work is the determination of the waiting time distribution of different queueing models in which batch processes are involved. Since the service time distributions are given, the sojourn time distribution can be determined and with it the probability of an on-time order fulfillment.

We conclude that the analysis of material flows by means of discrete time queueing methods enables a more detailed description of the system's behavior than analytical methods

| Example 1 | | | service time distributions | | | | |
|---|---|---|---|---|---|---|---|
| **Input** | $i$ | $a_i$ | $b_i$ (1) | $b_i$ (2) | $b_i$ (3) | $b_i$ (4) | $b_i$ (5) |
| | 0 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| | 1 | 0.070 | 0.000 | 0.324 | 0.074 | 0.050 | 0.206 |
| | 2 | 0.080 | 0.350 | 0.000 | 0.149 | 0.033 | 0.144 |
| | 3 | 0.110 | 0.175 | 0.000 | 0.315 | 0.660 | 0.104 |
| | 4 | 0.130 | 0.150 | 0.000 | 0.250 | 0.024 | 0.000 |
| | 5 | 0.150 | 0.115 | 0.475 | 0.000 | 0.043 | 0.175 |
| | 6 | 0.140 | 0.100 | 0.201 | 0.000 | 0.011 | 0.371 |
| | 7 | 0.110 | 0.040 | | 0.111 | 0.000 | |
| | 8 | 0.090 | 0.025 | | 0.101 | 0.179 | |
| | 9 | 0.070 | 0.025 | | | | |
| | 10 | 0.040 | 0.020 | | | | |
| | 11 | 0.010 | | | | | |
| | 12 | | | | | | |
| | 13 | | | | | | |
| | 14 | | | | | | |
| | mean value | 5.300 | 3.905 | 3.905 | 3.905 | 3.905 | 3.905 |
| | scv | 0.220 | 0.275 | 0.275 | 0.275 | 0.275 | 0.275 |
| | | utilization | 0.737 | 0.737 | 0.737 | 0.737 | 0.737 |
| **Discrete time** | Grassmann | **E(W)** | **2.243** | **2.079** | **2.230** | **2.266** | **2.121** |
| **approach** | | $\Delta_{max}$ | **0.073** | | | | |
| **2-parameter** | | Marchal | 2.239 | 2.239 | 2.239 | 2.239 | 2.239 |
| **-approximations** | | Δ | **0.002** | 0.077 | **0.004** | **0.012** | 0.056 |
| | | Kobayashi | 1.671 | 1.671 | 1.671 | 1.671 | 1.671 |
| | | Δ | 0.255 | 0.196 | 0.251 | 0.263 | 0.212 |
| | | Shanthikumar | 2.285 | 2.285 | 2.285 | 2.285 | 2.285 |
| | | Δ | 0.018 | 0.099 | 0.024 | 0.008 | 0.077 |
| | | Krämer-Langenbach-Belz | 2.019 | 2.019 | 2.019 | 2.019 | 2.019 |
| | | Δ | 0.100 | **0.029** | 0.095 | 0.109 | **0.048** |
| | | Page | 2.457 | 2.457 | 2.457 | 2.457 | 2.457 |
| | | Δ | 0.095 | 0.182 | 0.102 | 0.084 | 0.158 |

| Example 2 | | | service time distributions | | | | |
|---|---|---|---|---|---|---|---|
| **Input** | $i$ | $a_i$ | $b_i$ (1) | $b_i$ (2) | $b_i$ (3) | $b_i$ (4) | $b_i$ (5) |
| | 0 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| | 1 | 0.200 | 0.000 | 0.324 | 0.074 | 0.050 | 0.206 |
| | 2 | 0.300 | 0.350 | 0.000 | 0.149 | 0.033 | 0.144 |
| | 3 | 0.150 | 0.175 | 0.000 | 0.315 | 0.660 | 0.104 |
| | 4 | 0.050 | 0.150 | 0.000 | 0.250 | 0.024 | 0.000 |
| | 5 | 0.025 | 0.115 | 0.475 | 0.000 | 0.043 | 0.175 |
| | 6 | 0.000 | 0.100 | 0.201 | 0.000 | 0.011 | 0.371 |
| | 7 | 0.000 | 0.040 | | 0.111 | 0.000 | |
| | 8 | 0.000 | 0.025 | | 0.101 | 0.179 | |
| | 9 | 0.000 | 0.025 | | | | |
| | 10 | 0.000 | 0.02 | | | | |
| | 11 | 0.025 | | | | | |
| | 12 | 0.100 | | | | | |
| | 13 | 0.125 | | | | | |
| | 14 | 0.025 | | | | | |
| **Results** | mean value | 5.025 | 3.905 | 3.905 | 3.905 | 3.905 | 3.905 |
| | scv | 0.886 | 0.275 | 0.275 | 0.275 | 0.275 | 0.275 |
| | | utilization | 0.777 | 0.777 | 0.777 | 0.777 | 0.777 |
| **Discrete time** | Grassmann | **E(W)** | **8.587** | **8.687** | **8.593** | **8.572** | **8.652** |
| **approach** | | $\Delta_{max}$ | **0.013** | | | | |
| **2-parameter** | | Marchal | 7.829 | 7.829 | 7.829 | 7.829 | 7.829 |
| **-approximations** | | Δ | 0.088 | 0.099 | 0.089 | 0.087 | 0.095 |
| | | Kobayashi | 6.634 | 6.634 | 6.634 | 6.634 | 6.634 |
| | | Δ | 0.227 | 0.236 | 0.228 | 0.226 | 0.233 |
| | | Shanthikumar | 7.835 | 7.835 | 7.835 | 7.835 | 7.835 |
| | | Δ | 0.088 | 0.098 | 0.088 | 0.086 | 0.094 |
| | | Krämer-Langenbach-Belz | 7.884 | 7.884 | 7.884 | 7.884 | 7.884 |
| | | Δ | **0.082** | **0.092** | **0.083** | **0.080** | **0.089** |
| | | Page | 7.864 | 7.864 | 7.864 | 7.864 | 7.864 |
| | | Δ | 0.084 | 0.095 | 0.085 | 0.083 | 0.091 |

| Example 3 | | | service time distributions | | | | |
|---|---|---|---|---|---|---|---|
| **Input** | $i$ | $a_i$ | $b_i$ (1) | $b_i$ (2) | $b_i$ (3) | $b_i$ (4) | $b_i$ (5) |
| | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | 1 | 0.125 | 0.000 | 0.324 | 0.074 | 0.050 | 0.206 |
| | 2 | 0.125 | 0.350 | 0.000 | 0.149 | 0.033 | 0.144 |
| | 3 | 0.125 | 0.175 | 0.000 | 0.315 | 0.660 | 0.104 |
| | 4 | 0.125 | 0.150 | 0.000 | 0.250 | 0.024 | 0.000 |
| | 5 | 0.125 | 0.115 | 0.475 | 0.000 | 0.043 | 0.175 |
| | 6 | 0.125 | 0.100 | 0.201 | 0.000 | 0.011 | 0.371 |
| | 7 | 0.125 | 0.040 | | 0.111 | 0.000 | 0.000 |
| | 8 | 0.125 | 0.025 | | 0.101 | 0.179 | 0.000 |
| | 9 | | 0.025 | | | | |
| | 10 | | 0.020 | | | | |
| | mean value | 4.500 | 3.905 | 3.905 | 3.905 | 3.905 | 3.905 |
| | scv | 0.259 | 0.275 | 0.275 | 0.275 | 0.275 | 0.275 |
| | | utilization | 0.868 | 0.868 | 0.868 | 0.868 | 0.868 |
| **Discrete time** | Grassmann | **E(W)** | **6.435** | **6.241** | **6.416** | **6.462** | **6.285** |
| **approach** | | $\Delta_{max}$ | **0.030** | | | | |
| **2-parameter** | | Marchal | 6.307 | 6.307 | 6.307 | 6.307 | 6.307 |
| **-approximations** | | Δ | 0.020 | **0.011** | 0.017 | 0.024 | **0.004** |
| | | Kobayashi | 5.595 | 5.595 | 5.595 | 5.595 | 5.595 |
| | | Δ | 0.131 | 0.104 | 0.128 | 0.134 | 0.110 |
| | | Shanthikumar | 6.337 | 6.337 | 6.337 | 6.337 | 6.337 |
| | | Δ | **0.015** | 0.015 | **0.012** | **0.019** | 0.008 |
| | | Krämer-Langenbach-Belz | 6.162 | 6.162 | 6.162 | 6.162 | 6.162 |
| | | Δ | 0.042 | 0.013 | 0.040 | 0.047 | 0.020 |
| | | Page | 6.588 | 6.588 | 6.588 | 6.588 | 6.588 |
| | | Δ | 0.024 | 0.056 | 0.027 | **0.019** | 0.048 |

| Example 4 | | | service time distributions | | | | |
|---|---|---|---|---|---|---|---|
| **Input** | $i$ | $a_i$ | $b_i$ (1) | $b_i$ (2) | $b_i$ (3) | $b_i$ (4) | $b_i$ (5) |
| | 0 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0 |
| | 1 | 0.000 | 0.000 | 0.324 | 0.074 | 0.050 | 0.206 |
| | 2 | 0.000 | 0.350 | 0.000 | 0.149 | 0.033 | 0.144 |
| | 3 | 0.250 | 0.175 | 0.000 | 0.315 | 0.660 | 0.104 |
| | 4 | 0.400 | 0.150 | 0.000 | 0.250 | 0.024 | 0.000 |
| | 5 | 0.200 | 0.115 | 0.475 | 0.000 | 0.043 | 0.175 |
| | 6 | 0.100 | 0.100 | 0.201 | 0.000 | 0.011 | 0.371 |
| | 7 | 0.050 | 0.040 | | 0.111 | 0.000 | |
| | 8 | 0.000 | 0.025 | | 0.101 | 0.179 | |
| | 9 | 0.000 | 0.025 | | | | |
| | 10 | 0.000 | 0.020 | | | | |
| **Results** | mean value | 4.300 | 3.905 | 3.905 | 3.905 | 3.905 | 3.905 |
| | variability | 0.065 | 0.275 | 0.275 | 0.275 | 0.275 | 0.275 |
| | | utilization | 0.908 | 0.737 | 0.737 | 0.737 | 0.737 |
| **Discrete time** | Grassmann | **E(W)** | **5.781** | **5.469** | **5.739** | **5.824** | **5.565** |
| **approach** | | $\Delta_{max}$ | **0.054** | | | | |
| **2-parameter** | | Marchal | 5.854 | 5.854 | 5.854 | 5.854 | 5.854 |
| **-approximations** | | Δ | **0.013** | **0.070** | **0.020** | **0.005** | **0.052** |
| | | Kobayashi | 5.323 | 5.323 | 5.323 | 5.323 | 5.323 |
| | | Δ | 0.079 | 0.027 | 0.072 | 0.086 | 0.043 |
| | | Shanthikumar | 5.884 | 5.884 | 5.884 | 5.884 | 5.884 |
| | | Δ | 0.018 | 0.076 | 0.025 | 0.010 | 0.057 |
| | | Krämer-Langenbach-Belz | 5.518 | 5.518 | 5.518 | 5.518 | 5.518 |
| | | Δ | 0.045 | 0.009 | 0.038 | 0.053 | 0.009 |
| | | Page | 6.239 | 6.239 | 6.239 | 6.239 | 6.239 |
| | | Δ | 0.079 | 0.141 | 0.087 | 0.071 | 0.121 |

Table 2.1.: Analysis of the G/G/1-queue using standard 2-parameter approximations compared to discrete time methods

in the continuous time domain.

**Processes in Real Material Flow Systems are "Discrete"**

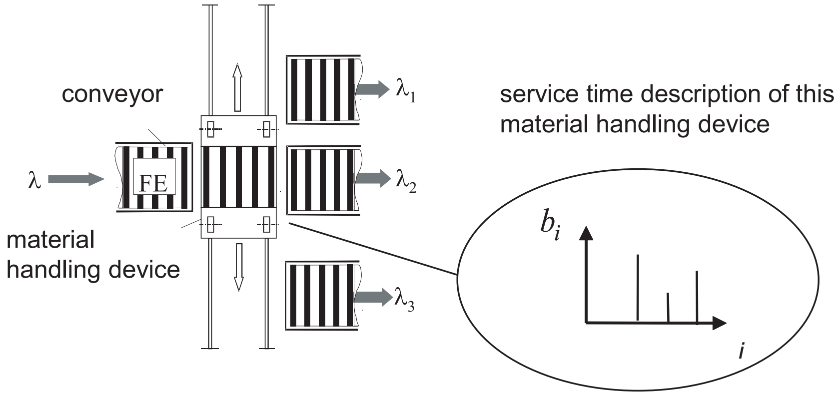


Figure 2.4.: Process description in material flow systems by means of discrete distributions; example: Service time description of a material handling device

The assumption that time is not continuous but discrete is essentially not a restriction for modeling material flow systems. For example, the travel time of a material handling device can only adopt a few time values, which can be very well described by a discrete distribution. Figure 2.4 demonstrates that, given the travel times for each possible direction and the probability choosing a direction, the service time distribution can be easily derived.

In contrast, the modeling of stochastic processes in the continuous time scale requires the existence of a theoretical distribution function or the description by their moments. The derivation of the theoretical function is time consuming and this function describes the real stochastic process with imprecision. Especially, the description of multi-modal functions is difficult. On the other hand, using discrete time queueing analysis arbitrary distribution functions, in the way they exist after an as-is analysis for a material flow system, can be used. Generally, the results of an as-is analysis of a material flow system are available in the form of histograms. The normalization of these histograms yields discrete distributions functions which are the input for the analysis.

We conclude that a discrete time queueing analysis represents real processes with a high degree of accuracy, since data ascertained in an as-is analysis is generally on a discrete basis. Furthermore, low effort is required in data acquisition.

**Efficiency**

In industrial practice, the analysis of material flow systems is often addressed via simulation. Simulation is a very powerful tool with an enormous degree of freedom regarding modeling. Simulation models can be used for any desired level of detail. However, simulation is very time consuming and therefore expensive. It requires a lot of time for modeling, validation and performing experiments. Numerous simulation runs are required for a single experiment in order to achieve correct results within a given statistical range.

Analytical approaches are well-suited to support the long-range planning of material flow systems in an early planning stage, in which the capacities are searched to minimize facility costs under the condition of cycle time targets. In this planning stage, detailed input data[2] is not available and often based on rough approximations. Therefore, a rough and extensive “what-if” analysis is required (see Hopp et al. (2002)). For this kind of analysis, analytical models are well suited since they require considerably less time for conduction the numerical experiments.

For example, to calculate the waiting, interdeparture and idle time distribution of a G/G/1-queue with batch arrivals (see Section 5.1), it takes a few milliseconds for a problem of reasonable size using the analytical approach and a few minutes for the corresponding simulation.[3]

We conclude that analytical methods are much more time efficient than simulation.

### Classification of Queueing Analysis in Discrete Time

We have so far explained the advantages of queueing analysis in discrete time. However, some of its limitations have also to be mentioned. We take for granted that the user considers both advantages and limitations to choose the right tool for his analysis.

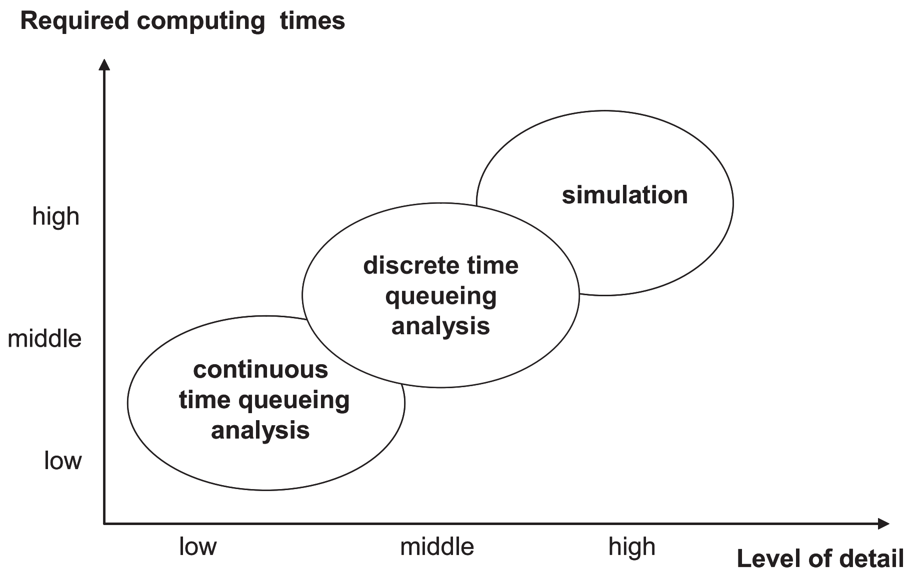


Figure 2.5.: Classification of queueing analysis in discrete time with regard to queueing analysis in continuous time and simulation

The level of detail and the accuracy of queueing analysis in discrete time is limited compared to simulation. The user is restricted to analytical models which have been de-

[2]Input data such as demand, processing times, quality rates, failure rates etc.

[3]Both the analytical calculations and the simulation runs are performed on a computer with an Intel Centrino processor. We used the simulation tool emPlant.

veloped so far. The input data has to be prepared such that it can be used as parameters for discrete time queueing models. For example, the breakdown behavior of a machine has to be included in the service time distribution. Furthermore, the models are based on assumptions in order to reduce the difficulty of developing. If an exact analytical solution cannot be found or requires too much computing time, analytical approximations are used. Thus, by the use of analytical approaches we get a simplified model of reality and the obtained results will deviate from the results which are observed in real processes. In contrast, using simulation highly detailed models can be developed which are quite close to the real world process behavior. Recall that this is very time consuming and the models are inclined to be quite large and intransparent. The requirements concerning data quality and availability increase and the model validation becomes more difficult.

If we compare discrete with continuous time queueing models, we recognize that continuous time queueing models require generally less computing time. Especially, 2-parameters approximations are easy to handle. Using discrete time queuing models, a higher level of detail and accuracy is often obtained at the expense of longer computing times. In several cases of our subsequent analysis, we propose combinatorial solutions to reach the renewal point in a renewal process. If the vectors representing discrete distributions become large, the number of possibilities to get from one renewal point to the next increase and therefore the time for computing.

Finally, if we take the level of detail and the required computing time as a criterion for the classification of queueing analysis in discrete time with regards to queueing analysis in continuous time and simulation, it results in the following classification as illustrated in Figure 2.5[4]. The modeler of material flow systems has to carefully consider the level of detail and accuracy which is required for his analysis, before he chooses his tool. He has to asses the advantages and limitations of each approach. Often, it is worthwhile to use both analytical methods and simulation for a systems analysis, since analytical methods perform an essential role in validating subsystems and special cases of a simulation model.

## 2.3. Chapter Conclusion

This chapter gave a brief introduction in probability theory focusing on the discrete time domain. We explained how the residual lifetime distribution of a discrete time renewal process can be derived, both for an observation immediately before and after discrete time instants.

We introduced three basic discrete time queueing models which enables the modeling of simple problems in material flow systems. Generally, material flows can be decomposed in three basic operations, namely the service, merge and split operation. For the analysis of these operations we referred to known approaches from the literature. However, these basic models are restricted to one-piece flows. In this work we will relax this assumption. Thus, we will present models for batch processes in order to increase the level of detail regarding modeling.

[4] Rall (1998) presents a similar figure, where discrete time queueing analysis is not considered.

Section 2.2 emphasized advantages of modeling material flow systems in discrete time. An analysis in the discrete time domain delivers advantages regarding accuracy, level of detail and efficiency. In addition, we observed that processes in material flow systems are often of a discrete time nature. Furthermore, we performed a numerical analysis of the G/G/1-queue in which we compared results using standard 2-parameter approximations compared to discrete time methods. We identified that the percentage of error of the standard 2-parameter approximations can be remarkable high and that moments higher than the second can influence the results significantly. We concluded this chapter discussing limitations of discrete time queuing analysis. This led to a classification of queueing analysis in discrete time.

# 3. Queueing Analysis of Batch Processes

*The larger the island of knowledge, the longer the shoreline of wonder.*

Ralph W. Sockman

In Section 3.1 we explain the reasons to build batches in material flow and production systems. We emphasize that batch building problems arise not only in these systems, but also in many other systems, like transportation and information systems. Thus, our analytical approaches are well suited to describe a wide variety of real world problems. In Section 3.2 we present a review of the literature which is done in the field of the analysis of batch processes in a stochastic environment and which is relevant to our work. In the literature review we distinguish between the analysis of batch arrival, batch service and batch building processes, and the analysis of queueing networks with respect to batch processes.

## 3.1. Batch Processes in Material Flow Systems

Many operations in material flow systems[1] are done in batches. The reason for building batches is clear and evident: capacity. In most cases, it is more efficient to transport or to process a batch of entities instead of transporting or processing a single entity. Often, the operation costs for processing or transporting one entity or a batch of entities are the same. Therefore, the operation costs per entity decrease with an increasing batch size.

In contrast, the batch building process causes a waiting process which results in an increase of inventory. Furthermore, the cycle time to produce one product type increases with increasing batch sizes. For example, the use of small batch sizes makes it possible to produce all different product types within one day, in comparison to several days required for the production involving huge batch sizes. Subsequently, the ability to react rapidly on demand changes decreases with increasing batch sizes. Thus, in material flow systems originates always the problem to determine the optimum batch size.

Let us exemplify the necessity of building batches in more detail by presenting four basic reasons which are related to material flow and production systems:

- **Transport/Handling**: To reduce the handling effort, a specific number of material units are gathered to one transport unit. The maximum batch size is given by

[1] Generally, if we talk about material flow systems, we mean first and foremost the material flow in a production or service facility

the capacity of the transport carrier. Examples of transport carriers are pallets, wire cages, plastic bins, etc.

- **Setup**: It is often necessary for industrial operations to set up machines at the beginning of a batch service. Setup operations are tool changes, cleaning operations or programming setups. The larger the batch size, the less setup operations are required and less capacity is lost. But in the same step, the inventory increases and the ability of the company to react to the changes in demand decreases.

- **Batch service**: There are different types of production operations in the industrial production environment, in which a certain amount of units is processed simultaneously. Examples are heating treatments in an oven or surface treatments of semiconductors in chemical washings.

- **Order Pooling**: The picking operation is a main task in warehouses and distribution centers. The picking time can be reduced significantly by a thoughtful grouping of the picking orders. The time per pick has to be minimized. This includes minimal ways for the picker through the picking zone and the pooling of same articles which belong to different customer orders.

Due to the fact that there are batch building processes because of fundamental economic interests, it is worthwhile to develop analytical methods in order to determine performance measures of high precision for material flow systems.

Moreover, we find batch processes in other systems besides material flow systems. The motivation to build batches remains the same, capacity. A further example are transportation systems. Goods and public passengers transportation require a specific amount of goods/passengers to be collected before the transport is released. This measure reduces the cost per good/passenger. Equally, we can model a traffic light which controls the traffic flow at a crossing by a batch server system. In this case, the batch service is rendered in terms of the ability of a group of cars to pass the phase of green light.

We know that there is limited capacity in information and communication systems. The memory, the processor performance, and the capacity of the communication network have all their physical limitations. Thus, it is common in such systems to collect packages of data before it is sent or processed.

Finally, the daily life is abound in examples of batch processes. For example, people arrive in groups for visiting a museum and the museum guide waits until a quota of persons to arrive before he starts his guiding tour. Furthermore, letters are delivered in batches at the post office for sorting and the subsequent distribution. Also, an elevator serves a “batch” of persons. The list is endless and too often the queues as well.

Later, when we introduce our analytical approaches for different batch processes, we will again describe briefly some appropriate application possibilities at the beginning of each model description. Before we start with the introduction of analytical models, we give a review of the literature which is done in this field.

## 3.2. Literature Review

There is a long tradition of studying batch processes, developed for the analysis of a great variety of applications. Bailey (1954) is the first who studies a queueing model for batch processes and its application to practical problems. Since this work, there is a large, scattered literature about batch queues. In most of the cases, classical transform techniques[2] are used to find the transforms of queue length and waiting time distribution. Chaudhry and Templeton (1983) were motivated to publish a book about batch queues, in which they summarize, synthesize and, in some instances, extend the major topics. They study different models of batch arrival, batch service, and multichannel batch queues in detail. Before analytical models about batch queues are presented, they give a comprehensive introduction to the basic techniques used in their book. For batch arrival queues they study methods with fixed and random batch sizes. Furthermore, the authors consider a set of batch service models in which the batch size is fixed or random or controlled by a service strategy. Finally, they discuss relations among different queueing systems.

Subsequently, we give a review over basic work studying batch queues. We distinguish between batch arrival, batch service and batch building processes. Finally, we refer to contributions about the analysis of queueing networks regarding batch processes.

### 3.2.1. Batch Arrival

The study of batch arrival queues begins with the work of Gaver (1959). He investigates the $\mathrm{M}^X$/G/1-queue in which a batch of random size $X$ enters the system following a Poisson process. The individual customers are processed by a single server system whose service time is generally distributed. He uses the embedded Markov chain technique which is introduced by Kendall (1951), (1953), and the renewal theory (see Feller (1968)) to discuss the waiting process of an arriving customer in the continuous time domain. The mean waiting time $E(W)$ and the variance of the waiting time $VAR(W)$ is derived from the transform of the waiting time. Furthermore, he presents the transform of the busy period duration and derives an expression for the number of departures during the busy period. The work of Gaver is completed by the contributions of Burke (1975) and Chaudhry (1979) who study the waiting time distribution and the distribution of the number of customers in the system at an arbitrary time instant. Ommeren (1990) presents approximations for the waiting time probabilities of individual customers in a $\mathrm{M}^X$/G/1-queue, which are easier to handle than exact solutions.

Brière and Chaudhry (1987) deliver numerical computations for the batch arrival queueing model $\mathrm{GI}^X$/M/1. They introduce an approach to calculate both the moments and the steady state distribution of the number of customers in the system at the customer arrival instant.

Chaudhry and Gupta (1997) introduce an approach in order to solve the discrete time $\mathrm{GI}^X$/Geom/1-queue with batch arrivals. They calculate the queue length and waiting

[2] It is common in queueing theory to calculate the transform of a random variable since the parameters of a random variable are often more easily computed from the transform.

time distribution for early and late arrivals. If a departure has precedence over an arrival, then it is referred to as early arrival system, otherwise it is referred to as late arrival system.

The $\mathrm{GI}^X$/Geom/m-queue in the discrete time domain is investigated by Chaudhry et al. (2001). They analyze the system's steady state using the embedded Markov chain technique and derive the average number of customers in the system. In addition, the waiting time distribution of a random customer of a batch in the queue is given.

### 3.2.2. Batch Service

The pioneer work to analyze batch services is done by Bailey (1954). He investigates the $\mathrm{M/G}^{[1,K]}$/1-queue in which at most $K$ customers can be served. When the service process ends, it is controlled how many customers are waiting in the queue. If at most $K$ customers are present in the queue, all customers can be served and the waiting room of the queue will be empty. If there are more than $K$ customers waiting, the number of customers in the queue is reduced by exactly $K$ and there are customers remaining in the queue. Bailey studies the queue length distribution by using the embedded Markov chain technique. The embedded Markov chain is generated by the length of the queue at instants just before the service takes place. Expressions of the mean waiting time, the mean and the variance of the queue length are derived. Bailey assumes that the server continues to serve even when there are no customers waiting in the queue. In such a case, the service is virtual. If a customer arrives at an empty system, he has to wait until the beginning of the succeeding service process. Bailey was motivated by the idea of designing a hospital outpatient department, where it was required to estimate the waiting time of arriving patients. He also refers to other applications in which batch services occur such as transport processes.

Bailey's work is followed by the contribution of Downtown (1955). Based on Bailey's results, he derived the Laplace transform of the waiting time distribution in equilibrium using the embedded Markov chain technique. He gives expressions for the mean and the variance of the waiting time. Further moments could be derived from the given transform if necessary.

Jaiswal (1960) uses the phase technique to calculate the waiting time distribution in the steady state case for the $\mathrm{M/G}^{[1,K]}$/1-queue. He assumes that the service time is composed of $R$ phases and the service time of each phase is independent and exponentially distributed. From this assumption it follows that Jaiswal's approach is less general than Downtown's.

In contrast to the previous authors, Foster and Nyunt (1961) investigate with the $\mathrm{M/G}^{[K,K]}$/1-queue a somewhat different system in which they assume that the customers are served in batches of exactly size $K$. Thus, the service process starts only if $K$ customers are available in the queue. They derive the equilibrium distribution of the number of customers in the system at the departure instant. Furthermore, they relate their results to the $\mathrm{E}_K$/G/1-queueing model, where E indicates an Erlang distribution with parameter $K$.

Neuts (1965) studies the busy period of the $\mathrm{M/G}^{[1,K]}$/1-queue. He derives the transform of the distribution of the busy period by means of an embedded Markov chain. Additionally, he shows that the busy period is equal to the time elapsed between successive visits to the state "zero customers in the system" in a semi-Markov process.

It is again Neuts (1967) who introduces a batch server system controlled by a control strategy called the minimum batch size rule. This strategy is driven by the number of customers in the queue. The minimum batch size rule means that the service is triggered immediately if there are at least a predefined number of $L$ waiting customers. Otherwise, the batch server, which has a capacity of $K$, remains idle until $L$ customers are accumulated in the queue. $L$ is an arbitrary constant and $1 \leq L \leq K$. Kendall's notation of the described system is M/G$^{[L,K]}$/1. Using the theory of embedded semi-Markov processes Neuts obtained a description of the output process, of the queue length, and the busy period in transform terms.

Chaudhry et al. (1987) present a computational analysis of the steady state probabilities of the M/G$^{[L,K]}$/1-queue. Thereby, they deliver insights into the system's behavior by means of numerical experiments in which they use the deterministic, uniform, Erlangian, and two-form hyperexponential distribution for the service time description.

Gold (1992) analyzes the M/G$^{[L,K]}$/1-S-system in which he considers a limited capacity $S$ of the waiting room. In contrast to Neuts (1967), he derives performance measures in the time domain and not in transform terms. Gold is motivated to develop models for production systems and in this context he presents methods for the M/G$^{[L,K]}$/1-S-system which operates both according to the "push" and "pull" mode. In addition to the M/G$^{[L,K]}$/1-S-system, he analyzes multi-channel systems, denoted by M/M$^{[L_i,K]}$/N-S, in which the service process is a Markov process. The state probabilities in equilibrium of the introduced systems are calculated by means of the embedded Markov chain technique. Furthermore, Gold provides insights into the system's characteristics by studying the mean waiting time, the mean number of customers per service start, and the blocking probability depending on both the system's utilization, the scv of the service time distribution, and the batch server control strategy.

Dümmler (1998) uses discrete time analysis to model the M/G$^{[L,K]}$/1-queue controlled by the minimum batch size rule. Since the arrival process is a Markov process, the system can be analyzed by the embedded Markov chain technique. He determines the distribution of the interdeparture time and size of the departing batch. We will drop the Markovian property of the arrival process in our work (see Chapter 5) and derive an approximation method for the analysis of G/G$^{[L,K]}$/1-queue. Moreover, we will present the waiting time distribution for this more general queueing system.

Powell and Humblet (1986b) and Powell (1985) use batch queueing systems for the investigation of vehicle dispatching strategies in transportation systems[3]. Their idea is to control the queue in order to avoid sending vehicles with uneconomical workloads. Powell and Humblet (1986b) analyze four basic dispatching strategies. First of all, they describe the strategy with no control, in which vehicles depart regardless of the queue length. This model corresponds to the basic model introduced by Bailey (1954). The second queueing control mechanism is the minimum batch size rule first treated by Neuts (1967). This strategy is referred to as a vehicle holding strategy. In addition, Powell

[3]To avoid confusion, Powell and Humblet (1986b) was formulated earlier than Powell (1985).

and Humblet (1986b) introduce the vehicle cancelation strategy. This means, if $x < K$ customers are present at the end of a service period, the departure of the vehicle is canceled in favor of waiting for an additional service period. Finally, they combine the vehicle holding and the vehicle cancelation strategy. If $x < K$ customers are waiting in the queue at the end of a service period, the vehicle is held for an additional time period $t$. If the number of waiting customers is at least $K$ after an additional time period is elapsed, the vehicle is dispatched, otherwise it is canceled. Powell and Humblet (1986b) and Powell (1985) also consider batch arrivals in their models. In Powell (1985), the relative performance of alternative vehicle dispatching strategies is discussed. In a later work, Powell (1986a) derived approximations for closed form moment formulas for batch arrival, bulk service queues by means of standard techniques as the embedded Markov chain. The approximations allow an application without any special knowledge of transforms or complex variables.

Alfa (2005) investigates a batch server system in the discrete time domain in which the interarrival and service times are general distributed. The service is provided in batches, depending on the number of customers waiting when the server becomes free. The interarrival and service times are written as remaining time representations using phase type distributions. Thus, Alfa represents them as remaining time Markov chains. Subsequently, he analyzes the system by means of matrix-geometric methods. For an introduction in matrix-geometric methods see Neuts (1981).

### 3.2.3. Batch Building

We devote a whole chapter of our work to the analysis of the batch building process. In this chapter we present approaches of three different batch building modes. Although the time analysis of the batch building process in a stochastic environment is very crucial for the evaluation of material flow systems, we have to recognize that it has attracted little attention in the literature so far.

Bitran and Tirupati (1989) investigate the $G/G^{[K,K]}/1$-system which is decomposed in a batch building node and in a server node. The arriving customers are collected to a fixed size at the batch building node before they are transferred immediately further to the server node. They compute the first two moments of the interdeparture time after the batch building operation.

For analyzing the optimal batching in a wafer fabrication facility, Fowler et al. (2002) take a multiproduct G/G/c-queue with batch processing. Like Bitran and Tirupati, they also use a decomposition approach. Thus, the queueing system consists of a batch building node, a sever node and a "unbatch" node. In the scope of their analysis, they derive the mean waiting time of a customer arriving at the batch building node.

Meng and Heragu (2004) study in their work the batch building operation in which a given amount of customers is collected. However, their analysis is restricted to the description of the batch building process by the first two moments.

All the above cited contributions concerning the batch building process are done in the continuous time domain.

### 3.2.4. Network Analysis

Material flow systems can be modeled by means of general queueing networks (see Reiser and Kobayashi (1974), Chandy and Sauer (1978), Kuehn (1979), Shanthikumar and Buzacott (1981), and Whitt (1983)). They propose a parametric decomposition approach which is used to approximately calculate performance measures such as the mean number of customers and the mean waiting time at each queueing system. These results are aggregated to network-wide performance measures. The methodology of the analysis of general queueing network relies on the following two notions: Assumption that the nodes of the network can be treated as stochastically independent and the 2-parameter approximation of the arrival process at each node provides reasonably accurate results.

Karmarkar et al. (1985) explicitly considers batch processes in a parametric decomposition approach for the analysis of queueing networks. However, they use M/M/1-queueing systems which generally do not meet the characteristics of real material flow systems. He holds the batch size of each operation on a product type constant.

Calabress and Hausmann (1991) develop a model to investigate the interactions of batch sizing decisions and the routing of the batches through a job shop which is modeled by a closed queueing network. The server stations are described by M/M/c-queues.

Bitran and Tirupati (1989) derive approximations for the departure process from a server station with batch processing in multiproduct queues. They assume that the interarrival and service time within each product class is iid and present methods to calculate the scv of the interdeparture time and the distribution of the size of the departing batch. The obtained solution can be used to analyze queueing networks with batch processing.

An optimized queueing network model to support the design of new and reconfigured semiconductor fabrication facilities is suggested by Hopp et al. (2002). Their solution minimizes the facility costs under the condition that the required throughput is satisfied for all products and the mean manufacturing cycle times for all products are short enough to meet the customer expectations. For estimating the manufacturing cycle times they use G/G/m queueing systems in which batching and setup operations are considered. The approximation results for the G/G/m queue are computed by the approach introduced by Whitt (1993).

Furthermore, Curry and Deuermeyer (2002), and Meng and Heragu (2004) analyze general queueing networks with batch processes. The research of these authors is based on the methods of the Queueing Network Analyzer (QNA) by Whitt (1983) and they derive analytical expressions for the three network operations (merge, split, departure) which are required to calculate the scv of the interarrival time at each node. Curry and Deuermeyer (2002) consider in their paper three different batch service models: one in which batches of constant size arrive at a service station and the customers of the batch are served individually, another in which setup occurs, and lastly one in which batches of constant sizes are served. Meng and Heragu (2004) achieve the same results with a somewhat different approach. They introduce the concept of a relative batch size which puts the batch size of an operation relative to the batch size of the succeeding operation.

In our work we will derive different queueing models which include batch processes. In contrast to the above named authors who develop models for queueing networks with batch processing in the continuous time domain, we do our analysis in the discrete time domain. We derive for each model the waiting time distribution, which can be used for a performance evaluation, and the interdeparture time distribution, which can be taken for a network analysis.

## 3.3. Chapter Conclusion

In this chapter we exemplified the necessity to build batches. Thus, the main reasons related to material flow systems are: Transport/handling, setup, batch service, and order pooling. By means of numerous application examples we underlined the relevance to develop analytical models for a performance evaluation of batch processes.

Furthermore, we gave a review of literature about the analysis of batch queues. The literature review was structured in batch arrival, batch service, batch building, and network analysis. It became apparent that there is a lack of appropriate batch queueing models for the analysis of batch processes. Especially, a detailed analytical description of batch building processes is missing so far. In Schleyer and Furmans (2005) we presented already an initial approach for the analysis of the batch building process until a defined collecting capacity is reached. However, there is much research work left to analyze batch processes under general distribution assumptions. In addition, an analysis of batch processes focusing on the calculation of the probability of an on-time order fulfillment is still missing as well. The logical consequence is the current work which provides a toolkit of batch models in the discrete time domain in order to allow a more detailed analysis and to achieve a better understanding of batch processes.

# 4. Batch Building

*The important thing is not to stop questioning. Curiosity has its own reason for existing.*

Albert Einstein

We highlighted in the previous chapter the economical necessity to build batches. Therefore, the correct batch building decisions are of prime importance for planning and operating material flow systems. As such, optimal batch sizes and an appropriate batch building mode have to be determined.

There are planning decisions which are strongly related to the selection of correct batch sizes.

- For example, what should be the capacity of the transport carriers for a given material flow system? After the transport carrier capacity is determined and the carriers are bought, it is not possible to annul this decision in short-term without enormous expenses.
- Furthermore, what type of machine is the best for a given production system? In the first case, should we choose the one with a greater capacity, but longer setup times? In the second case, should we choose the one with a lower capacity, but lower setup times? In the first case, the batch sizes are large in order to utilize the available capacity and to avoid setup operations. This results in low processing costs per unit, but also in a poor ability to react quickly to demand changes. The inventory level has to be large to sustain a given service level. In the second case, the processing costs per unit are obviously greater. However, the inventory can be reduced and the ability to react rapidly to demand changes increases.

Since real world material flow systems are exposed to stochastic events as demand changes, machine failures, scrap etc., the decision makers must be aware of the consequences arising from their decisions. In this chapter we provide a detailed discrete time analysis of different batch building processes in such a stochastic environment.

It is worthwhile to mention that collecting stations exist not only in material flow systems but also in information systems. An example is the collection of a specific amount of data before these data are released for processing. The reasons for batch building processes in information systems are the same as in production systems, namely an economic utilization of the available resources. We even find many batch building processes in our daily life. Buses, trains and ferries transport "batches" of passengers accumulated at the waiting stations. It becomes evident that the batch building process is a very basic process. Therefore, the economic need to analyze these processes calls for the development of analytical methods for a detailed quantitative performance evaluation of the batch building process in a stochastic environment.

In contrast to the literature known so far, we perform a more detailed analysis of the batch building process. Our work is not restricted to a mean value and a second moment analysis, since we derive both the interdeparture time and the waiting time distribution caused by the batch building operation. The waiting time distribution allows a performance analysis of the batch building operation and the interdeparture time distribution enables us to include the batch building process in a queueing network, where the departure process at the collecting station is the arrival process of the succeeding station. Furthermore, we allow batch arrivals instead of single arrivals at the collecting station.

The current chapter is structured as follows: Section 4.1 introduces basic notations for modeling batch processes used throughout the work. In Sections 4.2 and 4.3 we present exact models for two basic batch building modes, the *capacity rule* and the *timeout rule*. In a subsequent discussion in Section 4.4 we give further insights into the system's behavior of the introduced batch building modes. The relation between input parameters and results is investigated in order to understand the stochastic nature of batch building processes better. Furthermore, Section 4.5 presents a batch building control strategy which is called the *minimum batch size rule*. This introduced analytical approach is exact as well. We show how this approach can be applied to optimization problems.

## 4.1. Modeling of Batch Processes

The batch building process happens at the so called *collecting station* (see Figure 4.1). In the following, material units, jobs or information arriving at the collecting station are referred to as customers, as it is common in queueing theory.

We model the batch building process as general as possible. Thus, we allow batch arrivals at the collecting station and therefore, the arrival process is described by two discrete random variables, $A$, the interarrival time and $Y$, the batch size. Both random variables, $A$ and $Y$, are iid and their distributions are

$$P(A = i) = a_i \quad \forall i = 1, \ldots, a_{max}, \tag{4.1}$$

and

$$P(Y = i) = y_i \quad \forall i = 1,...,y_{max}. \tag{4.2}$$

Given the mean interarrival time $E(A)$, we can determine the batch arrival rate by

$$\lambda_{batch} = \frac{1}{E(A)}. \tag{4.3}$$

The arrival rate $\lambda$ results from the product of the mean batch size and the batch arrival rate:

$$\lambda = E(Y) \cdot \lambda_{batch} = \frac{E(Y)}{E(A)}. \tag{4.4}$$

Let us begin with the analysis of the collecting process using the capacity rule.

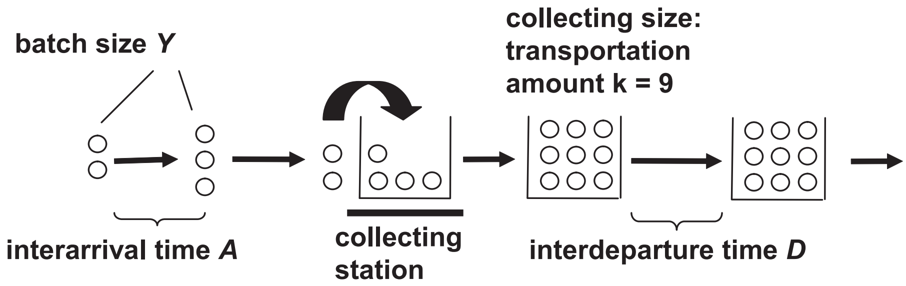


Figure 4.1.: Collecting process using the capacity rule: Exact $k$ customers are collected at the collecting station

## 4.2. Batch Building - Capacity Rule

In this section we introduce our first batch building mode called *capacity rule*. This rule means that a specific amount of $k$ is always collected at the collecting station. We denote $k$ as the collecting size. In material flow applications $k$ is for example the capacity of the transportation carrier. The collecting process under the capacity rule is illustrated in Figure 4.1.

Given $k$, $a_i$ and $y_i$, the waiting time distribution of an arriving customer $w_i$ and the interdeparture time distribution of the collected batches $d_i$ can be determined. Both $d_i$ and $k$ can be used to describe the arrival stream for the succeeding node in a network. In the following, we present an exact analytical approach for deriving $d_i$ and $w_i$.

### 4.2.1. Interdeparture Time Distribution

Since we allow the arrival of batches of stochastic size at the collecting station, a number of combinatorial possibilities to fill the collecting batch arises.

The arriving batch which fills the collecting batch to size $k$ "sees" at its arrival instant $k - i$ collected customers and must have a size of at least $i$ customers $(1 \leq i \leq k)$. If its size is greater than $i$, there are remaining customers after filling the collecting batch. These customers will be used for the subsequent collecting process. The greater the amount of remaining customers, denoted by $R_y$, the lesser the amount of customers needed to fill the subsequent batch and the expected time for filling decreases. The time for collecting $k$ customers and therefore the interdeparture time of the collected batch depends on the amount of remaining customers. Thus, we have to analyze the probability that an amount of $R_y = i$ remaining customers arises. The distribution of the number of remaining customers is called the remainder distribution and is denoted by $r_{y,i}$, $i = 0, \ldots, y_{max} - 1$. We assume that the maximum batch size of an arriving batch, $y_{max}$, is at most the collecting size $k$. This implies that at least one arrival is required to build a batch.

Let us consider the $n$th collecting process. The probability of getting a remainder $R_y^n = i$ after the completion of the $n$th collecting process depends on the remainder of the preceding $(n-1)$th collecting process, but not on the $(n-2)$th collecting process. Thus, we identify a Markov process and calculate $r_{y,i}$ by use of a discrete homogeneous Markov chain. A remainder of size $i$ corresponds to the state $R_y = i$. The maximal remainder follows from the maximum batch size of an arriving batch, $y_{max}$, minus one. This situation arises if a batch of size $y_{max}$ arrives and encounters $k-1$ customers at the collecting station. Therefore, the state number of the considered discrete Markov chain is finite and the state space is defined by a subset of the set $S = \{0,1,\ldots,y_{max}-1\}$.

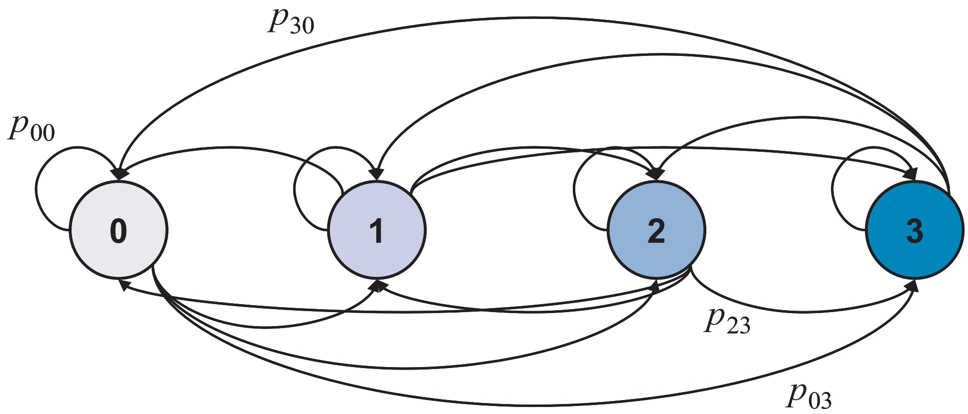


Figure 4.2.: Discrete Markov chain with possible states and transitions for a chosen example: The batch size distribution is given by $y_i$, with $y_i > 0$ for $i = 1,\ldots,y_{max} = 4$.

The transition from the current state to the succeeding state takes places only at discrete points of time. The probability of the transition from state $i$ to state $j$ is called the transition probability $p_{ij}$. The transition probability $p_{ij}$ is independent of $n$, what is referred to as a homogenous Markov chain (see Kleinrock (1975)). It follows that the transition probabilities are stationary and we denote

$$p_{ij} = P(R_y^{n+1} = j | R_y^n = i), \;\; \forall i = 0,\ldots,y_{max}-1, \;\; \forall j = 0,\ldots,y_{max}-1. \tag{4.5}$$

Due to the given description of the collecting process, we can derive the remainder distribution $r_{y,i}$ by the use of a discrete homogeneous Markov chain.

Here, it is worthwhile to mention that various stochastic problems in material flow systems can be solved by a discrete homogeneous Markov chain. An application can be found in Lippolt (2003) who calculates the expected travel times in automated storage/retrieval systems with double-deep storage.

Figure 4.2 shows a discrete homogeneous Markov chain with all possible states and transitions of a collecting process. The batch size distribution is given by $y_i$, with $y_i > 0$ for $i = 1,\ldots,y_{max} = 4$. In this example, there are four states and there are direct transitions possible from every state to each other state.

In some special cases, it would be possible to construct a periodic Markov chain for our collecting process, but this would be quite a rare phenomenon in the context of real applications. Let us explain this using the following example: The collecting size is $k = 3$ and the batch size is Dirac distributed with $y_2 = 1$. Thus, the remainder

$R_y$ is periodically alternating between 1 and 0. The resulting periodic Markov chain is depicted in Figure 4.3.

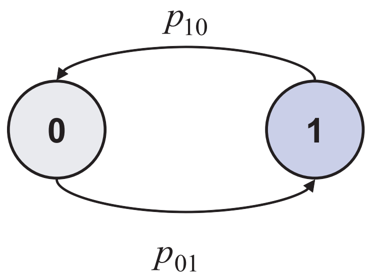


Figure 4.3.: Example of a discrete periodic Markov chain with period 2

Thus, we assume that the discrete Markov chain applied for modeling the collecting process is of an aperiodic nature. Due to Markov, an irreducible[1] and aperiodic Markov chain with a finite state space is ergodic (see Gnedenko and König (1983)). For an ergodic Markov chain, the limit $R_y = \lim_{n\to\infty} R_y^n$ exists and is independent of the initial probability distribution. The stochastic process is stationary for $n \to \infty$ and we can calculate the steady state distribution using the stationary equations defined as follows (see Gross and Harris (1985)):

$$P(R_y = s) = r_{y,s} = \sum_{u=0}^{y_{max}-1} p_{us} r_{y,u} \quad \forall s = 0, \ldots, y_{max} - 1. \tag{4.6}$$

A further equation results from the sum of the steady state probabilities:

$$\sum_{s=0}^{y_{max}-1} r_{y,s} = 1. \tag{4.7}$$

In order to analyze the batch collecting problem we get an over-determined equation system with $y_{max} + 1$ equations, so that one equation from (4.6) has to be omitted.

Before solving the stationary equations, the transition probabilities $p_{ij}$ have to be known, which can be computed by some combinatorics. All possibilities for reaching state $j$ starting from state $i$ are considered. Thus, we obtain with

$$p_{ij} = y_{j+k-i} + \sum_{l=1}^{y_{max}-j} y_{j+l} \sum_{u=1}^{\lceil k/a_{min} \rceil} y_{k-l-i}^{u\otimes} \quad \forall i = 0, \ldots, y_{max} - 1, \quad \forall j = 0, \ldots, y_{max} - 1 \tag{4.8}$$

the equation for the transition probabilities.[2] The expression $y_{j+k-i}$ is only greater than 0 if $j + k - i \leq y_{max}$, otherwise $y_{j+k-i} = 0$.

Based on $r_i$, the number of arriving batches $N_a$ necessary to form one collecting batch are determined in the following. We assume that a collecting process has resulted in a

[1] Irreducible means that every state can be reached directly or indirectly from every other state.

[2] $y_j^{u\otimes}$ is the probability that the sum of $u$ random variables, all described by $y_i$, is $j$ time units; $a_{min}$ is the minimal value of A

remainder of size $R_y = m$. Therefore, at least $k - m$ customers are required to fill the subsequent collecting batch of size $k$. We have to analyze the probability that $N_a = l$ arrivals are required to fill the $k - m$ places. Let us denote this conditional probability by $P(N_a = l | R_y = m)$. We obtain

$$P(N_a = l | R_y = m) = \sum_{j=k-m}^{y_{max}} y_j \quad \text{if } l = 1 \tag{4.9}$$

$$P(N_a = l | R_y = m) = \sum_{i=1}^{k-l-m+1} y_{k-i-m}^{(l-1)\otimes} \sum_{j=i}^{y_{max}} y_j \quad \forall l > 1. \tag{4.10}$$

If $l > 1$, the first part of equation (4.10) describes the probability that $l - 1$ arrivals fill the collecting batch to $k - i$, $i < k - m$ under the condition that the remainder is of size $m$. Then, the batch size $Y = j$ of the $l$th arrival has to be greater than or equal to $i$. Now, the next collecting process starts with a remainder of size $j - i$.

Given $r_{y,i}$ and the results of Equations (4.9) and (4.10), the law of total probability leads to the distribution of the number of arrivals required to fill a collecting batch. We derive

$$P(N_a = l) = \sum_{m=0}^{k-1} P(N_a = l | R_y = m) \cdot r_{y,m}. \tag{4.11}$$

The result of the $l$-fold convolution of $a_i$ weighted with the probability $P(N_a = l)$ leads to the interdeparture time distribution $d_i$. It yields

$$d_i = P(N_a = l) \cdot a_i^{l\otimes}. \tag{4.12}$$

Considering a single arrival stream instead of a batch arrival stream, the derivation of $d_i$ via the discrete homogenous Markov chain is not necessary, because there are no remaining customers after the departure of a collected batch. Thus, we obtain $d_i$ easily by the $k$-fold convolution of $a_i$, since $k$ arrivals are required to fill the collecting batch. Therefore, we obtain

$$d_i = a_i^{k\otimes}. \tag{4.13}$$

### 4.2.2. Waiting Time Distribution

A collecting process causes a waiting process for the arriving customers. We make the same assumptions for the derivation of the waiting time distribution as we did for the derivation of the interdeparture time distribution.

Observing the $n$th collecting process, we recognize that all these customers of the arriving batch which fill the collecting batch to the amount $k$ do not have to wait. We identify them as “lucky” customers. Those customers who are remaining for the collection of the succeeding batch have to wait until the succeeding $(n + 1)$th collecting process is completed. We define them as “unlucky” customers. The described situation is illustrated in Figure 4.4.

The other arriving customers have to wait depending on the fact that they arrive at the beginning or at the end of a collecting process. If a collected batch is composed of $l$ batch arrivals, the waiting time of an individual customer depends on the following events:

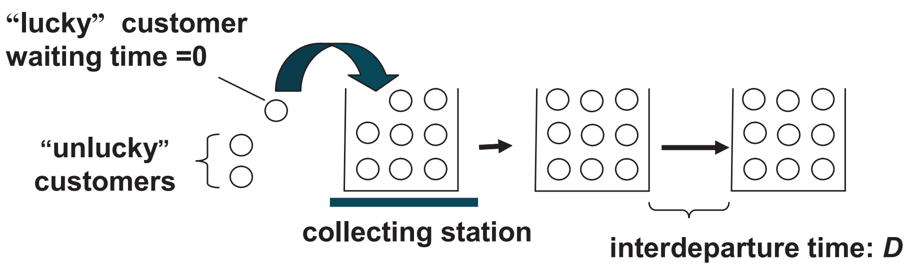


Figure 4.4.: Collecting process: “Lucky” customers who complete the batch and have no waiting time, and “unlucky” customers who have to wait at the collecting station until the next batch is completed

- the event that customers can be immediately transferred to the succeeding station
$\rightarrow$ “lucky” customers; their waiting time is zero
- the event that there are customers remaining
$\rightarrow$ “unlucky” customers; their waiting time is the duration of a complete collecting process
- the event that customers arrive with the $i$th batch arrival with $i < l$
$\rightarrow$ these customers have to wait for $l - i$ further batch arrivals until the collecting process is completed

We conclude that the distribution of the number of arrivals to fill one collecting batch $P(N_a = l)$ (see Equation (4.11)) is required for the derivation of the waiting time distribution.

Figure 4.5 illustrates an example, where four arrivals are used to fill a batch of nine units. The waiting times of the customers as multiples of the interarrival time are highlighted.

First, we compute $E(R|N_a = l)$, the mean number of remaining customers under the condition that $l$ arrivals are required for the collecting process. These are the “unlucky” customers who have to wait for the duration of a complete collecting process. Using expressions $P(N_a = l|R_y = m)$ and $P(N_a = l)$ known from Equations (4.9), (4.10) and (4.11), we obtain by the use of Bayes formula

$$P(R_y = m|N_a = l) = \frac{P(N_a = l|R_y = m) \cdot r_{y,m}}{P(N_a = l)}, \tag{4.14}$$

and furthermore

$$E(R_y|N_a = l) = \sum_{m=1}^{y_{max}-1} P(R_y = m|N_a = l) \cdot m. \tag{4.15}$$

The customers who arrive with the $i$th arrival, $i < l$, have to wait for further $l-i$ arrivals until the $n$th collecting process is completed. For these customers the waiting time distribution results from the $(l-i)$-fold convolution of the interarrival time distribution. Thus, we have to consider the probability that an arbitrary customer belongs to the $i$th arrival under the condition that $l$ arrivals, $l > i$, are required to fill the collecting batch.

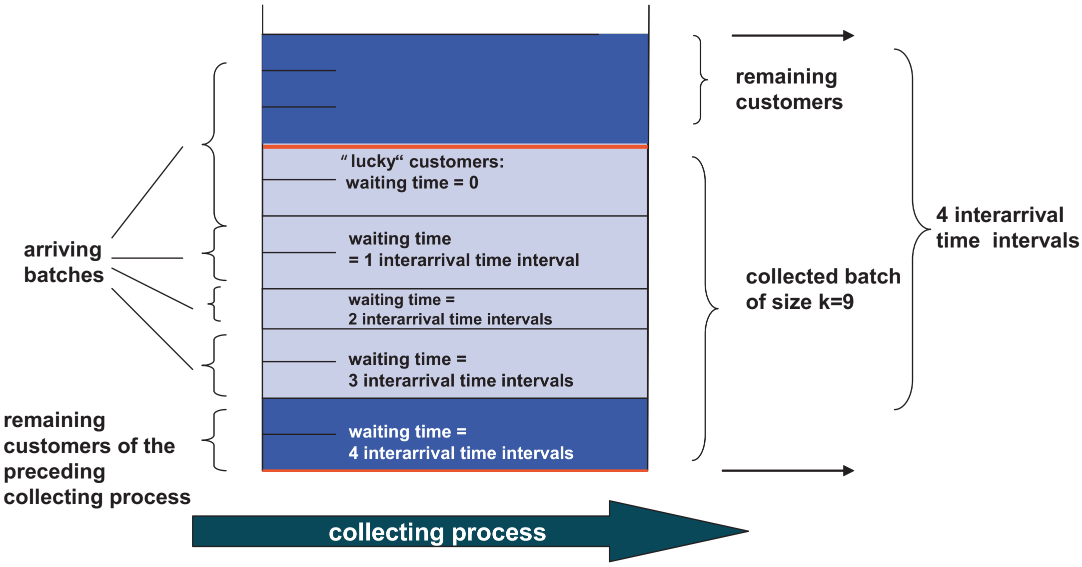


Figure 4.5.: Example of a batch building process with $l = 4$ batch arrivals to fill up a batch of size $k = 9$. Illustrated is the waiting time of the arriving customers as multiples of the interarrival time.

$E(Y|N_a = l)$ is referred to as the expected batch size of an arriving batch under the condition that $l > 1$ arrivals are required to fill the collecting batch. This leads to

$$E(Y|N_a = l) = \frac{1}{P(N_a = l)} \sum_{m=0}^{y_{max}-1} r_{y,m} \sum_{i=y_{min}}^{k-m-y_{min}} y_{k-i-m} \cdot (k-i-m) \sum_{j=i}^{y_{max}} y_j \quad \text{if } l = 2 \tag{4.16}$$

and

$$E(Y|N_a = l) = \frac{1}{P(N_a = l)} \sum_{m=0}^{y_{max}-1} r_{y,m} \sum_{x=y_{min}}^{k-m-y_{min}\cdot(l-2)} y_x \cdot x \sum_{i=y_{min}}^{k-m-x-y_{min}\cdot(l-2)} \sum_{j=i}^{y_{max}} y_{k-i-m-x}^{(l-2)\otimes} \cdot y_j \quad \text{if } l > 2. \tag{4.17}$$

Due to the fact that the batch size $y$ is iid, the expected size of all the arriving batches $E(Y|N_a = l)$ is equal for the first to the $(l-1)$th batch arrival $(l \geq 2)$.

The quotient of $E(Y|N_a = l)$ and $k$ is the sought probability that an arriving customer belongs to the $i$th arrival with $i = 1, \ldots, l-1$ and $l > 1$. Thus, we determine the waiting time distribution of an individual customer under the condition that $N_a = l$ arrivals are required for one collecting process by

$$P(W = 0|N_a = l) = \left(1 - \frac{E(R_y|N_a = l) + (l-1) \cdot E(Y|N_a = l)}{k}\right) \tag{4.18}$$

and

$$P(W = i|N_a = l) = a_i^{l\otimes} \cdot \frac{E(R_y|N_a = l)}{k} + \sum_{j=1}^{l-1} a_i^{(l-j)\otimes} \cdot \frac{E(Y|N_a = l)}{k} \tag{4.19}$$

if $i \geq 1$.

The law of total probability leads to the waiting time distribution given as follows:

$$\begin{aligned} w_i = P(W = i) = \sum_{l=1}^{l_{max}} P(W = i|N_a = l) \cdot P(N_a = l) \qquad & i = 0,1,\ldots \\ & l_{max} = \lceil \frac{k}{y_{min}} \rceil, \end{aligned} \tag{4.20}$$

If only single arrivals are considered instead of batch arrivals, the calculation of the waiting time distribution can be done quite easily. Generally, if a customer arrives at the collecting station, he encounters $x$ waiting customers, with $0 \leq x < k$. If $0 \leq x < k - 1$, he has to wait for the missing $k - x - 1$ customers. On the other hand, if $x = k - 1$, the customer does not have to wait, because the collected batch is filled to size $k$ and is transferred immediately to the succeeding station. The probability that an arriving customer encounters $x$ customers is $1/k$. Thus, the waiting time distribution can be determined by

$$\begin{aligned} w_i &= P(W = i) = \frac{1}{k} \sum_{j=1}^{k-1} a_i^{(k-j)\otimes} \qquad \text{if } i \geq 1 \\ w_0 &= P(W = 0) = \frac{1}{k} \qquad \text{if } i = 0. \end{aligned} \tag{4.21}$$

In the subsequent section we introduce a further batch building mode, the *timeout rule*. We present an exact analytical approach to calculate the interdeparture and waiting time distribution.

## 4.3. Batch Building - Timeout Rule

In the preceding section we analyzed the batch building process under the capacity rule, where a fixed collecting size $k$ was given. Under the timeout rule, we assume that the time for the batch building process is given by the time $t_{out}$ and there is no limiting capacity. This means that the arriving customers are collected within a constant time $t_{out}$. As in the preceding section the arrival stream is described by $a_i$, $i = 1, \ldots, a_{max}$ and $y_i$, $i = 1, \ldots, y_{max}$.

For example, this type of collecting mode is common in production systems, especially in lean production systems, where it is referred to as "milkrun". The term milkrun describes a transport mode, where predetermined loading and unloading stations are visited by a transport vehicle according to a predefined time table. The milkrun route is repeated in regular small time intervals. With the following method it is possible to calculate the number of units collected at the loading stations within a determined time interval. The results can be used for the buffer dimensioning at the loading stations and for the configuration of a suitable frequency of milkruns. Figure 4.6 illustrates the concept of a milkrun.

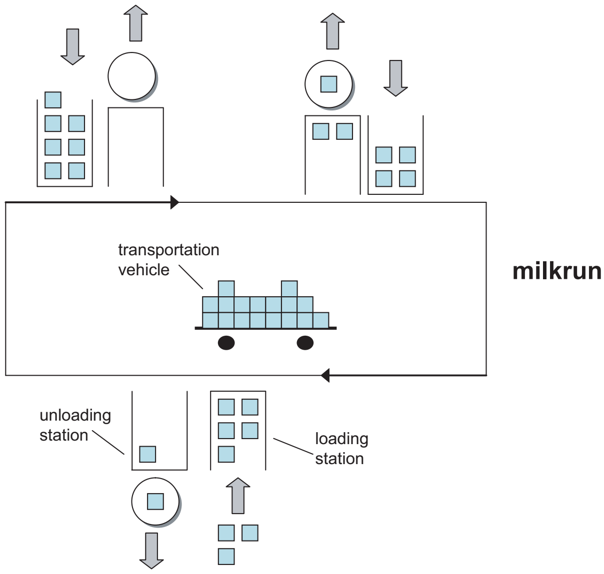


Figure 4.6.: Example for an application of the collecting process using the timeout rule: The milkrun transport mode: Predetermined loading and unloading stations are visited by a transportion vehicle according to a predefined time table

Moreover, the following analytical approach can also be used for the investigation of stochastic processes in other fields, like traffic research. Let us consider a bus timetable in which visits at predetermined bus stops on a predetermined route are listed. At each of these stops, passengers arrive in stochastic interarrival time intervals within the time interval $t_{out}$. Furthermore, they often arrive in groups, that means in batches. Now, with the given arrival process at each bus stop, we can compute the required capacity for the buses on their routes and the required stop frequency at a bus stop with the objective to fulfill a predefined service level.

Since $t_{out}$ is constant, the departure process at the collecting station is described by a constant interdeparture time of $D = t_{out}$ $(d_{t_{out}} = 1)$ and a stochastic batch size $Y_d$ distributed by $y_{d,i}$. Subsequently, we investigate the collecting process according to the timeout rule and derive $y_{d,i}$ and $w_i$.

### 4.3.1. Batch Size Distribution of the Departing Batch

The batch size of the collected batch depends on the number of arrivals within $t_{out}$ and the sizes of the arriving batches. If large batches arrive in short intervals, the collected batch becomes huge. A collecting process according to the timeout rule is illustrated in Figure 4.7.

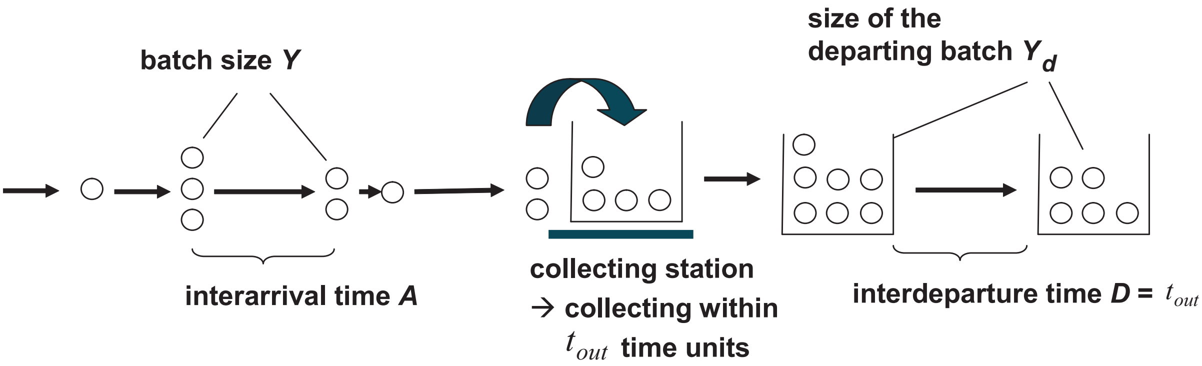


Figure 4.7.: Collecting process according to the timeout rule: Arriving customers are collected within a constant time $t_{out}$

In addition, Figure 4.8 emphasizes the time behavior of the collecting process in more detail. Possible events of the $n$th, $n > 1$, collecting process are depicted on a time axis. Subsequently, we consider the $n$th collecting process. The time instant at which the $n$th collecting process ends is denoted by $\delta_{out}^n$, which corresponds to the time instant at which the $(n + 1)$th collecting process starts. The time instant at which the arriving batch sees the collecting station empty, is denoted by $\tau^n$ (see Figure 4.8). The time interval from $\tau^n$ to $\delta_{out}^{n-1}$ is either a complete interarrival time interval or a residual of it. We refer to this time period as the residual interarrival time interval $R_a$, distributed by $r_{a,i}$, $i = 1, \ldots, a_{max}$. The greater the residual interarrival time interval of the $n$th collecting process $R_a^n = \tau^n - \delta_{out}^{n-1}$, the less time units remain for collection. Therefore, the batch size of the collected batch depends on the residual interarrival time interval length. Thus, we have to analyze the probability that a residual interarrival time interval of length $i$ arises.

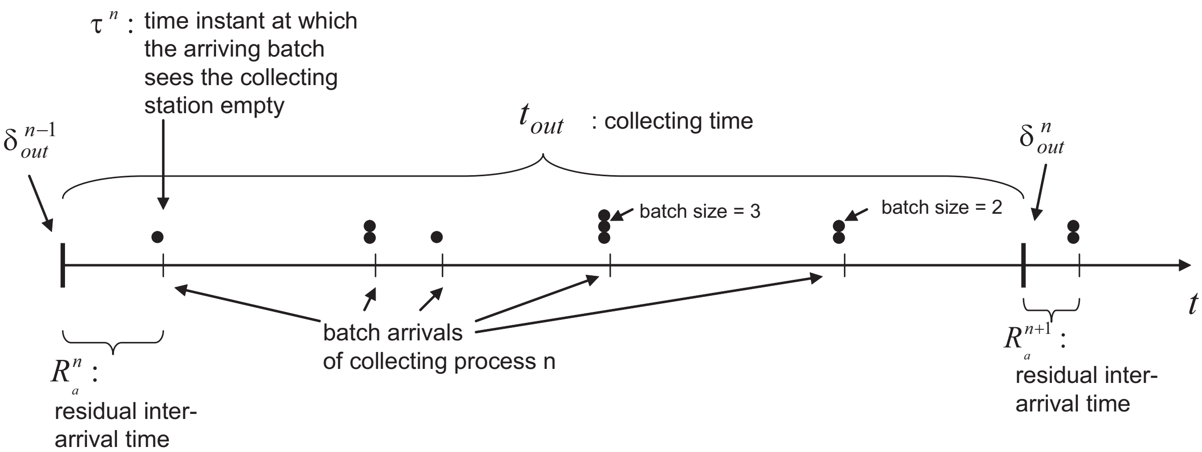


Figure 4.8.: Time axis of a collecting process according to the timeout rule

Since the $n$th collecting process depends on $R_a^n$ and not on $R_a^{n-1}$, we identify a Markov process. This observation is analogous to the observation we made in the analysis of the batch building process according to the capacity rule (see Section 4.2). Now, we observe a residual time at the beginning of the collecting process, whereas in Section 4.2 we observed a residual amount. Again, the considered collecting process can be modeled by a discrete homogenous Markov chain.

The state number is finite and the state space is defined by a subset of the set $S = \{1,\ldots,a_{max}\}$. Analogous to the collecting process according to the capacity rule, we have an irreducible and an aperiodic Markov chain with a finite state space (see the argumentation in Section 4.2). Thus, it follows that the Markov chain is ergodic and the steady state distribution which corresponds the residual interarrival time distribution, $r_{a,i}$, can be computed by the Equations (4.6) and (4.7).[3] The transition probabilities can be determined by

$$p_{ij} = a_{j+t_{out}-i} + \sum_{l=0}^{a_{max}-j} a_{j+l} \sum_{n=1}^{\lceil t_{out}/a_{min} \rceil} a^{n\otimes}_{t_{out}-l-i} \qquad \forall i = 1,\ldots,a_{max} \quad \forall j = 1,\ldots,a_{max}. \tag{4.22}$$

By computing $p_{ij}$ we have to take into account the time interval between the last batch arrival of a collecting process and the first batch arrival of the succeeding collecting process. The time interval between these two events must have a length of at least $j$ time units. The expression $a_{j+t_{out}-i}$ in Equation (4.22) is only greater than 0 if $j + t_{out} - i \leq a_{max}$, otherwise $a_{j+t_{out}-i} = 0$. We assume that $a_{max} \leq t_{out}$, which guarantees that at least one arrival occurs within $t_{out}$. Otherwise, Equation (4.22) has to be extended for the case that no arrival occurs within $t_{out}$.

The number of batch arrivals within the time $t_{out}$, denoted by $N_a$, can be calculated based on $r_{a,i}$. First, we determine $P(N_a = l|R_a = s)$, the probability that $l$ batches arrive within $t_{out}$ under the condition that $R_a$ is $s$ time units, $1 \leq s \leq a_{max}$. We obtain

$$P(N_a = l|R_a = s) = \sum_{j=t_{out}-s+1}^{a_{max}} a_j \quad \text{if } l = 1 \tag{4.23}$$

$$P(N_a = l|R_a = s) = \sum_{i=0}^{j=t_{out}-s-1} a^{(l-1)\otimes}_{t_{out}-i-s} \sum_{j=i+1}^{a_{max}} a_j \quad \text{if } l > 1. \tag{4.24}$$

- Equation (4.23) describes the case that exactly one batch arrival is recorded within $t_{out}$. This occurs if $A$ is greater than $t_{out} - s$ time units.
- Using Equation (4.24) we calculate the probability that $l > 1$ batches arrive within the collecting time $t_{out}$ and the $(l+1)$th batch arrives when $t_{out}$ is already passed.

Given $P(N_a = l|R_a = s)$ and $P(R_a = s)$ by solving Equation (4.22), we get $P(N_a = l)$ by

$$P(N_a = l) = \sum_{s=1}^{a_{max}} P(N_a = l|R_a = s) \cdot r_{a,s}. \tag{4.25}$$

[3]Note that the smallest residual interarrival time is one due to the definition made previously. This leads in comparison with the capacity rule to some small differences in the derivation of the transition probabilities and of the number of batch arrivals within the time $t_{out}$.

The result of the $l$-fold convolution of $y_i$ weighted by the probability $P(N_a = l)$ leads to the batch size distribution of the collected batch $y_{d,i}$. We compute

$$y_{d,i} = P(N_a = l) \cdot y_i^{l\otimes}. \tag{4.26}$$

The batch size distribution $y_{d,i}$ and $t_{out}$ describe the departure process at the collecting station, which can be used for the analysis of the succeeding node in a queueing network.

### 4.3.2. Waiting Time Distribution

The considered collecting process causes a waiting process for the arriving customers, which is bounded to $t_{out} - 1$ time units. If a batch arrives at the collecting station immediately one time unit after the preceding collecting process was completed, these customers have to wait for the maximum possible time of $t_{out} - 1$ time units. In contrast, the customers who arrive exactly at the end of the collecting process have to wait zero time units. In this case, we observe the favorable situation that all arriving customers can be transferred to the succeeding station without any delay.

At first, we derive the waiting time distribution of an arbitrary customer depending on the residual interarrival time interval, $R_a$.

The arriving batch which sees the collecting station empty arrives $s$ time units ($s = 1, \ldots, a_{max} - 1$) after the completion of the preceding collecting process. Therefore, the customers of this batch have to wait exactly $t_{out} - s$ time units with the probability of one. We conclude

$$P(W^1 = t_{out} - s | R_a = s) = 1. \tag{4.27}$$

Analyzing the second arriving batch, we derive with

$$P(W^2 = t_{out} - i | R_a = s) = a_{i-s} \quad i = s+1, \ldots, t_{out} \tag{4.28}$$

its waiting time distribution. In general, this yields

$$P(W^l = t_{out} - i | R_a = s) = a_{i-s}^{(l-1)\otimes} \qquad i = s+1, \ldots, t_{out},\ l = 2, \ldots, l_{max}. \tag{4.29}$$

$P(W^l = t_{out} - i | R_a = s)$, $l > 1$ is a defective distribution, since $i$ is bounded by $t_{out}$. The index $l$ is bounded by the maximum possible number of batch arrivals within $t_{out}$, which can be calculated by

$$l_{max} = \lceil \frac{t_{out}}{a_{min}} \rceil, \tag{4.30}$$

where $a_{min}$ is the minimum interarrival time. For the waiting time of an arbitrary customer under the condition of $R_a = s$, we obtain the following equivalence relation:

$$\begin{aligned} &P(W = t_{out} - i | R_a = s) \\ &\quad \sim \begin{cases} 1 & \text{if } i = s \\ \sum_{l=2} P(W^l = t_{out} - i | R_a = s) = \sum_{l=2}^{l_{max}} a_{i-s}^{(l-1)\otimes} & \text{if } i > s \end{cases} \\ &\quad = P(W^* = t_{out} - i | R_a = s). \end{aligned} \tag{4.31}$$

By resolving the dependence on $R_a$ we obtain

$$P(W^* = t_{out} - i) = \sum_{s=1}^{a_{max}} P(W^* = t_{out} - i | R_a = s) \cdot r_{a,s}. \tag{4.32}$$

Finally, we have to normalize $P(W^* = t_{out} - i)$ and compute the waiting time by

$$P(W = t_{out} - i) = \frac{P(W^* = t_{out} - i)}{\sum_{i=1}^{t_{out}} P(W^* = t_{out} - i)} \quad i = 1, \ldots, t_{out}. \tag{4.33}$$

We note additionally that

$$P(N_a \geq l | R_a = s) = \begin{cases} P(W^l = t_{out} - s | R_a = s) = 1 & \text{if } l = 1 \\ \sum_{i=s+1}^{t_{out}} P(W^l = t_{out} - i | R_a = s) & \text{if } l > 1 \end{cases} \tag{4.34}$$

represents the probability that at least $l$ batches arrive within the collecting time $t_{out}$ under the condition of $R_a = s$. Due to

$$\begin{aligned} E(N_a | R_a = s) &= \sum_{l=1}^{l_{max}} P(N_a \geq l | R_a = s) \\ &= 1 + \sum_{l=2}^{l_{max}} \sum_{i=s+1}^{t_{out}} P(W^l = t_{out} - i | R_a = s) \\ &= 1 + \sum_{l=2}^{l_{max}} \sum_{i=s+1}^{t_{out}} a_{i-s}^{(l-1)\otimes} \\ &= 1 + \sum_{i=s+1}^{t_{out}} P(W^* = t_{out} - i | R_a = s) \\ &= \sum_{i=s}^{t_{out}} P(W^* = t_{out} - i | R_a = s) \end{aligned} \tag{4.35}$$

we derive the mean number of batch arrivals within $t_{out}$ under the condition of $R_a = s$. Further, we obtain

$$E(N_a) = \sum_{i=1}^{t_{out}} P(W^* = t_{out} - i) \tag{4.36}$$

and we can simplify Equation (4.33) as follows:

$$w_{t_{out}-i} = P(W = t_{out} - i) = \frac{P(W^* = t_{out} - i)}{E(N_a)} \quad i = 1, \ldots, t_{out}. \tag{4.37}$$

Finally, it is worthwhile to mention that the batch size distribution has no influence on the waiting time distribution of an arbitrary customer since the batch size is iid.

**Additional note to Equation (4.35):**

Proof of

$$E(N_a|R_a = s) = \sum_{l=1}^{l_{max}} P(N_a \geq l|R_a = s). \tag{4.38}$$

Generally, we conclude that

$$\begin{aligned}
\sum_{l=1}^{l_{max}} P(x \geq l) &= \sum_{l=1}^{l_{max}} \left( \sum_{k=l}^{l_{max}} P(x = k) \right) \\
&= \sum_{k=1}^{l_{max}} P(x = k) + \sum_{k=2}^{l_{max}} P(x = k) + \sum_{k=3}^{l_{max}} P(x = k) + \dots \\
&= P(x = 1) + 2 \cdot \sum_{k=2}^{l_{max}} P(x = k) + \sum_{k=3}^{l_{max}} P(x = k) + \dots \\
&= P(x = 1) + 2 \cdot P(x = 2) + 2 \cdot \sum_{k=3}^{l_{max}} P(x = k) \\
&+ \sum_{k=3}^{l_{max}} P(x = k) + \dots \\
&= P(x = 1) + 2 \cdot P(x = 2) + 3 \cdot \sum_{k=3}^{l_{max}} P(x = k) + \dots \\
&= \sum_{l=1}^{l_{max}} l \cdot P(x = l) = E(x).
\end{aligned} \tag{4.39}$$

## 4.4. System's Behavior of the Basic Batch Building Modes

After the presentation of the analytical approaches for the batch building mode under the capacity and the timeout rule, it is intended to provide further insights into the system's behavior of these basic batch building modes through the following discussions. These insights aid in obtaining a profound understanding of the stochastic nature of batch building processes in real life applications.

In Section 4.4.1 we explain in detail why $w_i$ under the timeout rule can be easily determined by $1/t_{out}$, $i = 0, \dots, t_{out} - 1$. This in turn leads to intuitive solutions for the calculation of $R_y$ at the capacity rule and of $R_a$ at the timeout rule. It can be proved that this intuitive approach is correct by applying the previous approach for the determination of $R_y$ and $R_a$ via the discrete Markov chain. Both approaches lead to the same numerical results. However, the intuitive approach leads to a significant reduction in computing times of nearly 50%.

Furthermore, we provide in Sections 4.4.2 and 4.4.4 insights into the system's performance depending on the input parameters, which can be used to support management decisions regarding the design and operation of material flow systems. The equivalencies between the capacity and the timeout rule are shown in Section 4.4.3.

### 4.4.1. Waiting Time Behavior under the Timeout Rule

By studying the numerical results of $w_i$ under the timeout rule, it is remarkable that the waiting time distribution can be easily calculated by

$$w_i = P(W = i) = \frac{1}{t_{out}} \qquad i = 0, \ldots, t_{out} - 1. \tag{4.40}$$

To confirm this observation we have to transform the derived solution in Equation (4.37) to $1/t_{out}$. However, for this transformation a closed solution of $r_{a,i}$ is required, which we present subsequently by an intuitive approach.

We consider the batch arriving instants on a time axis which continues to infinity (see Figure 4.9). It is assumed that the interarrival time distribution $a_i$ has a scv greater than zero ($c_A^2 > 0$). The time instants of the timeouts are also depicted in Figure 4.9.

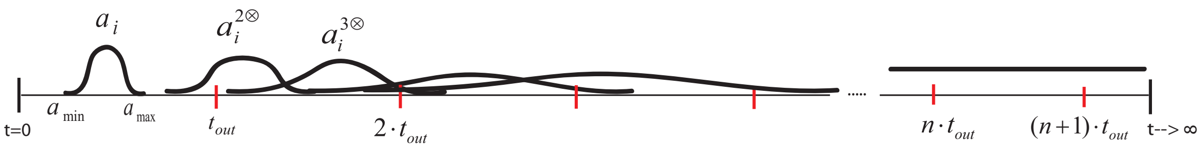


Figure 4.9.: Probability to meet a large number by the sum of $n$ iid random variables, where $n$ is a large number

We consider a truncated section of the time axis after sufficient time is elapsed. We notice that an arrival event coincides with an arbitrary time instant within the collecting period with the same probability as with any other time instant within the same collecting period. This is due to the fact that the probability of meeting a very large number $x$ resulting from the sum of a sufficient amount of iid random variables is the same that of meeting the number $x + 1$.[4] Therefore, the probability to meet a number $x$ with $n \cdot k \leq x < (n + 1) \cdot k$ ($n$ is a large number) is uniformly distributed within the interval $[n \cdot k; (n + 1) \cdot k)$.

See Figure 4.9 for an illustration.[5] The probability that the arrival meets a time instant is sketched. Thus, the first arrival takes place within a time interval beginning at $a_{min}$ and ending at $a_{max}$. The next time period which can be met by an arrival is between $2 \cdot a_{min}$ and $2 \cdot a_{max}$. Its probability is given by the 2-fold convolution of $a_i$. The value range of

[4]If all possible values of the iid random variable are multiples of $l$, we can consequently meet only numbers of multiples of $l$. If the value scale is then divided by $l$, our observation is again valid.

[5]We used a continuous distribution for $a_i$ since it simplifies the drawing of the figure and makes it better readable. Furthermore, our observation is valid both in the discrete and continuous time domain.

the $k$-fold convolution superposes on the value range of the $(k-1)$-fold convolution if $k$ is great enough. This superposition is apparent in Figure 4.9 if $k = 3$ and increases with an increasing $k$. It is to be noted that independent of the distribution of $a_i$ $(c_A^2 > 0)$ the $k$-fold convolution of $a_i$ is normally distributed if $k$ tends to infinity. Due to the symmetry of the normal distribution, the height of the area formed by the superposed graph and the time axis is uniformly high if sufficient time is elapsed. Therefore, each time instant between two subsequent timeout instants which results from the sum of a suitable amount of arrival intervals has the same probability to become an arrival instant. The time between an arrival and the timeout instant is the waiting time of the arriving customers. Due to the collecting time length of $t_{out}$, a waiting time distribution of $w_i = 1/t_{out}$, $i = 0, \ldots, t_{out} - 1$ results.

In the following, we prove that Equation (4.37) leads to $w_i = 1/t_{out}$, $i = 0, \ldots, t_{out} - 1$. We have identified that every arrival instant within the collecting time period is equally probable. It follows that every possible timeout instant placed between two succeeding batch arrivals is equally probable as well. Therefore, $R_a$, the time between a timeout instant and the succeeding arrival instant, corresponds to the residual lifetime of a renewal process observed immediately after the event occurrence. We illustrate our observation in Figure 4.10.

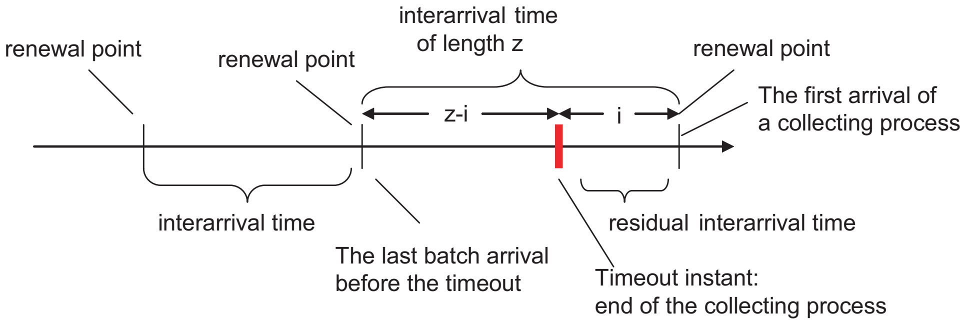


Figure 4.10.: Renewal process at the batch building process according to the timeout rule

The derivation of the distribution of the residual lifetime of a renewal process observed immediately after discrete time instants is presented in Section 2.1.2. This leads to the distribution of $R_a$ derived by

$$r_{a,i} = \frac{1}{E(A)} \cdot (1 - \sum_{k=0}^{i-1} a_k) \qquad i = 1, \ldots, a_{max}. \tag{4.41}$$

Expression

$$\overline{a}_i = (1 - \sum_{k=0}^{i-1} a_k) \tag{4.42}$$

describes the probability that the interarrival time interval has a length of at least $i$ and is used to ease the notation. Thus, Equation (4.41) is simplified by

$$r_{a,i} = \frac{\overline{a}_i}{E(a_{batch})} \qquad i = 1, \ldots, a_{max}. \tag{4.43}$$

It follows to proof that

$$P(W = t_{out} - i) = \frac{P(W^* = t_{out} - i)}{E(N_a)} = \frac{1}{t_{out}} \qquad i = 1, \dots, t_{out}.$$

We conclude

$$\begin{aligned} P(w = t_{out} - i) &= \frac{P(W^* = t_{out} - i)}{E(N_a)} \\ &= \frac{P(W^* = t_{out} - i | R_a = s) \cdot r_{a,s}}{E(N_a)} \\ &= \begin{cases} \frac{1}{E(N_a)} \left[ r_{a,i} + \sum_{s=1}^{i-1} r_{a,s} \cdot \sum_{l=2}^{l_{max}} a_{i-s}^{(l-1)\otimes} \right] & 1 \le i \le r_{a,max} \\ \frac{1}{E(N_a)} \left[ \sum_{s=1}^{r_{a,max}} r_{a,s} \cdot \sum_{l=2}^{l_{max}} a_{i-s}^{(l-1)\otimes} \right] & r_{a,max} < i \le t_{out} - 1 \end{cases} \\ &= \begin{cases} \frac{1}{E(N_a) \cdot E(A)} \left[ \overline{a}_i + \sum_{s=1}^{i-1} \overline{a}_s \cdot \sum_{l=2} a_{i-s}^{(l-1)\otimes} \right] & 1 < i \le r_{a,max} \\ \frac{1}{E(N_a) \cdot E(A)} \left[ \sum_{s=1}^{r_{a,max}} \overline{a}_s \cdot \sum_{l=2} a_{i-s}^{(l-1)\otimes} \right] & r_{a,max} < i \le t_{out} - 1 \end{cases} \end{aligned} \tag{4.44}$$

In addition, we use

$$\xi^j = \sum_{l=1}^{l_{max}} a_j^{l\otimes}, \tag{4.45}$$

which describes the probability that the sum of at least one random variable, here described by $a_i$, is $j$. Finally, since

$$E(N_a) = \frac{t_{out}}{E(A)}, \tag{4.46}$$

we simplify Expression (4.44) to

$$P(w = t_{out} - i) = \begin{cases} \frac{1}{t_{out}} \left[ \overline{a}_i + \sum_{s=1}^{i-1} \overline{a}_s \cdot \xi^{i-s} \right] & 1 < i \le r_{a,max} \\ \frac{1}{t_{out}} \left[ \sum_{s=1}^{r_{a,max}} \overline{a}_s \cdot \xi^{i-s} \right] & r_{a,max} < i \le t_{out} - 1. \end{cases} \tag{4.47}$$

Equation (4.47) is equal to $1/t_{out}$ if the expression inside the brackets equals 1. For our proof we use the subsequent expression which is obtained by

$$\xi^j = \sum_{i=1}^{j} a_i \xi^{j-i} = a_1 \xi^{j-1} + a_2 \xi^{j-2} + \dots + a_j \qquad j = 1, \dots \tag{4.48}$$

Let us first analyze the case of $1 < i \le r_{a,max}$. We obtain

$$\overline{a}_i + \sum_{s=1}^{i-1} \overline{a}_s \cdot \xi^{i-s} = \overline{a}_i + \overline{a}_1 \xi^{i-1} + \overline{a}_2 \xi^{i-2} + \dots + \overline{a}_{i-2} \xi^2 + \overline{a}_{i-1} \xi^1.$$

$$
\begin{aligned}
&= \overline{a}_i + \underbrace{\overline{a}_1}_{=1}(a_1\xi^{i-2} + a_2\xi^{i-3} + \ldots + a_{i-2}\xi^1 + a_{i-1}) \\
&+ \overline{a}_2\xi^{i-2} + \ldots + \overline{a}_{i-2}\xi^2 + \overline{a}_{i-1}\xi^1 \\
&= \overline{a}_i + a_{i-1} + \overline{a}_2\xi^{i-2} + a_1\xi^{i-2} + \ldots + \overline{a}_{i-1}\xi^1 + a_{i-2}\xi^1 \\
&= \overline{a}_{i-1} + \underbrace{\overline{a}_1}_{=1}\xi^{i-2} + \ldots + \overline{a}_{i-2}\xi^1 \\
&= \overline{a}_{i-1} + (a_1\xi^{i-3} + a_2\xi^{i-4} + \ldots + a_{i-3}\xi^1 + a_{i-2}) + \ldots + \overline{a}_{i-2}\xi^1 \\
&= \overline{a}_{i-2} + \underbrace{\overline{a}_1}_{=1}\xi^{i-3} + \ldots + \overline{a}_{i-3}\xi^1 \\
&\qquad\qquad \vdots \\
&= \overline{a}_2 + \overline{a}_1\xi^1 = \overline{a}_2 + a_1 = \overline{a}_1 = 1 \qquad q.e.d.
\end{aligned}
$$

For $r_{a,max} < i \leq t_{out} - 1$ we obtain

$$
\begin{aligned}
&\sum_{s=1}^{r_{a,max}} \overline{a}_s \cdot \xi^{i-s} \\
=&\overline{a}_1\xi^{i-1} + \overline{a}_2\xi^{i-2} + \overline{a}_3\xi^{i-3} + \ldots + \overline{a}_{r_{a,max}-1}\xi^{i-r_{a,max}-1} + \overline{a}_{r_{a,max}}\xi^{i-r_{a,max}} \\
=&\overline{a}_1(a_1\xi^{i-2} + a_2\xi^{i-3} + \ldots + a_{r_{a,max}-1}\xi^{i-r_{a,max}} + a_{r_{a,max}-1}\xi^{i-r_{a,max}-1}) \\
&+ \overline{a}_2\xi^{i-2} + \overline{a}_3\xi^{i-3} + \ldots + \overline{a}_{r_{a,max}}\xi^{i-r_{a,max}} \\
=&\overline{a}_1\xi^{i-2} + \overline{a}_2\xi^{i-3} + \ldots + \overline{a}_{r_{a,max}-1}\xi^{i-r_{a,max}} + \overline{a}_{r_{a,max}}\xi^{i-r_{a,max}-1} \\
&\qquad\qquad \vdots
\end{aligned}
$$

After $u$ steps, with $u = i - r_{a,max} > 0$, we get

$$
\begin{aligned}
=&\overline{a}_1\xi^{i-u} + \overline{a}_2\xi^{i-u-1} + \ldots + \overline{a}_{r_{a,max}-1}\xi^{i-r_{a,max}-u+2} + \overline{a}_{r_{a,max}}\xi^{i-r_{a,max}-u+1} \\
=&\overline{a}_1(a_1\xi^{i-u-1} + a_2\xi^{i-u-2} + \ldots + a_{r_{a,max}-2}\xi^{i-r_{a,max}-u+2} \\
&+ a_{r_{a,max}-1}\xi^{i-r_{a,max}-u+1} + a_{r_{a,max}}) + \overline{a}_2\xi^{i-u-1} + \overline{a}_3\xi^{i-u-2} + \\
&\ldots + \overline{a}_{r_{a,max}-1}\xi^{i-r_{a,max}-u+2} + \overline{a}_{r_{a,max}}\xi^{i-r_{a,max}-u+1} \\
=&a_{r_{a,max}} + \overline{a}_1 \underbrace{\xi^{i-u-1}}_{\xi^{r_{a,max}-1}} + \overline{a}_2 \underbrace{\xi^{i-u-2}}_{\xi^{r_{a,max}-2}} + \ldots + \overline{a}_{r_{a,max}-2} \underbrace{\xi^{i-r_{a,max}-u+2}}_{\xi^2} \\
&+ \overline{a}_{r_{a,max}-1} \underbrace{\xi^{i-r_{a,max}-u+1}}_{\xi^1} \\
=&a_{r_{a,max}} + \overline{a}_1(a_1\xi^{r_{a,max}-2} + a_2\xi^{r_{a,max}-3} + \\
&\ldots + a_{r_{a,max}-3}\xi^2 + a_{r_{a,max}-2}\xi^1 + a_{r_{a,max}-1}) \\
&+ \overline{a}_2\xi^{r_{a,max}-2} + \ldots + \overline{a}_{r_{a,max}-2}\xi^2 + \overline{a}_{r_{a,max}-1}\xi^1 \\
=&a_{r_{a,max}} + a_{r_{a,max}-1} + \overline{a}_1\xi^{r_{a,max}-2} + \overline{a}_2\xi^{r_{a,max}-3} + \ldots + \overline{a}_{r_{a,max}-3}\xi^2 \\
&+ \overline{a}_{r_{a,max}-2}\xi^1 \\
=&\overline{a}_{r_{a,max}-1} + \overline{a}_1\xi^{r_{a,max}-2} + \overline{a}_2\xi^{r_{a,max}-3} + \ldots + \overline{a}_{r_{a,max}-3}\xi^2 + \overline{a}_{r_{a,max}-2}\xi^1 \\
&\qquad\qquad \vdots \\
&\overline{a}_{r_{a,max}-r_{a,max}+2} + \overline{a}_1\xi^1 = \overline{a}_2 + \overline{a}_1 = 1 \qquad q.e.d.
\end{aligned}
$$

The above conducted proof shows that the waiting time of one arbitrary customer is $1/t_{out}$ and that $R_a$ can be described by a closed solution obtained by Equation(4.41).

Next, we identify the influence of the input parameters on the waiting time if the capacity rule is used. We show that the mean waiting time depends only on the first moments of the arrival process and the collecting size $k$. In contrast, quantiles as the 95% and 99%-quantile depend on moments higher than the first.

### 4.4.2. Waiting Time Behavior under the Capacity Rule

The insights obtained in the previous subsection can be applied to the capacity rule as well. The number of remaining customers at the end of a collecting process, denoted by $R_y$, can also be calculated by a closed solution. Now, the considered renewal process is described by $y_i$, $i > 0,1,\ldots,y_{max}-1$. The residual lifetime of $y_i$ corresponds to the number of "unlucky" customers. Now, the observation of the residual lifetime takes places immediately before discrete time instants[6]. Thus, the values of the residual lifetime range from $i = 0,\ldots,y_{max}-1$ and we get analogous to Equation (4.41)

$$r_i = \frac{1}{E(Y)} \cdot (1 - \sum_{k=0}^{i} y_k), \quad i = 0,\ldots,y_{max}-1. \tag{4.49}$$

This fact leads to some interesting insights about the batch building process under the capacity rule (see Figure 4.11).

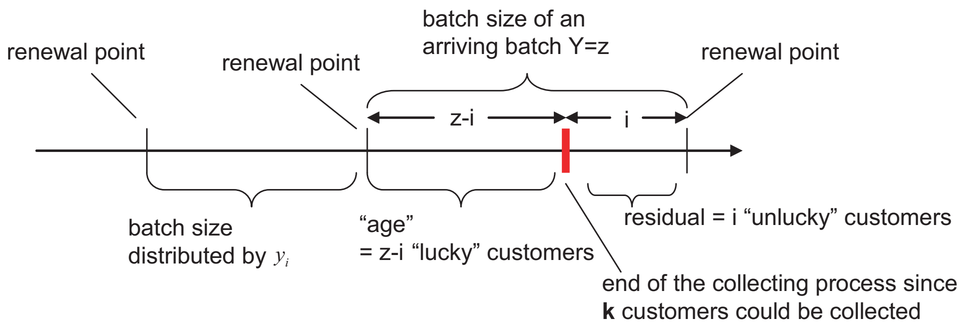


Figure 4.11.: Renewal process at the batch building process according to the capacity rule

If the residual lifetime distribution of a renewal process is known, the distribution of the age is also known (see Section 2.1.2). Since the residual lifetime is observed immediately before discrete time instants, it implies that the age is observed immediately after discrete time instants. Thus, we obtain the distribution of the age by

$$u_i = \frac{1}{E(Y)} \cdot (1 - \sum_{k=0}^{i-1} y_k), \quad i = 1,\ldots,y_{max}. \tag{4.50}$$

[6]Recall considering the timeout rule the observation of the residual lifetime takes place immediately after discrete time instants.

In our analysis of the collecting process under the capacity rule the age corresponds to the number of "lucky" customers who can be transferred immediately further to the succeeding station. Comparing the distribution of the "lucky" customers (see Equation (4.50)) with the "unlucky" customers (see Equation (4.49)), we recognize that both distributions are equal unless the value range is shifted by one:

$$\Rightarrow u_i = r_{i+1}. \tag{4.51}$$

Thus, it follows that the mean number of "lucky" customers is exactly one greater than the mean number of "unlucky" customers.

Subsequently, we show that $E(W)$ is only dependent on the collecting size $k$, the mean batch size $E(Y)$, and the mean interarrival time $E(A)$. The scv of the interarrival time $c_A^2$ and scv of the batch size $c_Y^2$ have no influence on $E(W)$. $E(W)$ leads to the expected work in progress (WIP) by use of Little's Law (Little (1961)) and is an important performance indicator for the collecting process.

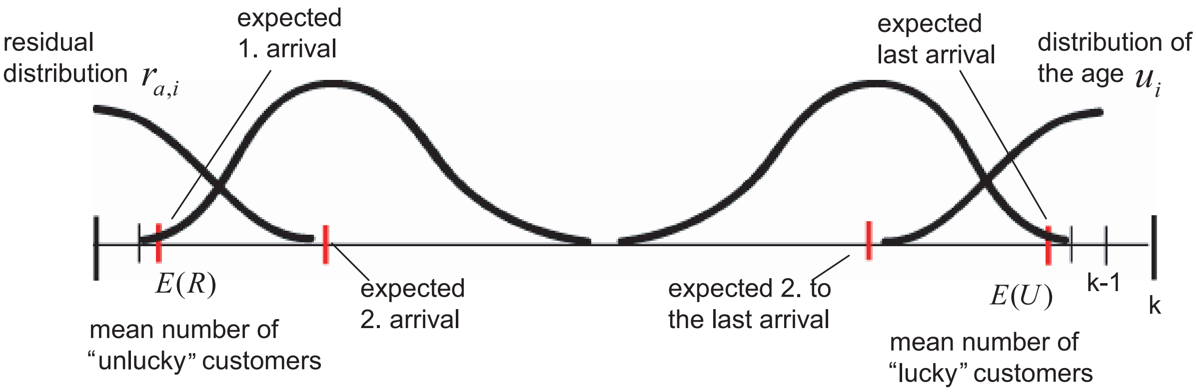


Figure 4.12.: Batch building under the capacity rule: Mean value analysis of the waiting time

The derivation of $E(W)$ is depicted in Figure 4.12. Since the number of "lucky" and "unlucky" customers are described by the same distribution unless the value range is shifted by one, the probability that the first arrival sees $i$, $i = 0, \ldots, y_{max} - 1$ customers at the collecting station is equal to the probability that the last arrival sees $k - 1 - i$, $i = 0, \ldots, y_{max} - 1$ customers. This fact and the mean number of customers observed by the first batch arrival, denoted by $E(R)$, and the mean number of customers observed by last batch arrival, denoted $E(U)$, are sketched out in Figure 4.12. Consequently, the probability that the second arrival sees $i$, $i = 1, \ldots, \min(2 \cdot y_{max} - 1, k - 1)$ customers is equal to the probability that the second last customer sees $k-1-i$ $i = 1, \ldots, \min(2 \cdot y_{max} - 1, k - 1)$ customers. This observation continues accordingly. Therefore, one arbitrary customer has to wait on an average for the arrival of $(k - 1)/2$ customers. Considering that the mean batch size of an arrival is $E(Y)$, we obtain the mean waiting time by

$$E(W) = \frac{(k-1)}{2 \cdot E(Y)} \cdot E(A). \tag{4.52}$$

This result is in accordance with the observation reported by Fowler et al. (2002) who present an expression for $E(W)$ for a collecting process with single arrivals.

Equation (4.52) shows that the variability of the arrival process, i.e. both the scv of the interarrival time, $c_A^2$, and of the batch size, $c_Y^2$, has no influence on the mean waiting time of a customer at the collecting station. However, for the performance of a logistic system it is crucial that a process can be performed on-time with a sufficiently high probability (e.g. 95%). Therefore, let us discuss the influence of unstable arrival processes, described by increasing values of $c_A^2$ and $c_Y^2$, on the 95% and 99%-quantile of the waiting time distribution.

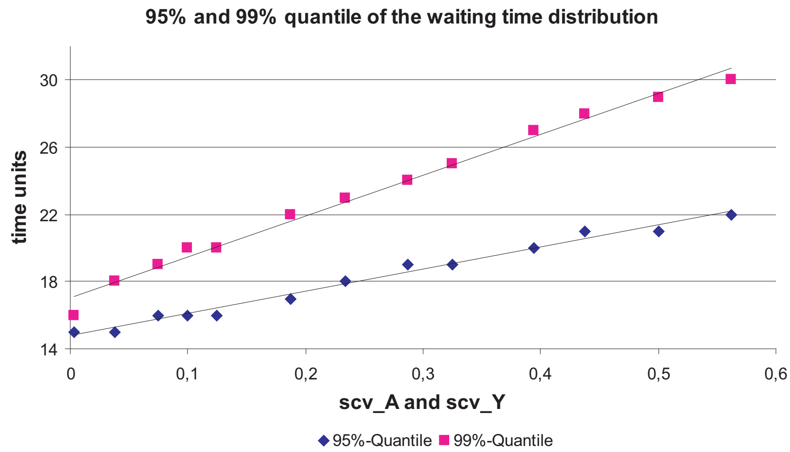


Figure 4.13.: The influence of the variability of the batch arrival process on the 95% and 99%-quantile of the waiting time distribution for a given example

We performed a set of experiments, where we increased simultaneously $c_A^2$ and $c_Y^2$. We observe that the 95% and 99%-quantile of the waiting time distribution increase with increasing $c_A^2$ and $c_Y^2$. The influence on the 99%-quantile is greater than on the 95%-quantile. This is depicted in Figure 4.13.[7] We conclude that an unstable process behavior of the arrival stream increases clearly the quantiles of the waiting time distribution. This means in turn that stable arrival processes with **low values of $c_A^2$ and $c_Y^2$ improve the probability of the on-time order fulfillment** of material flow systems.

### 4.4.3. Equivalence between the Basic Batch Building Modes: Capacity vs. Timeout Rule

In our analytical studies of the basic batch building algorithms, we identify some equivalences between the capacity rule and the timeout rule. Table 4.1 gives an overview of the equivalences between both basic batch building modes.

For the study of the capacity rule, we determine the number of arrivals for collecting $k$ customers, and for the study of the timeout rule, we determine the number of arrivals

[7]The input values for the different distributions of $a_i$ and $y_i$ can be found in Table A.1 in the Appendix.

| | Capacity Rule | | Timeout Rule |
|---|---|---|---|
| parameters | interarrival time $a_i$ | $\Leftrightarrow$ | batch size $y_i$ |
| | batch size $y_i$ | $\Leftrightarrow$ | interarrival time $a_i$ |
| | collecting size $k$ | $\Leftrightarrow$ | collecting time $t_{out}$ |
| results | waiting time $w_i$ $i = 0,1,\ldots$ | $\Leftrightarrow$ | number of still missing customers $(0,1,\ldots)$ |
| | number of still missing customer $(0,\ldots,k-1)$ | $\Leftrightarrow$ | waiting time $w_i$ $i = 0,\ldots,t_{out}$ |
| | time for the batch collecting process $b_i$ | $\Leftrightarrow$ | size of the collected batch $y_{d,i}$ |

Table 4.1.: Equivalence between the basic batch building modes: Capacity vs. timeout rule

within $t_{out}$ time units. In both cases, the number of arrivals obtained are results of a combinatorial solution using the batch size distribution in the first case and the interarrival time distribution in the other case (see Equations (4.8) and (4.22)). We remarked that the value ranges of $R_y$, $(0,...,y_{max} - 1)$ and $R_a$, $(1,...,a_{max})$ are slightly different. If we define the value range of $R_a$ by $(0,...,a_{max} - 1)$, the approach for calculating the number of arrivals of one collecting process would be analogous. However, we maintain the definition which we made for $R_a$ since this meets the real process description better. We conclude that in the case of the timeout rule $d_i$ corresponds to $y_{d,i}$ of the capacity rule and vice versa. We have only to exchange $k$ for $t_{out}$, $a_i$ for $y_i$ and $y_i$ for $a_i$, and we obtain the same result for $y_{d,i}$ as before for $d_i$. A numerical example is presented in Table 4.2, in which $e_i$ denotes the distribution of the number of still missing customers; the waiting time calculated using the capacity rule corresponds to the number of still missing customers calculated using the timeout rule; finally, the waiting time calculated using the timeout rule is equivalent to the number of still missing customers calculated using the capacity rule. For the timeout rule, we calculate $w_i$ by $1/t_{out}$ and for the capacity rule, we compute the number of still missing customers by $1/k$.

| | Capacity Rule | | | | | | | Timeout Rule | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | input | | | results | | | | input | | | results | | | |
| i | $a_i$ | $y_i$ | $k$ | $d_i$ | $y_{d,i}$ | $w_i$ | $e_i$ | $a_i$ | $y_i$ | $t_out$ | $d_i$ | $y_{dep,i}$ | $w_i$ | $e_i$ |
| 0 | 0 | 0 | 8 | 0.0000 | 0.0000 | 0.2063 | 0.1250 | 0 | 0 | 8 | 0.0000 | 0.0000 | 0.1250 | 0.2063 |
| 1 | 0.5 | 0.3 | | 0.0000 | 0.0000 | 0.1250 | 0.1250 | 0.3 | 0.5 | | 0.0000 | 0.0000 | 0.1250 | 0.1250 |
| 2 | 0.3 | 0.4 | | 0.0034 | 0.0000 | 0.1371 | 0.1250 | 0.4 | 0.3 | | 0.0000 | 0.0034 | 0.1250 | 0.1371 |
| 3 | 0.2 | 0.3 | | 0.0386 | 0.0000 | 0.1491 | 0.1250 | 0.3 | 0.2 | | 0.0000 | 0.0386 | 0.1250 | 0.1491 |
| 4 | | | | 0.0945 | 0.0000 | 0.1214 | 0.1250 | | | | 0.0000 | 0.0945 | 0.1250 | 0.1214 |
| 5 | | | | 0.1551 | 0.0000 | 0.0986 | 0.1250 | | | | 0.0000 | 0.1551 | 0.1250 | 0.0986 |
| 6 | | | | 0.1851 | 0.0000 | 0.0713 | 0.1250 | | | | 0.0000 | 0.1851 | 0.1250 | 0.0713 |
| 7 | | | | 0.1773 | 0.0000 | 0.0443 | 0.1250 | | | | 0.0000 | 0.1773 | 0.1250 | 0.0443 |
| 8 | | | | 0.1411 | 1.0000 | 0.0250 | | | | | 1.0000 | 0.1411 | | 0.0250 |
| 9 | | | | 0.0958 | | 0.0127 | | | | | | 0.0958 | | 0.0127 |
| 10 | | | | 0.0568 | | 0.0056 | | | | | | 0.0568 | | 0.0056 |
| 11 | | | | 0.0296 | | 0.0023 | | | | | | 0.0296 | | 0.0023 |
| 12 | | | | 0.0138 | | 0.0009 | | | | | | 0.0138 | | 0.0009 |
| 13 | | | | 0.0058 | | 0.0003 | | | | | | 0.0058 | | 0.0003 |
| 14 | | | | 0.0022 | | 0.0001 | | | | | | 0.0022 | | 0.0001 |
| 15 | | | | 0.0008 | | | | | | | | 0.0008 | | |
| 16 | | | | 0.0002 | | | | | | | | 0.0002 | | |
| 17 | | | | 0.0001 | | | | | | | | 0.0001 | | |

Table 4.2.: An example to illustrate the equivalence between the batch building mode under the capacity and the timeout rule; $e_i$ denotes the distribution of the number of still missing customers

In the following, we analyze the influence of the input parameters of a batch building operation on the variability of the departure process. The study of these influences supports the design of stable processes with low variabilities in material flow systems.

### 4.4.4. Departure Process Behavior

Recall that the description of the departure process of one node can be used as the arrival process for the succeeding node in a queueing network. This allows the analysis of queueing networks in which a collecting process is embedded. We will investigate the influence of the input parameters on the departure process, in particular with focus on the analysis of the stability of the departure processes described by the scv of the interdeparture time distribution and the scv of the collected batch size distribution.

Since we showed the equivalence between the capacity and the timeout rule, it is sufficient to analyze one of these methods. In the following, we choose the capacity rule.

By the use of an approximate expression for the determination of $c_D^2$, we can show that $c_Y^2$ has a greater influence on the stability of the departure process than $c_A^2$. If the timeout rule is chosen, this relation is exactly inverse. This should be considered for the design and planning of material flow systems in real applications.

It is obvious that $c_D^2$ is positively correlated with $c_A^2$ and $c_Y^2$.

First, we investigate the correlation between $c_D^2$ and $c_A^2$. We assume that there is a single arrival stream and therefore $c_Y^2 = 0$. It follows that a collecting process is always composed of $k$ arrivals and yields a collecting time distribution $d_i$ resulting from the $k$-fold convolution of $a_i$. Therefore, the scv of the interdeparture time distribution is obtained by

$$c_D^2 = \frac{c_A^2}{k}, \tag{4.53}$$

(see also Fowler et al. (2002)). Furthermore, in order to derive the dependence of $c_D^2$ on $c_Y^2$, we assume that $c_A^2 = 0$, so that $c_A^2$ has no influence on $c_D^2$. Recall the analysis done in Section 4.2.1 in which we concluded that $d_i$ can be derived from the probability that $N_a$ batches are required for a collecting operation (see Expressions (4.9), (4.10), (4.11) and (4.12)). If we assume that the interdeparture time is always one time unit, the departure process depends only on $y_i$, and $c_D^2$ corresponds to the scv of $P(N_a = l)$, denoted by $c_N^2$. We identified earlier that $P(N_a = l)$ depends on the distribution of the number of remaining customers at the beginning of each collecting process, $R_y$. The lower the mean number of arrivals required for one collecting process, $E(N)$, which can be calculated by

$$E(N) = \frac{k}{E(Y)}, \tag{4.54}$$

the greater is the influence of $R_y$ on $P(N_a = l)$. Since $R_y$ has influence on $c_D^2$ as well, it is not possible to derive easily an exact expression for the dependence of $c_D^2$ on $c_A^2$ as we did in Expression (4.53).

Subsequently, we present a simple approximation for which the calculation of $P(N_a = l)$ is not required and the quantitative influence of the input parameters on $c_D^2$ becomes

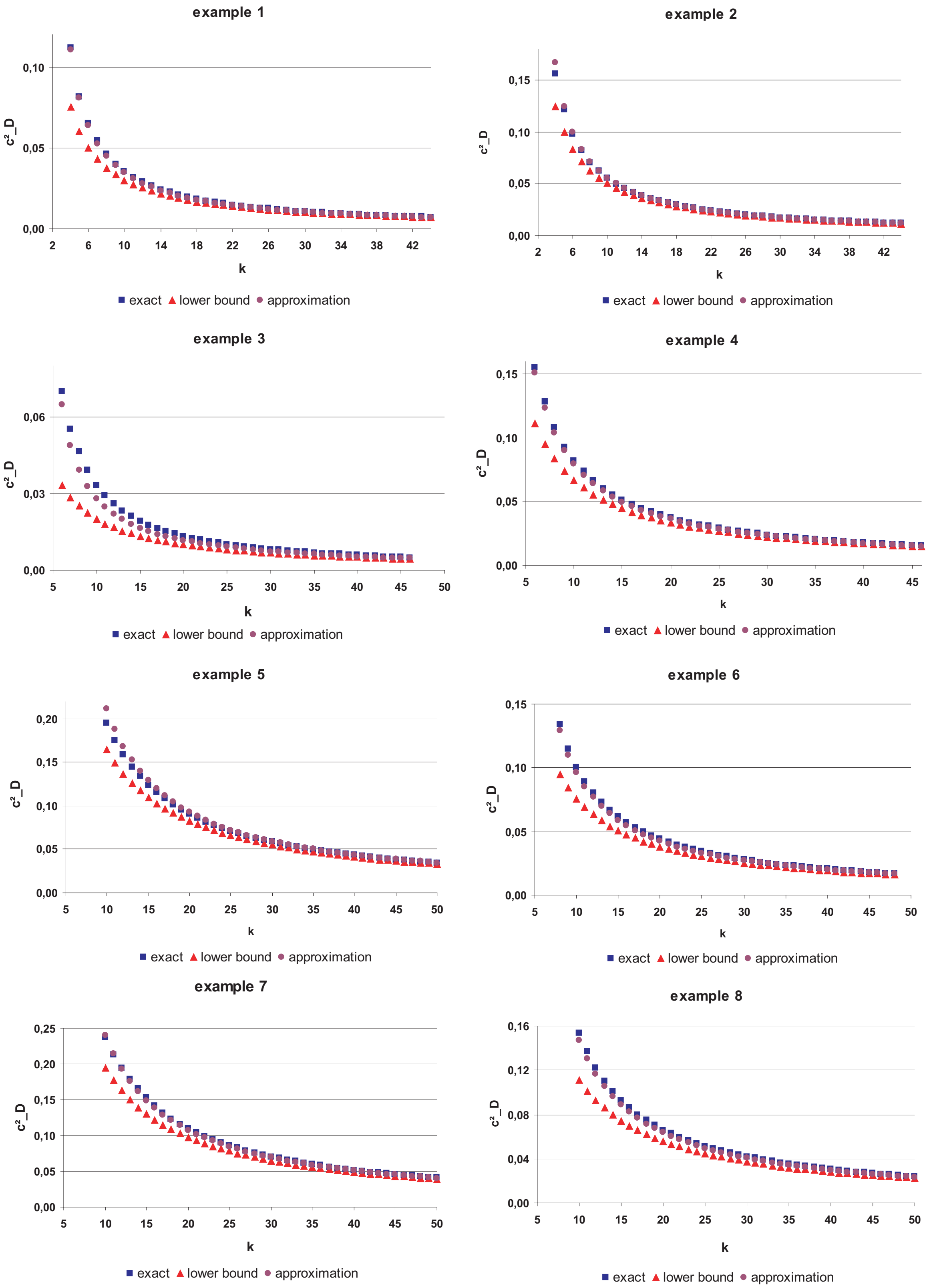


Figure 4.14.: $c_D^2$, depending on $k$; comparison of approximations to exact results; the input values are given in Table A.2 in the Appendix

nevertheless apparent. Since the greater is $k$ the greater is $E(N)$ (see Equation (4.54)). Thus, we conjecture that $c_D^2$ decreases with increasing $E(N)$ respectively $k$. This is analogous to the dependence of $c_D^2$ on $c_A^2$ as we showed in Equation (4.53). In a numerical study we identify Expression (4.55) as a lower bound[8]:

$$c_D^2 > \frac{c_Y^2}{E(N)} = \frac{c_Y^2 \cdot E(Y)}{k}. \tag{4.55}$$

In a set of numerical experiments we obtain an useful approximation by

$$c_D^2 \approx \frac{c_Y^2 \cdot E(Y)}{k - \frac{E(Y)}{\sqrt{2 \cdot c_Y^2}}}. \tag{4.56}$$

In Figure 4.14 we present different examples and illustrate the dependence of $c_D^2$ on $k$. The exact results computed by the approach presented previously and the results obtained by the approximation (4.56) are compared. The list with the input parameters for the presented examples can be found in Table A.2 in the Appendix.

Recall that if $d_i$ is computed, $P(N_a = l)$ is determined first which depends only on $y_i$. This causes the variability of the departure process induced by $y_i$. In the second calculation step, $P(N_a = l)$ is multiplied by the $l$-fold convolution of $a_i$. This induces the variability caused by $a_i$. The two steps are independent of each other and lead to an approximation for $c_D^2$ obtained by

$$c_D^2 \approx \frac{c_A^2 \cdot E(Y)}{k} + \frac{c_Y^2 \cdot E(Y)}{k - \frac{E(Y)}{\sqrt{2 \cdot c_Y^2}}}, \tag{4.57}$$

where we use approximation (4.56).

Finally, the quintessence of this section is the following: Since Equation (4.55) is a lower bound and the influence of $c_A^2$ on $c_D^2$ is independent of the influence of $c_Y^2$ on $c_D^2$, we conclude that the influence of $c_Y^2$ on the stability of the departure process is greater than the influence of $c_A^2$. This conclusion is reversed if the timeout rule is applied. Thus, the influence of $c_A^2$ on the stability of the departure process is greater than that of $c_Y^2$. This conclusion has to be kept in mind for the design of material flow systems.

## 4.5. Batch Building - Minimum Batch Size Rule

With the capacity and the timeout rule we investigated two basic batch building modes. In real material flow systems we can find various modifications of both basic batch building modes. For example, when applying the timeout rule, there could be a maximum capacity or a required minimum batch size. Equally, when applying the capacity rule, the collecting time could be bounded by a maximum time. Subsequently, we will analyze one of this modified batch building rules, where we choose the minimum batch size rule.

[8] It was not possible yet to proof mathematically Expression (4.55) as lower bound. However, in a great variety of experiments we did not find any counter-example which shows the opposite.

Analyzing all possible modifications of the presented basic batch building modes would go beyond the scope of this work. This is left open for future research. However, the procedure for deriving performance measures for all this modifications would be similar. So, the subsequent analysis could be taken as foundation for the study of further batch building modes.

The *minimum batch size rule* guarantees a minimum collecting size of $L$ customers and works as follows. The collecting process according to the minimum batch size rule lasts for at least $t_{out}$ time units. When $t_{out}$ ends and less than $L$ customers were collected, the batch building process continues until the required number of $L$ customers is attained. If at least $L$ customers were accumulated within $t_{out}$, the batch building process is completed and the collected batch is transferred immediately further. See Figure 4.15 for an illustration. The principle of the minimum batch size rule is introduced by Neuts (1967) who applies this rule to control a M/G$^{[L,K]}$/1-queue.

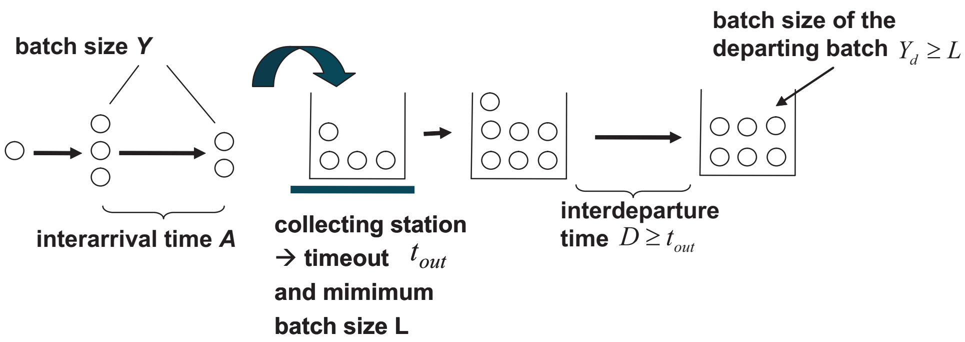


Figure 4.15.: Batch building: Collecting process according to the minimum batch size rule

The collecting process under the minimum batch size rule is a batching mode well-suited for transportation problems. It ensures a minimum utilization of the transport carrier or the transport vehicle. The higher the fixed costs of transportation, the higher is the targeted utilization. This leads to a value of $L$ close to $K$.

In addition to the description of material flow phenomena, the minimum batch size rule is applicable to problems in daily life. An example is the description of guided tours in museums or for exhibitions. A guide offers a museum tour which lasts $t_{out}$ time units. Within the guiding period of $t_{out}$ time units, the visitors arrive in stochastic time intervals and often in batches to attend the guided tour. When the time $t_{out}$ is elapsed, the guide checks if a minimum number of $L$ visitors have arrived. If not, he waits until a sufficient number of $L$ visitors is reached. If the guide chooses the minimum group size $L$ to be too low, the average number of visitors per tour will be low and the guide will have a poor yield of his efforts. If he chooses $L$ too high, the waiting time for the visitors will arise. This might lead to frustrated visitors with negative effects on the business of the museum.

The arrival process at the collecting station is, as in the preceding sections, described by the interarrival time distribution $a_i$ and the batch size distribution $y_i$. Both distributions are iid.

Subsequently, we perform a detailed time analysis of the batch building process under the minimum batch size rule and present an exact approach. The distributions of the size of the departing batch (Section 4.5.1), of the interdeparture time (Section 4.5.2) and of the waiting time of an arbitrary customer (Section 4.5.3) are determined. Section 4.5.4 shows the influence of the minimum batch size $L$ on the interdeparture time, the collected batch size and the waiting time. Furthermore, we detect a paradox: $E(W)$ can decrease with an increasing process instability. The reason for this phenomenon is explained. Finally, a simple example of an optimization problem is presented.

### 4.5.1. Batch Size Distribution of the Departing Batch

Let us discuss the $n$th batch building process. We consider the beginning of the collecting process. At this instant, the last batch arrival is dated back to at least zero and maximum $a_{max} - 1$ time units and the next batch arrival lies at least one and maximum $a_{max}$ time units ahead. The time between the start of the collecting process and the first batch arrival is a residual of the interarrival time interval. As in Section 4.3, this time period is called residual interarrival time and is denoted by the random variable $R_a$, distributed by $r_{a,i}$, $i = 1,\ldots,a_{max}$. If the $(n-1)$th collecting process ends with a batch arrival, the next batch arrives after a full interarrival time interval, described by the random variable $A$. This happens always if the batch building time is longer than $t_{out}$ time units, or if there is a batch arrival at the timeout instant and the amount of collected customers is $\geq L$. If the batch building time is $t_{out}$ time units and there is no batch arrival at the timeout instant, the next batch arrives in less than a full interarrival time interval but in at least one time unit.

Now, we analyze the number of collected customers within one collecting process, denoted by $Y_d$. Under the condition that $R_a = s$, $s = 1,\ldots,a_{max}$, we determine the number of arriving batches at the instant when $t_{out}$ is elapsed. $R_a = s$ implies that the first batch arrives after $s$ time units, measured from the beginning of the collecting process. If a batch arrives at the timeout instant of the $n$th collecting process, its customers are included in the $n$th collecting process. We obtain

$$P(N_a = l | R_a = s)_{t_{out}} = \sum_{i=0}^{t_{out}-s-1} a_{t_{out}-s-i}^{(l-1)\otimes} \sum_{j=i+1}^{a_{max}} a_j \quad \forall s = 1,\ldots,a_{max}, \tag{4.58}$$

where $N_a$ is the number of batch arrivals. Combining Equation (4.58) with $y_i$, we obtain the probability of $Y^{t_{out}} = x$ collected customers within $t_{out}$ time units under the condition that $R_a = s$ by

$$P(Y^{t_{out}} = x | R_a = s)_{t_{out}} = \sum_{l=0}^{l_{max}} y_x^{l\otimes} \cdot P(N_a = l | R_a = s)_{t_{out}} \quad \forall s = 1,\ldots,a_{max}, \quad l_{max} = \lceil t_{out}/a_{min} \rceil . \tag{4.59}$$

The collecting process is completed if at least $L$ customers are collected. If less than $L$ customers are collected within $t_{out}$, the collecting process continues. The probability that there are less than $L$ customers collected when $t_{out}$ is elapsed is

$$P(Y^{t_{out}} < L|R_a = s) = \sum_{x=0}^{L-1} P(Y^{t_{out}} = x|R_a = s) \quad \forall s = 1, \ldots, a_{max}. \tag{4.60}$$

For solving Equations (4.58), (4.59) and (4.60) we have to compute the distribution of $R_a$. Based on a discrete Markov chain (see Section 4.2.1 and 4.3.1), we solve the equilibrium equation system

$$P(R_a = s) = r_{a,s} = \sum_{u=1}^{a_{max}} p_{us} \cdot r_{a,u}, \tag{4.61}$$

where $p_{us}$ is the transition probability. We obtain

$$\begin{aligned} p_{us} = {} & P(Y^{t_{out}} \geq L|N_a = 1) \cdot a_{t_{out}-u+s} + \sum_{l=2}^{l_{max}} P(Y^{t_{out}} \geq L|N_a = l) \sum_{n=0}^{t_{out}-u-1} a^{(l-1)\otimes}_{t_{out}-u-n} \cdot a_{n+s} \\ & + P(Y^{t_{out}} < L|R_a = u) \cdot a_s \\ & \forall s = 1, \ldots, a_{max} \quad \forall u = 1, \ldots, a_{max}. \end{aligned} \tag{4.62}$$

For the derivation of Equation (4.62) we have to distinguish if at least $L$ customers can be collected within $t_{out}$ or not. The case that the minimum batch size $L$ is attained at the timeout instant is considered by the first two summands of Equation (4.62). The probability that at least $L$ customers are collected within $t_{out}$ on the basis of $l$ batch arrivals is given by

$$P(Y^{t_{out}} \geq L|N_a = l) = \sum_{x=0}^{y^{t_{out}}_{max}-L} y^{l\otimes}_{L+x} \quad \forall s = 1, \ldots, a_{max}, \quad y^{t_{out}}_{max} = l_{max} \cdot y_{max}. \tag{4.63}$$

The third summand of Equation (4.62) describes the case that less than $L$ customers are collected at the instant when the timeout is passed and the collecting process continues until the number of $L$ customers is attained. From this it follows that the collecting process ends with an arrival. Hence, the batch which completes this collecting process will be transported further immediately and the time interval to the next arrival is a complete interarrival time interval. The transition probability for the described case is computed by the product of $P(Y^{t_{out}} < L|R_a = u)$, given by Equation (4.60), and $a_s$, $s = 1, \ldots, a_{max}$.

Using Equations (4.61) and (4.59) leads to

$$P(Y^{t_{out}} = x) = \sum_{s=1}^{a_{max}} P(Y^{t_{out}} = x|R_a = s) \cdot r_{a,s}, \tag{4.64}$$

which is the probability that the number of collected customers at the timeout instant is $x$. If $x < L$, the collecting process continues until the required minimum number of $L$ customers is attained. Let us assume that exactly one customer is missing at the timeout instant. If the size of the succeeding arriving batch is $z$, it follows that the batch size of the collected batch is $L-1+z$. Generally, if there are $L-x$ customers missing, we have to consider possible combinations to attain the number $L$ by $q = 1,\dots,q_{max}$ additional arrivals. Thus, we derive the distribution of the batch size at the collecting station by

$$\begin{aligned} y_{d,L+z} &= P(Y^{t_{out}} = L+z) + \sum_{x=1}^{L-1} y_{L-x+z} \cdot P(Y^{t_{out}} = x) \\ &+ \sum_{i=1}^{y_{max}-z} \sum_{q=2}^{q_{max}} \sum_{x=1}^{L-1-i} y_{L-x-i}^{(q-1)\otimes} \cdot y_{i+z} \cdot P(Y^{t_{out}} = x) \quad \forall z = 0,\dots,y_{max}-1 \\ &\qquad q_{max} = \lceil (L - \frac{t_{out} \cdot y_{min}}{a_{max}})/y_{min} \rceil. \end{aligned} \tag{4.65}$$

After deriving the batch size distribution of the departing batch, we present in the next section an approach for computing the interdeparture time distribution at the collecting station.

### 4.5.2. Interdeparture Time Distribution

The interdeparture time $D$ corresponds to the collecting time. The fact that the collecting process lasts at least $t_{out}$ time units is trivial. The process lasts exactly $t_{out}$ time units if at least $L$ customers are collected at the timeout instant, otherwise it lasts more. If there are still $L-x$, $x = L-1,L-2,\dots,1$ customers missing at the timeout instant, the additional time to complete the collecting process has to be analyzed. For this analysis we introduce $R_a^{t_{out}}$, $s > 1,\dots,a_{max}$ which describes the residual interarrival time at the timeout instant (recall that $R_a$ is the residual interarrival time measured at the beginning of a collecting process). Both $R_a$ and $R_a^{t_{out}}$ are illustrated by means of Figure 4.16.

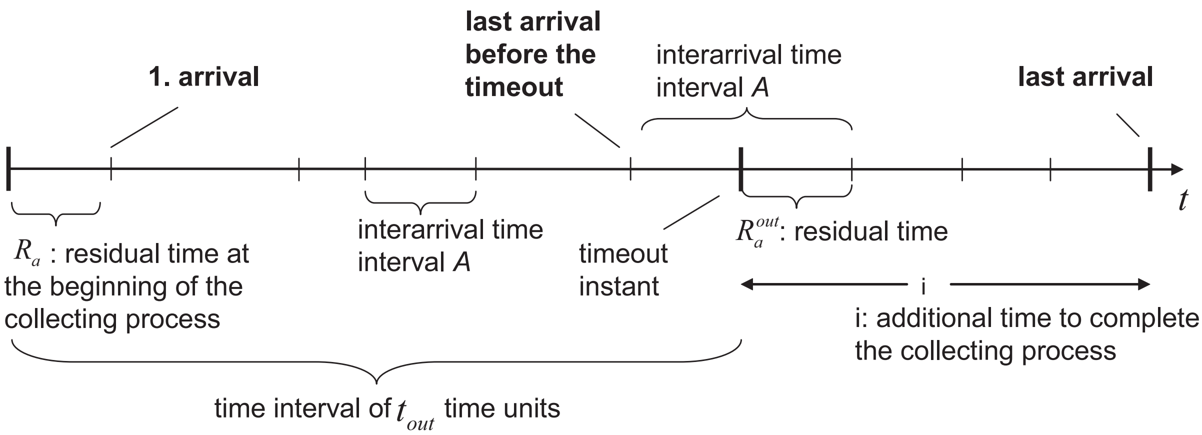


Figure 4.16.: Illustration of the residual times $R_a$ and $R_a^{t_{out}}$ at the collecting process under the minimum batch size rule

$P(R_a^{t_{out}} = s|Y_y^{t_{out}} = x)$ is the probability that the next batch arrives $s$ time units after the timeout instant, in the case that there are $Y_y = x$ customers collected at the timeout instant. If an arrival takes place at the timeout instant, the next arrival occurs in an interarrival time interval $A$. We compute the probability for the event that $R_a^{t_{out}} = s$ and that there are $Y^{t_{out}} = x$ customers collected at the timeout instant. Combinatorics yields

$$P(R_a^{t_{out}} = s \wedge Y^{t_{out}} = x) = \sum_{u=1}^{a_{max}} r_{a,u} \sum_{l=1}^{l_{max}} \sum_{j=1}^{t_{out}-u-1} a_{t_{out}-u-j}^{(l-1)\otimes} \cdot a_{s+j} \cdot y_x^{l\otimes} \quad \forall s > 1, \dots, a_{max}. \tag{4.66}$$

Furthermore, we obtain

$$P(R_a^{t_{out}} = s | Y^{t_{out}} = x) = \frac{P(R_a^{t_{out}} = s \wedge Y^{t_{out}} = x)}{P(Y^{t_{out}} = x)}. \tag{4.67}$$

Before calculating the interdeparture time $D$, we have to consider how many arrivals are required to complete the collecting process if $z = L - x$, $x = 1, \dots, L-1$ customers are still missing. The number of still missing arrivals is denoted by $N_c$. Therefore, we derive

$$P(N_c = q | Z = z) = \begin{cases} \sum_{i=0}^{y_{max}-z} y_{z+i} & \text{if } q = 1 \\ \sum_{j=1}^{z-1} \sum_{i=0}^{y_{max}-j} y_{z-j}^{(q-1)\otimes} \cdot y_{j+i} & \text{if } q > 1, \end{cases} \tag{4.68}$$

where $P(N_c = q | Z = z)$ is the probability that $N_c = q$, $q \geq 1$ arrivals are sufficient to complete the collecting process under the condition that $Z = z$, $z \geq 1$ customers are missing. Under the condition that $Y^{t_{out}} = L - z$ customers are present at the timeout instant, we obtain the interdeparture time using Equations (4.67) and (4.68) by

$$\begin{aligned} &P(D = t_{out} + i | Y^{t_{out}} = L - z) \\ &\quad = P(N_c = 1 | Z = z) \cdot P(R_a^{t_{out}} = i | Y^{t_{out}} = L - z) \\ &\quad + \sum_{q=2}^{z} \sum_{s=1, i>s}^{a_{max}} P(N_c = q | Z = z) \cdot a_{i-s}^{(q-1)\otimes} \cdot P(R_a^{t_{out}} = s | Y^{t_{out}} = L - z) \\ &\quad \forall i = 1, \dots \text{ and } \forall z = 1, \dots, L-1. \end{aligned} \tag{4.69}$$

The first term of Expression (4.69) considers the case that one arrival is sufficient to complete the collecting process and the second term describes the case that at least two arrivals are required. After solving Equation (4.69), it follows for the distribution of $D$, denoted by $d_i$, $i = t_{out}, t_{out}+1, \dots$,

$$d_{t_{out}} = \sum_{x=L}^{x_{max}} P(Y^{t_{out}} = x) \qquad x_{max} = \max\{L + y_{max} - 1, l_{max} \cdot y_{max}\}, \tag{4.70}$$

which describes the case that at least $L$ customers can be collected within $t_{out}$. If not, we get

$$d_{t_{out}+j} = \sum_{z=1}^{L-1} P(D = t_{out} + j | Y^{out} = L - z) \cdot P(Y^{t_{out}} = L - z) \quad \forall j = 1, \dots \tag{4.71}$$

Subsequently, the waiting time distribution for one arbitrary customer who arrives at the collecting station is derived.

### 4.5.3. Waiting Time Distribution

Using the minimum batch size rule we have to regard whether the batch building process ends after $t_{out}$ time units or continues until a collecting size of at least $L$ customers is reached. Let us consider the state at the timeout instant in which $Y^{t_{out}} = x$ customers are collected. If $x \geq L$, the collecting process is completed, otherwise it continues. The probability that an arbitrary customer is part of a batch building process which requires $q$ additional arrivals is denoted by $P(G_c = q)$, $q = 0,1,\ldots$. If $q = 0$, at least $L$ customers are present at the timeout instant and no additional arrival is required to complete the collecting process.

First, we determine the waiting time distribution for an arbitrary customer, assuming that he is part of a collecting process where $q$ additional arrivals are required. The number of batch arrivals within $t_{out}$ is denoted by $l$, $(l = 1,2,\ldots)$. Thus, we observe that $l + q$ batches arrive to form one collecting batch with the size of at least $L$. Considering one collecting process, the waiting time for the customers who arrive with batch $k$ $(1 \leq k \leq l + q)$ is denoted by $W^{k,l+q}$.

A process description of a batch building process according to the minimum batch size rule is shown in Figure 4.17. The waiting times of the arriving customers are illustrated. The $l$ batches which arrive before $t_{out}$ and the $q$ batches which arrive after $t_{out}$ are sketched. We use the same indexes $s$,$i$,$j$,$h$,$g$, and $z$ in this figure as subsequently in the equations.

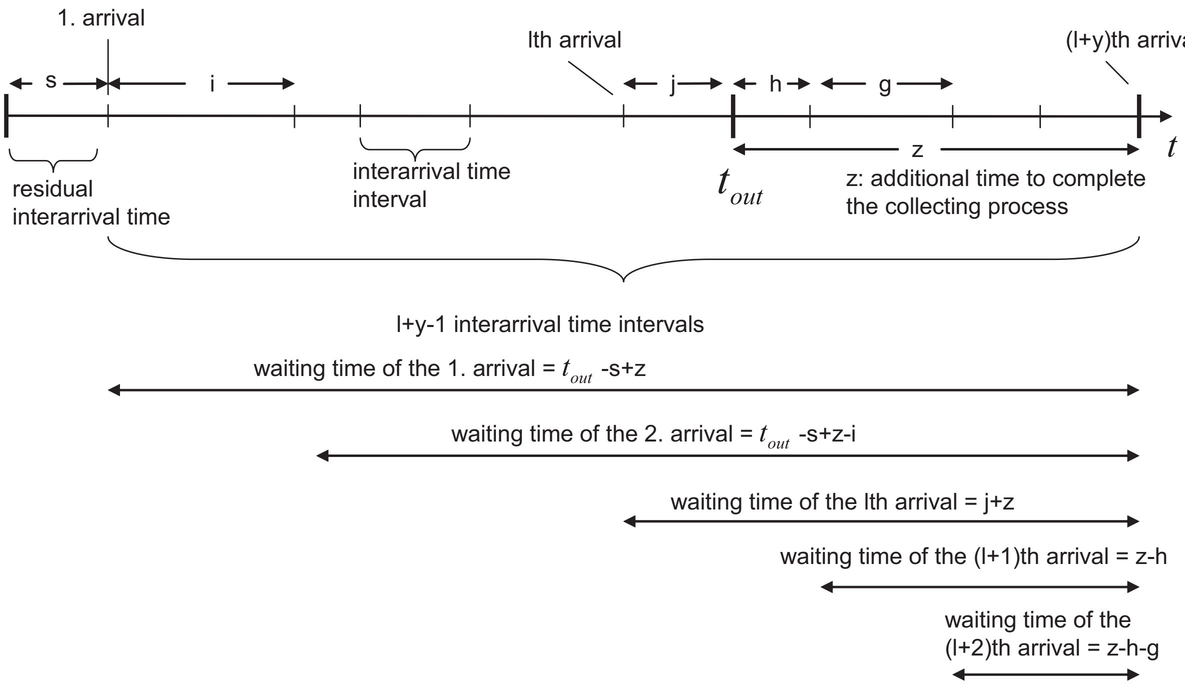


Figure 4.17.: Illustration of the waiting time if batch building according to the minimum batch size rule is applied

For the derivation of the waiting time distribution we choose a similar approach as in Section 4.3.2. Given a collecting time of $D = t_{out} + z$ with $z \geq 0$, we determine all the possibilities that a batch arrives at the time instant $i$ with $i = 1, \ldots, t_{out} + z$, which causes a waiting time of $t_{out} + z - i$. We have to regard the given constraints of a required minimum batch size and the fact that a residual interarrival time of length $s$ arises at the beginning of a collecting process. The waiting time for one arrival has to be weighted by its batch size. This finally leads to the waiting time of one individual customer.

Now, let us assume that $G_c = 0$ and that there are $l$ arrivals within $t_{out}$. If the batch size of the collected batch is $L + m$, the $L + m$ customers are uniformly distributed to the $l$ arrivals. The mean batch size of an arriving batch is $(L + m)/l$. The waiting time $W^{k,l}$ of the customers who arrive with the $k$th arrival ($k = 1, \ldots, l$) is proportional to the following expressions. For $l = 1$ we obtain

$$P(W^{1,1} = t_{out} - s | G_c = 0) \sim \sum_{s=1}^{a_{max}} r_{a,s} \cdot \overline{a}_{t_{out}-s+1}. \tag{4.72}$$

The residual interarrival time distribution, $r_{a,s}$, is calculated by Equation (4.61). Note that the first batch arrives always after $s$ time units (see also Figure 4.17). If $l \geq 2$, it yields

$$P(W^{k,l} = t_{out} - s | G_c = 0) \sim \sum_{s=1}^{a_{max}} r_{a,s} \sum_{j=0}^{t_{out}-s-1} a^{(l-1)\otimes}_{t_{out}-s-j} \cdot \overline{a}_{j+1} \quad \text{if } k = 1, \tag{4.73}$$

$$P(W^{k,l} = t_{out} - s - i | G_c = 0) \sim \sum_{s=1}^{a_{max}} r_{a,s} \sum_{j=0}^{a_{max}-1} \sum_{i=1}^{t_{out}-s-j-1} a_i^{(k-1)\otimes} \cdot a^{(l-k)\otimes}_{t_{out}-s-i-j} \cdot \overline{a}_{j+1} \quad \text{if } 2 \leq k \leq l-1 \tag{4.74}$$

and

$$P(W^{k,l} = j | G_c = 0) \sim \sum_{s=1}^{a_{max}} r_{a,s} \sum_{j=0}^{t_{out}-s-1} a^{(l-1)\otimes}_{t_{out}-s-j} \cdot \overline{a}_{j+1} \quad \text{if } k = l \tag{4.75}$$

Next, we assume that $G_c = 1$. This implies that, after the instant when the timeout is elapsed, one additional arrival is required to complete the collecting process. The collecting process is composed of $l + 1$ batch arrivals and lasts $t_{out+z}$ time units ($z \geq 1$).

If $l = 1$, there is exactly one arrival within $t_{out}$ and a total of two batch arrivals. Thus, the waiting time for the first arrival is $t_{out} - s + z$ and for the second it is zero. Due to a batch size of $L - n$ of the first arrival and $n + m$ of the second, the waiting times have to be weighted by the batch size. The waiting times are proportional to following expressions:

$$P(W^{1,2} = t_{out} - s + z | G_c = 1) \sim \sum_{s=1}^{a_{max}} r_{a,s} \sum_{z=1}^{t_{out}-s-1} a_{t_{out}-s+z} \left[ \sum_{n=1}^{L-1} \sum_{m=0}^{y_{max}-n} y_{L-n} \cdot y_{n+m} \cdot (L-n) \right] \tag{4.76}$$

and

$$P(W^{2,2} = 0|G_c = 1) \sim \sum_{s=1}^{a_{max}} \sum_{z=1}^{t_{out}-s-1} r_{a,s} \cdot a_{t_{out}-s+z} \left[\sum_{n=1}^{L-1} \sum_{m=0}^{y_{max}-n} y_{L-n} \cdot y_{n+m} \cdot (n+m)\right]. \tag{4.77}$$

If $l \geq 2$ and $G_c = 1$, the batches from 1 to $l$ have together a size of $L - n$ customers uniformly distributed to the $l$ batches. Furthermore, batch $l + 1$ has a size of $L + m$. The waiting time of the $k$th arrival ($k = 1, \ldots, l, l + 1$) has to be weighted by its batch size. It yields

$$P(W^{k,l+1} = t_{out} - s + z|G_c = 1) \sim \sum_{s=1}^{a_{max}} r_{a,s} \sum_{j=0}^{t_{out}-s-1} a_{t_{out}-s-j}^{(l-1)\otimes} \sum_{z=1}^{a_{max}-j} a_{j+z} \left[\sum_{n=1}^{L-1} \sum_{m=0}^{y_{max}-n} y_{L-n}^{l\otimes} \cdot y_{n+m} \cdot \frac{(L-n)}{l}\right] \quad \text{if } k = 1, \tag{4.78}$$

$$P(W^{k,l+1} = t_{out} - s + z - i|G_c = 1) \sim \sum_{s=1}^{a_{max}} r_{a,s} \sum_{j=0}^{a_{max}-1} \sum_{i=1}^{t_{out}-s-j-1} a_i^{(k-1)\otimes} \cdot a_{t_{out}-s-j}^{(l-k)\otimes} \sum_{z=1}^{a_{max}-z} a_{j+z} \left[\sum_{n=1}^{L-1} \sum_{m=0}^{y_{max}-n} y_{L-n}^{l\otimes} \cdot y_{n+m} \cdot \frac{(L-n)}{l}\right] \quad \text{if } 2 \leq k \leq l-1, \tag{4.79}$$

$$P(W^{k,l+1} = j + z|G_c = 1) \sim \sum_{s=1}^{a_{max}} r_{a,s} \sum_{j=0}^{t_{out}-s-1} a_{t_{out}-s-j}^{(l-1)\otimes} \sum_{z=1}^{a_{max}-j} a_{j+z} \left[\sum_{n=1}^{L-1} \sum_{m=0}^{y_{max}-n} y_{L-n}^{l\otimes} \cdot y_{n+m} \cdot \frac{(L-n)}{l}\right] \quad \text{if } k = l \tag{4.80}$$

and

$$P(W^{k,l} = 0|G_c = 1) \sim \sum_{s=1}^{a_{max}} r_{a,s} \sum_{j=0}^{t_{out}-s-1} a_{t_{out}-s-j}^{(l-1)\otimes} \sum_{z=1}^{a_{max}-j} a_{j+z} \left[\sum_{n=1}^{L-1} \sum_{m=0}^{y_{max}-n} y_{L-n}^{l\otimes} \cdot y_{n+m} \cdot (n+m)\right] \quad \text{if } k = l+1 \tag{4.81}$$

Finally, we assume that $G_c = q$, $q \geq 2$. We observe a total of $l + q$ batch arrivals. The batches from 1 to $l+q-1$ have together a size of $L-n$ customers uniformly distributed to the $l+q-1$ batches. The last arriving batch, batch $l+1$, has a size of $L+m$. As we mentioned before, the waiting time of the $k$th arrival ($k = 1, \ldots, l+q$) has to be weighted by its size.

If $l = 1$, we obtain

$$P(W^{1,1+q} = t_{out} - s + z | G_c = q) \sim \sum_{s=1}^{a_{max}} r_{a,s} \sum_{h=1}^{t_{out}-s-1} a_{t_{out}-s+h} \sum_{z=h+1}^{z_{max}} a_{z-h}^{(q-1)\otimes} \left[ \sum_{n=1}^{L-1} \sum_{m=0}^{y_{max}-n} y_{L-n}^{q\otimes} \cdot y_{n+m} \cdot \frac{(L-n)}{q} \right] \tag{4.82}$$

$$z_{max} = d_{max} - t_{out},$$

$$P(W^{2,1+q} = z - h | G_c = q) \sim \sum_{s=1}^{a_{max}} r_{a,s} \sum_{h=1}^{t_{out}-s-1} a_{t_{out}-s+h} \sum_{z=h+1}^{z_{max}} a_{z-h}^{(q-1)\otimes} \left[ \sum_{n=1}^{L-1} \sum_{m=0}^{y_{max}-n} y_{L-n}^{q\otimes} \cdot y_{n+m} \cdot \frac{(L-n)}{q} \right], \tag{4.83}$$

$$P(W^{k,1+q} = z - h - g | G_c = q) \sim \sum_{s=1}^{a_{max}} r_{a,s} \sum_{h=1}^{t_{out}-s-1} a_{t_{out}-s+h} \sum_{g=1}^{g_{max}} a_{g}^{(k-l-1)\otimes} \sum_{z=h+g+1}^{z_{max}} a_{z-h-g}^{(q+l-k)\otimes} \left[ \sum_{n=1}^{L-1} \sum_{m=0}^{y_{max}-n} y_{L-n}^{q\otimes} \cdot y_{n+m} \cdot \frac{(L-n)}{q} \right] \tag{4.84}$$

$$\text{if } l+2 \leq k \leq l+q-1 \quad g_{max} = z_{max} - 1$$

and[9]

$$P(W^{l+q,1+q} = 0 | G_c = q) \sim \sum_{s=1}^{a_{max}} r_{a,s} \sum_{h=1}^{t_{out}-s-1} a_{t_{out}-s+h} \sum_{z=h+1}^{z_{max}} a_{z-h}^{(q-1)\otimes} \left[ \sum_{n=1}^{L-1} \sum_{m=0}^{y_{max}-n} y_{L-n}^{q\otimes} \cdot y_{n+m} \cdot (n+m) \right] \tag{4.85}$$

Furthermore, if $q \geq 2$ and $l \geq 2$, we get the following expressions:

$$P(W^{k,l+q} = t_{out} - s + z | G_c = q) \sim \sum_{s=1}^{a_{max}} r_{a,s} \sum_{j=0}^{t_{out}-s-1} \sum_{h=1}^{a_{max}-j} a_{t_{out}-s+j}^{(l-1)\otimes} \cdot a_{j+h} \sum_{z=h+1}^{z_{max}} a_{z-h}^{(q-1)\otimes} \left[ \sum_{n=1}^{L-1} \sum_{m=0}^{y_{max}-n} y_{L-n}^{q\otimes} \cdot y_{n+m} \cdot \frac{(L-n)}{q+l-1} \right] \tag{4.86}$$

$$\text{if } k = 1,$$

[9] $z_{max}$ denotes the maximum possible time elapsed from the timeout instant to the end of the collecting process

$$P(W^{k,l+q} = t_{out} - s + z - i | G_c = q) \sim \sum_{s=1}^{a_{max}} r_{a,s} \sum_{j=0}^{a_{max}-1} \sum_{i=1}^{t_{out}-s-j-1} a_i^{(k-1)\otimes} \sum_{h=1}^{a_{max}-j} a_{t_{out}-s-i-j}^{(l-k)\otimes} \cdot a_{j+h} \sum_{z=h+1}^{z_{max}} a_{z-h}^{(q-1)\otimes} \left[ \sum_{n=1}^{L-1} \sum_{m=0}^{y_{max}-n} y_{L-n}^{q\otimes} \cdot y_{n+m} \cdot \frac{(L-n)}{q+l-1} \right] \quad \text{if } 2 \leq k \leq l-1, \tag{4.87}$$

$$P(W^{k,l+q} = j + z | G_c = q) \sim \sum_{s=1}^{a_{max}} r_{a,s} \sum_{j=0}^{t_{out}-s-1} \sum_{h=1}^{a_{max}-j} a_{t_{out}-s+j}^{(l-1)\otimes} \cdot a_{j+h} \sum_{z=h+1}^{z_{max}} a_{z-h}^{(q-1)\otimes} \left[ \sum_{n=1}^{L-1} \sum_{m=0}^{y_{max}-n} y_{L-n}^{q\otimes} \cdot y_{n+m} \cdot \frac{(L-n)}{q+l-1} \right] \quad \text{if } k = l, \tag{4.88}$$

$$P(W^{k,l+q} = z - h | G_c = q) \sim \sum_{s=1}^{a_{max}} r_{a,s} \sum_{j=0}^{t_{out}-s-1} \sum_{h=1}^{a_{max}-j} a_{t_{out}-s+j}^{(l-1)\otimes} \cdot a_{j+h} \sum_{z=h+1}^{z_{max}} a_{z-h}^{(q-1)\otimes} \left[ \sum_{n=1}^{L-1} \sum_{m=0}^{y_{max}-n} y_{L-n}^{q\otimes} \cdot y_{n+m} \cdot \frac{(L-n)}{q+l-1} \right] \quad \text{if } k = l+1, \tag{4.89}$$

$$P(W^{k,l+q} = z - h - g | G_c = q) \sim \sum_{s=1}^{a_{max}} r_{a,s} \sum_{j=0}^{t_{out}-s-1} \sum_{h=1}^{a_{max}-j} a_{t_{out}-s+j}^{(l-1)} \cdot a_{j+h} \sum_{g=1}^{g_{max}} a_g^{(k-l-1)\otimes} \sum_{z=h+g+1}^{z_{max}} a_{z-h-g}^{(q+l-k)\otimes} \left[ \sum_{n=1}^{L-1} \sum_{m=0}^{y_{max}-n} y_{L-n}^{q\otimes} \cdot y_{n+m} \cdot \frac{(L-n)}{q+l-1} \right] \quad \text{if } l+2 \leq k \leq l+q-1 \tag{4.90}$$

and

$$P(W^{k,l+q} = 0 | G_c = q) \sim \sum_{s=1}^{a_{max}} r_{a,s} \sum_{j=0}^{t_{out}-s-1} \sum_{h=1}^{a_{max}-j} a_{t_{out}-s+j}^{(l-1)\otimes} \cdot a_{j+h} \sum_{z=h+1}^{z_{max}} a_{z-h}^{(q-1)\otimes} \left[ \sum_{n=1}^{L-1} \sum_{m=0}^{y_{max}-n} y_{L-n}^{q\otimes} \cdot y_{n+m} \cdot (n+m) \right] \quad \text{if } k = l+q \tag{4.91}$$

We derived the proportional values for $P(W^{k,l+q} = i | G_c = q)$, $i = 0,1,2,\ldots$ by the Expressions from (4.72) to (4.91). Let us denote these expressions which are proportional to $P(W^{k,l+q} = i|G_c = q)$ by $P^*(W^{k,l+q} = i|G_c = q)$. Thus, we normalize and compute

$$P(W = i|G_c = q) = \frac{\sum_{l=1}\sum_{k=1}^{l+q} P^*(W^{k,l+q} = i|G_c = q)}{\sum_{i=0}\sum_{l=1}\sum_{k=1}^{l+q} P^*(W^{k,l+q} = i|G_c = q)}. \tag{4.92}$$

By use of Equation (4.92) we calculate the probability that an arbitrary customer has to wait $i$ time units under the condition that he is part of a batch building process requiring $q$ additional arrivals to fulfill a minimum batch size of $L$. We recall that the probability that an arbitrary customer is part of a batch building process requiring $q$ additional arrivals is denoted by $P(G_c = q)$. For $q = 0$, we obtain

$$\begin{aligned} P(G_c = q) =& \frac{1}{E(Y_d)} \sum_{s=1}^{a_{max}} r_{a,s} \Bigg( \overline{a}_{t_{out}-s+1} \sum_{m=0, y_{L+m} \le y_{max}}^{y_{max}-L} y_{L+m} \cdot (L+m) \\ &+ \sum_{l=2}^{l_{max}} \sum_{j=0}^{t_{out}-s-1} a_{t_{out}-s-j}^{(l-1)\otimes} \cdot \overline{a}_{j+1} \sum_{m=0}^{y_{max}-1} y_{L+m}^{l\otimes} \cdot (L+m) \Bigg) \\ & \text{if } q = 0 \end{aligned} \tag{4.93}$$

The first summand of Equation (4.93) represents the case that there is exactly one arrival within $t_{out}$ and the second one represents the case that there are more than one. Recall if $q = 0$, at least $L$ customers arrive within $t_{out}$. The greater the collected batch size $L+m$ ($m = 0,1,\ldots$) the greater is the probability that an arbitrary customer is part of a batch building process requiring $q$ additional arrivals. Since $P(G_c = q)$ is proportional to the collected batch size $L+m$, we have to weight Expression (4.93) by the batch size of the collected batch $L+m$ in relation to $E(Y_d)$. Analogous to Equation (4.93), we derive the case for $q > 0$. It yields

$$\begin{aligned} P(G_c = q) =& \frac{1}{E(Y_d)} \sum_{s=1}^{a_{max}} r_{a,s} \Bigg( \overline{a}_{t_{out}-s+1} \sum_{n=1}^{L-1} \sum_{m=0}^{y_{max}-n} y_{L-n}^{q\otimes} \cdot y_{n+m} \cdot (L+m) \\ &+ \sum_{l=2}^{l_{max}} \sum_{j=0}^{t_{out}-s-1} a_{t_{out}-s-j}^{(l-1)\otimes} \cdot \overline{a}_{j+1} \sum_{n=1}^{L-1} \sum_{m=0}^{y_{max}-n} y_{L-n}^{(l+q-1)\otimes} \cdot y_{n+m} \cdot (L+m) \Bigg) \\ & \text{if } q \ge 1. \end{aligned} \tag{4.94}$$

Finally, we obtain the waiting time distribution by

$$w_i = P(W = i) = \sum_{q=0}^{q_{max}} P(W = i|G_c = q) \cdot P(G_c = q). \tag{4.95}$$

After deriving an exact approach for the analysis of the batch building process under the minimum batch size rule, we will present some numerical results in the following section.

### 4.5.4. Analysis

The derived method for the minimum batch size rule allows us to investigate the influence of the input parameters on the output. Thus, we study the influence of $L$ on the

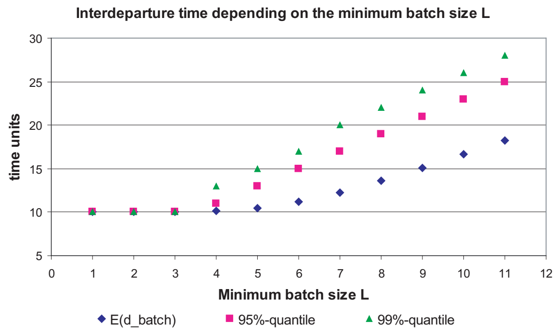


Figure 4.18.: Dependency of the interdeparture time on the minimum batch size L

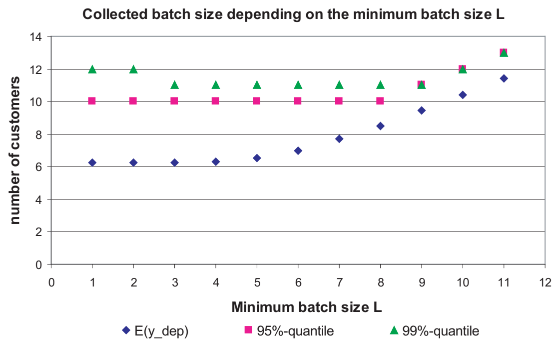


Figure 4.19.: Dependency of the collected batch size on the minimum batch size L

interdeparture time $D$, the collected batch size $Y_d$ and the waiting time $W$. The results are presented in Figures 4.18, 4.19 and 4.21. The input parameters and the obtained numerical results are printed in Table A.8 in the Appendix. If $L$ is chosen to be small enough, the collecting process ends always after $t_{out}$ time units. This corresponds to the timeout rule (see Section 4.3). This is the case in our example for $L = 1$ and $L = 2$.

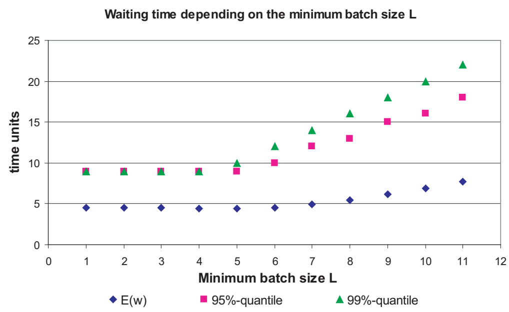


Figure 4.20.: Dependency of the waiting time on the minimum batch size L

It is interesting to note that the minimum mean waiting time is attained for a minimum batch size of $L = 5$ (see also the numerical results in the Appendix: Table A.8). This can be explained as follows: It is clear that greater $L$ results in a longer duration of the collecting process leading to an increasing waiting time. However, if we increase $L$ we increase also the proportion of customers having a waiting time of zero. This influence on $E(W)$ is weak, but leads to the fact that $E(W)$ is not minimal for a minimum size of $L$.

### The Paradox of Decreasing $E(W)$ with Increasing Process Instability

In a further experiment, we analyze the influence of the variability of the batch arrival process on performance measures. We study the behavior of $E(W)$, $\sigma_{w,0.95}$, and $\sigma_{w,0.99}$ depending on both $c_A^2$ and $c_Y^2$.[10] In the analyzed example, we choose a minimum batch

[10]The first and third moment of $a_i$ and $y_i$ were held constant. Thus, symmetrical distributions are

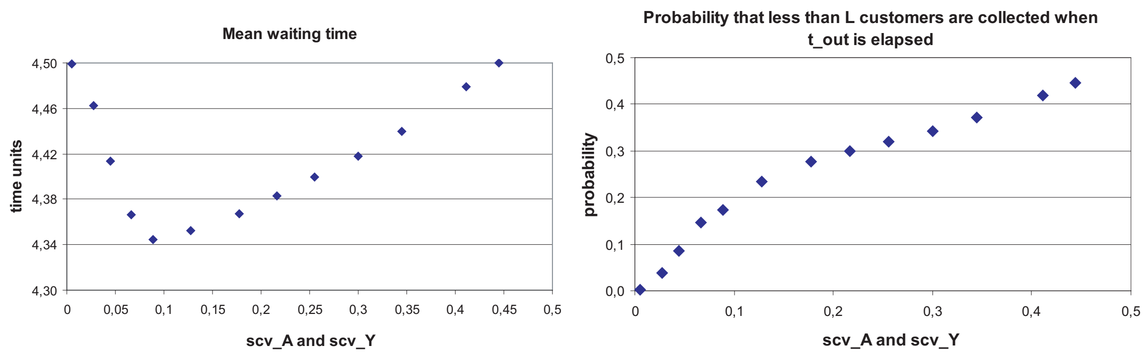


Figure 4.21.: Influence of the variability of the batch arrival process on performance measures: Left: Mean waiting time depending on $c_A^2$ and $c_Y^2$; right: 95% and 99%-quantile depending on depending on $c_A^2$ and $c_Y^2$

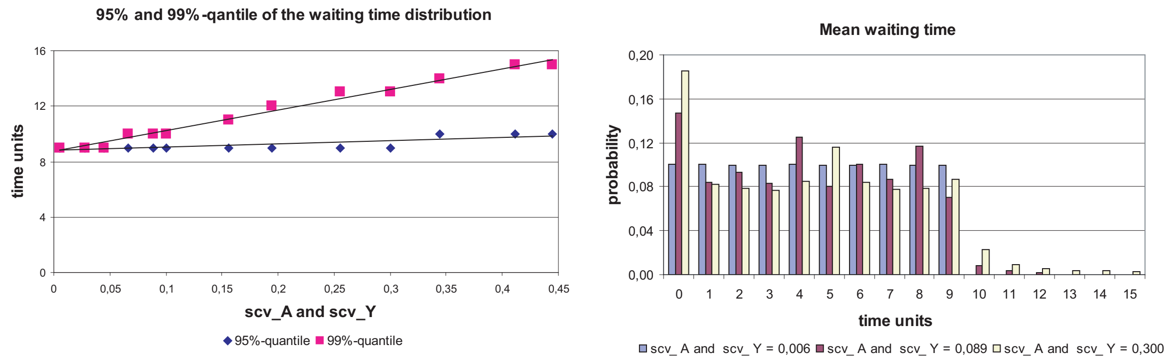


Figure 4.22.: Left: Probability that less than $L$ customers are collected when $t_{out}$ is elapsed depending on $c_A^2$ and $c_Y^2$; right: Waiting time distribution for different configurations of $c_A^2$ and $c_Y^2$

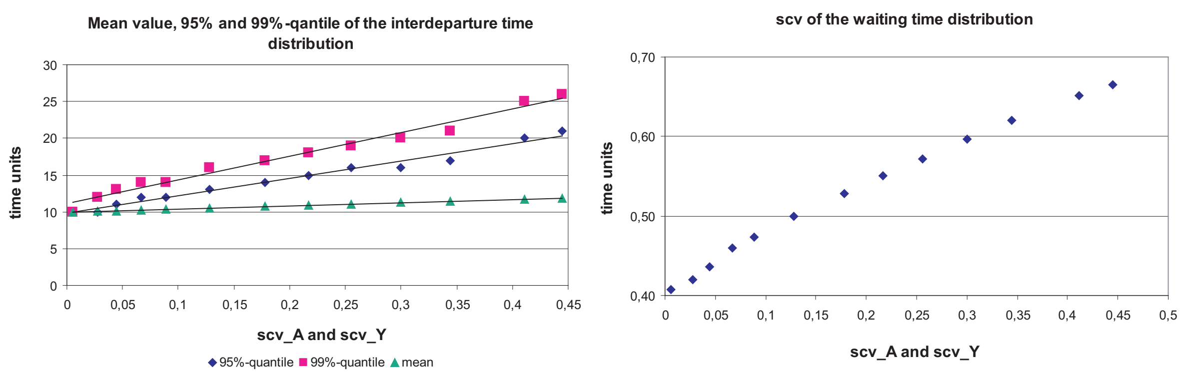


Figure 4.23.: Left: $c_W^2$ depending on $c_A^2$ and $c_Y^2$; right: Mean value, 95% and 99%-quantile of the interdeparture time depending on $c_A^2$ and $c_Y^2$

size of $L = 8$ and a timeout of $t_{out} = 10$. The graph on the left in Figure 4.21 shows $E(W)$ depending on both $c_A^2$ and $c_Y^2$. It is astonishing to note that $E(W)$ does not increase with an increasing variability within a certain range of $c_A^2$ and $c_Y^2$. In contrast, it is trivial to note that the probability that less than $L$ customers are collected when $t_{out}$ is elapsed increases with increasing $c_A^2$ and $c_Y^2$ (see right graph of Figure 4.21). This probability corresponds to $1 - d_{t_{out}}$ and in the chosen example it is almost zero for low values of $c_A^2$ and $c_Y^2$.

Having the system's behavior of the G/G/1-queue in mind, where increasing process instability causes always increasing $E(W)$, the phenomenon that $E(W)$ decreases with an increasing variability regarding the collecting process under the minimum batch size rule seems to be paradoxical. Let us explain this paradox by the following consideration:

First, let us consider the case that at least $L$ customers can be collected within $t_{out}$. This corresponds to the timeout rule discussed previously. In the considered numerical example, this situation is very probable for low values of $c_A^2$ and $c_Y^2$. Applying the timeout rule we know that $E(W)$ is computed by $(t_{out} - 1)/2$, that means 4.5 time units in the considered numerical case. The fact that $w_i = 1/t_{out}$, $i = 0,1,\ldots$ is illustrated in the graph on the right in Figure 4.22 for the case that $c_A^2$ and $c_Y^2$ is 0.0056.

If $c_A^2$ and $c_Y^2$ increases slightly, the probability that there are less then $L$ customers present when $t_{out}$ is elapsed increases as well (see the graph on the right in Figure 4.21). If there are less then $L$ customers at $t_{out}$, it is highly probable that one additional arrival is sufficient to terminate the collecting process. In this situation, $E(Y)$ customers arrive in the mean in order to complete the collecting process and they have all a waiting time of zero. Roughly considered, these customers arrive $E(A)/2$ time units after $t_{out}$. Thus, all these customers who have arrived before $t_{out}$ have an additional waiting time of $E(A)/2$, and in the considered numerical case this is 1.5 time units. In the worst case, $L - 1$ customers have to wait for an additional arrival. Since we expect $E(Y)$ customers with the subsequent arrival, the mean waiting time is given by

$$E(W) = \frac{(L-1)}{(L-1+E(Y))} \cdot \frac{(t_{out} - 1 + E(A))}{2}. \tag{4.96}$$

In order to show that $E(W)$ decreases, Equation (4.96) has to be smaller then $(t_{out} - 1)/2$. After some algebra, we obtain the following condition that $E(W)$ decreases with increasing $c_A^2$ and $c_Y^2$:

$$\frac{E(A)}{E(Y)} < \frac{t_{out} - 1}{L - 1} \approx \frac{t_{out}}{L} \Rightarrow \frac{t_{out} \cdot E(Y)}{E(A)} > L. \tag{4.97}$$

However, if $c_A^2$ and $c_Y^2$ exceed a certain threshold, $E(W)$ increases since the probability that more than at least two additional arrivals are required to finish a collecting process increases.

In summary, it can be ascertained that $E(W)$ can decrease with increasing variability of the arrival process within a certain range of $c_A^2$ and $c_Y^2$. Thus, $E(W)$ can be optimized depending on $c_A^2$ and $c_Y^2$. If the probability that there are less then $L$ customers available

chosen (see Table A.4 in the Appendix). The influence of the fourth and higher moments are not considered.

when $t_{out}$ is elapsed is high for low values of $c_A^2$ and $c_Y^2$, we observe no decreasing behavior of $E(W)$ depending on $c_A^2$ and $c_Y^2$. In this case, the number of arrivals within $t_{out}$, roughly given by $t_{out} \cdot E(Y)/E(A)$ is clearly less then $L$. In this case, the condition set up in Expression (4.97) is violated.

In contrast to the observation above, if we analyze “high percentage”-quantiles such as $\sigma_{0.95}$ and $\sigma_{0.99}$ depending on $c_A^2$ and $c_Y^2$, we note that $\sigma_{0.95}$ and $\sigma_{0.99}$ increases nearly linearly with increasing $c_A^2$ and $c_Y^2$ (see the left graph in Figure 4.22). Since the left graph of Figure 4.23 shows that the interdeparture time gets longer with increasing $c_A^2$ and $c_Y^2$, the probability to observe long waiting times increases. Furthermore, the scv of the waiting time distribution increases with $c_A^2$ and $c_Y^2$ too (see right graph of Figure 4.23). These contribute to the fact that increasing $c_A^2$ and $c_Y^2$ leads to increasing $\sigma_{0.95}$ and $\sigma_{0.99}$. Therefore, the responsible planers of material flow systems have to be aware that planning decisions based solely on $E(W)$ can be misleading in some cases. High performance material flow systems with short order sojourn times targets and a high degree of on-time order fulfillment can only be achieved by stable processes. In order to optimize the decision making process a thorough understanding of the stochastic system's behavior is required.

In order to conclude this discussion, three waiting time distributions for different values of $c_A^2$ and $c_Y^2$ are depicted in in the right graph of Figure 4.22. As above mentioned, if $c_A^2$ and $c_Y^2$ are very low, the system's behavior corresponds to the timeout rule and $w_i = 1/t_{out}$, $i = 0,1,\ldots$. If $c_A^2$ and $c_Y^2$ increases, the probability that the waiting time is zero and the probability that long waiting times arise increases too.

**Application: Optimization of the Batch Building Operation**

Subsequently, we present an optimization model for a basic transportation system which can be found in material handling systems and runs under the minimum batch size rule.

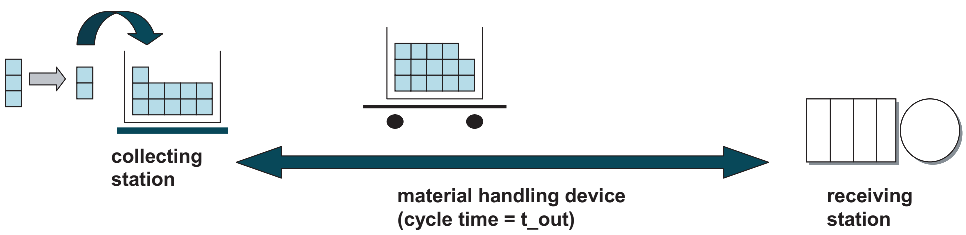


Figure 4.24.: Basic transportation system: A material handling device picks up collected material units at the collecting station to transfer them to the receiving station

Given is a material handling device which picks up the collected material units at a collecting station to transfer them to their destination station. The transportation time of the material handling device is $t_{out}$ time units. The material handling device, returning after $t_{out}$ time units, waits if less than $L$ material units are collected. We assume that the capacity of the material handling device is sufficient. The batch arrival stream at the collection station is given by $A$ and $Y$. The described system is illustrated by Figure 4.24.

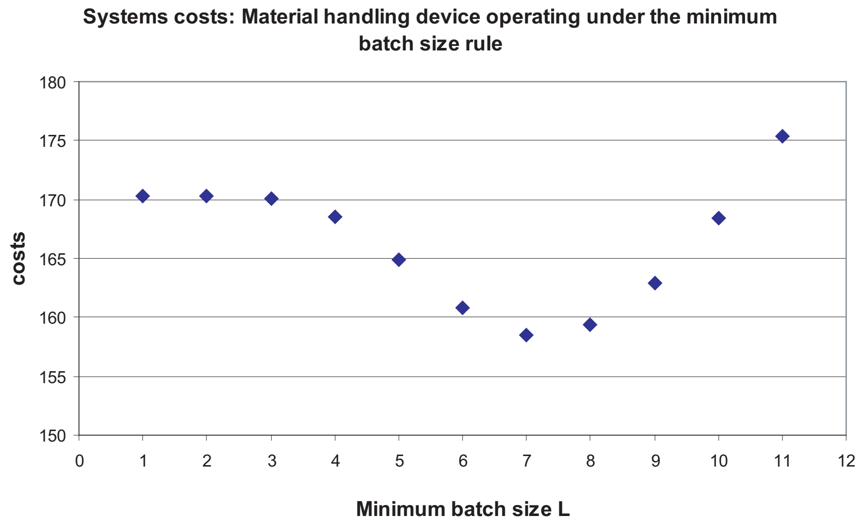


Figure 4.25.: The optimal minimum batch size L regarding the given systems costs

Subsequently we search the optimal minimum batch size $L$ such that the system's costs for the described system are minimal.

| | **arrival process** | | **system figures** | |
|---|---|---|---|---|
| i | $a_i$ | $y_i$ | | |
| 0 | 0.0 | 0.0 | transfer operation | |
| 1 | 0.2 | 0.6 | time ($= t_{out}$) | 10 |
| 2 | 0.4 | 0.3 | $c_{Trans}$ | 10 |
| 3 | 0.2 | 0.1 | $c_{In}$ | 25 |
| 4 | 0.2 | | time period | 100 |
| mean | 2.4 | 1.5 | | |

Table 4.3.: Input values for an optimization problem for a basic transportation system

Two types of costs can be observed: The costs for the transportation process $c_{Trans}$ and the inventory costs for one material unit within the time period $T$, denoted by $c_{In}$. The average number of transportation processes per time period, denoted by $E(N_{Trans})$, can be calculated by

$$E(N_{Trans}) = \frac{T}{E(D)}. \tag{4.98}$$

Furthermore, using Little's Law the average amount of inventory $E(N_{In})$ can be determined. We obtain

$$E(N_{In}) = \frac{E(Y)}{E(A)} \cdot E(W). \tag{4.99}$$

Finally, we get the total system's costs by

$$C_{tot} = c_{Trans} \cdot E(N_{Trans}) + c_{In} \cdot E(N_{In}). \tag{4.100}$$

If $c_{trans}$ and $c_{In}$ are given, we can compute the total costs depending on the minimum batch size $L$ using the analysis presented previously. A numerical example is presented for illustration. The input values for the transportation system are given in Table 4.3 and the system's costs are depicted in Figure 4.25, where the optimal minimum batch size is apparent for a value of $L = 7$.

## 4.6. Chapter Conclusion

In this chapter we presented a detailed discrete time analysis of batch building processes. We studied two basic batch building modes, the capacity and the timeout rule, and additionally a possible modification of these two basic batch building modes, called the minimum batch size rule. We summarize:

- Capacity rule: A given amount of $k$ customers is collected at the service station.
- Timeout rule: The duration for the batch building process is given by a timeout, denoted by $t_{out}$.
- Minimum batch size rule: The collecting process lasts at least $t_{out}$ time units. When $t_{out}$ ends and less than $L$ customers were collected, the batch building process continues until the required $L$ customers are attained.

For the named batch building modes we assumed a batch arrival stream, described by the iid random variables $A$ and $Y$. We presented exact solutions for $w_i$, $d_i$ and $y_{d,i}$. The process behavior of both the capacity and timeout rule is discussed in detail studying the influence of the input parameters on the waiting and departure process. Equivalences between the capacity and the timeout rule could be identified (see Table 4.1). We proved that $w_i$ under the timeout rule is easily given by $1/t_{out}$, $i = 0,1,\dots,t_{out}-1$. Furthermore, it is shown that $E(W)$ under the capacity rule is independent of the variability of the arrival process, however not the quantiles of $w_i$ as the 95% and 99%-quantile. If the capacity rule is applied, it is explained that the influence of $c_Y^2$ on the stability of the departure process is greater than the influence of $c_A^2$. Otherwise, if the timeout rule is applied, this is inverse.

Applying the minimum batch size rule, we detected and explained the paradox of decreasing $E(W)$ with increasing process instability within a certain range of $c_A^2$ and $c_Y^2$. We concluded this chapter with a numerical example discussing an optimization problem.

# 5. Batch Arrivals, Batch Service Queues, and Batch Split

*Millions saw the apple fall, but Newton asked why.*

Bernard Baruch

The focus in the previous chapter was on the batching process itself. In the subsequent chapter we analyze queueing systems in which a service process, described by a random variable, is performed. At first, an analytical approach for the performance analysis of the G/G/1-queue with batch arrivals is presented. Thereafter, we investigate batch server systems, in particular the batch server system running under the minimum batch size rule. In addition to service queues, we devote Section 5.5 to the stochastic split operation required to split a stochastic batch arrival stream.

## 5.1. G/G/1-Queue with Batch Arrivals

In the "basic" queueing models well known from the literature it is assumed that the customers arrive singly at a service facility. However, this assumption does not hold for arrival processes in the real world, which occur often in batches: The raw materials for a production facility are shipped by trucks and the material flow within the production facility is often performed by means of transportation carriers which contain more than one piece. Furthermore, customer orders arrive naturally in batches. Recall that there is an uncountable amount of queueing examples in information systems and in our everyday life, where the customers arrive in batches (see Section 3.1). From this it follows that suitable models for analyzing batch arrivals at a service station are required. This will be done by the introduction of analytical methods in order to compute performance measures for the G/G/1-queue with batch arrivals.

The analysis of the G/G/1-queue with batch arrivals is structured as follows. At first, in Section 5.1.1 we investigate the G/G/1-queue with batch arrivals of constant size and calculate the waiting time distribution. In Section 5.1.2 it is assumed that the batch size is an iid random variable and an approach for computing the waiting time distribution is presented. These methods for the determination of the waiting time distribution have been already introduced by Schleyer and Furmans (2006a). The departure process of the G/G/1-queue with batch arrivals is analyzed in Section 5.1.3. In order to dimension material flow buffers the distribution of the number of customers in the queue at the arrival instant has to be known. Section 5.1.4 is dedicated to this topic.

The G/G/1-queueing system with batch arrivals can be described as follows. The arrival stream is given by the interarrival time, $A$, and the batch size, $Y$. Both are random variables which are iid with $a_i$, $i = 1, \ldots, a_{max}$ and $y_i$, $i = 1, \ldots, y_{max}$. One customer can be served at the G/G/1-queueing station and the service time $B$ is iid with $b_i$, $i = 1, \ldots, b_{max}$. When the server is idle at the arrival instant, one arbitrary customer from the arriving batch can be served immediately. The remaining customers have to wait until the server becomes idle again. Then the next customer is served and so on. When a customers arrives at the same time increment as the service of a customer is finished, it is assumed that the served customer leaves the system before the arriving customer it enters.

### 5.1.1. Waiting Time Distribution: Batch Arrivals of Constant Batch Size

Ackroyd (1980), Grassmann and Jain (1989), and Tran-Gia (1996) present analytical methods for calculating the waiting time distribution of a G/G/1-queueing system, where the arrival stream is a single arrival stream and the arrival of each individual customers marks a renewal process. However, if we consider batch arrivals, we recognize that the arrival in batches influences the waiting time distribution of the "individual" customers the batch is composed of. Considering the individual customers within a batch (see Figure 5.1), it follows that the arrival stream is a correlated stochastic stream. This means that the time interval between the arrivals of two individual customers depends on the preceding time interval. For example, if every batch arrival consists of two customers, then every second time interval is equal to zero time units. Generally, if an individual customer is considered, a batch arrival of size $l$ results in a stochastic arrival stream with $l-1$ succeeding time intervals of length zero and one of length $A = i$ with probability $a_i$.

We present subsequently an approach which takes into account the influence of the batch size on the waiting time. We distinguish between two cases: The batch size is constant and the batch size is distributed stochastically.

At first, let us consider the case that the incoming batches consist always of exactly $Y = l$ customers, that means $y_l = 1$. This is illustrated in Figure 5.1.

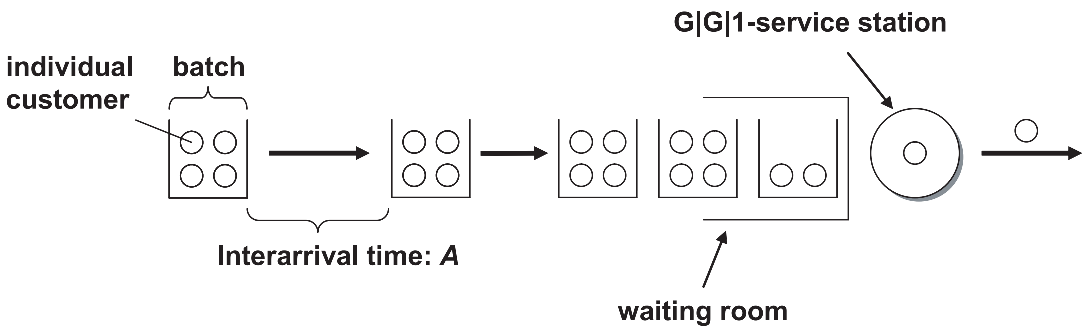


Figure 5.1.: G/G/1-queueing system at which batches of size $l = 4$ arrive

In order to calculate the waiting time distribution, the incoming batch is modeled as one individual customer. The service time of this customer results from $l$ sequential service operations of all customers the batch is composed of. Thus, we assume that all customers of the batch are in service, regardless of the customers still waiting for service. Figure 5.1 shows that exactly two customers of the batch which is in service waiting for their service. The discrete service time distribution for the batch is given by $b_{batch,i}$, $i = 1,\ldots,l \cdot b_{max}$, which results from the $l$-fold convolution of the service time distribution:

$$b_{batch,i} = b_i^{l\otimes}. \tag{5.1}$$

If one batch is modeled as an individual customer, we can determine the difference between the incoming work (time $B_{batch}$ required to serve the batch) and the outgoing work (time $A$ which is provided for service until the arrival of the next batch). Considering the $n$th batch we obtain:

$$c^n_{batch,i} = P(X^n_{batch} = B^n_{batch} - A^n = i \cdot t_{inc}) \quad \forall i = -a_{max} + 1,\ldots,l \cdot b_{max} - 1 \tag{5.2}$$

$X^n_{batch}$ is called working balance of the $n$th batch, which can have both positive and negative values. Since the interarrival time $A^n$ and the batch service time $B^n_{batch}$ are independent of each other, the probability distribution of the working balance is given by

$$c_{batch,i} = \sum_{j=1}^{l \cdot b_{max}} b_{batch,j} \cdot a_{j-i}. \tag{5.3}$$

We can omit the index $n$ in Equation (5.3), since the distributions of $A$ and $B_{batch}$ are independent of $n$. We use *Lindley's* equation in discrete form to determine the waiting time distribution of an incoming batch $w^I_{batch,i}$:

$$w^I_{batch,i} = \begin{cases} \sum_{j=0}^{\infty} w^I_{batch,j} \cdot c_{batch,i-j} & \forall i = 0,1,2,\ldots \\ 0 & \forall i < 0 \end{cases} \tag{5.4}$$

The algorithms of Grassmann and Jain (1989) based on the Wiener-Hopf factorization using ladder height distributions can be used to solve Equation (5.4). Grassmann and Jain present three algorithms and show the convergence of algorithm 1. Algorithm 1, applied to compute the waiting time distribution for an arbitrary arriving batch, $w^I_{batch,i}$, includes the following steps:

1. Initialize $\beta^0_i = 0, i = 1,2,\ldots,a_{max}$ and $\alpha^0_i = 0, i = 1,2,\ldots,l \cdot b_{max} - 1$
2. For $m = 0,1,2,\ldots$

   a)

   $$\beta^{m+1}_i = c_{batch,-i} + \sum_{j=1}^{\infty} \frac{\alpha^m_j \beta^m_{i+j}}{(1 - \beta^m_0)} \quad i = 0,1\ldots - c_{batch,min} \tag{5.5}$$

   b)

   $$\alpha^{m+1}_i = c_{batch,i} + \sum_{j=1}^{\infty} \frac{\alpha^m_{i+j} \beta^m_j}{(1 - \beta^m_0)} \quad i = 1\ldots c_{batch,max} \tag{5.6}$$

3. Iterate until $\max(|\alpha_i^m - \alpha_i^{m+1}|) < \epsilon$

4. It follows:

$$w^I_{batch,0} = 1 - \frac{\sum_{j=1}^{l \cdot b_{max}} \alpha_j}{1 - \beta_0} \tag{5.7}$$

$$w^I_{batch,i} = \frac{\sum_{j=1}^{l \cdot b_{max}} w_{batch,i-j}\alpha_j}{1 - \beta_0} \tag{5.8}$$

5. $\beta_i$ corresponds to the idle time distribution

In addition to the waiting time of the batch, $W^I_{batch}$, distributed by $w^I_{batch,i}$, the waiting time of an individual customer during the service of the batch itself has to be considered. We denote this additional waiting time by $W^{II}$ and its distribution by $w_i^{II}$. The customer first chosen from the batch for service has to wait only for the same amount of time which the whole batch has to wait for. The customer who is taken second has to wait for the service time of the first taken customer additionally. Therefore, the $l$th customer has to wait for the additional duration of $l-1$ service processes. Thus, the waiting time distribution for the $l$th customer is determined by the $(l-1)$-fold convolution of $b_i$.

The probability that an individual customer is placed at the $k$th position within a batch of size $l$, has to be considered. At constant batch sizes, this probability is $1/l$. The additional waiting time of an individual customer of a batch, who is in service, is calculated by

$$\begin{aligned} w_i^{II} &= \frac{1}{l}\sum_{j=1}^{l-1} b_i^{(l-j)\otimes} \quad \forall i = 1{,}2\ldots \\ w_i^{II} &= \frac{1}{l} \quad i = 0. \end{aligned} \tag{5.9}$$

The waiting time of an individual customer, $W$, is the sum of $W^I$ and $W^{II}$ since $W^{II}$ is independent of $W^I$. Thus, we get the waiting time distribution by the convolution of the distributions $w^I_{batch,i}$ and $w_i^{II}$:

$$w_i = P(W = i) = P(W^I_{batch} + W^{II} = i) = w^I_{batch,i} \otimes w_i^{II} \tag{5.10}$$

### 5.1.2. Waiting Time Distribution: Batch Arrivals of Stochastic Batch Size

Next, we develop a method for computing the waiting time distribution, $w_i$, considering batch arrivals of stochastic size. The batch size is a random variable, denoted by $Y$, and its distribution given by $y_i$, $i = 1, \ldots, y_{max}$. Figure 5.2 illustrates a G/G/1-queueing system in which the size of the arriving batch is between one and four.

In the first step, the whole batch is analyzed as an individual customer as we did in the case of batch arrivals of constant size. However, for the derivation of the batch service time distribution, $b_{batch,i}$, we have to take into account the stochastic batch size. For a

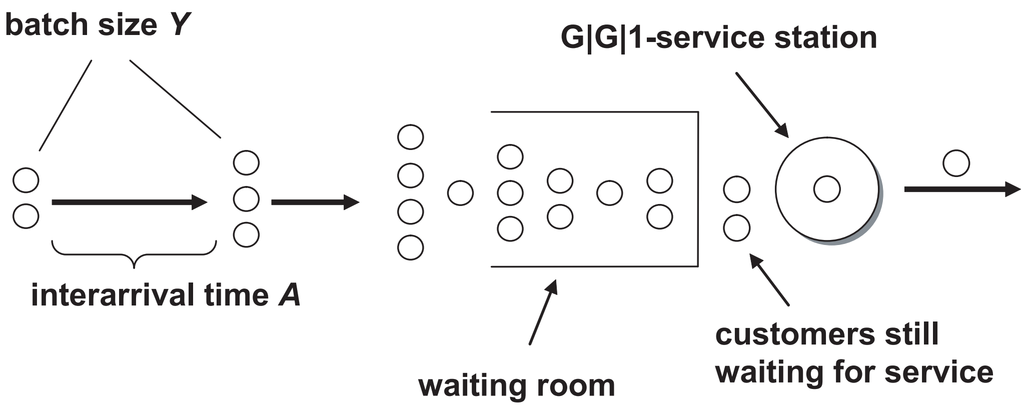


Figure 5.2.: G/G/1-queueing system, where the size of the arriving batch is stochastic

batch size of one, the service time distribution of a batch is the same as that of a single customer that is $b_i$. Therefore, for a batch size of $l$ the service time distribution results from the $l$-fold convolution of $b_i$. To obtain the service time distribution for an arbitrary batch, we have to weight the service time of a batch of size $l$ by the probability that a batch of size $l$ arrives at the service station. Thus, we obtain

$$b_{batch,i} = \sum_{l=1}^{y_{max}} y_l \cdot b_i^{l\otimes}. \tag{5.11}$$

Equation (5.3) yields the distribution of the working balance, $c_{batch,i}$, under the condition that the whole batch is considered as an individual customer. *Lindley*'s equation in discrete form can be set up by Equation (5.4). Next, we compute the waiting time distribution of the whole batch, $w^I_{batch}$, by Equations (5.5), (5.6), (5.7) and (5.8).

As described previously, an individual customer has to wait an additional time during the service of his batch depending on the sequence in which the individual customers are taken for service. In Figure 5.2 two customers are shown having to wait when their batch is already in service. If a customer is taken at the $m$th position of a batch, he has to wait an additional duration of $m-1$ service processes.

Thus, the distribution for this additional waiting time, denoted by $w_i^{II}$, depends on the position of an arbitrary individual customer and that this customer belongs to a batch of size $l$. The probability that an arbitrary individual customer belongs to a batch of size $l$, denoted by $q_l$, is proportional to the probability $y_l$ and to the size $l$ itself. We develop

$$q_l \sim l \cdot y_l \tag{5.12}$$

and this yields

$$q_l = C \cdot l \cdot y_l, \tag{5.13}$$

where $C$ is a constant. As the sum of the probabilities $q_l$ for $l = 1, \ldots, y_{max}$ has to be 1, it follows that

$$\sum_{l=1}^{y_{max}} q_l = 1 = C \sum_{l=1}^{y_{max}} l \cdot y_l = C \cdot E(Y)$$
$$\Rightarrow C = \frac{1}{E(Y)}. \tag{5.14}$$

Thus, we get

$$q_l = \frac{l \cdot y_l}{E(Y)}. \tag{5.15}$$

If an arbitrary customer belongs to a batch of size $l$ with probability $q_l$ and is taken out of the batch for service first, he does not have to wait for an additional time period. This customer is taken first with probability $1/l$. Accordingly, the probability that an arbitrary customer does not have to wait additionally can be specified by

$$P(W^{II} = 0) = w_0^{II} = \sum_{l=1}^{y_{max}} \frac{q_l}{l}. \tag{5.16}$$

For a customer who is taken at the $m$th position with $1 < m \leq l$ we develop Equation (5.17). The probability $q_l$ and the probability that a customer is taken from the $m$th position under the condition that the batch has a size of $l$ is taken into account. We obtain for the probability that the additional waiting time is $i$ time units

$$P(W^{II} = i) = w_i^{II} = \sum_{j=1}^{y_{max}} b_i^{j\otimes} \sum_{l=j+1}^{y_{max}} \frac{q_l}{l}. \tag{5.17}$$

Due to the independence of $W^{II}$ from $W^{I}$ the convolution of the distributions $w^{I}_{batch,i}$ and $w_i^{II}$ yields the waiting time distribution for the G/G/1-queueing system with batch arrivals of stochastic size. Thus, it follows that

$$w_i = w^{I}_{batch,i} \otimes w_i^{II}. \tag{5.18}$$

The sojourn time of a customer in the system is the sum of the waiting and service time. If $w_i$ is known, the sojourn time can be easily determined by

$$v_i = w_i \otimes b_i. \tag{5.19}$$

### 5.1.3. Interdeparture Time Distribution

Once the waiting time distribution is known, the interdeparture time distribution can be determined (see Grassmann and Jain (1988)). We have to distinguish between two cases. First, if customer $n+1$ arrives at the queueing system and has to wait, the interdeparture time between customer $n$ and $n+1$ corresponds to the service time. However, if customer

$n+1$ arrives at an empty system and initiates a busy period, his waiting time is zero and the interdeparture time between customer $n$ and $n+1$ is the sum of the service time and the idle time. Thus, we obtain

$$d_i = w_0 \cdot (\sum_{j=0} \beta_j \cdot b_{i-j}) + (1-w_0) \cdot b_i, \tag{5.20}$$

where $d_i$ is referred to as the interdeparture time distribution. In the following section we study the number of customers at the arrival instant.

### 5.1.4. Distribution of the Number of Customers at the Arrival Instant

In material flow systems, there is a limited space close to a machine or a group of machines for buffering unfinished goods. Therefore, it is of vital importance to analyze the number of customers in the queue at the arrival instant for the dimensioning of material flow buffers. It is crucial that there is enough free buffer capacity to receive arriving customers. If the distribution of the number of customers at the arrival instant is known, the required dimension of the buffer can be determined on the basis of a given confidence level $\sigma$, $0 \le \sigma < 1$.

With $\tau_n$ we denote the arrival instant of the $n$th customer, with $\delta_n$ the departure instant of the $n$th customer, and with $N(t)$ the number of customers at the time instant $t$, $t = 0,1,2,\ldots$. The following analytical approach is based on the idea that the probability to encounter $k$ customers in the system at the arrival instant of the $(n+k+1)$th customer, $\tau_{n+k+1}$, is equal to the probability to encounter $k$ customers at the departure instant of the $n$th customer, $\delta_n$. Therefore, we obtain

$$P\{N(\tau_{n+k+1}) \le k\} = P\{N(\delta_n) \le k\}. \tag{5.21}$$

Furmans and Zillus (1996) analyzes the distribution of the number of customers at the arrival instant in a G/G/1-queue with single arrivals. He explains the equivalence of $N(\tau_{n+k+1})$ and $N(\delta_n)$ as follows. It is assumed that $N(\tau_{n+k+1}) \le k$ and under this condition, the $n$th customer is already served before the $(n+k+1)$th customer arrives. Thus, it follows that there are no more than $k$ customers in the system immediately after the departure of customer $n$. Otherwise, if customer $n$ encounters no more than $k$ customers at the departure instant, $\delta_n$, it is impossible for customer $n+k+1$ to encounter more than $k$ customers at his arrival. Thus, we summarize:

$$\begin{array}{rlcll} \text{if} & N(\tau_{n+k+1}) \le k & \Rightarrow & N(\delta_n) \le k & \text{and} \\ \text{if} & N(\delta_n) \le k & \Rightarrow & N(\tau_{n+k+1}) \le k, & \\ \text{it follows that} & N(\tau_{n+k+1}) \le k & \Leftrightarrow & N(\delta_n) \le k. & \end{array} \tag{5.22}$$

At first, let us assume that customers arrive singly. If the $(n+k+1)$th customer encounters less than or equal to $k$ customers in the queueing system at his arrival, the time interval between the arrival of customer $n$ and $n+k+1$, denoted by $\tau_{n+k+1} - \tau_n$, has to be at least the sojourn time of the $n$th customer. The sojourn time is the sum of the waiting and service time. We conclude that

$$P\{N(\tau_{n+k+1}) \leq k\} = P\{N(\delta_n) \leq k\} = P\{\tau_{n+k+1} - \tau_n \geq W_n + B_n\}. \tag{5.23}$$

However, if batch arrivals are considered, the time interval $\tau_{n+k+1} - \tau_n$ and the waiting time of customer $n$ are dependent on the position $l$ of customer $n$ within his batch. For example, if the $n$th customer is in the last position within his batch, his waiting time and the time interval until the arrival of customer $n + k + 1$ is differently distributed as compared to the case that the $n$th customer is placed at the first position within his batch. Therefore, we have to consider the dependency of $P\{N(\delta_n) \leq k\}$ on the position $l$ of customer $n$ within his batch. Equation (5.23) has to be extended as follows:

$$P\{N(\delta_n(l)) \leq k\} = P\{[\tau_{n+k+1} - \tau_n](l) \geq W_n(l) + B_n\}. \tag{5.24}$$

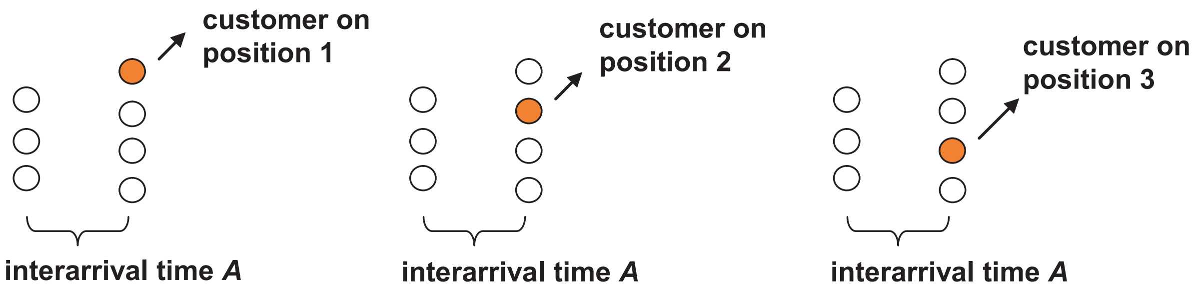


Figure 5.3.: Batch arrivals in a G/G/1-queue: Mode of numbering the individual customers of a batch

Let us number the individual customers of a batch using the following procedure. The customer who is chosen last to be served is assigned the number one. The customer who is chosen before is assigned the number two and so on. Finally, the customer who is chosen first is assigned the number of the batch size. See Figure 5.3 which clarifies the modus of numbering. We choose this mode, since this simplifies the derivation of the distribution of the interval length $[\tau_{n+k+1} - \tau_n](l)$. First of all, we compute the waiting time distribution depending on the position $l$, denoted by $w_i(l)$, $i = 0,1,\ldots$.

The probability that an arbitrary chosen customer is located at position $l$ is computed by

$$o_l = \sum_{m=l}^{y_{max}} \frac{q_m}{m} \qquad \forall l = 1,\ldots,y_{max}, \tag{5.25}$$

where $q_m$ denotes the probability that an arbitrary customer is an element of a batch of size $m$ and is calculated by Equation (5.15).[1]

If a customer is on the last position (that means $l = 1$), his waiting time depends on the size of his batch. Therefore, we calculate the probability that an arbitrary customer is

[1] It has to be noted that the probability to be on the last position of a batch is the same as to be on the first position. Furthermore, the probability to be on the second last position is the same as to be on the second position and so on. Thus, it is only dependent on the viewing angle.

element of a batch of size $m$ under the condition that this customer is at position $l$. The described probability is denoted by $P(Q = m|O = l)$ and is given by

$$P(Q = m|O = l) = \frac{P(O = l|Q = m) \cdot q_m}{o_l} \quad m = 1, \ldots, y_{max} \;\; l \leq m. \tag{5.26}$$

If a batch of size $m$ is given, an arbitrary customer is at position $l$, $1 \leq l \leq m$ with probability $1/m$. Thus we obtain

$$P(O = l|Q = m) = \frac{1}{m}. \tag{5.27}$$

Using the Equations (5.15) and (5.27) we simplify Equation (5.26) as

$$P(Q = m|O = l) = \frac{m \cdot y_m}{E(Y) \cdot m \cdot o_l} = \frac{y_m}{E(Y) \cdot o_l}. \tag{5.28}$$

The waiting time $W(l)$ is computed by the sum of the waiting time of the whole batch, $W^I_{batch}$, and of the additional waiting time of an individual customer who is placed at position $l$, $W^{II}(l)$. $W^I_{batch}$ is independent of $l$ and is obtained using the Equations from (5.4) to (5.8). The distribution of $W_{II}(l)$ is solved by

$$w_i^{II}(l) = P(Q = l|O = l) \quad i = 0 \tag{5.29}$$

$$w_i^{II}(l) = \sum_{m=1}^{y_{max}-l} P(Q = l + m|O = l) \cdot b_i^{m\otimes} \quad i = 1,2,\ldots. \tag{5.30}$$

The waiting time of a customer who is at position $l$ is zero if he is an element of a batch of size $l$. This is given by Equation (5.29). Equation (5.30) describes the case that a customer who is at position $l$ has to wait for $m > 0$ service processes since the service start of his batch. Analogous to Equation (5.18) we get

$$w_i(l) = w^I_{batch,i} \otimes w_i^{II}(l). \tag{5.31}$$

The sum of $W(l)$ and $B$ leads to the sojourn time of a customer depending on $l$, which is denoted by $V(l)$. The distribution of $V(l)$ is given by

$$v_i(l) = P(W(l) + B = i) = w_i(l) \otimes b_i \;\; i = 1,2,\ldots. \tag{5.32}$$

The length of the time interval $[\tau_{n+k+1} - \tau_n](l)$ is dependent on the position $l$. The distribution $\alpha_i^n(k,l)$ denotes the probability that the time interval $[\tau_{n+k+1} - \tau_n](l)$ is $i$ time units.

Figure 5.4 shows the time interval $[\tau_{n+k+1} - \tau_n](l)$ in which customer $n$ is located at position $l = 1$. The time interval $[\tau_{n+k+1} - \tau_n](l)$ is composed of multiple interarrival time intervals, each distributed by $a_i$. In Figure 5.4 $[\tau_{n+k+1} - \tau_n](l)$ is composed of five interarrival time intervals. The number of interarrival time intervals depends on the sizes of the arriving batches. For example, if batches with large sizes arrive within $[\tau_{n+k+1} - \tau_n](l)$, we encounter less arrivals in the considered time interval as if the batch sizes are small.

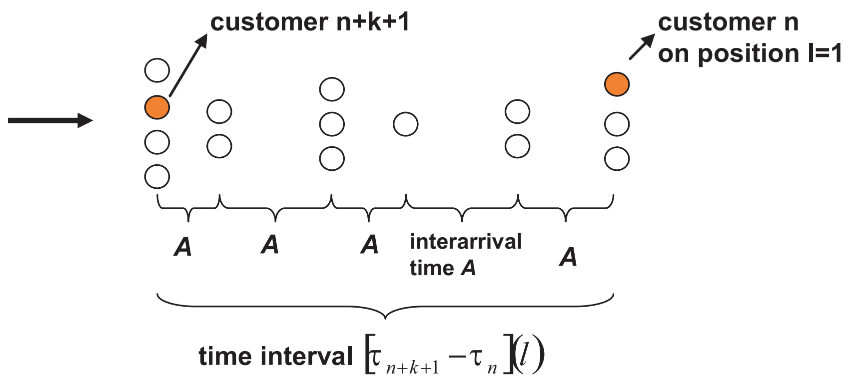


Figure 5.4.: Time interval between the arrival instant of customer $n$ and customer $n+k+1$ depending on the position $l$ of customer $n$

The distribution that the time interval $[\tau_{n+k+1} - \tau_n](l)$ is composed of $m$ interarrival time intervals is denoted by $\Phi^n_{k,m}(l)$. Since $\Phi^n_{k,m}(l)$ depends only on $l$ and $k$, we omit the index $n$ and write subsequently $\Phi_{k,m}(l)$. We determine the number of interarrival time intervals for $k = 0,1,\ldots,k_{max}$ and $l = 1,\ldots,y_{max}$, where $k_{max}$ is the next greater integer of the fraction of the maximum waiting time and minimum service time. We get

$$k_{max} = \left\lceil \frac{w^{max}}{b^{min}} \right\rceil . \tag{5.33}$$

If $l > k+1$, it means that customers $n$ and $n+k+1$ are located in the same batch and therefore the number of interarrival time intervals between the arrival of customer $n$ and $n+k+1$ is zero. The described situation is illustrated in Figure 5.5. It yields

$$\Phi_{k,m}(l) = \begin{cases} 1 & m = 0 \text{ and } l > k+1 \\ 0 & m > 0 \text{ and } l > k+1. \end{cases} \tag{5.34}$$

If $l = k+1$, customer $n$ is at position $k+1$ within his batch and customer $n+k+1$ is located on the first position of the succeeding batch. See Figure 5.6 to visualize this situation in which the time interval between customer $n$ and $n+k+1$ is exactly one interarrival time interval. We obtain

$$\Phi_{k,m}(l) = \begin{cases} 1 & m = 1 & \text{and } l = k+1 \\ 0 & \text{else} & \text{and } l = k+1. \end{cases} \tag{5.35}$$

If $l < k+1$, there is at least one interarrival time interval between the arrival of customer $n$ and $n+k+1$. If the succeeding batch has a size of at least $k-l+2$, there is exactly one interarrival time interval (see Figure 5.7). If $l < k+1$ and the size of the succeeding batch is less than $k-l+2$, the interval between the arrival of customer $n$ and $n+k+1$ consists of more than one interarrival time interval.

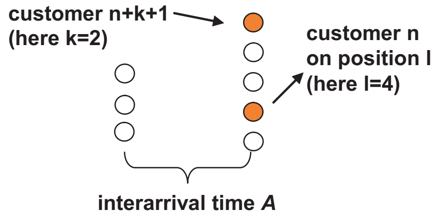


Figure 5.5.: Time interval between the arrival of customer $n$ and $n{+}k{+}1$; $l > k{+}1$

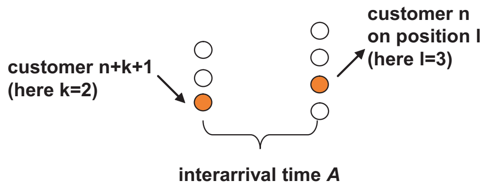


Figure 5.6.: Time interval between the arrival of customer $n$ and $n{+}k{+}1$; $l = k{+}1$

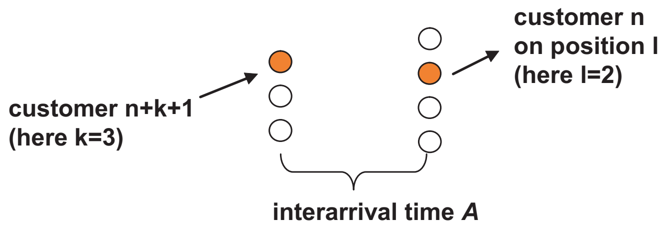


Figure 5.7.: Time interval between the arrival of customer $n$ and $n + k + 1$; $l < k + 1$

Given $l$ and $k$ with $l < k + 1$, it yields

$$\begin{aligned}
\Phi_{k,m}(l) &= 0 && m = 0 \text{ and } l < k+1 \\
\Phi_{k,m}(l) &= \sum_{j=k+1}^{y_{max}} y_{j-(l-1)} && m = 1 \text{ and } l < k+1 \\
\Phi_{k,m}(l) &= \sum_{i=1}^{y_{max}} y_{k+1-i-(l-1)}^{(m-1)\otimes} \sum_{j=i}^{y_{max}} y_j && m > 1 \text{ and } l < k+1.
\end{aligned} \tag{5.36}$$

Based on the results of Equations (5.34), (5.35) and (5.36), $\alpha_{k,i}^n(l)$ can be determined. Since $\alpha_{k,i}^n(l)$ can be denoted independently of $n$, we write subsequently $\alpha_{k,i}(l)$. The distribution of the interval length $[\tau_{n+k+1} - \tau_n](l)$ can be calculated by

$$\alpha_{k,i}(l) = \begin{cases} \Phi_{k,0}(l) & i = 0 \\ \sum_{m=1}^{\infty} \Phi_{k,m}(l) \cdot a_i^{m\otimes} & i > 0. \end{cases} \tag{5.37}$$

Hence, we can solve Equation (5.24) by

$$\begin{aligned}
P\{N(\delta_n(l)) \leq k\} &= P\{[\tau_{n+k+1} - \tau_n](l) \geq W^n(l) + B^n\} \\
&= P\{[\tau_{n+k+1} - \tau_n](l) \geq V^n(l)\} \\
&= \sum_{i=0}^{\infty} v_i^n(l) \left[ \sum_{j=i}^{\infty} \alpha_{k,j}(l) \right].
\end{aligned} \tag{5.38}$$

The probability that an arriving customer sees exactly $k$ customers in the queue depending on $l$ is determined by

$$P\{N(\delta_n(l)) = k\} = P\{N(\delta_n(l)) \leq k\} - P\{N(\delta_n(l)) < k\}. \tag{5.39}$$

The law of total probability leads to

$$P\{N(\delta_n) = k\} = \sum_{l=1}^{y_{max}} P\{N(\delta_n(l)) =)\} \cdot o_l. \tag{5.40}$$

Since we consider the system in the steady state, we can omit the index $n$ in Equation (5.38). This yields

$$P\{N(\delta(l)) \le k\} = \sum_{i=0}^{\infty} v_i(l) \left[ \sum_{j=i}^{\infty} \alpha_{k,j}(l) \right], \tag{5.41}$$

and finally leads to the distribution of the number of customers at the arrival instant of the G/G/1-queue with batch arrivals. The distribution of the number of customers at the arrival instant is denoted by $\eta_i$, $i = 0,1,\ldots$

In order to complete this section about the G/G/1-queue with batch arrivals, we present subsequently numerical results of the experiments we conducted.

### 5.1.5. Analysis

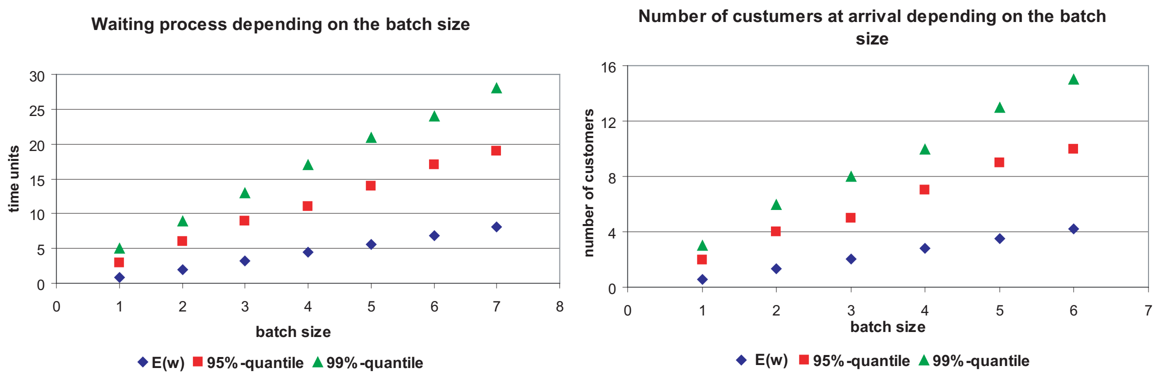


Figure 5.8.: Experiment 1: The influence of the batch size on performance measures of the G/G/1-queue; left: Waiting process; right: Number of customers at the arrival instant

Since the calculation of $w_i$ is exact within an $\epsilon$-environment (see Grassmann and Jain (1989)), the presented analytical approach is exact as well. In addition, the computing times are very short.[2] Therefore, it is possible to perform a large set of computations within a short time.

Subsequently, we present some numerical examples in which we analyze the influence of the input parameters on the output. At first, we investigate the influence of the batch size on the waiting process of a G/G/1-queue. We start with a batch size of one, the interarrival time distribution $a_i$ is (0.3,0.4,0.3), $i = 1,2,3$, and the service time distribution is (0.5,0.4,0.1), $i = 1,2,3$. This results in an utilization of $\rho = 0.8$. We

[2] A few milliseconds on an Intel Centrino processor

increase the batch size by one in each of the following experiments. The service time distribution is kept constant and the interarrival time is changed in such a way that only the batch size itself influences the waiting process. In the case that the batch size is two, we have to multiply the indices of $a_i$ by two and obtain (0,0.3,0,0.4,0,0.3), $i = 1, \ldots, 6$ as the new interarrival time distribution. If the batch size is three, we take an interarrival time distribution of (0,0,0.3,0,0,0.4,0,0,0.3),$i = 1, \ldots, 9$. As you can see in the left graph of Figure 5.8, the mean waiting time increases linearly with an increasing batch size. Since the presented approach enables the calculation of the waiting time distribution, its quantiles can be determined as well. In the left graph of Figure 5.8, $\sigma_{w,0.95}$ and $\sigma_{w,0.99}$ as quantiles of the waiting time distribution are shown as a function of the batch size. It is observable that $\sigma_{w,95}$ and $\sigma_{w,99}$ are more affected by an increasing batch size than $E(W)$.

Next, we study the influence of the batch size on the distribution of number of customers at the arrival instant, $\eta_i$. We perform the same experiment as described above and compute the mean number of customers at the arrival instant, $E(N(\tau))$, and the 95% and 99%-quantiles of $\eta_i$. The results are depicted in the right graph of Figure 5.8. In order to choose the suitable buffer size in front of the G/G/1-server with a given probability, the quantiles of $\eta_i$ are required.

In experiment 2, the influence of an increasing utilization on $E(w)$ and $\sigma_{v,0.95}$ is analyzed. We choose the quantiles of the sojourn time distribution, since this delivers the probability of the on-time order fulfillment of the G/G/1-queueing system with batch arrivals. For example, the 95%-quantile of $v_i$, $\sigma_{v,0.95}$, means that 95% of the orders can be fulfilled in less than or equal to $\sigma_{v,0.95}$ time units. The graphs of experiment 2 in Figure 5.9 illustrates that $E(w)$ and $\sigma_{v,0.95}$ increase disproportionately with an increasing utilization. The greater the scv of the batch size distribution, the greater $E(W)$ and $\sigma_{v,0.95}$.

In experiment 3, we investigate the effect of increasing utilization on the number of customers. The results are depicted in Figure 5.10 and are analogous to the results of experiment 2. In addition, we analyzed the 99%-quantile of $\eta_i$ depending on the utilization (see the left graph of Figure 5.11) and thereafter we calculated the probability that the queue is empty at arrival (see the right graph of Figure 5.11). The higher the utilization, the lower is this probability. The higher the scv of the batch size distribution, the lower the probability that the queue is empty at arrival.

Finally, in the experiments from 5 to 7 we study the dependence of $E(W)$, $\sigma_{v,0.95}$, $E(N(\tau))$ and $\sigma_{\eta,0.95}$ on the process stability, described by the scv of the batch size or interarrival time distribution. Thus, the Figures from 5.12 to 5.14 show that $E(W)$, $\sigma_{v,0.95}$, $E(N(\tau))$ and $\sigma_{\eta,0.95}$ increase with an increasing unstable behavior of arrival processes. Thereby, the 95%-quantiles are influenced greater than the mean values.

Therefore, **focusing on mean values, the consequences of varying scv's on the on-time order fulfillment will be underestimated**. This is a further argument using discrete time queueing models for the analysis of material flow problems.

Furthermore, stable arrival and service processes characterized by low values of the scv lead to a significant reduction in $E(W)$, $\sigma_{v,0.95}$, $E(N(\tau))$ and $\sigma_{\eta,0.95}$. Thus, it would be worthwhile for the management of material flow systems to make efforts to create stable

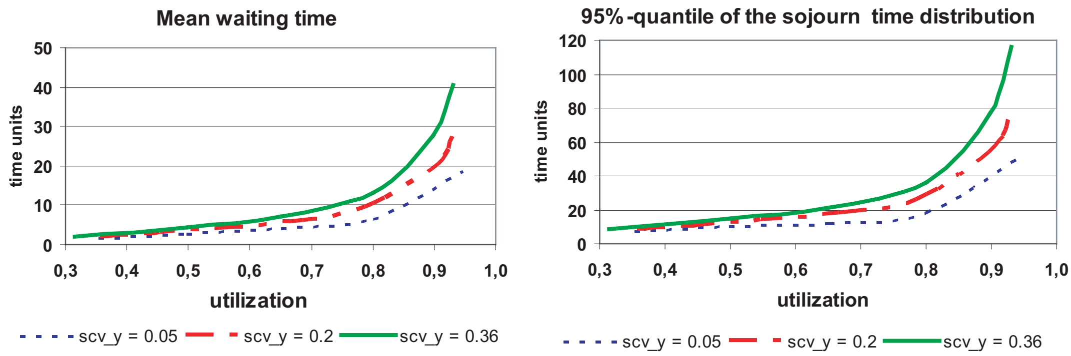


Figure 5.9.: Experiment 2: left: Mean waiting time depending on the utilization; right: 95%-quantile of the sojourn time distribution depending on the utilization

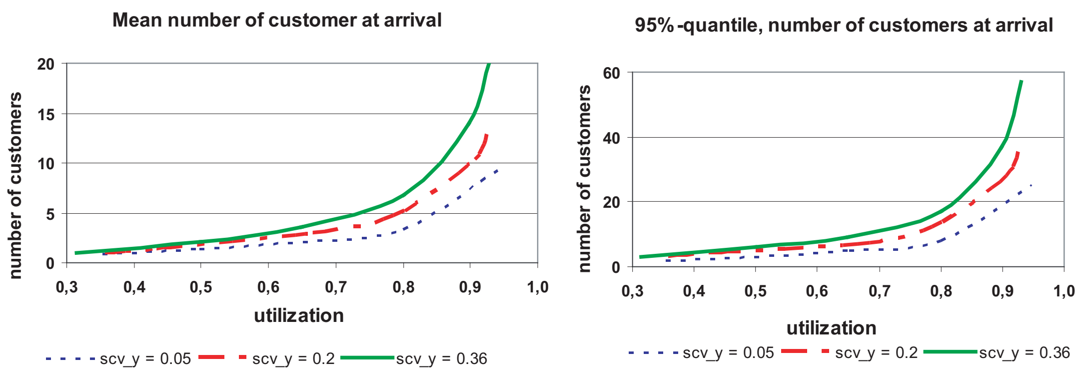


Figure 5.10.: Experiment 3: left: Mean number of customers in the queue at arrival depending on the utilization; right: 95%-quantile of $\eta_i$ depending on the utilization

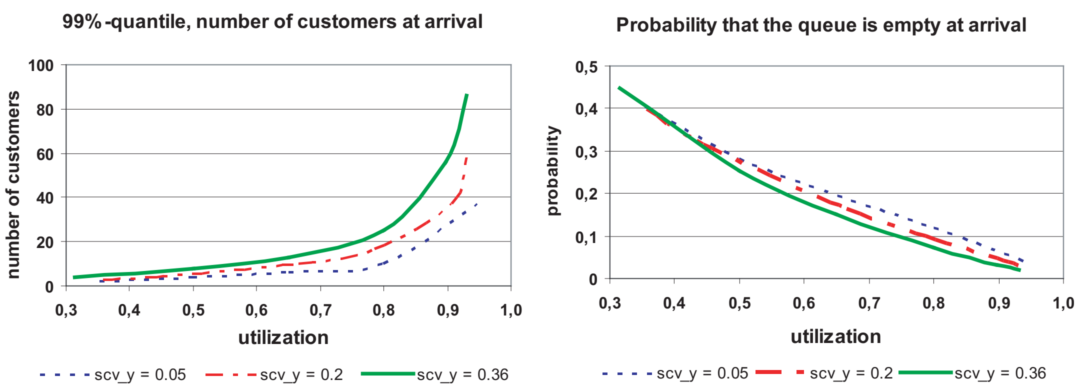


Figure 5.11.: Experiment 4: left: 99%-quantile of $\eta_i$ depending on the utilization; right: The probability that the queue is empty at arrival

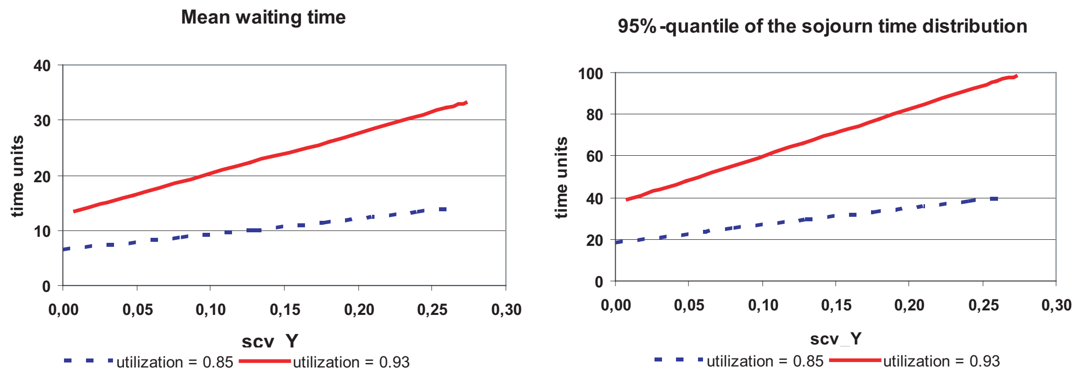


Figure 5.12.: Experiment 5: left: Mean waiting time depending on $c_Y^2$; right: 95%-quantile of the sojourn time distribution depending on $c_Y^2$

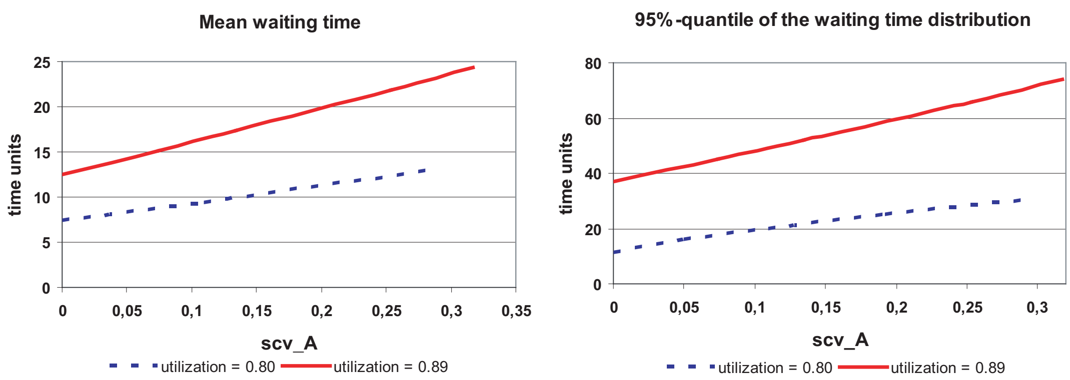


Figure 5.13.: Experiment 6: left: Mean waiting time depending on $c_A^2$; right: 95%-quantile of the sojourn time distribution depending on $c_A^2$

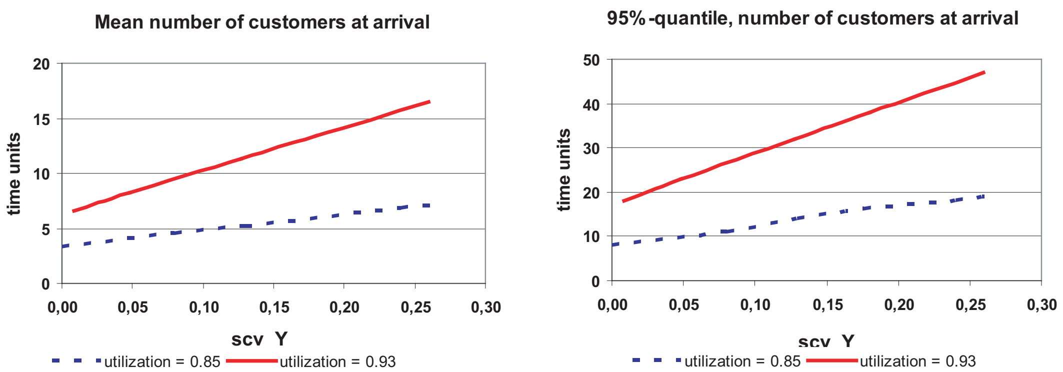


Figure 5.14.: Experiment 7: left: Mean number of customers in the queue at arrival depending on $c_Y^2$; right: 95%-quantile of $\eta_i$ depending on $c_Y^2$

processes. Examples of measures to create stable processes: Reduction of failure rates, reduction of scrap, introduction of equal batch sizes and customer collaboration with the objective to reduce demand fluctuations.

## 5.2. Batch Service Queues

A batch server is able to serve a specific number of customers simultaneously. The number of customers who can be served simultaneously is bounded by the server's capacity $K$. Batch service queues arise in a variety of settings in material flow systems due to the need of demand consolidations. Some examples from the manufacturing environment are furnace treatment in chemical washings, oven treatment in the semiconductor manufacturing, metallization steps, and painting operations. On the other hand, many transportation processes involving trucks, buses, ships, trains, and airplanes occur in batches. Conveying systems such as material handling devices, shuttles, automated guided vehicles, elevators, and gondulas also transfer batches. Furthermore, we can model a traffic light which controls the traffic flow at a crossing by a batch server queue. In this case, the batch service is rendered in terms of the ability of a group of cars to pass the phase of green light. Regardless of the various applications of batch server systems, the purpose is always to make an effectively use of the available capacity.

There are different variations of batch service systems. We discuss the *full batch policy* in Section 5.3, where always a batch of the maximal capacity $K$ is collected and served. Section 5.4 introduces the *minimum batch size policy* which is a server control strategy, where a service process is initiated only if at least $L$ customers are waiting in the queue.

## 5.3. Full Batch Policy

Firstly, we discuss the model for the full batch policy which can be decomposed into known subsystems.

The Kendall notation for the full batch policy is G/G$^{[K,K]}$/1. With this policy, a service operation is only initiated when a batch of size $K$ corresponding to the server's capacity is present in the queue. The G/G$^{[K,K]}$/1-queueing system consists of two subsystems connected in series. The first subsystem is a station collecting the arriving customers up to a number $K$. The second subsystem is the batch server itself, where the batch departing from the collecting station is processed. A G/G$^{[K,K]}$/1-system consisting of its two subsystems is illustrated in Figure 5.15.

We assume an arrival stream with interarrival time $A$, distributed by $a_i$, $i = 1{,}2, \ldots$ with either single arrival or batch arrival with batch size, $Y$ distributed by $y_i$, $i = 1{,}2, \ldots$. The service time $B$ is distributed by $b_i$, $i = 1{,}2, \ldots$. Each distribution, $a_i$, $y_i$ and $b_i$, is iid. The capacity of the batch server and subsequently the collecting size is $K$.

The G/G$^{[K,K]}$/1-system can be investigated by the analysis of the two named subsystems. The idea of separately analyzing these subsystems is known from the literature and can be found by Bitran and Tirupati (1989), and Fowler et al. (2002). In their work, they investigate batch processing in multi-product queues in the continuous time domain.

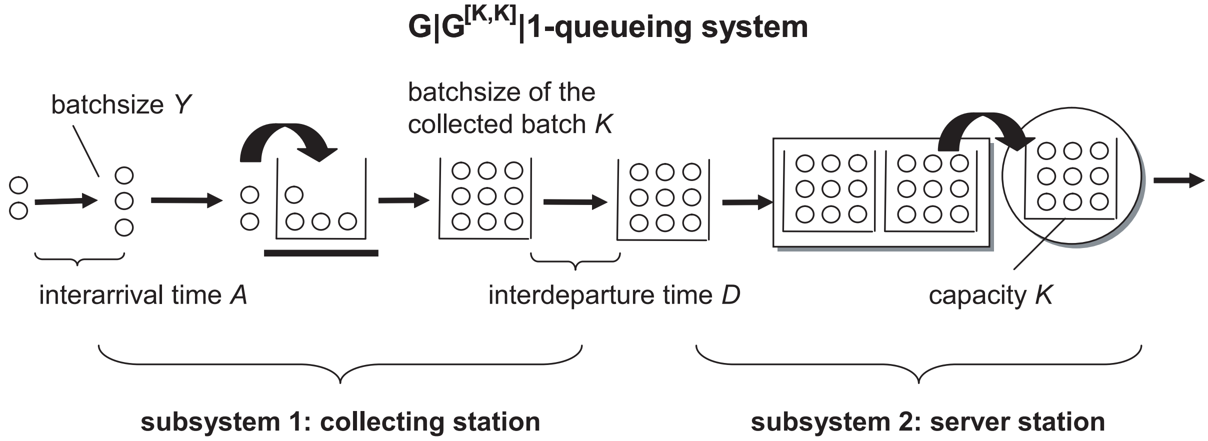


Figure 5.15.: $G/G^{[K,K]}/1$-queueing system which consists of two subsystems: The collecting station, where the batch is built and the batch server station, where the batch is processed

However, they investigate only single arrivals at the collecting station and perform a mean value analysis.

The discrete time analysis of the collecting process can be done by the presented batch building model using the capacity rule (see Section 4.2). This approach delivers the interdeparture time $D$, distributed by $d_i$, which corresponds to the interarrival time for the service station. In addition, we can calculate the waiting time distribution of an arbitrary customer at the collecting station. The batch server station can be modeled as an ordinary $G/G/1$-queue in which a batch corresponds to an individual customer. We suggest the approaches of Grassmann and Jain (1988), and Grassmann and Jain (1989) for the determination of the waiting and interdeparture time distribution. Finally, the waiting time $W$ of an arbitrary customer in a $G/G^{[K,K]}/1$-system can be computed by the sum of the waiting time of an arbitrary customer at the collecting station and the waiting time of the collected batch at the batch service station. The distribution of $W$ is determined by the convolution of both waiting time distributions.

The next section introduces a batch server system operating under the minimum batch size rule.

## 5.4. Batch Service Control Strategy - Minimum Batch Size Policy

A batch server operating under the minimum batch size rule works as follows. The batch server has a maximum capacity of $K$ customers. When the batch service ends and there are less than $L$ customers waiting, the server remains idle until $L$ customers are accumulated in the queue. If there are $\geq L$ and $\leq K$ customers waiting, the entire queue is served and if there are $> K$ customers accumulated, the queue length is reduced by $K$. The minimum batch size $L$ is an arbitrary constant between 1 and $K$. If $L = 1$, at least one single customer is able to trigger a service operation. This special case is called the "greedy" policy (see Bolch et al. (1998)). The Kendall notation of the described system is G/G$^{[L,K]}$/1. In contrast to the G/G$^{[K,K]}$/1-queue, we allow the arrival of single units. The analysis of the G/G$^{[L,K]}$/1-queue with batch arrivals is for future research. Note that the G/G$^{[K,K]}$/1-queue is a special case of the G/G$^{[L,K]}$/1-queue if $L = K$.

In particular, the G/G$^{[L,K]}$/1-queueing model can be applied for all the various transportation of manufacturing processes named above (see Section 5.2). Generally, for all the named batch service processes, it is not obligatory to operate only batches of a maximum size $K$. In contrast, a minimum batch size can be chosen in a way such that the optimal system configuration can be determined considering inventory and operation costs under the constraint of an on-time order fulfillment.

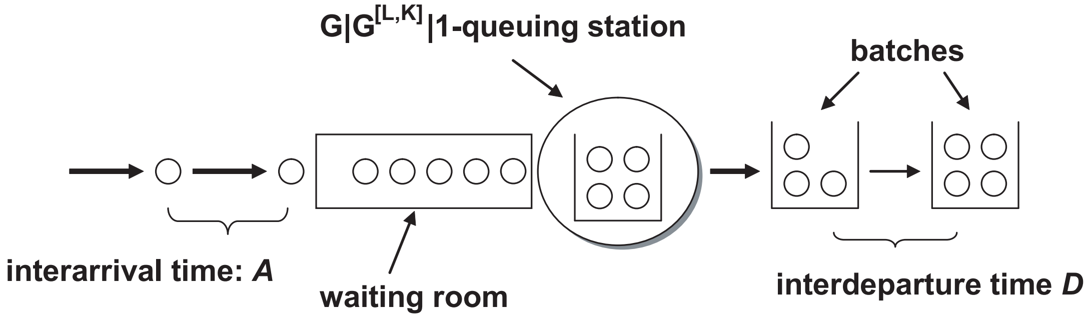


Figure 5.16.: The G/G$^{[L,K]}$/1-batch server system

Dümmler (1998) presents a model for the departure process of a discrete time M/G$^{[L,K]}$/1-system in which he assumes that the arrival process is a Markov process. In our work we drop the Markovian property of the arrival process and take Dümmler's approach as foundation for deriving an approximate method for the G/G$^{[L,K]}$/1-queue. In addition to Dümmler, we calculate the idle and waiting time distribution.

As in the previous models, the interarrival time is denoted by $A$ and the service time by $B$. Both $A$ and $B$ are iid and described by the distributions $a_i$ and $b_i$, $i = 1,2,\ldots$.

In Section 5.4.1 we analyze the number of customers at the departure instant which leads to the batch size and interdeparture time distribution presented in Section 5.4.2 and 5.4.3. Thereafter, Section 5.4.4 introduces an approach to compute the waiting time distribution.

### 5.4.1. Distribution of the Number of Customers at the Departure Instant

$N(\delta_{n-1})$ denotes the number of customers in the queue immediately after the departure of the $(n-1)$th batch. The distribution of $N(\delta_{n-1})$ is written as $\nu_x^{n-1}$, $x = 1, \ldots$. If $N(\delta_{n-1}) \leq K$, the queue will be empty when the server starts operating again. If there are more than $K$ customers waiting at the departure instant, the customers waiting in the queue are immediately reduced by $K$. Thus, the number of customers immediately after the $n$th service start is described by

$$H^n = \begin{cases} N(\delta_{n-1}) - K, & \text{if } \ N(\delta_{n-1}) > K \\ 0, & \text{if } \ N(\delta_{n-1}) \leq K. \end{cases} \tag{5.42}$$

We introduce the operator $\pi_m$ (see Kleinrock (1975) and Dümmler (1998)) which sums up all values of the distribution $f_x$ with $x < m$ and adds their sum to $\pi_m$. Thus, the operator is defined as

$$\pi_m(f_x) = \begin{cases} 0, & \text{if } \ x < m \\ \sum_{j=-\infty}^{m} f_j, & \text{if } \ x = m \\ f_x, & \text{if } \ x > m. \end{cases} \tag{5.43}$$

With $\Delta_m(f_x)$ we use an additional operator defined as

$$\Delta_m(f_x) = f_{m+x} \tag{5.44}$$

and is used to shift the elements of a distribution down by $m$ units. Thus, the distribution of $H^n$ can be written as

$$\varpi_x^n = \pi_0(\Delta_K(\nu_x^{n-1})). \tag{5.45}$$

Now, the time instant immediately after the start of the $n$th batch service, when $H^n$ customers are waiting in the queue, is considered. We denote with $Z^n$, the number of customers who arrive during the $n$th service period. If $N(\delta_{n-1}) < L$, it follows that the batch server was idle and waited until $L$ customers had arrived. Since the last customer arrived at the same time instant, when the $n$th service process started, the next customer will arrive after $A$ time units. Within the batch service time of $m$ time units exactly $x$ customers arrive if the sum of $x$ interarrival time intervals is less than or equal to $m$ time units and the $(x+1)$th customer arrives after at least $m+1$ time units elapsed since the service start. We derive with

$$P(Z^n = x | (B^n = m \wedge N(\delta_{n-1}) < L)) = \sum_{i=0}^{m-1} a_{m-i}^{x\otimes} \sum_{j=i+1}^{a_{max}} a_j \tag{5.46}$$

the probability that $x$ customers arrive during the $n$th batch service process under the condition that the service time lasts $m$ time units and that $N(\delta_{n-1}) < L$.

If $N(\delta_{n-1}) \geq L$, it follows that the server has not been idle before starting the $n$th service. Thus, after the start of the $n$th service the next customer arrives in an interval of shorter or equivalent length of an interarrival time interval. We approximate this residual interarrival time interval by the residual lifetime of a renewal process. A customer who arrives at the finish of the $n$th service process is included in $N(\delta_{n-1})$. Therefore, we consider arrival events immediately after discrete time instants for the calculation of the residual lifetime distribution, denoted by $r_{a,i}$. For deriving $r_{a,i}$ see Section 2.1.2. Under the condition that the batch service time is $m$ time units and $N(\delta_{n-1}) \geq L$, we obtain the probability of $x$ arriving customers during the $n$th service period by

$$P(Z^n = x | (B^n = m \wedge N(\delta_{n-1}) \geq L)) = \sum_{s=1}^{a_{max}} \sum_{i=0}^{m-s-1} r_{a,s} \cdot a_{m-s-i}^{(x-1)\otimes} \sum_{j=i+1}^{a_{max}} a_j. \tag{5.47}$$

Due to the independence of $B$ and $A$, we obtain

$$\begin{aligned} P(Z^n = x|N(\delta_{n-1}) < L) &= \sum_{m=x}^{b_{max}} P(Z^n = x|(B^n = m \wedge N(\delta_{n-1}) < L)) \cdot b_m \\ P(Z^n = x|N(\delta_{n-1}) \geq L) &= \sum_{m=x}^{b_{max}} P(Z^n = x|(B^n = m \wedge N(\delta_{n-1}) \geq L)) \cdot b_m, \end{aligned} \tag{5.48}$$

where we use Equations (5.46) and (5.47). Note that $b_m$ is independent of $n$. Considering that the server becomes idle with the probability $u^{n-1} = P(N(\delta_{n-1}) < L)$, we get the probability that $x$ customers arrive during the $n$th service period by

$$\begin{aligned} z_x^n = P(Z^n = x) = P(Z^n = x|N(\delta_{n-1}) < L) \cdot u^{n-1} \\ + P(Z^n = x|N(\delta_{n-1}) \geq L) \cdot (1 - u^{n-1}). \end{aligned} \tag{5.49}$$

From Equation (5.49) we can derive $N(\delta_n)$ which follows from the sum of $H^n$ and $Z^n$. Therefore, we write

$$\nu_x^n = \varpi_x^n \otimes z_x^n. \tag{5.50}$$

For the calculation of $\varpi_x^n$ we use Equation (5.45). It yields

$$\nu_x^n = \pi_0(\Delta_K(\nu_x^{n-1})) \otimes z_x^n. \tag{5.51}$$

Thus, we get $\nu_x^n$ in dependence on $\nu_x^{n-1}$. Assuming that there are zero customers in the queue at the analysis start, we can iterate starting with

$$\nu_x^0 = \begin{cases} 1, \text{ if } & x = 0 \\ 0, \text{ if } & x \neq 0. \end{cases} \tag{5.52}$$

From Equation (5.43) it follows that

$$\varpi_x^1 = \begin{cases} 1, \text{ if } & x = 0 \\ 0, \text{ if } & x \neq 0. \end{cases} \tag{5.53}$$

Using Equation (5.49) we compute $z_x^1$ and subsequently we get $\nu_x^1$ by Equation (5.51). We iterate over $n$ until we obtain the distribution of the number of customers in the queue at the departure instant:

$$\nu_x = \lim_{n \to \infty} \nu_x^n. \tag{5.54}$$

### 5.4.2. Batch Size Distribution of the Departing Batch

The distribution of the number of customers in a batch, denoted by $y_{d,x}$, $x = L, \dots, K$, can be derived using $\nu_x$, (see Dümmler (1998)). If there are less than $L$ customers in the queue at the end of a service process, the server will wait until $L$ customers are accumulated in the queue, resulting in a batch of size $L$. Whereas, a batch size of

$L < x < K$ is obtained if more than $L$ and less than $K$ customers are observed in the queue. Finally, at least $K$ customers in the queue at the end of a service process leads to a batch size of $K$. We obtain

$$y_{d,x} = \begin{cases} \sum_{j=0}^{L} \nu_j, & \text{if } x = L \\ \nu_x, & \text{if } L < x < K \\ \sum_{j=K}^{\infty} \nu_j, & \text{if } x = K. \end{cases} \tag{5.55}$$

### 5.4.3. Interdeparture Time Distribution

The interdeparture time distribution is given by $d_i$, which describes the time between two consecutive batch departures. If there are at least $L$ customers waiting in the queue at the end of a service process, the time to the next departure instant will be a service time interval. However, if there are less than $L$ customers in the queue, the time length to the next departure instant is the sum of the idle time period and the service time. Therefore, we derive firstly the distribution of the idle time, denoted by $\beta_i$, $i = 1, \ldots, \beta_{max}$. It is given by

$$\beta_i = \frac{1}{\sum_{x=0}^{L-1} \nu_x} \cdot \left[ r_{a,i} \cdot \nu_{L-1} + \sum_{l=1}^{L-1} \sum_{j=1}^{i-1} a_j^{l\otimes} \cdot r_{a,i-j} \cdot \nu_{L-(l+1)} \right], \tag{5.56}$$

where $\nu_x$, $x = 0,...,L-1$ delivers the probability that there are $L-x$ customers missing at the end of a service process in order to trigger the succeeding service process. Thus, the idle time can be calculated by the sum of one residual interarrival time period and $L-x-1$ full interarrival time periods (see Equation (5.56)). We have to divide the expression in the brackets of Equation (5.56) by the probability that an idle time period at the end of a service process arises. This probability is given by $\sum_{x=0}^{L-1} \nu_x$.

Given the idle time distribution, we can compute the interdeparture time distribution by

$$d_i = b_i \cdot \sum_{x=L}^{y_{max}} \nu_x + \sum_{j=1}^{i-1} \beta_j \cdot b_{i-j} \sum_{x=0}^{L-1} \nu_x. \tag{5.57}$$

### 5.4.4. Waiting Time Distribution

Subsequently, we present an approximate approach to calculate the waiting time distribution of an arbitrary customer, denoted by $w_i$. As in the derivation of $d_i$ we approximate the residual interarrival time by the residual lifetime of a renewal process.

The number of customers at the end of a batch service operation, distributed by $\nu_x$ (see Equation (5.54)) marks the initial situation for the subsequent analysis. First, we derive the waiting time depending on the fact that $N(\delta_n)$ customers can be observed at the end of the $n$th service process. We analyze the waiting time of the customers who arrive within the time period elapsed from the service start to its end. The principle procedure for deriving $w_i$ is similar to the derivation of $w_i$ for the collecting process according to the minimum batch size rule (see Section 4.5.3).

First, we consider the situation that $N(\delta_n) < L$. If $N(\delta_n) < L$, an idle time period follows and the time period from the beginning of the succeeding $(n+1)$th service process to the next arrival is a complete interarrival time interval, distributed by $a_i$. We number the arrivals starting at the beginning of a service period with $q = 0,1,\ldots$. The value of $l$ denotes the number of arrivals which are present in the queue at the end of a service process. If $l < L$, the $l$ customers in the queue have to wait for an additional time period until the $L-l$ still missing customers have arrived. If $L \leq l \leq K$, the waiting time for these $l$ customers ends with the completion of the service process. However, if $l > K$, the $q$th arrival with $q > K$ has to wait until the completion of further $\lceil q/K \rceil - 1$ service periods. The analytical expressions for the described cases are given in the following.

$P(W^q = i|N(\delta_n) < L)$ denotes the probability that the waiting time of the $q$th arrival after the start of the $(n+1)$th service period is $i$ time units under the condition that there were $N(\delta_n) < L$ customers present at the end of the preceding service process.

If $N(\delta_n) < L$, we get for $l = 1,\ldots,L-2$

$$\begin{aligned}
&P(W^l = k+j+i|N(\delta_n) < L) = P(W^{l+1} = k|N(\delta_n) < L) = P(W^L = 0|N(\delta_n) < L) \\
&\quad\sim \sum_{m=b_{min}}^{b_{max}} b_m \left[ \sum_{i=0}^{m-1} \sum_{j=1}^{a_{max}-i} a_{m-i}^{l\otimes} \cdot a_{i+j} \sum_{k=1}^{k_{max}} a_k^{(L-(l+1))\otimes} \right] \\
&\quad k_{max} = (L-(l-1)) \cdot a_{max}
\end{aligned} \tag{5.58}$$

$$\begin{aligned}
&P(W^q = k+j+m-s|N(\delta_n) < L) \\
&\quad\sim \sum_{m=b_{min}}^{b_{max}} b_m \left[ \sum_{s=1}^{m-1} a_s^{q\otimes} \sum_{i=0}^{m-s-1} \sum_{j=1}^{a_{max}-i} a_{m-i-s}^{(l-q)\otimes} \cdot a_{i+j} \sum_{k=1}^{k_{max}} a_k^{(L-(l+1))\otimes} \right] \\
&\quad \text{for } q = 1,\ldots,l-1
\end{aligned} \tag{5.59}$$

and

$$\begin{aligned}
&P(W^q = k-h|N(\delta_n) < L) \\
&\quad\sim \sum_{m=b_{min}}^{b_{max}} b_m \left[ \sum_{i=0}^{m-1} \sum_{j=1}^{a_{max}-i} a_{m-i}^{l\otimes} \cdot a_{i+j} \sum_{h=1}^{h_{max}} \sum_{k=h+1}^{k_{max}} a_h^{(q-(l+1))\otimes} a_{k-h}^{(L-q)\otimes} \right] \\
&\quad \text{for } q = l+2,\ldots,L-1, \quad h_{max} = k_{max} - 1,
\end{aligned} \tag{5.60}$$

In Figure 5.17 the waiting times of the arriving customers are illustrated if $N(\delta_n) < L$ and $l < L-1$.

For $l = L-1$, we obtain

$$\begin{aligned}
&P(W^q = j+m-s|N(\delta_n) < L) \sim \sum_{m=b_{min}}^{b_{max}} b_m \left[ \sum_{s=1}^{m-1} a_s^{q\otimes} \sum_{i=0}^{m-s-1} \sum_{j=1}^{a_{max}-i} a_{m-i-s}^{(l-q)\otimes} \cdot a_{i+j} \right] \\
&\quad \text{for } q = 1,\ldots,l-1
\end{aligned}$$

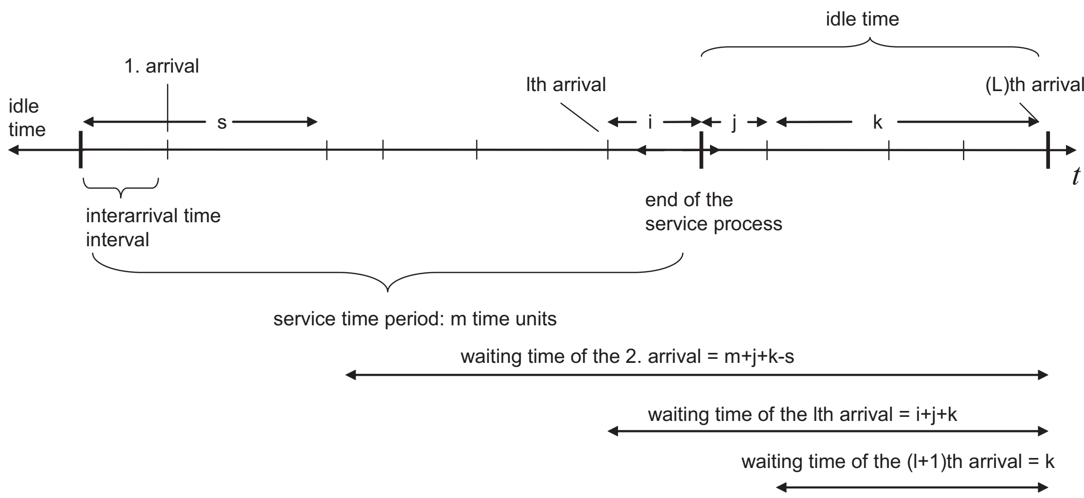


Figure 5.17.: G/G$^{[L,K]}$/1-batch server operating under the minimum batch size rule; illustration of the waiting times if $N(\delta_n) < L$ and $l < L-1$

(5.61)

and

$$P(W^l = j+i|N(\delta_n) < L) = P(W^L = 0|N(\delta_n) < L)$$
$$\sim \sum_{m=b_{min}}^{b_{max}} b_m \left[ \sum_{i=0}^{m-1} \sum_{j=1}^{a_{max}-i} a_{m-i}^{l\otimes} \cdot a_{i+j} \right]. \tag{5.62}$$

For $L \leq l \leq K$, there is no idle time after the completion of the $(n+1)$th service process. The $(n+2)$th service process starts immediately. Hence, we get

$$P(W^q = m-s|N(\delta_n) < L) \sim \sum_{m=b_{min}}^{b_{max}} b_m \left[ \sum_{s=1}^{m-1} a_s^{q\otimes} \sum_{i=0, i<a_{max}}^{m-s-1} a_{m-i-s}^{(l-q)\otimes} \cdot \overline{a}_{i+1} \right] \tag{5.63}$$
for $q = 1, \ldots, l-1$

and

$$P(W^l = i|N(\delta_n) < L) \sim \sum_{m=b_{min}}^{b_{max}} b_m \left[ \sum_{i=0, i<a_{max}}^{m-1} a_{m-i}^{l\otimes} \cdot \overline{a}_{i+1} \right]. \tag{5.64}$$

If $l > K$, the $q$th arriving customer, with $q > K$, has to wait for at least $\lceil q/K \rceil - 1$ complete service periods. It results in

$$P(W^q = m-s|N(\delta_n) < L) \sim \sum_{m=b_{min}}^{b_{max}} b_m \left[ \sum_{s=1}^{m-1} a_s^{q\otimes} \sum_{i=0, i<a_{max}}^{m-s-1} a_{m-i-s}^{(l-q)\otimes} \cdot \overline{a}_{i+1} \right] \tag{5.65}$$
for $q = 1, \ldots, K$,

$$P(W^q = m - s + t | N(\delta_n) < L)$$
$$\sim \sum_{t=1}^{t_{max}} b_t^{(\lceil q/K \rceil - 1)\otimes} \sum_{m=b_{min}}^{b_{max}} b_m \left[ \sum_{s=1}^{m-1} a_s^{q\otimes} \sum_{i=0, i<a_{max}}^{m-s-1} a_{m-i-s}^{(l-q)\otimes} \cdot \overline{a}_{i+1} \right] \tag{5.66}$$
$$\text{for } q = K+1, \ldots, l-1 \qquad t_{max} = (\lceil q/K \rceil - 1) \cdot b_{max}$$

and

$$P(W^l = i + t | N(\delta_n) < L) \sim \sum_{t=1}^{t_{max}} b_t^{(\lceil q/K \rceil - 1)\otimes} \sum_{m=b_{min}}^{b_{max}} b_m \left[ \sum_{i=0, i<a_{max}}^{m-1} a_{m-i}^{l\otimes} \cdot \overline{a}_{i+1} \right]. \tag{5.67}$$

Next, we discuss the waiting time if $L \leq N(\delta_n) \leq K$. We have to distinguish between the same cases again we did if $N(\delta_n) < L$. However, the time period from the start of a service process to the succeeding arrival instant is a residual of the interarrival time. We approximate this residual time, as in Section 5.4.1, by the residual lifetime of a renewal process. The residual lifetime is denoted by $r_{a,i}$. Otherwise, the waiting time can be derived analogically as before (see Equations from (5.58) to (5.67)).
For $l = 1$, we obtain

$$P(W^l = k + j + i | L \leq N(\delta_n) \leq K) = P(W^{l+1} = k | L \leq N(\delta_n) \leq K)$$
$$= P(W^L = 0 | L \leq N(\delta_n) \leq K) \sim \sum_{m=b_{min}}^{b_{max}} b_m \left[ \sum_{i=0}^{m-1} \sum_{j=1}^{a_{max}-i} r_{a,m-i} \cdot a_{i+j} \sum_{k=1}^{k_{max}} a_k^{(L-(l+1))\otimes} \right] \tag{5.68}$$

and

$$P(W^q = k - h | L \leq N(\delta_n) \leq K)$$
$$\sim \sum_{m=b_{min}}^{b_{max}} b_m \left[ \sum_{i=0}^{m-1} \sum_{j=1}^{a_{max}-i} r_{a,m-i} \cdot a_{i+j} \sum_{h=1}^{h_{max}} \sum_{k=h+1}^{k_{max}} a_h^{(q-(l+1))\otimes} a_{k-h}^{(L-q)\otimes} \right] \tag{5.69}$$
$$\text{for } q = l+2, \ldots, L-1.$$

For $l = 2, \ldots, L-2$, we get

$$P(W^1 = k + j + i - u | L \leq N(\delta_n) \leq K) + P(W^l = k + j + i | L \leq N(\delta_n) \leq K)$$
$$= P(W^{l+1} = k | L \leq N(\delta_n) \leq K) = P(W^L = 0 | L \leq N(\delta_n) \leq K)$$
$$\sim \sum_{m=b_{min}}^{b_{max}} b_m \left[ \sum_{u=1}^{a_{max}} r_{a,u} \sum_{i=0}^{m-u-1} \sum_{j=1}^{a_{max}-i} a_{m-u-i}^{(l-1)\otimes} \cdot a_{i+j} \sum_{k=1}^{k_{max}} a_k^{(L-(l+1))\otimes} \right], \tag{5.70}$$

$$P(W^q = k + j + m - u - s | L \leq N(\delta_n) \leq K)$$
$$\sim \sum_{m=b_{min}}^{b_{max}} b_m \left[ \sum_{u=1}^{a_{max}} r_{a,u} \sum_{s=1}^{m-u-1} a_s^{(q-1)\otimes} \sum_{i=0}^{m-u-s-1} \sum_{j=1}^{a_{max}-i} a_{m-i-u-s}^{(l-q)\otimes} \cdot a_{i+j} \sum_{k=1}^{k_{max}} a_k^{(L-(l+1))\otimes} \right]$$
$$\text{for } q = 2, \ldots, l-1$$

(5.71)

and

$$P(W^q = k-h|L \le N(\delta_n) \le K)$$
$$\sim \sum_{m=b_{min}}^{b_{max}} b_m \left[ \sum_{u=1}^{a_{max}} r_{a,u} \sum_{i=0}^{m-u-1} \sum_{j=1}^{a_{max}-i} a_{m-u-i}^{(l-1)\otimes} a_{i+j} \sum_{h=1}^{h_{max}} \sum_{k=h+1}^{k_{max}} a_h^{(q-(l+1))\otimes} a_{k-h}^{(L-q)\otimes} \right].$$
for $q = l+2, \ldots, L-1$.

(5.72)

In Figure 5.18 the waiting times for the case $L \le N(\delta_n) \le K$ and $l = 2, \ldots, L-2$ are depicted.

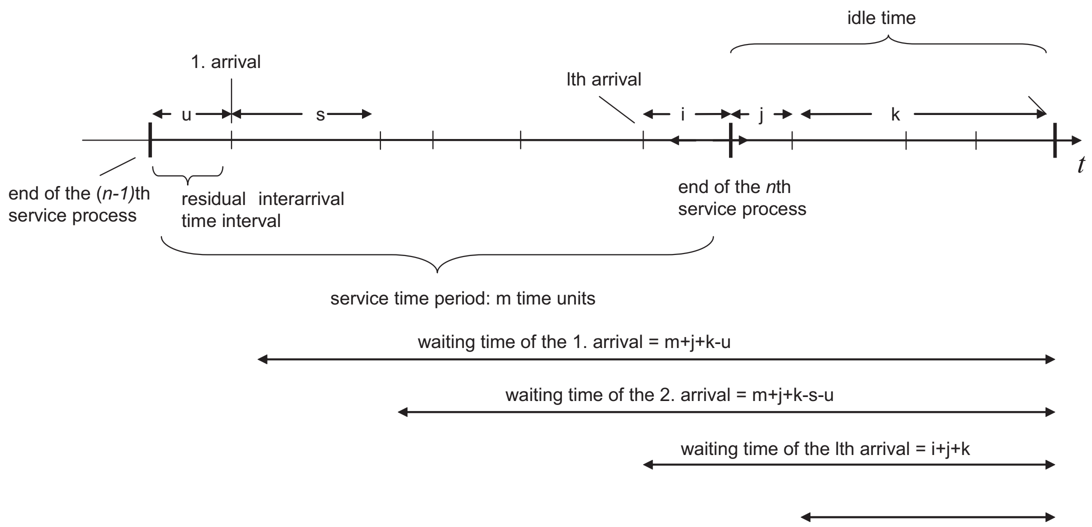


Figure 5.18.: G/G$^{[L,K]}$/1-batch server operating under the minimum batch size rule; illustration of the waiting times if $L \le N(\delta_n) \le K$ and $l = 2, \ldots, L-2$

Furthermore, for $l = L-1$, we obtain

$$P(W^1 = m-u|L \le N(\delta_n) \le K) = P(W^l = j+i|L \le N(\delta_n) \le K)$$
$$= P(W^L = 0|L \le N(\delta_n) \le K) \sim \sum_{m=b_{min}}^{b_{max}} b_m \left[ \sum_{u=1}^{a_{max}} r_{a,u} \sum_{i=0}^{m-u-1} \sum_{j=1}^{a_{max}-i} a_{m-u-i}^{(l-1)\otimes} \cdot a_{i+j} \right],$$

(5.73)

and

$$P(W^q = j+m-s|L \le N(\delta_n) \le K)$$
$$\sim \sum_{m=b_{min}}^{b_{max}} b_m \left[ \sum_{u=1}^{a_{max}} r_{a,u} \sum_{s=1}^{m-u-1} a_s^{(q-1)\otimes} \sum_{i=0}^{m-u-s-1} \sum_{j=1}^{a_{max}-i} a_{m-i-s}^{(l-q)\otimes} \cdot a_{i+j} \right] \quad (5.74)$$
for $q = 2, \ldots, l-1$,

For $L \leq l \leq K$, it yields

$$P(W^1 = m-u|L \leq N(\delta_n) \leq K) = P(W^l = i|L \leq N(\delta_n) \leq K)$$
$$\sim \sum_{m=b_{min}}^{b_{max}} b_m \left[ \sum_{u=1}^{a_{max}} r_{a,u} \sum_{i=0,i<a_{max}}^{m-u-1} a_{m-u-i}^{(l-1)\otimes} \cdot \overline{a}_{i+1} \right] \tag{5.75}$$

and

$$P(W^q = m-s|L \leq N(\delta_n) \leq K)$$
$$\sim \sum_{m=b_{min}}^{b_{max}} b_m \left[ \sum_{u=1}^{a_{max}} r_{a,u} \sum_{s=1}^{m-u-1} a_s^{(q-1)\otimes} \sum_{i=0,i<a_{max}}^{m-u-s-1} a_{m-u-i-s}^{(l-q)\otimes} \cdot \overline{a}_{i+1} \right] \tag{5.76}$$
for $q = 2,\ldots,l-1$.

If $l > K$, the $q$th arriving customer, with $q > K$, has to wait for additional $\lceil q/K \rceil - 1$ complete service periods after the end of the $(n+1)$th service process. It results in

$$P(W^1 = m-u|L \leq N(\delta_n) \leq K) \sim \sum_{m=b_{min}}^{b_{max}} b_m \left[ \sum_{u=1}^{a_{max}} r_{a,u} \sum_{i=0,i<a_{max}}^{m-u-1} a_{m-u-i}^{(l-1)\otimes} \cdot \overline{a}_{i+1} \right], \tag{5.77}$$

$$P(W^q = m-s-u|L \leq N(\delta_n) \leq K)$$
$$\sim \sum_{m=b_{min}}^{b_{max}} b_m \left[ \sum_{u=1}^{a_{max}} r_{a,u} \sum_{s=1}^{m-u-1} a_s^{(q-1)\otimes} \sum_{i=0,i<a_{max}}^{m-u-s-1} a_{m-u-i-s}^{(l-q)\otimes} \cdot \overline{a}_{i+1} \right] \tag{5.78}$$
for $q = 2,\ldots,K$,

$$P(W^q = m-u-s+t|L \leq N(\delta_n) \leq K)$$
$$\sim \sum_{t=1}^{t_{max}} b_t^{(\lceil q/K \rceil -1)\otimes} \sum_{m=b_{min}}^{b_{max}} b_m \left[ \sum_{u=1}^{a_{max}} r_{a,u} \sum_{s=1}^{m-u-1} a_s^{(q-1)\otimes} \sum_{i=0,i<a_{max}}^{m-u-s-1} a_{m-u-i-s}^{(l-q)\otimes} \cdot \overline{a}_{i+1} \right]$$
for $q = K+1,\ldots,l-1$
$$\tag{5.79}$$

and

$$P(W^l = i+t|L \leq N(\delta_n) \leq K)$$
$$\sim \sum_{t=1}^{t_{max}} b_t^{(\lceil l/K \rceil -1)\otimes} \sum_{m=b_{min}}^{b_{max}} b_m \left[ \sum_{u=1}^{a_{max}} r_{a,u} \sum_{i=0,i<a_{max}}^{m-u-1} a_{m-u-i}^{(l-1)\otimes} \cdot \overline{a}_{i+1} \right]. \tag{5.80}$$

Subsequently, we consider the case that there are already $z \geq 1$ customers waiting at the start of the $(n+1)$th service process. It follows that $N(\delta_n) = K+z$. The derivation of the waiting time is analogous to the case $L \leq N(\delta_n) \leq K$, except that there are $z$ arrivals less required to trigger the succeeding service process. If $z \geq K$, all the arriving customers during the $(n+1)$th service have to wait for at least a complete service period. The time period from the start of the $(n+1)$th service process to the first arrival instant is a residual of the interarrival time, distributed by $r_{a,i}$.

An idle time period occurring after the completion of the considered service process is possible if $l < L - z$. This case is modeled by the Expressions from (5.81) to (5.87). In the case of an idle time period, we have to differentiate between $l < L - 1 - z$ and $l = L - 1 - z$. If $l < L - 1 - z$, there is more than one customer missing at the end of the $(n+1)$th service process and if $l = L - 1 - z$, there is exactly one customer missing. Thus, for $l = 1$ and $l < L - 1 - z$, we obtain

$$\begin{aligned} &P(W^l = k + j + i | N(\delta_n) = K + z) = P(W^{l+1} = k | N(\delta_n) = K + z) \\ &= P(W^{L-z} = 0 | l = K + z) \sim \sum_{m=b_{min}}^{b_{max}} b_m \left[ \sum_{i=0}^{m-1} \sum_{j=1}^{a_{max}-i} r_{a,m-i} \cdot a_{i+j} \sum_{k=1}^{k_{max}} a_k^{(L-(l+1)-z)\otimes} \right] \\ &\qquad k_{max} = (L - (l+1) - z) \cdot a_{max} \end{aligned} \tag{5.81}$$

and

$$\begin{aligned} &P(W^q = k - h | N(\delta_n) = K + z) \\ &\quad \sim \sum_{m=b_{min}}^{b_{max}} b_m \left[ \sum_{i=0}^{m-1} \sum_{j=1}^{a_{max}-i} r_{a,m-i} \cdot a_{i+j} \sum_{h=1}^{h_{max}} \sum_{k=h+1}^{k_{max}} a_h^{(q-(l+1))\otimes} a_{k-h}^{(L-q-z)\otimes} \right] \\ &\quad \text{for } q = l+2, \ldots, L - z - 1. \end{aligned} \tag{5.82}$$

Furthermore, for $l = 2, \ldots, L - 2 - z$, it yields

$$\begin{aligned} &P(W^1 = k + j + i - u | N(\delta_n) = K + z) + P(W^l = k + j + i | N(\delta_n) = K + z) \\ &= P(W^{l+1} = k | N(\delta_n) = K + z) = P(W^{L-z} = 0 | N(\delta_n) = K + z) \\ &\quad \sim \sum_{m=b_{min}}^{b_{max}} b_m \left[ \sum_{u=1}^{a_{max}} r_{a,u} \sum_{i=0}^{m-u-1} \sum_{j=1}^{a_{max}-i} a_{m-u-i}^{(l-1)\otimes} \cdot a_{i+j} \sum_{k=1}^{k_{max}} a_k^{(L-(l+1)-z)\otimes} \right], \end{aligned} \tag{5.83}$$

$$\begin{aligned} &P(W^q = k + j + m - u - s | N(\delta_n) = K + z) \\ &\sim \sum_{m=b_{min}}^{b_{max}} b_m \left[ \sum_{u=1}^{a_{max}} r_{a,u} \sum_{s=1}^{m-u-1} a_s^{(q-1)\otimes} \sum_{i=0}^{m-u-s-1} \sum_{j=1}^{a_{max}-i} a_{m-i-u-s}^{(l-q)\otimes} \cdot a_{i+j} \sum_{k=1}^{k_{max}} a_k^{(L-(l+1)-z)\otimes} \right] \\ &\text{for } q = 2, \ldots, l - z - 1 \end{aligned} \tag{5.84}$$

and

$$\begin{aligned} &P(W^q = k - h | N(\delta_n) = K + z) \\ &\quad \sim \sum_{m=b_{min}}^{b_{max}} b_m \left[ \sum_{u=1}^{a_{max}} r_{a,u} \sum_{i=0}^{m-u-1} \sum_{j=1} a_{m-u-i}^{(l-1)\otimes} a_{i+j} \sum_{h=1}^{h_{max}} \sum_{k=h+1}^{k_{max}} a_h^{(q-(l+1))\otimes} a_{k-h}^{(L-q-z)\otimes} \right]. \\ &\quad \text{for } q = l+2, \ldots, L - z - 1. \end{aligned}$$

(5.85)

In Figure 5.19 the waiting times for the case $N(\delta_n) = K + z$ and $l = 2, \ldots ,L - 2 - z$ are depicted.

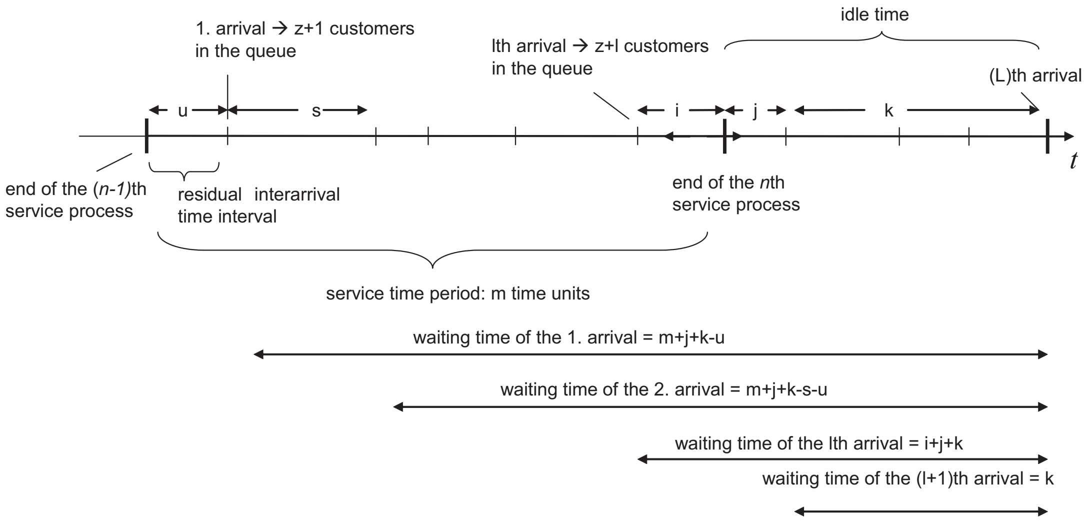


Figure 5.19.: G/G$^K$/1-batch server operating under the minimum batch size rule; illustration of the waiting times if $N(\delta_n) = K + z$ and $l = 2, \ldots ,L - 2 - z$

If $l = L-1-z$, there are $L-1$ customers waiting in the queue at the end of the $(n+1)$th service process. The service station will start to work when the next customer arrives. The waiting times in this case can be determined by

$$P(W^1 = m - u|N(\delta_n) = K + z) = P(W^l = j + i|N(\delta_n) = K + z)$$
$$= P(W^L = 0|N(\delta_n) = K + z) \sim \sum_{m=b_{min}}^{b_{max}} b_m \left[ \sum_{u=1}^{a_{max}} r_{a,u} \sum_{i=0}^{m-u-1} \sum_{j=1}^{a_{max}-i} a_{m-u-i}^{(l-1)\otimes} \cdot a_{i+j} \right], \tag{5.86}$$

and

$$P(W^q = j + m - s|N(\delta_n) = K + z)$$
$$\sim \sum_{m=b_{min}}^{b_{max}} b_m \left[ \sum_{u=1}^{a_{max}} r_{a,u} \sum_{s=1}^{m-u-1} a_s^{(q-1)\otimes} \sum_{i=0}^{m-s-u-1} \sum_{j=1}^{a_{max}-i} a_{m-i-s}^{(l-q)\otimes} \cdot a_{i+j} \right] \tag{5.87}$$

for $q = 2, \ldots ,l - 1$.

However, if $l \geq L-z$, it results in no idle time period after the completion of the $(n+1)$th service process.

If $L - z \leq l \leq K - z$, all customers who arrive during the $(n + 1)$th service period, can be served by the succeeding service process. We derive

$$P(W^1 = m - u|N(\delta_n) = K + z) = P(W^l = i|N(\delta_n) = K + z)$$
$$\sim \sum_{m=b_{min}}^{b_{max}} b_m \left[ \sum_{u=1}^{a_{max}} r_{a,u} \sum_{i=0,i<a_{max}}^{m-u-1} a_{m-u-i}^{(l-1)\otimes} \cdot \overline{a}_{i+1} \right] \tag{5.88}$$

and

$$P(W^q = m - s - u | N(\delta_n) = K + z)$$
$$\sim \sum_{m=b_{min}}^{b_{max}} b_m \left[ \sum_{u=1}^{a_{max}} r_{a,u} \sum_{s=1}^{m-u-1} a_s^{(q-1)\otimes} \sum_{i=0, i<a_{max}}^{m-u-s-1} a_{m-u-i-s}^{(l-q)\otimes} \cdot \overline{a}_{i+1} \right] \tag{5.89}$$
for $q = 2, \ldots, l-1$.

For $l > K - z$ and $z < K$, the customers who arrive at position $q = 1, \ldots, K - z$ during the $(n+1)$th service operation can be served by the succeeding service process. These customers who arrive at position $q > K - z$ have to wait for further $\lceil (q+z)/K \rceil - 1$ service processes. It yields

$$P(W^1 = m - u | N(\delta_n) = K + z) \sim \sum_{m=b_{min}}^{b_{max}} b_m \left[ \sum_{u=1}^{a_{max}} r_{a,u} \sum_{i=0, i<a_{max}}^{m-u-1} a_{m-u-i}^{(l-1)\otimes} \cdot \overline{a}_{i+1} \right], \tag{5.90}$$

$$P(W^q = m - s - u | N(\delta_n) = K + z)$$
$$\sim \sum_{m=b_{min}}^{b_{max}} b_m \left[ \sum_{u=1}^{a_{max}} r_{a,u} \sum_{s=1}^{m-u-1} a_s^{(q-1)\otimes} \sum_{i=0, i<a_{max}}^{m-u-s-1} a_{m-u-i-s}^{(l-q)\otimes} \cdot \overline{a}_{i+1} \right] \tag{5.91}$$
for $q = 2, \ldots, K - z$,

$$P(W^q = m - s - u + t | N(\delta_n) = K + z)$$
$$\sim \sum_{t=1}^{t_{max}} b_t^{(\lceil (q+z)/K \rceil - 1)\otimes} \sum_{m=b_{min}}^{b_{max}} b_m \left[ \sum_{u=1}^{u_{max}} r_{a,u} \sum_{s=1}^{m-u-1} a_s^{(q-1)\otimes} \sum_{i=0, i<a_{max}}^{m-u-s-1} a_{m-u-i-s}^{(l-q)\otimes} \cdot \overline{a}_{i+1} \right]$$
for $q = K - z + 1, \ldots, l-1$.
$$\tag{5.92}$$

and

$$P(W^l = i + t | N(\delta_n) = K + z)$$
$$\sim \sum_{t=1}^{t_{max}} b_t^{(\lceil (l+z)/K \rceil - 1)\otimes} \sum_{m=b_{min}}^{b_{max}} b_m \left[ \sum_{u=1}^{a_{max}} r_{a,u} \sum_{i=0, i<a_{max}}^{m-u-1} a_{m-u-i}^{(l-1)\otimes} \cdot \overline{a}_{i+1} \right]. \tag{5.93}$$

Finally, if $z \geq K$, it follows that the first arriving customer has to wait for additional $\lfloor z/K \rfloor$ complete service processes. Thus, we obtain

$$P(W^1 = m - u + t | N(\delta_n) = K + z)$$
$$\sim \sum_{t=1}^{t_{max}} b_t^{(\lfloor (z)/K \rfloor)\otimes} \sum_{m=b_{min}}^{b_{max}} b_m \left[ \sum_{u=1}^{a_{max}} r_{a,u} \sum_{i=0, i<a_{max}}^{m-u-1} a_{m-u-i}^{(l-1)\otimes} \cdot \overline{a}_{i+1} \right], \tag{5.94}$$

$$P(W^q = m-s-u+t|N(\delta_n) = K+z)$$
$$\sim \sum_{t=1}^{t_{max}} b_t^{(\lceil (q+z)/K \rceil -1)\otimes} \sum_{m=b_{min}}^{b_{max}} b_m \left[ \sum_{u=1}^{a_{max}} r_{a,u} \sum_{s=1}^{m-u-1} a_s^{(q-1)\otimes} \sum_{i=0, i<a_{max}}^{m-u-s-1} a_{m-u-i-s}^{(l-q)\otimes} \cdot \overline{a}_{i+1} \right]$$
$$\text{for } q = 2, \ldots, l-1 \tag{5.95}$$

and

$$P(W^l = i+t|N(\delta_n) = K+z)$$
$$\sim \sum_{t=1}^{t_{max}} b_t^{(\lceil (l+z)/K \rceil -1)\otimes} \sum_{m=b_{min}}^{b_{max}} b_m \left[ \sum_{u=1} r_{a,u}^{a_{max}} \sum_{i=0, i<a_{max}}^{m-u-1} a_{m-u-i}^{(l-1)\otimes} \cdot \overline{a}_{i+1} \right]. \tag{5.96}$$

We derived by means of the Expressions from (5.58) to (5.96) proportional values for $P(W^q = i|N(\delta_n) < L)$, $P(W^q = i|L \leq N(\delta_n) \leq K)$ and $P(W^q = i|N(\delta_n) = K+z)$. Let us denote these proportional expressions by $P^*(W^q = i|N(\delta_n) < L)$, $P(^*W^q = i|L \leq N(\delta_n) \leq K)$ and $P^*(W^q = i|N(\delta_n) = K+z)$. Thus, after normalization of these expressions we obtain

$$P(W = i|N(\delta_n) < L) = \frac{\sum_{q=1}^{\max\{\frac{b_{max}}{a_{min}}, L\}} P^*(W^q = i|N(\delta_n) < L)}{\sum_{i=0} \sum_{q=1}^{\max\{\frac{b_{max}}{a_{min}}, L\}} P^*(W^q = i|N(\delta_n) < L)}, \tag{5.97}$$

$$P(W = i|L \leq N(\delta_n) \leq K) = \frac{\sum_{q=1}^{\max\{\frac{b_{max}}{a_{min}}, L\}} P^*(W^q = i|L \leq N(\delta_n) \leq K)}{\sum_{i=0} \sum_{q=1}^{\max\{\frac{b_{max}}{a_{min}}, L\}} P^*(W^q = i|L \leq N(\delta_n) \leq K)} \tag{5.98}$$

and

$$P(W = i|N(\delta_n) = K+z) = \frac{\sum_{q=1}^{\max\{\frac{b_{max}}{a_{min}}, L\}} P^*(W^q = i|N(\delta_n) = K+z)}{\sum_{i=0} \sum_{q=1}^{\max\{\frac{b_{max}}{a_{min}}, L\}} P^*(W^q = i|N(\delta_n) = K+z)}. \tag{5.99}$$

Finally, we calculate the waiting time distribution by

$$\begin{aligned} w_i = P(W = i) &= P(W = i|N(\delta_n) < L) \cdot P(N(\delta_n) < L) \\ &+ P(W = i|L \leq N(\delta_n) \leq K) \cdot P(L \leq N(\delta_n) \leq K) \\ &+ \sum_{z=1}^{z_{max}} P(W = i|N(\delta_n) = K+z) \cdot P(N(\delta_n) = K+z) \\ &= P(W = i|N(\delta_n) < L) \cdot \sum_{x=0}^{L-1} \nu_x \\ &+ P(W = i|L \leq N(\delta_n) \leq K) \cdot \sum_{x=L}^{K} \nu_x \\ &+ \sum_{z=1}^{z_{max}} P(W = i|N(\delta_n) = K+z) \cdot \nu_{K+z}, \end{aligned} \tag{5.100}$$

where $\nu_x$ is obtained by Equation (5.54).

### 5.4.5. Analysis

Since we presented an approximate solution for the analysis of G/G$^{[L,K]}$/1-queue, we have to evaluate the quality of the introduced method. Therefore, we compare the analytical results derived by the described analytical models to results obtained by simulation. For all examples we calculate the values for $\nu_x$, $y_{d,x}$, $d_i$, and $w_i$. The resulting distributions of four different examples[3] are depicted in Figure 5.20 and 5.21. We find the results very promising since the deviations are small. Table 5.1 gives an overview of the obtained numerical results. In this table the mean values, scv, 95% and 99%-quantiles are listed. In addition, the deviations of the analytical results from the simulation results are presented.

Convinced of the accuracy of the analytical approach we analyze the influence of input parameters on output figures analogous to the analysis we performed in Section 4.5.4.

First, we investigate the dependency of performance figures on the minimum batch size $L$. Figure 5.22 shows the dependency of the waiting time on $L$ for a numerical example[4]. As expected, the waiting time increases disproportionately with increasing $L$. The same result is observed for the 95%-quantile. In addition, Figure 5.23 illustrates the mean batch size and the probability that the server becomes idle at the end of a service operation depending on $L$. Again, both output figures increase disproportionately with increasing $L$. To gain further insight into the system's behavior of the G/G$^{[L,K]}$/1-queue, we show in Figure 5.24 the waiting time distribution for three different configurations of $L$, namely, for $L = 5$, $L = 10$, and $L = 15$. Since the probability that the server becomes idle at the end of a service operation increases disproportionately with increasing $L$, the probability that the waiting time is zero increases too. When the server becomes idle it always follows that exactly one customer of the batch has the waiting time zero. If the number of customers at the departure instant is low, the waiting time for the customers who already arrived could be long if $L = 15$. This fact is also observable in Figure 5.24. It is required to note that the higher the utilization, the closer the system's behavior of the G/G$^{[L,K]}$/1-queue with low values of $L$ is to the system's behavior of the G/G$^{[L,K]}$/1-queue with high values of $L$. If the utilization is close to one, the G/G$^{[L,K]}$/1-queue serves almost batches of size $K$, which is close to the system's behavior of the G/G$^{[K,K]}$/1-queue.

Studying the batch building mode under the minimum batch size rule in Section 4.5.4 we detected a paradox of decreasing $E(W)$ with an increasing variability within a certain range of $c_A^2$ and $c_Y^2$. Now, we performed a set of numerical examples in which we increased $c_A^2$ and investigated the influence on different performance figures. In contrast to the batch building mode under the minimum batch size rule, the process behavior of the G/G$^{[L,K]}$/1-queue is as expected and shows no paradoxical behavior. In order to illustrate the process behavior of the G/G$^{[L,K]}$/1-queue depending on $c_A^2$, we present subsequently a numerical example. The input data of this example is given in Table ... in the Appendix.

Figure 5.25 shows the mean waiting time and the 95% and 99%-quantile of the waiting time distribution depending on $c_A^2$. These performance figures increase slightly disproportionately with increasing $c_A^2$. The behavior of the number of customer in the system at the departure instant is analogous (see Figure 5.26). Finally, it is observable in Figure 5.27 that both the mean batch size of the served batch and the mean interdeparture time

[3]The input values of theses four examples are given in Table A.5 in the Appendix

[4]We choose the input values of example 4 given in Table A.5 in the Appendix

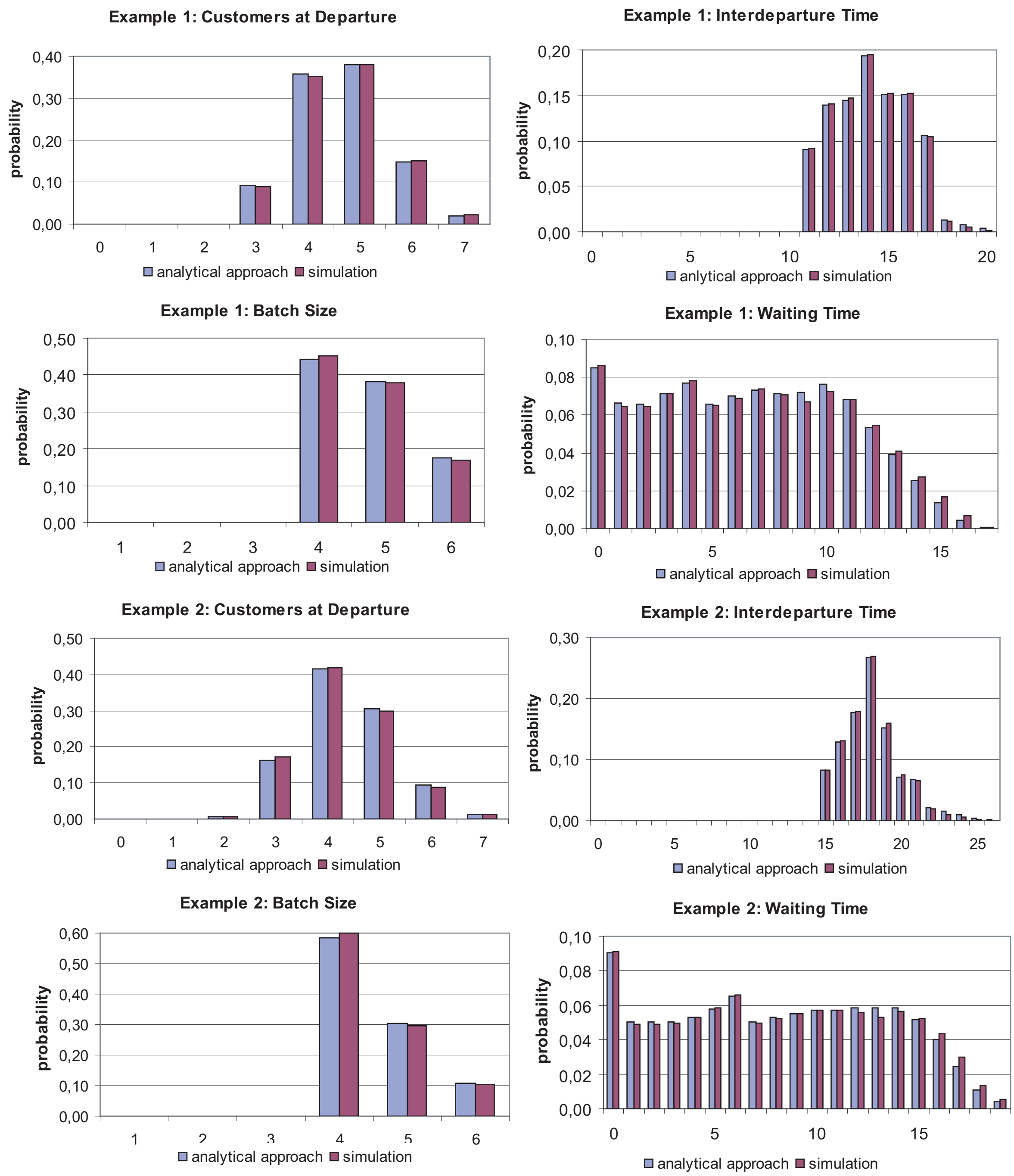


Figure 5.20.: Examples 1 and 2: Analysis of the G/G$^{[L,K]}$/1-queue; comparison of analytical results with simulation results

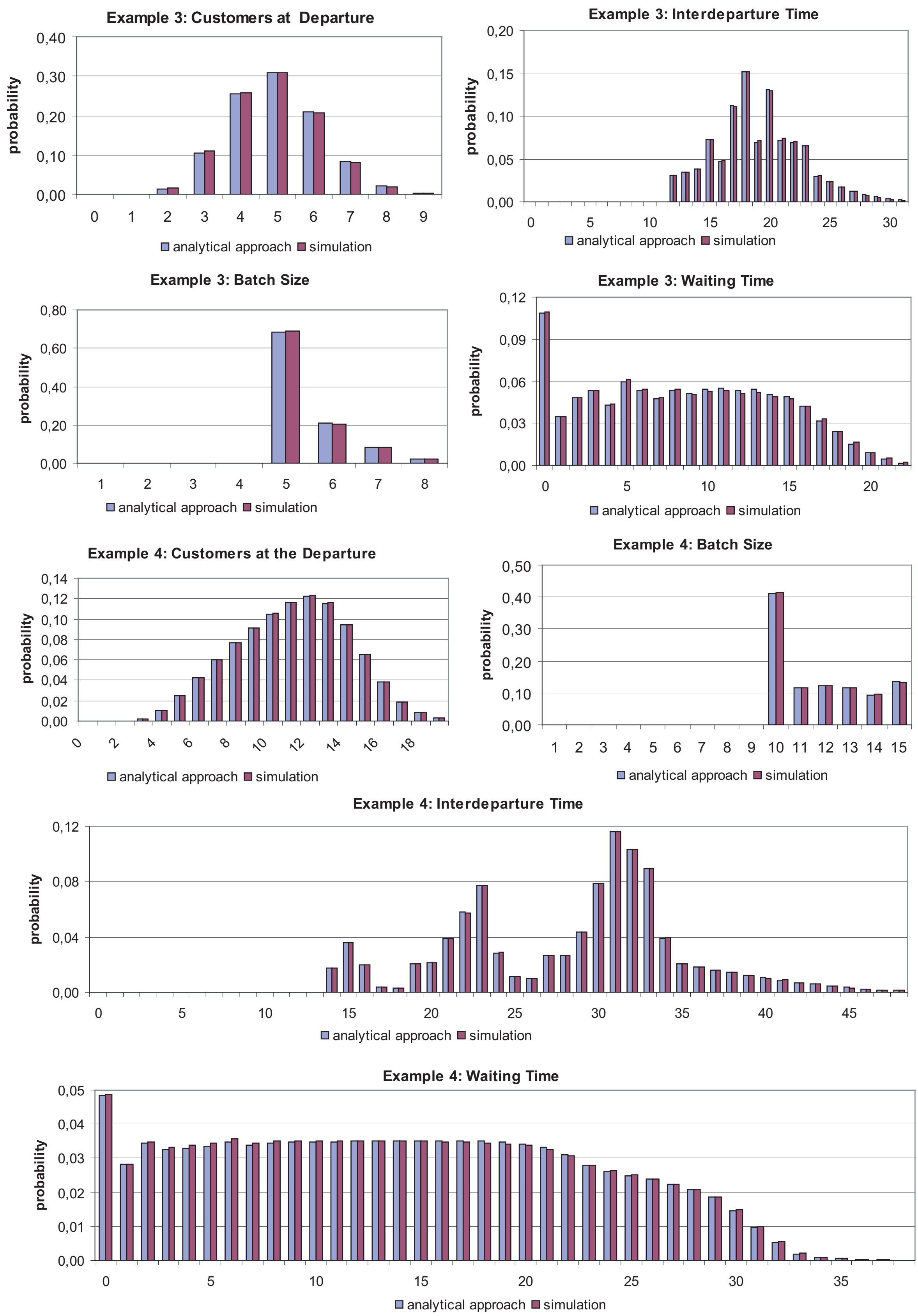


Figure 5.21.: Examples 3 and 4: Analysis of the G/G$^{[L,K]}$/1-queue; comparison of analytical results with simulation results

| example 1 | $N_d$ | | | | | $D$ | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | $E$ | $E^2$ | scv | $\sigma_{0.9}$ | $\sigma_{0.95}$ | $E$ | $E^2$ | scv | $\sigma_{0.9}$ | $\sigma_{0.95}$ |
| analy | 4.661 | 22.589 | 0.040 | 5 | 5 | 14.191 | 205.176 | 0.019 | 17 | 17 |
| sim | 4.643 | 22.400 | 0.039 | 5 | 5 | 14.149 | 203.760 | 0.018 | 17 | 17 |
| $\Delta_{abs}$ | 0.018 | 0.189 | 0.001 | | | 0.043 | 1.416 | 0.001 | | |
| $\Delta_{rel}$ | 0.004 | 0.008 | 0.018 | | | 0.003 | 0.007 | 0.047 | | |
| | $y_d$ | | | | | $W$ | | | | |
| | $E$ | $E^2$ | scv | $\sigma_{0.9}$ | $\sigma_{0.95}$ | $E$ | $E^2$ | scv | $\sigma_{0.9}$ | $\sigma_{0.95}$ |
| analy | 4.728 | 22.916 | 0.025 | 6 | 6 | 6.508 | 59.941 | 0.415 | 12 | 13 |
| sim | 4.717 | 22.787 | 0.024 | 6 | 6 | 6.577 | 61.389 | 0.419 | 12 | 14 |
| $\Delta_{abs}$ | 0.012 | 0.129 | 0.001 | | | 0.069 | 1.448 | 0.004 | | |
| $\Delta_{rel}$ | 0.002 | 0.006 | 0.029 | | | 0.011 | 0.024 | 0.010 | | |
| example 2 | $N_d$ | | | | | $D$ | | | | |
| | $E$ | $E^2$ | scv | $\sigma_{0.9}$ | $\sigma_{0.95}$ | $E$ | $E^2$ | scv | $\sigma_{0.9}$ | $\sigma_{0.95}$ |
| analy | 4,363 | 19,925 | 0,047 | 6 | 6 | 18,092 | 331,338 | 0,012 | 21 | 22 |
| sim | 4,331 | 19,632 | 0,047 | 6 | 6 | 18,008 | 327,745 | 0,011 | 21 | 21 |
| $\Delta_{abs}$ | 0,033 | 0,293 | 0,000 | | | 0,084 | 3,593 | 0,002 | | |
| $\Delta_{rel}$ | 0,007 | 0,015 | 0,003 | | | 0,005 | 0,011 | 0,130 | | |
| | $y_d$ | | | | | $W$ | | | | |
| | $E$ | $E^2$ | scv | $\sigma_{0.9}$ | $\sigma_{0.95}$ | $E$ | $E^2$ | scv | $\sigma_{0.9}$ | $\sigma_{0.95}$ |
| analy | 4,522 | 20,923 | 0,023 | 6 | 6 | 8,105 | 92,978 | 0,415 | 15 | 16 |
| sim | 4,503 | 20,728 | 0,022 | 6 | 6 | 8,193 | 95,241 | 0,419 | 15 | 17 |
| $\Delta_{abs}$ | 0,020 | 0,195 | 0,001 | | | 0,088 | 2,263 | 0,003 | | |
| $\Delta_{rel}$ | 0,004 | 0,009 | 0,024 | | | 0,011 | 0,024 | 0,008 | | |
| example 3 | $N_d$ | | | | | $D$ | | | | |
| | $E$ | $E^2$ | scv | $\sigma_{0.9}$ | $\sigma_{0.95}$ | $E$ | $E^2$ | scv | $\sigma_{0.9}$ | $\sigma_{0.95}$ |
| analy | 4.936 | 25.970 | 0.066 | 7 | 7 | 19.066 | 377.372 | 0.038 | 24 | 26 |
| sim | 4.916 | 25.752 | 0.065 | 7 | 7 | 19.019 | 374.983 | 0.037 | 24 | 25 |
| $\Delta_{abs}$ | 0.020 | 0.218 | 0.001 | | | 0.047 | 2.389 | 0.001 | | |
| $\Delta_{rel}$ | 0.004 | 0.008 | 0.008 | | | 0.002 | 0.006 | 0.037 | | |
| | $y_d$ | | | | | $W$ | | | | |
| | $E$ | $E^2$ | scv | $\sigma_{0.9}$ | $\sigma_{0.95}$ | $E$ | $E^2$ | scv | $\sigma_{0.9}$ | $\sigma_{0.95}$ |
| analy | 5.447 | 30.229 | 0.019 | 7 | 7 | 8.502 | 104.359 | 0.444 | 16 | 18 |
| sim | 5.436 | 30.089 | 0.018 | 7 | 7 | 8.489 | 104.643 | 0.452 | 16 | 18 |
| $\Delta_{abs}$ | 0.011 | 0.140 | 0.000 | | | 0.013 | 0.284 | 0.008 | | |
| $\Delta_{rel}$ | 0.002 | 0.005 | 0.025 | | | 0.001 | 0.003 | 0.019 | | |
| example 4 | $N_d$ | | | | | $D$ | | | | |
| | $E$ | $E^2$ | scv | $\sigma_{0.9}$ | $\sigma_{0.95}$ | $E$ | $E^2$ | scv | $\sigma_{0.9}$ | $\sigma_{0.95}$ |
| analy | 11.032 | 131.821 | 0.083 | 15 | 15 | 28.520 | 860.200 | 0.058 | 36 | 39 |
| sim | 11.083 | 132.394 | 0.078 | 16 | 16 | 28.513 | 859.626 | 0.057 | 36 | 39 |
| $\Delta_{abs}$ | 0.051 | 0.573 | 0.005 | | | 0.007 | 0.574 | 0.000 | | |
| $\Delta_{rel}$ | 0.005 | 0.004 | 0.064 | | | 0.000 | 0.001 | 0.003 | | |
| | $y_d$ | | | | | $W$ | | | | |
| | $E$ | $E^2$ | scv | $\sigma_{0.9}$ | $\sigma_{0.95}$ | $E$ | $E^2$ | scv | $\sigma_{0.9}$ | $\sigma_{0.95}$ |
| analy | 11.707 | 141.139 | 0.030 | 15 | 15 | 13.960 | 272.260 | 0.397 | 27 | 29 |
| sim | 11.762 | 141.776 | 0.025 | 15 | 15 | 14.002 | 273.353 | 0.394 | 26 | 29 |
| $\Delta_{abs}$ | 0.055 | 0.637 | 0.005 | | | 0.042 | 1.093 | 0.003 | | |
| $\Delta_{rel}$ | 0.005 | 0.005 | 0.171 | | | 0.003 | 0.004 | 0.007 | | |

Table 5.1.: Analysis of the G/G$^{[L,K]}$/1-queue; comparison of analytical results with simulation results

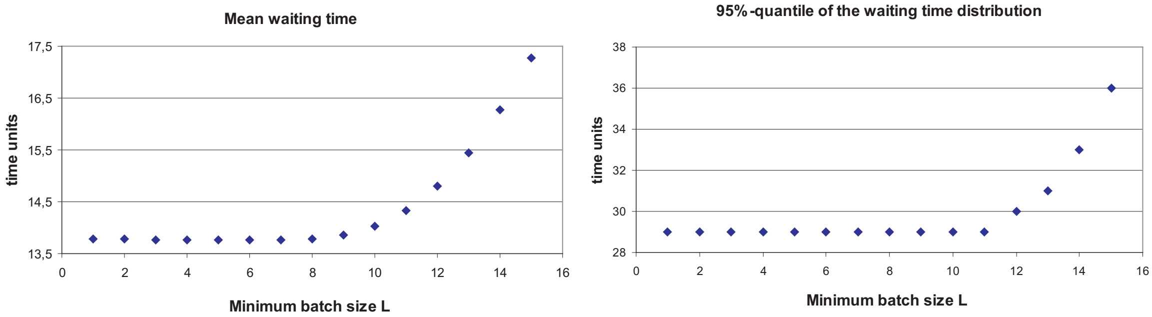


Figure 5.22.: Dependency of the waiting time on the minimum batch size L; left: Mean waiting time; right: 95%-quantile of the waiting time distribution

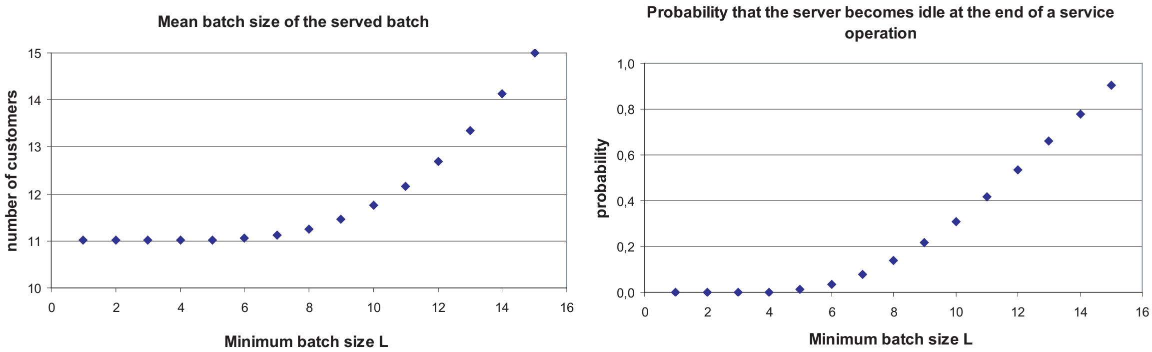


Figure 5.23.: Left: Mean batch size of the served batch depending on the minimum batch size L; right: Probability that the server becomes idle at the end of a service operation depending on the minimum batch size L

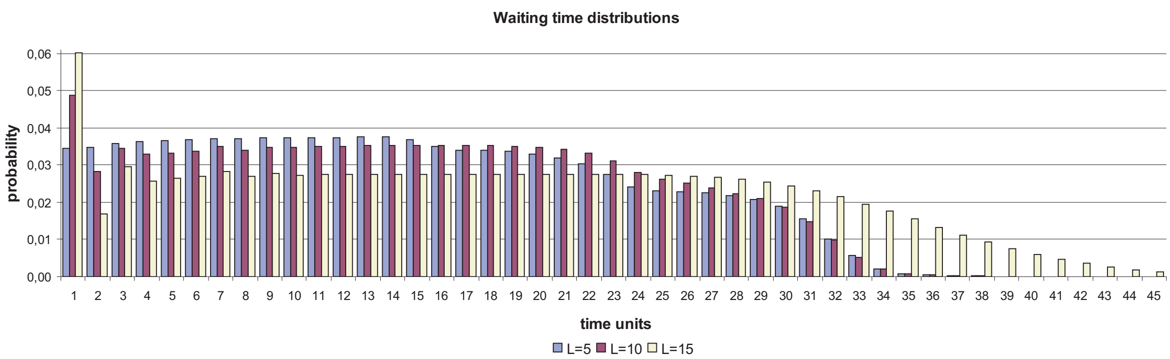


Figure 5.24.: Waiting time distribution for different configurations of the minimum batch size L (L=5, L=10, L=15)

increase linearly depending on $c_A^2$. Since the conservation of flows has to be guaranteed, $E(Y_d)$ and $E(D)$ are directly dependent on each other and the value of $E(D)/E(Y_d)$ has to corresponds to the value of $E(A)$.

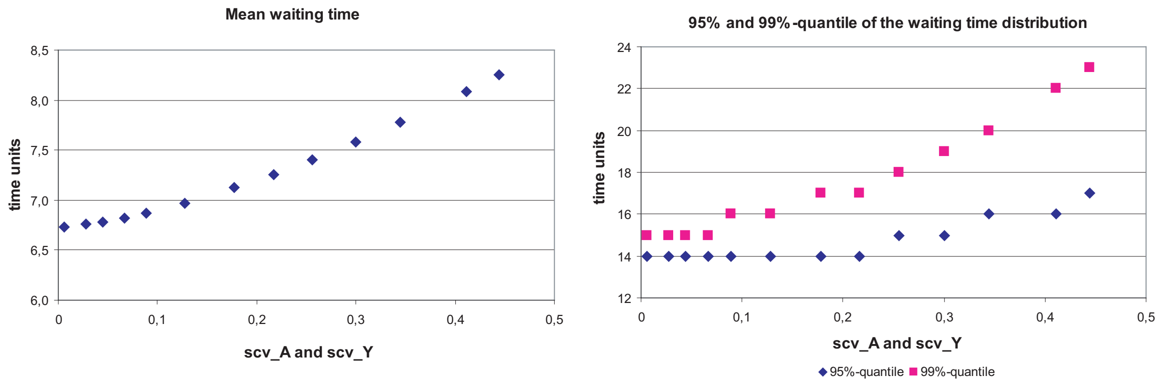


Figure 5.25.: Waiting time depending on the scv of the interarrival time distribution; left: Mean waiting time; right: 95% and 99%-quantile

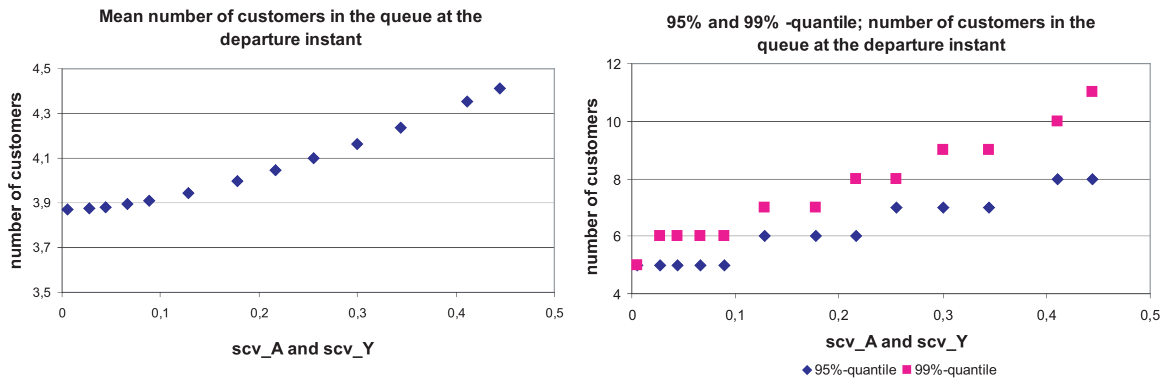


Figure 5.26.: Number of customers in the queue at the departure instant depending on the scv of the interarrival time distribution; left: Mean number of customers; right: 95% and 99%-quantile

### Application: Optimization of the $G/G^{[L,K]}/1$-queue

In Section 4.5.4 we modeled a material handling device which transfers collected material units from a collecting station to its destination. The material transfer is released if at least $L$ customers are collected. Now, we model an arbitrary batch server system which operates under the minimum batch size rule. This can be a service process in a distribution center, a batch production process, or a transportation process etc.

As in Section 4.5.4 we can evaluate the system's costs depending on $L$. Then, the optimal $L$ can be determined.

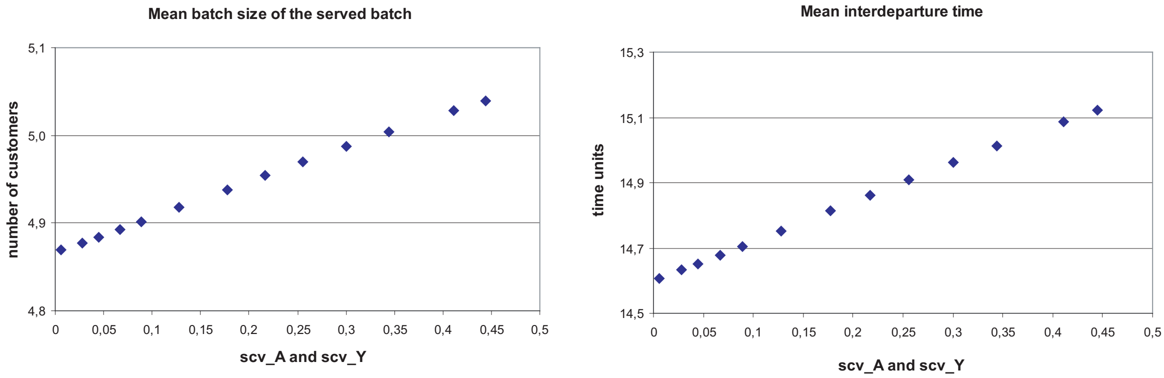


Figure 5.27.: Left: Probability that the server becomes idle at the end of a service operation depending on the scv of the interarrival time distribution; right: Mean interdeparture time depending on the scv of the interarrival time distribution

Two type of costs are observed, the operation costs and the inventory costs. The operation costs arise for each service operation and are represented by $c_O$. We consider all customers in the G/G$^{[L,K]}$/1-system as inventory. For each time period $T$ we observed costs of $c_{In}$ per customer.

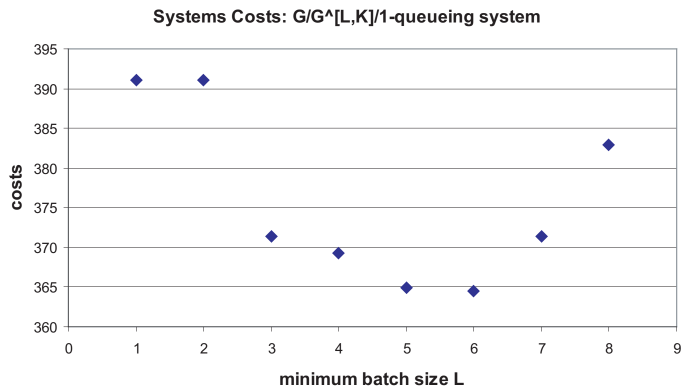


Figure 5.28.: Optimal minimum batch size $L$ regarding given systems costs

The expected number of service operations within $T$ is denoted by $E(N_O)$ and is given by

$$E(N_O) = \frac{T}{E(D)}. \tag{5.101}$$

Using Little's Law we obtain the mean number of customers in system, $E(N_{In})$, as follows:

$$E(N_{In}) = \frac{E(W)}{E(A)} + \rho \cdot E(Y_d), \tag{5.102}$$

where $E(W)/E(A)$ is the mean number of customers in the queue and $\rho \cdot E(Y_d)$ is the mean number of customers in the service station. Finally, we get the system's total costs by

$$C_{tot} = c_O \cdot E(N_O) + c_{In} \cdot E(N_{In}). \tag{5.103}$$

Subsequently, we present a small numerical example to illustrate the optimization problem. The input values for this example are given in Table A.7 in the Appendix. Figure 5.28 shows the total system's costs, in which the optimal minimum batch size is six.

# 5.5. Stochastic Split of Batches

In contrast to the analysis of the batch building operation introduced in Chapter 4, we present in this section the batch split operation which branches a batch arrival stream in several directions. The stochastic split operation with arrival of single customers is already described by Furmans (2004b). The model presented in this section extents the stochastic split for batch arrivals.

The split operation is modeled by a node in a network which has one input stream from node $h$ and several output streams. The input stream is described by the interarrival time distribution $a_i$, $i = 1, \dots, a_{max}$ and the batch size distribution $y_i$, $i = 1, \dots, y_{max}$. We assume that the split operation is stochastic. Modeling batch arrival streams, two different cases can be distinguished. In the first case, the entire batch is assigned to direction $j$ with probability $p_{hj}$, and in the second case, one customer of the arriving batch is assigned to direction $j$ with probability $p_{hj}$. The event that the batch/customer $n$ is assigned to direction $j$ is independent of the assignment of the batch/customer $n-1$. Subsequently, the split of entire batches is presented in Section 5.5.1 and the split of individual customers of a batch in Section 5.5.2.

## 5.5.1. Split of Entire Batches

Considering the split operation of entire batches, the method for analyzing the split operation of a single arrival stream can be taken without modifications briefly discussed in the following.

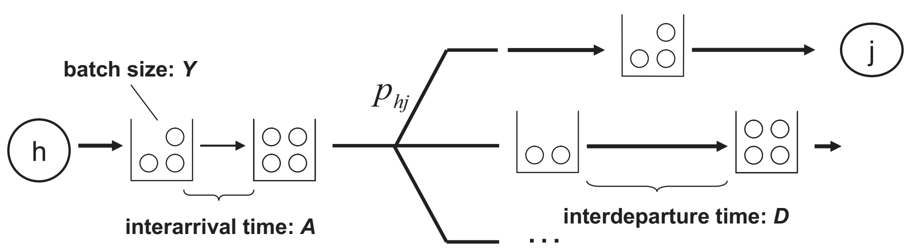


Figure 5.29.: Stochastic split of entire batches arriving at the split node

In order to mark the flow from node $h$ to $j$, we denote the interarrival stream from node $h$ to $j$ by $a_i^{h \to j}$, $i = 1, \dots, a_{max}$. It is assumed that the $n$th batch is directed to stream $j$.

The probability that the succeeding batch is assigned to the same stream $j$ is $p_{hj}$. This event generates an interarrival time of $A$ at node $j$, distributed by $a_i$. Furthermore, the probability that the $(n+1)$th batch is not assigned to node $j$ however the subsequent $(n+2)$th batch, is $(1-p_{hj})\cdot p_{hj}$. Thus, the generated interarrival time at node $j$ is the sum of two interarrival time intervals. The distribution of this interval is given by the 2-fold convolution of $a_i$. Accordingly, the interarrival time distribution of $a_i^{h\to j}$ is calculated by an iterative convolution of $a_i$, weighted with the probability of it's occurrence. Thus, it yields

$$a_i^{h\to j} = \sum_{x=0}^{\infty}(1-p_{hj})^x \cdot p_{hj} \cdot a_i^{(x+1)\otimes}. \tag{5.104}$$

The batch size distribution of stream $j$ is the same as of the given batch size distribution of stream $h$ since the batch size is not affected through the split operation.

## 5.5.2. Split of a Stream of Individual Customers

If one individual customer of the arriving batch is assigned to node $j$ with probability $p_{hj}$, the batch size is changed due to the split operation. The stochastic split of a stream of individual customers is depicted in Figure 5.30.

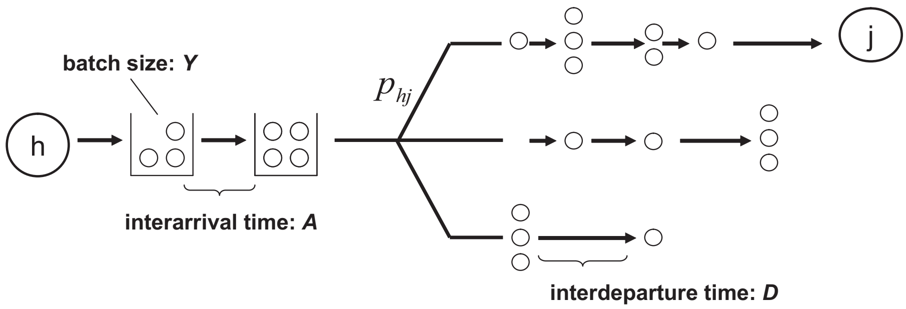


Figure 5.30.: Stochastic split of a stream of individual customers: The individual customers of a batch are assigned to one of the possible directions $j$

Let us analyze the $(n+1)$th batch arrival at the split node. If $x$ customers arrive simultaneously at the split node, $z = 0,\ldots,x$ customers are directed further to node $j$. If at least one customer from the $(n+1)$th batch is directed to node $j$, an interarrival time of $A$, distributed by $a_i$, at node $j$ is generated. The probability that no customer from a set of $x$ customers is chosen is denoted by $q(x)$ and is given by

$$q(x) = (1-p_{hj})^x \quad x = 1,\ldots,y_{max}. \tag{5.105}$$

Since the batch size is a random variable, the probability that no customer is directed to node $j$ is calculated by the summation of $q(x)$ weighted by $y_x$:

$$q = \sum_{x=1}^{y_{max}} y_x \cdot q(x). \tag{5.106}$$

The interarrival time $a_i^{h \to j}$ at node $j$ is a multiple of $a_i$. For example, if no customer of batch $n+1$ is transferred to direction $j$, but at least one customer of batch $n+2$, $a_i^{h \to j}$ results from the 2-fold convolution of $a_i$. For the general case, the interarrival time distribution can be developed by

$$a_i^{h \to j} = \sum_{l=0}^{\infty} q^l \cdot (1-q) \cdot a_i^{(l+1)\otimes}. \tag{5.107}$$

The batch size distribution $y_z^{h \to j}$ $(z = 1,...,y_{max})$ for the arrival stream at node $j$ is derived using the Bernoulli distribution. It has to be considered that the batch size after the split operation can vary between 1 and $x$. The probability that $z$ from $x$ customers are chosen has to be weighted by the event that a batch of size $x$ arrives at the split node. Therefore, the probability that exactly $z$ from $x$ arriving customers are chosen for the transfer to node $j$ is obtained by

$$y_z^{h \to j} = \sum_{z=1}^{x} \frac{y_x \binom{x}{z} p_{hj}^z (1-p_{ij})^{x-z}}{1-(1-p_{hj})^x}. \tag{5.108}$$

Since the summation index in Equation (5.107) goes to infinity, we have to truncate the summation after an appropriate number of steps. Therefore, the stochastic split is exact within an $\epsilon$-environment.

For example the sorting operation in warehouses and distribution centers can be modeled by the stochastic split of batches. Picked articles arrive in batches at the sorting area, where they are assigned to a customer order. The quality control can be modeled by the stochastic split of batches as well. Depending on the quality of the arriving "customer" (item), he (it) can be assigned to a succeeding node. For example, defect items are sorted out with probability $p_{h,j}$ and non defect items are transferred further into the storage area with probability $1 - p_{h,j}$.

Finally, we have to note that the approach of Furmans (2004a) for the merge of single arrival streams can be taken to model the merge of batch arrival streams as well. The interdeparture time distribution at the merge node can be approximated by the method of Furmans, in which he assumes that the merged flow is a renewal process. Under this assumption, the batch size distribution of the merged stream is determined by the combination of the batch size distributions of the arrival streams each weighted by the batch arrival rate. Given two arrival streams, indicated by $h$ and $j$, the batch size distribution of the merged stream, indicated by $merge$, can be calculated as follows:

$$y_i^{merge} = \frac{\lambda_{batch}^h \cdot y_i^h + \lambda_{batch}^j \cdot y_i^j}{\lambda_{batch}^h + \lambda_{batch}^j} \quad i = 1, \ldots \max\{y_{max}^h, y_{max}^j\}. \tag{5.109}$$

## 5.6. Chapter Conclusion

In the current chapter we introduced analytical models for the G/G/1-queue with batch arrivals, the G/G$^{[K,K]}$/1-queue, the G/G$^{[L,K]}$/1-queue and the batch split operation.

- G/G/1-queue with batch arrivals: We presented methods to compute the waiting time distribution and the distribution of the number of customers in the system at the arrival instant. The sojourn time distribution was derived from the waiting time distribution, which gives the probability that an order is fulfilled in a given time period.
- G/G$^{[K,K]}$/1-queue: This queueing system can be decomposed into two subsystems, namely a collecting station running under the capacity rule and a G/G/1-queueing system. Thus, analytical descriptions of these two subsystems can be applied to analyze the G/G$^{[K,K]}$/1-queue.
- G/G$^{[L,K]}$/1-queue: With this queueing system we introduced a server system in which a service operation is initiated if at least $L$ customers are present in the queue. Several distributions to describe performance figures were derived, such as the number of customers in the queue at the departure instant, batch size, interdeparture time, idle time, and waiting time. In contrast to the other methods in this work, we developed an approximate model for the G/G$^{[L,K]}$/1-queue.
- In order to model sorting processes in material flows systems the batch split operation was presented. We modeled both the split of entire batches and the split of individual customers within a batch.

For both the G/G/1-queue with batch arrivals and the G/G$^{[L,K]}$/1-queue we presented a set of numerical examples in which the dependence of the system behavior on the input parameters became apparent.

# 6. Analysis of Batch Processes in Queueing Networks

*Equipped with his five senses, man explores the universe around him and calls the adventure Science.*

Edwin Powell Hubble

Equipped with his five senses, man explores the universe around him and calls the adventure Science.

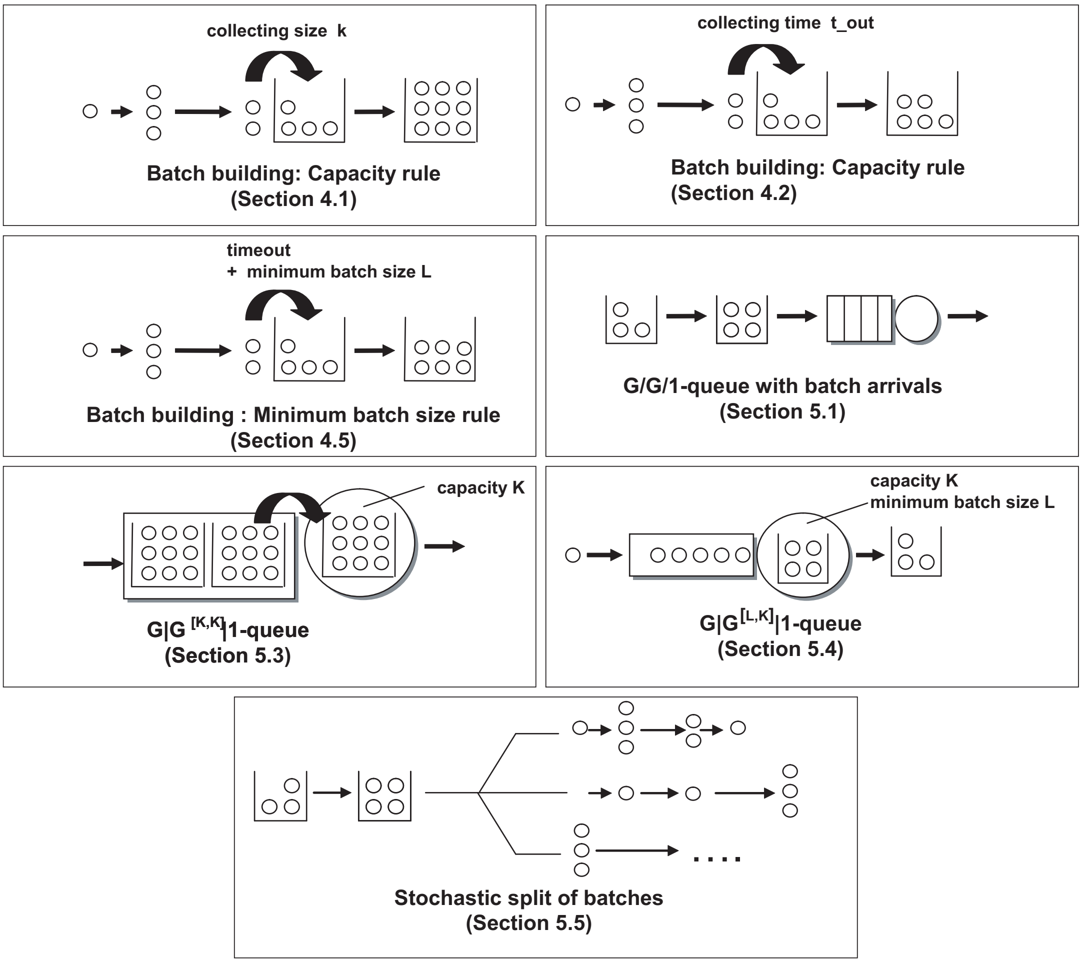


Figure 6.1.: Discrete time models for the analysis of batch processes

In the previous chapters we introduced different models for the analysis of batch processes. Figure 6.1 gives an over all overview of the different models presented. For each analytical problem we determined the interdeparture time distribution which enables us to investigate open queueing networks where batch processes are involved. Thus, a network can be composed from a given library of stochastic model elements. Thereafter, the network can be analyzed under various parameter configurations.

In modeling queueing networks, it is assumed that the network nodes are stochastically independent, which is a common assumption in queueing theory (see Whitt (1983) and Buzacott and Shanthikumar (1993)). It is to note that the departure stream of a queueing system is generally a correlated stochastic stream. Livny et al. (1993) investigates this effect using simulation.

In Section 6.1 we present a software solution which is well suited to model and analyze material flow networks. It is easy to handle and can be applied to practical problems. In Section 6.2 a numerical case analyzing the material flow in a warehouse is presented. The analytical results are compared to results obtained by simulation.

## 6.1. Software Tool

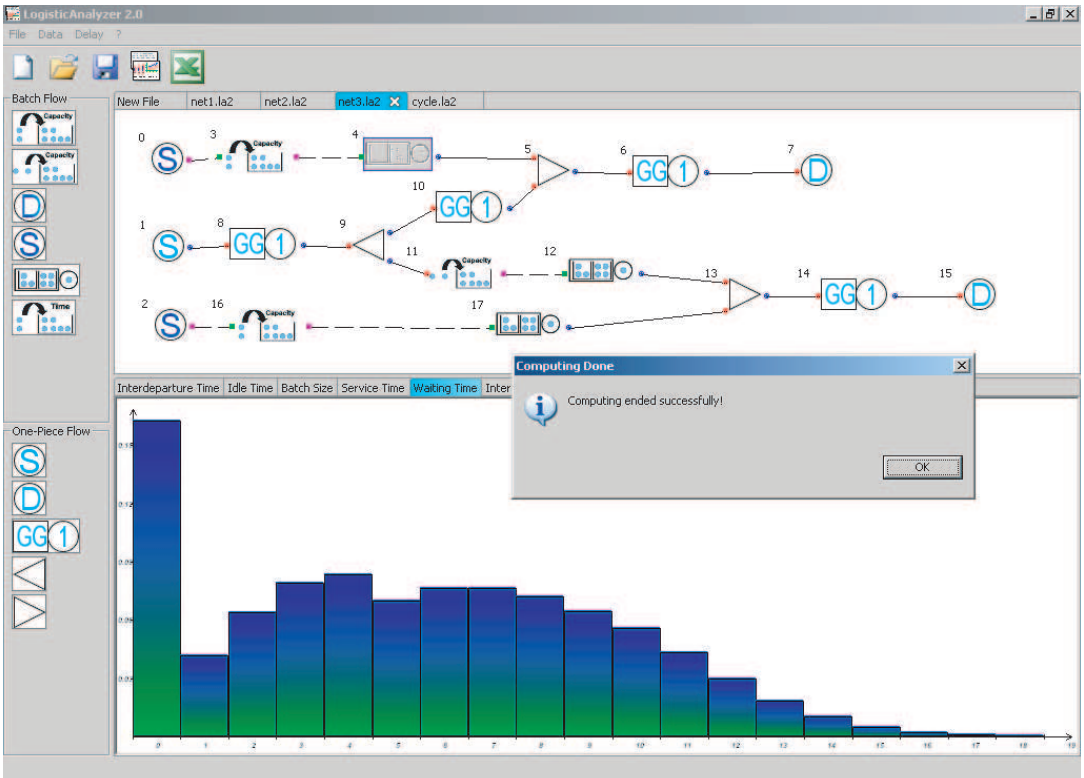


Figure 6.2.: Software solution for a numerical analysis of batch processes: Screenshot of the user interface

Unfortunately, the application of queueing models in industry is not quite widespread even though the basic stochastic phenomenons are easily understandable and traceable.

With the aim to simplify the usage of discrete time stochastic models, we developed a software tool. The user has access to a variety of algorithms via the user interface shown in Figure 6.2. This tool allows easy modeling, parameterizing and analyzing of material flow systems. Both one-piece flows and batch flows can be modeled. All stochastic model elements are clearly arranged in an object library, separated by elements for one-piece flows and batch flows. Arbitrary networks can be modeled via “drag and drop”. Only those network elements can be connected whose departure stream is compatible to the arrival stream of the succeeding element. For example, if the departure process is a batch flow, the arrival process of the succeeding element has to be a batch flow as well.

Network parameters such as arrival and service time distributions, collecting sizes, time-out etc. can be easily entered via provided windows. Distributions can be imported from Microsoft Excel-sheets. Thus, data obtained by an as-is analysis of a material flow system, generally available in the form of histograms, can be used directly. After a network is modeled and parameterized, it can be saved as a XML-file and uploaded for later use. The calculation results are illustrated using diagrams. In addition, all results can be exported as a Microsoft Excel-file for further analysis.

The presented software is developed in the Java programming language. The user interface is developed using SWT [1] due to performance reasons. We paid great attention to an easy extensibility during the development of the software solution. This is realized by a plug-in based software architecture. Thus, new algorithms can be connected to the API without any difficulty and they are immediately applicable after copying the algorithm's source code in the therefore directory.

In the next section we present a numerical case in which we used the introduced Software-Tool.

## 6.2. Numerical Case: Material Flow Network in a Warehouse

Using the earlier presented stochastic model elements, the sojourn time distribution of each element and therefore the sojourn time distribution for a customer in the network can be computed. Thus, we model an order flow in a warehouse by means of the introduced stochastic model elements. Let us recall that the order flow in warehouses and distribution centers should be designed in such a way that it guarantees the order fulfillment in a predetermined time with a chosen probability (e.g. 95%). In the numerical case, we compare the analytical results with results obtained by simulation. Figure 6.3 shows the order flow in our case. It is a rough process description which is often sufficient in an early planning stage. Likewise, the data available at this point is rough as well. For

[1] SWT is an open source widget toolkit for Java designed to provide efficient, portable access to the user-interface facilities of the operating systems on which it is implemented. The SWT implementation accesses the native GUI libraries of the operating system using JNI (Java Native Interface) in a manner that is similar to those programs written using operating system-specific APIs.

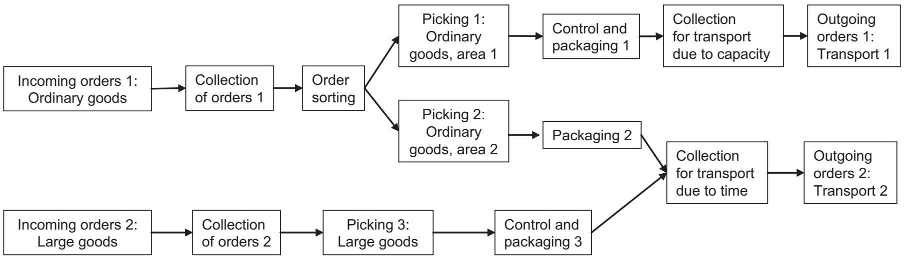


Figure 6.3.: Example of an order flow in a warehouse

these reasons, an analytical tool, like the one we present here, is a practical supplement to simulation. We identify processes like receiving incoming orders, differentiated by ordinary and large goods, collecting of orders, sorting, picking, control and packaging, and outgoing orders. Given the arrival stream of the incoming orders, service times of the underlying processes and batch sizes for the collecting processes, the sojourn time of orders in the system can be determined. Our approach enables the calculation of the probability of the on-time order fulfillment given sojourn time targets.

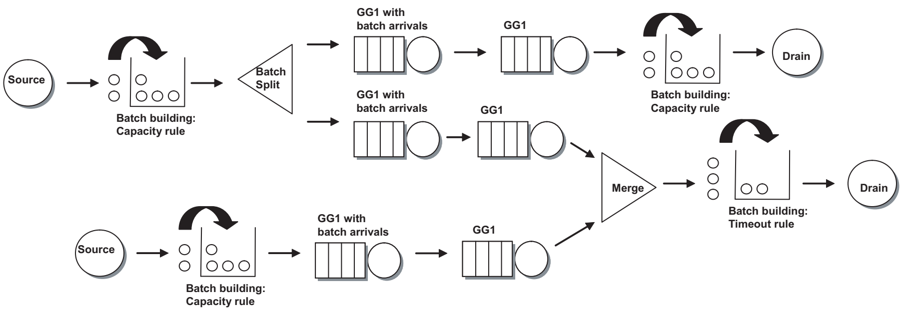


Figure 6.4.: Network of discrete time queueing elements for an order flow in a warehouse shown in Figure 6.3

The order flow of Figure 6.3 can be transferred to a queueing network using model elements presented previously. As such, we use the elements “batch building: capacity rule”, “stochastic split of batches”, “G/G/1-queue with batch arrivals”, “G/G/1-queue”, “stochastic merge” and “batch building: timeout rule”. The resulting network is illustrated in Figure 6.4. We identify three different order flows through the network, the first from “Incoming orders 1” to “Outgoing orders 1”,the second from “Incoming orders 1” to “Outgoing orders 2”, and the third from “Incoming orders 2” to “Outgoing orders 2”. We number them “order flow 1”, “order flow 2” and “order flow 3” accordingly. The input parameters for the network nodes are given in Table A.6 in the Appendix. The throughput of the three flows can be calculated by Equation (4.4) and it yields $\lambda_1 = 0.427$, $\lambda_2 = 0.284$ and $\lambda_3 = 0.405$ orders per time unit.

Since the departure stream of a queueing system is generally a correlated stochastic stream, we expect that there is a deviation between the analytical results and the results obtained by simulation.

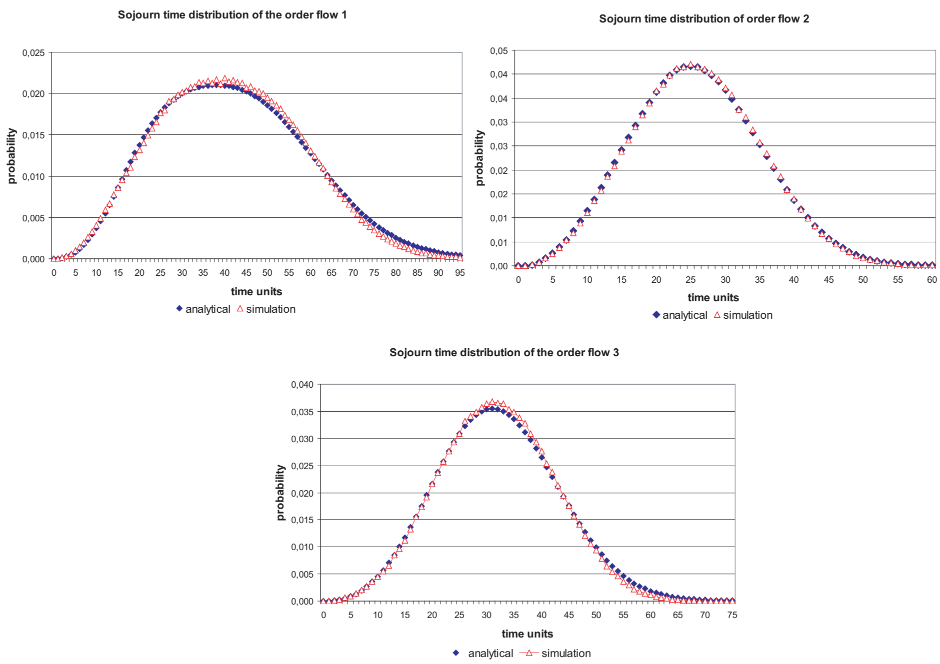


Figure 6.5.: The sojourn time distribution of the three order flows; analytical results versus simulation

In order to test the accuracy of our analytical case we use discrete event simulation. We computed the sojourn time distributions for all three flows. The results are illustrated in Figure 6.5. Due to correlation effects there is a deviation between the analytical and simulation results. This deviation increases if the network becomes larger and more complex. However, we find the results very promising and well suited for the analysis of material and information flows. In Table 6.1 the mean sojourn time of orders in the network, $E(v)$, the scv of the sojourn time distribution, $c_v^2$, and the 90%, 95% and 99%-quantile of the sojourn time distribution are given. Furthermore, its deviation from simulation results are illustrated.

| | order flow 1 | | | order flow 2 | | | order flow 3 | | |
|---|---|---|---|---|---|---|---|---|---|
| | analy | sim | %-error | analy | sim | %-error | analy | sim | %-error |
| mean | 42.05 | 41.46 | 1.42 | 25.86 | 25.95 | 0.32 | 32.20 | 31.89 | 0.98 |
| scv | 0.17 | 0.16 | 6.47 | 0.13 | 0.13 | 1.94 | 0.12 | 0.11 | 7.07 |
| 90-quantile | 66 | 64 | 3.03 | 24 | 24 | 0.00 | 35 | 34 | 2.13 |
| 95-quantile | 72 | 70 | 2.78 | 27 | 27 | 0.00 | 39 | 37 | 1.96 |
| 99-quantile | 86 | 81 | 5.81 | 33 | 33 | 0.00 | 47 | 44 | 5.08 |

Table 6.1.: Numerical case results; percentage error in the sojourn time; analytical approach versus simulation

Since the computing times for the analytical methods are low [2], it enables a "what-if" analysis for material flow systems which is necessary in the decision process at an early

[2]Varying milliseconds to seconds depending on the vector size of the input parameters.

planning stage. This is an enormous advantage compared to simulation which is a very time consuming approach.

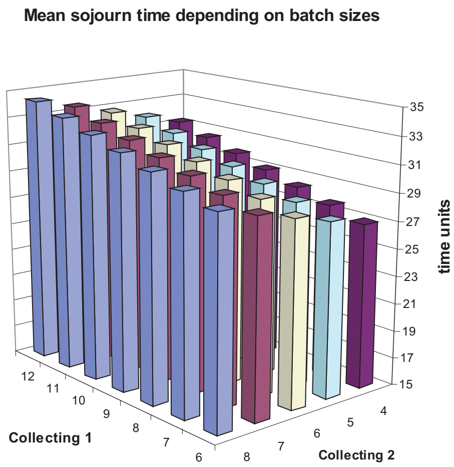


Figure 6.6.: The effect of reducing batch sizes on the mean sojourn time of an arbitrary order

Let us illustrate the possibility of a "what-if" analysis by a small example. We change the batch size resulting from "Collecting of orders 1" and "Collecting of orders 2" and calculate the resulting sojourn time distribution. We compared the performance of the different batch size configurations by means of the mean sojourn time of an arbitrary order. Given the batch size configuration of Table A.6, we reduced stepwise the batch size by one until the initial batch sizes were halved. The results are shown in Figure 6.6. It is obvious that the sojourn time decreases with decreasing batch sizes and by Little's law equally the work in progress. In contrast, small batch sizes lead to less efficiency, that means less use of the available capacity. This causes an increase in transports, setups and service operations. Subsequently, with given system's costs, an optimization model can be set up.

A frame of such an optimization model looks as follows:

min Total cost (storage, capacity, operation etc.)

subject to:
X%-quantile of the sojourn time $\leq$ Max sojourn time for all order types

Throughput = Required Throughput for all order types

However, the introduction of detailed optimization models for material flow systems where batch processes are involved have to be left open for future research.

For a more detailed description of modeling material handling systems using stochastic model elements in discrete time see Schleyer and Furmans (2006b).

## 6.3. Chapter Conclusion

In this chapter we exemplified how the stochastic models we have introduced in the current work can be used to analyze material flow networks. In this context we presented a software tool. We made efforts to develop an user-friendly solution which allows easy modeling, parameterizing and analyzing of material flow systems.

Using a numerical case we showed how an order flow in a warehouse is modeled and analyzed. We determined the sojourn time distribution for three given order flows. In addition, we tested the accuracy of the introduced analytical approach against simulation. The low deviation between analytical and simulation are very promising. Furthermore, we explained the possibility of "what-if" analysis using a small example. At the end of this chapter, a frame of an optimization model was sketched out.

# 7. Conclusion and Outlook

*He who enjoys doing and enjoys what he has done is happy.*

Johann Wolfgang von Goethe

Planning the material flow is a crucial task for many companies. Decisions about the appropriate capacities, buffer places, material flow layout, transport carrier etc. influence strongly the success of companies. Since material flow systems are exposed to stochastic events as demand changes, machine failures, scrap, varying processing times etc., suitable models are required for a system's analysis in order to make best decisions. In addition, the understanding of the system's behavior and of the dependencies between input parameters and output allows an appropriate reaction on system's changes. The purpose of the current work was to give such insights into the behavior of various batch processes occurring in material flow systems. Thus, a comprehensive toolkit of discrete time queueing models for different batch flow problems was provided. Now, it remains for the user to apply the appropriate tool for a given problem.

Furthermore, we focused on the determination of the probability of an on-time order fulfillment given time targets, since this is of vital importance for the design of material flow systems. We calculated the waiting and sojourn time distribution for each presented stochastic model using discrete time analysis. Quantiles such as $\sigma_{v,0.95}$ and $\sigma_{v,0.99}$, and thereby the probability of an on-time order fulfillment, can be estimated. Various numerical examples showed that $\sigma_{v,0.95}$ and $\sigma_{v,0.99}$ react more sensitively on parameter changes than $E(V)$. We concluded that an evaluation on the basis of distributions has essential advantages compared to the classical mean value analysis using 2-parameter approximations of general processes.

The presentation of analytical models was structured in two parts. First, the analysis of batch building modes and second, the analysis of server systems.

We provided a detailed time analysis of three different batch building modes: The capacity rule, timeout rule, and minimum batch size rule. All proposed approaches are exact. Recapitulating, we summarize the following insights gained through the analysis:

- There are equivalences between the capacity and timeout rule. Knowing this equivalences both batch building modes can be investigated by means of just one approach.

- We showed that the residual interarrival time distribution (timeout rule) respectively the remaining distribution (capacity rule) corresponds to the residual lifetime distribution of a renewal process.

- Applying the capacity rule, the influence of $c_Y^2$ on the stability of the departure process is greater than the influence of $c_A^2$. If the timeout rule is applied this conclusion is inverse.
- Applying the capacity rule, $E(W)$ is independent of $c_A^2$ and $c_Y^2$. However, quantiles such as $\sigma_{w,0.95}$ and $\sigma_{w,0.99}$ increase with increasing $c_A^2$ and $c_Y^2$. From this it follows that low values of $c_A^2$ and $c_Y^2$ improve the probability of an on-time order fulfillment.
- Applying the timeout rule, $E(W)$ is independent of the arrival process. $E(W)$ depends just on the timeout $t_{out}$.
- Applying the minimum batch size rule, we explained the paradox of decreasing $E(W)$ with increasing $c_A^2$ and $c_Y^2$ within a certain range of $c_A^2$ and $c_Y^2$. We exemplified the condition for the occurrence of this paradox.
- Applying the minimum batch size rule, we showed the influence of the minimum batch size $L$ on output figures. Furthermore, a collecting process running according the minimum batch size rule can be optimized by the appropriate choice of $L$.

In addition to the analysis of batch building processes, we investigated queueing systems such as the G/G/1-queue with batch arrivals, the G/G$^{[K,K]}$/1-queue, and the G/G$^{[L,K]}$/1-queue. Recapitulating, we summarize the following insights gained through the analysis:

- In addition to the waiting and departure process, the system's state of the G/G/1-queue with batch arrivals at the arrival instant was studied. The distribution of the number of customers at the arrival instant can be used for a buffer dimensioning in material flows.
- The analysis of G/G/1-queue with batch arrivals illustrated that $\sigma_{w,0.95}$, $\sigma_{v,0.95}$ and $\sigma_{\eta,0.95}$ react more sensitively on parameter changes than $E(W)$, $E(V)$ and $E(N(\tau))$.
- The G/G$^{[K,K]}$/1-queue can be decomposed into two subsystems, namely a collecting station running under the capacity rule and a G/G/1-queueing system. The G/G$^{[K,K]}$/1-queue can be evaluated by the analysis of these two subsystems.
- Approximating the residual interarrival time by the residual lifetime of a renewal process results in a well-fitting system description of the G/G$^{[L,K]}$/1-queue.
- In addition to the waiting and departure process, the system's state of the G/G$^{[L,K]}$/1-queue at the departure instant was studied.
- Analyzing the G/G$^{[L,K]}$/1-queue, we showed the influence of the minimum batch size $L$ on output figures. The G/G$^{[L,K]}$/1-queue can be optimized by the appropriate choice of $L$.

We showed how the presented discrete time models can be applied for a network analysis. A software solution for modeling and analyzing material flow networks was presented.

Finally, let us name some limitations of our approach and give an outlook on further research.

We explained that our models are well-suited to support the long range planning in an early planning stage, in which capacities are searched to minimize the facility costs given cycle time targets. Since queueing models are generally rough descriptions of the underlying reality, the application on a detailed planning level is limited. Then simulation or other methods from the operations research are more appropriate.

By means of a numerical case we showed how an order flow in a warehouse can be modeled and analyzed. Due to correlation effects there is a deviation between the analytical and simulation results. The deviation in this numerical example was small, however it increases if the network becomes larger and more complex. Therefore the application of discrete time queueing models to very large and complex networks is limited.

Using the batch building mode under the minimum batch size rule or the $G/G^{[L,K]}/1$-queue, the calculation of $w_i$ performs rapidly for problems of reasonable size. If the vectors describing the arrival and the service process become large, the computing time increases significantly. It seems to be within reach to overcome this limitation by a more rough analysis. Thus, approximations for $w_i$ are searched, which perform faster than the existing approaches and are nevertheless adequately accurate.

The current work focuses on the analysis of material flow systems. We mentioned that the approach can be used to support best decisions regarding the design of material flow systems. Future work could address setting up and solving optimizations problems using discrete time models.

The purpose of the work was to find new analytical approaches in the discrete time domain not known so far for the description of material flow processes. It is important to mention that there are still many problems left which can be addressed by further research; for example, the analysis of further batch building modes. As such, we studied not the timeout rule if a maximum capacity is given, and not the capacity rule if the collecting time is bounded by a maximum time. The presented approach of the minimum batch size rule can be taken as foundation for the analysis of further batch building modes. Further research can also be done in the field of discrete time priority queueing, discrete time inventory models (see Zillus (2003)), study of sorting strategies in material flows etc.. Furthermore, there is a need to develop application models using basic discrete time queueing methods. As such, the proposed batch queueing models can be used for an extension of the lot size decision model of Greiling (1997) considering cycle time targets. Furthermore, discrete time methods can be involved in the planning of hub and spoke networks (see Blunck (2003)). Finally, the proposed batch building models can be used after some modifications to analyze demand fluctuations in supply chains caused by lot size decisions.

# Bibliography


Ackroyd, H. M., 1980. Computing the waiting time distribution for the G/G/1 queue by signal processing methods. IEEE Transactions on Communications 38 (1), 52–58.

Alfa, A., 2005. Analysis of a GI/G$^Y$/1 system in discrete time. Stochastic Models 21, 185–213.

Bailey, N. T. J., 1954. On queueing processes with bulk service. Journal of the Royal Statistical Society 16, 80–87.

Bitran, G. R., Tirupati, D., 1989. Approximation for product departures from a single-server station with batch processing in multi-product queues. Management Science 35 (7), 851–878.

Blunck, S., 2003. Modellierung und Optimierung von Hub-and-Spoke-Netzen mit beschränkter Sortierkapazität. Dissertation, Fakultät für Maschinenbau der Universität Karlsruhe (TH).

Bolch, G., Greiner, S., de Meer, H., Trivedi, K., 1998. *Queueing Networks and Markov Chains: Modelling and Performance Performance Evaluation with Computer Science Applications*. John Wiley, New York.

Brière, G., Chaudhry, 1987. Computational analysis of single-server bulk-arrvial queues: $GI^X/M/1$. Queueing Systems 2, 173–185.

Bruneel, H., Kim, B. G., 1992. *Discrete-Time Models for Communication Systems Including ATM*. Kluwer Academic Publishers, Norwell, MA, USA.

Burke, P. J., 1975. Delays in single server queues with batch input. Operations Research 23, 830–833.

Buzacott, J. A., Shanthikumar, G., 1993. *Stochastic Models of Manufacturing Systems*. Prentice Hall, Englewood Cliffs, NJ.

Calabress, J. M., Hausmann, W. H., 1991. Simultaneous determination of lot sizes and routing mix in job shops. Management Science 37 (8), 1043–1059.

Chandy, K. M., Sauer, C. H., 1978. Approximate methods for analyzing queueing network models of computer systems. ACM Computing Surveys 10 (3), 281–317.

Chaudhry, M. L., 1979. The queueing system M$^X$/G/1 and its ramifications. Naval Res. Logist. Quart. 26, 667–674.

Chaudhry, M. L., Gupta, U., 1997. Queue-length and waiting-time distributions of discrete-time $\mathrm{GI}^X$/GEOM/1 queueing systems with early and late arrivals. Queueing Systems 25, 307–324.

Chaudhry, M. L., Gupta, U., Goswami, V., 2001. Modeling and analysis of discrete-time multiserver queues with batch arrivals: $\mathrm{GI}^X$/GEOM/M. Journal on Computing 13 (3), 172–180.

Chaudhry, M. L., Madhill, B. R., Brière, G., 1987. Computational analysis of steady-state probabilities of $\mathrm{M/G}^{[a,b]/1}$ and related nonbulk queues. Queueing Systems 2, 93–114.

Chaudhry, M. L., Templeton, J. G. C., 1983. *A First Course in Bulk Queues*. John Wiley, New York.

Curry, G. L., Deuermeyer, B. L., 2002. Renewal approximation for the departure process of batch systems. IIE Transactions 34, 95–104.

Downtown, F., 1955. Waiting time in bulk service queues. Journal of the Royal Statistical Society 17 (2), 256–261.

Dümmler, M., 1998. Analysis of the departure process of a batch service queueing system. Technical Report, Lehrstuhl für Informatik III, Universität Würzburg.

Feller, W., 1968. *An Introduction to Probability Theory and its Applications*. Vol. 1. Wiley, New York.

Foster, F. G., Nyunt, K. M., 1961. Queues with batch departures. Annals of Mathematical Statistics 32, 1324–1332.

Fowler, J., Phojanamongkolkij, N., Cochran, J., Montgomery, D., 2002. Optimal batching in a wafer fabrication facility using a multiproduct G/G/c model with batch processing. International Journal of Production Research 40 (2), 275–292.

Furmans, K., 2000. Bedientheoretische Methoden als Hilfsmittel der Materialflußplanung. Habilitatsionsschrift, Institut für Fördertechnik und Logistiksysteme, Universität Karlsruhe.

Furmans, K., 2004a. A framework of stochastic finite elements for models of material handling systems. In: Progress in Material Handling Research: 2004. 8. International Material Handling Research Colloquium, Graz, 2004.

Furmans, K., 2004b. Zeitdiskrete Modellierung von Logistikprozessen. Technical Report, Institut für Fördertechnik und Logistiksysteme, Universität Karlsruhe.

Furmans, K., Zillus, A., 1996. Modeling independent production buffers in discrete time queueing networks. In: Proceedings of Rensselaer's Fifth International Conference Computer Integrated Manufacturing and Automation Technology. pp. 275–280.

Gaver, D. P., 1959. Imbedded Markov chain analysis of a waiting-line process in continous time. Annals of Mathematical Statistics 30, 698–720.

Gnedenko, B. W., König, D., 1983. *Handbuch der Bedientheorie I.* Vol. 1. Akademie-Verlag, Berlin.

Gold, H., 1992. Performance Modelling of Batch Service Systems with Push and Pull Manufacturing Management Policies. PhD-Thesis, Universität Würzburg.

Grassmann, W. K., Jain, J. L., 1988. Numerical solution for the departure process from the GI/G/1 queue. Comput. Opns. Res. 15 (3), 293–296.

Grassmann, W. K., Jain, J. L., 1989. Numerical solutions of the waiting time distribution and idle time distribution of the arithmetic GI/G/1 queue. Operations Research 37 (1), 141–150.

Greiling, M., 1997. Verbesserung der Produktionslogistik durch Losgrößenharmonisierung - Ein bedientheoretischer Ansatz. Dissertation, Fakultät für Maschinenbau der Universität Karlsruhe (TH).

Gross, D., Harris, C., 1985. *Fundamentals of Queueing Theory*, 2nd Edition. John Wiley, New York.

Hopp, W. J., Spearman, M., Chayet, S., Donohue, K. L., Gel, E., 2002. Using an optimized queueing network model to support wafer fab design. IIE Transactions 34 (119-130).

Hübner, F., Tran-Gia, P., 1995. Discrete-time analysis of cell spacing in ATM systems. Telecommunications Systems 3, 379–395.

Jaiswal, N. K., 1960. Bulk service queueing problem. Operations Research 8, 139–143.

Karmarkar, U. S., Kerke, S., Kerke, S., 1985. Lotsizing in multi-item multi machine job shops. IIE Transactions 17 (3), 290–297.

Kendall, D. G., 1951. Some problems in the theory of queues. Journal of the Royal Statistical Society B, 13, 151–185.

Kendall, D. G., 1953. Stochastic processes occurring in the theory of queues and their analysis by the method of the imbedded Markov chain. Annals of Mathematical Statistics 24, 338–354.

Kleinrock, L., 1975. *Queueing Systems*. Vol. 1: Theory. John Wiley, New York; Chichester; Brisbane; Toronto.

Konheim, A. G., 1975. An elemntary solution of the queueing system G/G/1. SIAM J. Comput. (4), 540–545.

Krämer, W., Langenbach-Belz, M., 1976. Approximate formulae for the delay in the queueing system GI/G/1. In: Congressbook. 8th International Teletraffic Congress, Melbourne, pp. 235.1–235.8.

Kuehn, P. J., 1979. Approximate analysis of general networks by decomposition. IEEE Trans. Communication COM-27 (1), 113–126.

Lippolt, C., 2003. Spielzeiten in Hochregallagern mit doppeltiefer Lagerung. Dissertation, Fakultät für Maschinenbau der Universität Karlsruhe (TH).

Little, J. D., 1961. A proof for the queueing formula: $L = \lambda \cdot W$. Operations Research 9 (3), 383–387.

Livny, M., Melamed, B., Tsiolis, A. K., 1993. The impact of autocorrelation on queuing systems. Management Science 39, 322–339.

Marchal, W., 1976. An approximate formula for waiting time in single server queues. AIIE Transactions 8, 473 ff.

Meng, G., Heragu, S., 2004. Batch size modeling in a multi-item, discrete manufacturing system via an open queueing network. IIE Transactions 36, 743–753.

Neuts, M. F., 1965. The busy period of a queue with batch service. Operations Research 13, 815–819.

Neuts, M. F., 1967. A general class of bulk queues with poisson input. Annals of Mathematical Statistics 38 (2), 759–770.

Neuts, M. F., 1981. *Matrix-geometric Solutions in Stochastic Models.* The John Hopkins University Press, Baltimore.

Ommeren, J. V., 1990. Simple approximations for the batch-arrival $M^X/G/1$ queue. Operations Research 38 (4), 678–685.

Page, E. G., 1972. *Queueing Theory in Operations Research.* Crane Russak & Co.

Powell, W. B., 1985. Analysis of vehicle holding and cancellation strategies in bulk arrival, bulk service queues. Transportation Science 19 (4), 352–377.

Powell, W. B., 1986a. Appproximate, closed form moment formulas for bulk arrival, bulk service queues. Transportation Science 20 (1), 13–23.

Powell, W. B., Humblet, P., 1986b. The bulk service queue with a general control strategy: Theoretical analysis and a new computational procedure. Operations Research 34 (2), 267–274.

Rall, B., 1998. Analyse und Dimensioniereung von Materialflußsystemen mittels geschlossener Warteschlangennetze. Dissertation, Fakultät für Maschinenbau der Universität Karlsruhe (TH).

Reiser, M., Kobayashi, H., 1974. Accuracy of diffusion approximation for some queueing systems. IBM J. Research and Development 18, 110–124.

Schleyer, M., Furmans, K., 2005. Analysis of batch building processes. In: Proceedings of the 20th IAR Annual Meeting, Mulhouse. pp. 145–154.

Schleyer, M., Furmans, K., 2006a. An analytical method for the calculation of the waiting time distribution of a discrete time G/G/1-queueing system with batch arrivals. OR Spectrum(online first).

Schleyer, M., Furmans, K., 2006b. A framework of stochastic finite elements for models of material handling systems, submitted for publication to IIE Transactions (under review).

Shanthikumar, J. G., Buzacott, J. A., 1980. On the approximations to the single server queue. International Journal of Production Research 18 (6), 761–773.

Shanthikumar, J. G., Buzacott, J. A., 1981. Opem queueing network models of dynamic job shops. International Journal of Production Research 19 (3), 255–266.

Tran-Gia, P., 1996. *Analytische Leistungsbewertung verteilter Systeme*. Springer.

Whitt, W., 1983. Queueing network analyser. The Bell System Technical Journal 62 (9), 2779–2815.

Whitt, W., 1993. Approximations for the GI/G/m queue. Production and Operations Management 2 (2), 141–161.

Wolff, R. W., 1989. *Stochastic Modeling and the Theory of Queues*. Prentice Hall International Editions, Englewood Cliffs, NJ.

Zillus, A., 2003. Untersuchung der Wartezeit von Kundenaufträgen in der Supply Chain. Dissertation, Fakultät für Maschinenbau der Universität Karlsruhe (TH).

# A. Appendix: Input Data

## A.1. Input Data

| i | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **0** | 0.000 | 0.0 | 0.0 | 0.00 | 0.0 | 0.0 | 0.000 | 0.0 | 0.000 | 0.00 | 0.0 | 0.0 | 0.0 |
| **1** | 0.000 | 0.0 | 0.0 | 0.00 | 0.0 | 0.1 | 0.125 | 0.2 | 0.200 | 0.30 | 0.3 | 0.4 | 0.5 |
| **2** | 0.000 | 0.0 | 0.1 | 0.15 | 0.2 | 0.1 | 0.150 | 0.1 | 0.175 | 0.10 | 0.2 | 0.1 | 0.0 |
| **3** | 0.025 | 0.3 | 0.2 | 0.20 | 0.2 | 0.2 | 0.150 | 0.1 | 0.100 | 0.05 | 0.0 | 0.0 | 0.0 |
| **4** | 0.950 | 0.4 | 0.4 | 0.30 | 0.2 | 0.2 | 0.150 | 0.2 | 0.050 | 0.10 | 0.0 | 0.0 | 0.0 |
| **5** | 0.025 | 0.3 | 0.2 | 0.20 | 0.2 | 0.2 | 0.150 | 0.1 | 0.100 | 0.05 | 0.0 | 0.0 | 0.0 |
| **6** | | | 0.1 | 0.15 | 0.2 | 0.1 | 0.150 | 0.1 | 0.175 | 0.10 | 0.2 | 0.1 | 0.0 |
| **7** | | | | | | 0.1 | 0.125 | 0.2 | 0.200 | 0.30 | 0.3 | 0.4 | 0.5 |
| **mean** | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 |
| **scv** | 0.003 | 0.038 | 0.075 | 0.100 | 0.125 | 0.188 | 0.234 | 0.288 | 0.325 | 0.394 | 0.438 | 0.500 | 0.563 |

Table A.1.: Time discrete symmetric distributions for $A$ and $Y$ with different scv's. These distributions are used to demonstrate the influence of $c_A^2$ and $c_Y^2$ on the 95% and 99%-quantile of the waiting time distribution. The batch size $k$ was 15 in this set of experiments; the results are shown in Figure 4.13

| | example 1 | example 2 | example 3 | example 4 | example 5 | example 6 | example 7 | example 8 |
|---|---|---|---|---|---|---|---|---|
| **i** | $y_i$ | $y_i$ | $y_i$ | $y_i$ | $y_i$ | $y_i$ | $y_i$ | $y_i$ |
| **0** | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| **1** | 0.300 | 0.400 | 0.000 | 0.200 | 0.100 | 0.200 | 0.100 | 0.100 |
| **2** | 0.400 | 0.300 | 0.300 | 0.200 | 0.150 | 0.000 | 0.250 | 0.100 |
| **3** | 0.300 | 0.200 | 0.400 | 0.200 | 0.175 | 0.300 | 0.150 | 0.100 |
| **4** | | 0.100 | 0.300 | 0.200 | 0.250 | 0.150 | 0.100 | 0.200 |
| **5** | | | | 0.200 | 0.150 | 0.100 | 0.000 | 0.100 |
| **6** | | | | | 0.100 | 0.050 | 0.000 | 0.150 |
| **7** | | | | | 0.073 | 0.000 | 0.050 | 0.100 |
| **8** | | | | | 0.003 | 0.100 | 0.200 | 0.100 |
| **9** | | | | | | 0.050 | 0.100 | 0.050 |
| **10** | | | | | | 0.050 | 0.050 | |
| **mean** | 2.000 | 2.000 | 3.000 | 3.000 | 4.250 | 3.803 | 4.800 | 4.750 |
| **scv** | 0.150 | 0.250 | 0.067 | 0.222 | 0.387 | 0.199 | 0.406 | 0.234 |

Table A.2.: Examples for the calculation of the variability of the departure process (described by $c_D^2$ or $c_{Y_d}^2$) depending on $k$; the results are shown in Figure 4.14

| i | $a_i$ | $y_i$ | $t_{out}$ |
|---|---|---|---|
| **0** | 0.0 | 0.0 | 10 |
| **1** | 0.2 | 0.6 | |
| **2** | 0.4 | 0.3 | |
| **3** | 0.2 | 0.1 | |
| **4** | 0.2 | | |
| **mean** | 2.400 | 1.500 | |
| **scv** | 0.181 | 0.200 | |

Table A.3.: Input data in order to study the influence of the minimum batch size on the interdeparture time $D$, the collected batch size $Y_d$, and the waiting time $W$ applying the batch building mode under the minimum batch size rule; the results are illustrated in Figures from 4.18 to 4.21

| i | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **0** | 0.000 | 0.000 | 0.0 | 0.0 | 0.0 | 0.000 | 0.00 | 0.000 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| **1** | 0.000 | 0.000 | 0.0 | 0.0 | 0.0 | 0.100 | 0.15 | 0.200 | 0.25 | 0.30 | 0.35 | 0.45 | 0.50 |
| **2** | 0.025 | 0.125 | 0.2 | 0.3 | 0.4 | 0.175 | 0.20 | 0.175 | 0.15 | 0.15 | 0.15 | 0.05 | 0.00 |
| **3** | 0.950 | 0.750 | 0.6 | 0.4 | 0.2 | 0.450 | 0.30 | 0.250 | 0.20 | 0.10 | 0.00 | 0.00 | 0.00 |
| **4** | 0.025 | 0.125 | 0.2 | 0.3 | 0.4 | 0.175 | 0.20 | 0.175 | 0.15 | 0.15 | 0.15 | 0.05 | 0.00 |
| **5** | | | | | | 0.100 | 0.15 | 0.200 | 0.25 | 0.30 | 0.35 | 0.45 | 0.50 |
| **mean** | 3.000 | 3.000 | 3.000 | 3.000 | 3.000 | 3.000 | 3.000 | 3.000 | 3.000 | 3.000 | 3.000 | 3.000 | 3.000 |
| **scv** | 0.006 | 0.028 | 0.044 | 0.067 | 0.089 | 0.128 | 0.178 | 0.217 | 0.256 | 0.300 | 0.344 | 0.411 | 0.444 |

Table A.4.: Distributions which have the same mean value however a varying scv. This data is used to to study $E(W)$, $\sigma_{w,0.95}$ and $\sigma_{w,0.99}$ depending on both $c_A^2$ and $c_Y^2$ applying the batch building mode under the minimum batch size rule; the results are illustrated in Figures from 4.21 to 4.23.

| | Example 1 | | Example 2 | | Example 3 | | Example 4 | |
|---|---|---|---|---|---|---|---|---|
| | $L$ | 4 | $L$ | 4 | $L$ | 5 | $L$ | 10 |
| | $K$ | 6 | $K$ | 6 | $K$ | 8 | $K$ | 15 |
| | $a_i$ | $b_i$ | $a_i$ | $b_i$ | $a_i$ | $b_i$ | $a_i$ | $b_i$ |
| **0** | 0.0 | 0.00 | 0.0 | 0.00 | 0.0 | 0.00 | 0.000 | 0.000 |
| **1** | 0.0 | 0.00 | 0.0 | 0.00 | 0.0 | 0.00 | 0.250 | 0.000 |
| **2** | 0.3 | 0.00 | 0.2 | 0.00 | 0.3 | 0.00 | 0.400 | 0.000 |
| **3** | 0.4 | 0.00 | 0.2 | 0.00 | 0.3 | 0.00 | 0.175 | 0.000 |
| **4** | 0.3 | 0.00 | 0.2 | 0.00 | 0.1 | 0.00 | 0.075 | 0.000 |
| **5** | | 0.00 | 0.2 | 0.00 | 0.2 | 0.00 | 0.050 | 0.000 |
| **6** | | 0.00 | 0.2 | 0.00 | 0.1 | 0.00 | 0.050 | 0.000 |
| **7** | | 0.00 | | 0.00 | | 0.00 | | 0.000 |
| **8** | | 0.00 | | 0.00 | | 0.00 | | 0.000 |
| **9** | | 0.00 | | 0.00 | | 0.00 | | 0.000 |
| **10** | | 0.00 | | 0.00 | | 0.00 | | 0.000 |
| **11** | | 0.10 | | 0.00 | | 0.00 | | 0.000 |
| **12** | | 0.15 | | 0.00 | | 0.05 | | 0.000 |
| **13** | | 0.15 | | 0.00 | | 0.05 | | 0.000 |
| **14** | | 0.20 | | 0.00 | | 0.05 | | 0.025 |
| **15** | | 0.15 | | 0.10 | | 0.10 | | 0.050 |
| **16** | | 0.15 | | 0.15 | | 0.05 | | 0.025 |
| **17** | | 0.10 | | 0.20 | | 0.15 | | 0.000 |
| **18** | | | | 0.30 | | 0.20 | | 0.000 |
| **19** | | | | 0.15 | | 0.05 | | 0.025 |
| **20** | | | | 0.05 | | 0.15 | | 0.025 |
| **21** | | | | 0.05 | | 0.05 | | 0.050 |
| **22** | | | | | | 0.05 | | 0.075 |
| **23** | | | | | | 0.05 | | 0.100 |
| **24** | | | | | | | | 0.025 |
| **25** | | | | | | | | 0.000 |
| **26** | | | | | | | | 0.000 |
| **27** | | | | | | | | 0.025 |
| **28** | | | | | | | | 0.025 |
| **29** | | | | | | | | 0.050 |
| **30** | | | | | | | | 0.100 |
| **31** | | | | | | | | 0.150 |
| **32** | | | | | | | | 0.125 |
| **33** | | | | | | | | 0.100 |
| **34** | | | | | | | | 0.025 |
| **mean** | 3.000 | 14.000 | 4.000 | 17.600 | 3.500 | 15.400 | 2.425 | 26.700 |
| **scv** | 0.067 | 0.017 | 0.125 | 0.008 | 0.151 | 0.026 | 0.305 | 0.048 |

Table A.5.: Analysis of the G/G$^{[L,K]}$/1-queue: Input values for different examples; the results are illustrated in Figure 5.20 and 5.21

| | Income orders 1 | | Income orders 2 | | Collecting 1 | Collecting 2 | Order split |
|---|---|---|---|---|---|---|---|
| $i$ | $a_i$ | $y_i$ | $a_i$ | $y_i$ | $k = 12$ | $k = 8$ | $p_{i,j}$ |
| **0** | 0.00 | 0.0 | 0.00 | 0.0 | | | 0.6 |
| **1** | 0.00 | 0.1 | 0.10 | 0.5 | | | 0.4 |
| **2** | 0.40 | 0.4 | 0.05 | 0.4 | | | |
| **3** | 0.15 | 0.3 | 0.20 | 0.1 | | | |
| **4** | 0.10 | 0.2 | 0.30 | | | | |
| **5** | 0.15 | | 0.20 | | | | |
| **6** | 0.15 | | 0.10 | | | | |
| **7** | 0.05 | | 0.05 | | | | |
| **mean** | 3.650 | 2.600 | 3.950 | 1.600 | | | |
| **scv** | 0.212 | 0.124 | 0.150 | 0.172 | | | |

| | Picking 1 | Picking 2 | Picking large goods | Packaging 1 | Packaging 2 | Packaging 3 | Collecting transport |
|---|---|---|---|---|---|---|---|
| $i$ | $b_i$ | $b_i$ | $b_i$ | $b_i$ | $b_i$ | | $k = 20$ |
| **0** | 0.0 | 0.0 | 0.00 | 0.00 | 0.0 | 0.00 | |
| **1** | 0.5 | 0.3 | 0.40 | 0.55 | 0.6 | 0.50 | |
| **2** | 0.4 | 0.4 | 0.25 | 0.30 | 0.3 | 0.30 | |
| **3** | 0.1 | 0.3 | 0.30 | 0.05 | 0.1 | 0.15 | |
| **4** | | | 0.05 | 0.10 | | 0.05 | |
| **5** | | | | | | | |
| **6** | | | | | | | |
| **7** | | | | | | | |
| **mean** | 1.600 | 2.000 | 2.000 | 1.700 | 1.500 | 1.750 | |
| **scv** | 0.172 | 0.150 | 0.225 | 0.315 | 0.200 | 0.257 | |

Table A.6.: Input values for the numerical case: Material flow analysis in a warehouse; the results are illustrated in Figure 6.5 and Table 6.1

| | arrival and service process | | systems figures | |
|---|---|---|---|---|
| | $K$ | **8** | | |
| **i** | $a_i$ | $b_i$ | | |
| **0** | 0.0 | 0.00 | $c_O$ | 15 |
| **1** | 0.0 | 0.00 | $c_{In}$ | 40 |
| **2** | 0.3 | 0.00 | time period | 100 |
| **3** | 0.3 | 0.00 | | |
| **4** | 0.1 | 0.00 | | |
| **5** | 0.2 | 0.00 | | |
| **6** | 0.1 | 0.00 | | |
| **7** | | 0.00 | | |
| **8** | | 0.00 | | |
| **9** | | 0.00 | | |
| **10** | | 0.00 | | |
| **11** | | 0.00 | | |
| **12** | | 0.05 | | |
| **13** | | 0.05 | | |
| **14** | | 0.05 | | |
| **15** | | 0.10 | | |
| **16** | | 0.05 | | |
| **17** | | 0.15 | | |
| **18** | | 0.20 | | |
| **19** | | 0.05 | | |
| **20** | | 0.15 | | |
| **21** | | 0.05 | | |
| **22** | | 0.05 | | |
| **23** | | 0.05 | | |
| **mean** | **3.50** | **16.55** | | |

Table A.7.: Input values for a G/G$^{[L,K]}$/1-queue optimization problem; the results are illustrated in Figure 5.28

# A.2. Results

| | $D$ | | | $Y_d$ | | | $W$ | | |
|---|---|---|---|---|---|---|---|---|---|
| **L** | mean | 95%-quantile | 99%-quantile | mean | 95%-quantile | 99%-quantile | mean | 95%-quantile | 99%-quantile |
| **1** | 10.000 | 10 | 10 | 6.250 | 10 | 12 | 4.500 | 9 | 9 |
| **2** | 10.000 | 10 | 10 | 6.250 | 10 | 12 | 4.500 | 9 | 9 |
| **3** | 10.007 | 10 | 10 | 6.255 | 10 | 11 | 4.492 | 9 | 9 |
| **4** | 10.110 | 11 | 13 | 6.319 | 10 | 11 | 4.454 | 9 | 9 |
| **5** | 10.465 | 13 | 15 | 6.541 | 10 | 11 | 4.438 | 9 | 10 |
| **6** | 11.192 | 15 | 17 | 6.995 | 10 | 11 | 4.572 | 10 | 12 |
| **7** | 12.281 | 17 | 20 | 7.676 | 10 | 11 | 4.934 | 12 | 14 |
| **8** | 13.606 | 19 | 22 | 8.516 | 10 | 11 | 5.497 | 13 | 16 |
| **9** | 15.097 | 21 | 24 | 9.444 | 11 | 11 | 6.184 | 15 | 18 |
| **10** | 16.656 | 23 | 26 | 10.416 | 12 | 12 | 6.935 | 16 | 20 |
| **11** | 18.242 | 25 | 28 | 11.405 | 13 | 13 | 7.715 | 18 | 22 |

Table A.8.: Results which show the dependency of the interdeparture time $D$, the collected batch size $Y_d$, and the waiting time $W$ on the minimum batch size $L$ applying the batch building mode under the minimum batch size rule; the results are illustrated in Figures from 4.18 to 4.21

Universität Karlsruhe (TH)

Institut für Fördertechnik
und Logistiksysteme (IFL)

Scope of this doctoral thesis is the development of appropriate models for the evaluation of batch processes in material flow systems. The presented analytical methods support the long range planning in an early planning stage, in which capacities are determined to minimize the facility costs under the condition of cycle time targets. In this planning stage a rough and extensive ``what-if" analysis is required in order to find a competitive solution. We choose an analytical approach for a performance evaluation of material flow systems since it is more time efficient and allows deep insights into the general system's behavior.

Performance measures based on system averages are not sufficient to verify whether the requested shipping times can be met with an acceptable probability, which usually lies between 95% and 99%, possibly depending on order types. Therefore, for the evaluation of design alternatives in respect to their ability to reach the requested sojourn time from order entry to exit, discrete time queueing models are proposed.

Since there is still a lack of appropriate discrete time models for the analysis of material flow processes, we are motivated to find new solutions for problems in this field. Especially, models for the description of batch processes are missing. Therefore, we develop a variety of batch queueing models in the discrete time domain. The proposed models for the analysis of batch processes are: Batch building under the capacity rule, batch building under the timeout rule, batch building under the minimum batch size rule, G/G/1-queue with batch arrivals, G/G[K,K]/1-queue, G/G[L,K]/1-queue, and batch split. For each queueing system we provide methods for a detailed analysis of the waiting and departure process on the basis of discrete distributions.

ISSN: 0171-2772
ISBN: 978-3-86644-167-5

www.uvka.de