Electronic Communications of the EASST
Volume 080 (2021)

Conference on Networked Systems 2021
(NetSys 2021)

Discrete-time Analysis of Multicomponent GI/GI/1 Queueing Networks

Stefan Geissler, Stanislav Lange, Phuoc Tran-Gia, Tobias Hossfeld

4 pages

ECEASST Home Page: http://www.easst.org/eceasst/ ISSN 1863-2122

Discrete-time Analysis of Multicomponent GI/GI/1 Queueing Networks

Stefan Geissler[1], **Stanislav Lange**[2], **Phuoc Tran-Gia**[1], **Tobias Hossfeld**[1]

[1] University of Wuerzburg
[2] Norwegian University of Science and Technology

Abstract: In this work, we provide initial insights regarding the error introduced into multicomponent queueing systems by assuming the departure processes of arbitrary GI/GI/1-∞ queues to be renewal processes. To this end, we compute the sojourn time distribution as well as departure distributions of a linear chain of queueing components and compare the results to a simulation of the same system. By applying the renewal approximation, potential autocorrelations of the departure processes are lost. We investigate the magnitude of this error regarding both the sojourn time as well as interdeparture time distributions for a broad set of parameters. Although more indepth studies are needed, our results show that both distributions can be closely approximated, which allows the application of the model to asses the performance of real world NFV function chains.

Keywords: discrete-time analysis, queueing theory, queueing networks

1 Introduction

The complexity of both, applications as well as their underlying networks and compute architectures has increased substantially over the last few years. The abundance of cost-efficient cloud resources and the emergence of cloud-native microarchitectures, as well as the introduction of paradigms like network functions virtualization (NFV) and software-defined networking (SDN) have led to systems becoming significantly more complex. While legacy system are often designed as large, monolithic structures, modern systems consist of a multitude of largely independent actors that communicate with each other in order to perform their tasks. This microarchitecture paradigm allows cost-efficient resource scaling, improves system resilience, and shortens release cycles. However, the decomposition of functions into microservices (MS) introduces new challenges regarding the modeling and performance evaluation of such systems. Since requests traverse several MSs, the performance of the whole system is composed of the performance of chains of MSs. Finally, such models can be used during the dimensioning of complex, interconnected systems to establish required processing rates for each component as well as for the detection of bottlenecks in existing systems.

To this end, we propose an iterative approach to numerically compute the departure process of individual GI/GI/1-∞ MSs and show that by utilizing the resulting departure process as an arrival process to a subsequent system component, we can approximate several performance characteristics of downstream system components. The model is based on the assumption that the departure processes of individual components are renewal processes and do not exhibit any autocorrelation. This, in general, does not hold true in arbitrary systems.

Hence, in this work, we describe the approach itself as well as quantify the error introduced through the renewal approximation for different system configurations.

2 Related Work

Queueing stations of the type GI/GI/1-∞ as well as general queueing networks composed of multiple such queueing stations have been investigated in the past by, e.g. Tran-Gia [1], Kuehn [2], and Whitt [3]. Similarly, the evaluation of the waiting time distribution of GI/GI/1 systems has been evaluated previously [4–6]. Additionally, we have shown previously that the waiting time distribution can be determined numerically, even for more complex systems [7, 8].

3 Interdeparture Time Distribution Model

The model used in this work is based on the model proposed in [1]. To disambiguate between random variables (RVs) and distributions, we use the following convention: uppercase letters such as A denote RVs, their distribution is represented by $a(k)$. Accordingly, the model input is composed of the interarrival time distribution $a(k)$, as well as the service time distribution $b(k)$. Based on that, waiting time $w(k)$ and interdeparture time $d(k)$ can be calculated as follows.

To compute the interdeparture time distribution, we first need to compute the waiting time distribution $w(k)$. This can be done using Lindley's equation, in which $*$ denotes the convolution. Note that we assume the system to be in a stable state and can hence use $c(k)$, $a(k)$, and $b(k)$ without indices indicating their respective arrival.

$$w_{n+1}(k) = \pi_0(w_n(k) * c(k)) \text{ with } c(k) = a(-k) * b(k) \tag{1}$$

$$\pi_0(x(k)) = \begin{cases} x(k) & k > 0 \\ \sum_{i=-\infty}^{0} x(i) & k = 0 \\ 0 & k < 0 \end{cases}$$

The waiting time can then be used to compute the idle time distribution $i(k)$ via the virtual unfinished work $u^v(k)$. The idle time distribution describes the time a system is idle after a departure event.

$$u^v(k) = w(k) * c(k) = w(k) * a(-k) * b(k) \tag{2}$$

$$i(k) = K \cdot u^v(-k) \text{ with } K^{-1} = \sum_{j=1}^{\infty} u^v(-j)$$

Finally, the interdeparture time distribution $d(k)$ can be computed using the service time distribution $b(k)$, the idle probability P_E, and the idle time distribution $i(k)$.

$$d(k) = P_E \cdot (i(k) * b(k)) + (1 - P_E) \cdot b(k) \text{ with } P_E = \frac{E[A] - E[B]}{E[I]} \tag{3}$$

Analogously, the sojourn time distribution $s(k)$ of linear chains with n components can be approximated via $w(k)$ and $b(k)$. Note that this neglects potential autocorrelations in $w_i(k)$.

$$s(k) = w_1(k) * b_1(k) * w_2(k) * b_2(k) * \ldots * w_n(k) * b_n(k) \tag{4}$$

4 Evaluating Multicomponent Queueing Systems

In the following, we apply the model to iteratively compute the interdeparture time distribution of component i, which is then applied as an interarrival time distribution to component $i+1$. We assume a system of linearly concatenated GI/GI/1-∞ queueing components without cross-traffic (cf. Figure 1). Here, we assume the departure process to be a renewal process, which, in general, is not true. Instead, the departure process of a GI/GI/1-∞ queue must be assumed to exhibit autocorrelation. Subsequently, we quantify the error introduced by assuming D_1, D_2, D_3 to be renewal processes when it comes to a) the interdeparture time distribution and b) the distribution of the sojourn time for the whole system. In order to assess the error, the model results are compared to simulations with the same input parameters.

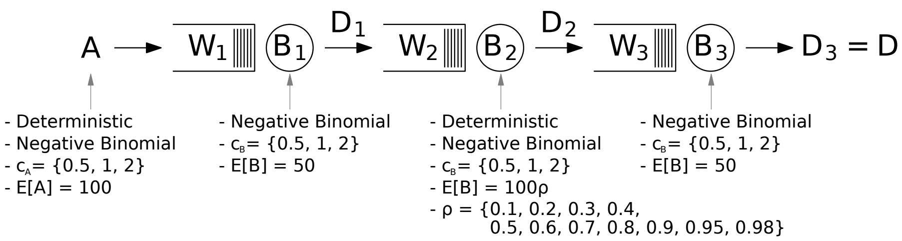
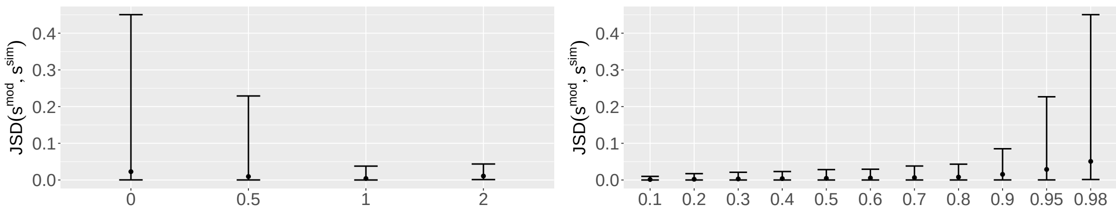


Figure 1: System and parameter overview.



(a) c_A (b) ρ

Figure 2: Mean, max, and min JSD of the sojourn time $s(k)$ for different values of c_A and ρ.

Figures 2 a) and b) show main effect plots of the Jenson-Shannon-Divergence (JSD) for different values of c_A and ρ. c_B has been omitted here since the effects are similar to the ones induced by c_A. The figures show all JSD values obtained by evaluating every parameter combination possible from the values presented in Figure 1.

The results show that the sojourn time can be closely approximated by assuming D_1, D_2, D_3 to be renewal processes. The 95% quantile of JSD values is 0.037 for the sojourn time. A similar evaluation regarding the interdeparture distribution of the last component has shown a 95% quantile of 0.0004. Regarding the sojourn time, the largest error with a JSD value of 0.45 has been observed for a deterministic arrival process and deterministic service time of the second processing unit in combination with high system load ($\rho = 0.98$). This is due to the fact that deterministic arrival and service times lead to maximum autocorrelation regarding the departure process. However, since deterministic distributions are rarely observed in technical systems, when filtering these processes out, the highest observed error was a JSD value of 0.13 for $c_A = 0.5$, $c_{B_1} = c_{B_3} = 2$, $c_{B_2} = 0.5$ and $\rho = 0.98$.

5 Conclusion

In this work, we provide preliminary insights into the linear concatenation of GI/GI/1-∞ queueing components. We apply existing models for the calculation of the interdeparture time distribution in order to iteratively compute the waiting time distribution of further components in the chain. Finally, we are able to compute the sojourn time for the total system and show that, for the parameter combinations evaluated in this work, a close approximation is possible. Moving forward, future work includes the extension of the parameter space to cover more cases as well as the introduction of new parameters such as the length of the chain as well as autoregressive arrival processes. Furthermore, the impact of cross-traffic and feedback loops needs to be investigated.

References

[1] P. Tran-Gia, "Discrete-time analysis for the interdeparture distribution of gi/g/1 queues," in *Proc. of the international seminar on Teletraffic analysis and computer performance evaluation*, 1986, pp. 341–357.

[2] P. Kuehn, "Approximate analysis of general queuing networks by decomposition," *IEEE Transactions on communications*, vol. 27, no. 1, pp. 113–126, 1979.

[3] W. Whitt, "The queueing network analyzer," *The bell system technical journal*, vol. 62, no. 9, pp. 2779–2815, 1983.

[4] J. Abate, G. L. Choudhury, and W. Whitt, "Calculation of the gi/g/1 waiting time distribution and its cumulants from pollaczek's formulas," *Archiv für Elektronik und Ubertragungstechnik*, vol. 47, no. 5/6, pp. 311–321, 1993.

[5] W. K. Grassmann and J. L. Jain, "Numerical solutions of the waiting time distribution and idle time distribution of the arithmetic gi/g/1 queue," *Operations Research*, vol. 37, no. 1, pp. 141–150, 1989.

[6] M. Murata and H. Miyahara, "An analytic solution of the waiting time distribution for the discrete-time gi/g/1 queue," *Performance evaluation*, vol. 13, no. 2, pp. 87–95, 1991.

[7] T. Zinner, S. Geissler, S. Lange, S. Gebert, M. Seufert, and P. Tran-Gia, "A discrete-time model for optimizing the processing time of virtualized network functions," *Computer Networks*, vol. 125, pp. 4–14, 2017.

[8] S. Geissler, T. Prantl, S. Lange, F. Wamser, and T. Hossfeld, "Discrete-time analysis of the blockchain distributed ledger technology," in *2019 31st International Teletraffic Congress (ITC 31)*, IEEE, 2019, pp. 130–137.