Queueing Systems (2022) 100:425–427
https://doi.org/10.1007/s11134-022-09776-5

# Policy robustness in queueing networks

**Itai Gurvich**[1] · **John J. Hasenbein**[2]

Received: 3 February 2022 / Accepted: 28 February 2022 / Published online: 2 April 2022


## 1 Introduction

Robustness has been, for decades now, a prominent topic in the optimization literature. Yet, it is only in the past few years that robust techniques have been imported and further developed for stochastic models, particularly stochastic processing networks. A reason for this late adoption might be that "non-robust" analysis of general stochastic networks already presents significant mathematical challenges. Hence, for much of the first 80 years of the study of queueing models robustness was not explicitly considered in most models that appeared in the literature.

Papers that do address robustness in queueing networks focus on two major robustness angles: *parameter robustness*, and *data robustness*. In the former group, the statistical structure for the arrival and service processes is known (e.g., it is a Poisson process) but the arrival and service rates (the parameters) are only known to lie within some uncertainty set; [4] is one of the earlier papers in this space. More recently, [5] studies an $M/M/s$ queue where only the mean and support of the arrival-rate distribution is known to the manager who must make staffing decisions.

Data robustness, in contrast, does not impose a parametric statistical structure and takes the view that the realized process (e.g., the sequence of inter-arrival times) belongs to a suitably defined uncertainty set and uses robust optimization to obtain approximations and bounds for performance metrics; see [1,3]. Further recent development along these lines appears in [2,10].

## 2 Problem statement

We would like to advocate, in addition to these angles, for a third notion of robustness in stochastic networks, namely *policy robustness*. Here, what is uncertain is the policy

✉ Itai Gurvich
i-gurvich@kellogg.northwestern.edu

John J. Hasenbein
jhas@mail.utexas.edu

[1] Kellogg School of Management, Northwestern University, Evanston, IL, USA

[2] Graduate Program in Operations Research & Industrial Engineering, University of Texas at Austin, Austin, TX, USA

that is used at each station in a stochastic network. In multiclass queueing networks (MQNs), the policy specifies the order in which jobs are served at each station. More generally, the policy could include decisions about setups, batching, and routing, for example. In MQNs, the "uncertainty set," as such, consists of a large, but well-specified, collection of scheduling policies.

This set can be as large as the collection of all non-idling policies. In order to further define a notion of policy robustness, one must specify a performance metric or network property that is of interest. To keep the exposition simple, we focus on stability (e.g., positive Harris recurrence).

We can now define a specific version of the **policy robustness problem**. In particular, for a MQN with a fixed topology, for what set of network parameters is the system stable, given an uncertainty set of policies?

Policy robustness is an appealing property of a network. It means that the network's manager or resources have flexibility to determine local prioritization without concern for first-order network-level objectives (like maximal throughput). The uncertainty set also provides a mathematical means to embody *decentralized control* for applications, such as telecommunications, in which centralized control is impractical.

In this general framing, this is not a new question. The notion of global stability [7,8] precisely concerns robustness (in the sense of stability) within the set of all non-idling policies. Finding parameters, e.g., service time means, for which a given network is policy-robust is the same as finding the *global stability* region.

Global stability is well understood for certain network topologies (feedforward networks), certain routing matrices (those corresponding to generalized Jackson networks), and limited-sized networks (specifically, two-station fluid networks). Beyond two-station MQNs, the global-stability region is fully characterized only for specific networks [7,8]; these papers highlight the difficulty of coming up with a general approach to global stability.

We advocate for a re-framing of global stability that connects more directly to robustness ideas in optimization and has the potential to advance the understanding of global stability. To create this connection, it is useful to consider uncertainty sets that are more restrictive (i.e., smaller than the set of all non-idling policies) but can lead to a more tractable analysis; see §1.

Ideally, a generalizable framework can also produce answers to problems that are in some sense "dual" to that of finding the global stability region (the set of policy robust parameters). One might be interested in characterizing network topologies that are policy robust. Meaning that, for these networks, any non-idling policy is stable as long as the usual traffic conditions hold, i.e., $\rho_j < 1$ for every station $j$ (see [6]).

## 3 Discussion

In forthcoming work [11], we have made progress on policy robustness for the case in which the uncertainty set contains all fixed-queue ratio policies (see [9]). We prove that the stability of a policy within this uncertainty set is inherited from the stability of "vertices" of the set. These vertices represent static buffer priority policies. In other words, if the parameters—specifically, mean service times—are such that *all* static buffer priority policies are stable, then so is any policy in the "interior" of the uncertainty set.

Our result relies on two, potentially generalizable, principles. First, because a Skorohod Problem (representing the fluid workload model) plays a key role in our analysis so do, in turn, the so-called reflection matrices. We show that certain "good" properties of the reflection matrices satisfy a (limited) form of convexity. If one of these properties is satisfied at *all* vertices, it is also satisfied at any point in the interior.

Second, this convexity property can be verified by framing verification of matrix properties as a robust optimization problem and applying known robust optimization arguments. In other words, the analysis reduces verification of stability to verification at corner points of the uncertainty set. In some cases, this leads to an explicit characterization of the (family of) parameters for which a network is policy robust.

Our hope is that the connections we draw between stability, convexity, and robust optimization can serve as the basis for, and motivate, a general framework toward policy robustness that applies, beyond MQNs, to flexible server networks, constrained multi-hop networks, and reflected Brownian motion models.

The ideas in this paper showcase the potential in bringing together ideas (and scholars) from applied probability and optimization to re-visit and make progress on fundamental questions in stochastic modeling.

## References

1. Bandi, C., Bertsimas, D., Youssef, N.: Robust queueing theory. Oper. Res. **63**(3), 676–700 (2015)
2. Bandi, C., Trichakis, N., Vayanos, P.: Robust multiclass queueing theory for wait time estimation in resource allocation systems. Manage. Sci. **65**(1), 152–187 (2019)
3. Bertsimas, D., Gamarnik, D., Rikun, A.A.: Performance analysis of queueing networks via robust optimization. Oper. Res. **59**(3), 455–466 (2011)
4. Chen, B.P.K., Henderson, S.G.: Two issues in setting call center staffing levels. Ann. Oper. Res. **108**(1), 175–192 (2001)
5. Chen, Y., Hasenbein, J.J.: Staffing large-scale service systems with distributional uncertainty. Queueing Syst. Theory Appl. **87**(1–2), 55–79 (2017)
6. Dai, J.G.: On positive Harris recurrence of multiclass queueing networks: a unified approach via fluid limit models. Ann. Appl. Probab. **5**, 49–77 (1995)
7. Dai, J.G., Hasenbein, J.J., VandeVate, J.H.: Stability of a three-station fluid network. Queueing Syst. Theory Appl. **33**, 293–325 (1999)
8. Dumas, V.: A multiclass network with non-linear, non-convex, non-monotonic stability conditions. Queueing Syst. Theory Appl. **25**, 1–43 (1997)
9. Gurvich, I., Whitt, W.: Service-level differentiation in many-server service systems: a solution based on fixed-queue-ratio routing. Oper. Res. **29**, 567–588 (2007)
10. Whitt, W., You, W.: Using robust queueing to expose the impact of dependence in single-server queues. Oper. Res. **66**(1), 184–199 (2018)
11. Zhao, F., Gurvich, I., Hasenbein, J.: Policy robustness: an optimization view of global stability. Working paper (2022)