

Capacity Allocation in Queueing Networks

Xuefeng GAO (ISYE)

Mentor: Prof Ton Dieker (ISYE)

Context:

A Queueing network is a collection of service centers. Customers arrive at the system at random epochs, and when they finish service at one center, they may be routed to another center to receive other kinds of services. Customers may need to wait in a queue if the service center cannot immediately process them. After they are done with all required service, they leave the system. Examples include computer data systems, call centers, and health care systems.

Problem:

The objective of the proposed research is to develop methods to do capacity optimization in a **large** feedforward queueing network. That is, given a cost constraint on service personnel, how do we allocate service capacity in the network in order to minimize our objective function, which is the total number of customers in the system? A feedforward network is a network where we can number the stations in an increasing order and customers always flow from lower numbers to high numbers of stations.

This problem is challenging but interesting. We have no explicit formulas for the objective as a function of the capacity allocation vector in general. Besides, numerical methods and simulations suffer from high computational cost especially in the case of large network. Therefore, it is our goal to **devise efficient and provable algorithms** to tackle the capacity allocation problem in large queueing network setting.

Motivation:

There are many applications of our problem. An illustrating example would be a call center of a bank. There can be many individual divisions even just for personal banking, such as check card, credit card, online banking technical support, etc. Customers may have several types of inquiries and their calls are routed in this large queueing network, handled by agents in each division. So it is a very important problem for the bank to know how many agents should be hired for each division.

Existing Methods:

1): When the data of the network satisfies a strong 'product form' condition, Wein [2] obtained analytical results for our problem. We work on the general case where there is no restriction on the network data.

2): When the allocation vector is given, there are two numerical methods for

analyzing a queueing network: QNA developed by W. Whitt and QNET by J. Dai and J. M. Harrison. But both methods are limited to small number of stations since the computational complexity grows very fast in the size of the network. Thus, they are impractical for analyzing large queueing networks.

Our Approaches:

We will work in the framework of Brownian network which provides reasonable approximation to the original queueing network while amenable to mathematical analysis. Besides, we want to explore the structural properties of our objective as a function of the capacities allocated to stations, such as monotonicity and convexity results. Ongoing work by Dieker *et. al* [4] develops results and algorithms for **small** networks and they use QNET when searching for optimal allocation rule. However, we are much more interested in **large** networks, which are more relevant in applications. In our case, it's inefficient for repeated evaluation of the objective function when searching for optimal allocation rule. Actually, there are no existing numerical methods that can be directly applied due to the high dimensionality. The main idea we current have involves three steps:

- 1): Decompose the large network into small subnetworks in appropriate ways. It's possible to use the idea of sequential bottleneck decomposition by Dai *et al.* [3] to come up with a candidate choice, but if the network has additional special structure, we may also be able to break down the network in other ways.
- 2): Allocate service capacity to those subnetworks. Essentially, we replace those multistation subnetworks as single stations. If the size of the "new" network is still large in the computational point of view, then sequential decomposition is necessary.
- 3): Allocate the capacity within each subnetwork. Since the subnetworks only have a few stations, we can allocate the capacity in a computationally efficient way.

References:

- [1] Harrison, J.M. R. J. Williams (1992). Brownian models of feedforward queueing networks: quasireversibility and product form solutions Ann. Appl. Probab. **2** 263-293
- [2] L. M. Wein, (1989). Capacity allocation in generalized Jackson networks. Oper. Res. Lett. **8** no.3, 143-146
- [3] J. G. Dai, V. Nguyen, and M. Reiman, (1994) Sequential bottleneck decomposition: An approximation method for generalized Jackson networks, Operations Research **42** 119-136
- [4] A. B. Dieker, S. Ghosh and M. S. Squillante (2009). Capacity allocation for Brownian feedforward stochastic networks. Ongoing