Facilitating Load-Dependent Queueing Analysis Through Factorization (Extended Abstract)

Giuliano Casale, Peter G. Harrison, Wai Hong Ong
Department of Computing
Imperial College London, UK
{g.casale, p.harrison, wai.ong11}@imperial.ac.uk

ABSTRACT

We construct novel exact and approximate solutions for mean-value analysis and probabilistic evaluation of closed queueing network models with limited load-dependent (LLD) nodes. In this setting, load-dependent functions are assumed to become constant after a finite queue-length threshold. For single-class models, we provide an explicit formula for the normalizing constant that applies to models with arbitrary LLD functions, whilst retaining constant complexity with respect to the total population size. From this result, we then derive corresponding closed-form solutions for the multiclass case and show that these yield a novel mean value analysis approach for LLD models. Significantly, this allows us to determine exactly the *correction factor* between a load-independent solution and a limited load-dependent one, enabling the reuse of state-of-the-art methods for load-independent models in the analysis of load-dependent networks.

Keywords

Normalizing constant, queueing network, closed system

1. INTRODUCTION

Despite a long history of work on this subject [6], when limited load-dependent (LLD) stations are considered in multiclass closed networks, which are natural representations of systems with finite levels of parallelism, they remain difficult to analyze. Commonly employed evaluation techniques, such as mean value analysis (MVA) [8], are significantly less efficient in the LLD setting than in the case of models that include only fixed-rate stations and infinite servers, i.e., the so-called *load-independent* (LI) case. For example, in LI models, mean queue lengths alone can be efficiently used to determine recursively the network equilibrium performance, as is routinely done in the classic MVA algorithm [8]. Conversely, the solution of a LLD model requires the entire queue-length distribution at each station to be carried through the load-dependent MVA algorithm recursion (MVA-LD) [9], significantly increasing computational requirements. Besides being inefficient, this approach can introduce numerical instabilities [7].

In this abstract, we report the key results derived in [4], which revisits the problem. The paper determines in particular novel exact and approximate solutions for product-form LLD closed queueing networks, both in the single-class and multiclass cases. For a full derivation of the results, as well as several additional contributions, we point the reader to [4].

1.1 Reference model

We study closed multiclass queueing networks that admit a product-form solution [1]. The queueing network model under study has M LLD queues and R job classes. Each class is populated by N_r jobs, with $N = N_1 + \ldots + N_R$. The state space of the underlying Markov process is $\mathcal{S}(\boldsymbol{N}) = \{\boldsymbol{n} = (n_{1,1}, \ldots, n_{M,R}) \mid n_{k,r} \geq 0, \sum_{k=1}^{M} n_{k,r} = N_r\}$, where we focus only on the marginal states in which $n_{k,r}$ denotes the total number of class-r jobs residing at queue k, either queueing or receiving service.

For queue k, we denote by $\theta_{k,r}$ the mean service demand of class r, which is the product of the mean number of visits with the mean service time of a class r job at queue k. If queue k is load-dependent then the service demand of the job in service is scaled by a load-dependent factor $\alpha_k(n_k)$ if the queue has $n_k = \sum_{r=1}^{R} n_{k,r}$ resident jobs. Throughout the paper we consider load-dependent scaling factors $\alpha_k(n_k)$ that meet the LLD condition, i.e., $\alpha_k(n_k) = \alpha_k(s_k)$, $\forall n_k \geq s_k$. For finite populations, this assumption does not reduce generality, as one may consider a general load-dependent model as having $s_k = N$, $\forall k$.

With the above definitions, the equilibrium distribution of the network is then given by [1]

$$\pi(\boldsymbol{n}) = \frac{1}{H_\theta(\boldsymbol{N})} \prod_{k=1}^{M} \frac{n_k!}{\alpha_k(n_k)} \prod_{r=1}^{R} \frac{\theta_{k,r}^{n_{k,r}}}{n_{k,r}!} \qquad \boldsymbol{n} \in \mathcal{S}(\boldsymbol{N}) \quad (1)$$

where $\boldsymbol{N} = (N_1, \ldots, N_R)$. The normalizing constant $H_\theta(\boldsymbol{N})$ in (1) ensures that probabilities sum to unity. Furthermore, mean performance metrics can be directly derived from it [2].

2. SINGLE CLASS LLD MODELS

Closed-form expressions for normalizing constants in LI models are derived in [5, Eq. (29)], which obtains a closed-form solution for single-class LLD models with single-server and multi-server stations. All such results are generalized by the following expression obtained in [4], which applies to arbitrary LLD models.

THEOREM 2.1. *In a single-class LLD closed queueing network with M stations*

$$h_{\boldsymbol{\theta}}(N) = \sum_{\boldsymbol{0} \leq \boldsymbol{v} < \boldsymbol{s}} g_{\boldsymbol{\sigma}}(N - v) \Phi_{\boldsymbol{\theta}}(\boldsymbol{v}) \quad (2)$$

where $\boldsymbol{v} = (v_1, \ldots, v_M)$, $v = |\boldsymbol{v}|$, $\boldsymbol{s} = (s_1, \ldots, s_M)$, and we define $\Phi_{\boldsymbol{\theta}}(\boldsymbol{v}) = \prod_{k=1}^{M} \phi_k(v_k)$, and

$$\phi_k(v_k) = \frac{\theta_k^{v_k}}{\prod_{t=1}^{v_k} \alpha_k(t)} \left(1 - \frac{\alpha_k(v_k)}{\alpha_k(s_k)}\right)$$

for $0 \leq v_k < s_k$.

In the above result, $g_{\boldsymbol{\sigma}}(N - v)$ represents the normalizing constant of a LI model with scaled demands $\boldsymbol{\sigma} = (\sigma_1, \ldots, \sigma_M)$, where $\sigma_i = \theta_i/\alpha_i(s_i)$ and total job population $N - v$.

It is possible to derive from the above several novel expressions for the LLD normalizing constant also in the multiclass setting. This is made possible by a finite difference relation connecting the single class normalizing constant $h_{\boldsymbol{\theta}}(N)$ with its multiclass counterpart $H_{\boldsymbol{\theta}}(\boldsymbol{N})$. In particular, [4] derives novel integral expressions, such as formulations on the unit simplex and novel expressions leveraging the Norlund-Rice integral form for finite differences. Notably, the latter requires one to consider integrands that are themselves single-class normalizing constants, but with complex service demands. As such, specialized results are obtained in [4] to evaluate normalizing constants with complex demands as well as their complex derivatives.

3. MULTICLASS LLD MODELS

As shown in [4], it is also possible to establish a relationship similar to (2) in the multiclass setting through the following theorem.

THEOREM 3.1. *The normalizing constant for a model with M LLD queueing stations and R classes, having demands $\theta_{k,r}$ and scaling functions α_k, admits the following reduced convolution expression*

$$H_{\boldsymbol{\theta}}(\boldsymbol{N}) = \sum_{v=0}^{V} \sum_{\substack{\boldsymbol{d} \geq 0: \\ |\boldsymbol{d}| = v}} G_{\boldsymbol{\sigma}}(\boldsymbol{N} - \boldsymbol{d}) E_{\boldsymbol{\theta}}(\boldsymbol{d}) \qquad (3)$$

where $\boldsymbol{d} = (d_1, \ldots, d_R)$ and $V = \min(N, \sum_{k=1}^{M}(s_k - 1))$. Here, $G_{\boldsymbol{\sigma}}$ is the multiclass normalizing constant of a LI model with scaled demands $\sigma_{k,r} = \theta_{k,r}/\alpha_k(s_k)$, and we define

$$E_{\boldsymbol{\theta}}(\boldsymbol{d}) = \sum_{\boldsymbol{v} \in \boldsymbol{S}(\boldsymbol{d})} \prod_{i=1}^{M} \frac{v_i!}{\prod_{k=1}^{v_i} \alpha_i(k)} \left(1 - \frac{\alpha_i(v_i)}{\alpha_i(s_i)}\right) \prod_{r=1}^{R} \frac{\theta_{i,r}^{v_{i,r}}}{v_{i,r}!} \qquad (4)$$

$E_{\boldsymbol{\theta}}(\boldsymbol{d})$ may itself be seen as a LLD multiclass normalizing constant for a model with demands $\theta_{k,r}$ and suitably defined scaling functions [4].

Using Theorem 3.1, [4] establishes in particular the exact correction factor between the LLD model under study and the related LI model with normalizing constant $G_{\boldsymbol{\sigma}}(\boldsymbol{N})$.

THEOREM 3.2 (EXACT LLD CORRECTION). *The normalizing constant for a model with M LLD queueing stations and R classes, having demands $\theta_{k,r}$ and scaling functions α_k, may be obtained from the normalizing constant of a related fixed-rate model with scaled demands $\boldsymbol{\sigma}$ as follows*

$$H_{\boldsymbol{\theta}}(\boldsymbol{N}) = \Gamma(\boldsymbol{N}) G_{\boldsymbol{\sigma}}(\boldsymbol{N}) \qquad (5)$$

where the LLD correction factor is the quantity

$$\Gamma(\boldsymbol{N}) = \sum_{v=0}^{V} \sum_{\substack{\boldsymbol{d} \geq 0: \\ |\boldsymbol{d}| = v}} \prod_{(\boldsymbol{s},r) \in P(\boldsymbol{d},\boldsymbol{N})} X_r^{\boldsymbol{\sigma}}(\boldsymbol{s}) E_{\boldsymbol{\theta}}(\boldsymbol{d}) \qquad (6)$$

where $P(\boldsymbol{d}, \boldsymbol{N}) = \{(\boldsymbol{s}, r) \,|\, \boldsymbol{s} = (N_1, \ldots, N_{r-1}, n_r, N_{r+1} - d_{r+1}, \ldots, N_R - d_R) : \forall r = 1, \ldots, R; \forall n_r = N_r - d_r, \ldots, N_r\}$, V is defined in Theorem 3.1, $X_r^{\boldsymbol{\sigma}}(\boldsymbol{N})$ is the mean system throughput of class r in the LI model, and $E_{\boldsymbol{\theta}}(\boldsymbol{d})$ as in (4).

A notable consequence of the last result is the following characterization of the class-r system throughput.

THEOREM 3.3. *The exact relationship between the mean system throughput $X_r(\boldsymbol{N})$ of a LLD model and the corresponding metric $X_r^{\boldsymbol{\sigma}}(\boldsymbol{N})$ in a fixed-rate model with scaled demands $\boldsymbol{\sigma}$, is given by*

$$X_r(\boldsymbol{N}) = \frac{\Gamma(\boldsymbol{N} - 1_r)}{\Gamma(\boldsymbol{N})} X_r^{\boldsymbol{\sigma}}(\boldsymbol{N}) \qquad (7)$$

for all classes $r = 1, \ldots, R$, and in which $\Gamma(\cdot)$ is the LLD correction factor defined in (6).

Stemming from this result, [4] derives a novel approximate mean-value analysis method, called *reduction heuristic*, that is shown to produce more accurate approximations for LLD models than existing techniques. This is shown over several thousands of experiments reported in [4].

4. REFERENCES

[1] F. Baskett, K. M. Chandy, R. R. Muntz, and F. G. Palacios. Open, closed, and mixed networks of queues with different classes of customers. *JACM*, 22:248–260, 1975.

[2] S. C. Bruell and G. Balbo. *Computational Algorithms for Closed Queueing Networks*. North-Holland, 1980.

[3] S. C. Bruell, G. Balbo, and P. V. Afshari. Mean value analysis of mixed, multiple class BCMP networks with load dependent service stations. *Perform. Eval.*, 4:241–260, 1984.

[4] Giuliano Casale, Peter G. Harrison, and Wai Hong Ong. Facilitating load-dependent queueing analysis through factorization. *Performance Evaluation*, 152:102241, 2021.

[5] J. J. Gordon. The evaluation of normalizing constants in closed queueing networks. *Operations Research*, 38(5):863–869, 1990.

[6] S. S. Lavenberg. A perspective on queueing models of computer performance. *Perform. Eval.*, 10(1):53–76, 1989.

[7] M. Reiser. Mean-value analysis and convolution method for queue-dependent servers in closed queueing networks. *Perform. Eval.*, 1:7–18, 1981.

[8] M. Reiser and S. S. Lavenberg. Mean-value analysis of closed multichain queueing networks. *JACM*, 27(2):312–322, 1980.

[9] C. H. Sauer. Computational algorithms for state-dependent queueing networks. *ACM Trans. Comp. Syst.*, 1:67–92, 1983.