

Corrected phase-type approximations for the workload of the MAP/G/1 queue with heavy-tailed service times

Citation for published version (APA):

Vatamidou, E., Adan, I. J. B. F., Vlasiou, M., & Zwart, A. P. (2013). *Corrected phase-type approximations for the workload of the MAP/G/1 queue with heavy-tailed service times*. Poster session presented at International Conference on Levy Processes, 15-19 July 2013, Wrocław, Poland, Wrocław, Poland.

Document status and date:

Published: 01/01/2013

Document Version:

Publisher's PDF, also known as Version of Record (includes final page, issue and volume numbers)

Please check the document version of this publication:

- A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
- The final author version and the galley proof are versions of the publication after peer review.
- The final published version features the final layout of the paper including the volume, issue and page numbers.

[Link to publication](#)

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license above, please follow below link for the End User Agreement:

www.tue.nl/taverne

Take down policy

If you believe that this document breaches copyright please contact us at:

openaccess@tue.nl

providing details and we will investigate your claim.

Corrected-PH approximations for the workload of the MAP/G/1 queue with heavy-tailed services

Eleni Vatamidou, Ivo Adan, Maria Vlasiou, and Bert Zwart



EURANDOM

Introduction

What makes the evaluation of the workload in a MAP/G/1 queue (FIFO) a challenging problem:

- significant correlations between arrivals, and
- heavy-tailed service times.

Here, we construct very accurate approximations of the workload with a small relative error.

Model description

Parameters related to arrivals:

- MArP with N states
- \mathbf{P} (trans. prob.), π (stat. dist.)
- Λ (exp. trans. rates), \mathbf{Q} (prob. real arriv. cust.)

Parameters related to service times:

- $G_\epsilon(t) = (1 - \epsilon)F_p(t) + \epsilon F_h(t)$, $\epsilon \in [0, 1)$, service time of real customer (mixture motivated by statistical analysis)
- $\tilde{\mathbf{G}}_\epsilon(s) = \tilde{G}_\epsilon(s)\mathbf{Q} + (\mathbf{I} - \mathbf{Q})$ (LST), μ_ϵ (mean)

Stability condition:

$$\pi (\Lambda^{-1} - \mu_\epsilon \mathbf{Q}) \mathbf{e} > 0,$$

LT of steady-state workload:

$$\tilde{\Phi}_\epsilon(s) = [\tilde{\phi}_{\epsilon,1}(s), \dots, \tilde{\phi}_{\epsilon,N}(s)].$$

Theorem. *There exists \mathbf{u} s.t. [2]*

$$\tilde{\Phi}_\epsilon(s) \left(\tilde{\mathbf{G}}_\epsilon(s)\mathbf{P}\Lambda + s\mathbf{I} - \Lambda \right) = s\mathbf{u}_\epsilon,$$

$$\tilde{\Phi}_\epsilon(0)\mathbf{e} = 1.$$

Target: total workload

$$\tilde{v}_\epsilon(s) = \frac{s \cdot \mathbf{u}_\epsilon \cdot \text{adj} \left(\tilde{\mathbf{G}}_\epsilon(s)\mathbf{P}\Lambda + s\mathbf{I} - \Lambda \right) \mathbf{e}}{\det \left(\tilde{\mathbf{G}}_\epsilon(s)\mathbf{P}\Lambda + s\mathbf{I} - \Lambda \right)}.$$

Approach

Perturbation on a PH base model. Steps:

1. PH approximation as base.

(a) Set $\epsilon = 0$ & $\tilde{F}_p(s) = q_n(s)/p_m(s)$, (polynomials $q_n(s), p_m(s)$).

(b) Evaluate \mathbf{u} , determinant and adjoint of $\tilde{\mathbf{G}}(s)\mathbf{P}\Lambda + s\mathbf{I} - \Lambda$.

(c) For $\text{Re}(y_j), \text{Re}(x_j) > 0, j = 1, \dots, mr$,

$$\tilde{v}(s) = \frac{\mathbf{u}\mathbf{e} \prod_{j=1}^{mr} (s + y_j)}{\prod_{j=1}^{mr} (s + x_j)}.$$

(d) $\mathbb{P}(V > t) = \mathcal{L}^{-1}\{\tilde{v}(s)\}$.

2. Take $G_\epsilon(t) = F_p(t) + \epsilon(F_h(t) - F_p(t))$ and find the perturbed parameters (ϵ pert. param.).

3. Find $\tilde{v}_\epsilon(s)$, by keeping only up to ϵ -order terms, i.e.,

$$\tilde{v}_\epsilon(s) = \tilde{v}(s) + \epsilon \tilde{v}(s)k(s) + O(\epsilon^2),$$

$k(s)$ is well-defined on \mathbb{R}^+ .

Results

Definition. CORRECTED-PH APPROXIMATIONS

$$\mathbb{P}(\hat{V}_\epsilon > t) := \mathbb{P}(V > t) + \epsilon \frac{1}{\mathbf{u}\mathbf{e}} \left(\alpha \mathbb{P}(V > t) + \beta (\mu_p \mathbb{P}(V + B^e > t) - \mu_h \mathbb{P}(V + C^e > t)) + \gamma (\mu_p \mathbb{P}(V + V' + B^e > t) - \mu_h \mathbb{P}(V + V' + C^e > t)) + \sum_{j=1}^{mr} \alpha_j \mathbb{P}(V + E_{y_j} > t) + \sum_{j=1}^{mr} \beta_j (\mu_p \mathbb{P}(V + B^e + E_{y_j} > t) - \mu_h \mathbb{P}(V + C^e + E_{y_j} > t)) + \sum_{j=1}^{mr} \gamma_j (\mu_p \mathbb{P}(V + V' + B^e + E_{y_j} > t) - \mu_h \mathbb{P}(V + V' + C^e + E_{y_j} > t)) \right).$$

Example. Erlang-2 arrival process with rate 5. Service times mixture of an Exp(3) distribution and a heavy-tailed one [1] with $\tilde{F}_h(s) = 1 - \frac{s}{(2+\sqrt{s})(1+\sqrt{s})}$ for $\epsilon = 0.1$.

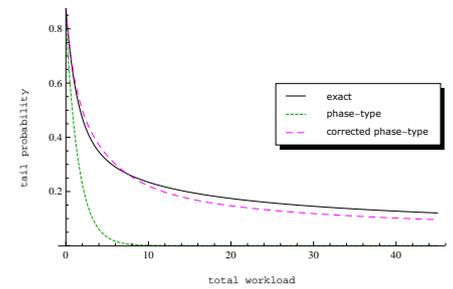


Figure 1: Comparison of exact total workload with approximations, for load 0.875

Conclusions

The approximations provide:

- correct tail behavior, and
- small absolute and relative error.

Future work: tandem queues.

References

- [1] ABATE, J. AND WHITT, W. (1999). Explicit $M/G/1$ waiting-time distributions for a class of long-tail service-time distributions. *Oper.Res.Lett.* 25, 25–31.
- [2] ADAN, I.J.B.F. AND KULKARNI, V.G. (2003). Single-server queue with Markov-dependent inter-arrival and service times. *Queueing Syst. Theory. App.* 45, 113–134.