Fluctuation Analysis in Parallel Queues with Hysteretic Control

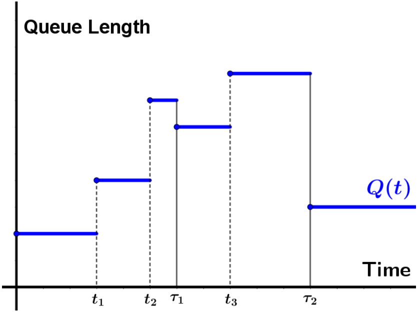
Ryan T. White*, Jewgeni H. Dshalalow, Ahmed Merie

Florida Institute of Technology

AMS Fall Southeastern Sectional Meeting 2019
University of Florida

Queue Length Process

- $Q(t)$ on a filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t), \mathbb{P})$
- Entering customers: batches X arrive at $t_1 < t_2 < \dots$
- Exiting customers: batches Y depart at $\tau_1 < \tau_2 < \dots$



QUEUE TYPE: $M^X/G^Y/1$ with several operational policies

Two Common Problems with Queueing Systems

- Switchovers can be costly

- Resources are wasted when the system is off

GOAL: Minimize switchovers and complete secondary tasks when possible while serving primary customers efficiently.

N-Policy: Reducing Switchovers

POLICY:

1. If $Q(t) < N$ and the system is off, wait until it reaches N.

2. Else, serve customers.

- Classical switchover mitigation technique (Yadin and Naor [1963])
  - $Q(t)$ small $\Longrightarrow$ queue is likely exhausted quickly and system turns off
  - $Q(t)$ large $\Longrightarrow$ queue will persist, system works continuously

BIG CON: customers must wait sometimes

r-R-Quorum: Completing Secondary Tasks

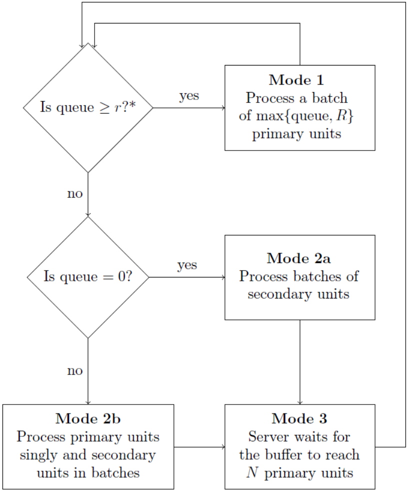
POLICY:

1. If $Q(t) = 0$, batch secondary service
2. If $0 < Q(t) < r$, parallel service. Single primary service, batch secondary service
3. If $Q(t) > r$, primary service to batch of $\max\{\text{queue}, R\}$

- Related to some classical queueing ideas
  - r-quorum (e.g. Neuts [1967])
  - Hysteretic control (e.g. Loris-Tegham [1978])

MAIN BENEFITS: less primary waiting, secondary work done

The Queueing System



Fluctuation Analysis and Operational Calculus

- For each mode, we will seek a functional of changes during a mode

$$\Phi(u,v,w,\theta) = \mathbb{E}\left[u^{\text{primary served}} v^{\text{secondary served}} w^{\text{primary arrivals}} e^{-\theta(\text{duration})}\right]$$

- We apply modified z-transforms

$$D_p(\cdot)(x) = \sum_{p=0}^{\infty} x^p(\cdot)(1-x) \text{ to } \Phi = \sum_{P\in\mathcal{P}} \Phi_P = \sum_{P\in\mathcal{P}} \mathbb{E}\left[(\cdots)\mathbb{1}_P\right]$$

for a countable partition $\mathcal{P}$ of Ω, e.g., for $j,k \in \mathbb{Z}_{\geq 0}$,

$$P_{jk} = \{\text{primary service finishes at } j\text{th cycle, secondary finishes at } k\text{th cycle}\}$$

- Operational calculus technique

$$\Phi \xrightarrow{D_p D_q} \Psi \xrightarrow{\text{Assumptions on the system}} \Psi \text{ (convenient form)} \xrightarrow{D_y^{-1} D_x^{-1}} \Phi \text{ (tractable)}$$

Results For Mode 2A

- Secondary Service + Primary Arrivals
- We find Φ as a function of a transform of the increment of primary arrivals, secondary batch size, secondary service time upon service completions,

$$\mathbb{E}\left[z^{\pi_1} w^{v_1} e^{-\theta s_1}\right]$$

under a modified z-transform

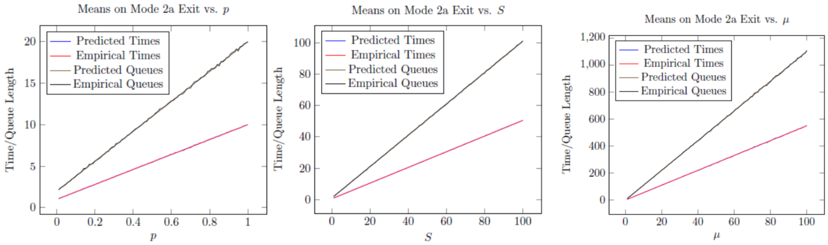
Means for Mode 2A

- Secondary batch sizes geometric(p) $\Longrightarrow$

$$\begin{array}{ccccc} \mathbb{E}[s_\rho] & = & (1-p+pS) & \times & \mathbb{E}[s_1] \\ \text{mode duration} & = & \text{secondary batches served} & \times & \text{secondary service time} \end{array}$$

$$\begin{array}{ccccccc} \mathbb{E}[\Pi_\rho] & = & \lambda & \times & \mathbb{E}[s_\rho] & \times & \mathbb{E}[\mathfrak{a}_1] \\ \text{queue length} & = & \text{arrivals/time} & \times & \text{mode duration} & \times & \text{arriving batch size} \end{array}$$

- Secondary service times exponential(μ) $\Longrightarrow$



Results for Mode 2B

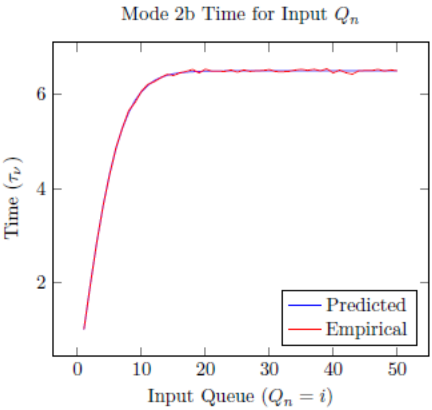
- Single Primary Service + Secondary Service + Primary Arrivals
- We find Φ as a function of a transform of the increment of the same upon each primary service completion,

$$\mathbb{E}\left[u^{x_1} v^{y_1} w^{p_1} e^{-\theta \Delta_1}\right]$$

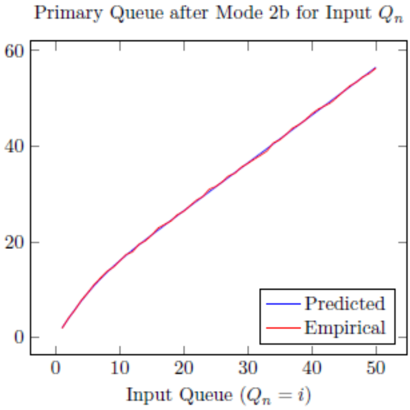
under a composition of modified z-transforms

Special Case Results for Mode 2B

mean mode time
$\mathbb{E}[\tau_\nu]$



mean queue length at mode exit
$\mathbb{E}[Q_n - X_\nu + \Pi_\nu]$



Mode 3

- Primary Arrivals
- Wait until $Q(t) \geq N$ (maybe instantaneous)
- Similar results
    - Transform of queue length and time on mode exit
    - Moments in special cases

Queueing Results

- Transition probability matrix (irreducible and aperiodic)
- Ergodicity conditions (conditions for TPM to be recurrent positive similar to Abolnikov and Dukhovny [1991])
- Stationary distribution (as a solution to a simple linear system)
- Mean stationary service cycle

Dshalalow, Merie, and White [2019], Dshalalow and Merie [2018]

References

L. Abolnikov and A. Dukhovny. Markov chains with transition delta-matrix: ergodicity conditions, invariant probability measures and applications. *Journal of Applied Mathematics and Stochastic Analysis*, 4(4):335–355, 1991.

J. H. Dshalalow and A. Merie. Fluctuation analysis in queues with several operational modes and priority customers. *TOP*, 26(2):309–333, 2018. doi: 10.1007/s11750-018-0478-3.

J. H. Dshalalow, A. Merie, and R. T. White. Fluctuation analysis in parallel queues with hysteretic control. *Methodology and Computing in Applied Probability*, pages 1–33, 2019. doi: 10.1007/s11009-019-09701-z.

J. Loris-Tegham. Hysteretic control of an M/G/1 queueing system with two service time distributions and removable server. In *Point Processes and Queueing Problems*, volume 24 of *Colloquia Mathematica Societatis Janos Bolyai*, pages 291–305. North-Holland, Amsterdam, 1978.

M. F. Neuts. A general class of bulk queues with Poisson input. *The Annals of Mathematical Statistics*, 38(3):759–770, 1967.

M. Yadin and P. Naor. Queueing systems with a removable service station. *Journal of the Operational Research Society*, 14(4):393–405, 1963. doi: 10.1057/jors.1963.63. URL `https://doi.org/10.1057/jors.1963.63`.