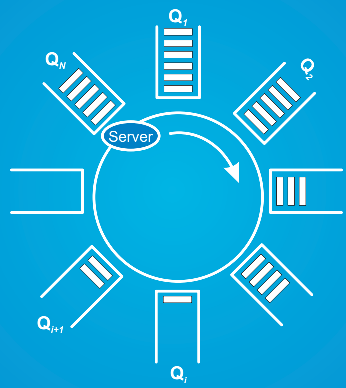
Fairness and Efficiency in Waiting Times for Polling Models



Sandra van Wijk
A.C.C.v.Wijk@tue.nl

Joint work with Ivo Adan, Onno Boxma, and Adam Wierman

Technische Universiteit **Eindhoven** University of Technology

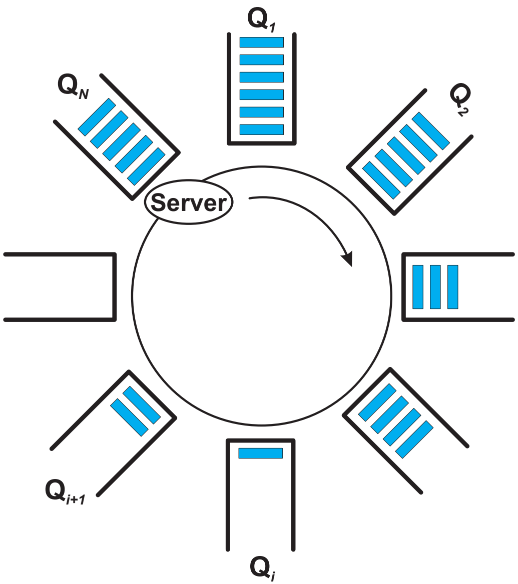
Second Israeli–Dutch Workshop on Queueing Theory

Wednesday September 29, 2010

Introduction

Polling Model

- N queues,
- Arrivals at Q_i: Poisson(λ_i),
- Service time at Q_i: B_i,
- Load at Q_i: $\rho_i = \lambda_i \mathsf{E}(B_i)$,
- Visit time at Q_i: V_i,
- Switch-over time from Q_i to Q_{i+1}: S_i,
- Waiting time customer at Q_i: W_i.

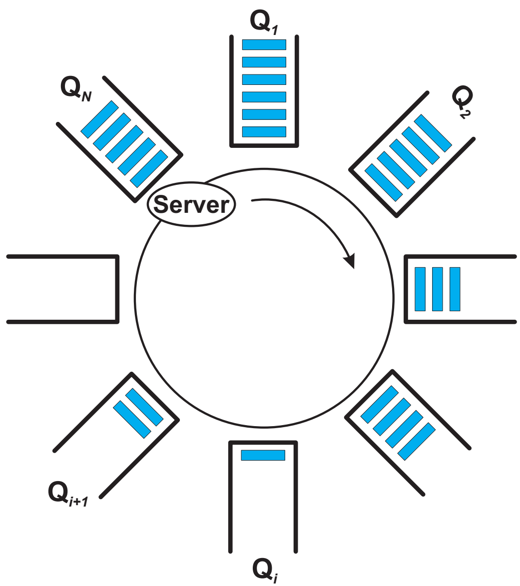


Cycle: $V_1 - S_1 - V_2 - S_2 - \ldots - V_N - S_N$.

Ordinary service discipline: e.g. exhaustive, (glob.) gated.

Applications: telecommunication, repairman, production, etc.

Fairness vs. Efficiency



Fairness: (other definitions exist)

$$\max_i \mathsf{E}(W_i) - \min_j \mathsf{E}(W_j)$$

Efficiency:

$$\sum_i \rho_i \mathsf{E}(W_i)$$

Typically:

- efficient service disciplines are unfair (e.g. exhaustive)
- fair service disciplines are inefficient

Fairness vs. Efficiency (II)

Fair service disciplines (less efficient), e.g.:

- gated
- two stage gated (Park et al. 2005; Van der Mei and Resing 2008)
- elevator polling glob. gated (Altman et al. 1992)

Previous work:

- one cycle gated, one cycle exhaustive (Boxma et al. 2008)

$\rightarrow$ Introduce κ-Gated Discipline

Introduction

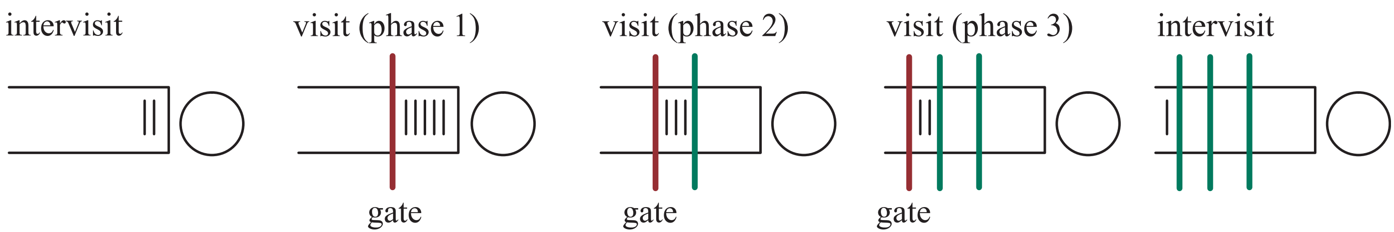
κ-Gated Discipline

κ-Gated discipline is hybrid version of:

- exhaustive: efficient but not fair
- gated: more fair but inefficient

Parameter: $\kappa = (\kappa_1, \kappa_2, \ldots, \kappa_N)$

Serve queue i subsequently (at most) κ_i 'times' gated



Technische Universiteit Eindhoven University of Technology

κ-Gated Discipline (II)

Goal: set κ to minimize

$$\gamma(\alpha) = \max_i \mathsf{E}(W_i) - \min_j \mathsf{E}(W_j) + \alpha \sum_i \rho_i \mathsf{E}(W_i)$$

Outline

- Pseudo conservation law
- Waiting time distributions
- Mean waiting times
- Fluid limits $\rightarrow$ heuristic setting for κ
- Performance heuristic

Analysis

Pseudo Conservation Law

expression for $\sum_i \rho_i \mathsf{E}(W_i)$ (cf. Boxma and Groenendijk 1987)

$$\sum_i \rho_i \mathsf{E}(W_i) = \rho \frac{\sum_i \rho_i \mathsf{E}(R_{B_i})}{1-\rho} + \rho\, \mathsf{E}(R_S) + \frac{\mathsf{E}(S)}{2\,(1-\rho)} \left(\rho^2 - \sum_i \rho_i^2\right) + \sum_i \mathsf{E}(M_i).$$

where

$$\mathsf{E}(M_i) = \rho_i^{\kappa_i+1} \frac{1-\rho_i}{1-\rho_i^{\kappa_i}} \frac{\mathsf{E}(S)}{1-\rho}.$$

Efficiency: $\sum_i \mathsf{E}(M_i)$

Waiting Time Distributions

using Multi-Type Branching Processes (cf. Resing 1993)

Queue length process: N-type branching process with immigration.

Each customer present effectively replaced (i.i.d.) by random population with pgf $h_i(z_1, \ldots, z_N)$:

$$h_i^{\text{(1-gated)}}(\underline{z}) = h_i^{\text{(gated)}}(\underline{z}) = \beta_i \left(\sum_{j=1}^{N} \lambda_j (1 - z_j) \right);$$

$$h_i^{(m\text{-gated})}(\underline{z}) = \beta_i \left(\sum_{j=1, j \neq i}^{N} \lambda_j (1 - z_j) + \lambda_i \left(1 - h_i^{((m-1)\text{-gated})}(\underline{z}) \right) \right), \; m = 2, 3, \ldots.$$

We derive joint and marginal queue length distributions, and waiting time distributions.

Mean Waiting Times

(cf. Boon et al. 2009)

κ-Gated discipline fits into the framework of a polling model with smart customers (arrival rate depends on server position).

Introduce extra queues and route customers and route to correct queue:

$$V_1^{(1)} - V_1^{(2)} - \ldots - V_1^{(\kappa_1)} - S_1 - V_2^{(1)} - V_2^{(2)} - \ldots - V_2^{(\kappa_2)} - S_2 - \ldots - S_{N-1} - V_N^{(1)} - \ldots - V_N^{(\kappa_N)} - S_N$$

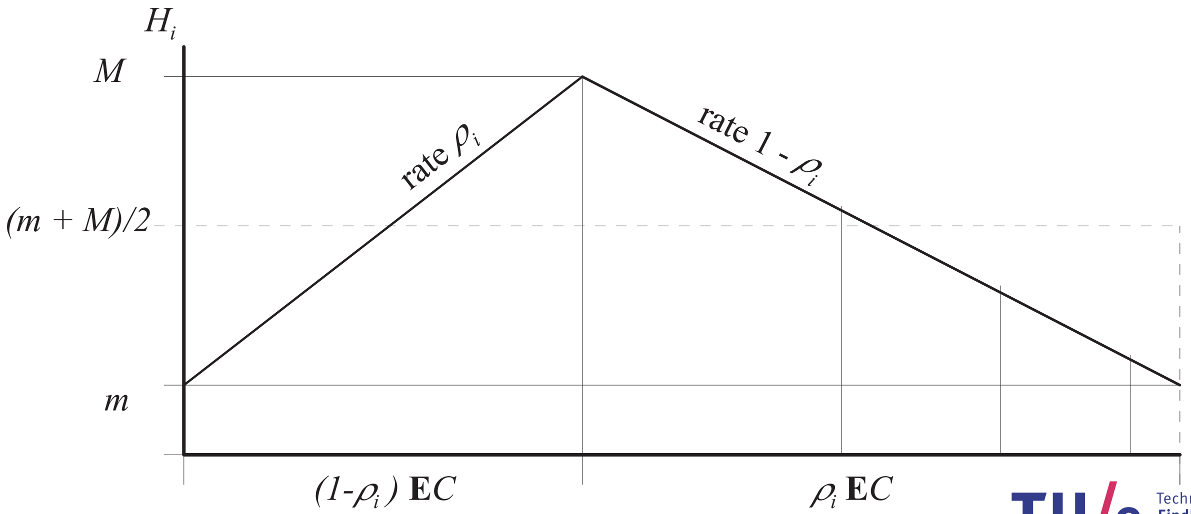
Now Mean Value Analysis for polling models gives mean waiting times in an easy way.

Fluid limits

scale: $\lambda_i \to \infty$ and $\mathsf{E}(B_i) \to 0$ while keeping the workload $\lambda_i\, \mathsf{E}(B_i) = \rho_i$ fixed

Gives closed form expression for (approximation of) $\mathsf{E}(W_i)$:

$$\mathsf{E}\left(W_i^{fluid}\right) = \frac{m+M}{2\,\rho_i} = (1+\rho_i^{\kappa_i})\frac{1-\rho_i}{2\,(1-\rho_i^{\kappa_i})}\mathsf{E}(C).$$



Fluid limits $\rightarrow$ heuristic

Maximal fairness using fluid limits:

$$\mathsf{E}\left(W_1^{fluid}\right) = \mathsf{E}\left(W_2^{fluid}\right) = \ldots = \mathsf{E}\left(W_N^{fluid}\right)$$

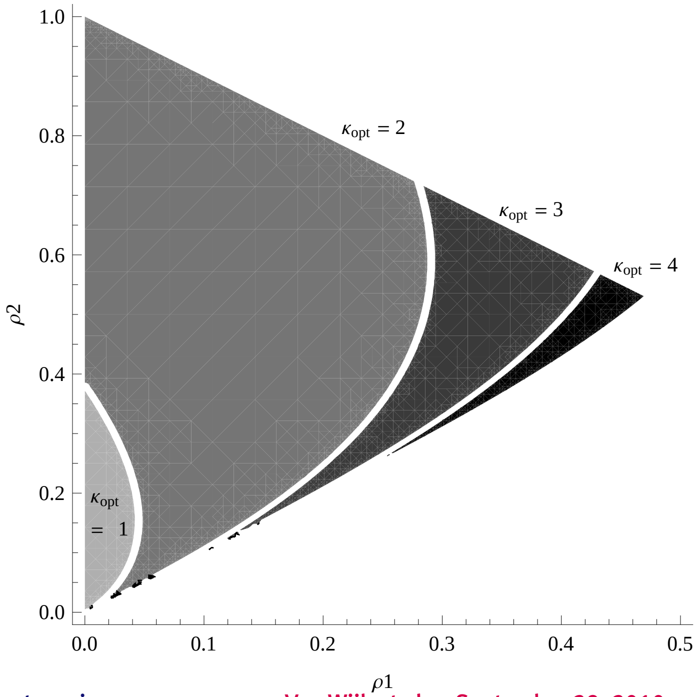
$\rightarrow$ family of solutions for $\kappa_1, \kappa_2, \ldots, \kappa_N$.

Take most efficient solution:

$$\begin{cases} \text{For all } i \text{ such that } i = \arg\min \rho_i, \text{ let } \kappa_i = \infty; \\ \text{For all } j = 1, 2, \ldots, N \text{ where } j \neq i, \text{ let } \kappa_j = \log_{\rho_j} \frac{\rho_j - \rho_i}{2 - \rho} \end{cases}$$

Fluid limits $\to$ heuristic (II)

$$\begin{cases} \text{For all } i \text{ such that } i = \arg\min \rho_i, \text{ let } \kappa_i = \infty; \\ \text{For all } j = 1, 2, \ldots, N \text{ where } j \neq i, \text{ let } \kappa_j = \log_{\rho_j} \frac{\rho_j - \rho_i}{2 - \rho} \end{cases}$$



Results

Numerical results

Example for 2 queues: $\lambda_1 = 0.35,\ \lambda_2 = 0.25,$
$B_i \sim \mathsf{exp}(1),\ S_i \sim \mathsf{exp}(2).$

Heuristic settings: $\kappa = (3, \infty)$.

κ_1	κ_2	$\mathsf{E}(W_1)$	$\mathsf{E}(W_2)$	Δ	$\sum \mathsf{E}(M_i)$	$\gamma(0)$	$\gamma(1)$	$\gamma(2)$	$\gamma(5)$
1	1	9.3	8.6	0.6	1.8	0.6	2.5	4.3	9.9
∞	1	**5.1**	9.5	4.3	0.6	4.3	5.0	5.6	7.5
1	∞	9.7	**5.6**	4.0	1.2	4.1	5.3	6.5	10.2
2	∞	6.7	6.0	0.7	0.3	0.7	1.0	1.4	2.3
3	∞	6.0	6.2	0.2	0.1	0.3	**0.4**	**0.5**	**0.8**
∞	∞	5.6	6.4	0.8	**0.0**	0.9	0.9	0.9	0.9
Elev.GG		11.5	11.5	**0.0**	3.9	**0.0**	3.9	7.9	39.3

Testbed with over 4,500 instances: heuristics performs very well

Conclusion

Summary

Introduced κ-gated service discipline for polling systems.

Waiting times (distribution and means), PCL, fluid limits.

Heuristic setting for κ for ‘fairness and efficiency’, performs well.

A.C.C. van Wijk, I.J.B.F. Adan, O.J. Boxma and A. Wierman,

Fairness in Waiting Times for Polling Models with the κ-Gated Service Discipline,

In preparation.