Bad luck when joining the shortest queue

Hans Blanc

Dept. Econometrics & Operations Research

Tilburg University, The Netherlands

Eindhoven, July 2007

Overview of presentation

- Problem description
- Method of analysis
- Results
- Conclusion

Problem description

- Systems with two or more servers in parallel
- Separate queues in front of the servers
- Poisson arrivals, exponential services
- Customers join one of shortest queues upon arrival
  - ∗ No jockeying; FIFO within each queue
  - ∗ System stable ($\rho < 1$); stationary situation
- Frequent observation: overtaking by later arriving customers in other queues: probability of bad luck

Analysis

- Global balance equations for multidimensional birth-death process
- Numerical solution by Power-Series Algorithm
  - ∗ Blanc, *Operation Research* 1992
  - – Insert power-series expansions in load ρ
  - – Recursive scheme for computing coefficients
  - – Improve convergence of series by ϵ-algorithm
- Result: joint queue-length distribution as function of load
  - – $p(\mathbf{n}) \doteq p(n_1, \ldots, n_c) = \sum_{k=0}^{\infty} \rho^{k+n_1+\cdots+n_c} b(k; \mathbf{n})$

Analysis

- W_n: waiting time given n customers in queue
- Conditional probability of bad luck (given queue lengths upon arrival): $\Pr\{\min_{i=1,\ldots,c} W_{n_i} < W_{n_j}\}$
  - ∗ Involves repeated integration of Erlang densities
  - ∗ Can be reduced to finite sums
- Unconditional probability of bad luck:

$$P_{\mathrm{BL}} \doteq \sum_{n_1} \cdots \sum_{n_c} p(\mathbf{n})\, \gamma_j(\mathbf{n})\, \Pr\{\min_{i=1,\ldots,c} W_{n_i} < W_{n_j}\}$$

  - – $\gamma_j(\mathbf{n})$: probability of joining queue j given $\mathbf{n}$: $0, \frac{1}{c}, \ldots, \frac{1}{2}, 1$
- Power-series expansion $p(\mathbf{n})$: power-series expansion P_{BL}

Probability of bad luck (conditional)

- Case: two servers with equal service rates
- Customer finds n_j customers in queue j, $j = 1, 2$
- Customer joins queue 1
- Conditional probability of bad luck: $\Pr\{W_{n_2} < W_{n_1}\}$

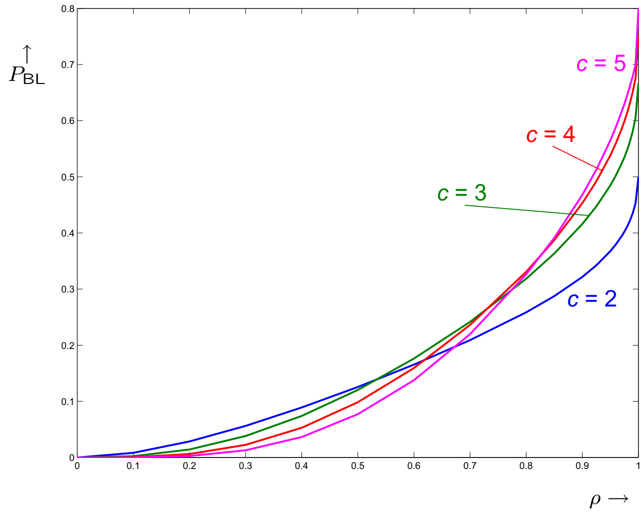
$n_2 \backslash n_1$	1	2	3	4	5	6
6	0.0156	0.0625	0.1445	0.2539	0.3770	0.5000
5	0.0313	0.1094	0.2266	0.3633	0.5000	0.6230
4	0.0625	0.1875	0.3438	0.5000	0.6367	0.7461
3	0.1250	0.3125	0.5000	0.6563	0.7734	0.8555
2	0.2500	0.5000	0.6875	0.8125	0.8906	0.9375
1	0.5000	0.7500	0.8750	0.9375	0.9688	0.9844

Probability of bad luck (conditional)

- Case: three servers with equal service rates
- Customer finds n_j customers in queue j, $j = 1, 2, 3$
- Customer joins queue 1 where $n_1 = 2$
- Conditional probability of bad luck: $\Pr\{W_{n_2} \vee W_{n_3} < W_{n_1}\}$

$n_3 \backslash n_2$	2	3	4	5	6
6	0.5066	0.3271	0.2117	0.1431	0.1045
5	0.5158	0.3448	0.2379	0.1764	0.1431
4	0.5364	0.3813	0.2887	0.2379	0.2117
3	0.5802	0.4527	0.3813	0.3448	0.3271
2	0.6667	0.5802	0.5364	0.5158	0.5066

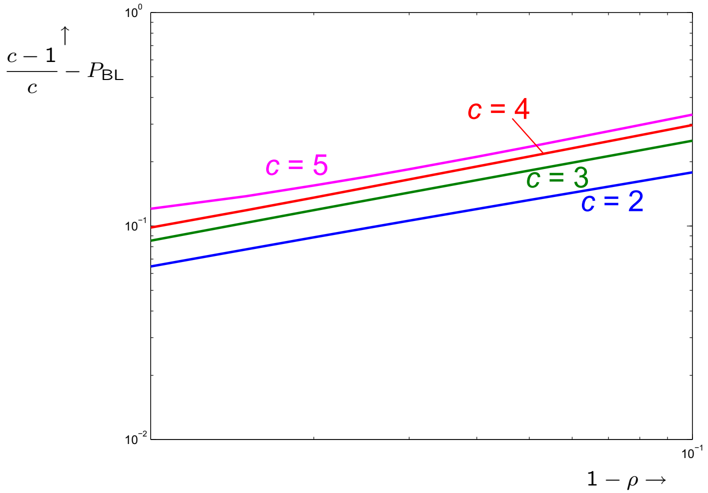
Probability of bad luck (unconditional)



Probability of bad luck as function of load ρ

Symmetric systems, $c = 2, 3, 4, 5$ servers.

Probability of bad luck in heavy traffic



Heavy traffic

$0.9 < \rho < 0.99$
$0.01 < 1-\rho < 0.1$

logarithmic scale.

Symmetric systems.

Probability of bad luck in heavy traffic

Asymptote of probability as $\rho \uparrow 1$:

- general form: $P_{BL} \sim \frac{c-1}{c} - A(1-\rho)^q$
- for $c = 2$: $P_{BL} \sim \frac{1}{2} - 0.51\,(1-\rho)^{0.45}$
- for $c = 3$: $P_{BL} \sim \frac{2}{3} - 0.75\,(1-\rho)^{0.47}$
- for $c = 4$: $P_{BL} \sim \frac{3}{4} - 0.88\,(1-\rho)^{0.48}$

Probability of bad luck in light traffic

Asymptote of probability as $\rho \downarrow 0$:

- general form (from power-series expansions):

$$P_{BL} \sim \frac{c^{c-2}}{(c-2)!}\,\rho^c - (c^3 - c^2 - c + 2)\,\frac{c^{c-2}}{c!}\,\rho^{c+1} + O(\rho^{c+2})$$

- for $c = 2$: $P_{BL} \sim \rho^2 - 2\,\rho^3 + O(\rho^4)$
- for $c = 3$: $P_{BL} \sim 3\rho^3 - \frac{17}{2}\rho^4 + O(\rho^5)$
- for $c = 4$: $P_{BL} \sim 8\rho^4 - \frac{92}{3}\rho^5 + O(\rho^6)$

Bad luck when joining the shortest queue

Remarks on computation: time-consuming procedure

- computation time Power-Series Algorithm:
  - ∗ strongly increases with c (coefficients up to same power of load)
    - – although not as fast as for comparable multidimensional birth-death process
  - ∗ coefficients up to higher power of load required due to atypical heavy-traffic asymptote
- computation time conditional probability of bad luck:
  - ∗ also strongly increases with c (multiplicity of sums)

Bad luck when joining the shortest queue

Variants on symmetric systems:

- servers with unequal service rates (customers don't know)
  - ∗ worse in light traffic
    - – about equal probability of joining queue at slower server
  - ∗ better in heavy traffic
    - – most customers served by faster servers
- systems with dedicated customers for some or all servers
  - ∗ smaller probability of bad luck for customers joining shortest queue
    - – reduces probability of finding all queues equally short
    - – the more so as dedicated traffic is more asymmetric

Bad luck when joining the shortest queue

Conclusion

In symmetric systems:

- Probability of bad luck may well exceed 0.5
  - ∗ although optimal to join shortest queue: bad luck
  - ∗ but only occurs in quite heavily loaded systems
  - ∗ at the same load, larger probability with more servers if $\rho \uparrow 1$
    - – clear due to limit $(c-1)/c$
  - ∗ at the same load, smaller probability with more servers if $\rho \downarrow 0$
    - – more servers: larger probability of zero waiting time
- Conditional probability of bad luck given state upon arrival
  - ∗ largest if all queues equally short: $(c-1)/c$