Performance analysis and decomposition results for some dynamic priority schemes in 2-class queues

Prathamesh Mayekar
J. Venkateswaran
Manu K. Gupta
N. Hemachandra

Industrial Engineering and Operations Research
Indian Institute of Technology Bombay

May 26, 2015

Overview

Introduction

- D2D communication problems can be modeled by multi-class tandem queuing network (See [Mumtaz and Rodriguez, 2014])
- Quality of service is important aspect in D2D communication problems
- Quality of service differentiation can be achieved by:
  - Strict priority queuing systems
  - Dynamic priority queuing systems

Dynamic priority queing systems

- Performance measures of dynamic priority queueing system are not known analytically
- Following dynamic priority queueing system are analyzed via simulation:
  - Relative priority
  - EDD dynamic priority

Relative priority queuing system

- Proposed by Moshe Haviv and Van Der Wal (see [Haviv and van der Wal, 2007])
- A positive parameter p_i is associated with each class i
- Arrivals are assumed to be independent Poisson with rate λ_i
- Service can have any general expression

Relative priority queuing system

- n_j jobs of class j on service completion
- Probability that class i job commences service:

$$\frac{n_i p_i}{\sum_{j=1}^{N} n_j p_j}, \quad 1 \le i \le N$$

Earliest Due Date (EDD) Dynamic Priority

- Proposed by Henry M. Goldberg (see [Goldberg, 1977])
- Each class i has a constant urgency number u_i
- Customer from class i arriving at time t_i has urgency number $t_i + u_i$
- Customer with minimum value of $\{t_i + u_i\}$ commences service

Performance analysis

- Performance measures
  - Tail probability of waiting time
  - Switching frequency
- Simulation Details
  - Simulation has been done using Simpy package in Python (see [Matloff, 2008])
  - Simulation run time :11,000 time units
  - Warm up period : 1,000 time units

Tail Probability for relative priority queuing system

- Parameter setting:
  $\lambda_1 = 6, \ \lambda_2 = 8, \ \mu = 15, \ p_1 = 0.3$ and $p_2 = 0.7 = 1 - p_1$ and t is varying from 0.3 to 3

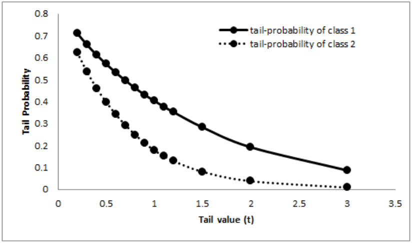

Figure: Change in $P(W_1 > t)$ and $P(W_2 > t)$ by varying t

Tail Probability for relative priority queuing system

- Parameter setting: $t = 0.8$, $\lambda_1 = 6$, $\lambda_2 = 8$, $\mu = 15$ and p_1 is varied from 0.1 to 0.9

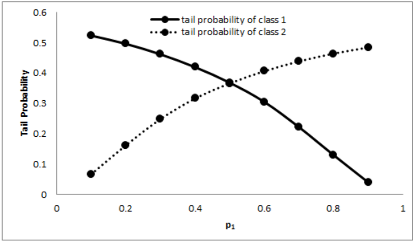

Figure: Change in $P(W_1 > t)$ and $P(W_2 > t)$ by varying p_1

- Tail probabilities of class 1 and class 2 customers are equal at $p_1 = 0.5$

Tail Probability for relative priority queuing system

- Parameter setting:
  $t = 0.5, \ \lambda_2 = 8, \ \mu = 15, \ p_1 = 0.3 \ p_2 = 0.7$ and λ_1 is varied from 1to6

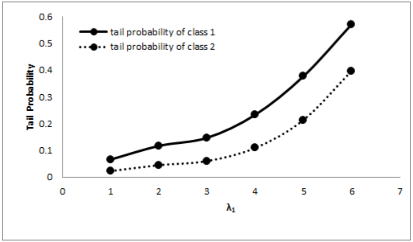

Figure: Change in $P(W_1 > t)$ and $P(W_2 > t)$ by varying λ_1

- Tail probability increases with higher rate for higher values of λ_1.

Tail probability of EDD Dynamic Priority

- Parameter setting: $\lambda_1 = 6, \ \lambda_2 = 8, \ u_1 = 2, \ u_2 = 5$ and $\mu = 15$

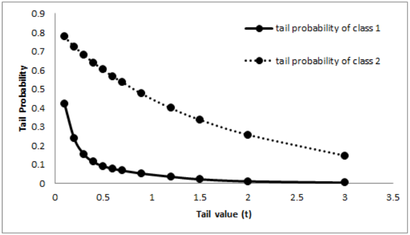

Figure: Change in $P(W_1 > t)$ and $P(W_2 > t)$ by varying t

Tail probability of EDD Dynamic Priority

- Parameter setting: $\lambda_1 = 6$, $\lambda_2 = 8$, $u_2 = 5$, $t = 0.2$ and $\mu = 15$

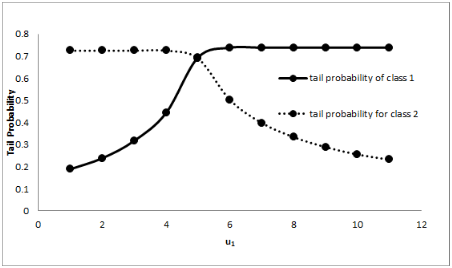

Figure: Change in $P(W_1 > t))$ and class 2 $P(W_2 > t)$ by varying u_1

- Tail probability remains constant for certain range of u_1

Tail probability of EDD Dynamic Priority

- Parameter setting: $\lambda_2 = 8,\ u_1 = 2,\ u_2 = 5,\ t = 0.2$ and $\mu = 15$

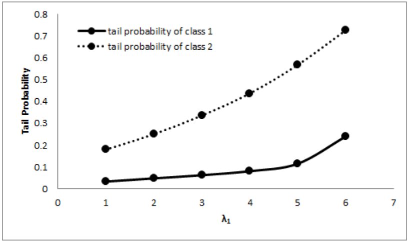

Figure: Change in $P(W_1 > t)$ and class 2 $P(W_2 > t)$ by varying λ_1

Switching frequency

$$\text{Switching frequency} = \frac{\text{Ratio of number of service switches among different classes}}{\text{the total number of customers served}}$$

Switching frequency for relative priority queuing system

- Parameter setting: $\lambda_1 = 4, \ \lambda_2 = 8, \ \mu = 15$

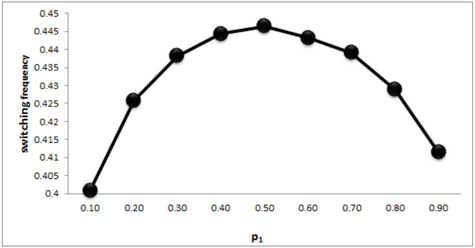

Figure: Switching frequency Vs p_1

- Switching frequency is a concave function when plotted against p_1
- Switching frequency maximum at $p_1 = 0.5$

Switching frequency for EDD Dynamic Priority

- Parameter setting: $\lambda_1 = 4,\ \lambda_2 = 8,\ \mu = 15,\ u_2 = 50$

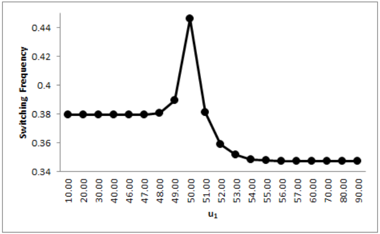

Figure: Switching frequency Vs u_1

- Switching frequeny changes only in a certain interval when u_1 close to u_2
- Switching frequency is maximum at $u_1 = u_2$ (global first come first serve)

Relative priority queuing network

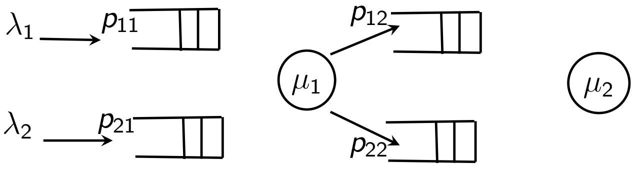

Figure: A two class two stage queueing network

Decomposition in waiting time for relative priority queuing network

Conjecture

A two class two stage queueing network with Poisson arrivals is decomposable in terms of mean waiting time when relative priority is used across classes in both stages for exponential service time distribution at first stage.

Verification of decompositon result

- Simulation Details(except for heavy traffic case)
  - Simulation has been done using Simpy package in Python
  - Simulation run time :11,000 time units
  - Warm up period:1,000 time units
  - Five replications of each experimental setting
  - Half width calculated using t-distribution at 95% level of confidence

Verification by varying relative priority parameters

- Parameter setting :
  $\lambda_1 = 4, \ \lambda_2 = 8, \ \mu_1 = 15, \ \mu_2 = 16, p_{11} = 0.3$ and $p_{12} = 0.6$

Replication	Differences at stage 1		Differences at stage 2	
	Class 1	Class 2	Class 1	Class 2
1	0.0078	0.0064	0.0014	-0.0112
2	-0.0073	-0.0002	0.0088	-0.0009
3	0.0068	0.0051	0.0082	-0.0003
4	-0.0046	-0.0003	-0.0025	-0.0139
5	-0.0058	-0.0014	0.0155	0.0065
mean	-0.0006	0.0019	0.0063	-0.0040
halfwidth	0.0091	0.0045	0.0088	0.0105

Verification by varying relative priority parameters

- Parameter setting :
  $\lambda_1 = 4,\ \lambda_2 = 8,\ \mu_1 = 15,\ \mu_2 = 16,\ p_{11} = 0.8$ and $p_{12} = 0.9$

Replication	Differences at stage 1		Differences at stage 2	
	Class 1	Class 2	Class 1	Class 2
1	0.0056	0.0217	-0.0049	0.0037
2	0.0007	0.0045	-0.0056	0.0117
3	0.0023	0.0050	-0.0045	0.0010
4	-0.0002	0.0065	-0.0039	-0.0039
5	-0.0015	-0.0027	-0.0026	0.0079
mean	0.0014	0.0070	-0.0044	0.0041
halfwidth	0.0035	0.0112	0.0014	0.0075

- The confidence interval of difference of class 1 waiting time does not contain zero
- Although the difference with respect to the theoretical waiting time is of the order of 10^{-1}

Verification for low traffic intensity

- Parameter setting : $p_{11} = 0.5$, $p_{12} = 0.4$ $\mu_1 = 15$, $\mu_2 = 16$, $\lambda_1 = 1$ and $\lambda_2 = 1$

Replication	Differences at stage 1		Differences at stage 2	
	Class 1	Class 2	Class 1	Class 2
1	0.00026	0.00009	0.00039	0.0002
2	-0.00029	-0.00053	-0.00045	-0.0001
3	0.00043	0.00067	0.00025	0.0010
4	0.00053	0.00071	0.00002	-0.0039
5	0.0002	-0.00042	-0.00038	0.0079
mean	0.000229	.000105	-0.000034	0.000147
halfwidth	0.000400	0.000729	0.000469	0.000216

Verification for moderate traffic intensity

- Parameter setting : $p_{11} = 0.5$, $p_{12} = 0.4$ $\mu_1 = 15$, $\mu_2 = 16$, $\lambda_1 = 4$ and $\lambda_2 = 1$

Replication	Differences at stage 1		Differences at stage 2	
	Class 1	Class 2	Class 1	Class 2
1	-0.0015	0.0002	0.0004	-0.0003
2	0.0012	-0.0006	0.0011	0.0009
3	0.0009	0.0013	0.0002	-0.0003
4	-0.0009	-0.0009	-0.0001	-0.0012
5	0.0014	0.0016	0.0016	0.0019
mean	0.00019	0.00056	0.00066	0.00020
halfwidth	0.00162	0.00124	0.00087	0.00147

Verification for heavy traffic intensity

- Simulation run length is 100000 with a warmup period of 20000
- Parameter setting : $p_{11} = 0.8$, $p_{12} = 0.9$, $\mu_1 = 15$, $\mu_2 = 14.5$, $\lambda_1 = 6$ and $\lambda_2 = 8$

Replication	Differences at Stage 1		Differences at Stage 2	
	Class 1	Class 2	Class 1	Class 2
1	-0.0033	-0.0265	-0.0207	-0.0894
2	-0.0053	-0.0133	-0.0403	-0.2671
3	-0.0046	-0.0193	-0.0020	0.0547
4	0.0008	0.0048	0.0175	0.2585
5	-0.0106	-0.0448	-0.0252	-0.1466
mean	-0.0046	-.0198	-0.0141	-0.0380
halfwidth	0.0050	0.0225	0.0277	0.2510

Application of decompositon result

- Consider the linear waiting time cost minimization problem as follows:

$$\textbf{P1:} \quad \min_{\mathcal{F}} \quad c_1 \bar{W}_1 + c_2 \bar{W}_2$$

- Relative dynamic priority is *complete* ([Gupta et al., 2014])
  - $\bar{W}_1$ and $\bar{W}_2$ are mean waiting time of class 1 and class 2 respectively
  - c_1 and c_2 be the cost associated with class 1 and class 2 respectively
- Thus can be reduced to following problem:

$$\min_{p_{11},\ p_{12}} \quad c_1 \bar{W}_1 + c_2 \bar{W}_2$$

Application of decompositon result

- By using the decomposition result:

$$\bar{W}_1(p_{11}, p_{12}) = \bar{W}_{11}(p_{11}) + \bar{W}_{12}(p_{12})$$

$$\bar{W}_2(p_{11}, p_{12}) = \bar{W}_{21}(p_{11}) + \bar{W}_{22}(p_{12})$$

  $\bar{W}_{ij}$ for $i, j = 1, 2$ is mean waiting time for class i in stage j

- The optimization problem can be decomposed into simpler optimization problem as shown:

$$\min_{p_{11}} \quad c_1 \bar{W}_{11}(p_{11}) + c_2 \bar{W}_{21}(p_{11}) +$$

$$\min_{p_{12}} \quad c_1 \bar{W}_{12}(p_{12}) + c_2 \bar{W}_{22}(p_{12})$$

Departure process of relative priority queue

Conjecture

A two class single stage exponential queueing system has Poisson departure process when relative dynamic priority is implemented across classes

- Various graphical tests have performed to verify the conjecture

Verification through histograms

- Histograms of inter-departure time at first stage for both classes are plotted
- Histograms are plotted for a bin width of 0.01 time units

Verification through histograms

- Parameter setting:
  $\lambda_1 = 6,\ \lambda_2 = 8,\ \mu_1 = 15,\ \mu_2 = 16,\ p_{11} = 0.8,\ \ p_{12} = 0.9$

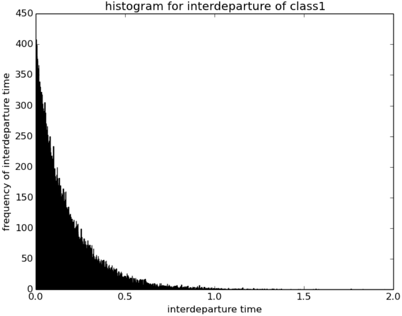

Figure: Estimated density of interdeparture time for class1 customers

Verification through histograms

- Parameter setting:
  $\lambda_1 = 6,\ \lambda_2 = 8,\ \mu_1 = 15,\ \mu_2 = 16,\ p_{11} = 0.8,\ \ p_{12} = 0.9$

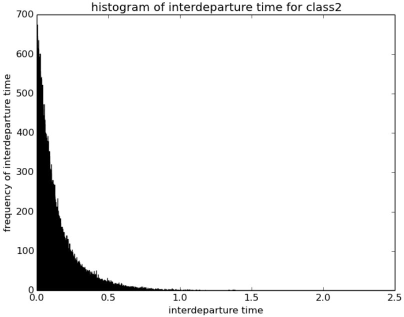

Figure: Estimated density of interdeparture time for class2 customers

Verification through Q-Q plot

- In a Q-Q plot, we ordered the data points are ploted against the theoretical data quantiles of the distribution
- The points in the Q-Q plot will approximately lie on the line $y = x$, if distributions being compared are similar (see[Chambers et al.,])

Verification through Q-Q plot

- Parameter setting: $\lambda_1 = 1.0,\ \lambda_2 = 2.0,\ \mu_1 = 15.0,\ \mu_2 = 16.0,\ p_{11} = 0.6,\ p_{12} = 0.8$ for both class 1 and class 2 customers

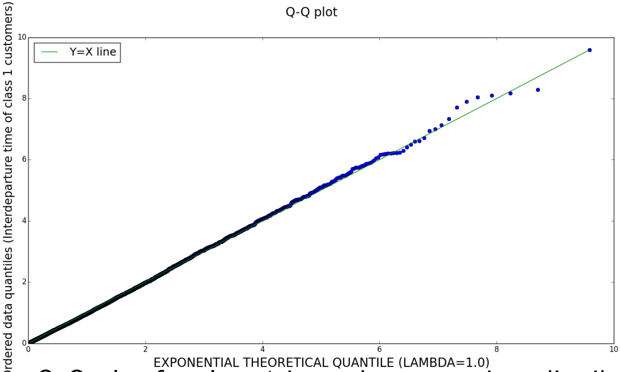

Figure: Q-Q plot for class 1 inter-departure time distribution.

Verification through Q-Q plot

- Parameter setting: $\lambda_1 = 1.0,\ \lambda_2 = 2.0,\ \mu_1 = 15.0,\ \mu_2 = 16.0,\ p_{11} = 0.6,\ p_{12} = 0.8$ for both class 1 and class 2 customers

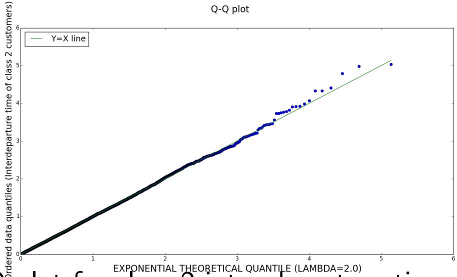

Figure: Q-Q plot for class 2 inter-departure time distribution.

References I

Chambers, J. M., Cleveland, W. S., Kleiner, B., and Tukey, P. A.
Graphical methods for data analysis. 1983.
Wadsworth & Brooks, Pacific Grove, California, pages 82–87.

Goldberg, H. M. (1977).
Analysis of the earliest due date scheduling rule in queueing systems.
Mathematics of Operations Research, 2(2):145–154.

References II

Gupta, M. K., Hemachandra, N., and Venkateswaran, J. (2014).
On completeness and equivalence of some dynamic priority schemes.
Technical report, IIT Bombay.
http://www.ieor.iitb.ac.in/files/faculty/nh/completeTR.pdf.

Haviv, M. and van der Wal, J. (2007).
Waiting times in queues with relative priorities.
Operations Research Letters, 35:591 – 594.

References III

Matloff, N. (2008).
Introduction to discrete-event simulation and the simpy language.
Davis, CA. Dept of Computer Science. University of California at Davis. Retrieved on August, 2:2009.

Mumtaz, S. and Rodriguez, J. (2014).
Smart Device to Smart Device Communication.
Springer.

Thank You