McMaster University

A Stable Mean Value Analysis Algorithm for Closed Systems with Load-dependent Queues

Lei Zhang and Douglas G. Down

Computing and Software
McMaster University
Hamilton, ON, Canada

October 28, 2016

McMaster University

Motivation

Challenges for adopting Mean Value Analysis (MVA) [1] for Flow-Equivalent Server (FES) [2]:

- Load-dependent MVA algorithm suffers from numerical instability issues:
  - The computations of the state probabilities at each node.
  - Negative values of mean performance metrics can be produced.
- The literature is lacking efficient solutions for the numerical instability of MVA.

Contributions

- The Stable MVA (SMVA) algorithm, which is an efficient approximate solution for closed networks with load-dependent queues.
- An extended multi-class model used to determine class-level mean performance metrics.

The Conditional MVA Algorithm

Casale [3] introduced the Conditional MVA (CMVA) algorithm:

- Avoid the computations of the state probabilities.
- Pros - exact solution.
- Cons - time and space complexities - $O(MN^2)$, complexities for the original MVA grow as $O(MN)$.

Seidmann's Approximation

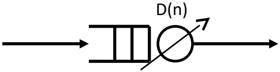
Seidmann's approximation [4]:

- Applicable for multi-server queues in which all servers are load independent.
- Pros - time and space complexities - $O(MN)$.
- Cons - an approximate solution,
  assumption can be easily violated.

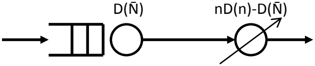
Basic Idea

The basic idea of SMVA is to replace a load-dependent queue with two tandem servers:

- The first one is a load-independent queue.
- The second node is a load-dependent delay centre.



(a) A generic load-dependent queue



(b) Approximating queues in SMVA

Service Demand - 1

The first load-independent queue with service demand:

$$D^q(n) = D(\bar{N}).$$

Assumption: There exists a finite $\bar{N}$ such that $D(n) = D(\bar{N})$ for all $n \geq \bar{N}$. This assumption is reasonable for many systems, in particular when $D(n)$ becomes sufficiently close to $D(\bar{N})$.

Service Demand - 2

The second load-dependent delay center with service demand:

$$D^d(n) = \begin{cases} nD(n) - D(\bar{N}), & \text{if } n < \bar{N} \\ (\bar{N} - 1)D(\bar{N}), & \text{if } n \geq \bar{N}. \end{cases}$$

Assumption: To make sure the service demands of the delay centre are positive, we assume that $nD(n) \geq D(\bar{N})$, for $n < \bar{N}$.

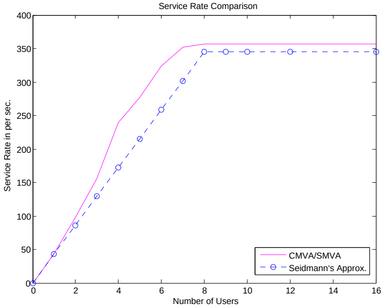
Discussion

- Although the delay centre is load dependent, there is no need to calculate its state probabilities because it does not have a queue.
- Under light load, the two tandem servers behave as a server which has service demand $D(n)$.
- If there are jobs waiting in the first queue, the time spent by a job in the approximating node is dominated by the time spent at the first queue. The node behaves as a server which has service demand $D(\bar{N})$.

Experimental Settings

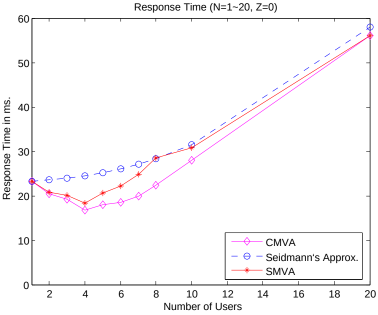
- Testbed on a quad-core system.
- We employ TPC-W [5] to generate the workload.
- The results (the mean response time and the throughput) of SMVA are compared to those of CMVA and Seidmann's approximation.

Service rate curves:

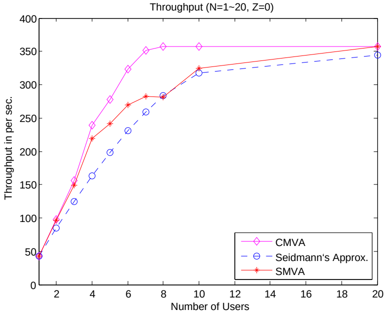


McMaster University

Experimental Results - 1



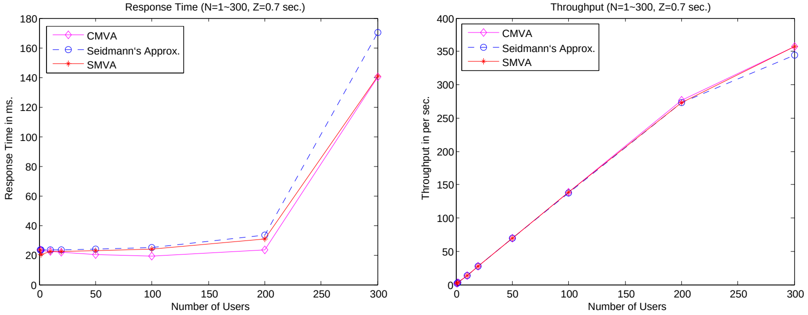
(a) Response time



(b) Throughput

Figure: Comparison with $Z=0$

Experimental Results - 2

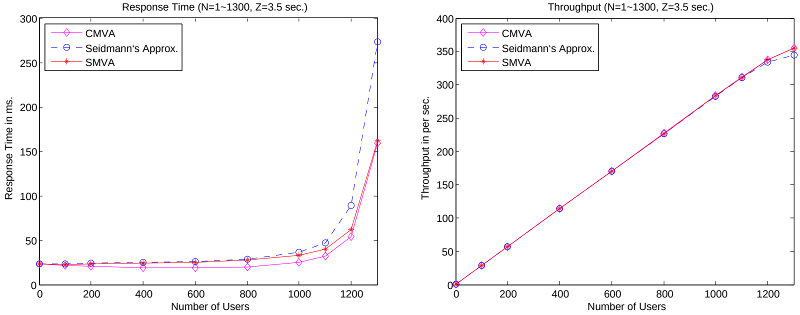


(a) Response time

(b) Throughput

Figure: Comparison with $Z=0.7$ seconds

Experimental Results - 3



(a) Response time (b) Throughput

Figure: Comparison with Z=3.5 seconds

Discussion - 1

Compared to the CMVA algorithm and Seidmann's approximation, the SMVA algorithm has two advantages:

- The time and space complexities of SMVA are a significant improvement over CMVA - $O(MN)$ versus $O(MN^2)$.
- Compared to Seidmann's approximation, SMVA is better able to handle cases when the service demands of a load-dependent node do not have a linear relationship.

Discussion - 2

Additional observations about SMVA:

- SMVA works as well as the CMVA algorithm for both light and heavy loads.
- The errors of SMVA increase when the system is under intermediate loads.

Discussion - 3

Two reasons leading to the error of SMVA under intermediate loads:

- It assumes that all the jobs are being processed at the server when n jobs are in the system.
- The delay centre can “delay” a job too long, in particular when the time that a job spends at the first queue does not dominate the total time.

McMaster University

Multi-class Model

Consider that there are C classes of transactions, where the job population vector is given by $\vec{N} = (n_1, n_2, \ldots, n_C)$, so that $N \equiv \sum_{c=1}^{C} n_c$.

The service demand of class c is given by

$$D_c^q(n) = D_c(\bar{N}), \ n = 1 \ldots N.$$

The service demand at the delay centre becomes

$$D_c^d(n) = \begin{cases} nD_c(n) - D_c(\bar{N}), & n < \bar{N} \\ (\bar{N} - 1)D_c(\bar{N}), & n \geq \bar{N}. \end{cases}$$

McMaster University

Conclusions and Future Work

Conclusions:

- SMVA - an efficient approximation to address the numerical instability of the load-dependent MVA algorithm for closed queueing networks.

Potential directions for future work:

- Improve the accuracy under intermediate workload.
- Keep the time and space complexities as $O(MN)$.

References I

[1] M. Reiser and S. S. Lavenberg.
Mean-value analysis of closed multichain queuing networks.
Journal of the ACM, vol. 27, pages 313-322, 1980.

[2] K. M. Chandy, U. Herzog and L. Woo
Parametric analysis of queuing networks.
IBM Journal of Research and Development, vol. 19, pages 36-42, 1975.

[3] G. Casale.
A note on stable flow-equivalent aggregation in closed networks.
Queueing Systems, vol. 60, pages 193-202, 2008.

McMaster University

References II

[4] A. Seidmann, J. Paul and S. Shalev-Oren.
Computerized closed queueing network models of flexible manufacturing systems: A comparative evaluation.
Large Scale Systems, vol. 12, pages 91-107, 1987.

[5] T. Horvath.
TPC-W J2EE Client Implementation.
http://www.cs.virginia.edu/~th8k/downloads, last accessed 24 September 2016.

Thank you very much!