Considerations in workload characterization for Parallel Access Volumes

Thomas Begin
University Pierre & Marie Curie
Paris, France
thomas.begin@lip6.fr

Alexandre Brandwajn
Pallas International Corporation
San Jose, CA, USA
alexb@pallas-international.com

In this paper we consider issues that may arise in the I/O workload characterization for Parallel Access Volumes. Limiting the characterization of I/O device service time to the mean and the standard deviation is sufficient to correctly predict the average I/O time in a realistic model of a single-exposure device. Things are less simple, however, when one considers devices with multiple parallel servers. We use two examples to illustrate the need for a more detailed workload characterization with multiple servers. We also briefly consider methods for matching real-life workload distributions to phase-type distributions, used to enable numerical solutions of queues with multiple servers, and we point out potential issues related to their use in modeling the performance of Parallel Access Volumes.

1 Introduction

In the area of mainframe I/O, the Parallel Access Volume (PAV) feature [IBM1999] allows a heavily used logical device (or volume) to be replaced by a set of volumes serving a queue of I/O requests. We refer to such a set of parallel volumes as multiple exposures. The goal of multiple exposures is to reduce the expected queueing time of I/O requests directed to a heavily used device.

In the case of a device with a single exposure, using common simple assumptions on the arrivals of requests, we need to know only the mean and the standard deviation of the I/O service time to predict the average queueing (IOSQ) time. Indeed, the Pollaczek-Khintchine formula [ALL1990] involves only the first two moments of the service time distribution. With multiple exposures, such a limited knowledge of the I/O service time is not sufficient. This has potential implications on the workload characterization for Capacity Planning with Parallel Access Volumes.

Our main goal of this paper is to bring this issue to the attention of performance specialists working in the area of I/O subsystem performance.

In the next section, we present two examples to illustrate the potential inadequacy of a workload characterization limited to the mean and the standard deviation. In Section 3, we point out the limitations of two approaches that have been proposed in the literature to represent the workload in models used to evaluate the performance of systems with multiple exposures. We also point the shortcomings of most approximations one might be tempted to use in the context of PAVs. Section 4 concludes this paper.

2 Examples of IOSQ time with multiple exposures

Under simple assumptions on the arrival process of I/O requests, the average IOSQ time at a single-exposure I/O device can be obtained using the *M/G/1* queueing model e.g. [BRA1983]. The well-known Pollaczek-Khintchine formula [ALL1990] shows that the average queueing time in an *M/G/1* queue depends only on the mean and the standard deviation of the service time. Thus, with a single-exposure device, it is sufficient to characterize the first two moments of the I/O workload. When one deals with Parallel Access Volumes, one might be tempted to keep this simple characterization of the I/O service time. Unfortunately, the underlying queueing model then becomes the so-called *M/G/c* queue, and in such a system the average queueing time depends on more than just the mean and the standard deviation of the service time.

A couple of simple examples illustrate the fact that one cannot rely on the first two moments of the service time to correctly predict the average IOSQ time in the case of multiple exposures. Our numerical results have been obtained using discrete-event simulation at 95% confidence level.

2.1 A high-level description of a cached I/O with PAVs

As our first example, we consider a high-level abstraction of the read performance of a cached storage controller with PAVs. For a cache hit, assuming a fixed record size, we view the service time as constant, and for a cache miss the information must be read from the underlying physical devices. We consider two sets of parameters with different distributions for the miss service time: uniform and truncated exponential [JAW2004]. Both resulting overall I/O service time distributions have the same mean and coefficient of variation, shown in Table 1, but different higher order properties. Since the coefficient of variation is the ratio of the standard deviation to the mean, the two distributions have the same mean and standard deviation.

Table 1. I/O service time parameters in two storage subsystems.

	Service time distribution with mean of 3.5 and coefficient of variation of 2.48		
	Hit ratio (probability)	Hit service time	Miss service time
Dist. A	0.999	3.34	Truncated exponential mean: 300, max: 1000
Dist. B	0.9	1	Uniform [2,50]

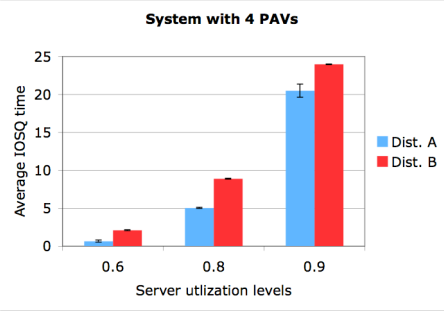

Fig 1a. Results for 4 PAVs.

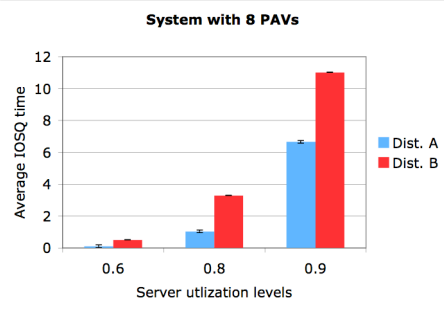

Fig 1b. Results for 8 PAVs.

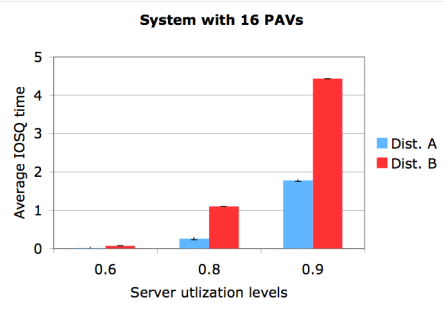

Fig 1c. Results for 16 PAVs.

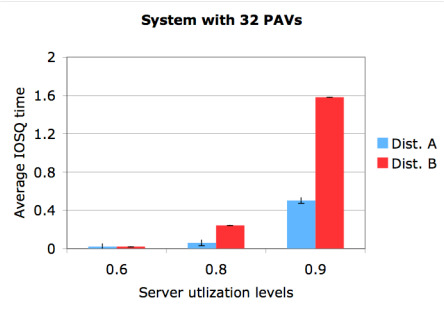

Fig 1d. Results for 32 PAVs.

Fig. 1. Average IOSQ time for a multi-exposure volume for service time distributions from Table 1.

Fig. 1 shows the average IOSQ time for different numbers of PAVs and volume utilization levels of 0.6, 0.8, and 0.9. We observe that the influence of higher properties of the service time distribution can be important. It has been our experience that, as the number of servers increases, this influence tends to become more visible for higher levels of server utilization. Also, higher order properties tend to be more important for service time distribution with higher coefficients of variation. It is interesting to note that, in this example, the relative difference

in the average IOSQ time can be over 100% while the coefficient of variation of the I/O service time is quite moderate (2.48).

We observe that the influence on the average IOSQ time of properties of the service time distribution beyond its mean and standard deviation peaks for some value of the number of PAVs, and then decreases as the number of PAVs increases. Since we see markedly different IOSQ times for two service time distributions with the same means and the same coefficients of variation, it is clear that a PAV model that accounts only for the mean (such as the *M/M/c* queue) cannot be expected to consistently produce correct results.

2.2 A more specific description of a cached I/O with PAVs

As our second example, we consider a cached I/O device with 8 Parallel Access Volumes, and we assume again that all accesses are read requests. The service time for a hit consists of an overhead followed by the record transfer time. Both times are taken to be constants in our example. The service time for a miss comprises an overhead, device orientation (seek plus rotational latency), followed by transfer of data to the cache and to the host. Our goal is to show that differences in higher order properties lead to visible difference in performance even when the distributions are structurally similar.

We present two possible workloads, referred to as Dist. 1 and Dist. 2 (see Table 2 in the Appendix), with different hit ratios, device orientation times, as well as different transfer times, chosen so as to have the same overall average I/O service time and the same standard deviation. As illustrated in Fig. 2a, the resulting average IOSQ times can be quite different even though the first two moments of the I/O service time are identical (mean I/O service time of 1.195 and standard deviation of 2.324). Similarly, the effect of different service time distributions can be observed if one is interested in more detailed I/O performance metric, such as, e.g., the probability that there are 2 or more requests waiting in the queue, as shown in Fig. 2b.

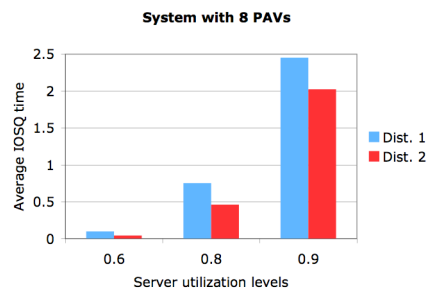

Fig 2a. Average IOSQ time

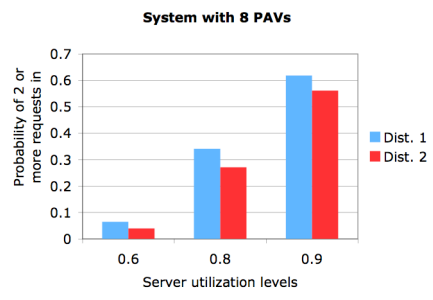

Fig 2b. Steady-state probability of having two or more requests in queue

Fig. 2. Performance of a multi-exposure volume for service time distributions from Table 2.

In the next section, we look at some practical implications of our observations.

3 Practical implications

To the best of our knowledge, there is no simple closed-form solution for the *M/G/c* queue and no equivalent to the Pollaczek-Khintchine formula. Hence, a good approximation to evaluate the average queueing time would certainly be most useful. A number of approximations have been proposed for the *M/G/c* queue [KIM1983, KIM1986, KIM1995, KIM1996, BOX1979, BJO1964, TIJ1981, MIY1986, NOZ1978, WOL1989, SCH1978, SMI2003, MA1995]. Unfortunately, virtually none of them attempts to account for the fact that the average queueing time may depend on more than just the mean and the standard deviation of the service time. Therefore, one must approach the results of such approximations with a fair dose of caution.

Since there is no closed-form solution for the *M/G/c* queueing system, such queues must be solved using discrete-event simulation or numerical approaches, e.g., [NEU1981, LAT1999, SEE1986, BRA2009]. The latter methods use a phase representation of service time distributions [OSO2006, COX1961, ALL1990]. Indeed, any distribution can be represented arbitrarily closely by a distribution of this type. There is an abundant literature on matching real-life distributions using phase-type distributions [BOB2005, OSO2006, ASM1996]. For a good bibliography on this subject, the reader can refer to the work of Osogami and Harchol-Balter [OSO2006].

However, the existing distribution matching techniques do not always produce reliable results when applied to the problem at hand. Indeed, assume that we are able to obtain (either through measurements or through analysis) the distribution of the I/O service time. To predict the expected queueing time with Parallel Access Volumes using a numerical solution method we would first have to represent the I/O service time distribution as a phase-type distribution. Distribution matching methods can be broadly divided into two groups: one that concentrates on correctly matching the first three moments of the real-life distribution e.g. [BOB2005, OSO2006], and one that focuses on approximating the general shape of the distribution e.g. [ASM1996]. It has been our experience that both approaches can lead to significant inaccuracies when their results are used to evaluate the performance of PAVs. This seems to be particularly true when the underlying service time distribution is multi-modal, as is often the case with cached I/O devices.

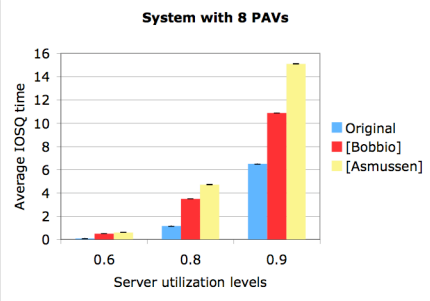

Fig 3a. Average IOSQ time with a distribution from Table 1 (miss service uniform) and approximations [BOB2005, ASM1996]

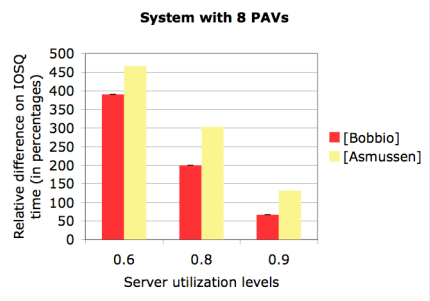

Fig 3b. Relative difference in the average IOSQ time using approximate phase-type distributions [BOB2005, ASM1996]

Fig. 3. Performance of an I/O device with 8 PAVs using a service time distribution from Table 1 and approximations.

This is illustrated in Figure 3 where we show the average IOSQ for a system with 8 PAVs for the high-level description of I/O service time considered in Section 2.1. The miss service time is assumed to be uniformly distributed with the parameter values given in Table 1. In Figure 3a we compare the average IOSQ time obtained using the original I/O service time distribution (labeled “Original”) with that obtained using the phase-type distributions produced by the moment matching method of Bobbio et al. [BOB2005], as well as the phase-type distribution produced by the shape matching method of Asmussen et al. [ASM1996]. Figure 3b shows the corresponding relative difference in the average IOSQ time. We observe that the relative difference can exceed 100%.

4 Conclusions

We have considered some issues that may arise in the context of Parallel Access Volumes used to relieve I/O request queueing. We used two examples to underscore the fact that the I/O workload characterization limited to the mean and the standard deviation, frequently thought of as adequate with single-exposure volumes, is in fact inadequate when one considers multiple parallel servers. This is even more so if one considers a more realistic request arrival process than the Poisson process assumed in the *M/G/c* queue.

We have also shown that one has to approach with caution existing methods for matching real-life workload distributions to phase-type distributions used to enable numerical solutions of queues with multiple servers. For a class of distribution matching methods the issue is that they correctly reproduce only a very limited subset of moments of the real distribution. While a distribution is known to be totally defined by all its moments (if they exist) [ALL1990], we have not been able to determine how many moments, or exactly which characteristics of the real-life workload distribution, are needed to correctly reproduce the expected performance of Parallel Access Volumes or similar systems with parallel servers.

5 References

[ALL1990] Allen, A. O. 1990. Probability, Statistics, and Queueing Theory with Computer Science Applications. Academic Press, 2nd edition.

[ASM1996] Asmussen, S., Nerman, O., and Olsson, M. 1996. Fitting Phase-Type Distributions via the EM Algorithm. Scandinavian Journal of Statistics. 23, 4, 419-441.

[BJO1964] Björklund, M., and Elldin, A. 1964. A practical method of calculation for certain types of com-plex common control systens. Ericsson Technics. 20, 3-75.

[BOB2005] Bobbio, A., Horváth, A., and Telek, M. 2005. Matching three moments with minimal acyclic phase type distributions. Stoch. Models, 21, 2-3, 303-326.

[BOX1979] Boxma, O. J., Cohen, J. W., and Huffels, N. 1979. Approximations of the Mean Waiting Time in an M/G/s Queueing System. Operations Research. 27, 6, 1115-1127.

[BRA1983] Brandwajn, A. 1983. Models of DASD Subsystems with Multiple Access Paths: A Throughput-Driven Approach. IEEE Trans. Comput. 32, 5 (May. 1983),
451-463.

[BRA2009] Brandwajn, A., and Begin T. 2009. Preliminary Results on a Simple Approach to G/G/c-like Queues. In Proceedings: ASMTA 2009, Madrid, Spain.

[COX1961] Cox, D. R., and Smith, W. L. 1961. Queues. John Wiley, New York.

[IBM1999] IBM Enterprise Storage Server, IBM Publication SG24-5665, 1999.

[JAW2004] Jawitz, J.W. 2004. Moments of truncated continuous univariate distributions. Advances in Water Resources. 27, 3, 269-281.

[KIM1983] Kimura, T. 1983. Diffusion Approximation for an M/G/m Queue. Operations Research. 31, 2, 304-321.

[KIM1986] Kimura, T. 1986. A two-moment approximation for the mean waiting time in the GI/G/s queue. Management Science. 32, 751-763.

[KIM1995] Kimura, T. 1995. Approximations for multi-server queues: System interpolations. Queueing Systems. 17, 3-4, 347-382.

[KIM1996] Kimura, T. 1996. Optimal buffer design of an M/G/s queue with finite capacity. Stochastic Models. 12, 1, 165-180.

[LAT1999] Latouche , G., and Ramaswami, V. 1999. Introduction to Matrix Analytic Methods in Stochastic Modeling. ASA-SIAM.

[MA1995] Ma, B., N. W., and Mark, J. W. 1995. Approximation of the Mean Queue Length of an M/G/c Queueing System. Operations Research. 43, 1, 158-165.

[MIY1986] Miyazawa, M. 1986. Approximation of the Queue-Length Distribution of an M/GI/s Queue by the Basic Equations. Journal of Applied Probability. 23, 2, 443-458.

[NEU1981] Neuts, M. 1981. Matrix-geometric solutions in stochastic models: an algorithmic approach. Johns Hopkins Univ. Press.

[NOZ1978] Nozaki, S. A., and Ross, S. M. 1978. Approximations in Finite-Capacity Multi-Server Queues with Poisson Arrivals. Journal of Applied Probability. 15, 4, 826-834.

[OSO2006] Osogami, T. and Harchol-Balter, M. 2006. Closed form solutions for mapping general distributions to quasi-minimal PH distributions. Performance Evaluation. 63, 6, 524-552.

[SCH1978] Schweitzer, P., and Konheim, A. 1978. Buffer overflow calculations using an infinite-capacity model. Stochastic Processes and Their Applications. 6, 267-276.

[SEE1986] Seelen, L. P. 1986. An Algorithm for Ph/Ph/c Queues. European Journal of the Operations Research Society. 23, 118-127.

[SMI2003] Smith, J. M. 2003. M/G/c/K blocking probability models and system performance. Performance Evaluation. 52, 4, 237-267.

[TIJ1981] Tijms, H. C., Van Hoorn, M. H., and Federgruen, A. 1981. Approximations for the Steady-State Probabilities in the M/G/c Queue. Advances in Applied Probability. 13, 1, 186-206.

[WOL1989] Wolff, R. 1989. Stochastic Modeling and the Theory of Queues. Prentice-Hall, New Jersey.

6 Appendix

Table 2. Two service time distributions with the same first two moments (used in our second example)

	Service time distribution with mean of 1.195 and coefficient of variation of 1.945						
	Hit ratio	Hit overhead	Hit transfer	Miss overhead	Miss seek time	Miss rotational delay	Miss transfer time
Dist. 1	0.9	0.1	0.4	0.6	Exponential of rate: 2.0 truncated at: 4.0	Uniform [0, 11.11] (i.e. 5400 RPM)	0.8
Dist. 2	0.9944	0.642	0.4	1.765	Exponential of rate: 9.3×10^{-3}, truncated at: 51.8	Uniform [0 , 4.0] (i.e. 15 000 RPM)	0.8