Send Orders for Reprints to reprints@benthamscience.ae

324 *The Open Cybernetics & Systemics Journal,* 2015, *9,* 324-328

Open Access

Performance Analysis of RAID in Different Workload

Zhang Dule*, Ji Xiaoyun, He Miao and Zhu Huaijie

China Petroleum Pipleline Engineering Corporation, Langfang, Hebei, 065000, P.R. China

Abstract: A performance evaluation model is built for the RAID system with queuing network. With MVA method we develop, validate and apply an analytic performance model for disks arrays configured as a RAID 5. The results show that the model can basically express the trend of the performance with different load.

Keywords: Analytical performance model, array cache, disk array, I/O performance evaluation, parallel I/O, RAID.

1. INTRODUCTION

The disk array which has the characteristics of large storage capacity and high performance is the typical secondary storage. With the increasing importance of disk arrays, models for evaluating their performance have become increasingly important. Unfortunately, the actual system is very complex, the performance is affected by many factors such as hard disk performance, RAID level, different load types, RAID controller etc. Thus, defining an analytic performance model proved far more difficult than expected.

In this paper, we develop, validate and apply an analytic performance model for disks arrays configured as a RAID 5. In order to the simplicity and computational efficiency, the performance model uses the average value for fork-join networks analysis technology. Another advantage is that the predictions of model are right verified in a real system, the HP Smart Array P410i. As verified in a real array, the predicted values under the tested workloads are, on average, accurate within 3.32% of the real values. Hence, our model is simple and intuitive.

2. RELATED WORK

Response time and throughput are the key performance metrics of disk array. Response time is the total amount of time taken to respond by an I/O request at the disk array. Throughput is the rate at which the disk array can service Input/Output operations, it is usually measured in Input/Output operations per second (IOPS).

Research on the model of RAID performance can be traced back to last century 70's. Simulation techniques and analytic techniques are the most common method built Performance models. Papers [1] mainly summarizes four primary categories: (1) simulation method; (2) analytic models that deal with fork-join networks and neglect queuing effects; (3) analytic models that deal with fork-join networks and neglect its synchronization; and (4) restricted queuing

E-mail: cppe_zhangdl@cnpc.com.cn

models that the modeling technique is based on fork-join synchronization using the MVA techniques.

The following lists analytical models of disk arrays that are published in recent years. Paper [2] describes an object RAID system by using Object-based storage Devices and presents queuing models with cache to analyze the system; Paper [3] presents a model of I/O service time and the best performance; Paper [4] presents an analytical model that incorporates the factors of the real system, such as array caches, array controllers, disk controllers, and so on; Paper [5] presents a performance evaluation model that is built for the RAID software system with queuing network. Hence, the performance models for disk arrays in the papers are almost very complicated.

In this paper, we use the top-down design method to modeling RAID system. We analyze and study the principle and process of the RAID system, and identify the key components, and explicitly modeling them. The performance model will allow us to compute the entire system's throughput and response time by using the MVA technique.

3. PERFORMANCE MODEL

A real disk array is composed of hardware and operating system, the hardware of the disk array is made up of several parts, for example disk caches with their own internal caches, array caches, host interfaces, array controllers, disk interface, disk (SAS II) controllers, internal buses, etc. So, it is a very complex. In order to develop an analytic performance model, let us identify the key components of the system. Let us analyze and study how the disk array services its workload. All I/O requests from the client are first submitted to the array controller. A read request is straight away returned from the array cache if the array cache is hit in; otherwise it is submitted to the disks. A write request is first written into the array cache and later submitted to the disks. Hence, All requests are first processed by the array controllers and submitted to the array cache and array disks. Based on this analysis, three components including the array controller, the array cache and the array disks are identified as the key components, so the model is made up three nodes.

1874-110X/15

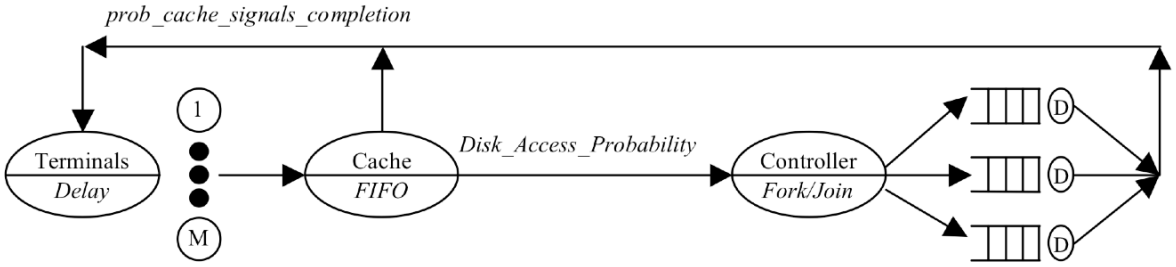

Fig. (1). Queuing network model of disk arrays with cache.

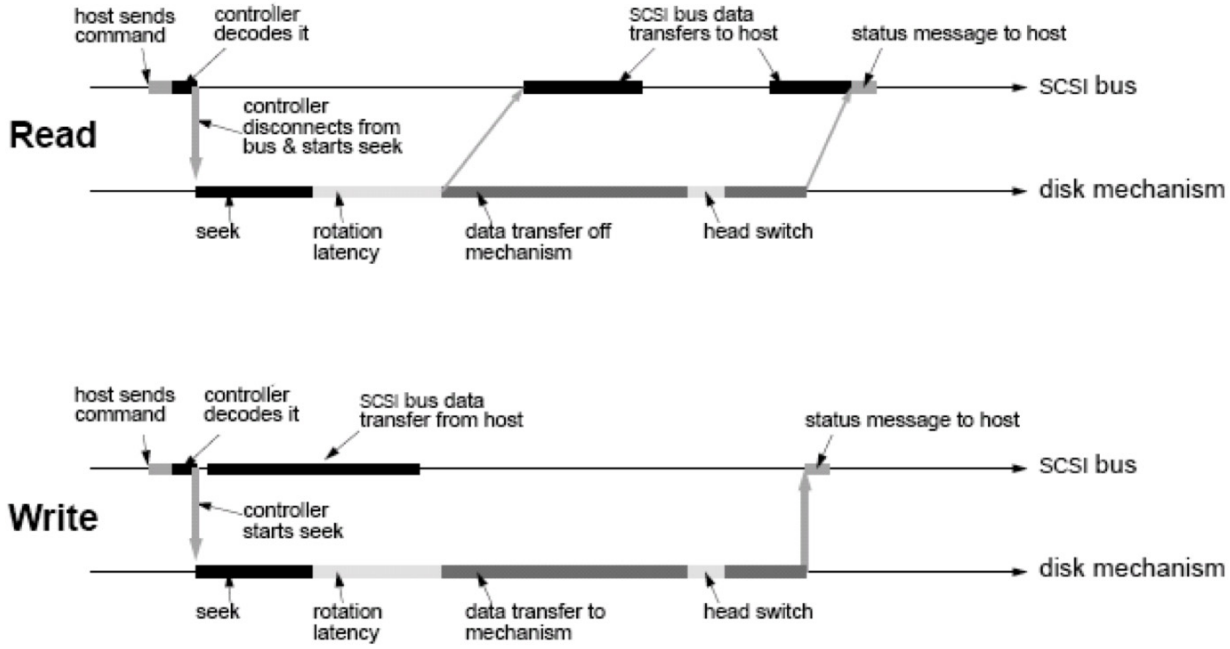

Fig. (2). Overlap of bus phases and mechanism.

We abstracts the array cache into a single queuing server, and abstracts the array controller into a fork-join queue, and abstracts each disk into a queuing server.

Fig. (1) shows a fork-join network model of disk arrays with cache that models with M jobs that cycle between the M terminals and the disk-array.

3.1. CPU Service Time

Based on equipment using single CPU and pressure test, CPU node is abstracted into FCFS queuing node. The CPU service time for all tasks is a constant and is given by T_{CPU_Delay}= constant.

3.2. Cache Service Time

In a real system, cache is a smaller, faster memory that stores data for improving overall system performance, the access time that an I/O request can be served by the array cache is faster than served directly by a set of disk drivers. Hence, the more I/O requests hit the array cache, the faster the disk array performs. The Cache service time is the total amount of time which the data are transferred between the array cache and the client, it is given by [6]:

$$ST_{buf}(b) = \frac{b}{V_{RAM}}$$

3.3. Disks Service Time

The disks are a data storage device and are one of the most important parts in the disk array, it has the characteristics of slow speed and high reliability, it is used for storing and retrieving data. Thus, the disk service time is the main constitution of the response time, it is necessary to get the predictions of disk service time. Paper [7] falls into the SAS disk drive model and working principle, it is given by Fig. (2).

The disks service time (T_s) is the sum of disk positioning time(T_p) [4], transfer time(T_t), and parallelism overhead for N disks (To) [8].

$T_s=T_p+T_t+T_o$

Where, the Variable T_s, T_p, T_t, T_o is:

$$T_t(b) = \frac{b}{V_{disk}}$$

$$T_p = \begin{cases} a + \dfrac{b}{\sqrt{1 + disk\_queue}} & disk\_queue > 3 \\ c + d * disk\_queue & disk\_queue \le 3 \end{cases}$$

where the constants are a =3.53 ms, b =8.81 ms, c =-2.73ms and d =3.68ms.

$$T_O = (0.5 - \frac{1}{n+1}) \times S$$

3.4. Performance Model

According to the analysis result, the performance model used by the MVA technique is expressed as:

$R(m)=T_{wait}(m)+T_{service}(m)$

$T_{wait}(m)=R(m-1)$

$R(0)=0$

$R(m)=m*T_{service}(1)+(m-1)*T_{service}(2)+\ldots\ldots+T_{service}(m)$

$=\sum_{i=1}^{m}(m-i+1)\times T_{service}(i)$

$T\_service(m)=T_{CPU}(m)+T_{cache}(m)+P(m)*T_{disk}(m)$

Where the variable *m* means number of I/O workload streams, it varies from 1 to M.

The *R(m)* means response time that I/O request *m* is serviced by the disk array.

The *P(m)* means the probability that the request *m* can be served directly by a set of disk drivers, it depends on the cache size of array disk, and on the I/O workload and so on, the predictions of *P(m)* is very difficult.

$$P(m) = 1\text{-}hit\_probability = \begin{cases} 1 & \text{served from a disk directly} \\ 0 & otherwise \end{cases}$$

The $T_{disk}(m)$ means the disks service time, under the same workload, the same the disk services is, namely, $T_{disk}(m)=T_{disk}(m-1)=T_{disk}$, the response time of the request m is given by

$$R(m) = \frac{m\times(m+1)}{2} \times T_{CPU} + \frac{m\times(m+1)}{2} \times \frac{B_{Size}}{V_{Cache}} + \frac{(\sum_{i=1}^{m} P_i + 1) \times \sum_{i=1}^{m} P_i}{2} \times T^{m}_{Disk}$$

The response time of the disk arrays R_m is given by:

$$R_m = \frac{R(m)}{m} = \frac{m+1}{2} \times T_{CPU} + \frac{m+1}{2} \times \frac{B_{Size}}{V_{Cache}} + \frac{(\sum_{i=1}^{m} P_i + 1) \times \sum_{i=1}^{m} P_i}{2\times m} \times T^{m}_{Disk}$$

3.5. Read Model

The HP Smart Array P410i uses a read-ahead technology to improve overall system performance by predicting when sequential read requests will follow [9]. With the cache hitting, the read requests is served by the cache, the data is returned from the array cache and the disk array control signals service completion. Otherwise, the request is served by a set of disks and all disks I/Os for the request finish task, the request is finish.

3.6. Write Model

The HP Smart Array P410i uses a write-back caching technology (FBWC) to improve overall system performance

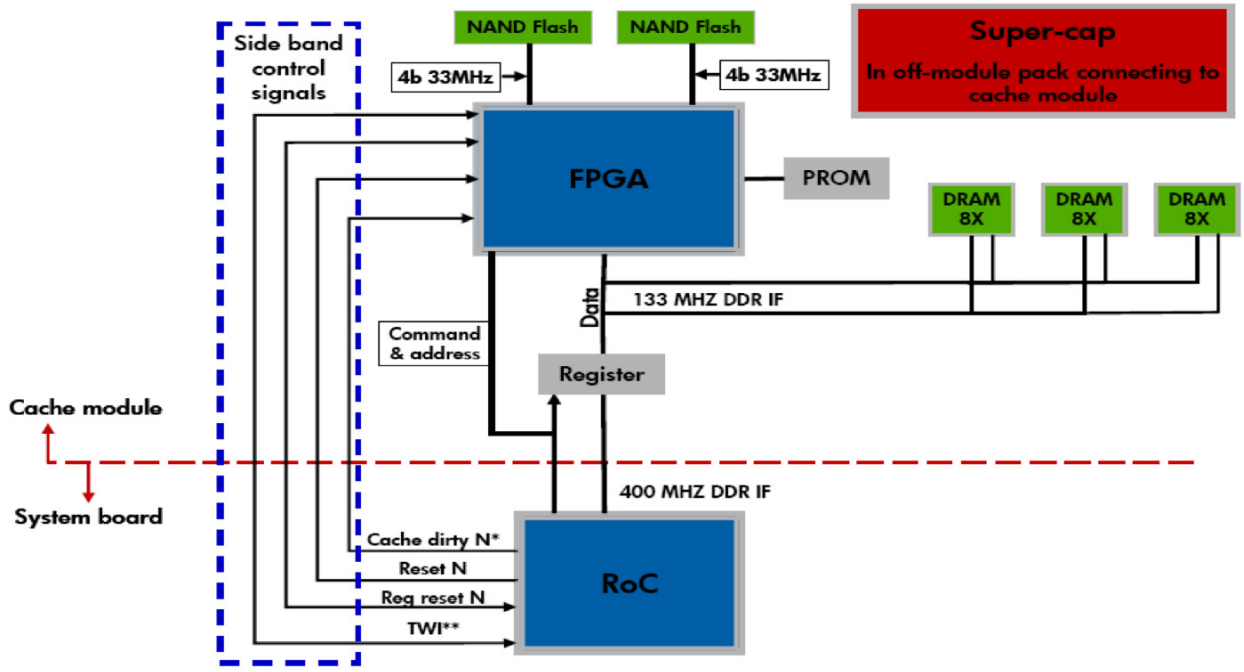

Fig. (3). FBWC block diagram.

by writing directly to the cache, this writing approaches is more complex to implement. With technique of write-back caching, write coalescing, full-stripe writes, and read-modify-write, The HP Smart Array P410i significantly improves performance for I/O operations. FBWC architecture is shown in Fig. (3) [9].

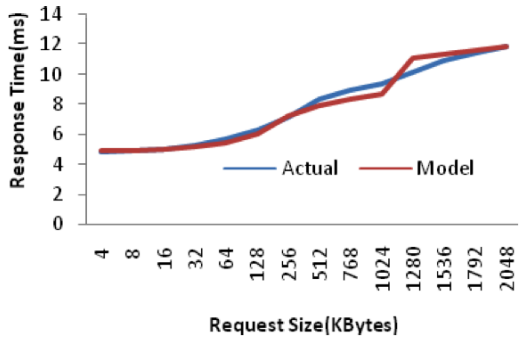

Fig (4). Predicted and Measured Response time for Job=1.

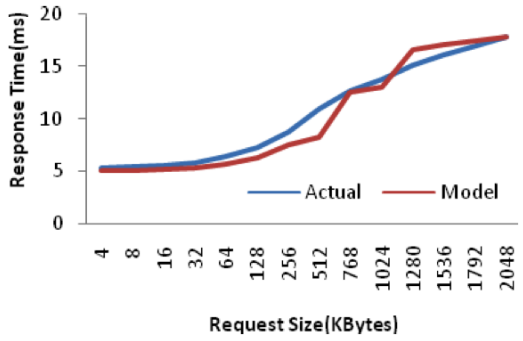

Fig (5). Predicted and Measured Response time for Job=2.

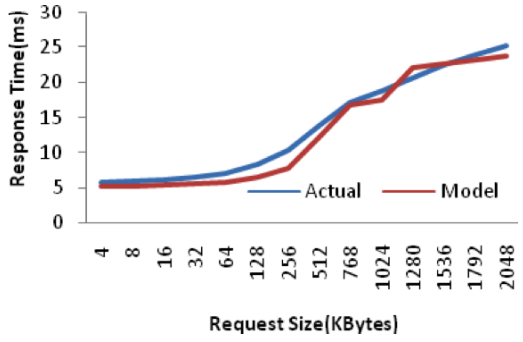

Fig (6). Predicted and Measured Response time for Job=3.

4. EMPIRICAL VALIDATION

To validate the analytic model developed in the previous section, we compare the predictions of analytic performance model against the simulation results over the range of system and workload parameters investigated. The disk model is based upon the HP Smart Array P410i.

We use HP DL380G5 server as a client. A HP DL380G5 server with four 3.2GHz Intel E4250 processor and 16GB main memory, running the Microsoft Windows Server 2008 operating system, is used to generated the workloads and access the HP Smart Array P410i.

A HP Smart Array P410i array with 3.5inch 300GB SASII disks, one controller, and 256 MB of cache. HP Smart Array P410i containing 5 disks. The LUN is configured at RAID 5 and uses a stripe unit size of 256 KB.

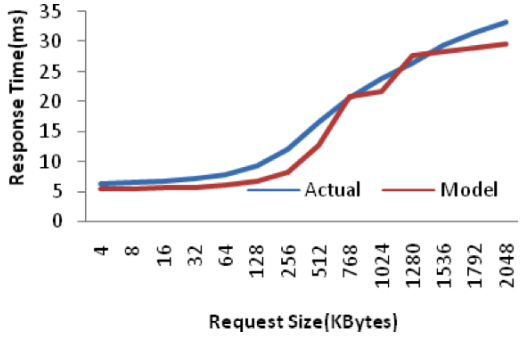

Fig (7). Predicted and Measured Response time for Job=4.

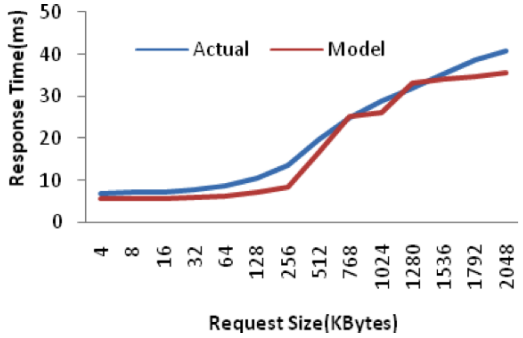

Fig (8). Predicted and Measured Response time for Job=5.

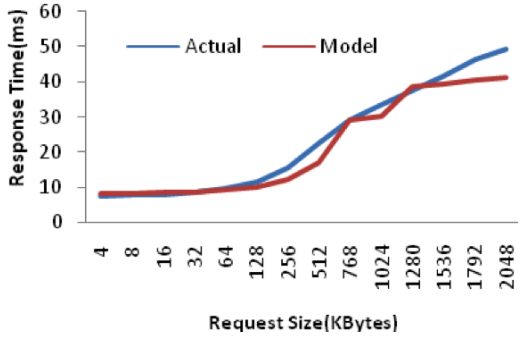

Fig (9). Predicted and Measured Response time for Job=6.

We use synthetic workloads with I/O request sizes ranging from 4 KB to 2048 KB. The number of jobs generating requests range from 1 to 8. Iometer is selected for the simulation tool.

Iometer is a storage subsystem benchmark and troubleshooting tool. Iometer is one of the most commonly tools. It can show the CPU utilization rate, the maximum throughput, the response of the disk system. In order to simulate the actual environment, Iometer allows users set different parameters, for example, the access type (such as sequential, ran-

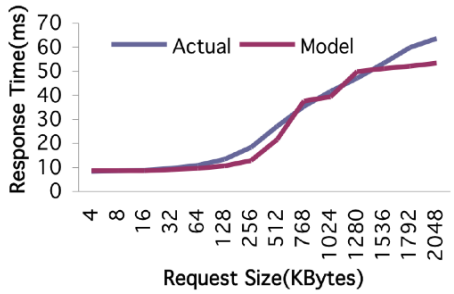

Fig (10). Predicted and Measured Response time for Job=8.

dom), the block size, the queue depth and so on. It has the following characteristics: simple operation, recording the test script, the graphical user interface. Thus, the software can accurately reflect effectively the performance of the storage system.

Figs. (**4-10**) present the predicted response time and the measured response time for read and write workloads for a range of request sizes, multiprogramming levels, think times, and sequentially degrees.

CONCLUSION

The paper presents an analytic performance model that deal with fork-join networks and incorporates the effects of a real array system. We first identify and model the key components, then model the disk array as fork-join network, derive the response time and throughput calculation formula by using the parallel MVA technique.

The validity of the model with a specific disk array(the HP Smart P410i) is proved for a variety of synthetic workloads, request size range from 4KiB to 2048KiB, the number of job range from 1 to 8.

Through estimation, the predictions of model indicate that the real throughput values for P410i under the random and sequential workloads is quite similar. Across all workloads and experiment sets, the predicted values are, on average, within 3.32% of the real values. However, the difference between actual and predicted values under the specific workloads(request size=SUS and number of jobs >2) is much higher (24.23%), These error margins are reasonable since the details of stripe unit and optimizations are not revealed by the manufacturers.

CONFLICT OF INTEREST

The authors confirm that this article content has no conflict of interest.

ACKNOWLEDGEMENTS

Declared none.

REFERENCES

[1] E.K. Lee, and R.H. Katz, "*An Analytic Performance Model of Disk Arrays and its Applications*," Tech. Rep. UCB/CSD91/660, University of California at Berkeley, 1991.

[2] G. Liu, J. Zhou, M. Jiang, and K. Wang, "Performance analysis of object RAID", *Computer Science*, vol. 32, no. 12, pp. 135-137, 2005.

[3] Q. Chen, and J. Zhang, "Analyse of I/O Service time in high performance disk array," *Mini-micro Systems*, vol. 21, no. 3, pp. 235-237, 2000.

[4] E. Varki, A. Merchant, J. Xu, and X. Qiu, "Issues and challenges in the performance analysis of real disk arrays," *IEEE Transactions on Parallel and Distributed Systems*, vol. 15, no. 6, pp. 559-574, 2004.

[5] Y. Zhang, Y. Hu, and T. Jiang, "Performance evaluation model for RAID storage systems based on queuing Network," *Journal of Changchun University of Technology (Natural Science Edition)*, vol. 31, no. 4, pp. 471-475, 2010.

[6] G. Xie, J. Liu, G. Wang, X. Liu, and J. Liu, "Performance evaluation model for Exp-RAID system based on MCQN," *Journal of Computer Research and Development*, vol. 45, no. z1, pp. 207-211, 2008.

[7] C. Ruemmler, and J. Wilkes, "An introduction to disk drive modeling," *Computer*, vol. 27, no. 3, pp. 17-28, 1994.

[8] K. Zhou, J. Zhang, and D. Feng, "Analyzing of I/O response Ttme and throughout of RAID with cache," *Microelectronics & Computer*, vol. 20, no. 8, pp. 66-68, 2003.

[9] HP Smart Array controller technology [EB/OL], http://wenku.baidu.com/link?url=dKzIzWaIbSZwx_qTN7YB9kJCAq5v1JlRU-HW58CNGuTyPZ4xQl0ZS74f6iwdSDS4XHHgIj6casFb2tiGeAH_vXTxdLto1UmRDJ4J1okLTBO

Received: September 16, 2014 Revised: December 23, 2014 Accepted: December 31, 2014