Front. Electr. Electron. Eng. China (2006) 4: 470–474
DOI 10.1007/s11460-006-0090-4

RESEARCH ARTICLE

CUI Bao-jiang, LIU Jun, WANG Gang, LIU Jing

Study on I/O response time bounds of networked storage systems

Abstract In order to predict and improve the performance of networked storage systems, this paper explored the relationship between the system I/O response time and its performance factors by quantitative analytical method. Through analyzing data flow in networked RAID storage system, we established its analytical model utilizing closed queueing networks and studied the performance bounds of the system I/O response time. Experimental results show that the theoretical bounds are found to be in agreement with the actual performance bounds of the networked RAID storage system and reflect the dynamic trend of its actual performance. Furthermore, it concludes that the CPU processing power and cache hit rate of the central storage server are the key factors affecting the I/O response time as the concurrent jobs are lower, while the network bandwidth and cache hit rate of the central storage server become the key factors as the concurrent jobs go higher.

Keywords networked storage, performance modeling, queueing networks, I/O response time

1 Introduction

With the rapid development of network technology, the information is spreading fast. Current storage systems are hard to meet the needs of applications in performance. Therefore, performance study becomes a key issue in the area of networked storage systems.

The current work on performance evaluation of storage systems is mainly focused on direct attached storage (DAS) [1], but the factors affecting the performance of storage systems are more complicated due to the emergence of storage networking. In addition to DAS, it is related to the performance of the network and the host. Although much work has been done on the performance evaluation of networked storage systems, most of them are qualitative [2–4]. The quantitative study is still limited [5]. Therefore, a closed queueing networks model (CQNM) for networked RAID storage systems is proposed for the quantitative performance evaluation in combination with the theory of queueing networks. The CQNM can be used for quick analysis of the I/O response time bounds and key factors affecting the performance of networked storage systems.

The remainder of the paper is organized as follows: Sect. 2 presents the architecture of networked RAID storage system. In Sect. 3, the CQNM is proposed and the analytical model of I/O response time bounds is put forward based on CQNM. Section 4 makes the validation of our analytical model of I/O response time bounds, and the analyses of the performance factors utilizing the proposed model. Conclusions are made in Sect. 5.

2 Architecture of networked RAID storage system

The architecture of networked storage system based on two-level RAID is shown in Fig. 1. The storage devices distributed across the network are mapped into virtual disks at central storage server by IP storage protocol ENBD. These virtual disks are organized into various levels of RAID storage space by software RAID driver at central storage server. Thus distributed storage resources across the network form a virtual space with a single I/O space, available at central storage server.

The I/O requests from applications at central storage server to local virtual RAID storages are passed to remote storage server through network by ENBD client. When the ENBD server receives the packets, it resolves them into

Translated from *Journal on Communications*, 2006, 27(1): 69–74 (in Chinese)

CUI Bao-jiang(✉)
Information Security Centre 126#,
Beijing University of Posts and Telecommunications,
Beijing 100876, China
E-mail: cuibj@bupt.edu.cn

LIU Jun
Department of Information Science and Technology,
Tianjin University of Finance and Economics, Tianjin 300222, China

WANG Gang, LIU Jing
Department of Computer Science, Nankai University,
Tianjin 300071, China

original data and I/O commands. Then I/O commands perform the particular read/write operation on storage devices through the device file system or device driver. Finally the correspondent acknowledgements are fed back to the central storage server.

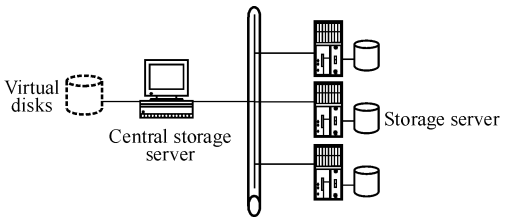

Fig. 1 Architecture of networked RAID storage system

3 Analytical model for networked RAID storage system

3.1 Closed queueing networks model

On the basis of the above data flow analysis in networked RAID storage system, the main components of systems are abstracted as service nodes in queueing networks. The CPUs at central storage server and storage server are abstracted as CPU service nodes. Network interface card (NIC) is abstracted as NIC service node. The network for transferring data between central storage server and storage server is abstracted as network transfer node. The disks attached to storage server are abstracted as disk I/O nodes. Assuming that each node is an M/M/1 queue, Fig. 2 shows the closed queueing networks model (CQNM) for networked RAID storage system.

In Fig. 2, the central storage server at the left is connected to the storage servers at the right through network. CPU service nodes are responsible for processing local applications and data. NIC service nodes send/receive data through NIC. Network transfer node transfers data through network. Disk I/O nodes take charge of the read/write operations.

3.2 Analytical model for performance bounds

This section presents the quantitative analysis for I/O response time bounds of networked RAID storage system based on the CQNM, using balanced job bounds(BJB) [6] method.

Assuming that CQNM consists of arbitrarily connected K nodes, D_i denotes the service demand of the ith node, where $i \in \{1, 2, \cdots, K\}$, $D_{\max} = \max\{D_i, i \in (1, 2, \cdots, K)\}$, $D_{\text{sum}} = \sum_{i=1}^{K} D_i$, $i \in (1, 2, \cdots, K)$, $D_{\text{avg}} = D_{\text{sum}}/K$. Then the I/O response time bound of the CQNM is described as follows:

$$\max (ND_{\max}, D_{\text{sum}}+(N-1)D_{\text{avg}}) <= R(N) \quad (1)$$

According to the principle of data processing in the networked RAID storage system, the analytical model of each service demand D_i is established as the following.

The service demands of CPU service nodes are classified into two cases: central storage server and storage server. The service demand D_{Cm} of CPU service node at central storage server is the time used for processing storage operations including the time required by CPU service node reading/writing virtual disks, the time spent by ENBD client processing data as cache miss and the time cost by TCP/IP processing overhead. Thus we derive the following:

$$D_{\text{Cm}} = T_{\text{mpro}} \frac{S_{\text{m}}}{\text{STU}_{\text{m}}} + (1-P_{\text{m}})\, T_{\text{mp}} \text{IP}_{\text{num}} \quad (2)$$

where, S_{m} is the number of bytes processed by a single job; T_{mpro} is the mean processing time of the single data unit including memory copy and read/write; STU_{m} is the size of stripe unit at central server; P_{m} is cache hit probability; $1-P_{\text{m}}$ represents probability of a job accessing disks at storage servers. Suppose that cache hit probability is zero for write operation because the file size is far more than system cache in the experiment. T_{mp} is the total of the time spent by ENBD client processing data in a single IP packet and the time cost by TCP/IP processing overhead. IP_{num} is the number of IP packets into which a singe job is fragmented, that is:

$$\text{IP}_{\text{num}} = \lceil S_{\text{mR}i}/\text{MSS} \rceil \quad (3)$$

where, MSS is the size of the largest TCP segment in Ethernet. $S_{\text{mR}i}$ represents the number of bytes processed by central server for a RAID i job. For instance, the RAID5 write includes data write and parity write, then:

$$S_{\text{mR5}} = S_{\text{m}}\, \text{STP}_{\text{m}} / (\text{STP}_{\text{m}} - 1) \quad (4)$$

where STP_{m} is the number of stripe units in a stripe at central server. For RAID5 read, the following holds:

$$S_{\text{mR5}} = S_{\text{m}} \quad (5)$$

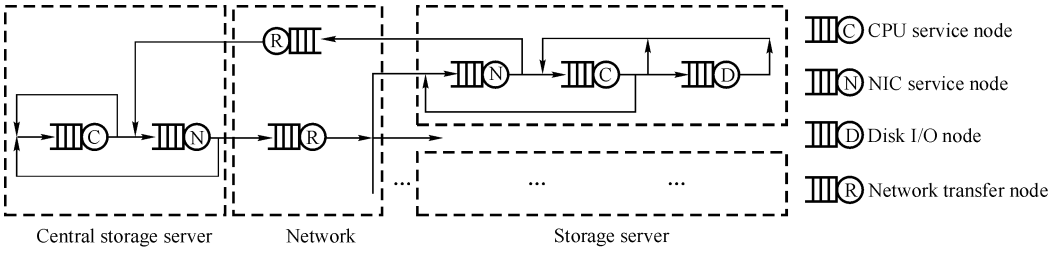

Fig. 2 The CQNM for networked RAID storage system

Similarly, the service demand D_{Csn} of CPU service nodes at storage servers is denoted:

$$D_{\mathrm{Csn}}=\frac{1-P_{\mathrm{m}}}{\mathrm{STP}_{\mathrm{m}}}\left(T_{\mathrm{snpro}}\frac{S_{\mathrm{sn}}}{\mathrm{STU}_{\mathrm{m}}}+T_{\mathrm{snp}}\mathrm{IP}_{\mathrm{snnum}}\right) \tag{6}$$

where, T_{snpro} is the mean time of storage server processing a single data unit; S_{sn} is the number of bytes processed by storage server for a RAID5 job from central server, $S_{\mathrm{sn}}=S_{\mathrm{mR5}}/\mathrm{STP}_{\mathrm{m}}$; T_{snp} is the time of storage server processing data in a single IP packet. $\mathrm{IP}_{\mathrm{snnum}}$ is the number of IP packets processed by storage server for a RAID5 job from central server, denoted by $\mathrm{IP}_{\mathrm{snnum}}=\mathrm{IP}_{\mathrm{num}}/\mathrm{STP}_{\mathrm{m}}$.

The service demand of NIC service node at central storage server is described by the mean time of a single job transmission through NIC, formulated by:

$$D_{\mathrm{Nm}}=(1-P_m)\frac{\mathrm{IP}_{\mathrm{num}}\mathrm{Frames}_{\mathrm{e}}}{\mathrm{TRate}_{\mathrm{e}}} \tag{7}$$

where, $\mathrm{TRate}_{\mathrm{e}}$ is the Ethernet transmission rate and $\mathrm{Frames}_{\mathrm{e}}$ is the size of the Ethernet frame.

Similar to D_{Nm}, the service demand of NIC service node at storage servers is derived:

$$D_{\mathrm{Nsn}}=\frac{1-P_{\mathrm{m}}}{\mathrm{STP}_{\mathrm{m}}}\mathrm{IP}_{\mathrm{snnum}}\frac{\mathrm{Frames}_{\mathrm{e}}}{\mathrm{TRate}_{\mathrm{e}}} \tag{8}$$

The service demand of network transfer node is the mean time cost by each job in the process of network transmission, described by:

$$D_{\mathrm{Net}}=(1-P_m)\frac{S_{\mathrm{mR5}}}{\mathrm{TRate}_{\mathrm{e}}} \tag{9}$$

The service demand of disk I/O nodes is the time required by each job performing disk operations at storage server, denoted by:

$$D_{\mathrm{d}}=\frac{1-P_{\mathrm{m}}}{\mathrm{STP}_{\mathrm{m}}}(1-P_{\mathrm{sn}})(\mathrm{Seek}+\mathrm{Iatency}+\frac{\mathrm{STU}_{\mathrm{m}}}{\mathrm{TRate}_{\mathrm{d}}})\frac{S_{\mathrm{sn}}}{\mathrm{STU}_{\mathrm{m}}} \tag{10}$$

where, P_{sn} is read cache hit probability at storage servers; $1-P_{\mathrm{sn}}$ is the probability of a job accessing disks for read operation. Since the file size is far more than system cache, cache hit probability for write operation can be considered as zero. The parameters Seek and Iatency denote average seek time and average latency, respectively. $\mathrm{TRate}_{\mathrm{d}}$ is the maximal sustained disk transfer rate.

Putting the aforementioned service demands into Eq. (1), the analytical model for the I/O response time bounds of networked RAID storage system is derived.

4 Performance testing and analysis

In this section, the validation of the analytical model is made in the existing experimental environment and the performance factors are discussed further.

The experimental settings of networked RAID storage system based RAID5 are composed of 4 PCs running on Linux 7.3 (its topology is shown in Fig. 1). The configuration of PC is PIII 650 CPU, 64 MB RAM, 10/100 Mbps D-Link DFE530TX NIC, 36 GB IBM DDYS-T36950 SCSI disk (4 MB disk buffer, 4.9 ms average seek time, 2.99 ms average latency, 35 MB/s the maximal sustained disk transfer rate). The stripe size is three. The stripe unit at storage servers is set 16 KB. The data size processed by a single job is 200 MB. The Ethernet bandwidth is 100 Mbps. The size of the largest TCP segment is 1 460 B and the Ethernet frame is 1 518 B. In addition, the CPU mean processing time T_{mpro}, T_{mp}, T_{snpro}, T_{snp} are the mean testing value 0.027 ms, 0.045 ms, 0.033 ms and 0.028 ms, respectively.

Figure 3 shows the write I/O response time comparisons between testing results and the theoretic performance bounds at different concurrent jobs. It indicates that the theoretic bounds given by the analytical model reflect the dynamic trend of the actual system performance and its performance bounds at various concurrent jobs.

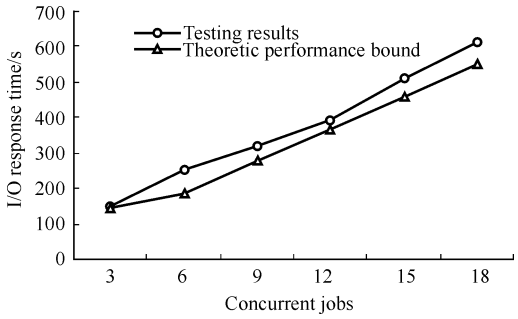

Fig. 3 Comparisons between testing results and the theoretic performance bound

On the basis of the aforementioned validation, further discussion for performance factors of networked RAID storage system is explored in the following section using the analytical model.

Figure 4 shows the relationship between the write I/O response time and the network bandwidth at different concurrent jobs. It is found that network bandwidth is the key factor affecting the I/O response time of the storage system as the concurrent jobs are higher, but it is less effective on the response time as the concurrent jobs are lower.

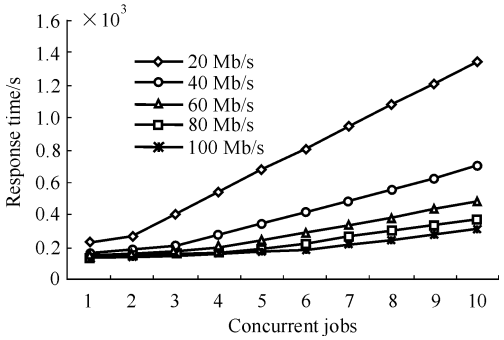

Fig. 4 Relationships between response time, concurrent jobs and network bandwidth

Figure 5 shows the relationship between the write I/O response time and cache hit rate of the central storage server at different concurrent jobs. Unlike the bandwidth, the cache hit rate has an obvious effect on the system response time at various jobs and it has nothing to do with the number of jobs.

Figure 6 shows the relationship between the write I/O response time and CPU processing power of the central storage server at different concurrent jobs. It indicates that system response time decreases with the enhancement of CPU processing power as the concurrent jobs are lower, while CPU processing power exhibits less effect on system response time as the concurrent jobs are higher.

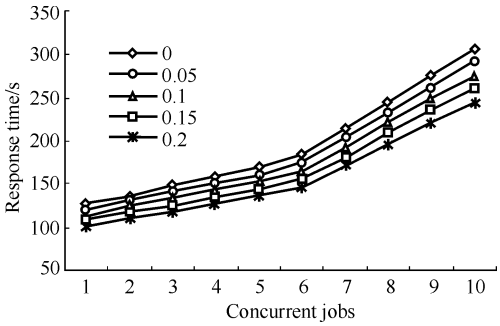

Fig. 5 Relationships between response time, concurrent jobs and cache hit rate

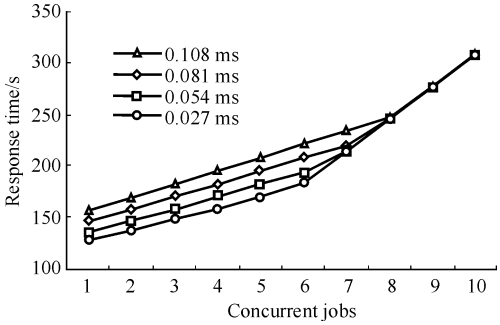

Fig. 6 Relationships between response time, concurrent jobs and central storage server CPU performance

The relationship between the write I/O response time and CPU processing power of the storage server at different concurrent jobs is shown in Fig. 7 and the relationship between the write I/O response time and the maximal sustained disk transfer rate of the storage server at different concurrent jobs is shown in Fig. 8. Unlike the previous performance factors, the CPU processing power and the disk performance of the storage servers demonstrate less effect on I/O response time.

From the aforementioned analysis results, it is found that the CPU processing power of the central server is the key factor affecting the I/O response time as the concurrent jobs are lower. The performance of networked RAID storage system can be improved by increasing the central server CPU processing power. And the network bandwidth displays less effect on the response time and is the non-key factor.

As the concurrent jobs are higher, the network bandwidth becomes the key factor and the central server CPU processing power is the non-key factor. The I/O response time is decreased effectively by enhancing the bandwidth.

Furthermore, the cache hit rate of the central storage server is the key factor at different concurrent jobs. CPU processing power and the disk performance of the storage server have less effect on I/O response time at different concurrent jobs, belonging to non-key factors.

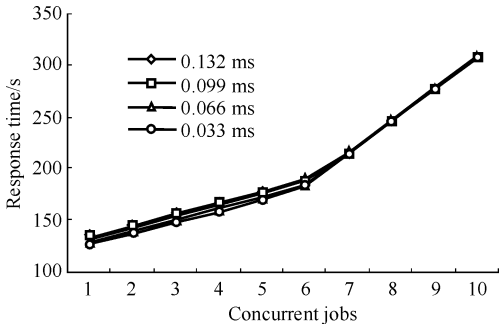

Fig. 7 Relationships between response time, concurrent jobs and storage server CPU performance

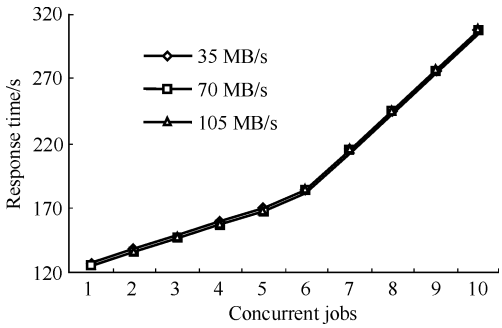

Fig. 8 Relationships between response time, concurrent jobs and the maximal sustained disk transfer rate

The aforementioned analytical model of networked RAID storage system offers an effective approach to system performance evaluation quantitatively. It can be used for optimization of system performance in the preliminary design of storage networking.

5 Conclusions

This paper focuses on I/O response time bounds of networked RAID storage system and performance factors in order to optimize the system performance. The main contribution is to establish the closed queueing networks model (CQNM) of the networked RAID storage system through analyzing the system data flow. The analytical model of I/O

response time bound for the storage system is proposed based on CQNM. Experimental results show that the theoretical bounds are found to be in agreement with the actual performance bounds of the networked RAID storage system, and reflect the dynamic trend of its actual performance and I/O response time bounds. Furthermore, it is concluded that the CPU processing power and cache hit rate of the central storage server are the key factors affecting the I/O response time as the concurrent jobs are lower, while the network bandwidth and cache hit rate of central storage server become the key factors as the concurrent jobs go higher. The CPU processing power and the maximal sustained disk transfer rate of the storage servers are non-key factors; these factors display less effect on system performance.

Acknowledgements This paper was granted by National Natural Science Foundation of China (No. 60273031, No. 90612001), Education Ministry Doctoral Research Foundation of China (No. 20020055021), Technological Development Project Foundation of Tianjin (No. 043800311).

References

1. Barve R., Shriver E., Gibbons P. B., Modeling and optimizing I/O throughput of multiple disks on a bus, Proceedings of Sigmetrics '98 / Performance '98, New York: ACM Press, 1998, 264–275
2. LU Ying-ping, David H. C. D., Performance study of iscsi-based storage subsystems, IEEE Communications Magazine, 2003, 41(8): 76–82
3. He Xu-bin, Beedanagarip, Zhou Dan, Performance evaluation of distributed iSCSI RAID, Proceedings of the 2003 IEEE/ACM International Workshop on Storage Network Architecture and Parallel I/O (SNAPI'03), New Orleans, LA, USA: 2003
4. Wee T. N., Hillyer B. K., Shriver E., Obtaining high performance for storage outsourcing, Proceedings of Conference on File and Storage Technologies (FAST '02), Monterey, California: 2002: 145–158
5. Zhu Yao-long, Zhu Shun-yu, Xiong Hui, Performance analysis and testing of the storage area network, 19th IEEE Symposium on Mass Storage Systems and Technologies, Maryland, USA: 2002
6. Lazowska E. D., Zahorjan J., Graham G. S. et al., Quantitative system performance: computer system analysis using queueing network models, Englewood Cliffs, NJ: Prentice-Hall, 1984