OPTIMIZING PREFETCH IN A RESIDENTIAL GATEWAY WITH NETWORKED STORAGE SYSTEMS

Eunsam Kim, Jonathan C.L. Liu

Department of Computer and Information Science and Engineering
University of Florida
Gainesville, FL, 32611, USA
{eunkim, jcliu}@cise.ufl.edu

ABSTRACT

VOD services require the storage system support with multiple disks. SCSI-based systems are usual choices. However, disks can be attached directly to networks these days. Fiber Channel Arbitrated Loop (FC-AL) is a leading network-accessible storage interface. As a residential service gateway, the FC-AL-based servers enable the stable delivery of high quality video to thousands of clients between external service providers and local clients. We have observed that, even though they performs better than conventional SCSI-based systems, all the disks in FC-AL-based VOD servers utilize only a small portion of their caches to a similar degree due to FC-AL fairness arbitration algorithm. In this paper, we analyze the impact of prefetching on the performance of the FC-AL-based systems according to disk block/cache size. Analysis is also focused on how to find the optimal number of blocks transmitted to the FC-AL from the disk cache per FC-AL arbitration among multiple disks. Finally, simulation results show that a 16-disk system provides over 800 concurrent video streams with quality guarantee, supporting our analytical model.

1. INTRODUCTION

With the recent advances in network technology, the number of high-speed networked homes has increased rapidly and the Internet service providers can offer enhanced services such as VOD services, not limited to high-speed broadband connections. However, it is not still easy to deliver high quality video over the internet due to limited resources and unstable network environments. By expanding home gateways, a residential service gateway enables such enhanced services for thousands of clients in a small area. On the other hand, there is another trend whereby a storage system can be attached directly to a network. One of the leading network-attached storage systems is the Fiber Channel Arbitrated Loop (FC-AL) [1] for Storage Area Networks (SANs). At a low cost, it can stably provide good quality video (e.g. DVD quality MPEG-2 stream) streaming services to thousands of clients in the internal networks. The architecture of our system is based on the FC-AL that directly connects several streaming servers and many disks, as shown in Fig. 1. We have observed that this system performs better than SCSI-based systems, but we can further improve the system performance.

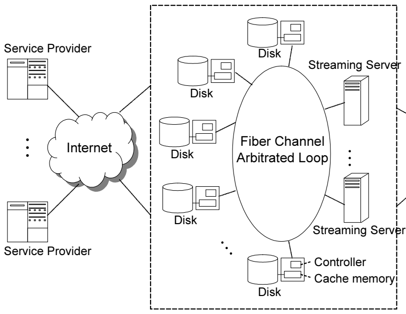

Fig. 1. Overall architecture of the FC-AL-based video servers as a residential service gateway.

In order to support many concurrent streams, storage systems should consist of multiple disks. SCSI-based storage systems are the usual choice. However, the SCSI employs a prioritized arbitration protocol for bus access. When multiple devices compete for bus access, the device with the highest priority always wins the arbitration. Thus, there were significant differences among the latencies according to the disks' priorities [4]. It implies that the lower priority disks can have blocks stay longer in the disk cache by transmitting blocks at a slower rate. Thus, it is difficult for the SCSI unfair bus arbitration to utilize the whole portion of disk cache, resulting in performance degradation. In order to prevent a situation where the lower priority devices may experience starvation, the FC-AL defines a fairness arbitration algorithm that enables all the devices participating in the arbitration to have an equal opportunity to have loop access [1, 4, 6]. Thus, the FC-AL fairness arbitration algorithm reduces the variation of the latencies for loop access among disks. As a result, the fairness algorithm enables all the disks to utilize their caches to a similar degree. Furthermore, the cache utilization of each disk is very low since disks transfer blocks one by one over the high-speed FC-AL and conventional caching approach generates only a very low disk cache hit ratio for VOD services. Therefore, video blocks stay in disk cache for a short time and are rarely hit by the future I/O requests. This means that most of the cache blocks are unused. Therefore, we are motivated by the fact that, unlike in an SCSI-based system, the disk cache can be further utilized in FC-AL-based VOD servers.

For VOD servers, prefetching is a better scheme to utilize the disk cache than conventional caching. Moreover, prefetching can reduce disk seek time significantly by reading blocks contiguously stored together. We will analyze how to determine the maximum number of prefetched blocks in each disk depending on disk block/cache sizes. In addition, in order to guarantee the cache space of each disk for prefetched blocks, we focus on how to find the optimal number of blocks transmitted over the FC-AL per arbitration among multiple disks. Finally, the experimental results support our analytical model, showing that a 16-disk system can provide more than 800 concurrent streams with quality guarantee. The remainder of this paper is organized as follows. Section 2 describes previous and related works on the FC-AL, disk cache, and video servers. Section 3 presents analysis of the prefetching. Section 4 shows the simulation results that validate the analysis. Finally, Section 5 offers our conclusions.

2. RELATED WORKS

Most of research on VOD services have been based on SCSI or distributed networks [7]. The research area on the FC-AL-based video streaming services is quite new [2, 4, 6, 8, 9]. Many papers [3, 5] have focused on further exploiting disk cache but few of them have handled multimedia data requiring large block size. [7] has studied the exploitation of cache memory in a hierarchy, such as host, disk array controller, and disk drives. However, it did not consider the impact of different storage interfaces on overall performance of prefetching schemes. The prefetching schemes at the disk cache level have tried to select blocks that are most likely to be requested in the near future. Thus, they used the locality of reference to effectively predict the disk access pattern for the conventional data, which is quite random. In this paper, we study the prefetching for video data, whose access pattern is sequential.

3. ANALYSIS OF PREFETCHING

Since each disk already has the prefetched blocks in its cache, it can participate in FC-AL arbitration as soon as it receives disk I/O requests. Thus, disks can acquire FC-AL access in decreasing order of the FC-AL priority since all the disks participate in FC-AL arbitration simultaneously at the beginning of each cycle and access window. Therefore, it is possible to determine the time when each disk acquires FC-AL access by calculating the transmission time of the blocks that the disks with higher priorities have sent out over the FC-AL. For the purpose of the analysis of the prefetching, we have assumed that the server schedules the requests with the worst seek and latency times. Although the estimated results may be conservative, the performance trend and analysis still can be clearly demonstrated. Table 1 lists the important symbols used for the analysis.

Table 1. Definition of important symbols

Symbol	Definition
${}_nB$	maximum number of blocks per cycle without prefetching
${}_nB_t$	total number of blocks per cycle with prefetching
${}_nB_p$	number of prefetched blocks per cycle
${}_nB_i$	number of immediate blocks per cycle
${}_nB_{ps}$	number of prefetched blocks per stream
${}_nB_{is}$	number of immediate blocks per stream
${}_nB_c$	disk cache capacity in terms of the number of blocks
${}_nB_{cp}$	total number of prefetched blocks the disk cache holds at the end of each cycle
${}_nB_a$	number of blocks transmitted to FC-AL per arbitration

3.1. Number of concurrent streams with prefetching

Given a block size (B), seek time (T_s), rotational latency (T_r), and data transfer rate (D_{xr}), the effective disk bandwidth (B_{eb}) can be computed with $D_{eb} = (B\ /\ (T_s + T_r + (B/D_{xr})$. To provide on-time delivery of video blocks, each block should be delivered to clients before finishing displaying the previous one. The period between the delivery deadlines of two consecutive blocks is called a cycle time. Thus, the cycle time (C) is determined by $C = B/R_p$ where R_p is playback rate. Then we can derive the maximum number of blocks each disk can read during one cycle without prefetching : ${}_nB = \left\lfloor \frac{D_{eb} * C}{B} \right\rfloor$. Let ${}_nB_i$ denote the number of immediate blocks per cycle (i.e., the non-prefetched blocks that require immediate transmission to the FC-AL within the same cycle in which they are read from the disks). It can be seen that the upper bound on ${}_nB_i$ should be ${}_nB$ when the system does not prefetch any block, and the lower bound should be one. Thus, ${}_nB_i$ can have the following integer values : $1 \leq {}_nB_i \leq {}_nB \;\;\Leftrightarrow\;\; {}_nB_i = \{{}_nB,\ {}_nB - 1,\ \ldots,\ 1\}$.

In the prefetching, the disks read more than one block belonging only to one stream at a time. The first block to be read should be an immediate block, followed by prefetched blocks. Thus, there is neither seek time nor rotational latency for the prefetched blocks since the blocks belonging to one stream are contiguously stored. In addition, in order to avoid the jitters, the total disk access time for the immediate and prefetched blocks should not exceed the cycle time. (1) describes such system design constraint.

$$C \geq {}_nB_i * \left(T_s + T_r + \frac{B}{D_{xr}}\right) + {}_nB_p * \frac{B}{D_{xr}} \quad (1)$$

To obtain the maximum concurrent streams (${}_nB_t = {}_nB_i + {}_nB_p$), we should find all the possible pairs of ${}_nB_i$ and ${}_nB_p$. From (1), it can be seen that ${}_nB_i$ is inversely proportional to ${}_nB_p$. Since there is neither seek time nor rotational latency for the prefetched block, ${}_nB_t$ becomes maximal when ${}_nB_p$ is maximal (${}_nB_i$ is minimal, i.e., one). However, ${}_nB_p$ should be restricted depending on the cache size. Since the streaming servers request one block for each stream per cycle, the number of immediate (${}_nB_{is}$) per stream should be one (${}_nB_{is} = 1$) and prefetched (${}_nB_{ps}$) blocks to read per stream is determined simply by ${}_nB_{ps} = \left\lceil \frac{{}_nB_p}{{}_nB_i} \right\rceil = k$.

Let ${}_nB_{ps}$ be k for the convenience of the representation. If $k > 1$, the disks should prefetch the blocks for the next and subsequent cycles. As k increases, the required cache size increases as a degree of k^2. The reason for this is that the disk cache should hold blocks for not more than k cycles until they are transmitted to the FC-AL. Suppose we should do k-cycle prefetching in i-th cycle. Let r denote the remnant of ${}_nB_p$ divided by ${}_nB_i$. In the case that r is zero, the disk cache should hold $k * {}_nB_i$ blocks for the $(i+1)$ cycle, $(k-1) * {}_nB_i$ blocks for the $(i+2)$ cycle and so on until reaching $(1 * {}_nB_i)$ blocks for the $(i+k)$ cycle. On the other hand, in the case that r is larger than zero, the disk cache should hold ${}_nB_i * \sum_{j=1}^{k-1}(j)$ blocks and r blocks for each next cycle. Thus, ${}_nB_{cp}$ can be computed as follows.

$$r = {}_nB_p \bmod {}_nB_i,$$

$${}_nB_{cp} = \begin{cases} {}_nB_i * \left(\frac{k*(k+1)}{2}\right) & \text{if } r = 0 \\ r * k + {}_nB_i * \left(\frac{k*(k-1)}{2}\right) & \text{if } r \neq 0 \end{cases} \quad (2)$$

Assuming that the outgoing rate of blocks to the FC-AL from the disk cache should be equal to or higher than the incoming rate from the disk, the required cache blocks is largest at the beginning of each cycle. They include the following two kinds of blocks such as ${}_nB_{is} + {}_nB_{ps}$ and ${}_nB_{cp}$. Thus, we can determine the maximum ${}_nB_p$ which satisfies (3). The cache capacity in terms of the number of blocks (${}_nB_c$) is computed as $\lfloor \frac{M}{B} \rfloor$ where M is disk cache size.

$${}_nB_{is} + {}_nB_{ps} + {}_nB_{cp} \leq {}_nB_c \quad (3)$$

3.2. Number of blocks transmitted to the FC-AL per arbitration

The interval between two consecutive loop access times for each disk is determined by calculating the transmission time of the blocks that all the disks have sent out over the FC-AL with one arbitration. During the interval, each disk transmits ${}_nB_a$ blocks. Thus, the outgoing rate of ${}_nB_a$ blocks to the FC-AL from the disk cache (C_{or}) can be computed as (4) where T_{ar}, T_{opn} and T_{cls} denote the arbitration time, opening connection time and closing connection time, respectively. D is the total number of disks and T_{fc} is FC-AL overhead occurring to transmit one block.

$$C_{or} = \frac{{}_nB_a}{\left(T_{ar} + T_{opn} + T_{cls} + {}_nB_a * \left(T_{fc} + \frac{B}{L_{xr}}\right)\right) * D} \quad (4)$$

Likewise, since we can evaluate disk access time of blocks to read per stream, the incoming rate of one immediate block and subsequent prefetched blocks per stream (${}_nB_{is} + {}_nB_{ps}$) to the disk cache from the disk (C_{ir}) can be computed as (5).

$$C_{ir} = \frac{{}_nB_{is} + {}_nB_{ps}}{\left({}_nB_{is} * \left(T_s + T_r + \frac{B}{D_{xr}}\right) + {}_nB_{ps} * \frac{B}{D_{xr}}\right)} \quad (5)$$

To ensure that all the cache blocks are never replaced before being transmitted to FC-AL at least once, the outgoing rate should be equal to or higher than the incoming rate. Otherwise, disk bandwidth should be wasted since the blocks should be read again soon. Therefore, we should obtain maximum ${}_nB_a$ to satisfy (6).

$$C_{ir} \leq C_{or} \quad (6)$$

4. PERFORMANCE RESULTS AND VALIDATION

To validate the analysis, we have performed extensive simulation experiments with multiple disks. The disk model is based on the IBM DeskStar 60GXP. The average data transfer rate, seek time and rotational latency are 30.8 MB/s, 8.5 ms and 4.17 ms, respectively. The FC-AL has a data transfer rate of 200 MB/s. The per-node delay of the interface to forward a frame is 240 nanoseconds and propagation delay between two devices is 4 nanoseconds/meter. On average, the playback rate of each stream is 4 Mbps and the duration is 100 minutes. Each stream is assumed to have constant bit rate and tolerable jitter ratio is 1%. The streaming servers schedule disk I/O requests across the disks as evenly as possible so that the impact of our prefetching scheme interfere with other factors. We employ the average seek time and rotational latency for the analysis so that the average experimental results can be fairly compared with the analytical results.

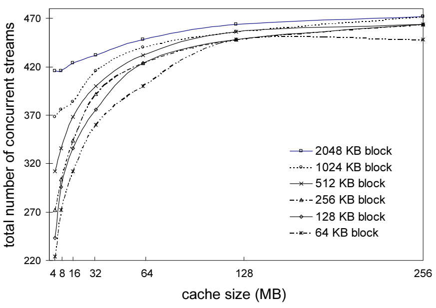

Fig. 2. Block size vs. cache size

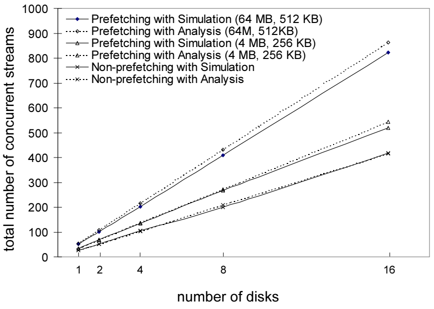

Fig. 3. Prefetching effect

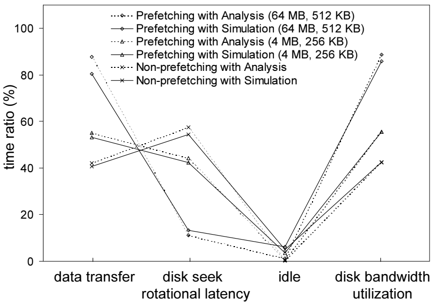

Fig. 4. Disk bandwidth utilization

4.1. Impact of disk block/cache sizes

While varying block size from 64 KB to 2 MB, the impact of the cache size from 4 MB to 256 MB with prefetching was examined in dual loop and eight disk configuration. In Fig. 2, at around a 64 MB cache for most of block sizes, the number of concurrent streams starts to be saturated. Before reaching the saturation point, cache size significantly affects the system performance. When a 64 KB block is used, we obtains a 78.5% performance improvement. It is also shown that the impact of cache size decreases as the block size increases. It is because the ratio of data transfer time is already high compared to that of disk seek time since large data contiguously stored can be read without seeking operations. However, as a larger block is employed, both streaming servers and clients should have larger buffers, and the startup latency should also increase.

4.2. Prefetching effect

We conducted three different experiments such as two prefetchings with a cost-optimal configuration in the previous section (64 MB cache and 512 KB block) and a common one (4 MB cache and 256 KB block), and non-prefetching. As shown in Fig. 3, compared to non-prefetching, two prefetching achieved about 100% and 30% improvement on average in terms of the total number of concurrent streams in up to 16 disk configuration. Fig. 4 shows the disk time ratio obtained from the analysis and experiments when eight disks are employed. The prefetching experiments enable 43% and 13% greater disk bandwidth utilization. This is because on average the prefetching eliminates the seek time and rotational latency of about 80% and 40% of the total blocks. In addition, from Figs. 3 and 4, we can see that the analytical results show the same trend as the experiments.

5. CONCLUSIONS

The FC-AL-based servers can be suitable as a part of a residential service gateway to provide high quality entertainment services to local clients in internal networks due to high performance and cost-benefit. We have analyzed how many concurrent streams the prefetching can increase depending on disk cache and blocks size in FC-AL-based servers. Also, we have estimated the optimal number of blocks transmitted from the disk cache to the FC-AL per FC-AL arbitration among multiple disks. We have also shown that the experimental results are consistent with our analytical model. Our future work is to develop efficient schemes to support VCR-like functions such as fast forward/rewind.

6. REFERENCES

[1] ANSI, “Fiber Channel Arbitrated Loop (FC-AL-2), 2001.

[2] Shenze Chen and Manu Thapar, “A Fibre Channel-based Architecture for Internet Multimedia Server Clusters,” In Proc. of International Algorithms and Architectures for Parallel, pp. 437-450,1997.

[3] Z. Dimitrijevic, R. Rangaswami , and E. Chang, “Virtual IO: Preemptible Disk Access,” ACM Multimedia, pp. 231 - 234, 2002.

[4] David H. C. Du, Tai-Sheng Chang, Jenwei Hsieh, Sangyup Shim, and Yuewei Wang, “Two Emerging Serial Storage Interfaces for Supporting Digital Libraries: Serial Storage Architecture (SSA) and Fiber Channel-Arbitrated Loop (FC-AL),” Multimedia Tools and Application, 10(2): 179-203, 2000.

[5] P. Goyal, D. Jadav, D. Modha, and R. Tewari, “CacheCOW: QoS for storage system caches,” In International Workshop on Quality of Service (IWQoS), pp. 498-516, 2003.

[6] John R. Heath and Peter J.Yakutis, “High-Speed Storage Area Networks Using a Fiber Channel Arbitrated Loop Interconnect,” IEEE Network, pp 51-56, 2000.

[7] A. Hondroulis, C. Harizakis, and P. Triantafillou, “Optimal Cache Memory Exploitation for Continuous Media: To Cache or to Prefetch?,” Multimedia Tools and Applications, 23(3):203-220, 2004.

[8] T. M. Ruwart, “Performance characterization of large and long fiber channel arbitrated loops,” In Proc. of IEEE Mass Storage Systems Symposium, pp. 11 - 21, 1999.

[9] Telikepalli, R., Drwiega, T. and Yan, J, “Storage area network extension solutions and their performance assessment,” IEEE Communications Magazine, 42(4):56-63, 2004.