Chapter III

Real-Time Disk Scheduling for Multimedia Data Retrieval

Ray-I Chang
IIS, Academia Sinica, Taiwan

Wei-Kuan Shih
CIS, NTHU, Taiwan

Ruei-Chuan Chang
CIS, NCTU, Taiwan

INTRODUCTION

The great progresses in computer and network technologies make applications of distributed multimedia ever more popular. Various commercial products have been presented in these years such as electronic commerce, video-on-demand, teleconferencing, tele-medicine, distance education and digital library. The basic architecture of a distributed multimedia system contains three parts—the server subsystem, the network subsystem and the client subsystem, as shown in Figure 1 (Chang, et al., 1997; Chang, 1996; Gemmell and Han, 1994; Rangan and Vin, 1993).

Server subsystem: Media data are stored in the storage devices of server subsystem, such as hard disks and CD-ROM, by data placement schemes (Wang, et al., 1997; Lougher and Shepherd, 1993; Vin and Rangan, 1991). When user requests are presented, media data must be retrieved from storage devices to system buffer under pre-specified timing requirements (Tindell and Burns, 1994).

Figure 1. The basic architecture of a distributed multimedia system
Network subsystem: Then, the retrieved data are transmitted from server to client by a real-time network transmission scheduler. Generally, the available bandwidth and queue length in network are limited.

Client subsystem: At last, client uses a template buffer to regulate the received data by introducing the suitable start-up latency. Therefore, the media stream can be played under guaranteed QoS (quality-of-service).

Different from the data block required in conventional computation systems, media data (such as video and audio) are notably in large volume and have strict timing requirements for presentation and synchronization. The data rates required for smoothly presenting different media data are shown in Table 1. As the volumes of media data are large, the disk scheduler has a pronounced effect to the system performance. In this chapter, we study the real-time disk scheduling problem for supporting multimedia data retrieval.

A real-time disk scheduler considers four input parameters for each task $T_i$ (the data location $a_i$, the data capacity $b_i$, the ready time $r_i$ and the deadline $d_i$) to decide two timing factors (start-time $e_i$ and fulfill-time $f_i$) scheduled to retrieve media data. Without loss of generality, a non-real-time task can be viewed as a real-time task with an unlimited deadline (Stankovic and Buttazzo, 1995).

**Definition 1:**

*Data location:* the start location at which the required data block is stored.

*Data capacity:* the data size that is retrieved from the start location.

*Ready time:* the earliest time at which the task is ready to be started.

*Deadline:* the latest time at which the task must be completed.

*Start-time:* the time at which the task is actually started.

*Fulfill-time:* the time at which the task is actually completed.

A magnetic disk is generally made up of rotating disks, access heads and an assembly arm that moves the access heads into specified cylinders. The request for disk service, said $T_i$, specifies a pair of necessary information such as the location $a_i$ and the amount $b_i$ of data to be transferred. To serve this request, first the disk arm must take a seek time to move the disk arm to the appropriate cylinder. After that, the system must wait until the desired block rotates under its read-write head. This waiting time is called latency time. Then, a switch time is taken for switching the controller’s attention from one disk head to the other within the same cylinder. Finally, the amount of data transferred between the disk and buffer will take a transfer time. A disk service time $c_i$ is the summation of a seek time, a latency time, a head switch time and a transfer time (King, 1990; Ruemmler and Wilkes, 1994).

$$c_i = \text{seek time} + \text{latency time} + \text{switch time} + \text{transfer time}$$

Conventionally, a disk scheduler was designed to maximize the data throughput in serving the presented tasks. However, in a time-critical application, the first goal of a disk scheduler is to meet the specified real-time constraints (Shih et al., 1992; Mok, 1983). Maximizing the data throughput is also important, but a real-time task is meaningless if the

<table>
<thead>
<tr>
<th>phone voice</th>
<th>CD music</th>
<th>MPEG audio</th>
<th>MPEG video</th>
<th>HDTV video</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.064 M bits/sec</td>
<td>1.40 M bits/sec</td>
<td>0.384 M bits/sec</td>
<td>0.42 M bytes/sec</td>
<td>81.0 M bytes/sec</td>
</tr>
</tbody>
</table>

Table 1. The required data rates for presenting different media data
Semi-Supervised Multimodal Fusion Model for Social Event Detection on Web Image Collections
Zhenguo Yang, Qing Li, Zheng Lu, Yun Ma, Zhiguo Gong, Haiwei Pan and Yangbin Chen (2015). *International Journal of Multimedia Data Engineering and Management* (pp. 1-22).
[www.igi-global.com/article/semi-supervised-multimodal-fusion-model-for-social-event-detection-on-web-image-collections/135514?camid=4v1a](www.igi-global.com/article/semi-supervised-multimodal-fusion-model-for-social-event-detection-on-web-image-collections/135514?camid=4v1a)