

2022년 2학기 운영체제실습 13주차

I/O Scheduler

System Software Laboratory

School of Computer and Information Engineering
Kwangwoon Univ.

Contents

- 디스크 스케줄링
- I/O 스케줄러 종류
- IOZone

- 실습 1

Linux I/O Scheduler

■ 디스크 스케줄링의 목표

- 디스크 서비스 시간
 - = 탐색시간(seek time) + 회전시간(rotation delay) + 전송시간(transfer time)
- 탐색시간과 회전지연시간의 최소화
 - 회전지연시간은 운영체제 입장에서는 고려하기 어려움
 - 탐색시간의 최소화에 주력

■ 요청(Request)

- I/O 스케줄의 단위
- struct request로 표현

■ 요청 큐(Request queue)

- 요청들의 대기 큐
- struct request_queue로 표현

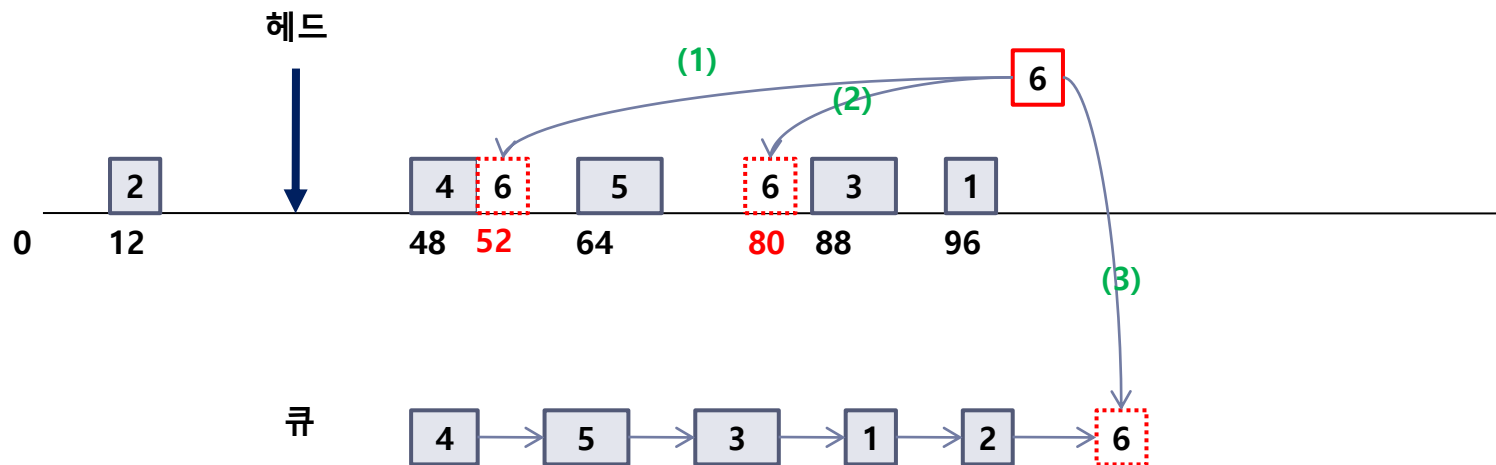
■ I/O scheduler의 기능

- merging, sorting

Linux I/O Scheduler

▪ Elevator scheduler

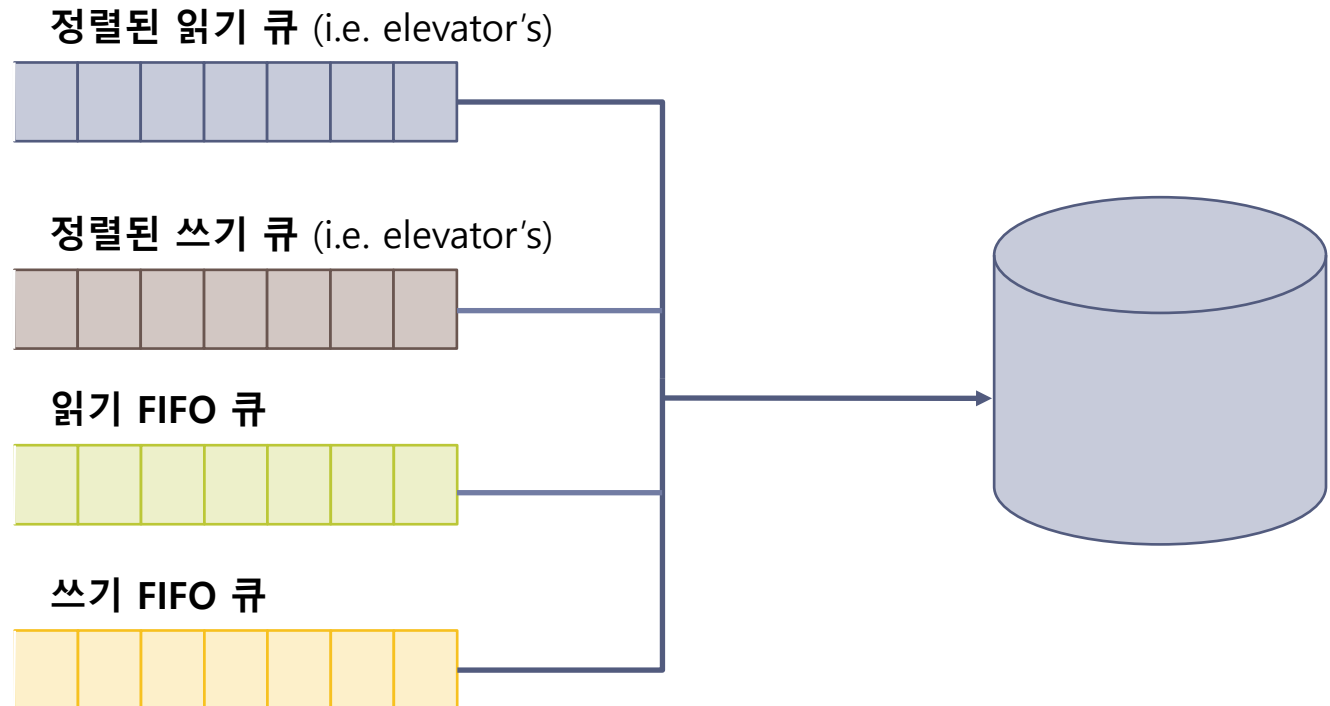
- Linux Kernel 2.4에서 사용
- 디스크 헤드의 이동 방향을 고려
- Mmerge
 - Front, Back
- Insertion point
 - (1) 기존 요청과 인접한 요청이 들어오면 병합
 - (2) 물리적으로 적당한 위치에 삽입
 - (3) 큐에 오랫동안 기다린 요청이 있을 경우, 큐의 맨 끝에 삽입



Linux I/O Scheduler

■ Deadline I/O scheduler

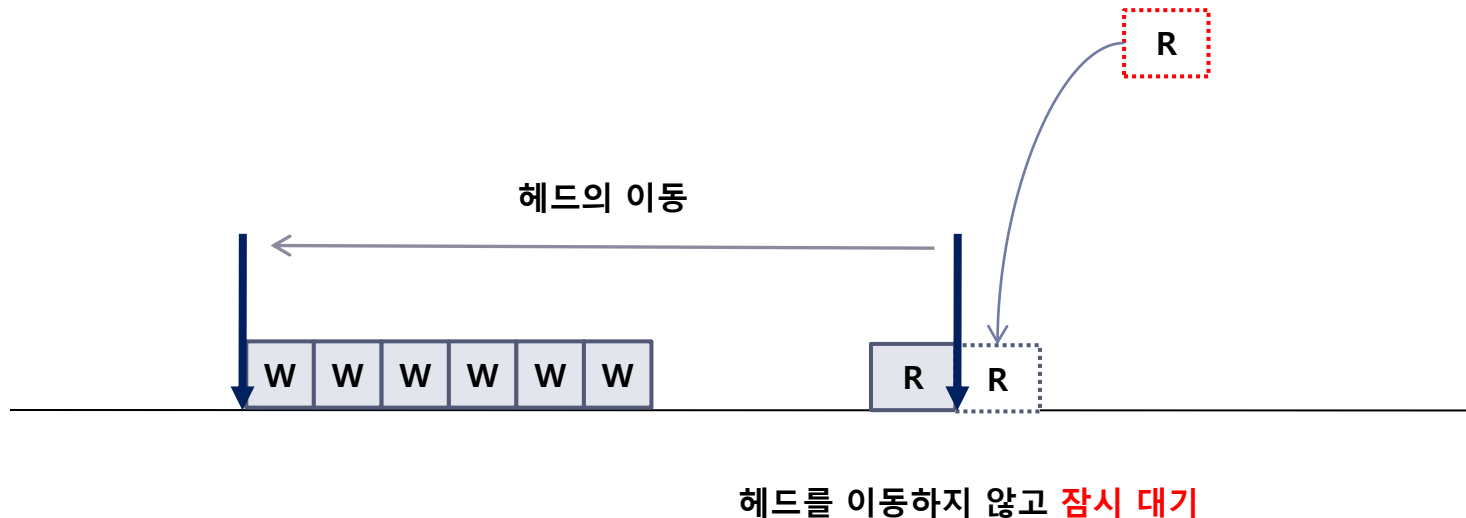
- 4개의 큐 사용
 - 정렬된 읽기 큐, 정렬된 쓰기 큐, 입력 FIFO 큐, 출력 FIFO 큐
 - Deadline이 지난 요청이 없을 경우에 정렬된 큐에서 요청을 꺼내서 처리
- 모든 요청은 마감시간을 가짐
 - 읽기 요청(500ms), 쓰기 요청(5s)
- **기아 현상(starvation) 방지**
 - 읽기 우선 정책



Linux I/O Scheduler

Anticipatory I/O scheduler

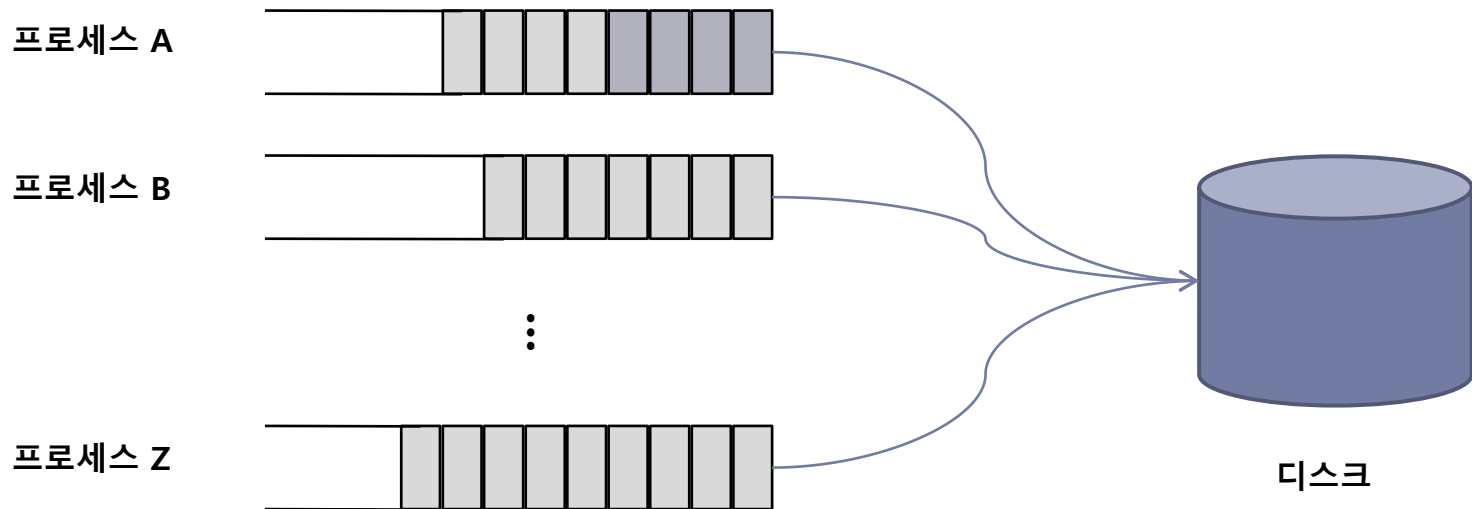
- 데드라인 I/O 스케줄러에 기반
- 인접한 위치에서 짧은 시간 간격으로 읽기 요청하는 경우 빈번
- 읽기 요청을 처리한 후, 헤드를 이동하지 않고 잠시동안 새로운 요청 기다림
 - 6ms(설정 가능한 값)
 - 예측 실패 시 **오버헤드(Overhead)**



Linux I/O Scheduler

■ CFQ(Complete Fair Queueing) I/O scheduler

- 입출력을 요청하는 모든 프로세스들에 대해 디스크 I/O 대역폭을 공평하게 할당하는 것을 보장하는 기법
- I/O를 요청한 프로세스 별로 큐를 할당
- 각 프로세스에 관한 큐는 섹터 순으로 정렬
- 각 큐는 라운드 로빈으로 처리
 - 4개씩 요청을 처리(설정 가능한 값)



Linux I/O Scheduler

- **NOOP(NO OPreation) I/O scheduler**
 - 인접한 요청 병합만 수행하고 그 외에 아무 작업을 하지 않음
 - Request FIFO 큐만 유지
 - Random access 하는 device를 위한 스케줄러
 - 탐색에 대한 부담이 없음 → 정렬이 필요 없음
 - 플래시 메모리에서 주로 사용

Linux I/O Scheduler 확인/선택 방법

- Scheduler 확인

```
sslab@ubuntu:~$ cat /sys/block/sda/queue/scheduler
noop deadline [cfq]
sslab@ubuntu:~$
```

- Scheduler 변경

- noop로 변경

```
sslab@ubuntu:~$ echo noop | sudo tee /sys/block/sda/queue/scheduler
noop
sslab@ubuntu:~$ cat /sys/block/sda/queue/scheduler
[noop] deadline cfq
sslab@ubuntu:~$
```

Benchmark – IO Zone

■ 파일시스템 벤치마크 프로그램

- Linux, Window, NFS client 환경에서의 I/O 동작의 성능 측정
- read, write, re-read, re-read, re-write, read backwards, read strided, fread, fwrite, random read, pread, mmap, aio_read, aio_write 등 다양한 테스트 제공

■ IOZone 설치

- \$ sudo apt-get install iozone3

```
sslab@ubuntu:~$ sudo apt-get install iozone3
Reading package lists... Done
Building dependency tree
Reading state information... Done
The following NEW packages will be installed:
  iozone3
0 upgraded, 1 newly installed, 0 to remove and 98 not upgraded.
Need to get 417 kB of archives.
After this operation, 742 kB of additional disk space will be used.
Get:1 http://us.archive.ubuntu.com/ubuntu xenial/multiverse amd64 iozone3 amd64 429-3 [417 kB]
Fetched 417 kB in 6s (62.5 kB/s)
Selecting previously unselected package iozone3.
(Reading database ... 216606 files and directories currently installed.)
Preparing to unpack .../iozone3_429-3_amd64.deb ...
Unpacking iozone3 (429-3) ...
Processing triggers for man-db (2.7.5-1) ...
Setting up iozone3 (429-3) ...
sslab@ubuntu:~$
```

Benchmark – IO Zone

■ IO Zone Options

- -R : Excel report 생성
- -b[filename] : Excel과 호환 가능한 binary file 생성
- -i : test 실행 (ex. -i 1 -i 2)

Number	Description	Number	Description	Number	Description
0	write/re-write	1	read/re-read	2	random-read/write
3	read-backwards	4	re-write-record	5	strid-read
6	fwrite/re-fwrite	7	fread/re-fread	8	random-mix
9	pwrite/re-pwrite	10	pread/re-pread	11	writew/re-pwritew
12	preadv/re-preadv				

- -l(대문자 i) : 파일에 대한 모든 operation은 buffer cache를 우회 및 disk 직접 이동(i.e. transferring data synchronously)

Benchmark – IO Zone

■ IO Zone Options

- -r #

: record size 설정(기본 4k)

ex. -r 512m → record size 512MB로 설정

name	Description
k	Size in KB
m	Size in MB
g	Size in GB

- -s(소문자)

: file size 설정

- -t

: throughput test에 사용할 thread 또는 process 개수 설정

- -F

: throughput test에서 각 thread 또는 process 파일 이름

Benchmark – IO Zone

■ IOZone 사용 예제

■ 조건

- I/O 실험 결과에 대한 excel file로 작성 및 file name : file_name.xls
- read, write, random write 연산 수행
- buffer cache 거치지 않고 연산 수행
- record size : 512KB
- file size : 512MB
- thread or process 개수 : 1
- thread or process의 파일 경로 : ~/iozone_test

```
$ iotest -R -i 0 -i 1 -i 2 -l -r 512k -s 512m -t 1 -F ~/iozone_test -b file_name.xls
```

실습 1

- 스케줄러 별 성능 테스트
- Scheduler
 - noop
- IOZone 사용 예제
 - 조건
 - I/O 실험 결과에 대한 excel file로 작성 : file name : ex_noop.xls
 - Read, write, random write 연산 수행
 - Buffer cache 거치지 않고 연산 수행
 - Record size : 5KB
 - file size : 1GB
 - Thread or process 개수 : 1
 - 파일 경로 : ~/iozone_test

실습 1

■ 실습 방법

1. 스케줄러 확인 후 변경
2. 매 실험 전 캐시 및 버퍼 비워서 실험에 영향 주는 요소 제거

\$ rm -rf ~/iozone_test

Remove temporary directories and files

\$ sync

Linux command to flush file system buffer

\$ echo 3 | sudo tee /proc/sys/vm/drop_caches

Linux commands to free pagecache, dentries, and inodes

3. IOZone 실행

실습 1

■ 결과

- 스케줄러 확인

```
sslab@ubuntu:~$ cat /sys/block/sda/queue/scheduler  
[noop] deadline cfq
```

- 성능 확인(일부분)

```
Children see throughput for 1 random readers = 73539.50 kB/sec  
Parent sees throughput for 1 random readers = 73532.10 kB/sec  
Min throughput per process = 73539.50 kB/sec  
Max throughput per process = 73539.50 kB/sec  
Avg throughput per process = 73539.50 kB/sec  
Min xfer = 1048575.00 kB  
  
Children see throughput for 1 random writers = 74257.95 kB/sec  
Parent sees throughput for 1 random writers = 74248.86 kB/sec  
Min throughput per process = 74257.95 kB/sec  
Max throughput per process = 74257.95 kB/sec  
Avg throughput per process = 74257.95 kB/sec  
Min xfer = 1048575.00 kB
```

```
"Throughput report Y-axis is type of test X-axis is number of processes"  
"Record size = 5 kBytes "  
"Output is in kBytes/sec"
```

```
" Initial write " 58294.48  
"      Rewrite " 62545.08  
"      Read " 84063.67  
"      Re-read " 100809.77  
"      Random read " 73539.50  
"      Random write " 74257.95
```

```
iozone test complete.
```

```
sslab@ubuntu:~$ ls -l  
total 112  
drwxr-xr-x 2 sslab sslab 4096 Oct 21 18:47 Desktop  
drwxr-xr-x 2 sslab sslab 4096 Aug 31 2021 Documents  
drwxr-xr-x 3 sslab sslab 4096 Aug 31 2021 Downloads  
--rw-rw-r-- 1 sslab sslab 2806 Nov 22 20:42 ex_noop.xls
```


Assignment 5

- 제출 기한: 2022. 11.24(목) ~ 2022.12.9(금) 23:59:59
- Delay 없음
- 업로드 양식에 어긋날 경우 감점 처리
- Hardcopy 제출하지 않음