CSE 421/521 - Operating Systems
Fall 2014

LECTURE - XXI

FILE SYSTEMS

Tevfik Koşar

University at Buffalo

November 6th, 2014

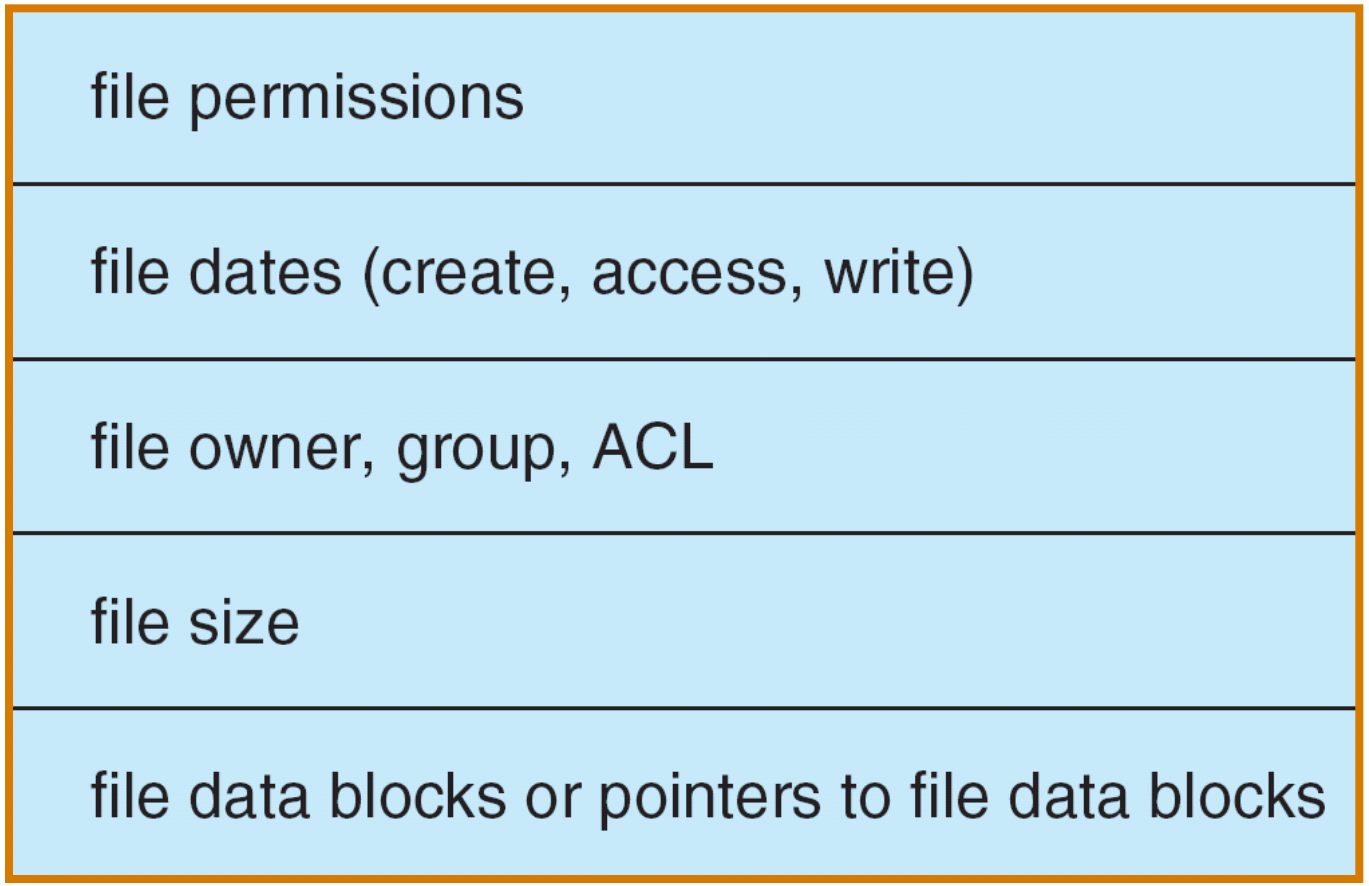
File Systems

- An interface between users and files
- Provides organized and efficient access to data on secondary storage:

1. Organizing data into files and directories and supporting primitives to manipulate them (create, delete, read, write etc)
2. Improve I/O efficiency between disk and memory (perform I/O in units of blocks rather than bytes)
3. Ensure confidentiality and integrity of data

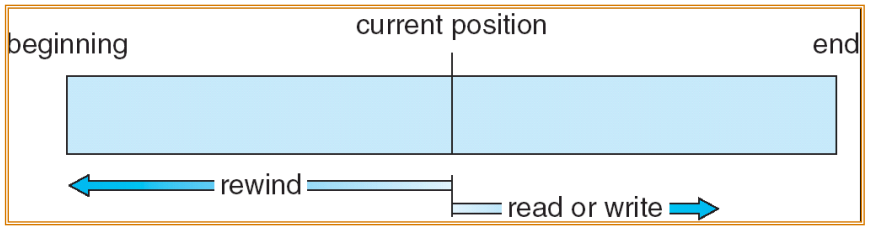
- Contains file structure via a File Control Block (FCB)
  - Ownership, permissions, location..

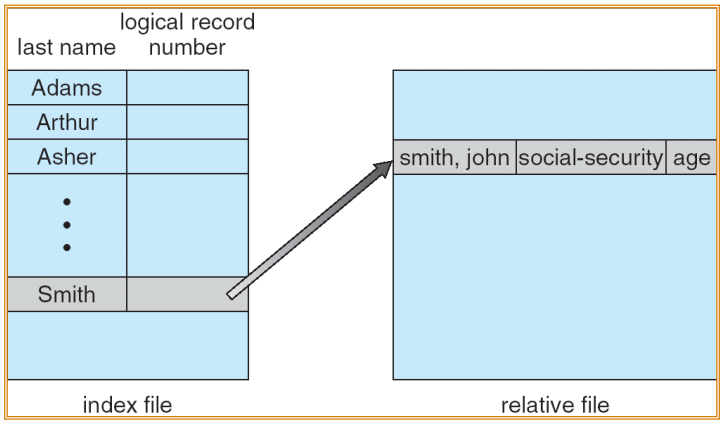
A Typical File Control Block



File access methods

- Some file systems provide different access methods that specify ways the application will access data
  - sequential access
    - read bytes one at a time, in order
  - direct access
    - random access given a block/byte #
  - record access
    - file is array of fixed-variable-sized records
  - indexed access
    - FS contains an index to a particular field of each record in a file apps can find a file based on value in that record (similar to DB)





Directories

Directories provide:

- a way for users to organize their files
- a convenient file name space for both users and FS's

Most file systems support multi-level directories

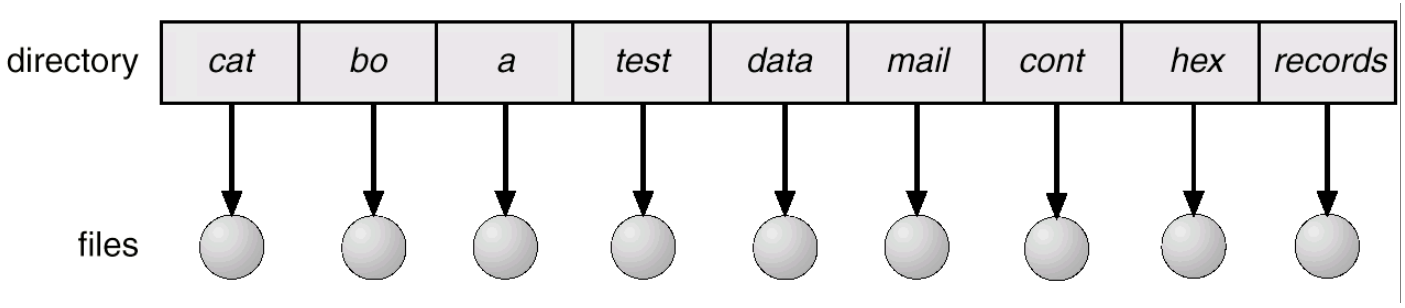
- naming hierarchies (/, /usr, /usr/local, /usr/local/bin, …)

A directory is typically just a file that happens to contain special metadata

- directory = list of (name of file, file attributes)
- attributes include such things as:
  size, protection, location on disk, creation time, access time, …
- the directory list is usually unordered (effectively random)
  when you type "ls", the "ls" command sorts the results for you

Directories

- Single-level directory structure
  - ✓ simplest form of logical organization: one global or **root** directory containing all the files
  - ✓ problems
    - global namespace: unpractical in multiuser systems
    - no systematic organization, no groups or logical categories of files that belong together

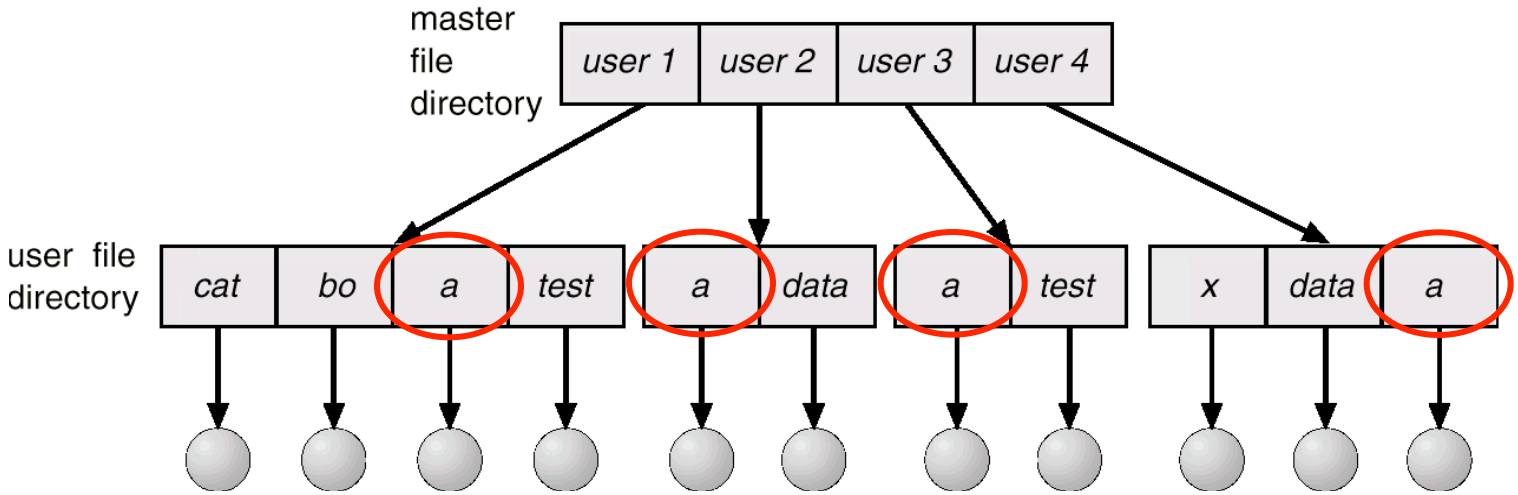


Single-level directory

Silberschatz, A., Galvin, P. B. and Gagne. G. (2003) *Operating Systems Concepts with Java (6th Edition).*

Directories

- Two-level directory structure
  - ✓ in multiuser systems, the next step is to give each user their own private directory
  - ✓ avoids filename confusion
  - ✓ however, still no grouping: not satisfactory for users with many files

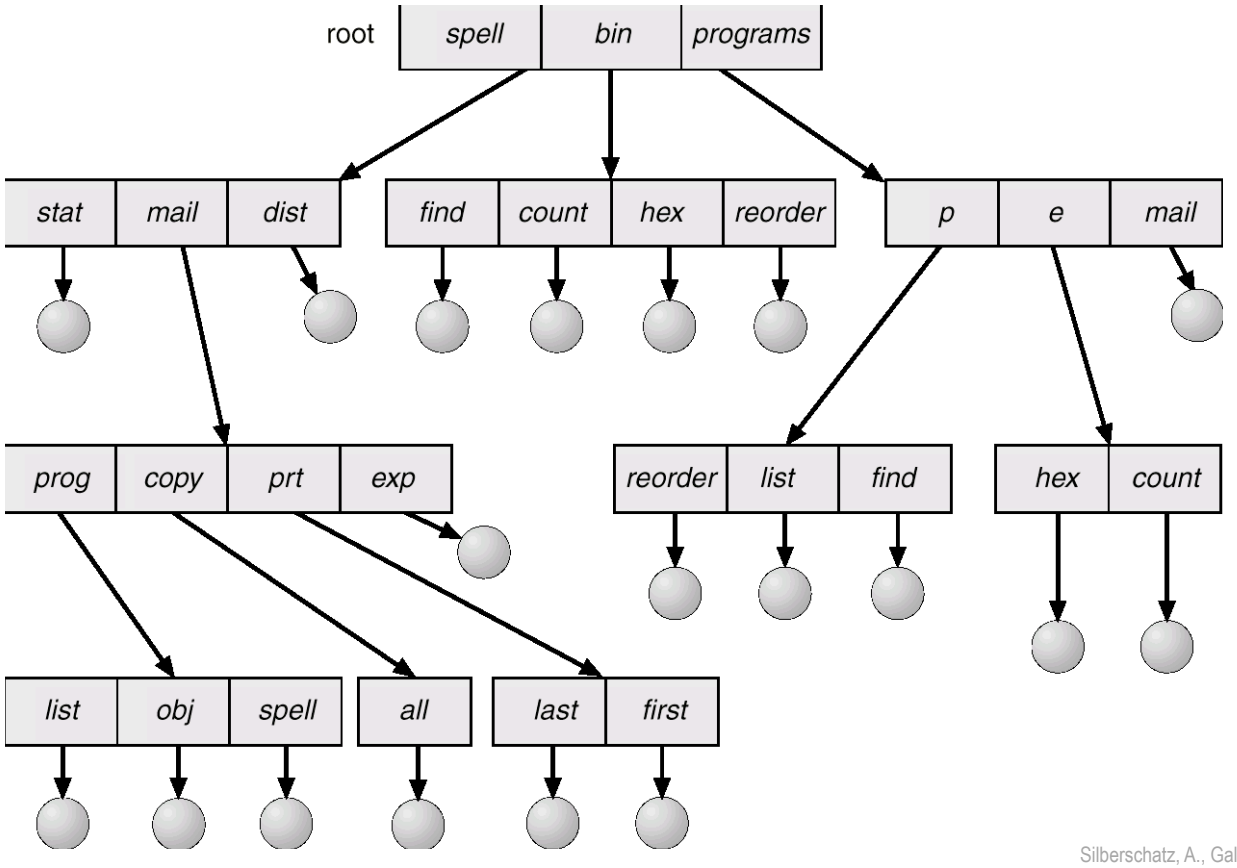


Two-level directory

Silberschatz, A., Galvin, P. B. and Gagne. G. (2003)
Operating Systems Concepts with Java (6th Edition).

Directories

- Tree-structured directory structure



Tree-structured directory

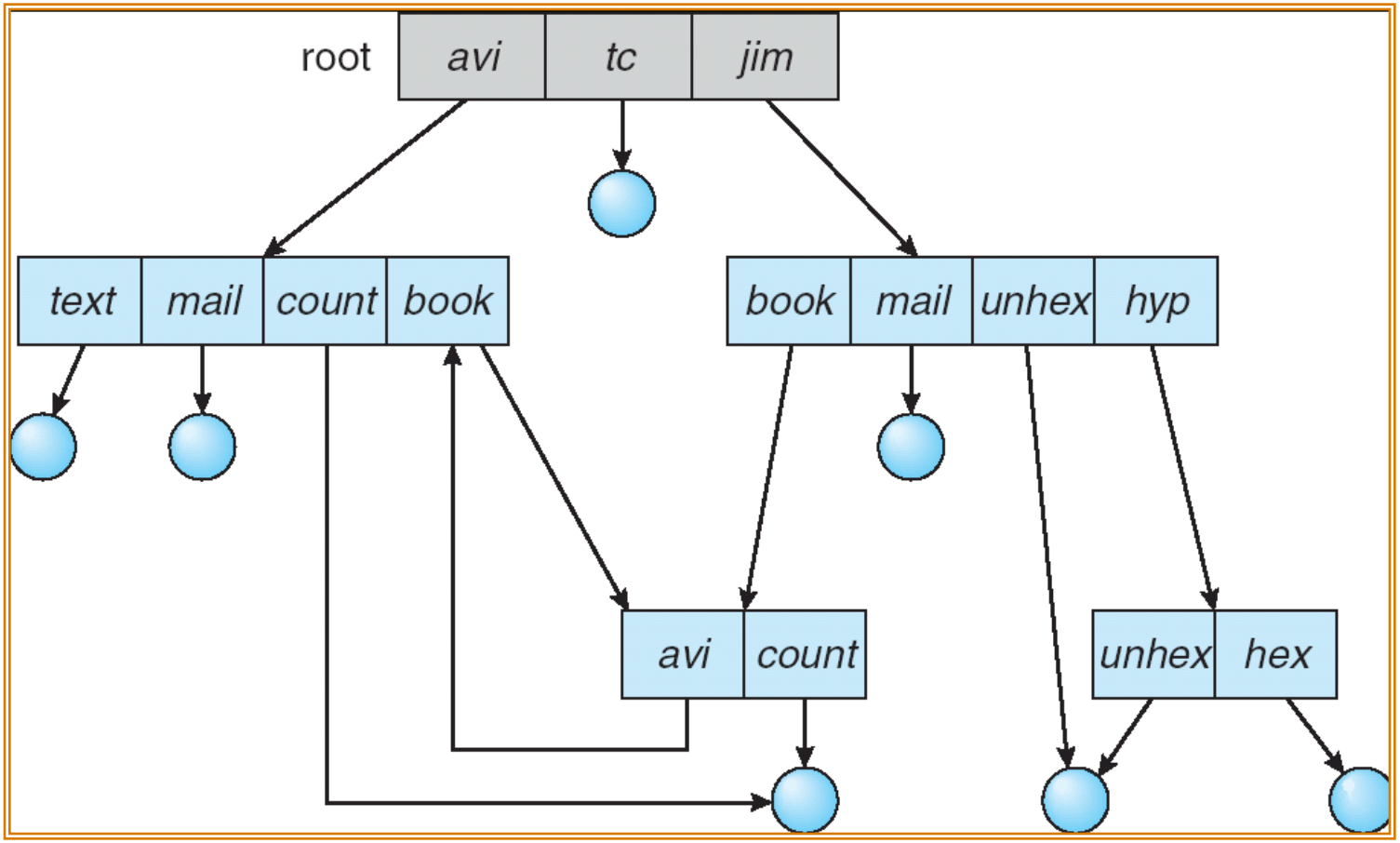
Silberschatz, A., Galvin, P. B. and Gagne. G. (2003) *Operating Systems Concepts with Java (6th Edition).*

Directories

- Tree-structured directory structure
  - ✓ natural extension of the two-level scheme
  - ✓ provides a general hierarchy, in which files can be grouped in natural ways
  - ✓ good match with human cognitive organization: tendency to categorize objects in embedded sets and subsets
  - ✓ navigation through the tree relies on **pathnames**
    - absolute pathnames start from the root, example: /jsmith/academic/teaching/cs446/assignment4/grades
    - relative pathnames start at from a current **working directory**, example: assignment4/grades
    - the current and parent directory are referred to as . and ..

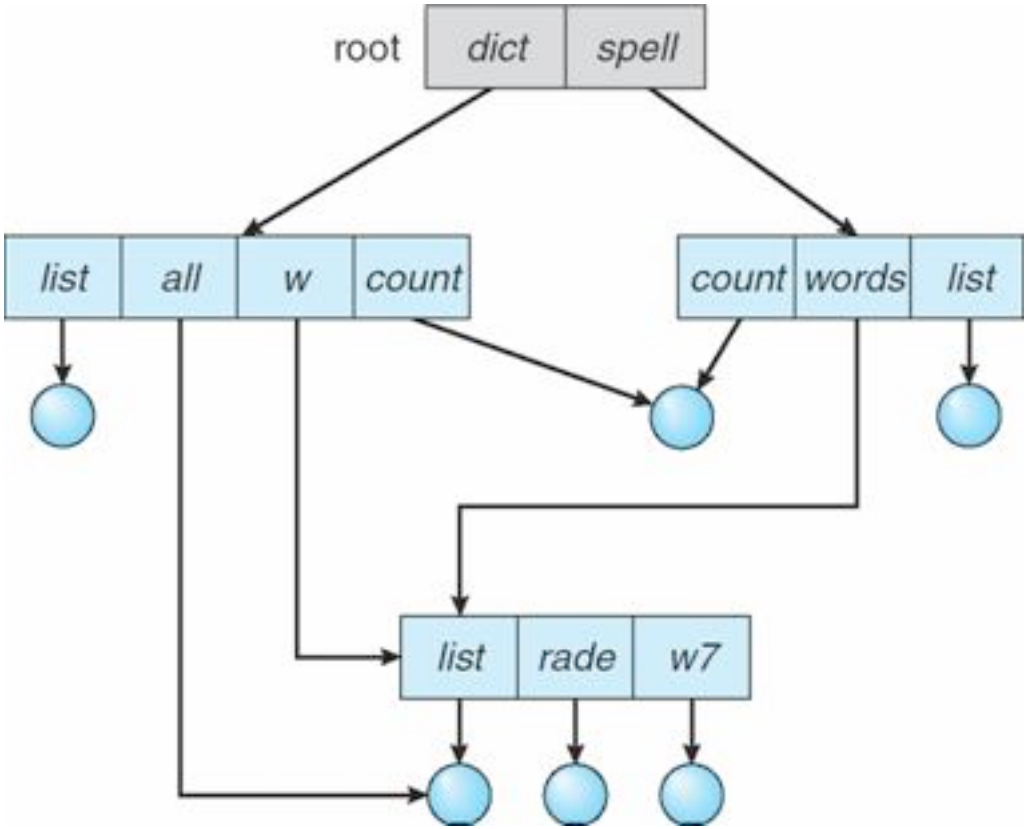
Directories

- General Graph directory structure



Directories

- A-cyclic Graph directory structure



A-cyclic graph structured directory

Silberschatz, A., Galvin, P. B. and Gagne. G. (2003)
Operating Systems Concepts with Java (6th Edition).

Pathname Translation

- Let's say you want to open "/one/two/three"

```
fd = open("/one/two/three", O_RDWR);
```

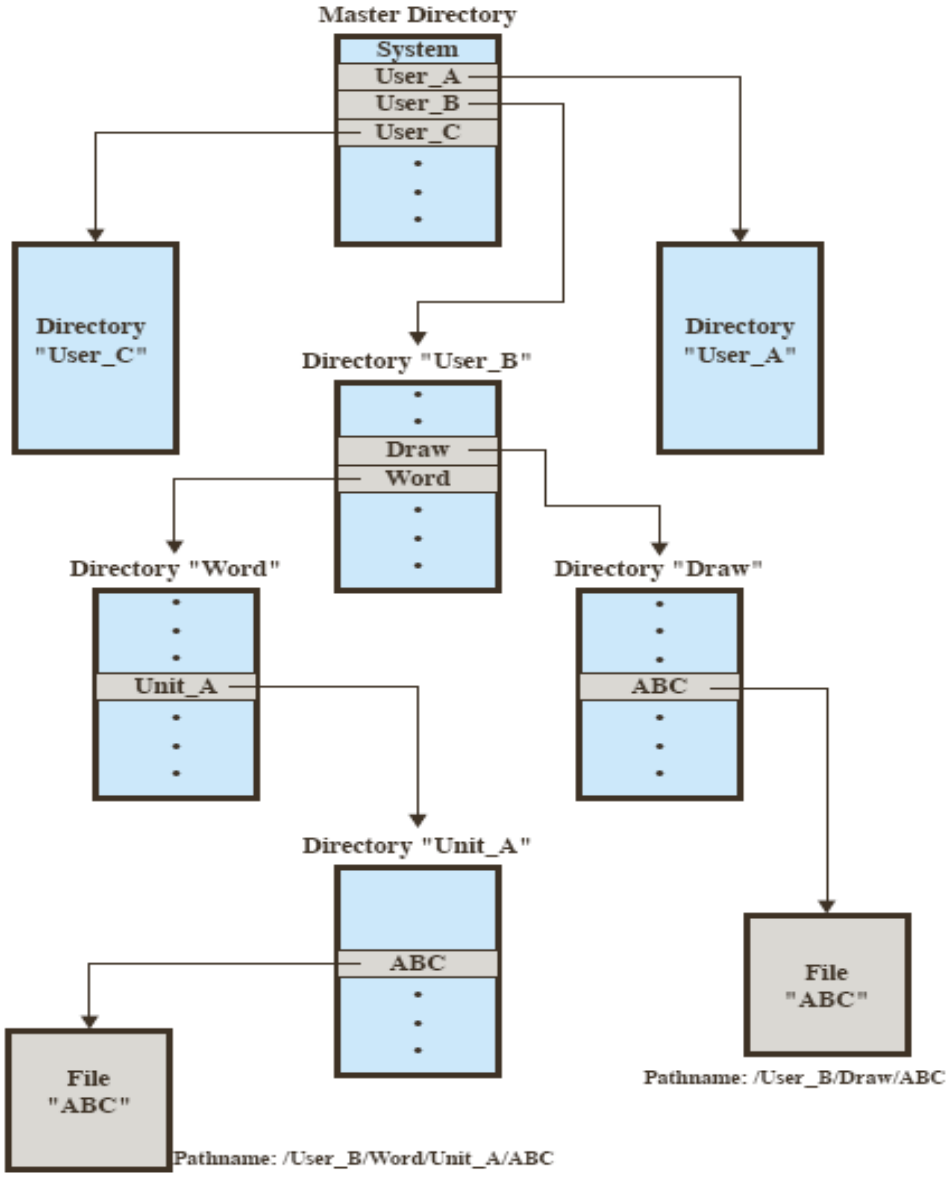
- What goes on inside the file system?
  1. open directory "/" (well known, can always find)
  2. search the directory for "one", get location of "one"
  3. open directory "one", search for "two", get location of "two"
  4. open directory "two", search for "three", get loc. of "three"
  5. open file "three"
  6. (of course, permissions are checked at each step)

- FS spends lots of time walking down directory paths
  - this is why open is separate from read/write (session state)
  - OS will cache prefix lookups to enhance performance
    - /a/b, /a/bb, /a/bbb all share the "/a" prefix

Directory Implementation

- **Linear list** of file names with pointer to the data blocks.
  - simple to program
  - time-consuming to execute
- **Hash Table** – linear list with hash data structure.
  - decreases directory search time
  - **collisions** – situations where two file names hash to the same location
  - fixed size

Directory Implementation



Stallings, W. (2004) *Operating Systems: Internals and Design Principles (5th Edition).*

UNIX Directories

- Directory is a special file that contains list of names of files and their inode numbers
- to see contents of a directory:

```
$ls -1ia .
9535554 .
9535489 ..
9535574 .bash_history
9535555 bin
9535584 .emacs.d
9535560 grading
9535803 hw1
9535571 test
9535801 .viminfo
```

Directories - System View

- user view vs system view of directory tree
  - representation with “dirlists (directory files)”
- The real meaning of “A file is in a directory”
  - directory has a link to the inode of the file
- The real meaning of “A directory contains a subdirectory”
  - directory has a link to the inode of the subdirectory
- The real meaning of “A directory has a parent directory”
  - “..” entry of the directory has a link to the inode of the parent directory

Example inode listing

$ ls -iaFR demodir:

865 . 193 .. 277 a/ 520 c/ 491 y

demodir/a:

277 . 865 .. 402 x

demodir/c:

520 . 865 .. 651 d1/ 247 d2/

demodir/c/d1:

651 . 520 ..

demodir/c/d2:

247 . 520 .. 680 z

Please show the system representation (system view) of this directory tree.

Link Counts

- The kernel records the number of links to any file/ directory.
- The *link count* is stored in the inode.
- The *link count* is a member of *struct stat* returned by the *stat* system call.

Change Links

- What will be the resulting changes in directory tree?

```
$ cp demodir/a/x demodir/c/xcopy

$ ln demodir/a/x demodir/c/d1/xlink

$ ln -s demodir/a/x demodir/c/d1/slink

$ mv demodir/y  demodir/a/y
```

Implementing “pwd”

1. “.” is 247
   chdir ..

2. 247 is called “d2”
   “.” is 520
   chdir ..

3. 520 is called “c”
   “.” is 865
   chdir ..

4. 865 is called “demodir”
   “.” is 193
   chdir ..

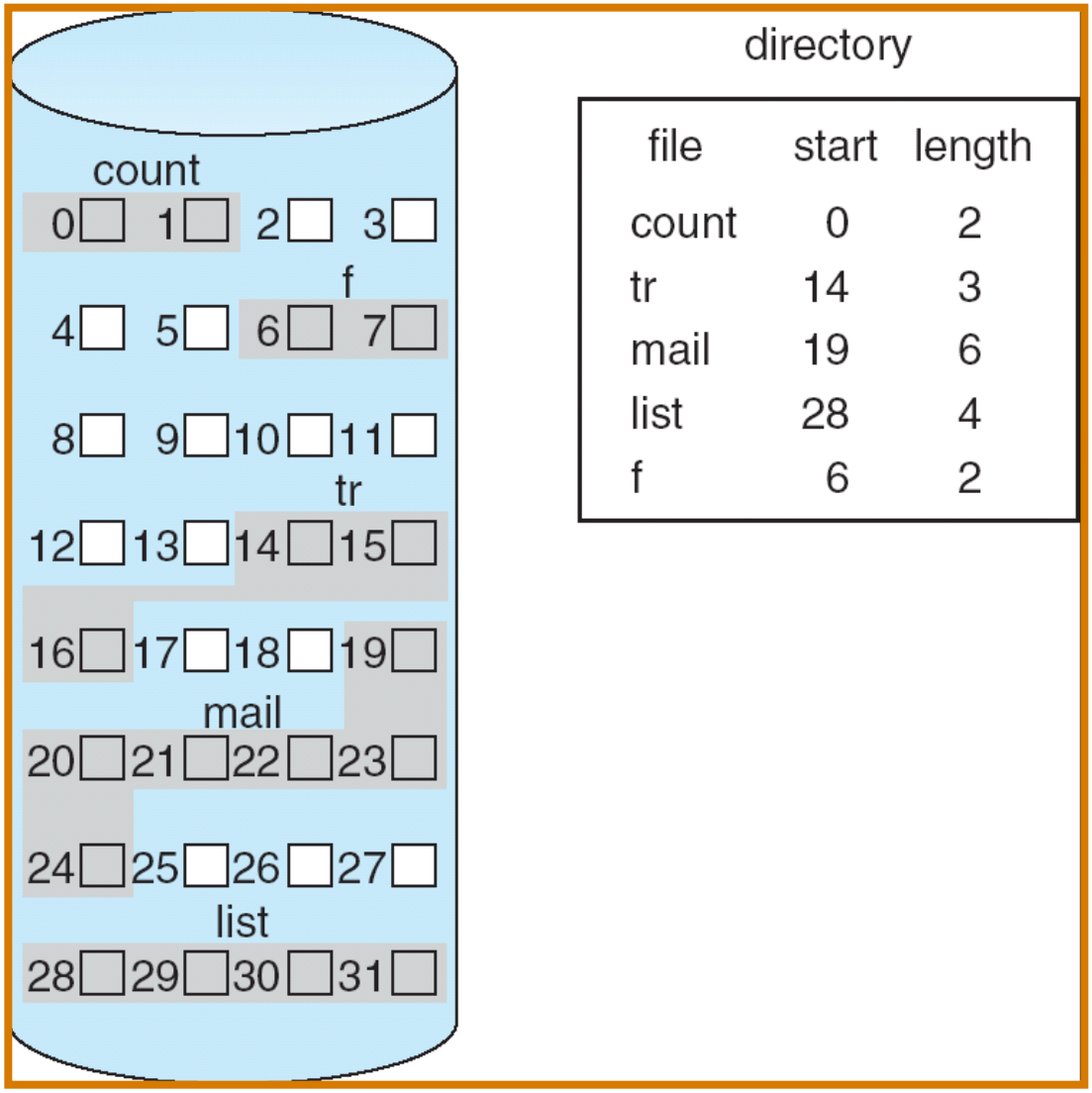
Allocation Methods

- An allocation method refers to how disk blocks are allocated for files:
- **Contiguous allocation**
- **Linked allocation**
- **Indexed allocation**

Contiguous Allocation

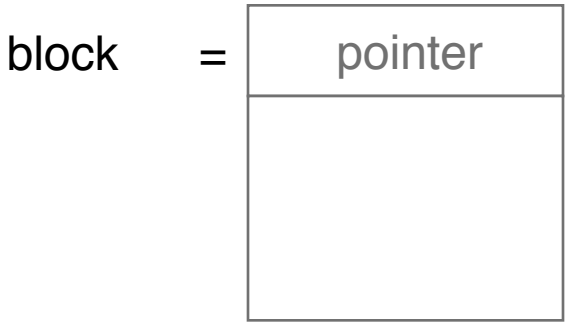
- Each file occupies a set of contiguous blocks on the disk
- + Simple – only starting location (block #) and length (number of blocks) are required
- - Wasteful of space (dynamic storage-allocation problem - fragmentation)
- - Files cannot grow

Contiguous Allocation of Disk Space



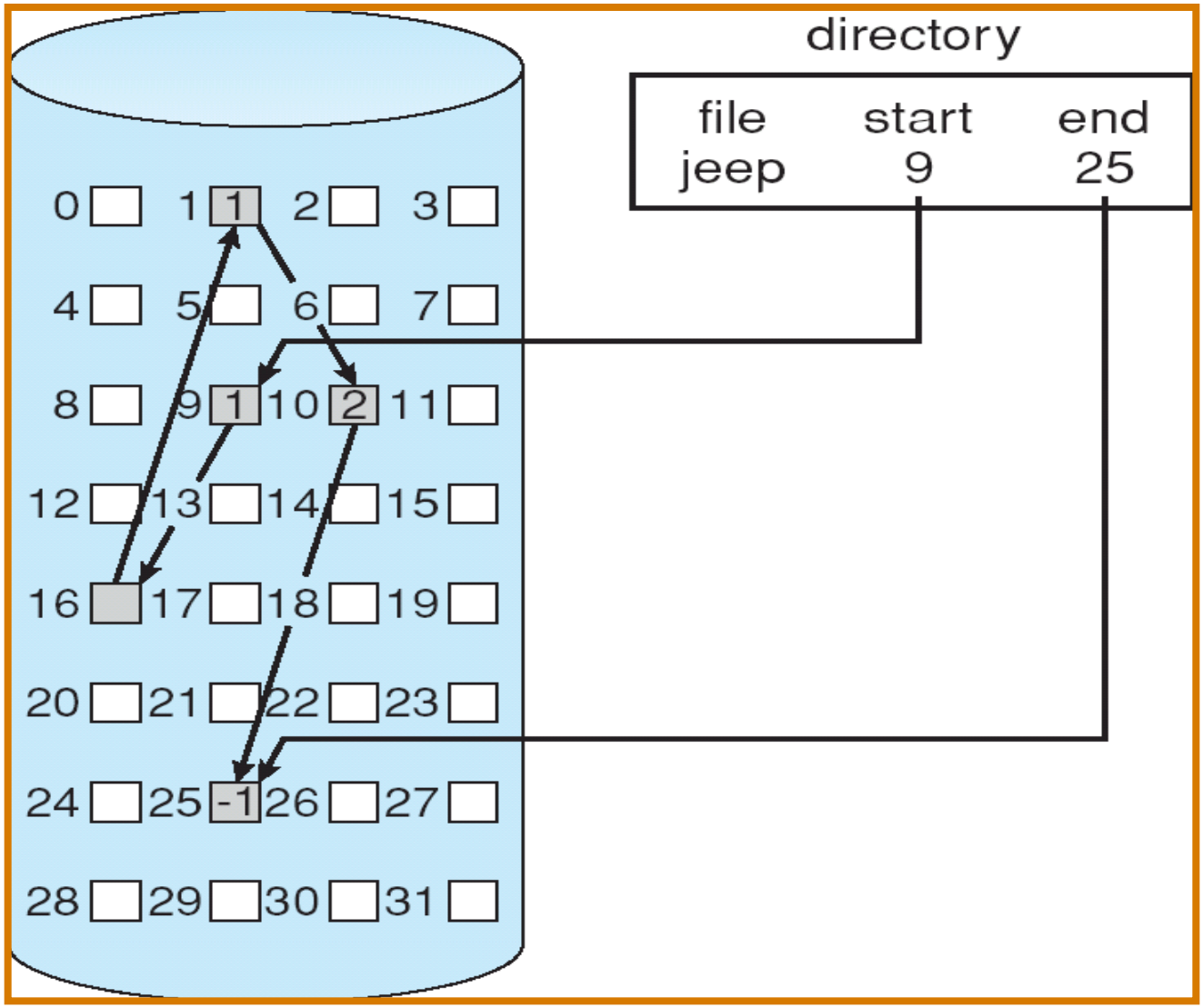
Linked Allocation

- Each file is a linked list of disk blocks: blocks may be scattered anywhere on the disk.

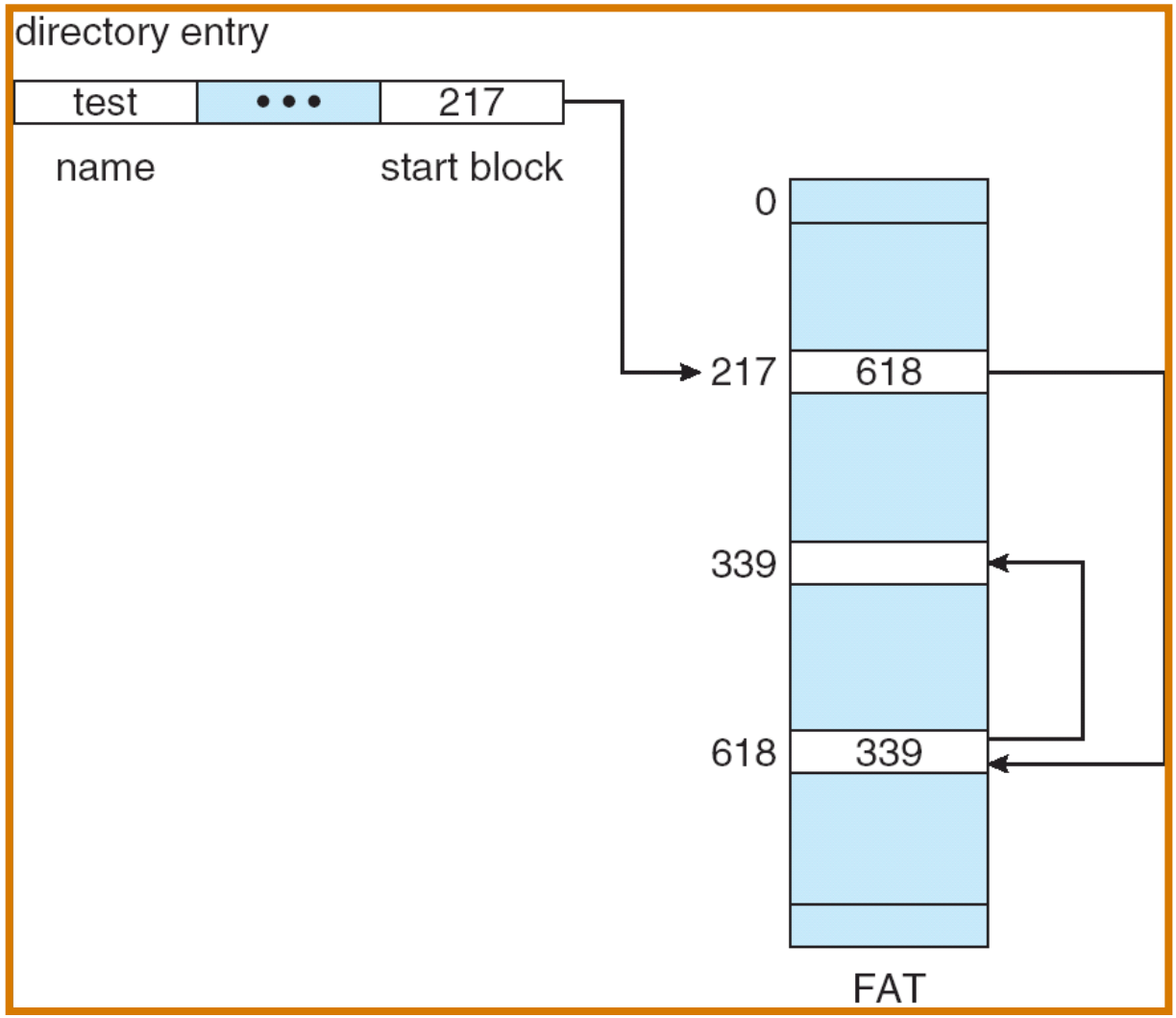


+ Simple – need only starting address
+ Free-space management system – no waste of space
+ Defragmentation not necessary for file allocation
- No random access
- Extra space required for pointers
- Reliability: what if a pointer gets corrupted?

Linked Allocation

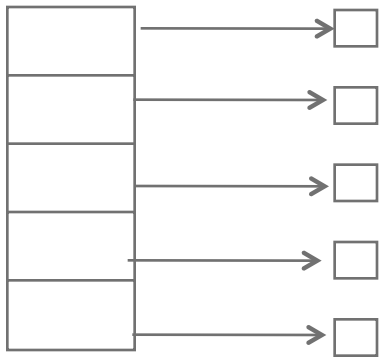


File-Allocation Table



Indexed Allocation

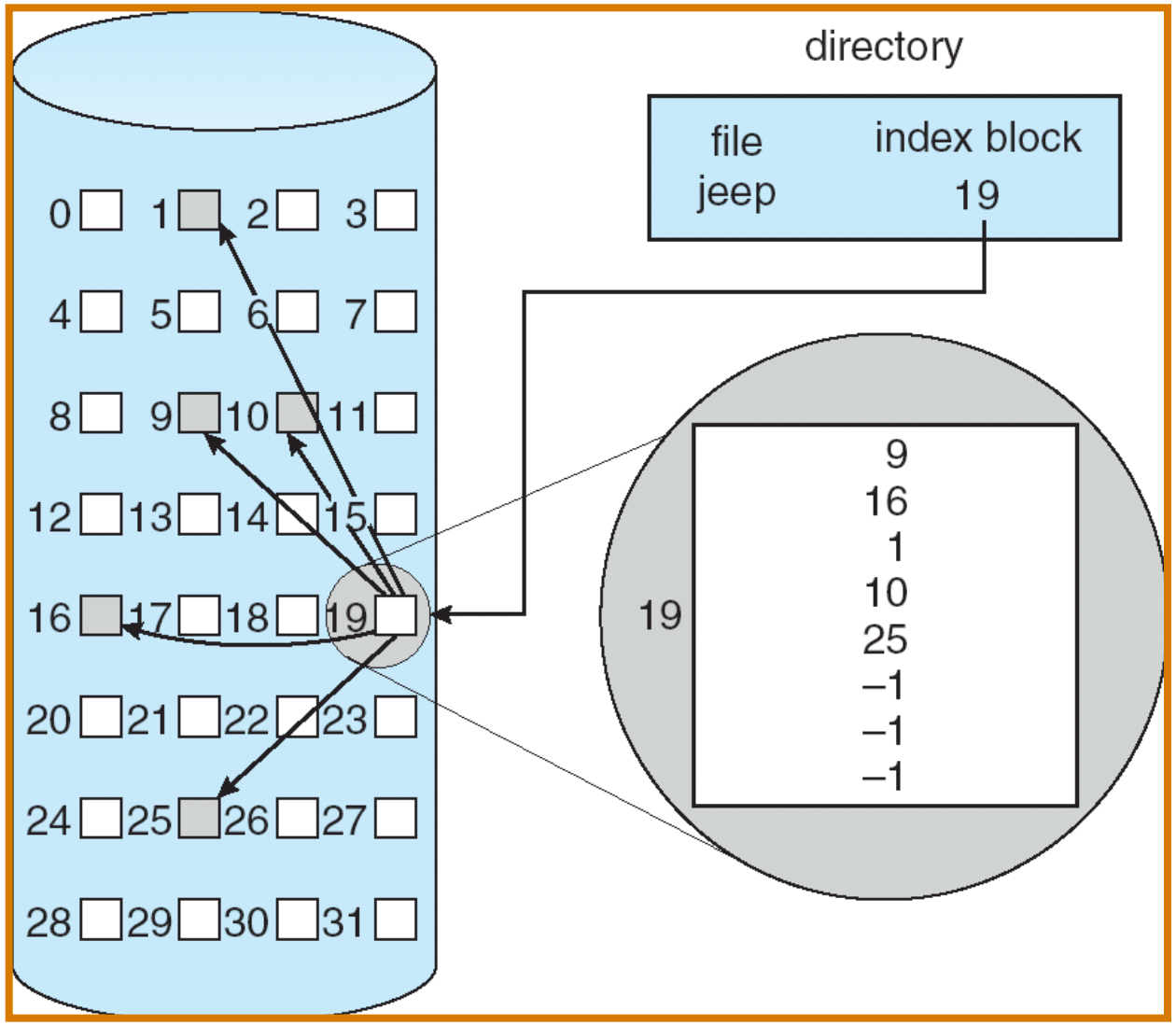
- Brings all pointers together into the *index block*, to allow random access to file blocks.
- Logical view.



index table

+ Supports direct access
+ Prevents external fragmentation
- High pointer overhead --> wasted space

Example of Indexed Allocation

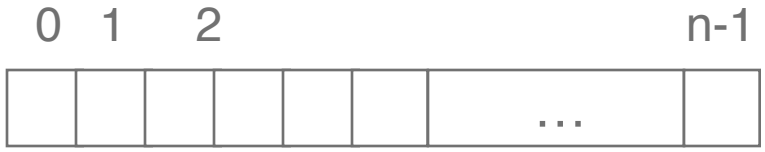


Free Space Management

- Disk space limited
- Need to re-use the space from deleted files
- To keep track of free disk space, the system maintains a free-space list
  - Records all free disk blocks
- Implemented using
  - Bit vectors
  - Linked lists

Free-Space Management (Cont.)

- Bit vector (*n* blocks)
  - Each block is represented by 1 bit
  - 1: free, 0: allocated

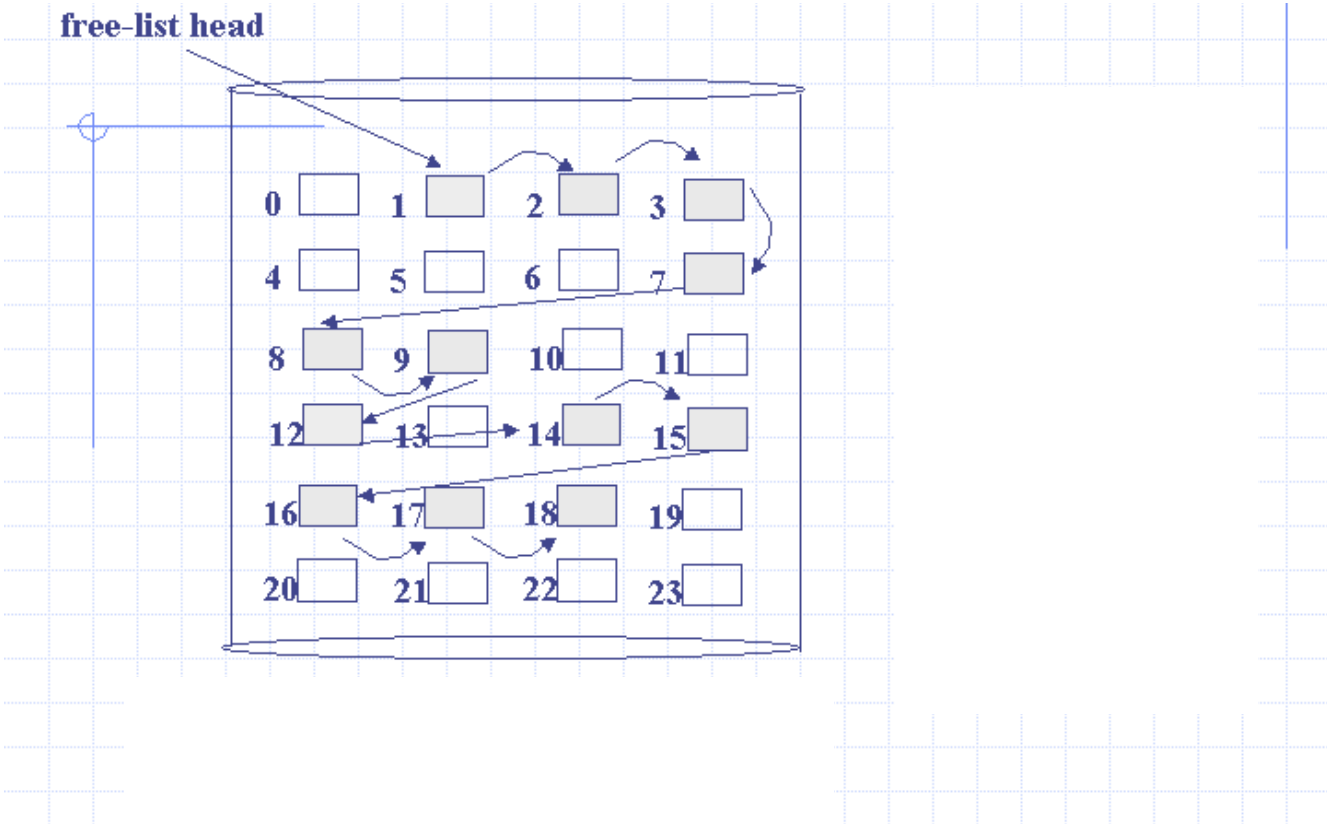


$$bit[i] = \begin{cases} 1 \Rightarrow block[i] \text{ free} \\ 0 \Rightarrow block[i] \text{ occupied} \end{cases}$$

- e.g. 000011111000100010010000

Free-Space Management (Cont.)

- Linked List Approach



Free-Space Management (Cont.)

- Bit map requires extra space
  - Example:

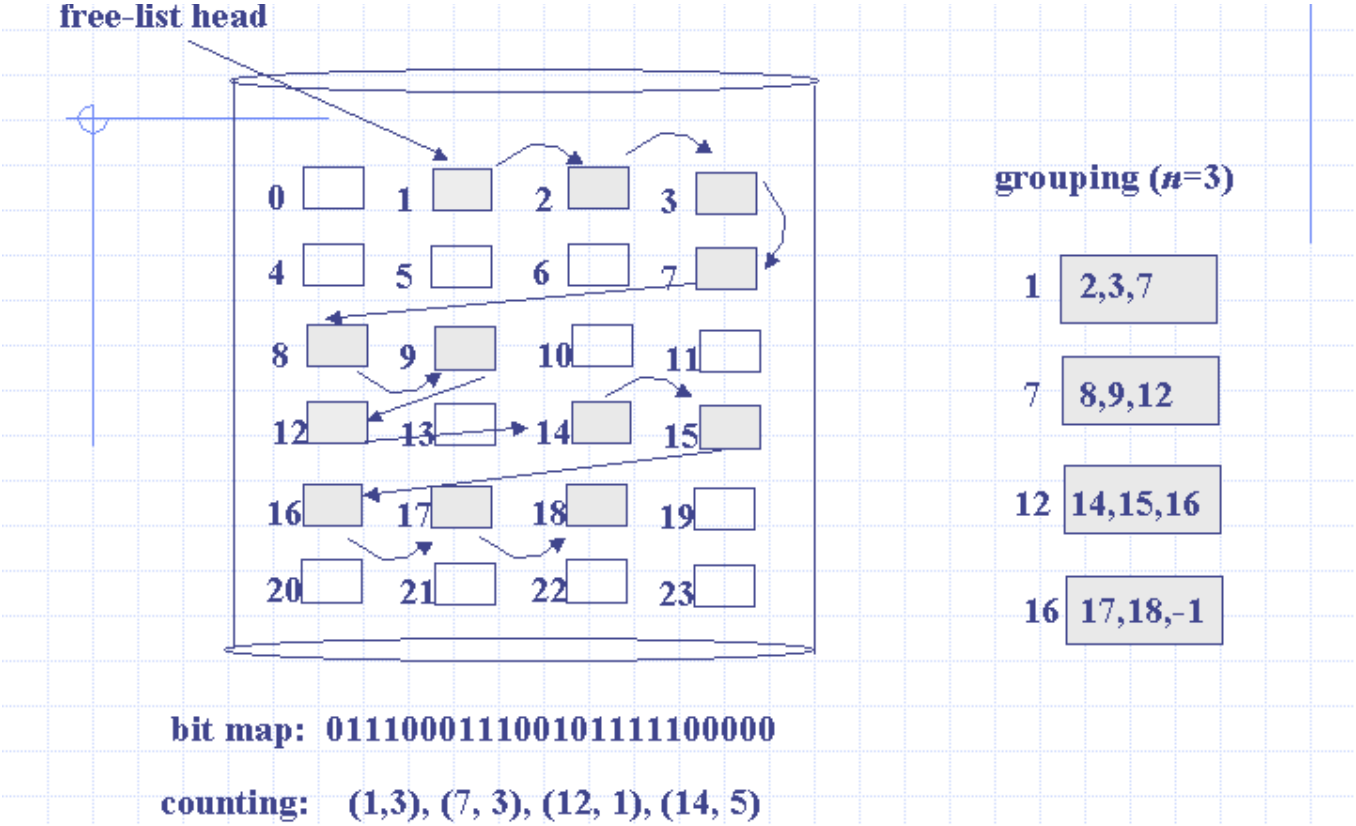
    block size = 2^{12} bytes

    disk size = 2^{30} bytes (1 gigabyte)

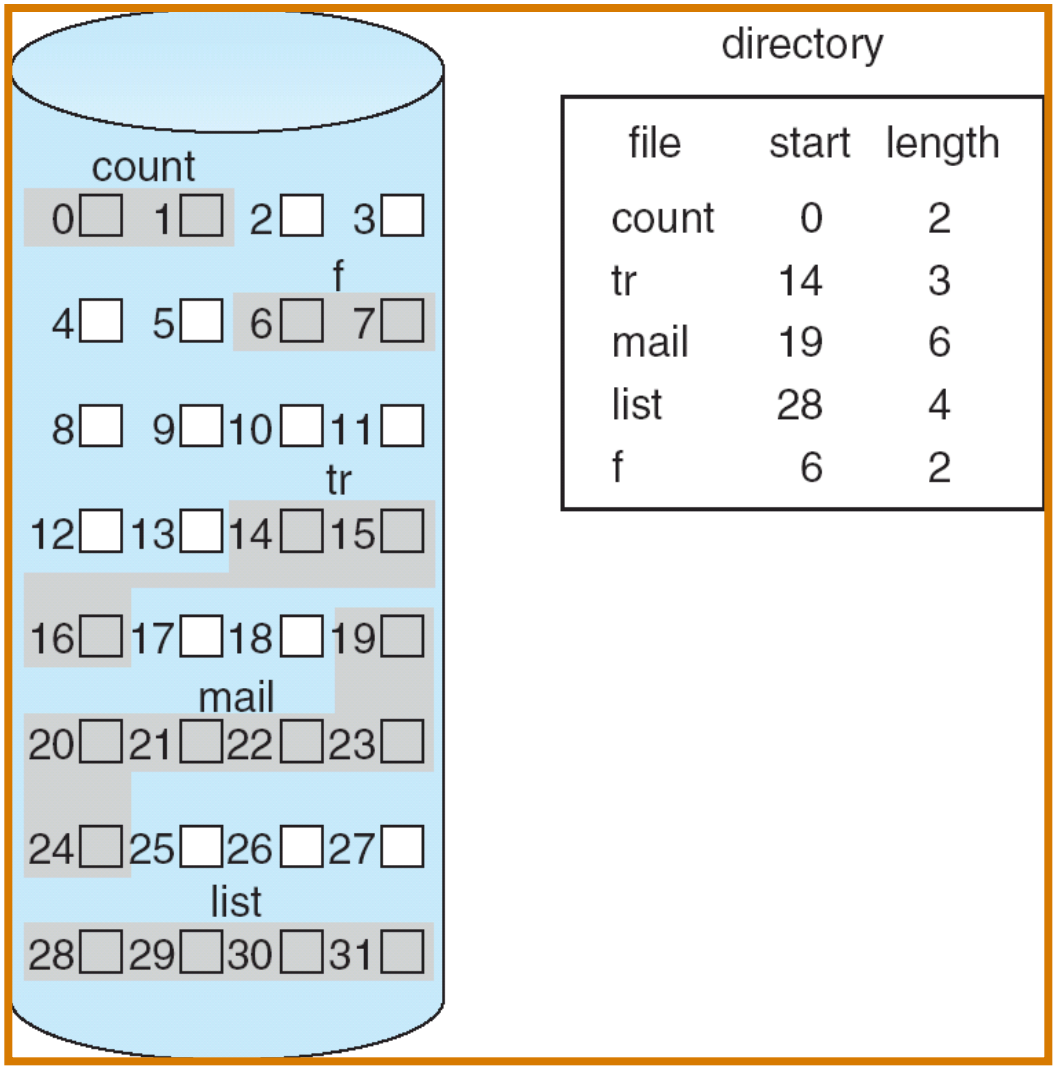
    $n = 2^{30}/2^{12} = 2^{18}$ bits (or 32K bytes)
- Easy to get contiguous files
- Linked list (free list)
  - Cannot get contiguous space easily
  - requires substantial I/O
- Grouping
  - Modification of free-list
  - Store addresses of n free blocks in the first free block
- Counting
  - Rather than keeping list of n free addresses:
    - Keep the address of the first free block
    - And the number n of free contiguous blocks that follow it

Free-Space Management (Cont.)

- Linked List



Exercise

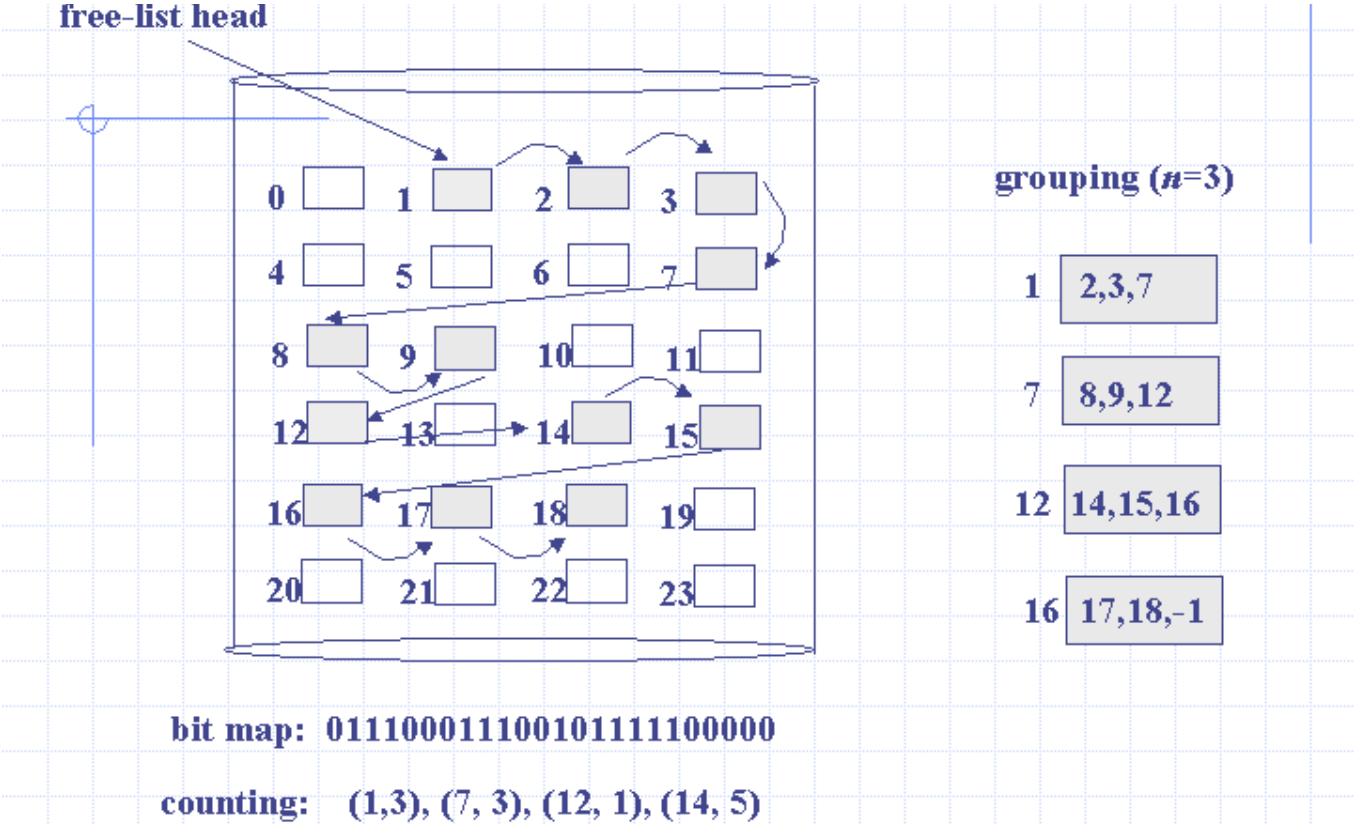


directory

file	start	length
count	0	2
tr	14	3
mail	19	6
list	28	4
f	6	2

Free-Space Management (Cont.)

- Linked List



Acknowledgements

- “Operating Systems Concepts” book and supplementary material by A. Silberschatz, P. Galvin and G. Gagne
- “Operating Systems: Internals and Design Principles” book and supplementary material by W. Stallings
- “Modern Operating Systems” book and supplementary material by A. Tanenbaum
- R. Doursat and M. Yuksel from UNR