

OPERATING SYSTEMS

LECTURE #5: FILE MANAGEMENT

Written by David Goodwin
based on the lecture series of Dr. Dayou Li
and the book *Understanding Operating Systems 4th ed.*
by I.M.Flynn and A.Mclver McHoes (2006)

DEPARTMENT OF COMPUTER SCIENCE AND TECHNOLOGY,
UNIVERSITY OF BEDFORDSHIRE.



OPERATING SYSTEMS, 2013

25th FEBRUARY 2013



Outline

Lecture #5 File Management

David Goodwin
University of
Bedfordshire

Introduction

Interaction with the
file manager

Files

Physical storage
allocation

Directories

File system

Access

Data compression
summary

- 1 Introduction
- 2 Interaction with the file manager
- 3 Files
- 4 Physical storage allocation
- 5 Directories
- 6 File system
- 7 Access
- 8 Data compression
- 9 summary



Lecture #5 File Management

David Goodwin
University of
Bedfordshire

Introduction

Interaction with the
file manager

Files

Physical storage
allocation

Directories

File system

Access

Data compression
summary

3

INTRODUCTION



Introduction

Lecture #5 File Management

David Goodwin
University of
Bedfordshire

Introduction

Interaction with the
file manager

Files

Physical storage
allocation

Directories

File system

Access

Data compression
summary

4

- Responsibilities of the file manager
 - ① Keep track of where each file is stored
 - ② Use a policy that will determine where and how the files will be stored, making sure to efficiently use the available storage space and provide efficient access to the files.
 - ③ Allocate each file when a user has been cleared for access to it, and then record its use.
 - ④ Deallocate the file when the file is to be returned to storage, and communicate its availability to others who may be waiting for it.

46



Definitions

Lecture #5 File Management

David Goodwin
University of
Bedfordshire

Introduction

Interaction with the
file manager

Files

Physical storage
allocation

Directories

File system

Access

Data compression
summary

5

field is a group of related bytes that can be identified by the user with a name, type, and size.

record is a group of related fields.

file is a group of related records that contains information to be used by specific application programs to generate reports. This type of file contains data and is sometimes called a “flat file” because it has no connections to other file; unlike databases it has no dimensionality.

database appear to be a type of file, but databases are more complex. They are groups of related files that are interconnected at various levels to give flexibility of access to the data stored.

program files contain instructions.

data files contain data.

directories are listings of filenames and their attributes.



Lecture #5 File
Management

David Goodwin
University of
Bedfordshire

Introduction

Interaction with the
file manager

Files

Physical storage
allocation

Directories

File system

Access

Data compression

summary

6

INTERACTION WITH THE FILE MANAGER



Interacting with the File Manager

Lecture #5 File Management

David Goodwin
University of
Bedfordshire

Introduction

Interaction with the
file manager

Files

Physical storage
allocation

Directories

File system

Access

Data compression

summary

7

- User communicates via commands, either embedded in user's program or submitted interactively by user.
 - embedded commands are:
 - OPEN / CLOSE - pertain to the availability of a file for the program invoking it.
 - READ / WRITE - are I/O commands.
 - MODIFY - specialised WRITE command for existing data files, allows for appending records or for rewriting selected records in their original place in the file.
 - interactive commands are:
 - CREATE / DELETE - deal with the system's knowledge of the file (some systems have the first instance of SAVE command, others OPEN NEW, OPEN... FOR OUTPUT.
 - RENAME - allows users to change the name of an existing file.
 - COPY - lets users make duplicate copies of existing files.
- These commands (and many more) are designed to be simple to make them **device independent**.



Interaction with the File Manager

Lecture #5 File Management

David Goodwin
University of
Bedfordshire

Introduction

Interaction with the
file manager

Files

Physical storage
allocation

Directories

File system

Access

Data compression
summary

8

- Example - READ instruction:

- ① Move the read/write heads of the cylinder where the record is to be found.
- ② Wait for the rotational delay until the sector containing the desired record passes under the read/write head.
- ③ Activate the appropriate read/write head and read the record.
- ④ Transfer the record to main memory.
- ⑤ Send a flag to indicate that the device is free to satisfy another request.



Volume Configuration

Lecture #5 File Management

David Goodwin
University of
Bedfordshire

Introduction

Interaction with the
file manager

Files

Physical storage
allocation

Directories

File system

Access

Data compression

summary

9

- Each storage unit is considered a **volume**.
- Each volume in the system is given a name and other descriptive information contained on an easy-to-access place (innermost part of CD, first sector of outermost track of a disk pack). Volume Descriptor:
 - Creation date - date when volume was created.
 - Pointer to directory area - indicates first sector where directory is stored.
 - Pointer to file area - indicates first sector where file is stored.
 - File system code - used to detect volumes with incorrect formats.
 - Volume name - user-allocated name.
- The **Master file directory (MFD)** is stored immediately after the volume descriptor
 - lists the names and characteristics of every file contained in that volume.



Lecture #5 File Management

David Goodwin
University of
Bedfordshire

Introduction

Interaction with the
file manager

Files

Physical storage
allocation

Directories

File system

Access

Data compression
summary

10

FILES



File attributes

Lecture #5 File Management

David Goodwin
University of
Bedfordshire

Introduction

Interaction with the
file manager

Files

Physical storage
allocation

Directories

File system

Access

Data compression

summary

11

- Information about files is kept in the directory structure, which is also maintained on the disk.
 - Name – only information kept in human-readable form.
 - Identifier – non-human readable unique tag (usually a number).
 - Type – needed for systems that support different types.
 - Location – pointer to file location on device.
 - Size – current file size.
 - Protection – controls who can do reading, writing, executing.
 - Time, date, and user identification – data for protection, security, and usage monitoring.



File operations

Lecture #5 File Management

David Goodwin
University of
Bedfordshire

Introduction

Interaction with the
file manager

Files

Physical storage
allocation

Directories

File system

Access

Data compression
summary

12

- Create
- Write
- Read
- Reposition within file – file seek
- Delete
- Truncate
- $\text{Open}(F_i)$ – search the directory structure on disk for entry F_i , and move the content of entry to memory.
- $\text{Close}(F_i)$ – move the content of entry F_i in memory to directory structure on disk.



File types

Lecture #5 File Management

David Goodwin
University of
Bedfordshire

Introduction

Interaction with the
file manager

Files

Physical storage
allocation

Directories

File system

Access

Data compression
summary

13

- OS recognises and supports file types
 - helps prevent user mistakes
 - convenient by automatically doing various jobs after a command
 - must define every file type allowed and difficult to create new file types
- Commonly implemented by using an extension name
- Creator attribute used to identify the file type and invoke the creator application program
- Internal file structure can be indicated by file types



File Types

Lecture #5 File Management

David Goodwin
University of
Bedfordshire

Introduction

Interaction with the
file manager

Files

Physical storage
allocation

Directories

File system

Access

Data compression

summary

14

- Two components are common to most filenames:
 - ① relative filename
 - ② extension (called a suffix in UNIX/Linux)
- A complete filename identifies a file's **absolute filename**; being the long name that includes path information.
- there are restrictions on filename with each operating system:
 - MS-DOS - allows names from one to eight alphanumeric characters, but without spaces
 - most modern operating systems do not have restrictions on filenames.
- An extension usually 2 or 3 characters long and is separated by a period (.). The purpose is to identify the file type or its contents.
- NOTE: some extensions are associated with more than one application



File types – name and extension

Lecture #5 File Management

David Goodwin
University of
Bedfordshire

Introduction

Interaction with the
file manager

Files

Physical storage
allocation

Directories

File system

Access

Data compression
summary

15

file type	usual extension	function
executable	exe, com, bin or none	read to run machine- language program
object	obj, o	compiled, machine language, not linked
source code	c, cc, java, pas, asm, a	source code in various languages
batch	bat, sh	commands to the command interpreter
text	txt, doc	textual data, documents
word processor	wp, tex, rrf, doc	various word-processor formats
library	lib, a, so, dll, mpeg, mov, rm	libraries of routines for programmers
print or view	arc, zip, tar	ASCII or binary file in a format for printing or viewing
archive	arc, zip, tar	related files grouped into one file, sometimes com- pressed, for archiving or storage
multimedia	mpeg, mov, rm	binary file containing audio or A/V information



Access methods

Lecture #5 File Management

David Goodwin
University of
Bedfordshire

Introduction

Interaction with the
file manager

Files

Physical storage
allocation

Directories

File system

Access

Data compression

summary

16

- Sequential access
 - Easiest
 - File searched from its beginning until record is found
- Direct (relative/random) access
 - Only on direct access storage devices
 - Records identified by their relative address to the beginning of the file
- Indexed access
 - Combines the best of sequential and direct access



Access methods - direct access example

Lecture #5 File Management

David Goodwin
University of
Bedfordshire

Introduction

Interaction with the
file manager

Files

Physical storage
allocation

Directories

File system

Access

Data compression
summary

17

- Records identified by logical address, then relative to the beginning of the file.
- User identifies a field in the record format and designates it as the **key field** because it uniquely identifies each record.
- the program used to store the data follows a set of instructions called a **hashing algorithm**, that transforms each key into a number, the record's logical address.
- This is given to the File Manager, which takes the necessary steps to translate the logical address into a physical address (cylinder, surface, and record numbers), preserving the file organisation.



Lecture #5 File Management

David Goodwin
University of
Bedfordshire

Introduction

Interaction with the
file manager

Files

Physical storage
allocation

Directories

File system

Access

Data compression
summary

18

PHYSICAL STORAGE ALLOCATION



Physical storage allocation

Lecture #5 File Management

David Goodwin
University of
Bedfordshire

Introduction

Interaction with the
file manager

Files

Physical storage
allocation

Directories

File system

Access

Data compression

summary

19

- Contiguous storage
 - records stored one after another
 - any record can be found and read by knowing its starting address and size
 - ease of direct access, every part of the file is stored in the same compact area
 - file cant be expanded unless there is empty space immediately following it - also causing fragmentation



Physical storage allocation

Lecture #5 File Management

David Goodwin
University of
Bedfordshire

Introduction

Interaction with the
file manager

Files

Physical storage
allocation

Directories

File system

Access

Data compression

summary

20

- Noncontiguous storage
 - files use any space available on disk
 - records stored in a contiguous manner only if there's enough empty space
 - any remaining records, and all other additions to the file, are stored in other sections of the disk
 - sometimes called **extents** of the file - linked together with pointers in one of two ways
 - at storage level - each extent points to the next one in sequence
 - at directory level - each extent is listed with its physical address, size, and a pointer to the next extent. A null pointer indicates it's the last one.
 - doesn't support direct access

46



Physical storage allocation

Lecture #5 File Management

David Goodwin
University of
Bedfordshire

Introduction

Interaction with the
file manager

Files

Physical storage
allocation

Directories

File system

Access

Data compression

summary

21

- Indexed storage
 - brings together into an index block, the pointers linking every extent of that file
 - every file has its own index block
 - doesn't improve storage
 - supports direct access



Lecture #5 File
Management

David Goodwin
University of
Bedfordshire

Introduction

Interaction with the
file manager

Files

Physical storage
allocation

Directories

File system

Access

Data compression
summary

22

DIRECTORIES



Directory structure

Lecture #5 File Management

David Goodwin
University of
Bedfordshire

Introduction

Interaction with the
file manager

Files

Physical storage
allocation

Directories

File system

Access

Data compression

summary

23

- Directory is a collection of nodes containing information about all files.
 - Users are concerned with only the logical directory and its structure.
 - Users can ignore the problems of physically allocating file space.
 - Multiple operating systems are allowed on a computer system.
 - A disk can be split into partitions or volumes, viewed as virtual disks.
 - Separate disks can be grouped into a large logical structure.
 - Each partition contains information about files in it, kept in entries in a device directory or volume table of contents.
- Both the directory structure and the files reside on disk.

46



A typical file system organisation

Lecture #5 File Management

David Goodwin
University of
Bedfordshire

Introduction

Interaction with the
file manager

Files

Physical storage
allocation

Directories

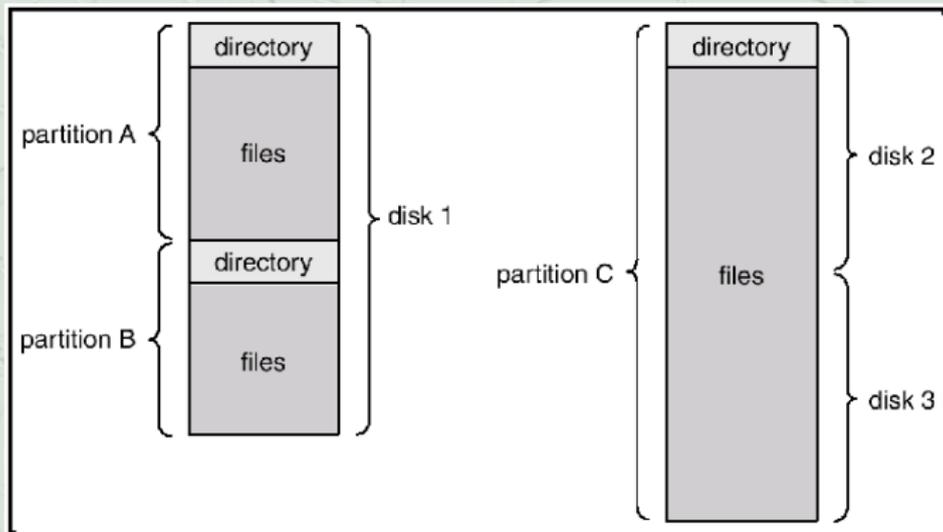
File system

Access

Data compression

summary

24





Information on a device directory

Lecture #5 File Management

David Goodwin
University of
Bedfordshire

Introduction

Interaction with the
file manager

Files

Physical storage
allocation

Directories

File system

Access

Data compression
summary

25

- Name
- Type
- Address
- Current length
- Maximum length
- Date last accessed (for archival)
- Date last updated (for dump)
- Owner ID (who pays)
- Protection information



Operations performed on a directory

Lecture #5 File Management

David Goodwin
University of
Bedfordshire

Introduction

Interaction with the
file manager

Files

Physical storage
allocation

Directories

File system

Access

Data compression
summary

26

- Search for a file
- Create a file
- Delete a file
- List a directory
- Rename a file
- Traverse the file system



Effective organisation of a directory

Lecture #5 File Management

David Goodwin
University of
Bedfordshire

Introduction

Interaction with the
file manager

Files

Physical storage
allocation

Directories

File system

Access

Data compression
summary

27

- **Efficiency** – locating a file quickly
- **Naming** – convenient to users
 - Different users can have the same name for different files.
 - The same file can have several different names.
- **Grouping** – logical grouping of files by properties, (e.g., all Java programs, all games, ...)



Single-level directory

Lecture #5 File Management

David Goodwin
University of
Bedfordshire

Introduction

Interaction with the
file manager

Files

Physical storage
allocation

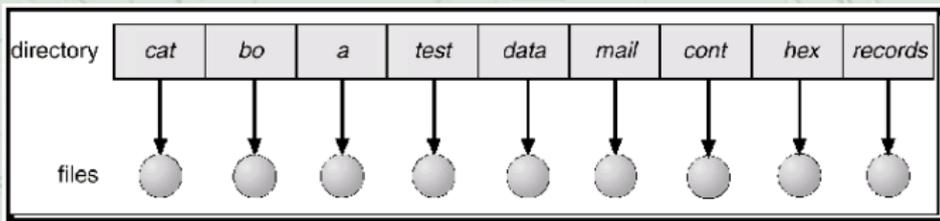
Directories

File system

Access

Data compression
summary

- A single-level directory for all users



- Naming problem
- Grouping problem

28



Tree-structured directories

Lecture #5 File Management

David Goodwin
University of
Bedfordshire

Introduction

Interaction with the
file manager

Files

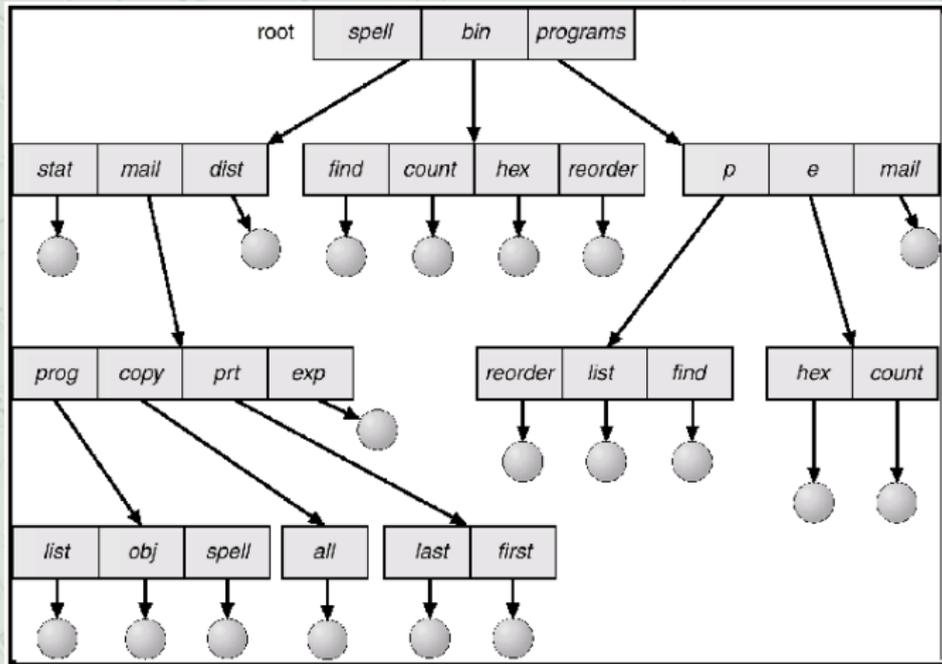
Physical storage
allocation

Directories

File system

Access

Data compression
summary



29



Tree-structured directories

Lecture #5 File Management

David Goodwin
University of
Bedfordshire

Introduction

Interaction with the
file manager

Files

Physical storage
allocation

Directories

File system

Access

Data compression
summary

30

- Efficient searching
- Grouping capability
- Current directory (working directory)
cd /spell/mail/prog
type list
- Absolute path name begins at the root
- Relative path name defines a path from current directory
- Creating a new file done in current directory
- Delete a file
rm <file-name>
- Creating a new subdirectory is done in current directory
mkdir <dir-name>

46



Acyclic graph directories

Lecture #5 File Management

David Goodwin
University of
Bedfordshire

Introduction

Interaction with the
file manager

Files

Physical storage
allocation

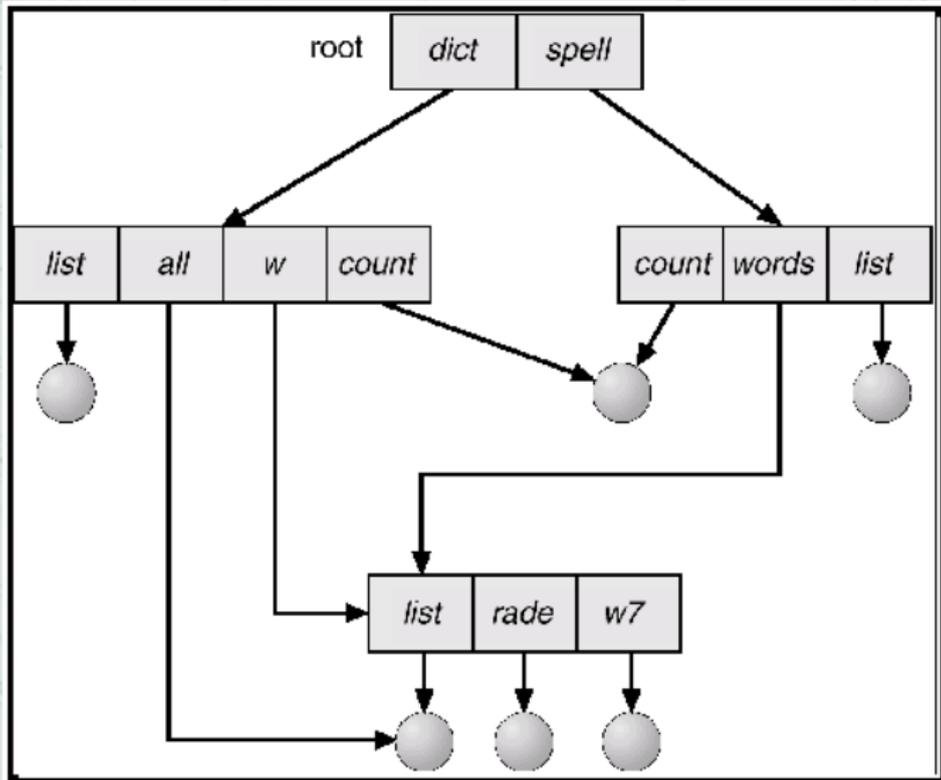
Directories

File system

Access

Data compression
summary

31





Acyclic graph directories

Lecture #5 File Management

David Goodwin
University of
Bedfordshire

Introduction

Interaction with the
file manager

Files

Physical storage
allocation

Directories

File system

Access

Data compression
summary

32

- There are shared subdirectories and files.
- Different absolute path names (aliasing).
- If words deletes list → dangling pointer.

Solutions:

- Backpointers, so we can delete all pointers.
- Backpointers using a daisy chain organisation.
- Entry-hold-count solution.



General graph directories

Lecture #5 File Management

David Goodwin
University of
Bedfordshire

Introduction

Interaction with the
file manager

Files

Physical storage
allocation

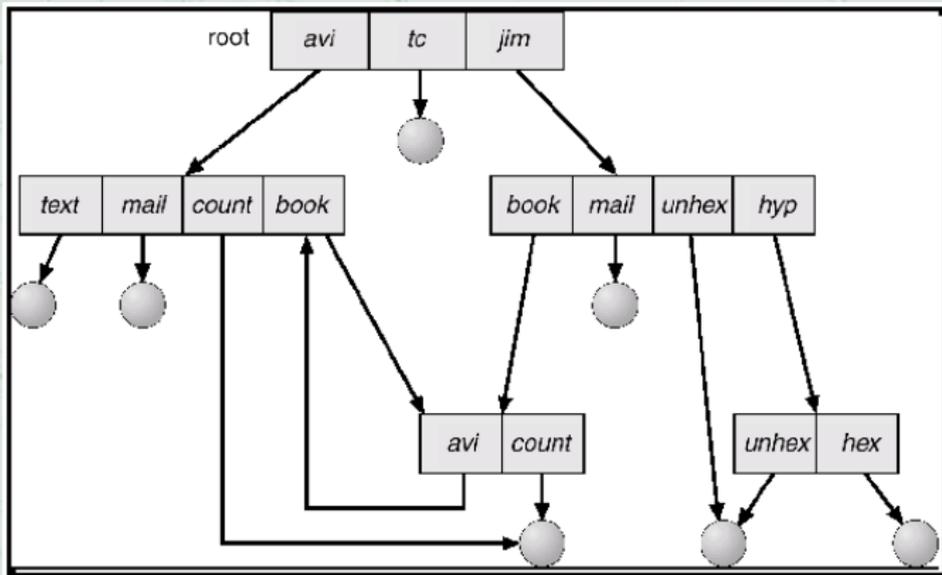
Directories

File system

Access

Data compression

summary



33



General graph directories

Lecture #5 File Management

David Goodwin
University of
Bedfordshire

Introduction

Interaction with the
file manager

Files

Physical storage
allocation

Directories

File system

Access

Data compression
summary

34

- How do we guarantee no cycles?
 - Allow only links to file not subdirectories.
 - Garbage collection.
 - Every time a new link is added, a cycle detection algorithm is used to determine whether it is OK.



Lecture #5 File Management

David Goodwin
University of
Bedfordshire

Introduction

Interaction with the
file manager

Files

Physical storage
allocation

Directories

File system

Access

Data compression
summary

FILE SYSTEM

35



File system mounting

Lecture #5 File Management

David Goodwin
University of
Bedfordshire

Introduction

Interaction with the
file manager

Files

Physical storage
allocation

Directories

File system

Access

Data compression
summary

- A file system must be mounted before it can be accessed.
- An unmounted file system (device) is mounted at a mount point.
- Typical example - UNIX systems
e.g. user's directories can be mounted as
/home or /users
and thus a user's directory may become
/home/user01 or /users/user01
where /home or /users is the mount point

36



File system mounting

Lecture #5 File Management

David Goodwin
University of
Bedfordshire

Introduction

Interaction with the
file manager

Files

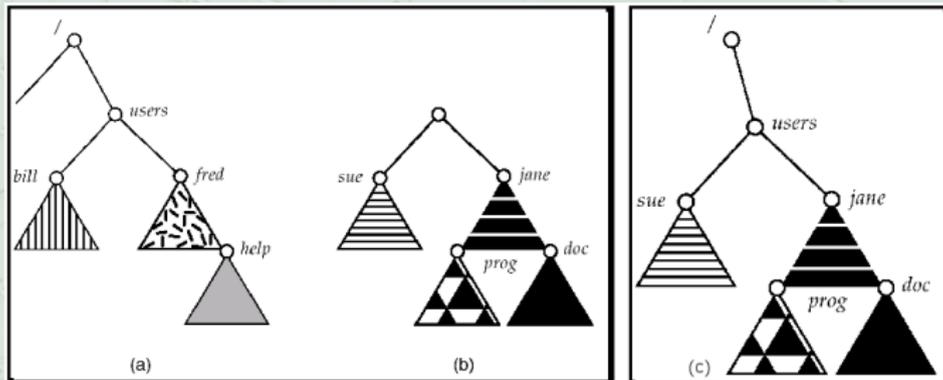
Physical storage
allocation

Directories

File system

Access

Data compression
summary



- a) Existing
- b) Unmounted partition
- c) Mounted at users/



File sharing

Lecture #5 File Management

David Goodwin
University of
Bedfordshire

Introduction

Interaction with the
file manager

Files

Physical storage
allocation

Directories

File system

Access

Data compression
summary

38

- Sharing of files on multi-user systems is desirable.
- Sharing may be done through a protection scheme.
- On distributed systems, files may be shared across a network.
- Network File System (NFS) is a common distributed file-sharing method.



Protection

Lecture #5 File Management

David Goodwin
University of
Bedfordshire

Introduction

Interaction with the
file manager

Files

Physical storage
allocation

Directories

File system

Access

Data compression
summary

- File owner/creator should be able to control:
 - What can be done
 - By whom
- Types of access
 - Read
 - Write
 - Execute
 - Append
 - Delete
 - List

39



Lecture #5 File Management

David Goodwin
University of
Bedfordshire

Introduction

Interaction with the
file manager

Files

Physical storage
allocation

Directories

File system

Access

Data compression
summary

40

ACCESS



Access lists and groups

Lecture #5 File Management

David Goodwin
University of
Bedfordshire

Introduction

Interaction with the
file manager

Files

Physical storage
allocation

Directories

File system

Access

Data compression

summary

41

- Mode of access: read, write, execute
- Three classes of users
 - ① a) owner access 7 → 1 1 1 (RWX)
 - ② b) group access 6 → 1 1 0 (RWX)
 - ③ c) public access 1 → 0 0 1 (RWX)
- Ask manager to create a group (unique name), say G, and add some users to the group.
- For a particular file (say game) or subdirectory, define an appropriate access.

```
owner group public
   |     |     |
   +-----+
   chmod 761 game
```

Attach a group to a file

```
chgrp G game
```



Lecture #5 File Management

David Goodwin
University of
Bedfordshire

Introduction

Interaction with the
file manager

Files

Physical storage
allocation

Directories

File system

Access

Data compression
summary

42

DATA COMPRESSION



Data compression

Lecture #5 File Management

David Goodwin
University of
Bedfordshire

Introduction

Interaction with the
file manager

Files

Physical storage
allocation

Directories

File system

Access

Data compression

summary

43

- Technique to save space in files
- Records with repeated characters
ADAMSbbbbbbbbbb → ADMSb10
300000000 → 3#8
- Repeated terms represented with symbols
Smith, Betty; Smith, Gino; Smithberger, John; Smithbren, Ali
→
Smith, Betty; 7Gino; 5berger, John; 6ren, Ali
- Front-end compression
- Trade-off - Storage space saving at cost of processing time



Lecture #5 File Management

David Goodwin
University of
Bedfordshire

Introduction

Interaction with the
file manager

Files

Physical storage
allocation

Directories

File system

Access

Data compression

summary

44

SUMMARY



Key Terms

Lecture #5 File Management

David Goodwin
University of
Bedfordshire

Introduction

Interaction with the
file manager

Files

Physical storage
allocation

Directories

File system

Access

Data compression

summary

45

absolute filename
contiguous storage
current directory
data compression
data file
database
device independent
direct record organisation
directory
extension
field
file



Key Terms

Lecture #5 File Management

David Goodwin
University of
Bedfordshire

Introduction

Interaction with the
file manager

Files

Physical storage
allocation

Directories

File system

Access

Data compression

summary

46

file descriptor
hashing algorithm
indexed sequential record organisation
key field
logical address
master file directory
noncontiguous storage
program file
record
relative address
relative filename
sequential record organisation
volume