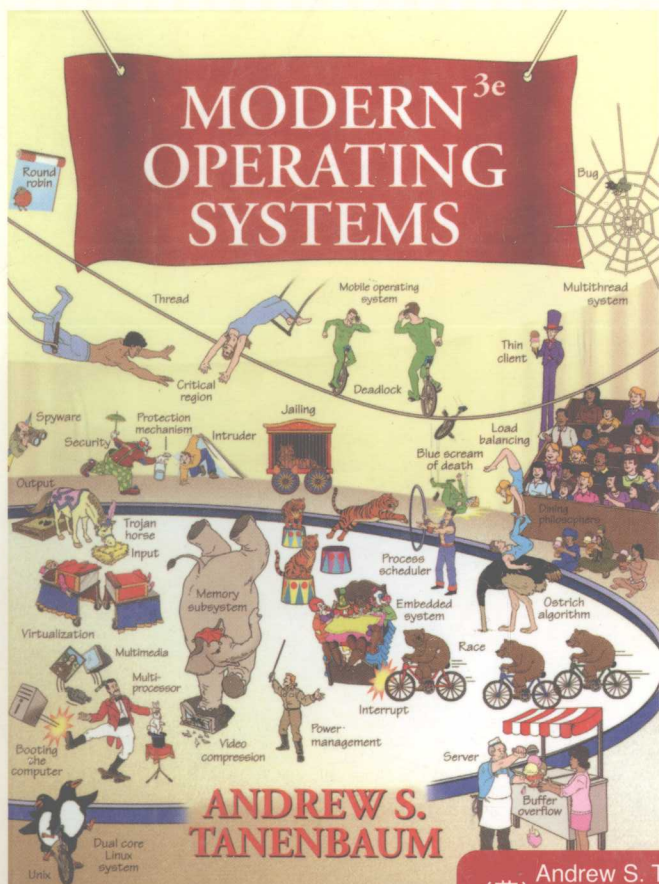


# 现代操作系统

(英文版 · 第3版)



Andrew S. Tanenbaum 著  
(荷) Vrije大学

TP316/Y57

2009.

经 典 原 版 书 库

# 现代操作系统

(英文版·第3版)

**Modern Operating Systems**

(Third Edition)

(荷) Andrew S. Tanenbaum 著  
Vrije大学



机械工业出版社  
China Machine Press

English reprint edition copyright © 2009 by Pearson Education Asia Limited and China Machine Press.

Original English language title: *Modern Operating Systems, Third Edition* (ISBN 978-0-13-600663-3) by Andrew S. Tanenbaum, Copyright © 2008 Pearson Education, Inc.

All rights reserved.

Published by arrangement with the original publisher, Pearson Education, Inc., publishing as Prentice Hall.

For sale and distribution in the People's Republic of China exclusively (except Taiwan, Hong Kong SAR and Macau SAR).

本书英文影印版由Pearson Education Asia Ltd.授权机械工业出版社独家出版。未经出版者书面许可，不得以任何方式复制或抄袭本书内容。

仅限于中华人民共和国境内（不包括中国香港、澳门特别行政区和中国台湾地区）销售发行。

本书封面贴有Pearson Education（培生教育出版集团）激光防伪标签，无标签者不得销售。

版权所有，侵权必究。

本书法律顾问 北京市展达律师事务所

本书版权登记号：图字：01-2009-1357

图书在版编目（CIP）数据

现代操作系统（英文版·第3版）/（荷）塔嫩鲍姆（Tanenbaum, A. S.）著.  
—北京：机械工业出版社，2009.3

（经典原版书库）

书名原文：Modern Operating Systems, Third Edition

ISBN 978-7-111-26527-6

I. 现… II. 塔… III. 操作系统—英文 IV. TP316

中国版本图书馆CIP数据核字（2009）第031374号

机械工业出版社（北京市西城区百万庄大街22号 邮政编码 100037）

责任编辑：迟振春

北京京师印务有限公司印刷

2009年3月第1版第1次印刷

150mm×214mm · 34.5印张

标准书号：ISBN 978-7-111-26527-6

定价：69.00元

凡购本书，如有倒页、脱页、缺页，由本社发行部调换  
本社购书热线：（010）68326294

# 出版者的话

文艺复兴以降，源远流长的科学精神和逐步形成的学术规范，使西方国家在自然科学的各个领域中取得了垄断性的优势；也正是这样的传统，使美国在信息技术发展的六十多年间名家辈出、独领风骚。在商业化的进程中，美国的产业界与教育界越来越紧密地结合，计算机学科中的许多泰山北斗同时身处科研和教学的最前线，由此而产生的经典科学著作，不仅肇划了研究的范畴，还揭示了学术的源变，既遵循学术规范，又自有学者个性，其价值并不会因年月的流逝而减退。

近年，在全球信息化大潮的推动下，我国的计算机产业发展迅猛，对专业人才的需求日益迫切。这对计算机教育界和出版界都既是机遇，也是挑战；而专业教材的建设在教育战略上显得举足轻重。在我国信息技术发展时间较短的现状下，美国等发达国家在其计算机科学发展的几十年间积淀和发展的经典教材仍有许多值得借鉴之处。因此，引进一批国外优秀计算机教材将对我国计算机教育事业的发展起到积极的推动作用，也是与世界接轨、建设真正的世界一流大学的必由之路。

机械工业出版社华章分社较早意识到“出版要为教育服务”。自1998年开始，华章分社就将工作重点放在了遴选、移译国外优秀教材上。经过多年的不懈努力，我们与Pearson, McGraw-Hill, Elsevier, MIT, John Wiley & Sons, Cengage等世界著名出版公司建立了良好的合作关系，从他们现有的数百种教材中甄选出Andrew S. Tanenbaum, Bjarne Stroustrup, Brian W. Kernighan, Dennis Ritchie, Jim Gray, Alfred V. Aho, John E. Hopcroft, Jeffrey D. Ullman, Abraham Silberschatz, William Stallings, Donald E. Knuth, John L. Hennessy, Larry L. Peterson等大师名家的一批经典作品，以“计算机科学丛书”为总称出版，供读者学习、研究及珍藏。大理石纹理的封面，也正体现了这套丛书的品位和格调。

“计算机科学丛书”的出版工作得到了国内外学者的鼎力襄助，国内的专家不仅提供了中肯的选题指导，还不辞劳苦地担任了翻译和审校的工作。

作；而原书的作者也相当关注其作品在中国的传播，有的还专程为其书的中译本作序。迄今，“计算机科学丛书”已经出版了近两百个品种，这些书籍在读者中树立了良好的口碑，并被许多高校采用为正式教材和参考书籍。其影印版“经典原版书库”作为姊妹篇也被越来越多实施双语教学的学校所采用。

权威的作者、经典的教材、一流的译者、严格的审校、精细的编辑，这些因素使我们的图书有了质量的保证。随着计算机科学与技术专业学科建设的不断完善和教材改革的逐渐深化，教育界对国外计算机教材的需求和应用都将步入一个新的阶段，我们的目标是尽善尽美，而反馈的意见正是我们达到这一终极目标的重要帮助。华章分社欢迎老师和读者对我们的工作提出建议或给予指正，我们的联系方式如下：

华章网站：[www.hzbook.com](http://www.hzbook.com)

电子邮件：[hzsj@hzbook.com](mailto:hzsj@hzbook.com)

联系电话：(010) 88379604

联系地址：北京市西城区百万庄南街1号

邮政编码：100037



# PREFACE

The third edition of this book differs from the second edition in numerous ways. To start with, the chapters have been reordered to place the central material at the beginning. There is also now more of a focus on the operating system as the creator of abstractions. Chapter 1, which has been heavily updated, introduces all the concepts. Chapter 2 is about the abstraction of the CPU into multiple processes. Chapter 3 is about the abstraction of physical memory into address spaces (virtual memory). Chapter 4 is about the abstraction of the disk into files. Together, processes, virtual address spaces, and files are the key concepts that operating systems provide, so these chapters are now placed earlier than they previously had been.

Chapter 1 has been heavily modified and updated in many places. For example, an introduction to the C programming language and the C run-time model is given for readers familiar only with Java.

In Chapter 2, the discussion of threads has been revised and expanded reflecting their new importance. Among other things, there is now a section on IEEE standard Pthreads.

Chapter 3, on memory management, has been reorganized to emphasize the idea that one of the key functions of an operating system is to provide the abstraction of a virtual address space for each process. Older material on memory management in batch systems has been removed, and the material on the implementation of paging has been updated to focus on the need to make it handle the larger address spaces now common and also the need for speed.

Chapters 4–7 have been updated, with older material removed and some new material added. The sections on current research in these chapters have been rewritten from scratch. Many new problems and programming exercises have been added.

Chapter 8 has been updated, including some material on multicore systems. A whole new section on virtualization technology, hypervisors, and virtual machines, has been added with VMware used as an example.

Chapter 9 has been heavily revised and reorganized, with considerable new material on exploiting code bugs, malware, and defenses against them.

Chapter 10, on Linux, is a revision of the old Chapter 10 (on UNIX and Linux). The focus is clearly on Linux now, with a great deal of new material.

Chapter 11, on Windows Vista, is a major revision of the old Chap. 11 (on Windows 2000). It brings the treatment of Windows completely up to date.

Chapter 12 is new. I felt that embedded operating systems, such as those found on cell phones and PDAs, are neglected in most textbooks, despite the fact that there are more of them out there than there are PCs and notebooks. This edition remedies this problem, with an extended discussion of Symbian OS, which is widely used on Smart Phones.

Chapter 13, on operating system design, is largely unchanged from the second edition.

Numerous teaching aids for this book are available. Instructor supplements can be found at [www.prenhall.com/tanenbaum](http://www.prenhall.com/tanenbaum). They include PowerPoint sheets, software tools for studying operating systems, lab experiments for students, simulators, and more material for use in operating systems courses. Instructors using this book in a course should definitely take a look.

In addition, instructors should examine GOAL (Gradiance Online Accelerated Learning), Pearson's premier online homework and assessment system. GOAL is designed to minimize student frustration while providing an interactive teaching experience outside the classroom. With GOAL's immediate feedback, hints, and pointers that map back to the textbook, students will have a more efficient and effective learning experience. GOAL delivers immediate assessment and feedback via two kinds of assignments: multiple choice Homework exercises and interactive Lab work.

The multiple-choice homework consists of a set of multiple choice questions designed to test student knowledge of a solved problem. When answers are graded as incorrect, students are given a hint and directed back to a specific section in the course textbook for helpful information.

The interactive Lab Projects in GOAL, unlike syntax checkers and compilers, check for both syntactic and semantic errors. GOAL determines if the student's program runs but more importantly, when checked against a hidden data set, verifies that it returns the correct result. By testing the code and providing immediate feedback, GOAL lets you know exactly which concepts the students have grasped and which ones need to be revisited.



Instructors should contact their local Pearson Sales Representative for sales and ordering information for the GOAL Student Access Code and Modern Operating Systems, 3e Value Pack (ISBN: 0135013011).

A number of people helped me with this revision. First and foremost I want to thank my editor, Tracy Dunkelberger. This is my 18th book and I have worn out a lot of editors in the process. Tracy went above and beyond the call of duty on this one, doing things like finding contributors, arranging numerous reviews, helping with all the supplements, dealing with contracts, interfacing to PH, coordinating a great deal of parallel processing, generally making sure things happened on time, and more. She also was able to get me to make and keep to a very tight schedule in order to get this book out in time. And all this while she remained chipper and cheerful, despite many other demands on her time. Thank you, Tracy. I appreciate it a lot.

Ada Gavrilovska of Georgia Tech, who is an expert on Linux internals, updated Chap. 10 from one on UNIX (with a focus on FreeBSD) to one more about Linux, although much of the chapter is still generic to all UNIX systems. Linux is more popular among students than FreeBSD, so this is a valuable change.

Dave Probert of Microsoft updated Chap. 11 from one on Windows 2000 to one on Windows Vista. While they have some similarities, they also have significant differences. Dave has a great deal of knowledge of Windows and enough vision to tell the difference between places where Microsoft got it right and where it got it wrong. The book is much better as a result of his work.

Mike Jipping of Hope College wrote the chapter on Symbian OS. Not having anything on embedded real-time systems was a serious omission in the book, and thanks to Mike that problem has been solved. Embedded real-time systems are becoming increasingly important in the world and this chapter provides an excellent introduction to the subject.

Unlike Ada, Dave, and Mike, who each focused on one chapter, Shivakant Mishra of the University of Colorado at Boulder was more like a distributed system, reading and commenting on many chapters and also supplying a substantial number of new exercises and programming problems throughout the book.

Hugh Lauer also gets a special mention. When we asked him for ideas about how to revise the second edition, we weren't expecting a report of 23 single-spaced pages, but that is what we got. Many of the changes, such as the new emphasis on the abstractions of processes, address spaces, and files are due to his input.

I would also like to thank other people who helped me in many ways, including suggesting new topics to cover, reading the manuscript carefully, making supplements, and contributing new exercises. Among them are Steve Armstrong, Jeffrey Chastine, John Connelly, Mischa Geldermans, Paul Gray, James Griffioen, Jorrit Herder, Michael Howard, Suraj Kothari, Roger Kraft, Trudy Levine, John Masiyowski, Shivakant Mishra, Rudy Pait, Xiao Qin, Mark Russinovich, Krishna Sivalingam, Leendert van Doorn, and Ken Wong.



The people at Prentice Hall have been friendly and helpful as always, especially including Irwin Zucker and Scott Disanno in production and David Alick, ReeAnne Davies, and Melinda Haggerty in editorial.

Finally, last but not least, Barbara and Marvin are still wonderful, as usual, each in a unique and special way. And of course, I would like to thank Suzanne for her love and patience, not to mention all the *druiven* and *kersen*, which have replaced the *sinaasappelsap* in recent times.

Andrew S. Tanenbaum

## ABOUT THE AUTHOR

Andrew S. Tanenbaum has an S.B. degree from M.I.T. and a Ph.D. from the University of California at Berkeley. He is currently a Professor of Computer Science at the Vrije Universiteit in Amsterdam, The Netherlands, where he heads the Computer Systems Group. He was formerly Dean of the Advanced School for Computing and Imaging, an interuniversity graduate school doing research on advanced parallel, distributed, and imaging systems. He is now an Academy Professor of the Royal Netherlands Academy of Arts and Sciences, which has saved him from turning into a bureaucrat.

In the past, he has done research on compilers, operating systems, networking, local-area distributed systems and wide-area distributed systems that scale to a billion users. His main focus now is doing research on reliable and secure operating systems. These research projects have led to over 140 refereed papers in journals and conferences. Prof. Tanenbaum has also authored or co-authored five books which have now appeared in 18 editions. The books have been translated into 21 languages, ranging from Basque to Thai and are used at universities all over the world. In all, there are 130 versions (language + edition combinations).

Prof. Tanenbaum has also produced a considerable volume of software. He was the principal architect of the Amsterdam Compiler Kit, a widely-used toolkit for writing portable compilers. He was also one of the principal designers of Amoeba, an early distributed system used on a collection of workstations connected by a LAN and of Globe, a wide-area distributed system.

He is also the author of MINIX, a small UNIX clone initially intended for use in student programming labs. It was the direct inspiration for Linux and the platform on which Linux was initially developed. The current version of MINIX, called MINIX 3, is now focused on being an extremely reliable and secure operating system. Prof. Tanenbaum will consider his work done when no computer is equipped with a reset button. MINIX 3 is an on-going open-source project to which you are invited to contribute. Go to [www.minix3.org](http://www.minix3.org) to download a free copy and find out what is happening.

Prof. Tanenbaum's Ph.D. students have gone on to greater glory after graduating. He is very proud of them. In this respect he resembles a mother hen.

Tanenbaum is a Fellow of the ACM, a Fellow of the IEEE, and a member of the Royal Netherlands Academy of Arts and Sciences. He has also won numerous scientific prizes, including:

- the 2007 IEEE James H. Mulligan, Jr. Education Medal
- the 2003 TAA McGuffey Award for Computer Science and Engineering
- the 2002 TAA Texty Award for Computer Science and Engineering
- the 1997 ACM/SIGCSE Award for Outstanding Contributions to Computer
- the 1994 ACM Karl V. Karlstrom Outstanding Educator Award

He is also listed in *Who's Who in the World*. His home page on the World Wide Web can be found at URL <http://www.cs.vu.nl/~ast/>.

# CONTENTS

## PREFACE

v

## 1 INTRODUCTION

1

### 1.1 WHAT IS AN OPERATING SYSTEM? 3

1.1.1 The Operating System as an Extended Machine 4

1.1.2 The Operating System as a Resource Manager 6

### 1.2 HISTORY OF OPERATING SYSTEMS 7

1.2.1 The First Generation (1945–55) Vacuum Tubes 7

1.2.2 The Second Generation (1955–65) Transistors and Batch Systems 8

1.2.3 The Third Generation (1965–1980) ICs and Multiprogramming 10

1.2.4 The Fourth Generation (1980–Present) Personal Computers 15

### 1.3 COMPUTER HARDWARE REVIEW 19

1.3.1 Processors 19

1.3.2 Memory 23

1.3.3 Disks 26

1.3.4 Tapes 27

1.3.5 I/O Devices 27

1.3.6 Buses 30

1.3.7 Booting the Computer 33

1.4	THE OPERATING SYSTEM ZOO	33
1.4.1	Mainframe Operating Systems	34
1.4.2	Server Operating Systems	34
1.4.3	Multiprocessor Operating Systems	34
1.4.4	Personal Computer Operating Systems	35
1.4.5	Handheld Computer Operating Systems	35
1.4.6	Embedded Operating Systems	35
1.4.7	Sensor Node Operating Systems	36
1.4.8	Real-Time Operating Systems	36
1.4.9	Smart Card Operating Systems	37
1.5	OPERATING SYSTEM CONCEPTS	37
1.5.1	Processes	38
1.5.2	Address Spaces	40
1.5.3	Files	40
1.5.4	Input/Output	43
1.5.5	Protection	44
1.5.6	The Shell	44
1.5.7	Ontogeny Recapitulates Phylogeny	46
1.6	SYSTEM CALLS	49
1.6.1	System Calls for Process Management	52
1.6.2	System Calls for File Management	56
1.6.3	System Calls for Directory Management	57
1.6.4	Miscellaneous System Calls	58
1.6.5	The Windows Win32 API	59
1.7	OPERATING SYSTEM STRUCTURE	62
1.7.1	Monolithic Systems	62
1.7.2	Layered Systems	63
1.7.3	Microkernels	64
1.7.4	Client-Server Model	67
1.7.5	Virtual Machines	67
1.7.6	Exokernels	71
1.8	THE WORLD ACCORDING TO C	72
1.8.1	The C Language	72
1.8.2	Header Files	73
1.8.3	Large Programming Projects	74
1.8.4	The Model of Run Time	75
1.9	RESEARCH ON OPERATING SYSTEMS	76

- 1.10 OUTLINE OF THE REST OF THIS BOOK 77
- 1.11 METRIC UNITS 78
- 1.12 SUMMARY 79

## **2 PROCESSES AND THREADS**

**83**

- 2.1 PROCESSES 83
  - 2.1.1 The Process Model 84
  - 2.1.2 Process Creation 86
  - 2.1.3 Process Termination 88
  - 2.1.4 Process Hierarchies 89
  - 2.1.5 Process States 90
  - 2.1.6 Implementation of Processes 91
  - 2.1.7 Modeling Multiprogramming 93
- 2.2 THREADS 95
  - 2.2.1 Thread Usage 95
  - 2.2.2 The Classical Thread Model 100
  - 2.2.3 POSIX Threads 104
  - 2.2.4 Implementing Threads in User Space 106
  - 2.2.5 Implementing Threads in the Kernel 109
  - 2.2.6 Hybrid Implementations 110
  - 2.2.7 Scheduler Activations 111
  - 2.2.8 Pop-Up Threads 112
  - 2.2.9 Making Single-Threaded Code Multithreaded 114
- 2.3 INTERPROCESS COMMUNICATION 117
  - 2.3.1 Race Conditions 117
  - 2.3.2 Critical Regions 119
  - 2.3.3 Mutual Exclusion with Busy Waiting 120
  - 2.3.4 Sleep and Wakeup 125
  - 2.3.5 Semaphores 128
  - 2.3.6 Mutexes 130
  - 2.3.7 Monitors 134
  - 2.3.8 Message Passing 140
  - 2.3.9 Barriers 144

- 2.4 SCHEDULING 145
  - 2.4.1 Introduction to Scheduling 145
  - 2.4.2 Scheduling in Batch Systems 152
  - 2.4.3 Scheduling in Interactive Systems 154
  - 2.4.4 Scheduling in Real-Time Systems 160
  - 2.4.5 Policy versus Mechanism 161
  - 2.4.6 Thread Scheduling 162
- 2.5 CLASSICAL IPC PROBLEMS 163
  - 2.5.1 The Dining Philosophers Problem 164
  - 2.5.2 The Readers and Writers Problem 167
- 2.6 RESEARCH ON PROCESSES AND THREADS 168
- 2.7 SUMMARY 169

## **3 MEMORY MANAGEMENT**

**175**

- 3.1 NO MEMORY ABSTRACTION 176
- 3.2 A MEMORY ABSTRACTION: ADDRESS SPACES 179
  - 3.2.1 The Notion of an Address Space 180
  - 3.2.2 Swapping 181
  - 3.2.3 Managing Free Memory 184
- 3.3 VIRTUAL MEMORY 188
  - 3.3.1 Paging 189
  - 3.3.2 Page Tables 193
  - 3.3.3 Speeding Up Paging 194
  - 3.3.4 Page Tables for Large Memories 198
- 3.4 PAGE REPLACEMENT ALGORITHMS 201
  - 3.4.1 The Optimal Page Replacement Algorithm 202
  - 3.4.2 The Not Recently Used Page Replacement Algorithm 203
  - 3.4.3 The First-In, First-Out (FIFO) Page Replacement Algorithm 204
  - 3.4.4 The Second-Chance Page Replacement Algorithm 204
  - 3.4.5 The Clock Page Replacement Algorithm 205
  - 3.4.6 The Least Recently Used (LRU) Page Replacement Algorithm 206
  - 3.4.7 Simulating LRU in Software 207
  - 3.4.8 The Working Set Page Replacement Algorithm 209

3.4.9	The WSClock Page Replacement Algorithm	213
3.4.10	Summary of Page Replacement Algorithms	215
3.5	DESIGN ISSUES FOR PAGING SYSTEMS	216
3.5.1	Local versus Global Allocation Policies	216
3.5.2	Load Control	218
3.5.3	Page Size	219
3.5.4	Separate Instruction and Data Spaces	221
3.5.5	Shared Pages	221
3.5.6	Shared Libraries	223
3.5.7	Mapped Files	225
3.5.8	Cleaning Policy	226
3.5.9	Virtual Memory Interface	226
3.6	IMPLEMENTATION ISSUES	227
3.6.1	Operating System Involvement with Paging	227
3.6.2	Page Fault Handling	228
3.6.3	Instruction Backup	229
3.6.4	Locking Pages in Memory	230
3.6.5	Backing Store	231
3.6.6	Separation of Policy and Mechanism	233
3.7	SEGMENTATION	234
3.7.1	Implementation of Pure Segmentation	237
3.7.2	Segmentation with Paging: MULTICS	238
3.7.3	Segmentation with Paging: The Intel Pentium	242
3.8	RESEARCH ON MEMORY MANAGEMENT	247
3.9	SUMMARY	248

## **4 FILE SYSTEMS**

**255**

4.1	FILES	257
4.1.1	File Naming	257
4.1.2	File Structure	259
4.1.3	File Types	260
4.1.4	File Access	262
4.1.5	File Attributes	263



- 4.1.6 File Operations 264
- 4.1.7 An Example Program Using File System Calls 265
- 4.2 DIRECTORIES 268
  - 4.2.1 Single-Level Directory Systems 268
  - 4.2.2 Hierarchical Directory Systems 268
  - 4.2.3 Path Names 269
  - 4.2.4 Directory Operations 272
- 4.3 FILE SYSTEM IMPLEMENTATION 273
  - 4.3.1 File System Layout 273
  - 4.3.2 Implementing Files 274
  - 4.3.3 Implementing Directories 280
  - 4.3.4 Shared Files 283
  - 4.3.5 Log-Structured File Systems 285
  - 4.3.6 Journaling File Systems 287
  - 4.3.7 Virtual File Systems 288
- 4.4 FILE SYSTEM MANAGEMENT AND OPTIMIZATION 292
  - 4.4.1 Disk Space Management 292
  - 4.4.2 File System Backups 298
  - 4.4.3 File System Consistency 304
  - 4.4.4 File System Performance 307
  - 4.4.5 Defragmenting Disks 311
- 4.5 EXAMPLE FILE SYSTEMS 312
  - 4.5.1 CD-ROM File Systems 312
  - 4.5.2 The MS-DOS File System 318
  - 4.5.3 The UNIX V7 File System 321
- 4.6 RESEARCH ON FILE SYSTEMS 324
- 4.7 SUMMARY 324

## **5 INPUT/OUTPUT**

**329**

- 5.1 PRINCIPLES OF I/O HARDWARE 329
  - 5.1.1 I/O Devices 330
  - 5.1.2 Device Controllers 331

5.1.3	Memory-Mapped I/O	332
5.1.4	Direct Memory Access (DMA)	336
5.1.5	Interrupts Revisited	339
5.2	PRINCIPLES OF I/O SOFTWARE	343
5.2.1	Goals of the I/O Software	343
5.2.2	Programmed I/O	344
5.2.3	Interrupt-Driven I/O	346
5.2.4	I/O Using DMA	347
5.3	I/O SOFTWARE LAYERS	348
5.3.1	Interrupt Handlers	348
5.3.2	Device Drivers	349
5.3.3	Device-Independent I/O Software	353
5.3.4	User-Space I/O Software	359
5.4	DISKS	360
5.4.1	Disk Hardware	361
5.4.2	Disk Formatting	376
5.4.3	Disk Arm Scheduling Algorithms	379
5.4.4	Error Handling	382
5.4.5	Stable Storage	385
5.5	CLOCKS	388
5.5.1	Clock Hardware	388
5.5.2	Clock Software	390
5.5.3	Soft Timers	393
5.6	USER INTERFACES: KEYBOARD, MOUSE, MONITOR	394
5.6.1	Input Software	394
5.6.2	Output Software	399
5.7	THIN CLIENTS	415
5.8	POWER MANAGEMENT	417
5.8.1	Hardware Issues	418
5.8.2	Operating System Issues	419
5.8.3	Application Program Issues	424
5.9	RESEARCH ON INPUT/OUTPUT	425
5.10	SUMMARY	426