

MODERN OPERATING SYSTEMS

SECOND EDITION

ANDREW S. TANENBAUM

*Vrije Universiteit
Amsterdam, The Netherlands*



PRENTICE HALL

UPPER SADDLE RIVER, NEW JERSEY 07458

Library of Congress Cataloging-in-Publication Data

Tanenbaum, Andrew S.

Modern operating systems / Andrew S. Tanenbaum.--2nd ed.
p. cm.

Includes bibliographical references and index.

ISBN 0-13-031358-0

1. Operating systems (Computers) I. Title

QA76.76.O63 T359 2001

005.4'3--dc21

Vice President and Editorial Director, ECS: **Marcia Horton**

Publisher: **Alan Apt**

Associate Editor: **Toni D. Holm**

Editorial Assistant: **Amy K. Todd**

Vice President and Director of Production and Manufacturing, ESM: **David W. Riccardi**

Executive Managing Editor: **Vince O'Brien**

Managing Editor: **David A. George**

Senior Production Editor: **Camille Trentacoste**

Director of Creative Services: **Paul Belfanti**

Creative Director: **Carole Anson**

Art Director: **Heather Scott**

Cover Art: **Don Martinetti**

Cover Design: **Joseph Sengotta**

Assistant to Art Director: **John Christiana**

Interior Illustration: **Patricia Gutiérrez**

Interior Design and Typesetting; Cover Illustration Concept: **Andrew S. Tanenbaum**

Manufacturing Manager: **Trudy Piscioti**

Manufacturing Buyer: **Pat Brown**

Marketing Manager: **Jennie Burger**



© 2001 by Prentice-Hall, Inc.

Upper Saddle River, New Jersey 07458

The authors and publisher of this book have used their best efforts in preparing this book. These efforts include the development, research, and testing of the theories and programs to determine their effectiveness. The authors and publisher make no warranty of any kind, expressed or implied, with regard to these programs or to the documentation contained in this book. The authors and publisher shall not be liable in any event for incidental or consequential damages in connection with, or arising out of, the furnishing, performance, or use of these programs.

Many of the designations used by manufacturers and sellers to distinguish their products are claimed as trademarks and registered trademarks. Where those designations appear in this book, and Prentice Hall and the authors were aware of a trademark claim, the designations have been printed in initial caps or all caps. All product names mentioned remain trademarks or registered trademarks of their respective owners.

All rights reserved. No part of this book may be reproduced, in any form or by any means, without permission in writing from the publisher.

Printed in the United States of America

10 9 8 7 6 5 4

ISBN 0-13-031358-0

Prentice-Hall International (UK) Limited, *London*

Prentice-Hall of Australia Pty. Limited, *Sydney*

Prentice-Hall Canada Inc., *Toronto*

Prentice-Hall Hispanoamericana, S.A., *Mexico*

Prentice-Hall of India Private Limited, *New Delhi*

Prentice-Hall of Japan, Inc., *Tokyo*

Pearson Education Asia Pte. Ltd., *Singapore*

Editora Prentice-Hall do Brasil, Ltda., *Rio de Janeiro*

**MODERN
OPERATING SYSTEMS**

SECOND EDITION

Other bestselling titles by Andrew S. Tanenbaum

Structured Computer Organization, 4th edition

This widely-read classic, now in its fourth edition, provides the ideal introduction to computer architecture. It covers the topic in an easy-to-understand way, bottom up. There is a chapter on digital logic for beginners, followed by chapters on microarchitecture, the instruction set architecture level, operating systems, assembly language, and parallel computer architectures.

Computer Networks, 3rd edition

This widely-read classic, now in its third edition, provides the ideal introduction to today's and tomorrow's networks. It explains in detail how modern networks are structured. Starting with the physical layer and working up to the application layer, the book covers a vast number of important topics, including wireless communication, fiber optics, data link protocols, Ethernet, routing algorithms, network performance, security, DNS, electronic mail, USENET news, the World Wide Web, and multimedia. The book has especially thorough coverage of TCP/IP and the Internet.

Operating Systems: Design and Implementation, 2nd edition

This popular text on operating systems is the only book covering both the principles of operating systems and their application to a real system. All the traditional operating systems topics are covered in detail. In addition, the principles are carefully illustrated with MINIX, a free POSIX-based UNIX-like operating system for personal computers. Each book contains a free CD-ROM containing the complete MINIX system, including all the source code. The source code is listed in an appendix to the book and explained in detail in the text.

Distributed Operating Systems

This text covers the fundamental concepts of distributed operating systems. Key topics include communication and synchronization, processes and processors, distributed shared memory, distributed file systems, and distributed real-time systems. The principles are illustrated using four chapter-long examples.

To Suzanne, Barbara, Marvin, and the memory of Bram and Sweetie π

PREFACE

The world has changed a great deal since the first edition of this book appeared in 1992. Computer networks and distributed systems of all kinds have become very common. Small children now roam the Internet, where previously only computer professionals went. As a consequence, this book has changed a great deal, too.

The most obvious change is that the first edition was about half on single-processor operating systems and half on distributed systems. I chose that format in 1991 because few universities then had courses on distributed systems and whatever students learned about distributed systems had to be put into the operating systems course, for which this book was intended. Now most universities have a separate course on distributed systems, so it is not necessary to try to combine the two subjects into one course and one book. This book is intended for a first course on operating systems, and as such focuses mostly on traditional single-processor systems.

I have coauthored two other books on operating systems. This leads to two possible course sequences.

Practically-oriented sequence:

1. Operating Systems Design and Implementation by Tanenbaum and Woodhull
2. Distributed Systems by Tanenbaum and Van Steen

Traditional sequence:

1. Modern Operating Systems by Tanenbaum
2. Distributed Systems by Tanenbaum and Van Steen

The former sequence uses MINIX and the students are expected to experiment with MINIX in an accompanying laboratory supplementing the first course. The latter sequence does not use MINIX. Instead, some small simulators are available that can be used for student exercises during a first course using this book. These simulators can be found starting on the author's Web page: www.cs.vu.nl/~ast/ by clicking on Software and supplementary material for my books .

In addition to the major change of switching the emphasis to single-processor operating systems in this book, other major changes include the addition of entire chapters on computer security, multimedia operating systems, and Windows 2000, all important and timely topics. In addition, a new and unique chapter on operating system design has been added.

Another new feature is that many chapters now have a section on research about the topic of the chapter. This is intended to introduce the reader to modern work in processes, memory management, and so on. These sections have numerous references to the current research literature for the interested reader. In addition, Chapter 13 has many introductory and tutorial references.

Finally, numerous topics have been added to this book or heavily revised. These topics include: graphical user interfaces, multiprocessor operating systems, power management for laptops, trusted systems, viruses, network terminals, CD-ROM file systems, mutexes, RAID, soft timers, stable storage, fair-share scheduling, and new paging algorithms. Many new problems have been added and old ones updated. The total number of problems now exceeds 450. A solutions manual is available to professors using this book in a course. They can obtain a copy from their local Prentice Hall representative. In addition, over 250 new references to the current literature have been added to bring the book up to date.

Despite the removal of more than 400 pages of old material, the book has increased in size due to the large amount of new material added. While the book is still suitable for a one-semester or two-quarter course, it is probably too long for a one-quarter or one-trimester course at most universities. For this reason, the book has been designed in a modular way. Any course on operating systems should cover chapters 1 through 6. This is basic material that every student should know.

If additional time is available, additional chapters can be covered. Each of them assumes the reader has finished chapters 1 through 6, but Chaps. 7 through 12 are each self contained, so any desired subset can be used and in any order, depending on the interests of the instructor. In the author's opinion, Chaps. 7 through 12 are much more interesting than the earlier ones. Instructors should tell their students that they have to eat their broccoli before they can have the double chocolate fudge cake dessert.

I would like to thank the following people for their help in reviewing parts of the manuscript: Rida Bazzi, Riccardo Bettati, Felipe Cabrera, Richard Chapman, John Connely, John Dickinson, John Elliott, Deborah Frincke, Chandana Gamage, Robbert Geist, David Golds, Jim Griffioen, Gary Harkin, Frans Kaashoek, Muk-

kai Krishnamoorthy, Monica Lam, Jussi Leiwo, Herb Mayer, Kirk McKusick, Evi Nemeth, Bill Potvin, Prasant Shenoy, Thomas Skinner, Xian-He Sun, William Terry, Robbert Van Renesse, and Maarten van Steen. Jamie Hanrahan, Mark Russinovich, and Dave Solomon were enormously knowledgeable about Windows 2000 and very helpful. Special thanks go to Al Woodhull for valuable reviews and thinking of many new end-of-chapter problems.

My students were also helpful with comments and feedback, especially Staas de Jong, Jan de Vos, Niels Drost, David Fokkema, Auke Folkerts, Peter Groenewegen, Wilco Ibes, Stefan Jansen, Jeroen Ketema, Joeri Mulder, Irwin Oppenheim, Stef Post, Umar Rehman, Daniel Rijkhof, Maarten Sander, Maurits van der Schee, Rik van der Stoel, Mark van Driel, Dennis van Veen, and Thomas Zeeman.

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

Andrew S. Tanenbaum

CONTENTS

PREFACE

xvi

1 INTRODUCTION

1

- 1.1. WHAT IS AN OPERATING SYSTEM? 3
 - 1.1.1. The Operating System as an Extended Machine 3
 - 1.1.2. The Operating System as a Resource Manager 5

- 1.2. HISTORY OF OPERATING SYSTEMS 6
 - 1.2.1. The First Generation (1945-55) 6
 - 1.2.2. The Second Generation (1955-65) 7
 - 1.2.3. The Third Generation (1965-1980) 9
 - 1.2.4. The Fourth Generation (1980-Present) 13
 - 1.2.5. Ontogeny Recapitulates Phylogeny 16

- 1.3. THE OPERATING SYSTEM ZOO 18
 - 1.3.1. Mainframe Operating Systems 18
 - 1.3.2. Server Operating Systems 19
 - 1.3.3. Multiprocessor Operating Systems 19
 - 1.3.4. Personal Computer Operating Systems 19
 - 1.3.5. Real-Time Operating Systems 19
 - 1.3.6. Embedded Operating Systems 20
 - 1.3.7. Smart Card Operating Systems 20

- 1.4. COMPUTER HARDWARE REVIEW 20
 - 1.4.1. Processors 21
 - 1.4.2. Memory 23
 - 1.4.3. I/O Devices 28
 - 1.4.4. Buses 31

- 1.5. OPERATING SYSTEM CONCEPTS 34
 - 1.5.1. Processes 34
 - 1.5.2. Deadlocks 36
 - 1.5.3. Memory Management 37
 - 1.5.4. Input/Output 38
 - 1.5.5. Files 38
 - 1.5.6. Security 41
 - 1.5.7. The Shell 41
 - 1.5.8. Recycling of Concepts 43

- 1.6. SYSTEM CALLS 44
 - 1.6.1. System Calls for Process Management 48
 - 1.6.2. System Calls for File Management 50
 - 1.6.3. System Calls for Directory Management 51
 - 1.6.4. Miscellaneous System Calls 53
 - 1.6.5. The Windows Win32 API 53

- 1.7. OPERATING SYSTEM STRUCTURE 56
 - 1.7.1. Monolithic Systems 56
 - 1.7.2. Layered Systems 57
 - 1.7.3. Virtual Machines 59
 - 1.7.4. Exokernels 61
 - 1.7.5. Client-Server Model 61

- 1.8. RESEARCH ON OPERATING SYSTEMS 63

- 1.9. OUTLINE OF THE REST OF THIS BOOK 65

- 1.10. METRIC UNITS 66

- 1.11. SUMMARY 67

2 PROCESSES AND THREADS**71**

- 2.1. PROCESSES 71
 - 2.1.1. The Process Model 72
 - 2.1.2. Process Creation 73
 - 2.1.3. Process Termination 75
 - 2.1.4. Process Hierarchies 76
 - 2.1.5. Process States 77
 - 2.1.6. Implementation of Processes 79
- 2.2. THREADS 81
 - 2.2.1. The Thread Model 81
 - 2.2.2. Thread Usage 85
 - 2.2.3. Implementing Threads in User Space 90
 - 2.2.4. Implementing Threads in the Kernel 93
 - 2.2.5. Hybrid Implementations 94
 - 2.2.6. Scheduler Activations 94
 - 2.2.7. Pop-Up Threads 96
 - 2.2.8. Making Single-Threaded Code Multithreaded 97
- 2.3. INTERPROCESS COMMUNICATION 100
 - 2.3.1. Race Conditions 100
 - 2.3.2. Critical Regions 102
 - 2.3.3. Mutual Exclusion with Busy Waiting 103
 - 2.3.4. Sleep and Wakeup 108
 - 2.3.5. Semaphores 110
 - 2.3.6. Mutexes 113
 - 2.3.7. Monitors 115
 - 2.3.8. Message Passing 119
 - 2.3.9. Barriers 123
- 2.4. CLASSICAL IPC PROBLEMS 124
 - 2.4.1. The Dining Philosophers Problem 125
 - 2.4.2. The Readers and Writers Problem 128
 - 2.4.3. The Sleeping Barber Problem 129
- 2.5. SCHEDULING 132
 - 2.5.1. Introduction to Scheduling 132
 - 2.5.2. Scheduling in Batch Systems 138
 - 2.5.3. Scheduling in Interactive Systems 142
 - 2.5.4. Scheduling in Real-Time Systems 148
 - 2.5.5. Policy versus Mechanism 149
 - 2.5.6. Thread Scheduling 150

- 2.6. RESEARCH ON PROCESSES AND THREADS 151
- 2.7. SUMMARY 152

3 DEADLOCKS

159

- 3.1. RESOURCES 160
 - 3.1.1. Preemptable and Nonpreemptable Resources 160
 - 3.1.2. Resource Acquisition 161
- 3.2. INTRODUCTION TO DEADLOCKS 163
 - 3.2.1. Conditions for Deadlock 164
 - 3.2.2. Deadlock Modeling 164
- 3.3. THE OSTRICH ALGORITHM 167
- 3.4. DEADLOCK DETECTION AND RECOVERY 168
 - 3.4.1. Deadlock Detection with One Resource of Each Type 168
 - 3.4.2. Deadlock Detection with Multiple Resource of Each Type 171
 - 3.4.3. Recovery from Deadlock 173
- 3.5. DEADLOCK AVOIDANCE 175
 - 3.5.1. Resource Trajectories 175
 - 3.5.2. Safe and Unsafe States 176
 - 3.5.3. The Banker's Algorithm for a Single Resource 178
 - 3.5.4. The Banker's Algorithm for Multiple Resources 179
- 3.6. DEADLOCK PREVENTION 180
 - 3.6.1. Attacking the Mutual Exclusion Condition 180
 - 3.6.2. Attacking the Hold and Wait Condition 181
 - 3.6.3. Attacking the No Preemption Condition 182
 - 3.6.4. Attacking the Circular Wait Condition 182
- 3.7. OTHER ISSUES 183
 - 3.7.1. Two-Phase Locking 183
 - 3.7.2. Nonresource Deadlocks 184
 - 3.7.3. Starvation 184
- 3.8. RESEARCH ON DEADLOCKS 185
- 3.9. SUMMARY 185

4 MEMORY MANAGEMENT**189**

- 4.1. BASIC MEMORY MANAGEMENT 190
 - 4.1.1. Monoprogramming without Swapping or Paging 190
 - 4.1.2. Multiprogramming with Fixed Partitions 191
 - 4.1.3. Modeling Multiprogramming 192
 - 4.1.4. Analysis of Multiprogramming System Performance 194
 - 4.1.5. Relocation and Protection 194

- 4.2. SWAPPING 196
 - 4.2.1. Memory Management with Bitmaps 199
 - 4.2.2. Memory Management with Linked Lists 200

- 4.3. VIRTUAL MEMORY 202
 - 4.3.1. Paging 202
 - 4.3.2. Page Tables 205
 - 4.3.3. TLBs—Translation Lookaside Buffers 211
 - 4.3.4. Inverted Page Tables 213

- 4.4. PAGE REPLACEMENT ALGORITHMS 214
 - 4.4.1. The Optimal Page Replacement Algorithm 215
 - 4.4.2. The Not Recently Used Page Replacement Algorithm 216
 - 4.4.3. The First-In, First-Out 217
 - 4.4.4. The Second Chance Page Replacement Algorithm 217
 - 4.4.5. The Clock Page Replacement Algorithm 218
 - 4.4.6. The Least Recently Used 218
 - 4.4.7. Simulating LRU in Software 220
 - 4.4.8. The Working Set Page Replacement Algorithm 222
 - 4.4.9. The WSClock Page Replacement Algorithm 225
 - 4.4.:. Summary of Page Replacement Algorithms 227

- 4.5. MODELING PAGE REPLACEMENT ALGORITHMS 228
 - 4.5.1. Belady's Anomaly 229
 - 4.5.2. Stack Algorithms 229
 - 4.5.3. The Distance String 232
 - 4.5.4. Predicting Page Fault Rates 233

- 4.6. DESIGN ISSUES FOR PAGING SYSTEMS 234
 - 4.6.1. Local versus Global Allocation Policies 234
 - 4.6.2. Load Control 236
 - 4.6.3. Page Size 237
 - 4.6.4. Separate Instruction and Data Spaces 239

- 4.6.5. Shared Pages 239
- 4.6.6. Cleaning Policy 241
- 4.6.7. Virtual Memory Interface 241
- 4.7. IMPLEMENTATION ISSUES 242
 - 4.7.1. Operating System Involvement with Paging 242
 - 4.7.2. Page Fault Handling 243
 - 4.7.3. Instruction Backup 244
 - 4.7.4. Locking Pages in Memory 246
 - 4.7.5. Backing Store 246
 - 4.7.6. Separation of Policy and Mechanism 247
- 4.8. SEGMENTATION 249
 - 4.8.1. Implementation of Pure Segmentation 253
 - 4.8.2. Segmentation with Paging: MULTICS 254
 - 4.8.3. Segmentation with Paging: The Intel Pentium 257
- 4.9. RESEARCH ON MEMORY MANAGEMENT 262
- 4.10. SUMMARY 262

5 INPUT/OUTPUT

269

- 5.1. PRINCIPLES OF I/O HARDWARE 269
 - 5.1.1. I/O Devices 270
 - 5.1.2. Device Controllers 271
 - 5.1.3. Memory-Mapped I/O 272
 - 5.1.4. Direct Memory Access 276
 - 5.1.5. Interrupts Revisited 279
- 5.2. *PRINCIPLES OF I/O SOFTWARE* 282
 - 5.2.1. Goals of the I/O Software 283
 - 5.2.2. Programmed I/O 284
 - 5.2.3. Interrupt-Driven I/O 286
 - 5.2.4. I/O Using DMA 287
- 5.3. I/O SOFTWARE LAYERS 287
 - 5.3.1. Interrupt Handlers 287
 - 5.3.2. Device Drivers 289

- 5.3.3. Device-Independent I/O Software 292
- 5.3.4. User-Space I/O Software 298
- 5.4. DISKS 300
 - 5.4.1. Disk Hardware 300
 - 5.4.2. Disk Formatting 315
 - 5.4.3. Disk Arm Scheduling Algorithms 318
 - 5.4.4. Error Handling 322
 - 5.4.5. Stable Storage 324
- 5.5. CLOCKS 327
 - 5.5.1. Clock Hardware 328
 - 5.5.2. Clock Software 329
 - 5.5.3. Soft Timers 332
- 5.6. CHARACTER-ORIENTED TERMINALS 333
 - 5.6.1. RS-232 Terminal Hardware 334
 - 5.6.2. Input Software 336
 - 5.6.3. Output Software 341
- 5.7. GRAPHICAL USER INTERFACES 342
 - 5.7.1. Personal Computer Keyboard, Mouse, and Display Hardware 343
 - 5.7.2. Input Software 347
 - 5.7.3. Output Software for Windows 347
- 5.8. NETWORK TERMINALS 355
 - 5.8.1. The X Window System 356
 - 5.8.2. The SLIM Network Terminal 360
- 5.9. POWER MANAGEMENT 363
 - 5.9.1. Hardware Issues 364
 - 5.9.2. Operating System Issues 365
 - 5.9.3. Degraded Operation 370
- 5.10. RESEARCH ON INPUT/OUTPUT 371
- 5.11. SUMMARY 372

6 FILE SYSTEMS

379

- 6.1. FILES 380
 - 6.1.1. File Naming 380
 - 6.1.2. File Structure 382
 - 6.1.3. File Types 383
 - 6.1.4. File Access 385
 - 6.1.5. File Attributes 386
 - 6.1.6. File Operations 387
 - 6.1.7. An Example Program Using File System Calls 389
 - 6.1.8. Memory-Mapped Files 391

- 6.2. DIRECTORIES 393
 - 6.2.1. Single-Level Directory Systems 393
 - 6.2.2. Two-level Directory Systems 394
 - 6.2.3. Hierarchical Directory Systems 395
 - 6.2.4. Path Names 395
 - 6.2.5. Directory Operations 398

- 6.3. FILE SYSTEM IMPLEMENTATION 399
 - 6.3.1. File System Layout 399
 - 6.3.2. Implementing Files 400
 - 6.3.3. Implementing Directories 405
 - 6.3.4. Shared Files 408
 - 6.3.5. Disk Space Management 410
 - 6.3.6. File System Reliability 416
 - 6.3.7. File System Performance 424
 - 6.3.8. Log-Structured File Systems 428

- 6.4. EXAMPLE FILE SYSTEMS 430
 - 6.4.1. CD-ROM File Systems 430
 - 6.4.2. The CP/M File System 435
 - 6.4.3. The MS-DOS File System 438
 - 6.4.4. The Windows 98 File System 442
 - 6.4.5. The UNIX V7 File System 445

- 6.5. RESEARCH ON FILE SYSTEMS 448

- 6.6. SUMMARY 448