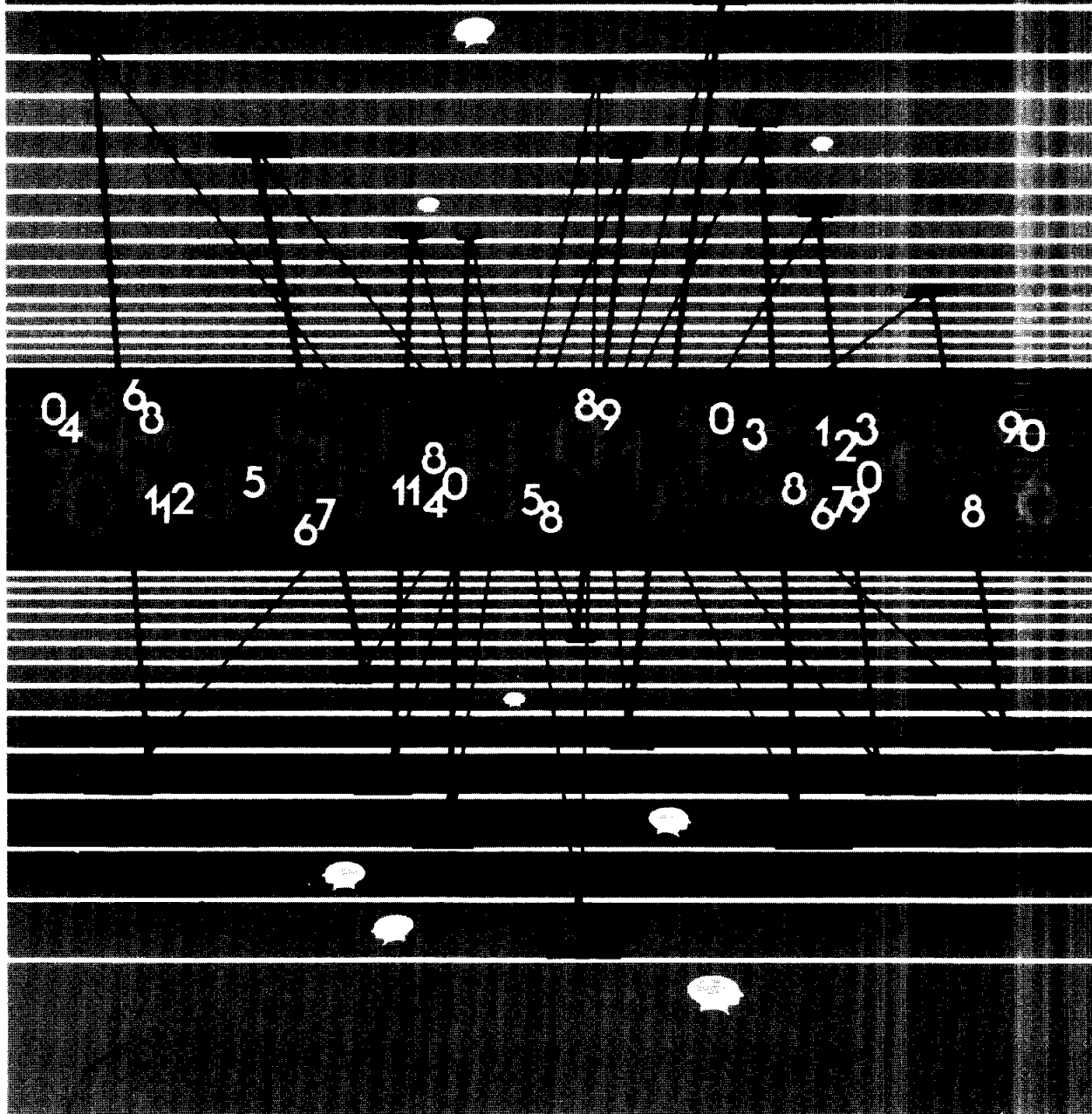


C 1989

# Reliable Distributed System Software

John A. Stankovic

IEEE CATALOG NUMBER EHO230-3  
LIBRARY OF CONGRESS NUMBER 85-60382  
IEEE COMPUTER SOCIETY ORDER NUMBER 570  
ISBN 0-8186-0570-7



Published by IEEE Computer Society Press  
1109 Spring Street  
Suite 300  
Silver Spring, MD 20910

Copyright and Reprint Permissions: Abstracting is permitted with credit to the source. Libraries are permitted to photocopy beyond the limits of U.S. copyright law for private use of patrons those articles in this volume that carry a code at the bottom of the first page, provided the per-copy fee indicated in the code is paid through the Copyright Clearance Center, 29 Congress Street, Salem, MA 01970. Instructors are permitted to photocopy isolated articles for noncommercial classroom use without fee. For other copying, reprint or republication permission, write to Director, Publishing Services, IEEE, 345 E. 47 St., New York, NY 10017. All rights reserved. Copyright © 1985 by The Institute of Electrical and Electronics Engineers, Inc.

IEEE Catalog Number EHO230-3  
Library of Congress Number 85-60382  
IEEE Computer Society Order Number 570  
ISBN 0-8186-0570-7 (Paper)  
ISBN 0-8186-4570-9 (Microfiche)

4321  
1-11-85

Order from: IEEE Computer Society	IEEE Service Center
Post Office Box 80452	445 Hoes Lane
Worldway Postal Center	Piscataway, NJ 08854
Los Angeles, CA 90080	

## Acknowledgements

Some of the material found in Sections 1 and 2 was jointly developed by Prof. Kohler, Prof. Ramamritham, and me. I thank them for their work in these areas.



## Preface

This tutorial text organizes and presents issues and concepts related to reliability in distributed systems software. In this presentation, software includes communication protocols, logical interprocess communication (IPC) support, distributed programming languages, distributed operating systems, and distributed databases. While there exists a considerable body of knowledge on reliability techniques for hardware, communication protocols, and distributed databases, there has been significantly less work on reliability in distributed operating systems. In fact, much of the distributed operating system work only briefly or implicitly addresses reliability. This tutorial text attempts to cover the broad spectrum of reliability techniques used in distributed system software including distributed operating systems.

A reader can expect to learn what reliability is, what reliability techniques are used in the different areas of distributed system software, and how reliability techniques can be better applied across all areas of distributed systems software (especially in the distributed operating system area). The text is primarily written for computer scientists and systems programmers who need to understand, design, or implement reliable system software. Hardware designers also can benefit from knowledge of the software reliability techniques by obtaining a better overall system perspective on reliability.

The text first presents an overview of reliability (hardware and software) as well as an overview of "general" distributed computer systems research. Reliability techniques for each subarea of distributed systems software is then presented including: the communication subnet (Section 3), the operating system area (Sections 4–7 inclusive), and the database area (Section 8). Techniques in the operating system area are divided into subareas including logical IPC and distributed programming languages (Section 4), distributed control (Section 5), structuring distributed systems for reliability (Section 6), and a summary collection of software reliability techniques (Section 7). The text concludes with several case studies of reliable systems (Section 9). An extensive set of references and a bibliography are also provided.

John A. Stankovic



# Table of Contents

<b>Preface</b> .....	iii
<b>Acknowledgements</b> .....	v
<b>Section 1: Introduction</b> .....	1
1.1: Overview of Reliability. ....	2
1.2: Early Work and Surveys .....	9
Recovery Blocks in Action: A System Supporting High Reliability.....	11
<i>T. Anderson and R. Kerr (Proceedings of the 2nd International Conference on Software Engineering, 1976, pages 447-457)</i>	
Fault Tolerant Operating Systems. ....	22
<i>P.J. Denning (Computing Surveys, December 1976, pages 359-389)</i>	
Software Reliability—Status and Perspectives.....	53
<i>C.V. Ramamoorthy and F.B. Bastani (IEEE Transactions on Software Engineering, July 1982, pages 354-371)</i>	
<b>Section 2: Distributed Systems Software Issues</b> .....	71
2.1: Overview of Current Work and Critical Issues in Distributed System Software. ....	73
<b>Section 3: Reliability in the Communications Subnet</b> .....	77
3.1: Overview .....	79
<b>Section 4: Reliable Interprocess Communication.</b> .....	81
4.1: Overview .....	82
4.2: Reliable Interprocess Communication and Distributed Programming Languages .....	85
The Design of a Reliable Remote Procedure Call Mechanism .....	87
<i>S.K. Shrivastava and F. Panzieri (IEEE Transactions on Computers, July 1982, pages 692-697)</i>	
Transactions: A Construct for Reliable Distributed Computing .....	93
<i>A.Z. Spector and P.M. Schwarz (ACM Operating System Review, April 1983, pages 18-35)</i>	
Guardians and Actions: Linguistic Support for Robust, Distributed Programs.....	111
<i>B. Liskov and R. Scheifler (Proceedings of the Ninth Symposium on Principles of Programming Languages, 1982, pages 7-19)</i>	
NIL: An Integrated Language and System for Distributed Programming .....	124
<i>R.E. Strom and S. Yemini (ACM SIGPLAN Notices, June 1983, pages 73-82)</i>	
<b>Section 5: Decentralized Control.</b> .....	133
5.1: Overview .....	134
5.2: Executive Decentralized Control .....	139
Decentralized Executive Control of Computers .....	141
<i>E.D. Jensen (Proceedings of the 3rd International Conference on Distributed Computing Systems, 1982, pages 31-35)</i>	
Distributed Systems—Towards a Formal Approach .....	146
<i>G. Le Lann (Proceedings IFIP Congress, August 1977, pages 155-160)</i>	

**Section 6: Structuring Distributed Systems for Reliability, Relocatability, Small Protection Domains, Object Based Systems** . . . . . 153

6.1: Overview . . . . . 154

6.2: Structuring Distributed Systems . . . . . 157

Process Structuring, Synchronization, and Recovery Using Atomic Actions . . . . . 159

*D.B. Lomet (Proceedings of the ACM Conference on Language Design for Reliable Software, SIGPLAN Notices, March 1977, pages 128-137)*

LOCUS: A Network Transparent, High Reliability Distributed System . . . . . 169

*G. Popek, B. Walker, J. Chow, D. Edwards, C. Kline, G. Rudisin, and G. Thiel (Proceedings of the 8th Symposium on Office Systems Principles, December 1981, pages 169-177)*

Structuring Distributed Systems for Recoverability and Crash Resistance . . . . . 178

*S.K. Shrivastava (IEEE Transactions on Software Engineering, July 1981, pages 436-447)*

**Section 7: Software Reliability Techniques** . . . . . 191

7.1: Overview . . . . . 192

7.2: Miscellaneous Software Reliability Techniques . . . . . 196

A Framework for Software Fault Tolerance in Real-Time Systems . . . . . 198

*T. Anderson and J.C. Knight (IEEE Transactions on Software Engineering, May 1983, pages 355-364)*

Watchdog Processors and Structural Integrity Checking . . . . . 208

*D.J. Lu (IEEE Transactions on Computers, July 1982, pages 681-685)*

Reaching Agreement in the Presence of Faults . . . . . 213

*M. Pease, R. Shostak, and L. Lamport (Journal of the ACM, April 1980, pages 228-234)*

State Restoration in Systems of Communicating Processes . . . . . 220

*D.L. Russell (IEEE Transactions on Software Engineering, March 1980, pages 183-194)*

Optimistic Recovery: An Asynchronous Approach to Fault-Tolerance in Distributed Systems . . . . . 232

*R.E. Strom and S. Yemini (The Proceedings of the Fourteenth International Conference on Fault-Tolerant Computing, 1984, pages 374-379)*

Redundancy in Data Structures: Improving Software Fault Tolerance . . . . . 238

*D.J. Taylor, D.E. Morgan, and J.P. Black (IEEE Transactions on Software Engineering, November 1980, pages 585-594)*

**Section 8: Reliable Distributed Databases** . . . . . 249

8.1: Replication, Crash Recovery, and Nested Transactions . . . . . 250

Reliability Issues for Fully Replicated Distributed Databases . . . . . 252

*H. Garcia-Molina (Computer, September 1982, pages 34-42)*

A Formal Model of Crash Recovery in a Distributed System . . . . . 261

*D. Skeen and M. Stonebraker (IEEE Transactions on Software Engineering, May 1983, pages 219-228)*

Resilient Distributed Computing . . . . . 271

*L. Svobodova (IEEE Transactions on Software Engineering, May 1984, pages 257-268)*

8.2: Case Studies in Reliable Distributed Databases. . . . . 283

The Recovery Manager of the System R Database Manager . . . . . 285

*J. Gray, P. McJones, M. Blasgen, B. Lindsay, R. Lorie, T. Price, F. Putzolu, and I. Traiger (Computing Surveys, June 1981, pages 223-242)*

Reliability Mechanisms for SDD-1: A System for Distributed Databases. . . . . 305

*M. Hammer and D. Shipman (ACM Transactions on Database Systems, December 1980, pages 431-466)*

**Section 9: Case Studies of Reliable Systems** ..... 337

9.1: REBUS ..... 338

REBUS, A Fault-Tolerant Distributed System for Industrial Real-Time Control ..... 339

*J.M. Ayache, J.P. Courtiat, and M. Diaz (IEEE Transactions on Computers, July 1982, pages 637-647)*

9.2: SIFT ..... 350

Formal Specification and Mechanical Verification of SIFT: A Fault-Tolerant Flight Control System ..... 351

*P.M. Melliar-Smith and R.L. Schwartz (IEEE Transactions on Computers, July 1982, pages 616-630)*

9.3: The NonStop Operating System ..... 366

A NonStop Kernel ..... 367

*J.F. Bartlett (Proceedings of the Eighth Symposium on Operating Systems Principles, December 1981, pages 22-29)*

**Section 10: Bibliography** ..... 375

**Author and Subject Index** ..... 385

**Biography** ..... 389