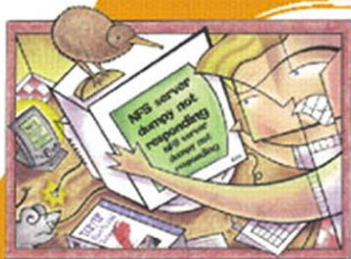
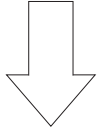


NFS Illustrated

Brent Callaghan



ADDISON-WESLEY PROFESSIONAL COMPUTING SERIES



NFS ILLUSTRATED

NFS Illustrated

Table of Contents

Contents

Preface

1 Introduction

1.1 File Access and File Transfer

1.2 Early File Access Protocols

1.3 ONC RPC

1.4 Organization of This Book

2 XDR External Data Representation

2.1 Protocols and Transportable Data

2.2 A Canonical Standard

2.3 XDR Unit

2.4 Primitive XDR Data Types

2.4.1 Integer

2.4.2 Unsigned Integer

2.4.3 Boolean

2.4.4 Hyper Integer and Unsigned Hyper Integer

2.4.5 Fixed-Length Opaque Data

2.4.6 Variable-Length Opaque Data

2.4.7 String

2.5 Structured XDR Data Types

2.5.1 Fixed-Length Array

2.5.2 Variable-Length Array

2.5.3 Discriminated Union

2.5.4 Linked Lists

2.6 XDR Language

Table of Contents

- 2.6.1 Notational Conventions
- 2.6.2 Lexical Notes
- 2.6.3 Syntax Information
- 2.6.4 Syntax Notes
- 2.6.5 Example of an XDR Data Description
- 2.6.6 XDR Language Variant in This Book

2.7 Summary

3 ONC RPC

- 3.1 Remote Procedure Call Model
- 3.2 RPC Transport Independence and Semantics
- 3.3 Program Numbers and Versioning
- 3.4 Requirements of the RPC Protocol
- 3.5 RPC Call Message
- 3.6 RPC Reply Message
- 3.7 Record-Marking Standard
- 3.8 Portmapper and Rpcbind
 - 3.8.1 Portmap Protocol
 - 3.8.2 Rpcbind Protocol
- 3.9 Summary

4 RPC Authentication

- 4.1 RPC Credential and Verifier
- 4.2 Authentication Flavors
- 4.3 Null Authentication
- 4.4 System Authentication
- 4.5 Diffie-Hellman Authentication
 - 4.5.1 Naming
 - 4.5.2 DH Authentication Verifiers
 - 4.5.3 Nicknames and Clock Synchronization
 - 4.5.4 DH Authentication Protocol

Table of Contents

4.5.5 DH Full Network Name Credential and Verifier

4.5.6 DH Nickname Credential and Verifier

4.5.7 Server Verifier

4.5.8 Diffie-Hellman Encryption

4.5.9 Weaknesses of DH Authentication

4.6 Kerberos Version 4 Authentication

4.6.1 Kerberos Naming

4.6.2 Kerberos-Based Authentication Protocol Specification

4.6.3 Kerberos Full Network Name Credential and Verifier

4.6.4 Kerberos Nickname Credential and Verifier

4.6.5 Server Verifier

4.6.6 Kerberos-Specific Authentication Status Values

4.6.7 Weaknesses of Kerberos Authentication

4.7 RPCSEC_GSS Authentication

4.7.1 RPCSEC_GSS Security Protocol

4.7.2 RPCSEC_GSS Credential

4.7.3 Context Creation

4.7.4 Successful Creation Response

4.7.5 Unsuccessful Context Creation

4.7.6 RPC Data Exchange

4.7.7 Data Integrity Protection

4.7.8 Data Privacy

4.7.9 Server Processing of RPC Data Requests

4.7.10 Servers Reply

4.7.11 RPCSEC_GSS Errors

4.7.12 Performance of RPCSEC_GSS

4.7.13 Snoop Trace of Session

4.8 Connection-Based Security

4.9 Summary

5 NFS Filesystem Model

Table of Contents

5.1 Filesystem and Filesystem Objects

5.2 Filehandles

5.3 Pathnames

5.3.1 Filename Component Handling

5.3.2 Pathname Evaluation

5.4 Stateless Server

5.4.1 Server Recovery

5.4.2 Idempotent Operations

5.4.3 Statelessness and High Availability

5.5 Summary

6 NFS Version 2

6.1 Common Data Types

6.2 Server Procedures

6.2.1 Procedure 0: NULLDo Nothing

6.2.2 Procedure 1: GETATTRGet File Attributes

6.2.3 Procedure 2: SETATTRSet File Attributes

6.2.4 Procedure 4: LOOKUPLook Up File Name

6.2.5 Procedure 5: READLINKRead from Symbolic Link

6.2.6 Procedure 6: READRead from File

6.2.7 Procedure 8: WRITEWrite to File

6.2.8 Procedure 9: CREATECreate File

6.2.9 Procedure 10: REMOVERemove File

6.2.10 Procedure 11: RENAMERename File

6.2.11 Procedure 12: LINKCreate Link to File

6.2.12 Procedure 13: SYMLINKCreate Symbolic Link

6.2.13 Procedure 14: MKDIRCreate Directory

6.2.14 Procedure 15: RMDIRRemove Directory

6.2.15 Procedure 16: READDIRRead from Directory

6.2.16 Procedure 17: STATFSGet Filesystem Attributes

7 NFS Version 3

Table of Contents

7.1 Changes from the NFS Version 2 Protocol

- 7.1.1 Deleted Procedures
- 7.1.2 Modified Procedures
- 7.1.3 New Procedures
- 7.1.4 Filehandle Size
- 7.1.5 Maximum Data Sizes
- 7.1.6 Error Return
- 7.1.7 File Type
- 7.1.8 File Attributes
- 7.1.9 Set File Attributes
- 7.1.10 32-Bit Clients/Servers and 64-Bit Clients/Servers

7.2 Basic Data Types

- 7.2.1 Sizes
- 7.2.2 Basic Data Types

7.3 Server Procedures

- 7.3.1 Procedure 0: NULLDo Nothing
- 7.3.2 Procedure 1: GETATTRGet File Attributes
- 7.3.3 Procedure 2: SETATTRSet File Attributes
- 7.3.4 Procedure 3: LOOKUPLook Up Filename
- 7.3.5 Procedure 4: ACCESSCheck Access Permission
- 7.3.6 Procedure 5: READLINKRead from Symbolic Link
- 7.3.7 Procedure 6: READRead from File
- 7.3.8 Procedure 7: WRITEWrite to File
- 7.3.9 Procedure 8: CREATECreate a File
- 7.3.10 Procedure 9: MKDIRCreate a Directory
- 7.3.11 Procedure 10: SYMLINKCreate a Symbolic Link
- 7.3.12 Procedure 11: MKNODCreate a Special Device
- 7.3.13 Procedure 12: REMOVERemove a File
- 7.3.14 Procedure 13: RMDIRRemove Directory
- 7.3.15 Procedure 14: RENAMERename File or Directory
- 7.3.16 Procedure 15: LINKCreate Link to an Object

Table of Contents

- 7.3.17 Procedure 16: REaddirRead from Directory
- 7.3.18 Procedure 17: REaddirPlusExtended Read from Directory
- 7.3.19 Procedure 18: FSStatGet Dynamic Filesystem Information
- 7.3.20 Procedure 19: FSInfoGet Static Filesystem Information
- 7.3.21 Procedure 20: PathconfRetrieve POSIX Information
- 7.3.22 Procedure 21: CommitCommit Cached Data on a Server to Stable Storage

8 NFS Implementation

- 8.1 File Attributes
- 8.2 Unsupported Procedures
- 8.3 Multiple Version Support
- 8.4 Last Close Problem
- 8.5 Crossing of Server Mountpoints
- 8.6 Problems with Hierarchical Mounts
- 8.7 Permission Issues
 - 8.7.1 Identifying Users
 - 8.7.2 Access to Open Files
 - 8.7.3 UID/GID Mapping
 - 8.7.4 Checking File Access
 - 8.7.5 Executable Files
- 8.8 Concurrent RPC Calls
- 8.9 Duplicate Request Cache
- 8.10 Synchronous Operations and Stable Storage
- 8.11 Adaptive Retransmission for UDP
- 8.12 Read-ahead and Write-behind
- 8.13 Write Gathering
- 8.14 Caching Policies
 - 8.14.1 Cached Objects
 - 8.14.2 Cache Consistency

Table of Contents

- 8.14.3 Close-to-Open Consistency
- 8.14.4 Read and Write Caching
- 8.14.5 Disk Caching
- 8.14.6 Disconnection and Reconnection

8.15 Connectathon

8.16 Summary

9 NFS MOUNT Protocol

9.1 Protocol Revisions

9.2 Transport

9.3 Authentication

9.4 Access Lists

9.5 Server Procedures

- 9.5.1 Procedure 0: NULLDo Nothing
- 9.5.2 Procedure 1: MOUNTPROC_MNTAdd MOUNT Entry
- 9.5.3 Procedure 2: MOUNTPROC_DUMPReturn MOUNT Entries
- 9.5.4 Procedure 3: MOUNTPROC_UMNTRemove MOUNT Entry
- 9.5.5 Procedure 4: MOUNTPROC_UMNTALL Remove All MOUNT Entries
- 9.5.6 Procedure 5: MOUNTPROC_EXPORTReturn Export List
- 9.5.7 Procedure 6: MOUNTPROC_EXPORTALLReturn Export List
- 9.5.8 Procedure 7: MOUNTPROC_PATHCONFPOSIX Pathconf Information

9.6 MOUNT Table

9.7 Submounts

9.8 Export Limitations

9.9 Summary

10 NFS Lock Manager Protocol

10.1 Monitored Locking

10.2 Advisory vs. Mandatory Locks

10.3 Exclusive and Nonexclusive Locks

10.4 Asynchronous Procedures and Callback

Table of Contents

- 10.5 DOS/Windows File Sharing Procedures
- 10.6 Server Crash Recovery
- 10.7 Lockd and Statd Implementation
- 10.8 Client Crash Recovery
- 10.9 Deadlock Detection
- 10.10 Locking Cached or Mapped Files
- 10.11 Transport and Authentication
- 10.12 Basic Data Types
- 10.13 Errors
- 10.14 Lock Manager Procedures
 - 10.14.1 Procedure 0: NULLDo Nothing
 - 10.14.2 Procedure 1: NLM_TESTTest for a Lock
 - 10.14.3 Procedure 2: NLM_LOCKCreate a Lock
 - 10.14.4 Procedure 3: NLM_CANCELCancel a Lock
 - 10.14.5 Procedure 4: NLM_UNLOCKRemove a Lock
 - 10.14.6 Procedure 5: NLM_GRANTEDLock Is Granted
 - 10.14.7 Procedure 20: NLM_SHAREShare a File
 - 10.14.8 Procedure 21: NLM_UNSHAREUnshare a File
 - 10.14.9 Procedure 22: NLM_NM_LOCKEstablish a Nonmonitored Lock
 - 10.14.10 Procedure 23: NLM_FREE_ALLFree All Locks
- 10.15 Network Status Monitor Protocol
 - 10.15.1 Procedure 0: SM_NULLDo Nothing
 - 10.15.2 Procedure 1: SM_STATCheck Status
 - 10.15.3 Procedure 2: SM_MONMonitor a Host
 - 10.15.4 Procedure 3: SM_UNMONUnmonitor a Host
 - 10.15.5 Procedure 4: SM_UNMON_ALLUnmonitor All Hosts
 - 10.15.6 Procedure 5: SM_SIMU_CRASHSimulate a Crash
 - 10.15.7 Procedure 6: SM_NOTIFYNotify a Host
- 10.16 Summary

Table of Contents

11 Automounting

11.1 Automounter as NFS Server

11.2 Problems with Symbolic Links

11.3 Automounting with Autofs

11.4 Automounter Maps

11.4.1 Master Map

11.4.2 Direct Map

11.4.3 Indirect Map

11.4.4 Executable Map

11.5 Offset Mounts

11.6 Multiple Mounts

11.7 Replica Servers

11.8 Map Variables and Key Substitution

11.9 MOUNT Options

11.10 Amd Automounter

11.10.1 Nonblocking Operation

11.10.2 Server Keepalives

11.10.3 Map Syntax

11.11 Summary

12 NFS Variants

12.1 Spritely NFS

12.1.1 Stateful Server

12.1.2 Callback

12.1.3 Write-behind

12.1.4 Recovery

12.1.5 Performance

12.1.6 Summary

12.2 NQNFS

12.2.1 Leases

Table of Contents

12.2.2 Recovery

12.2.3 Other Features

12.2.4 Summary

12.3 Trusted NFS

12.3.1 AUTH_MLS Credential

12.3.2 Extended Attributes

12.3.3 Filename Attributes

12.3.4 TNFS Interoperability

12.3.5 Summary

12.4 NASD NFS

12.4.1 NASD Storage

12.4.2 Locating the Data

12.4.3 Data Security

12.4.4 Summary

13 Other Distributed Filesystems

13.1 Remote File Sharing

13.1.1 Remote System Calls

13.1.2 RFS Naming

13.1.3 Security and UID/GID Mapping

13.1.4 Summary

13.2 Andrew File System

13.2.1 File Caching

13.2.2 Shared Namespace

13.2.3 Volume Movement

13.2.4 Read-only Replication

13.2.5 Security

13.2.6 Summary

13.3 DCE/DFS

13.3.1 Cache Consistency with Tokens

13.3.2 DFS Namespace

Table of Contents

13.3.3 Episode File System

13.3.4 Summary

13.4 SMB File Access

13.4.1 Namespace

13.4.2 Session Setup

13.4.3 PC File Semantics

13.4.4 Batched Requests

13.4.5 File Locking

13.4.6 Opportunistic Locks

13.4.7 The Samba Server

13.4.8 Summary

14 PC NFS

14.1 File Naming

14.2 File Attributes

14.3 Text Files

14.4 Symbolic Links

14.5 PCNFSD Protocol

14.5.1 Printing

14.5.2 Comment Strings

14.5.3 Transport and Authentication

14.6 PCNFSD Version 1

14.6.1 Procedure 1: PCNFSD_AUTHPerform User Authentication

14.6.2 Procedure 2: PCNFSD_PR_INITPrepare for Remote Printing

14.6.3 Procedure 3: PCNFSD_PR_STARTSubmit Print Job

14.7 PCNFSD Version 2

14.7.1 Procedure 1: PCNFSD2_INFODetermine Supported Services

14.7.2 Procedure 2: PCNFSD2_PR_INITPrepare for Remote Printing

14.7.3 Procedure 3: PCNFSD2_PR_STARTSubmit Job for Printing

14.7.4 Procedure 4: PCNFSD2_PR_LISTList Printers on Server

14.7.5 Procedure 5: PCNFSD2_PR_QUEUEList Printer Jobs Queued

Table of Contents

- 14.7.6 Procedure 6: PCNFSD2_PR_STATUSDetermine Printer Status
- 14.7.7 Procedure 7: PCNFSD2_PR_CANCELCancel a Print Job
- 14.7.8 Procedure 8: PCNFSD2_PR_ADMINPrinter Administration
- 14.7.9 Procedure 9: PCNFSD2_PR_REQUEUEChange Print Job Queue Position
- 14.7.10 Procedure 10: PCNFSD2_PR_HOLDHold a Print Job in the Queue
- 14.7.11 Procedure 11: PCNFSD2_PR_RELEASERelease Hold on a Print Job
- 14.7.12 Procedure 12: PCNFSD2_MAPIDTranslate Between Username and ID
- 14.7.13 Procedure 13: PCNFSD2_AUTHPerform User Authentication
- 14.7.14 Procedure 14: PCNFSD2_ALERTSend Message to Server Administrator

14.8 BWNFSD/HCLNFSD Protocol

14.9 Summary

15 NFS Benchmarks

15.1 Factors Affecting Performance

- 15.1.1 Memory
- 15.1.2 CPU
- 15.1.3 Network
- 15.1.4 Network Interfaces
- 15.1.5 Server Data Bus
- 15.1.6 NVRAM
- 15.1.7 Disk Controllers
- 15.1.8 Disk Spindles

15.2 Workload

- 15.2.1 Operation Mix
- 15.2.2 Working Set
- 15.2.3 File and Directory Size

15.3 Nfsstone

15.4 Nhfsstone

15.5 SFS 1.0 and 1.1

- 15.5.1 Running the SFS Benchmark

Table of Contents

15.5.2 Workload

15.5.3 Server Configuration

15.5.4 Cluster Challenge

15.6 SFS 2.0

15.7 Summary

16 WebNFS

16.1 Internet and NFS over TCP

16.2 Internet and NFS Version 3

16.3 Firewalls

16.4 Public Filehandle

16.5 Multicomponent LOOKUP

16.5.1 Pathname Evaluation and Latency

16.5.2 Initialization Overhead

16.5.3 Pathname Evaluation

16.5.4 Server Evaluation of Pathname

16.5.5 Symbolic Links

16.6 NFS URL

16.6.1 URL Structure

16.6.2 Absolute vs. Relative Paths

16.7 WebNFS Client Characteristics

16.7.1 Mountpoint Crossing

16.7.2 Fetching a File

16.7.3 Fetching a Directory

16.7.4 Fetching a Symbolic Link

16.7.5 Document Types

16.8 WebNFS Server Requirements

16.8.1 Evaluating a Multicomponent LOOKUP

16.8.2 Canonical vs. Native Path

16.8.3 Client and Server Port

Table of Contents

16.9 WebNFS Security Negotiation

16.10 Summary

17 NFS Version 4

17.1 An IETF Protocol

17.2 NFS Version 4 Working Group Charter

17.3 Improved Access and Good Performance on the Internet

17.3.1 Protocol Integration

17.3.2 Latency

17.3.3 Bandwidth

17.3.4 Efficient Caching

17.3.5 Proxy Caching

17.3.6 Scalability

17.3.7 Availability

17.4 Strong Security with Negotiation Built into the Protocol

17.4.1 Connection-Based Security

17.4.2 RPC-Based Security

17.5 Better Cross-Platform Interoperability

17.5.1 File Attributes

17.5.2 Representing User and Group IDs

17.5.3 Access Control Lists

17.5.4 File Locking

17.5.5 Internationalization

17.6 Designed for Protocol Extensions

17.7 Summary

References

Index