SMB remote file protocol (including SMB 3.02)

John Reed, Product Manager, NetApp

SNIA Legal Notice

- The material contained in this tutorial is copyrighted by the SNIA unless otherwise noted.
- Member companies and individual members may use this material in presentations and literature under the following conditions:
  - Any slide or slides used must be reproduced in their entirety without modification
  - The SNIA must be acknowledged as the source of any material used in the body of any document containing material from these presentations.
- This presentation is a project of the SNIA Education Committee.
- Neither the author nor the presenter is an attorney and nothing in this presentation is intended to be, or should be construed as legal advice or an opinion of counsel. If you need legal advice or a legal opinion please contact your attorney.
- The information presented herein represents the author's personal opinion and current understanding of the relevant issues involved. The author, the presenter, and the SNIA do not assume any responsibility or liability for damages arising out of any reliance on or use of this information.
  NO WARRANTIES, EXPRESS OR IMPLIED. USE AT YOUR OWN RISK.

Abstract and Learning Objectives

SNIA Education

Title: SMB remote file protocol (including SMB 3.02)

Abstract

- The SMB protocol evolved over time from CIFS to SMB1 to SMB2, with implementations by dozens of vendors including most major Operating Systems and NAS solutions. The SMB 3.0 protocol had its first commercial implementations by Microsoft, NetApp and EMC by the end of 2012, and many other implementations are in-progress. This SNIA Tutorial describes the basic architecture of the SMB protocol and basic operations, including connecting to a share, negotiating a dialect, executing operations and disconnecting from a share. The second part of the tutorial covers improvements in the version 2 of the protocol, including a reduced command set, support for asynchronous operations, compounding of operations, durable and resilient file handles, file leasing and large MTU support. The final part covers the latest changes in SMB3, including persistent handles (SMB Transparent Failover), active/active clusters (SMB Scale-Out), multiple connections per sessions (SMB Multichannel), support for RDMA protocols (SMB Direct), snapshot-based backups (VSS for Remote File Shares) opportunistic locking of folders (SMB Directory Leasing), and SMB encryption.

Objectives

- Understand the basic architecture of the SMB protocol family
- Enumerate the main capabilities introduced with SMB 2.0/2.1
- Describe the main capabilities introduced with SMB 3.0/SMB 3.02

Summary

- History
  - Remote file protocol
  - A brief history of CIFS, SMB, SMB2 and SMB3
  - SMB implementers
- Basics
  - The basics of SMB
  - SMB 2.0
  - SMB 2.1
- SMB 3.0
  - SMB Transparent Failover
  - SMB Scale-Out
  - SMB Witness
  - SMB Multichannel
  - SMB Direct
  - SMB Directory Leasing
  - SMB Encryption
  - VSS for Remote File Shares
- SMB 3.02

Remote file protocol

- Remote (not Local)
  - Access file across the wire (LAN, WAN)
- File (not Block)
  - Different semantics
- Protocol
  - Well-defined and documented
- Examples
  - NFS, SMB2, SMB3, WebDAV

A brief history of CIFS, SMB, SMB2, and SMB3

- SMB - 1980s
  - PC-DOS – 1984
  - LAN Manager – 1988
  - Implemented on Unix and other operating systems (part of the OS or as a suite like Samba)
- CIFS - 1996
  - Windows NT 4.0 – 1996
  - IETF draft – Common Internet File System – 1997
  - SNIA Technical Specification – 1999
- Back to SMB - 2000
  - Windows 2000 Extensions – 2000
  - Extensions for other implementations of SMB
- SMB 2.0 (or SMB2) - 2008
- SMB 2.1 (or SMB2.1) - 2010
- SMB 3.0 (or SMB3) – 2012
- SMB 3.02 (or SMB3) – 2013

CIFS as a generic term for SMB?

- CIFS means SMB as it existed in Windows NT 4
- However, the term "CIFS" is commonly used incorrectly to refer to more recent versions of SMB like SMB2, SMB2.1, or SMB3
- 'CIFS' is sometimes used as a marketing term to identify specific products, independent of the SMB version implemented
- Using the term 'CIFS' to refer to SMB 2.0 or SMB 3.0 is the equivalent to
  - Using POP to refer to IMAP (in e-mail protocols)
  - Using WEP to refer to WPA (in wireless security)

Negotiating SMB dialects

	SMB 3.02	SMB 3.0	SMB 2.1	SMB 2.0	SMB 1.0
SMB 3.02	SMB 3.02	SMB 3.0	SMB 2.1	SMB 2.0	SMB 1.0
SMB 3.0	SMB 3.0	SMB 3.0	SMB 2.1	SMB 2.0	SMB 1.0
SMB 2.1	SMB 2.1	SMB 2.1	SMB 2.1	SMB 2.0	SMB 1.0
SMB 2.0	SMB 2.0	SMB 2.0	SMB 2.0	SMB 2.0	SMB 1.0
SMB 1.0	SMB 1.0	SMB 1.0	SMB 1.0	SMB 1.0	SMB 1.0

Any references to CIFS usually mean SMB 1.0, but could be other versions.

SMB implementers (alphabetical order)

Apple

- MacOX X 10.2 Jaguar – CIFS/SMB 1 (via Samba)
- MacOS X 10.7 Lion – SMB 1 (via Apple's SMBX)
- MaxOS X 10.9 Mavericks – SMB 2.1 (default file protocol)

EMC

- Older versions – CIFS/SMB 1
- VNX – SMB 3.0
- Isilon OneFS 6.5 – SMB 2
- Isilon OneFS 7.0 – SMB 2.1
- Isilon OneFS 7.1.1 – SMB 3.0

Microsoft

- Microsoft LAN Manager – SMB
- Windows NT 4.0 – CIFS
- Windows 2000 – SMB 1
- Windows Server 2003 or Windows XP – SMB 1
- Windows Server 2008 or Windows Vista – SMB 2
- Windows Server 2008 R2 or Windows 7 – SMB 2.1
- Windows Server 2012 or Windows 8 – SMB 3.0
- Windows Server 2012 R2 or Windows 8.1 – SMB 3.02

NetApp

- Older versions – CIFS/SMB 1
- Data ONTAP 7.3.1 – SMB 2
- Data ONTAP 8.1 – SMB 2.1
- Data ONTAP 8.2 – SMB 3.0

Samba (Linux or others)

- Older versions – CIFS/SMB 1
- Samba 3.6 – SMB 2 (some SMB 2.1)
- Samba 4.1 – SMB 3.0

And many others…

- Most widely implemented remote file protocol in the world, available in ~every NAS and File Server

Information on this slide gathered from publicly available information as of March 2014.

Please contact the implementers directly to obtain the accurate, up-to-date information on their SMB implementation.

The basics of SMB

- Negotiating a dialect
- Connecting to a share
- Executing operations
- Disconnecting from a share

DIR \\FS.EXAMPLE.COM\SHARE1

From	To	Packet
CL	FS	SMB:C NEGOTIATE, Dialect = (Dialect List)
FS	CL	SMB2:R NEGOTIATE (0x0), GUID={8E4F0109-0E04-FD9C-434A-05881428984C}, Mid = 0
CL	FS	SMB2:C SESSION SETUP (0x1), Mid = 1
FS	CL	SMB2:R SESSION SETUP (0x1) ,SessionFlags=0x0, Mid = 1
CL	FS	SMB2:C TREE CONNECT (0x3), Path=\\fs.example.com\IPC$, Mid = 2
FS	CL	SMB2:R TREE CONNECT (0x3), TID=0x1, Mid = 2
CL	FS	DFS:Get DFS Referral Request, FileName: \fs.example.com\share1, MaxReferralLevel: 4
FS	CL	SMB2:R , Mid = 3 - NT Status: System - Error, Code = (412) STATUS_FS_DRIVER_REQUIRED → Not a DFS Namespace, just a file share
CL	FS	SMB2:C TREE CONNECT (0x3), Path=\\fs.example.com\share1, Mid = 4
FS	CL	SMB2:R TREE CONNECT (0x3), TID=0x5, Mid = 4
CL	FS	SMB2:C CREATE (0x5), Context=DHnQ, Context=MxAc, Context=QFid, Mid = 5
FS	CL	SMB2:R CREATE (0x5), Context=MxAc, Context=QFid, FID=0xFFFFFFFF00000001, Mid = 5
CL	FS	SMB2:C QUERY INFO (0x10), FID=0xFFFFFFFF00000001, InformationClass=Query FS Volume Info, FID=0xFFFFFFFF00000001, Mid = 6
FS	CL	SMB2:R QUERY INFO (0x10), Mid = 6
CL	FS	SMB2:C CREATE (0x5), Context=DHnQ, Context=MxAc, Context=QFid, Mid = 8
FS	CL	SMB2:R CREATE (0x5), Context=MxAc, Context=QFid, FID=0xFFFFFFFF00000005, Mid = 8
CL	FS	SMB2:C CLOSE (0x6), FID=0xFFFFFFFF00000001, Mid = 11
FS	CL	SMB2:R CLOSE (0x6), Mid = 11
CL	FS	SMB2:C QUERY INFO (0x10), FID=0xFFFFFFFF00000005, InformationClass=Query FS Full Size Info, FID=0xFFFFFFFF00000005, Mid = 12
FS	CL	SMB2:R QUERY INFO (0x10), Mid = 12
CL	FS	SMB2:C TREE DISCONNECT (0x4), TID=0x1, Mid = 13
FS	CL	SMB2:R TREE DISCONNECT (0x4), Mid = 13
CL	FS	SMB2:C TREE DISCONNECT (0x4), TID=0x5, Mid = 14
FS	CL	SMB2:R TREE DISCONNECT (0x4), Mid = 14
CL	FS	SMB2:C LOGOFF (0x2), Mid = 15
FS	CL	SMB2:R LOGOFF (0x2), Mid = 15

Note: CL= SMB Client, FS= SMB File Server

SMB 1.0

- CIFS as in the 1997 IETF draft
- Windows improvements (over time)
  - Kerberos authentication
  - Shadow copy
  - Server to server copy
  - Signing – MD5
- Non-Windows improvements (over time)
  - Improvements proposed and/or implemented by communities using CIFS/SMB on other operating systems including Unix and MacOS. Not part of any official standard.

SMB 2.0

- First major redesign of SMB
- Increased file sharing scalability
- Improved performance
  - Improved request compounding (reduced round trips)
  - Asynchronous operations (multiple packets in flight)
  - Larger reads/writes (more data in each packet)
- Security-related changes
  - Much smaller command set (from 75 to just 19)
  - SMB Durability provides limited network fault tolerance
  - Signing – Uses HMAC SHA-256 instead of old MD5

SMB 2.0 reduced command set

- Protocol negotiation, user auth and share access
  - NEGOTIATE, SESSION_SETUP, LOGOFF, TREE_CONNECT, TREE_DISCONNECT
- File, directory and volume access
  - CANCEL, CHANGE_NOTIFY, CLOSE, CREATE, FLUSH, IOCTL, LOCK, QUERY_DIRECTORY, QUERY_INFO, READ, SET_INFO, WRITE
- Other
  - ECHO, OPLOCK_BREAK

SMB 2.1

- File leasing improvements
  - File Leasing replaces Opportunistic Locking (oplocks)
  - Improves performance when frequently updating metadata
  - Uses local metadata caching, some forms of shared leases
- Large MTU support
  - Large message support increases throughput
  - Specially relevant for high bandwidth networks like 10GbE
- Peer Content Caching and Retrieval
  - Implemented as BranchCache in Windows
  - Open source implementation in Prequel from Red Hat

SMB 3.0

- Availability
  - SMB Transparent Failover
  - SMB Witness
  - SMB Multichannel
- Performance
  - SMB Scale-Out
  - SMB Direct (RDMA)
  - SMB Multichannel
  - Directory Leasing
  - BranchCache™ V2
- Backup
  - VSS for SMB File Shares
- Security
  - SMB Encryption – AES-CCM
  - Signing - AES-CMAC
- Management
  - PowerShell™ over WS-Man
  - SMI-S File

SMB Transparent Failover

Failover transparent to application

- SMB Server and SMB client handle failover gracefully
- Zero downtime – small IO delay during failover

Supports planned and unplanned failovers

- Hardware or Software Maintenance
- Hardware or Software Failures
- Load Rebalancing

Resilient for both file and directory operations

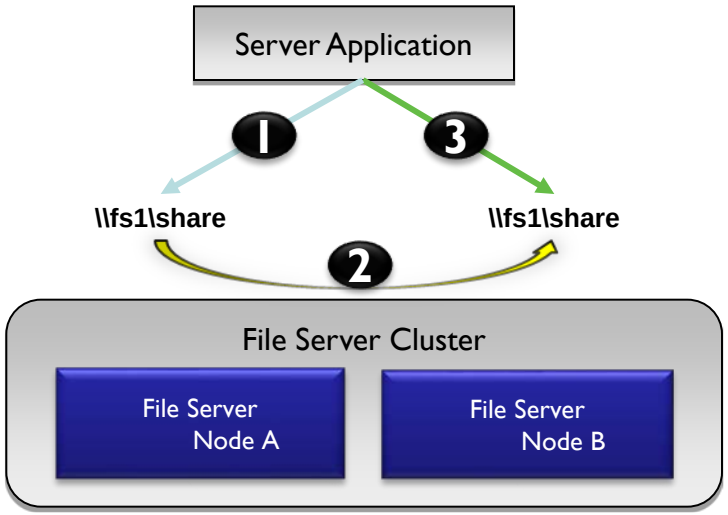
Requires:

- SMB Server in a Failover Cluster
- SMB Server and SMB client must implement SMB 3.0
- Shares enabled for 'Continuous Availability'

Impact to SMB before 3.0

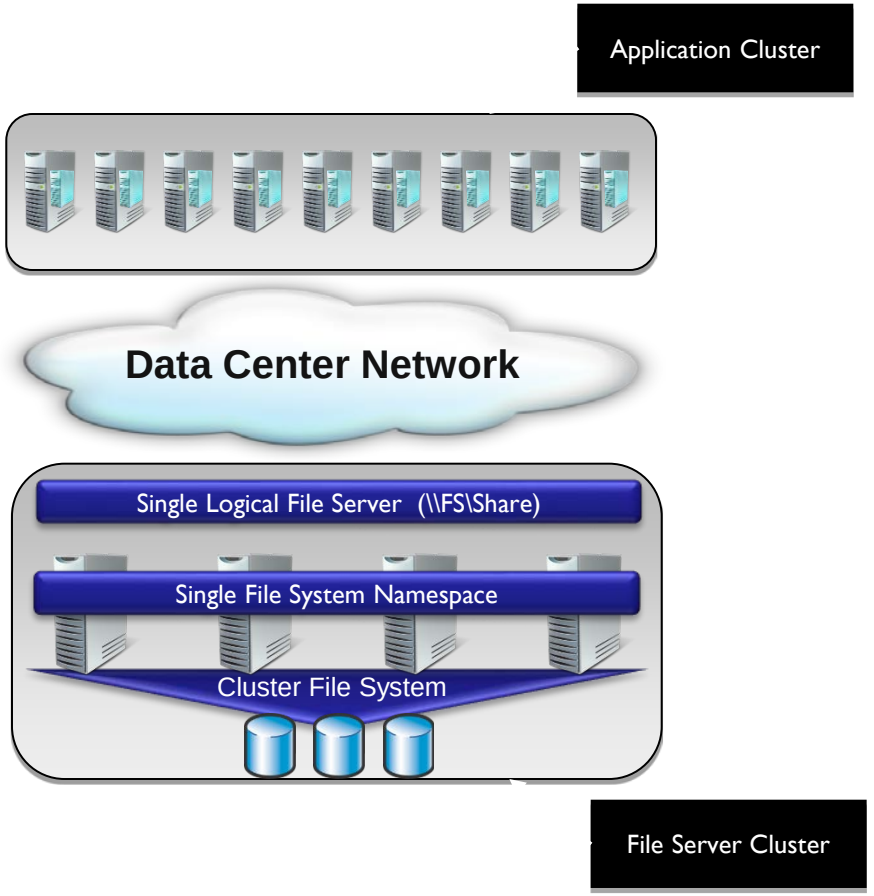
- Older clients can connect,
  but without the Transparent Failover capability

1. **Normal operation**
2. **Failover share - connections and handles lost, temporary stall of IO**
3. **Connections and handles auto-recovered Application IO continues with no errors**



SMB Scale-Out

- Targeted for server app storage
  - Example: Virtualization and Databases
  - Increase available bandwidth by adding cluster nodes
- Key capabilities:
  - Active/Active file shares
  - Fault tolerance with zero downtime
  - Fast failure recovery
- Impact to SMB before 3.0
  - SMB 2.x clients can connect, but without the failover capability
  - SMB1 clients not supported



SMB Direct (SMB over RDMA)

Advantages

- Scalable, fast and efficient storage access
- High throughput with low latency
- Minimal CPU utilization for I/O processing
- Load balancing, automatic failover and bandwidth aggregation via SMB Multichannel

Scenario

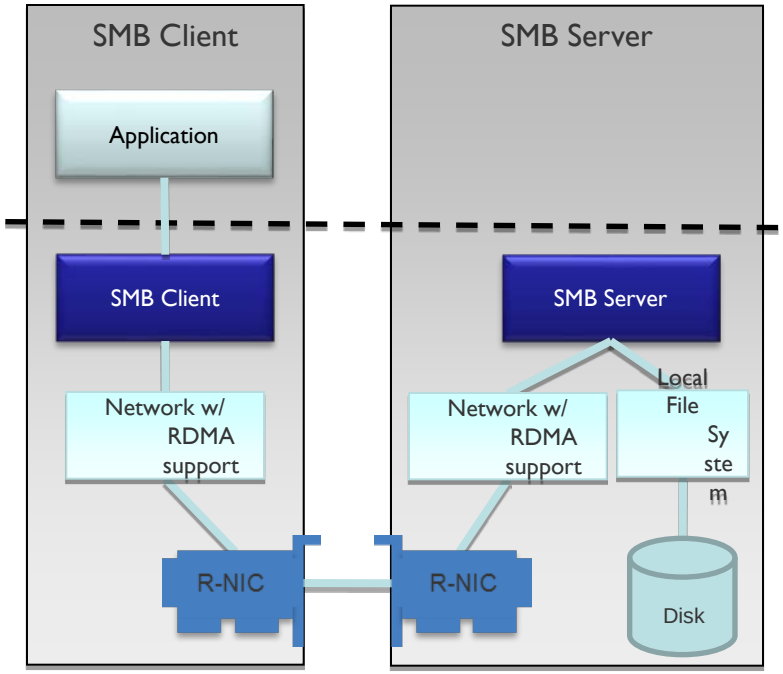
- High performance remote file access for application servers like Virtualization and Databases

Required hardware

- RDMA-capable network interface (R-NIC)
- Three types: iWARP, RoCE and Infiniband

Impact to SMB before 3.0

- Older clients can connect, but without the RDMA capability



SMB Multichannel

Full Throughput

- Bandwidth aggregation with multiple NICs
- Multiple CPUs cores engaged when NIC offers Receive Side Scaling (RSS)

Automatic Failover

- SMB Multichannel implements end-to-end failure detection
- Leverages NIC teaming if present, but does not require it

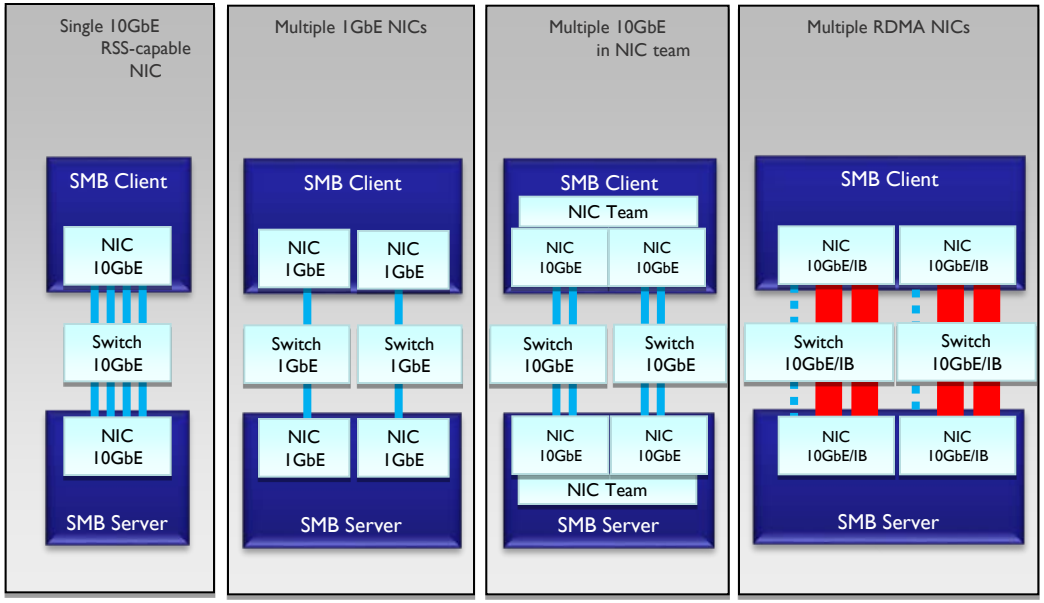
Automatic Configuration

- SMB detects and uses multiple paths

Impact to SMB before 3.0

- Older clients can connect, but without the Multichannel capability

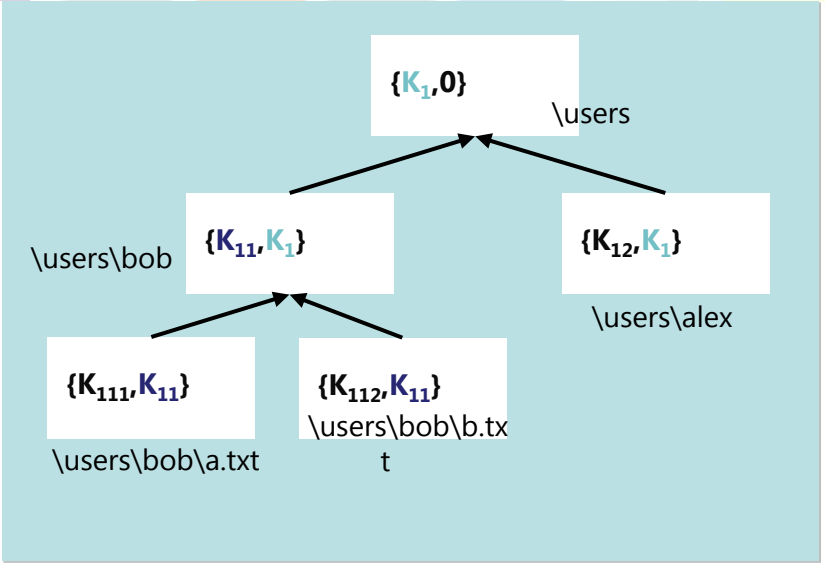
Sample Configurations



Vertical lines are logical channels, not cables

SMB Directory Leasing

- Reduces roundtrips from client to server
  - Metadata is retrieved from longer lived directory cache
  - Directory cache coherency is maintained due to the implementation of directory leases
  - Client gets notified if directory information on server changes
- Targeted at
  - HomeFolder (read/write with no sharing) scenarios
  - Publication (read-only with sharing) scenarios
- Metadata cache
  - Directory handles
  - Directory metadata
- Impact to SMB before 3.0
  - Older clients connect, but without the Directory Leasing capability



- Lease breaks when directory metadata is updated
  - Creation of new children
  - Rename of immediate child file/directory
  - Deletion/Modification of immediate children (manifests when handle is closed)
- Lease breaks when directory handle itself gets a sharing conflict
  - Another conflicting open to directory
  - Rename/deletion of a parent directory

SMB Encryption

End-to-end encryption of data in flight

- Protects data from eavesdropping/snooping attacks on untrusted networks
- Configured per share or for the entire server

Used in scenarios where data traverses untrusted networks

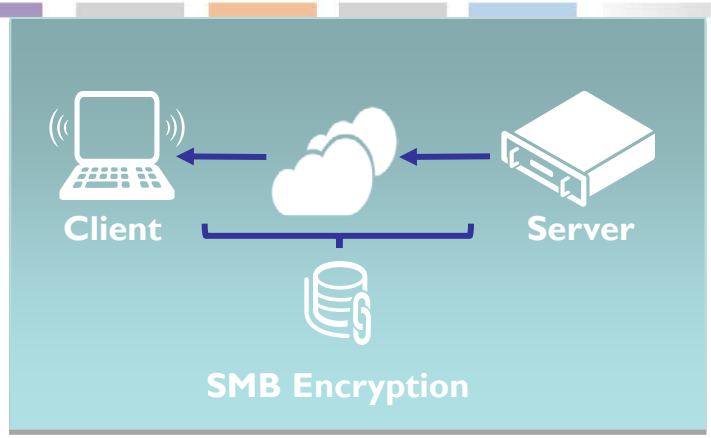
- Application workload over unsecured networks
- Branch Offices over WAN networks

Low deployment costs

- No IPSec required
- No Public Key Infrastructure (PKI) required
- No specialized hardware required

Impact to SMB versions before 3.0

- If encryption is turned on, older clients get "Access Denied" errors

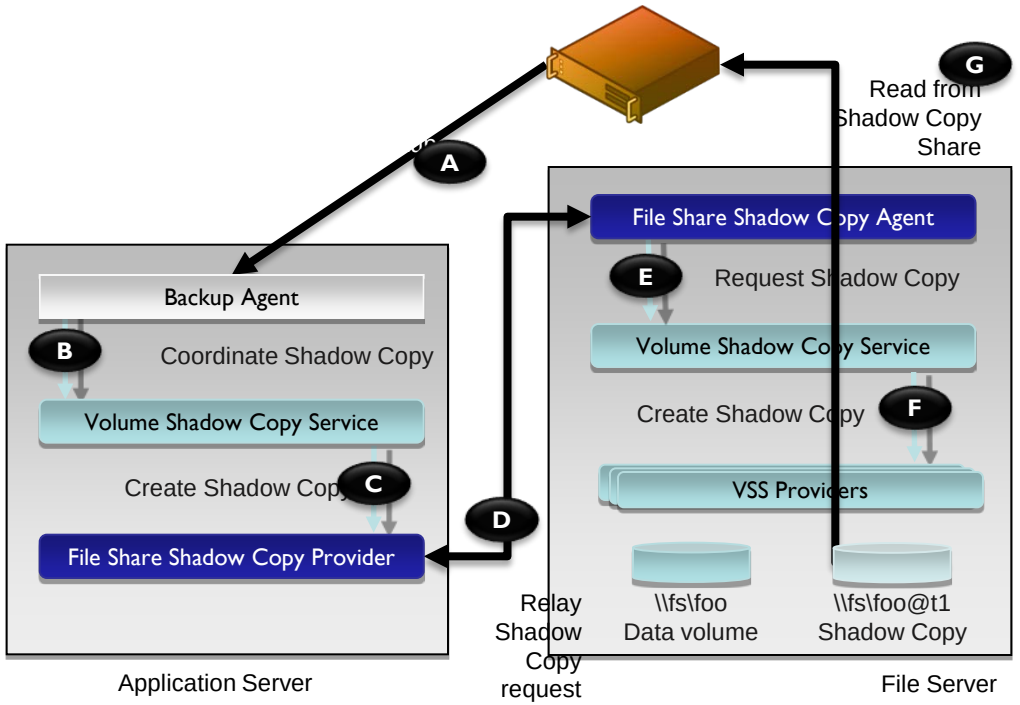


Algorithm

- AES CCM 128 bit
- No algorithm negotiation capability
- Will sign AND encrypt in the same step (independent of SMB Signing setting)
- AES acceleration provided by most new processors aids in performance
- Some CPUs that provide AES hardware acceleration.

VSS for SMB File Shares

- Supports backup and restore scenarios for application servers like Virtualization and Databases
- Application consistent shadow copies for server application data stored on SMB 3.0 file shares
- Full integration with Microsoft's Volume Shadowcopy Services (VSS) infrastructure
- Implemented by at least one vendor besides Microsoft



SMB Management

WMI objects introduced (accessible via WS-Management)

- Manages SMB shares, file server sessions and settings, client connections and settings
- Aimed at both System Administrator and Developers
- Covers both standalone and clustered file server and shares

Main objects and associated methods

- SMB Share: Get, New, Set and Remove
- SMB Share Access: Get, Grant, Revoke, Block and Unblock
- SMB Session: Get and Close
- SMB Open File: Get and Close
- SMB Configuration: Get and Set for Server and Client
- SMB Network Interfaces: Get for Server and Client
- SMB Connection: Get for Connection, Get for Multichannel Connection
- SMB Mappings: Get, New and Remove
- SMB Multichannel Constraints: Get, New and Remove

SMI-S File

- Main WMI objects mapped to SMI-File object model
- Initial support by Microsoft, NetApp, and EMC

WMI = Windows Management Instrumentation, implementation of DMTF standards (WBEM, CIM) on the Windows Platform.

WS-Management = Web Services Management. DMTF open standard for SOAP-based protocol for server management.

SMI-S = Storage Management Initiative – Specification. SNIA Storage Management Standard.

SMB 3.02

- Asymmetric Scale-Out File Server Clusters
  - Clients learn when File Server Cluster is not symmetric
  - Witness protocol changed to allow moving client per SMB share
  - In Windows, SMB clients automatically rebalance
- SMB Direct Remote Invalidation
  - Avoids specific invalidation requests, improving RDMA performance
  - Especially important for workloads with high rate of small IOs
- Unbuffered read/write operations
  - Per-request flags for read/write operations
- Remote Shared Virtual Disk Protocol
  - New protocol defines block semantics for shared virtual disk files
  - Implements SCSI over SMB (SMB protocol used here as a transport)

Links to protocol documentation

Specification	Description
[MS-CIFS]: Common Internet File System (CIFS) Protocol Specification	Specifies the Common Internet File System (CIFS) Protocol, a cross-platform, transport-independent protocol that provides a mechanism for client systems to use file and print services made available by server systems over a network.
[MS-SMB]: Server Message Block (SMB) Protocol Specification	Specifies the Server Message Block (SMB) Protocol, which defines extensions to the existing Common Internet File System (CIFS) specification that have been implemented by Microsoft since the publication of the [CIFS] specification.
[MS-SMB2]: Server Message Block (SMB) Protocol Versions 2 and 3 Specification	Specifies the Server Message Block (SMB) Protocol Versions 2 and 3, which support the sharing of file and print resources between machines and extend the concepts from the Server Message Block Protocol.
[MS-SMBD]: SMB Remote Direct Memory Access (RDMA) Transport Protocol Specification	Specifies the SMB Remote Direct Memory Access (RDMA) Transport Protocol, a wrapper for the existing SMB protocol that allows SMB packets to be delivered over RDMA-capable transports such as iWARP or Infiniband while utilizing the direct data placement (DDP) capabilities of these transports. Benefits include reduced CPU overhead, lower latency, and improved throughput.
[MS-SWN]: Service Witness Protocol Specification	Specifies the Service Witness Protocol, which enables an SMB clustered file server to notify SMB clients with prompt and explicit notifications about the failure or recovery of a network name and associated services.
[MS-FSRVP]: File Server Remote VSS Provider Protocol Specification	Specifies the File Server Remote VSS Protocol, an RPC-based protocol used for creating shadow copies of file shares on a remote computer, and for facilitating backup applications in performing application-consistent backup and restore of data on SMB shares.
[MS-RSVD]: Remote Shared Virtual Disk Protocol	Specifies the Remote Shared Virtual Disk Protocol, which supports accessing and manipulating virtual disks stored as files on an SMB3 file server. This protocol enables opening, querying, administering, reserving, reading, and writing the virtual disk objects, providing for flexible access by single or multiple consumers. It also provides for forwarding of SCSI operations, to be processed by the virtual disk.

Note: Protocols published by Microsoft, but available to anyone to implement in non-Windows platforms.

SNIA SMB2 / SMB3 Plugfest

SNIA Education

- SMB/SMV2/SMB3 PlugFest happens every year side-by-side with the Storage Developer Conference (SNIA SDC) in September
- Intense week of interaction across operating systems and SMB implementations.

Participants in the 2013 edition of the
SNIA SMB2 / SMB3 Plugfest
Santa Clara, CA – September 2013

Review

Objectives

- Understand the basic architecture of the SMB protocol family
- Enumerate the main capabilities introduced with SMB 2.0/2.1
- Describe the main capabilities introduced with SMB 3.0/SMB 3.02

Attribution & Feedback

The SNIA Education Committee would like to thank the following individuals for their contributions to this Tutorial.

Authorship History

Jose Barreto / September 2012

Updates:

Jose Barreto / October 2012
Jose Barreto / January 2014
John Reed / April 2014

Additional Contributors

SW Worth
Christopher Hertel
John Reed

Please send any questions or comments regarding this SNIA Tutorial to
tracktutorials@snia.org

Thank you!

Plugfest

- SMB/SMV2/SMB3 PlugFest happens every year side-by-side with the Storage Developer Conference (SNIA SDC) in September
- Intense week of interaction across operating systems and SMB implementations.

Participants of the 2012 edition of the SNIA SMB/SMB2/SMB3 Plugfest.
Santa Clara, CA – September 2012