Technical Report

SMB 3.0 Multichannel

Accelerate SMB 3.0 Performance for Applications

Brahmanna Chowdary Kodavali, Ron Pratt, NetApp
January 2019 | TR-4740

Abstract

Microsoft has introduced a new feature called Multichannel in the SMB 3.0 protocol in Windows Server 2012 and Windows 8, with the goal of improving the SMB3 protocol by addressing the performance and reliability limitations of SMB1 and SMB2. This technical report is an overview of the Multichannel feature in NetApp® ONTAP®, including its capabilities, best practices, and performance test results.

NetApp®

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

1 SMB 3.0 Multichannel Overview

NetApp ONTAP 9.4 introduces Multichannel, a new SMB 3.0 protocol feature. Multichannel was first introduced by Microsoft in Windows Server 2012 and Windows 8. The goal of the Multichannel feature in SMB3 is to improve the SMB3 protocol by addressing the performance and reliability limitations of SMB1 and SMB2.

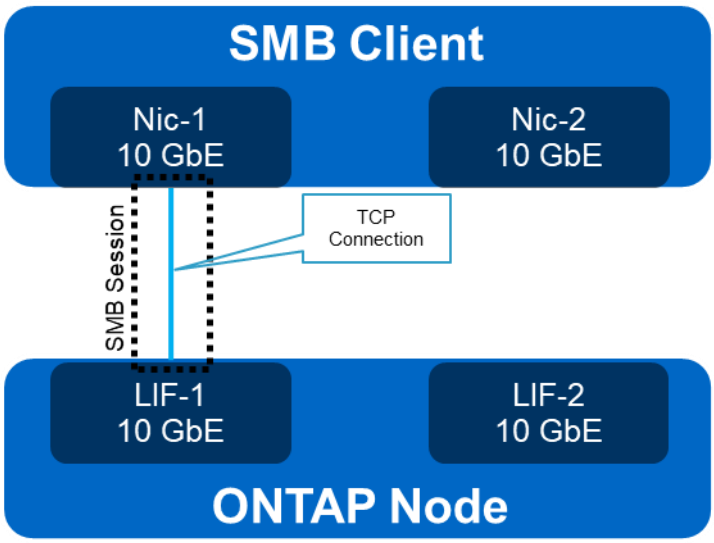
2 How Does SMB 3.0 Multichannel Work?

The Multichannel feature enables an SMB3 client to establish a pool of connections over a single network interface card (NIC) or multiple NICs and use them to send requests for a single SMB session. In contrast, SMB1 and SMB2, by design, require the client to establish one connection and send all the SMB traffic for a given session over that connection. This single connection limits the overall protocol performance that can be achieved from a single client.

2.1 Sessions and Connections

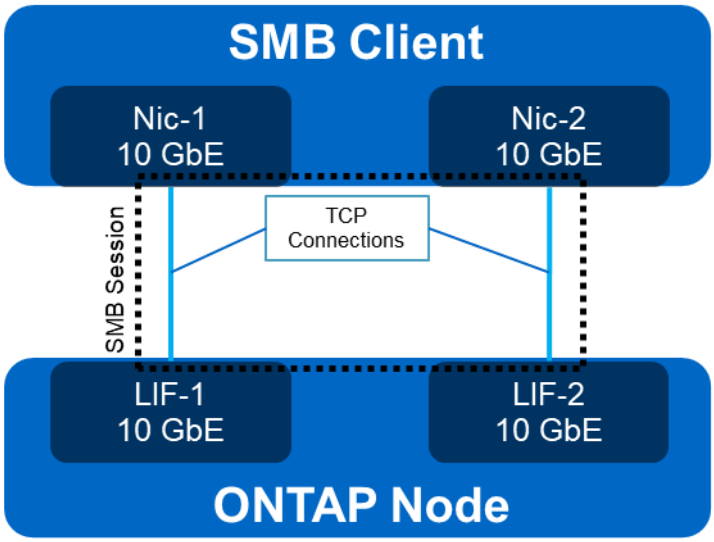
Performance limitations are especially apparent in environments in which both the client and the server have multiple NICs, which is often the case when the client is a Windows Server in a data center. Additionally, a single connection limits the reliability of the protocol. With the exception of SMB3 continuously available shares, any event that results in the loss of an SMB connection usually results in complete or partial disruption to the application.

Figure 1) SMB 1.0, SMB 2.0, and SMB 3.0 (without Multichannel) allow one TCP connection per SMB session.



Prior to SMB 3.0, the SMB protocol imposed a tight coupling between a CIFS session and the TCP connection over which the session is established. In other words, a CIFS session is established over one TCP connection, and all the SMB requests for that session are sent over it. Because throughput is limited to what can be achieved over a single TCP connection, a single NIC between a client and a server limits the throughput that can be achieved over a single CIFS session.

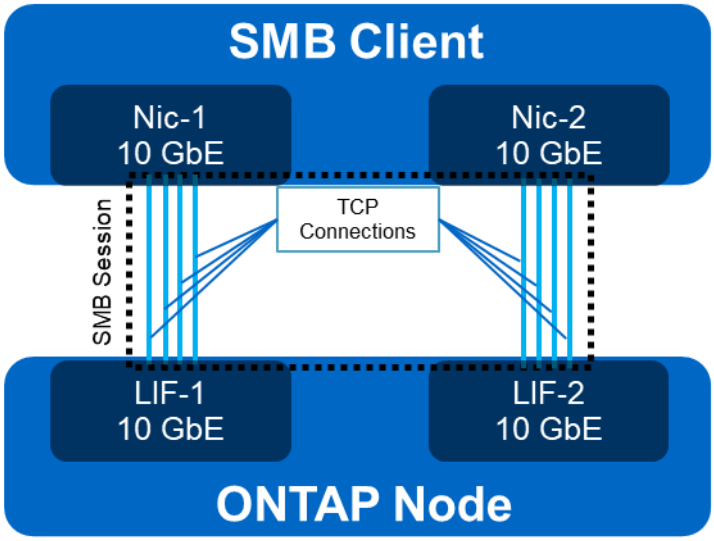
Figure 2) SMB 3.0 Multichannel allows multiple TCP connections per SMB session.



In addition to the throughput limitation, the current model lacks network fault tolerance, because a failure in a NIC or a switch, or a network glitch, can interrupt the session.

Multichannel enables an SMB3 client to establish multiple TCP connections to an SMB3 server, possibly over multiple NICs or even over a single receive-side scaling (RSS)-capable NIC, and associates a single CIFS session with the multiple connections. When more than one TCP connection is established, RSS-capable NICs can use more cores.

Figure 3) With RSS-enabled NICs, SMB 3.0 Multichannel creates four TCP connections per NIC.



2.2 Session Binding and Interface Discovery

Session binding is an important change in the SMB3 protocol that uses multiple connections for a single session. This mechanism enables the existing SMB3 session to associate with another connection.

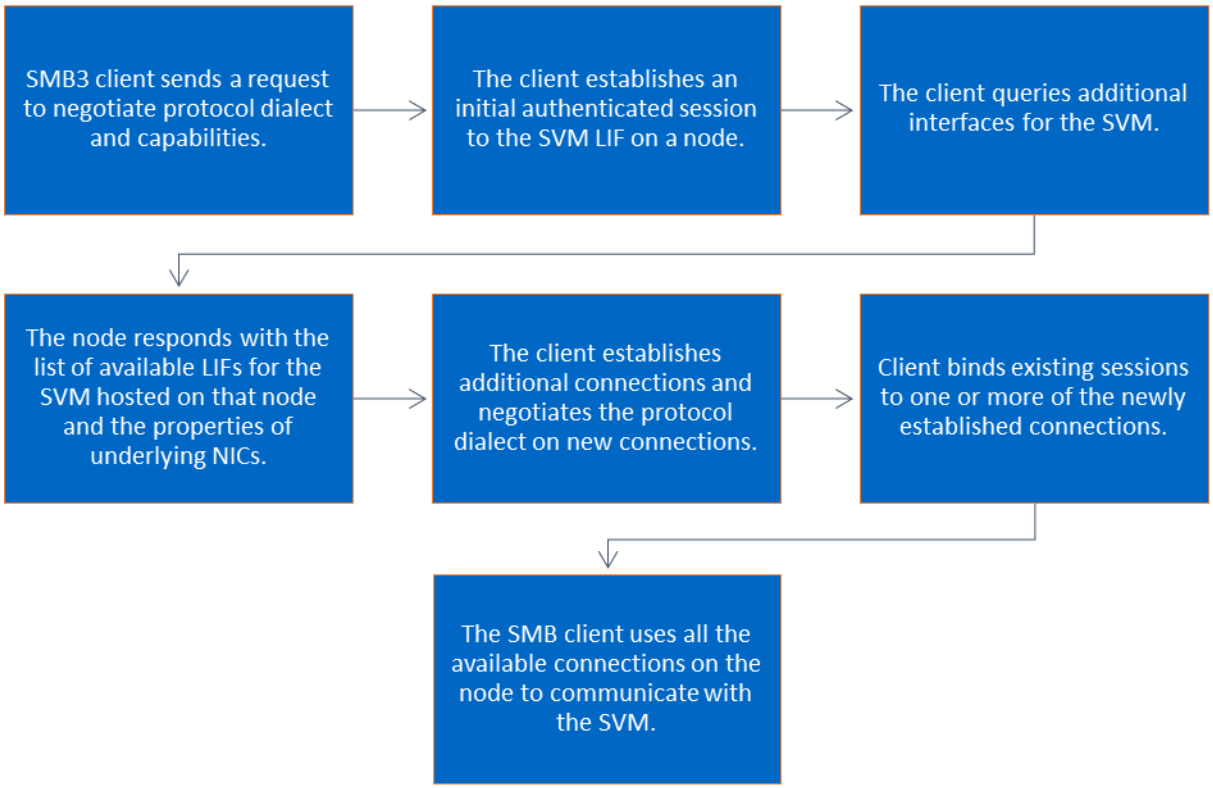
Initial session establishment in SMB3 is similar to the previous versions of the protocol. After a session is established, an SMB3 client performs interface discovery. It then establishes more connections and binds the existing session to these new connections.

The SessionSetup request goes through the normal authentication process just like a regular session setup request. The client is required to provide a security BLOB to complete NTLM/ or Kerberos authentication. Session binding is established after successful authentication.

Multichannel Workflow

Figure 4 describes the workflow for the SMB3 Multichannel feature.

Figure 4) Multichannel workflow for SMB3.



Interface Discovery

Interface discovery is a process to detect any changes in the network. It happens after the initial session setup and every 15 minutes after that. In the interface discovery process, the client sends a new input-output control (IOCTL) command (`fsctl_query_network_interfaces`) to query the list of network interfaces associated with the storage virtual machine (SVM). Each node responds to the request and returns the list of available LIFs for the SVM that are currently hosted on that node and the properties of underlying NICs. The properties returned are:

- The IP address
- The interface index
- The speed of the underlying port
- The RSS/remote direct memory access (RDMA) capability of the port

The client then matches the interfaces from the nodes with available interfaces on the client and determines how many connections to establish. After alternate connections are established, the client binds the existing session to each new connection and starts sending subsequent requests over new connections.

The algorithm to determine the number of connections and the selection of server interfaces depends on the client implementation. Windows clients by default establish up to four connections per interface and limit the total number of connections to a given server to 32.

RSS capability in ONTAP

All of the 10G NICs supported in ONTAP are RSS-capable. Even if the NIC is not RSS-capable, ONTAP simulates RSS by classifying incoming packets based on a software hash. In summary, with or without RSS, ONTAP is capable of leveraging multiple CPUs for processing incoming packets in both the driver and the network stack. Therefore, all interfaces are reported as RSS capable, even if the underlying NIC is not in fact RSS-capable.

Connection selection for requests and responses

After multiple connections are established, the client is allowed to send SMB3 requests pertaining to a given session over any associated connection. To achieve higher throughput, clients are expected to distribute requests in an implementation-specific manner. Windows clients use a round-robin mechanism. The response is sent over the same connection, and ONTAP doesn't select a connection for a response.

2.3 Nondisruptive Behavior

The Multichannel feature allows multiple connections to be associated with an SMB session. Therefore, loss of a subset of connections looks nondisruptive to the application or user, provided that there is at least one other active connection.

Session state and the open files are not impacted when only a subset of connections is lost. This is the key change in behavior introduced by Multichannel that makes the protocol more resilient to connection failures. In addition, on a connection loss, the client replays any outstanding requests over other connections that are available. Replay semantics are not specific to Multichannel and are supported in ONTAP 8.2 and later. As a result, the application or user does not experience disruption.

Windows clients attempt to reestablish a new connection each time an existing connection is lost.

When the last connection associated with a session is destroyed, the session is destroyed as if Multichannel were not present. Therefore, the client might see disruption with the same failure semantics as with SMB3 without Multichannel.

3 SMB Features Supported with Multichannel

SMB offers multiple features relating to security, performance, and resiliency, and Multichannel complements the performance effect of these features. Multichannel comes into the picture during the session setup process.

Features like node referrals and the witness protocol move the client connection, depending on specific criteria. After the client connection is moved to the partner node, the client negotiates the dialect and the session setup, so that Multichannel doesn't break or block any of those features.

Table 1) SMB features that work with SMB3 Multichannel.

SMB Feature	Works with Multichannel
Signing	Yes
Encryption	Yes
Node referrals	Yes
Witness	Yes
Continuously available shares	Yes
NetApp FPolicy® and antivirus	Yes (RPC and SMB communication can leverage Multichannel.)

4 Multichannel Behavior in Cluster Architecture

In an ONTAP cluster, a given SVM can have LIFs hosted on multiple nodes. In the current architecture, the CIFS runtime state (connections, sessions, open files, and so on) is a node-scoped, in-memory state that is not shared between multiple nodes. Node1 is not aware of CIFS sessions established on Node2.

Multichannel does not change the scope of the CIFS runtime state. Because a given CIFS session is visible to only one node, the SMB3 client can be allowed to bind the session to the connections that are established to the LIFs hosted on only that node. When the SMB3 client queries for a list of interfaces, the node returns only the locally hosted LIFs.

4.1 LIF Migration

The client can establish multiple connections to the LIFs that are on the same node and bind a single SMB3 session to all the connections. A Multichannel client connected over multiple LIFs on a node does not experience the disruption unless all the LIFs are migrated from the node. There is a possibility of one of the LIFs migrating to another node in a process initiated by an administrator, or perhaps as a result of auto-revert. When a LIF is migrated, existing connections associated with it are closed. When LIF migration results in connection loss, the client attempts to reestablish the connection immediately.

4.2 Takeover and Giveback

During the takeover or giveback process, LIFs fail over to the rest of the nodes or revert back to the home node in the cluster according to failover policy rules. From an SMB3 client perspective, all the connections are lost, and the client must reconnect to one of the interfaces to establish a new session. The behavior seen by the client is the same with and without Multichannel. If only some of the LIFs migrate away, then the behavior seen by the client is similar to a LIF migration.

Note: During LIF migration, takeover, and giveback, after a LIF moves out of the home node, the connection over that LIF is terminated on the home node. The partner node onto which the LIF has moved does not participate immediately in the current SMB session. On the next LIF query, the client establishes a connection over the LIF and associates with the corresponding SMB session.

5 Setting up Multichannel

Multichannel requires support on ONTAP as well as on the host or client establishing the SMB connection to ONTAP. The following list details these requirements:

- NetApp ONTAP version: 9.4 or later
- Microsoft Windows Server: Windows 2012 or later
- Microsoft Windows Client: Windows 8.0 or later
- SMB Protocol version: 3.0
- At least one of the following network interface card configurations is required:
  - Multiple NICs are available on a node and client to establish multiple TCP connections
  - One or more network adapters that support RSS
  - One of more network adapters configured with NIC teaming
  - One or more network adapters that support remote direct memory access (RDMA)

5.1 Enabling Multichannel on ONTAP

The Multichannel feature can be enabled or disabled by modifying the option `is-multichannel-enabled` under the `vserver cifs option` command directory:

```
vserver cifs*> options modify -vserver fsvcs -is-multichannel-enabled true
```

To verify these settings, run the following commands:

```
vserver cifs*> options show -vserver fsvcs -fields is-multichannel-enabled
vserver is-multichannel-enabled
------- ------------------------
fsvcs   true
```

On Windows, Multichannel is enabled by default. To enable or disable the option, use the following commands with Windows PowerShell:

```
Enable/disable multichannel feature. (Default: true)
Set-SmbClientConfiguration -EnableMultiChannel $true

Configure the maximum total number of connections per client/server pair using:
Set-SmbClientConfiguration –MaximumConnectionCountPerServer <n>

Configure the number SMB Multichannel connections per RSS-capable network interface
Set-SmbClientConfiguration -ConnectionCountPerRssNetworkInterface <n>
```

In addition to these commands, Multichannel adds two more CIFS options, `-max-connections-per-session` and `-max-lifs-per-session`, which are affected during upgrade and revert. The maximum value for the `-max-connections-per-session` parameter is 32. The maximum value for the parameter `-max-lifs-per-session` parameter is 256.

When the last node in the cluster is upgraded to the 9.4 release, the default values for all the Multichannel options are populated. By default, Multichannel is disabled.

When the first node is reverted from the 9.4-configured clusters, a revert check forces the user to disable the Multichannel feature on the node. Therefore, revert causes all Multichannel connections to be closed, which is disruptive to any clients using this feature.

Note: The Multichannel feature requires some enhancements to correctly handle locks that are usable over multiple connections. The Multichannel feature can work correctly only if the node hosting the LIF and the node hosting the volume are upgraded to 9.4. The Multichannel feature cannot be enabled in a mixed-node cluster. If the administrator tries to enable Multichannel with 9.4 nodes by modifying the `-is-multichannel-enabled` option, the operation fails if the effective cluster version is not 9.4.

5.2 Verify Multichannel Functionality

The Multichannel option in CIFS is to administer the configuration, but the functionality depends on multiple factors.

Multichannel works only if clients support SMB 3.0 or later, if they can negotiate the SMB 3.0 protocol, and if multiple NICs are available on a node, allowing the client to establish multiple TCP connections.

5.2.1 CLI Commands

Multichannel functionality can be verified through the `cifs session show` and the `cifs connection show` commands.

cifs session show

The `cifs session show` command provides the session ID and the connection count parameters. The number shown in the Connection Count column represents the number of connections created between the client and the node that are associated with a specific session ID. Each connection has a connection ID, but this command output is truncated and shows only one connection ID. The `cifs connection show` command provides all the connection details.

```
stg-lab-mc::> cifs session show

Node:    stg-lab-mc-02
Vserver: fsvcs
Connection Session                                      Open         Idle      Connection
ID         ID      Workstation      Windows User        Files        Time           Count
---------- ------- ---------------- ---------------- --------- ------------ ---------------
438731506  5635691983701270642                                     3s               8
                   10.10.56.80     VEGA\                    2
                                    Administrator

stg-lab-mc::> cifs session show -session-id 5635691983701270642 -fields connection-count
node                vserver session-id          connection-id connection-count
------------------- ------- ------------------- ------------- ----------------
stg-lab-mc-02        fsvcs   5635691983701270642 438731506     8
```

cifs connection show

The `cifs connection show` command with the `-session-id` parameter presents the details of the connections associated with that session ID: connection IDs, LIF details, and the workstation IP. Without the -s`ession-id` parameter, this command shows all the connections and associated `session-id` details.

```
stg-lab-mc::> cifs connection show -session-id 5635691983701270642
Node:    stg-lab-mc-02
Vserver: fsvcs
Connection Session                              Workstation
ID         IDs                   Workstation IP  Port        LIF IP
---------- --------------------- --------------- ----------- ---------------
438731506  5635691983701270642   10.10.56.80    55470       10.10.59.130
438731509  5635691983701270642   10.10.59.131   55475       10.10.59.130
438731510  5635691983701270642   10.10.59.131   55476       10.10.59.130
438731511  5635691983701270642   10.10.59.131   55477       10.10.59.130
438731512  5635691983701270642   10.10.59.131   55478       10.10.59.130
438731513  5635691983701270642   10.10.56.80    55479       10.10.59.130
438731514  5635691983701270642   10.10.56.80    55480       10.10.59.130
438731515  5635691983701270642   10.10.56.80    55481       10.10.59.130
8 entries were displayed.
```

6 Performance

One of the key goals of the Multichannel feature is to improve the performances for reads and writes. Multichannel accomplishes this by creating multiple connections across different interfaces and binding them to one session. It then uses algorithms like round robin (observed on packet traces) to efficiently multiplex the requests on this pool of connections. This naturally results in Multichannel sessions using CPU resources for a longer time compared to non-Multichannel sessions. Therefore, the performance of Multichannel sessions is higher.

We performed tests to compare the performance of SMB3 with Multichannel enabled or disabled. The first test was with multiple instances of SQL Server with an OLTP workload. The second test was performed with a single instance of SQL Server using an OLTP workload to demonstrate the overall performance gains achieved by enabling Multichannel on a server.

Figure 5 shows a storage-side performance comparison using multiple instances of SQL server with Multichannel enabled and disabled. The operations per second (Ops) achieved with Multichannel disabled was ~185,000 at a latency of 821µs. SMB3 with Multichannel enabled shows an improvement of ~60%, with ~280,000 Ops and a latency of ~960µs. The average operation size for all operations in both test runs was 8,000.

Figure 5) SMB Multichannel comparison—storage side.

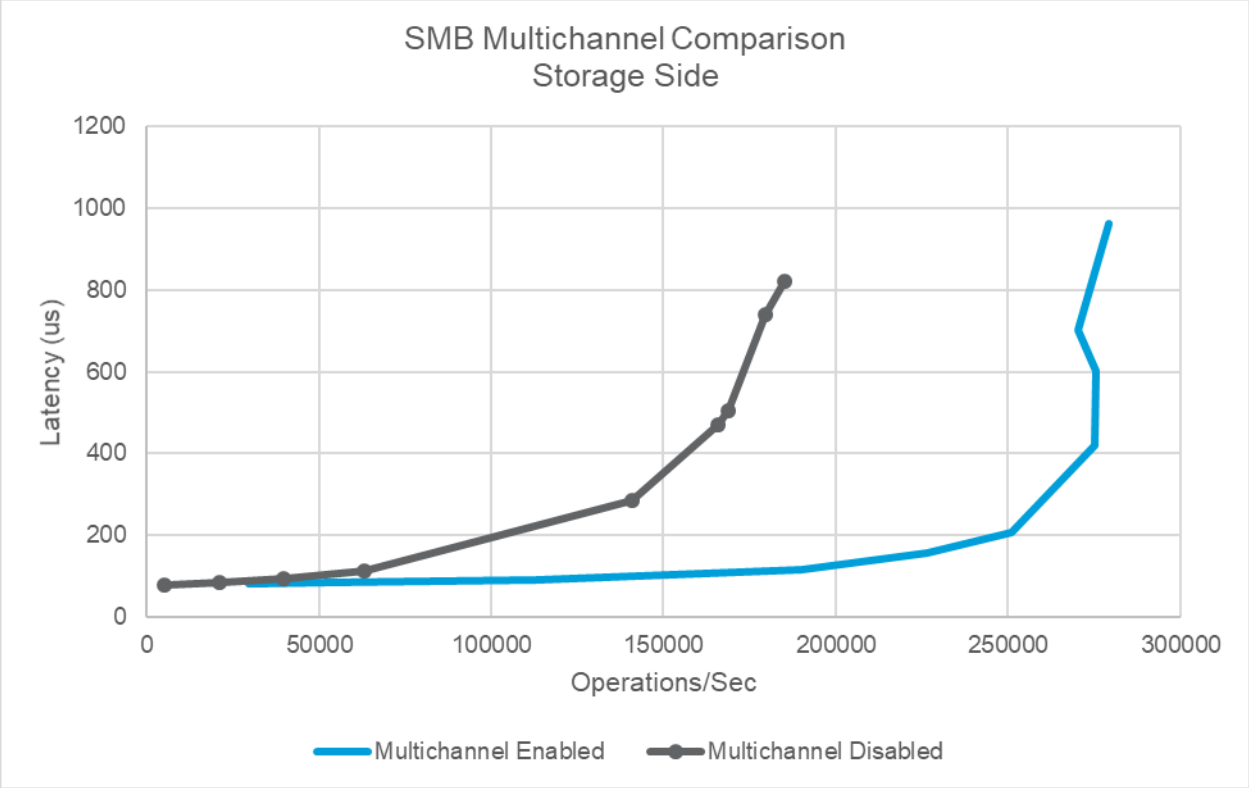
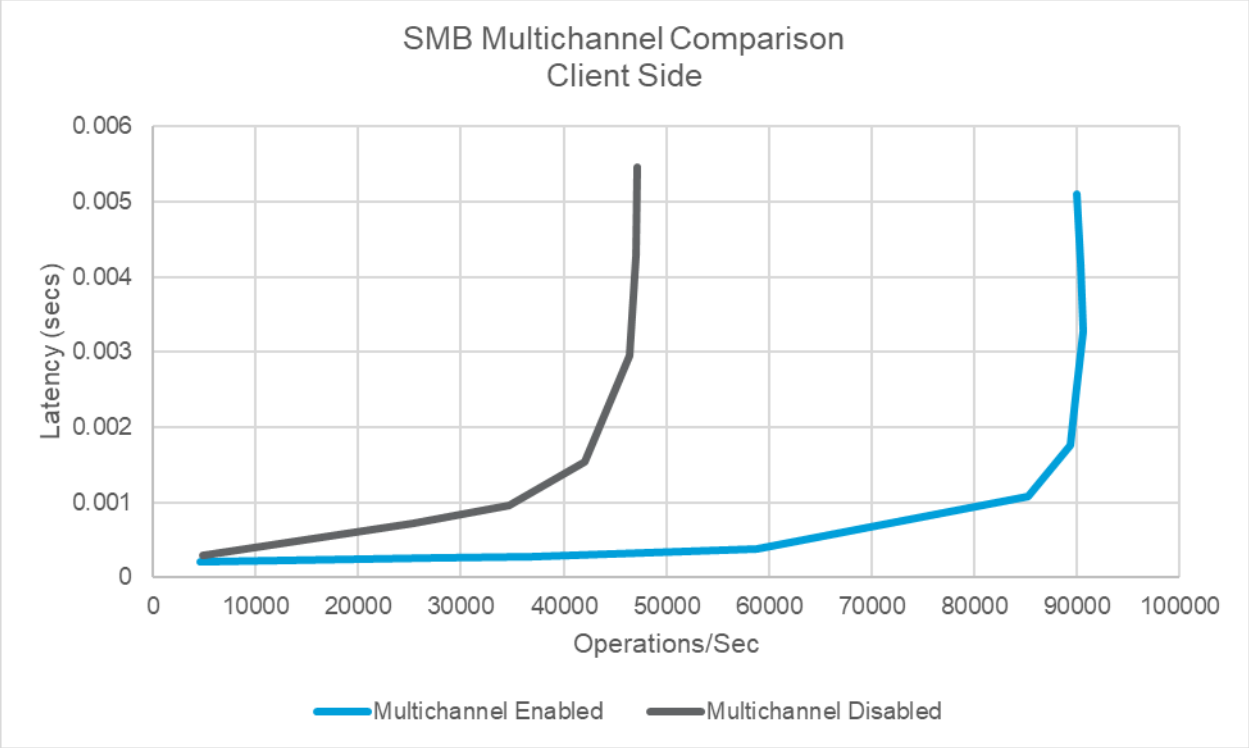


Figure 6 shows performance for a single SQL Server instance. The data was collected from Windows Performance Monitor counters and represents performance as seen from the SQL Server side. This graph shows that performance with SMB Multichannel enabled was almost 50% better at ~90,000 Ops at 5ms, with ~45,000 Ops at ~5ms with SMB Multichannel disabled. The average operations size for all operations was 8,000.

Figure 6) SMB Multichannel comparison—client side.



6.1 Multichannel Counters

Multichannel-related counters are in the `smb2_ctx` object. These are per network, context-based counters.

Table 2) Multichannel counters.

Counter Name	Counter Description
`session_bind_ops`	Number of SMB 3.0 Multichannel session binds between the client and server
`session_token_attempt`	Number of attempts to grab a session token for SMB 3.0 Multichannel connections
`session_token_granted`	Number of times a session token is granted for SMB 3.0 Multichannel connections
`session_token_denied`	Number of times a session token is denied for SMB 3.0 Multichannel connections
`max_time_session_token_held`	Maximum amount of time a session token is held for SMB 3.0 Multichannel connections
`max_time_command_spent_to_acquire_session_token`	Maximum amount of time an SMB 3.0 command spends to acquire a session token
`session_token_held_latency_histogram`	Histogram for total time that an SMB 3.0 Multichannel session token is held

`time_command_spent_to_acquire_session_token_histogram`	Histogram for total time taken by SMB 3.0 requests to acquire a Multichannel session token

7 Troubleshooting and Debugging

There are a number of reasons why the Multichannel feature might not work. Follow this workflow to identify the root cause and fix the problem. If this workflow does not correct your problem, initiate a support case with NetApp support.

7.1 Multiple connections are not established.

Possible cause: Multichannel is not enabled on the client and/or server.

How to check	How to resolve

Check whether Multichannel is enabled on the client side by using the following Windows PowerShell cmdlet:
`Get-SmbClientConfiguration | Select EnableMultichannel`
Check whether Multichannel is enabled on ONTAP by using the following command:
`vserver cifs options show -vserver vs1 -field is-multichannel-enabled`
	Enable Multichannel on the client by using the following Windows PowerShell cmdlet:
	`Set-SmbClientConfiguration -EnableMultiChannel $true`
	Enable Multichannel on ONTAP (SVM setting) by using the following command (advanced mode):
	`vserver cifs options modify -vserver vs1 -is-multichannel-enabled true`

Possible cause: Multichannel is enabled, but the client has an incompatible network interface configuration.

How to check	How to resolve

Verify that the client has multiple NICs and/or verify the RSS and RDMA capabilities of the NICs by using the following windows PowerShell cmdlets:
`Get-NetAdapter`
`Get-NetAdapterRSS`
`Get-NetAdapterRDMA`
`Get-NetAdapterHardwareInfo`
`Get-SmbClientNetworkInterface`
	The network interface adaptor must meet the hardware requirements listed in the "Setting up Multichannel" section of the Software and Hardware Requirements table.

Possible cause: Client is not running any read/write traffic or is running very little.

How to check	How to resolve
Check the client and server versions and determine the workload against the client/server. Multichannel on Windows Server kicks in as soon as read/write operations are issued because network fault tolerance is the key priority on the server. Windows client behavior is not different.	Multichannel is used only when there are few I/Os in flight at the same time. This is determined by the WindowSizeThreshold setting. The default value is 8, which means that it is triggered when there are at least eight packets in flight asynchronously. Single small file copy does not trigger Multichannel. Ideally, it does not start for every connection from a client, especially if it is doing just a small amount of work.

Note: For more information, see this Microsoft article.

7.2 Multichannel is working fine but the number of channels that are established is different from what is expected.

Possible cause: You might have chosen a different value than the default parameter settings.

How to check	How to resolve

Check the client-side settings by using the following Windows PowerShell cmdlet:
`Get-SmbClientConfiguration | select ConnectionCountPerRssNetworkInterface,MaximumConnectionCountPerServer`
The default value for `ConnectionCountPerRssNetworkInterface` is 4.
The default value for `MaximumConnectionCountPerServer` is 32.
Check the server-side settings by using following command:
`vserver cifs options show -vserver vs1 -fields max-connections-per-session, max-lifs-per-session`
The default value for `max-connections-per-session` is 32.
The default value for `max-lifs-per-session` is 256.
	If the values are not the default, consider reverting to the default settings.
	Multichannel is a client-driven feature. Therefore, depending upon the algorithm (round robin, shortest queue length, processor affinity, and so on), the client running it dynamically decides whether or not it needs more channels. The server, on the other hand, obliges if the settings are within system constraints**.**

8 Conclusion

With the introduction of applications like Microsoft SQL Server and Hyper-V over SMB 3.0, performance has become a key requirement for the SMB 3.0 protocol. ONTAP combined with Multichannel support brings both performance and resiliency to applications running on the SMB 3.0 protocol.

Where to Find Additional Information

To learn more about the information described in this document, refer to the following documents and/or websites:

- NetApp Documentation Centers
  https://mysupport.netapp.com/info/web/ECMLP2557637.html

Version History

Version	Date	Document Version History	Author(s)
Version 1.0	January 2019	Introduction of Multichannel	Brahmanna Chowdary Kodavali, Ron Pratt

Refer to the Interoperability Matrix Tool (IMT) on the NetApp Support site to validate that the exact product and feature versions described in this document are supported for your specific environment. The NetApp IMT defines the product components and versions that can be used to construct configurations that are supported by NetApp. Specific results depend on each customer's installation in accordance with published specifications.