The benefits of Fabric Vision technology for disaster recovery

Contents

Executive summary

Extending Fabric Vision technology over distance for disaster recovery delivers increased visibility, pinpoints problems and accelerates troubleshooting to maximize the performance of critical applications. This white paper provides an overview of the Fabric Vision features and functionalities that are included within IBM SAN b-type Fibre Channel extension solutions.

Significant challenges face storage administrators in today's data centers. Storage consumes a large quantity of IP network resources. Those IP networks lack adequate visibility of storage flows and are sometimes unreliable, overly complex and very inflexible. Greater complexity creates more opportunities for issues to occur. Storage and network administration provide different roles and responsibilities in an enterprise. Storage administrators do not manage IP networks, and network administrators do not manage storage. When issues arise, each role can tend to blame the other for the problems.

Recovery point objectives (RPOs) for mission-critical data typically require a couple of seconds or less. Such RPOs are difficult to maintain when data rates are high (10 Gbps or above) and when network problems are also present. Such degraded states are difficult to troubleshoot. Often multiple vendors are involved, such as storage vendors, storage network vendors, IP network vendors and wide area network (WAN) service providers. Such situations cost organizations considerable money and expose the business to data loss. The end result is that user expectations are not met.

To address these problems within the specific context of extension, IBM has introduced the IBM® System Storage® SAN42B-R extension switch with advanced Fabric Vision capabilities. Fabric Vision includes a number of monitoring, alerting and reporting tools specific to b-type extension. Additionally, diagnostic tools useful for determining IP network validation and health are available. The objective is to determine quickly the root cause of degraded situations or outages and expedite a return to normal operations as quickly as possible.

The situation

IBM customers pose the question, "How can we resolve support issues more quickly and effectively?" Often storage vendor support organizations report that they are contacted after the remote data replication (RDR) application is already down, resulting in emergency measures for disaster recovery. This situation is further aggravated by the inability to pinpoint quickly whether the problem is a network or storage issue. The IBM SAN b-type Fibre Channel (FC) products portfolio can help you proactively monitor and effectively troubleshoot your local FC connections, network device health, and the ability of your IP network to meet service level agreements (SLAs). Storage arrays are not capable of providing proactive warnings or identifying network problems.

Storage administrators face conditions such as:

- **Unreliable IP networks.** Storage administrators have to rely on the experience, capabilities, availability and accuracy of network administrators to provide feedback about networks during performance assessment and troubleshooting. This exchange of information can prolong the length of a degraded state.
- **Increased scale and complexity of data centers.** Virtualization of every type is used at every level, now more than ever. This means more opportunities for errors, with each error potentially affecting other processes. Pinpointing what went wrong is not always a simple task.
- **Asymmetric behavior.** Does outbound data take one path and data returns take another path? Are the paths equal in quality? Can the IP network properly load balance?
- **Different administration groups.** At least two administrative groups (storage and network) exist within an enterprise, and each group has different roles, training, expertise and responsibilities—possibly even different management personnel. These differences can cause difficulties for storage administrators and network administrators alike. Storage personnel might not understand the network infrastructure and requirements, and network personnel might not understand the storage infrastructure and requirements.
- **RPO at high data rates.** RPO has changed over time. The amount of data and the number of applications managed has grown significantly. This means that RPO times must shrink. The amount of data processed in even a short amount of time has become significant and represents critical business transactions.
- **Catch-up time from outage backlogs.** Data rates are in the 10 Gbps range for many companies, so when an outage or degraded state occurs, it is more difficult to reestablish a state of a fully synchronized consistent data rate useful for a database restart. If the amount of time it takes to regain a consistent state is exceedingly long, the exposure to disaster becomes a liability.

Now, if replication is stressed, is slower than expected, or might be suspended, what does the timeline to resolution look like? Refer to Figure 1. Generally, a storage administrator is not aware of any problems until an error condition occurs. Storage administration might then place a request into network operations to investigate why replication is not performing optimally across the WAN. More often than not, the IP network will immediately be cleared of any issues. After all, if the IP network is not down, other applications are fine and a ping can get through, and if links are not fully utilized then the assumption is that the network must be functional. To solve the quandary, it is likely that the storage administrator will open a support call with the storage array vendor to determine the cause of the issue. Rudimentary troubleshooting to validate basic issues will establish a baseline for the investigation. After that, any call that involves replication across an IP network requires establishing whether the problem exists in the network or in the array. At that point, the process of ruling out issues with the IP network commences.

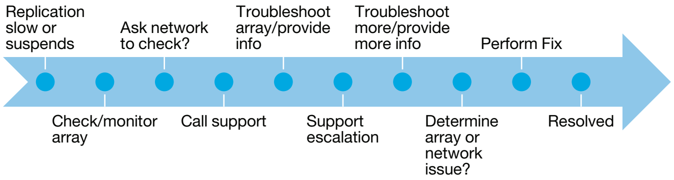

Replication slow or suspends
Check/monitor array
Ask network to check?
Call support
Troubleshoot array/provide info
Support escalation
Troubleshoot more/provide more info
Determine array or network issue?
Perform Fix
Resolved

Figure 1. A typical outage timeline

How does the IBM SAN42B-R extension switch address this type of situation? Given the many aspects of storage networks, every aspect needs to be continuously monitored and inspected for proper operation.

In this case, the SAN42B-R switch monitors the IP network in which its managed tunnel flows, then alerts the storage administrator proactively when conditions arise that indicate degraded states. Degraded states within the IP network include: transient congestion events, chronic out of order delivery, instability and frequent network rerouting, oversubscription, data integrity problems, excessive latency, excessive jitter and more. The capabilities of the SAN42B-R switch enable storage administrators to gain visibility into an infrastructure on which they rely, but have had no visibility into before now.

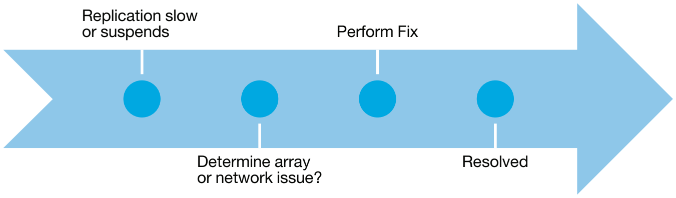

Replication slow or suspends
Determine array or network issue?
Perform Fix
Resolved

Figure 2. Preemptive outage timeline

This raises another problem. Storage administrators cannot spend all of their time monitoring networks. The network itself must have the intelligence to create an alert when an issue becomes suspect. IBM has incorporated decades of expertise and experience into improving network intelligence. When an alert is created by the Monitoring and Alerts Policy Suite, it is based on the experience and expertise of the MAPS technology. This allows storage administrators to regain valuable time.

Quicker time to resolution

Fabric Vision technology provides an innovative hardware and software solution that simplifies monitoring, maximizes network availability and helps to dramatically reduce costs. Featuring monitoring, management and diagnostic capabilities, Fabric Vision technology enables administrators to preempt problems before they impact operations. This all helps organizations meet SLAs, primarily the SLA of maintaining the desired RPO.

This section lists the features and functionality that have been incorporated into IBM SAN b-type Fibre Channel extension. These features help preempt support issues altogether—or, when an issue cannot be totally preempted, these features help assure its quick resolution.

IBM Network Advisor dashboard

Every storage administrator approaches the task of managing and troubleshooting the environment differently. The IBM Network Advisor software management tool has a customizable dashboard, as shown in Figure 3. The customizable dashboard displays the monitors, counters and status indicators of importance to you. The IBM Network Advisor tool comes with over one hundred dashboard widgets. If the information you require is not available through an already existing widget, the Network Advisor comes with the tools to create it. When you know exactly what is happening in your network, the goal of continuous uptime becomes a reality.

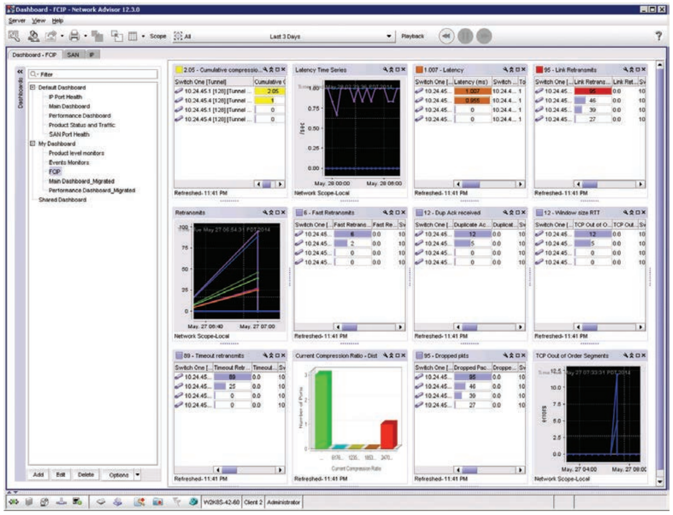
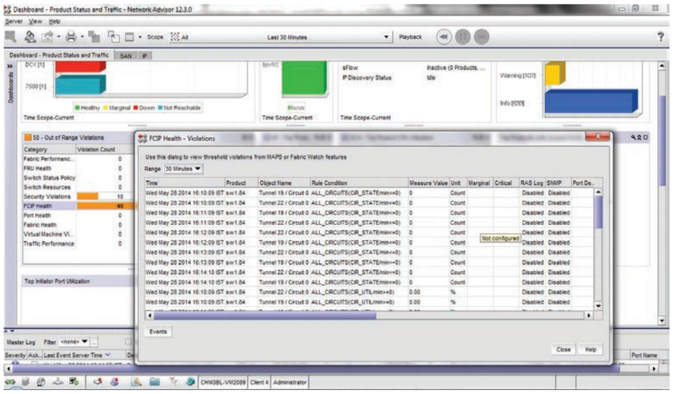


Figure 3. IBM Network Advisor dashboard

If a dashboard item is indicating an event or status, clicking that item will allow you to drill down into more specific information. In the example shown in Figure 4, a window showing FCIP health violations is open to display more detailed information. The detailed information shown here includes:

- Timestamp
- Device name
- Tunnel and circuit designation
- The rule that was violated
- The offending value that was measured, and the units of that measurement

You can also see if it was registered in the RAS log, Simple Network Management Protocol (SNMP), or email, as well as the level of violation (marginal or critical) and other information.



50 - Out of Range Violations

Category	Violation Count
Fabric Performanc...	0
FRU Health	0
Switch Status Policy	0
Switch Resources	0
Security Violations	10
FCIP Health	40
Port Health	0
Fabric Health	0
Virtual Machine Vi...	0
Traffic Performance	0

FCIP Health - Violations

Use this dialog to view threshold violations from MAPS or Fabric Watch features

Range: 30 Minutes

Time	Product	Object Name	Rule Condition	Measure Value	Unit	Marginal	Critical	RAS Log	SNMP	Port De...
Wed May 28 2014 16:10:09 IST	sw1.84	Tunnel 19 / Circuit 0	ALL_CIRCUITS(CIR_STATE/min<=0)	0	Count			Disabled	Disabled	
Wed May 28 2014 16:10:09 IST	sw1.84	Tunnel 22 / Circuit 0	ALL_CIRCUITS(CIR_STATE/min<=0)	0	Count			Disabled	Disabled	
Wed May 28 2014 16:11:09 IST	sw1.84	Tunnel 19 / Circuit 0	ALL_CIRCUITS(CIR_STATE/min<=0)	0	Count			Disabled	Disabled	
Wed May 28 2014 16:11:09 IST	sw1.84	Tunnel 22 / Circuit 0	ALL_CIRCUITS(CIR_STATE/min<=0)	0	Count			Disabled	Disabled	
Wed May 28 2014 16:12:09 IST	sw1.84	Tunnel 22 / Circuit 0	ALL_CIRCUITS(CIR_STATE/min<=0)	0	Count		Not configured	Disabled	Disabled	
Wed May 28 2014 16:12:09 IST	sw1.84	Tunnel 19 / Circuit 0	ALL_CIRCUITS(CIR_STATE/min<=0)	0	Count			Disabled	Disabled	
Wed May 28 2014 16:13:09 IST	sw1.84	Tunnel 22 / Circuit 0	ALL_CIRCUITS(CIR_STATE/min<=0)	0	Count			Disabled	Disabled	
Wed May 28 2014 16:13:09 IST	sw1.84	Tunnel 19 / Circuit 0	ALL_CIRCUITS(CIR_STATE/min<=0)	0	Count			Disabled	Disabled	
Wed May 28 2014 16:14:10 IST	sw1.84	Tunnel 22 / Circuit 0	ALL_CIRCUITS(CIR_STATE/min<=0)	0	Count			Disabled	Disabled	
Wed May 28 2014 16:14:10 IST	sw1.84	Tunnel 19 / Circuit 0	ALL_CIRCUITS(CIR_STATE/min<=0)	0	Count			Disabled	Disabled	
Wed May 28 2014 16:10:09 IST	sw1.84	Tunnel 22 / Circuit 0	ALL_CIRCUITS(CIR_UTIL/min>=0)	0.00	%			Disabled	Disabled	
Wed May 28 2014 16:10:09 IST	sw1.84	Tunnel 19 / Circuit 0	ALL_CIRCUITS(CIR_UTIL/min>=0)	0.00	%			Disabled	Disabled	

Figure 4. IBM Network Advisor dashboard with FCIP violations pane

Command line interface dashboard

The command line interface (CLI) dashboard is available for users that prefer to use the CLI instead of the IBM Network Advisor tool. All information from each monitoring category can be obtained from the CLI, including any out-of-range conditions and the rules that were triggered. Error statistics provide at-a-glance views of status and the various contributing conditions to that status. Historical switch information for up to the last seven days provides a variety of error counters, giving users instant visibility into problem areas and facilitating the decision process towards proper resolution and planning.

Monitoring and Alerts Policy Suite

The Monitoring and Alerts Policy Suite (MAPS) is a simple to configure and use solution for policy-based threshold monitoring and alerting. MAPS proactively monitors the health and performance of infrastructure to ensure application uptime and availability. By using prebuilt rules and policy-based templates, MAPS simplifies threshold configuration, monitoring and alerting. Organizations can configure one, multiple or all fabrics at once using common rules and policies, or they can customize policies for specific ports, switch elements and items—all through a single dialog. As shown in Figure 5, the integrated dashboard displays an overall switch health report, along with details on any out-of-range conditions. Administrators can quickly pinpoint potential issues and easily identify trends and other aberrant behaviors occurring within their fabric. MAPS for extension monitors and generates alerts for the following conditions:

- Tunnel/Trunk state change
- Tunnel/Trunk Overall throughput
- Tunnel/Trunk PTQ throughput
- Tunnel/Trunk PTQ DupAck (duplicate acknowledgements)
- Tunnel/Trunk PTQ packet loss
- Tunnel/Trunk PTQ slow starts
- Circuit state change
- Circuit utilization
- Circuit packet loss
- Circuit round trip time
- Circuit jitter

For the per priority TCP QoS (PTQ) monitors, there is a separate monitor for each priority: class F, high, medium and low.

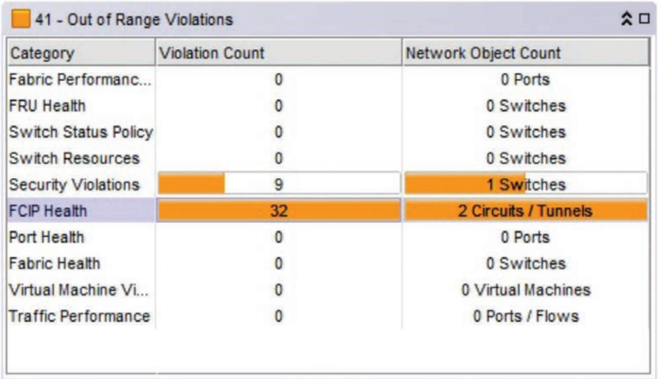
This set of monitors enables the detection of just about any degraded IP network condition that might occur. If the network goes down for a period longer than the keepalive timeout value for a circuit, that event is detected and an alert is processed. If the IP network experiences transient congestion that results in either excessive jitter or packet loss, that event is detected and an alert is processed. If any one of the PTQ priorities suffers low throughput, packet loss or out-of-order events, that issue is detected and an alert is processed. If the network reroutes, causing the RTT to change for the worse, that event is detected and an alert is processed. MAPS provides comprehensive detection of various network events and the ability to make those events known to storage administrators. Storage administrators can use this information to enforce IP network SLAs.

Tunnels are a managed transport between two b-type extension endpoints. There are many reasons for using a tunnel as a data transport across infrastructure, including granular load balancing, lossless failover/failback, higher availability, encryption, bandwidth pooling, protocol optimization, congestion management, quality of service marking and enforcement, monitoring, reporting and network diagnostics. A tunnel uses extension trunking and may consist of multiple member circuits, which may traverse one or more service providers, each with a distinct SLA. IBM provides storage administrators with a single point of management and service provider SLA validation. Tunnels are point-to-point, and the endpoints of a tunnel are the same endpoints as the trunk.

There is a hierarchy of extension connectivity. Virtual E_Ports (VE_Ports, or tunnel endpoints,) are in the top tier of the network and define the endpoints of tunnels. VE_Ports contain one or more circuits. Each circuit has four WAN-optimized TCP (WO-TCP) sessions, one for each priority (class F, high, medium and low). MAPS is integrated into each of these tiers. Any indication of a problem in a lower tier poses a problem for its associated upper tiers.

For example, if a problem exists with the IP network path that the medium PTQ WO-TCP session is using, the tunnel in which that circuit is a member will likely also have a problem. This alerts administrators when thresholds are exceeded.

MAPS for tunnels or trunks (VE_Ports) monitors and performs actions based on throughput and state change.



41 - Out of Range Violations

Category	Violation Count	Network Object Count
Fabric Performanc...	0	0 Ports
FRU Health	0	0 Switches
Switch Status Policy	0	0 Switches
Switch Resources	0	0 Switches
Security Violations	9	1 Switches
FCIP Health	32	2 Circuits / Tunnels
Port Health	0	0 Ports
Fabric Health	0	0 Switches
Virtual Machine Vi...	0	0 Virtual Machines
Traffic Performance	0	0 Ports / Flows

Figure 5. MAPS summary pane

Sudden failures and gradual degradation detection

MAPS can monitor both sudden failures and gradually deteriorating conditions. For example, MAPS can alert users if a cyclic redundancy check (CRC) error counter suddenly increases to five per minute or gradually increases to five per day. This is useful for monitoring service provider SLAs. Service provider infrastructure is often immense, and it is difficult for service providers to monitor every optic, cable and device in their network. Fabric Vision provides you with the tools you need to ensure your paths are within the promised SLA. Refer to Figure 6, which illustrates FCIP TCP retransmits.

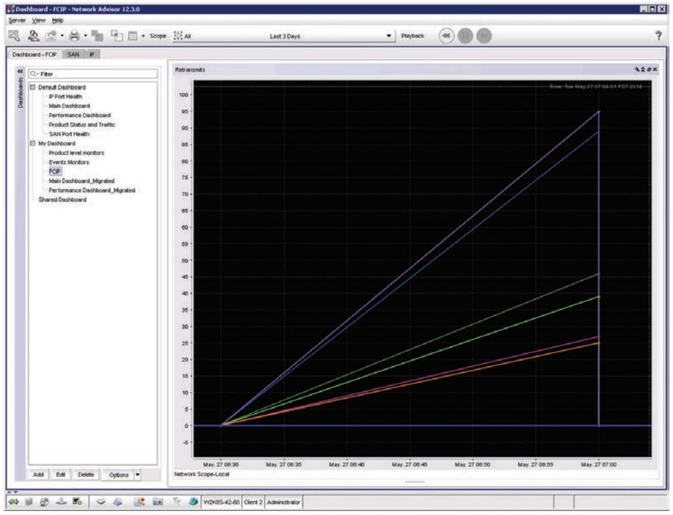


Figure 6. Increasing extension retransmits over 30 minutes

Policy-based monitoring

Policy-based monitoring involves predefined monitoring groups with prevalidated monitoring policies. Multiple monitoring categories enable monitoring of the following: tunnels, circuits, VE_Ports, overall switch status, switch ports, small form-factor pluggables (SFPs), port blades, core blades, switch power supplies, fans, temperature sensors, security policy violations, fabric reconfigurations, scaling limits, CPU, memory use, traffic performance and more.

Predefined monitoring policies are tiered to provide the best starting place for your particular environment and administrative style: aggressive, moderate and conservative. A monitoring group's policy tier can be set individually and, if needed, you can fine-tune your specific environment. Individual items within a group can be set to the appropriate tier for customization. For example, when setting a style for FICON-associated extension connections you might select an aggressive style, and for remote data replication (RDR) extension connections you might select a moderate style.

Custom monitoring groups

You can create custom monitoring groups, such as switch ports attached to high-priority applications and switch ports attached to low-priority applications. Monitoring of each group happens according to the group's unique rules, high-priority applications and low-priority applications.

Flexible rules

Rule flexibility means you can monitor a given counter for various threshold values, taking different actions when each value is crossed. For example, you can monitor a CRC error counter at a switch port and generate a RAS log entry when the error rate reaches two per minute, send an email notification when the error rate is at five per minute, and fence a port when the error rate exceeds ten per minute.

MAPS: actions and alerts

MAPS generates various alerts and actions. Actions include circuit fencing, port fencing, port decommissioning and setting status. Alerts include RAS log messages, SNMP traps and email notifications.

Circuit fencing

Specific to the IBM System Storage SAN42B-R extension switch, the IBM System Storage SAN06B-R extension switch, and the b-type extension blade, when certain thresholds are reached, a circuit can be taken offline. This is an important action for b-type extension, because a degraded circuit causes all circuits belonging to the same VE_Port to be degraded as well. This might result in a total throughput that is less than it would be without the degraded circuit. It is necessary to isolate a degraded circuit from the remaining clean circuits to maintain overall optimal throughput.

Because data is delivered to the upper layer protocol (ULP) in order, a degraded circuit will cause all member circuits to degrade as well. If a trunk has two circuits, and one circuit is degraded such that it requires retransmits to complete successful transmission, the data sent on the clean circuit must wait for the retransmissions before delivering to the ULP. This means that both the clean and degraded circuits will go no faster than the degraded circuit. The degraded circuit may effectively be delivering only a small fraction of the bandwidth apportioned to it, depending on the degree of degradation. By fencing the degraded circuit, the clean circuit operates at the full bandwidth apportioned to it.

Port fencing

The port fencing action takes the port offline immediately when user-defined thresholds are exceeded. Supported port types include VE_Ports as well as physical ports (E_Ports and F_Ports). This action is valid only for conditions evaluated by the actual port.

Port decommissioning

Port decommissioning acts in addition to port fencing. Port decommissioning allows ports to be gracefully shut down. When certain monitored statistics cross defined thresholds, ports are decommissioned, similar to port fencing but without the abrupt traffic disruption.

SFP marginal

The SFP marginal action sets the state of the affected SFP transceiver in the MAPS dashboard to “down.” This action does not bring the SFP transceiver down. It affects only what is displayed in the dashboard. This action is valid only in the context of advanced SFP groups.

RAS log messages

Following an event, MAPS adds an entry to the internal event log for each switch involved. The RAS log stores detailed event information but does not actively send alerts. The IBM Network Advisor tool can be used to collect the RAS log information from multiple switches to a common point. It is easier to check numerous switches from the IBM Network Advisor instance.

SNMP traps

In environments where you have a high number of messages coming from a variety of switches, you might want to receive them in a single location and view them using a GUI. In this type of scenario, SNMP notifications may be the most efficient notification method. You can avoid logging in to each switch individually, as you need to do for error log notifications.

When specific events occur on a switch, SNMP generates a message (called a “trap”) that notifies a management station using SNMP. Log entries can also trigger SNMP traps if the SNMP agent is configured. When the SNMP agent is configured to a specific error message level, error messages at that level trigger SNMP traps. An SNMP trap forwards the following information to an SNMP management station:

- Name of the element whose counter registered the event
- Class, area and index number of the threshold that the counter crossed
- Event type
- Value of the counter that exceeded the threshold
- State of the element that triggered the alarm
- Source of the trap

Email alert

An email alert sends information about the event to one or more specified email addresses. The email alert specifies the device or devices and the threshold and describes the event, much like an error message.

Flow Vision diagnostic tool

The Flow Vision diagnostic tool enables administrators to identify, monitor and analyze specific application and data flows in order to maximize performance, avoid congestion and optimize resources. Flow Vision includes Flow Monitor, Flow Mirroring and Flow Generator. MAPS can be used in conjunction with Flow Monitoring to alert the administrator when thresholds are exceeded for monitored flows.

Flow Monitor

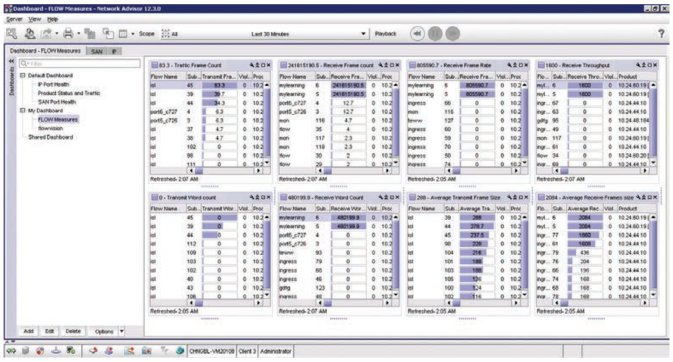
Flow Monitor enables you to monitor all the traffic passing through E_Ports, EX_Ports, F_Ports, and ISL_Ports by using hardware-supported flow parameters. Users gain comprehensive visibility into application flows within a fabric, including the ability to discover flows automatically. Define your own flows to monitor using combinations of ingress and egress ports, source and destination devices, Logical unit number (LUN) and frame types. The monitoring provided on the IBM SAN42B-R switch identifies resource contention or congestion that is impacting application performance. Flow Monitor provides support for learning or manually defining the following types of extension flows:

- Top talkers
- Flow within a fabric from a host to a target or LUN on a given port
- Flows inside Virtual Switch Logical Fabrics
- Flows passing through E_Ports (Inter-Switch Links, or ISLs)
- Flows passing through F_Ports (end device)
- Frame-based flows
- EX_Ports FC Routing (FCR) flows (routed)
  - Edge-to-edge
  - Edge-to-backbone
- Interfabric flows passing through backbone E_Ports
- Flows passing through XISL_Ports in a Virtual Fabric environment
- NPIV flows from a host to monitor VM-to-LUN performance

For specified flows, captured statistics provide insight into application performance. Monitoring various frame types at switch ports gives deeper insights into storage I/O patterns for a LUN, reservation conflicts and I/O errors. SCSI read/write frame counts and SCSI read/write data statistics are supported on F_Ports when either the source or destination device is directly connected to the switch. Integration with MAPS enables threshold-based monitoring and alerting. Statistics include the following (refer to Figure 7):

- Transmitted and received frame counts
- Transmitted and received throughput rates
- SCSI read/write frame counts
- SCSI reads and writes per second (IOPS)
- Monitored frame types:
  - SCSI Aborts–SCSI Read–SCSI Write
  - SCSI reserve
  - Rejected frames
  - Many others

Figure 7 shows the dashboard for flow measurements from Flow Monitor. Multiple flows are monitored simultaneously, and their various characteristics are displayed.



83.3 - Traffic Frame Count

Flow Name	Sub...	Transmit Fra...	Viol...	Proc
isl	45	83.3	0	10.2
isl	39	39.7	0	10.2
isl	44	34.3	0	10.2
port6_c727	4	6.3	0	10.2
port5_c726	3	6.3	0	10.2
isl	37	4.7	0	10.2
isl	38	4.7	0	10.2
isl	102	0	0	10.2
isl	98	0	0	10.2
isl	111	0	0	10.2

Refreshed- 2:07 AM

241615190.5 - Receive Frame count

Flow Name	Sub...	Receive Fra...	Viol...	Proc
mylearning	6	241615190.5	0	10.2
mylearning	5	241615190.5	0	10.2
port6_c727	4	12.7	0	10.2
port5_c726	3	12.7	0	10.2
mon	116	4.7	0	10.2
flow	35	4	0	10.2
mon	117	2.3	0	10.2
mon	118	2.3	0	10.2
flow	30	2	0	10.2
flow	29	2	0	10.2

Refreshed- 2:07 AM

805590.7 - Receive Frame Rate

Flow Name	Sub...	Receive Fra...	Viol...	Proc
mylearning	6	805590.7	0	10.2
mylearning	5	805590.7	0	10.2
ingress	66	0	0	10.2
mon	116	0	0	10.2
fewrw	127	0	0	10.2
ingress	60	0	0	10.2
ingress	59	0	0	10.2
ingress	70	0	0	10.2
ingress	50	0	0	10.2
ingress	74	0	0	10.2

Refreshed- 2:05 AM

1600 - Receive Throughput

Flo...	Sub...	Receive Thro...	Viol...	Product
myl...	6	1600	0	10.24.60.19
myl...	5	1600	0	10.24.60.19
ingr...	67	0	0	10.24.44.10
ingr...	63	0	0	10.24.44.10
gdfg	95	0	0	10.24.48.184
ingr...	49	0	0	10.24.44.10
mon	117	0	0	10.24.60.19
ingr...	61	0	0	10.24.44.10
flow	34	0	0	10.24.60.20
ingr...	69	0	0	10.24.44.10

Refreshed- 2:07 AM

0 - Transmit Word count

Flow Name	Sub...	Transmit Wor...	Viol...	Proc
isl	45	0	0	10.2
isl	39	0	0	10.2
isl	44	0	0	10.2
isl	112	0	0	10.2
isl	109	0	0	10.2
isl	103	0	0	10.2
isl	102	0	0	10.2
isl	40	0	0	10.2
isl	43	0	0	10.2
isl	106	0	0	10.2

Refreshed- 2:05 AM

480199.9 - Receive Word Count

Flow Name	Sub...	Receive Wor...	Viol...	Proc
mylearning	6	480199.9	0	10.2
mylearning	5	480199.9	0	10.2
port6_c727	4	0	0	10.2
port5_c726	3	0	0	10.2
fewrw	93	0	0	10.2
ingress	79	0	0	10.2
ingress	68	0	0	10.2
ingress	46	0	0	10.2
gdfg	123	0	0	10.2
ingress	48	0	0	10.2

Refreshed- 2:07 AM

288 - Average Transmit Frame Size

Flow Name	Sub...	Average Tra...	Viol...	Proc
isl	39	288	0	10.2
isl	44	278.7	0	10.2
isl	45	237.5	0	10.2
isl	98	228	0	10.2
isl	104	216	0	10.2
isl	101	188	0	10.2
isl	103	188	0	10.2
isl	105	136	0	10.2
isl	100	124	0	10.2
isl	102	116	0	10.2

Refreshed- 2:05 AM

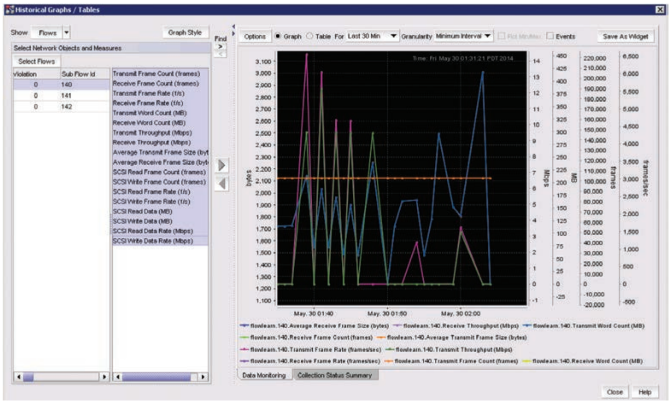
2084 - Average Receive Frames size

Flo...	Sub...	Average Rec...	Viol...	Product
myl...	6	2084	0	10.24.60.19
myl...	5	2084	0	10.24.60.19
ingr...	77	1860	0	10.24.44.10
ingr...	61	1608	0	10.24.44.10
ingr...	79	436	0	10.24.44.10
ingr...	76	204	0	10.24.44.10
ingr...	66	196	0	10.24.44.10
ingr...	74	168	0	10.24.44.10
ingr...	68	168	0	10.24.44.10
ingr...	78	168	0	10.24.44.10

Refreshed- 2:05 AM

Figure 7. Flow Monitor measurements dashboard

The flow monitor shown in Figure 8 charts various flows simultaneously. The number of violations (left side) and the respective scales (right side) per the data are displayed.



Historical Graphs / Tables

Show: Flows

Select Network Objects and Measures

Violation	Sub Flow Id
0	140
0	141
0	142

- Transmit Frame Count (frames)
- Receive Frame Count (frames)
- Transmit Frame Rate (f/s)
- Receive Frame Rate (f/s)
- Transmit Word Count (MB)
- Receive Word Count (MB)
- Transmit Throughput (Mbps)
- Receive Throughput (Mbps)
- Average Transmit Frame Size (byt
- Average Receive Frame Size (byt
- SCSI Read Frame Count (frames)
- SCSI Write Frame Count (frames)
- SCSI Read Frame Rate (f/s)
- SCSI Write Frame Rate (f/s)
- SCSI Read Data (MB)
- SCSI Write Data (MB)
- SCSI Read Data Rate (Mbps)
- SCSI Write Data Rate (Mbps)

● Graph ○ Table For: Last 30 Min Granularity: Minimum Interval □ Plot Min/Max □ Events

Figure 8. SAN, storage, disaster recovery and network administration

MAPS for Flow Monitor

MAPS monitors flows that are established within Flow Vision and generates alerts based on user-defined rules. This enables users to monitor and be alerted when established thresholds are exceeded for a particular flow. You can use MAPS for Flow Monitor to identify slow-drain devices.

Flow Generator

Flow Generator is a traffic generator for pretesting and validating infrastructure. Flow Generator is hardware-specific (8 Gbps and 16 Gbps switches and directors) and can be used across the b-type extension infrastructure. Various components and devices can be tested, including internal switch connections, cable plant, patch panels, new switches and blades, servers, storage and IP networks. Flow Generator allows users to:

- Configure a FC/FICON-capable port as a simulated device that can transmit frames at 16 Gbps line rate
- Emulate a FC/FICON fabric without having any hosts, targets or testers
- Pretest the entire SAN

Flow Generator creates a special port type called SIM ports, which are used to generate and receive simulated traffic. SIM ports behave like normal E_Ports, EX_Ports or F_Ports but are used only for testing purposes. SIM ports originate and terminate Flow Generator line-rate traffic without the need for external traffic generators or real host and target. Flow Generator generates standard or custom frames with user-specified sizes and patterns. Flow Generator supports predefined or learned flows to generate traffic between configured SIM ports. As an example use case, you can create a traffic flow between a SID and DID to validate routing and throughput across an FCIP ISL.

WAN-optimized TCP in SAN b-type Fibre Channel products

WAN-optimized TCP (WO-TCP) is a high-powered and aggressive TCP stack used for the fast and efficient transport of large data sets across enterprise WAN IP infrastructures. To this end, WO-TCP provides an exclusive streams-based mechanism. Multiple data streams (flows) can be transported autonomously without the need for separate TCP stacks for each individual stream. Because each stream is autonomous, there is no head of line blocking (HoLB) in the event that one stream becomes slow and TCP windowing performs flow control. All other flows can continue to run at rate, even when one flow is running well below rate. Additionally, when data needs to be sent without a delivery guarantee—for example, during tests and other special conditions—WO-TCP can designate streams as “nonguaranteed,” and the behavior becomes identical to user datagram protocol (UDP). This special mode of WO-TCP is essential for testing IP networks while maintaining the same TCP headers used by the b-type extension. This ensures that IP network testing sees the same extension traffic that it normally would see. WAN tool uses WO-TCP.

WAN tool

The WAN tool allows you to generate traffic at a specified rate in Kbps over a pair of IP addresses to test the network link for issues such as maximum throughput, congestion, loss percentage, out of order delivery and other network conditions. The main purpose of this tool is to determine the health of a link before deploying it for use as a circuit in an FCIP tunnel.

The WAN tool is an IP-specific tool for testing the WAN-side infrastructure on the SAN42B-R extension switch. The WAN tool creates data flows that use the same circuits configured in a tunnel or trunk. Since WAN tool uses the same circuit, all the characteristics of that circuit remain viable during testing, including jumbo frames/path maximum transmission unit (PMTU), VLAN, IPv4/IPv6, and IPsec. If a circuit in a trunk is selected with the WAN tool, the trunk’s other circuits remain online and operational while the selected circuit is decommissioned for testing.

WO-TCP is an advanced TCP stack with unique abilities. WO-TCP can logically behave just like UDP. Testing an IP network for proper SLA requires a UDP-like transport to provide constant bit rates, no reordering and no retransmits. If the IP network reorders, you will see it. If the IP network drops packets, you will see it. Traditional TCP hides these issues, making any diagnostic tool useless. The WAN tool uses the same TCP headers as the tunnel, except that all retransmit mechanisms are disabled, exposing aberrant network conditions. To test a specific protocol such as FCIP, you need that protocol’s TCP headers to traverse the network and pass through security devices that may exist. Additionally, when used with the WAN tool, WO-TCP prevents traffic windowing due to network error conditions. This means that the WAN tool always drives traffic at circuit rates despite network errors that would normally cause traditional TCP to close its windows. IBM SAN b-type Fibre Channel products provide WO-TCP technology in the world’s leading data transport TCP stack. The WAN tools allow you to:

- See throughput through the IP WAN
- See all packet drops
- See all packet reordering
- Drives test traffic at specified rate
- See round trip time (RTT)

Fabric Performance Impact monitoring

Fabric Performance Impact (FPI) monitoring uses pre-defined thresholds and alerts in conjunction with MAPS to detect and alert administrators automatically to severe levels of latency, and identifies slow drain devices that might impact the network. This feature uses advanced monitoring capabilities and intuitive MAPS dashboard reporting to indicate various latency severity levels, pinpointing exactly which devices are causing or impacted by a bottlenecked port.

ClearLink diagnostics

ClearLink diagnostics, sometimes referred to as D_Port, is a b-type Gen5 Fibre Channel exclusive feature. It ensures optical and signal integrity for Gen 5 Fibre Channel optics and cables, simplifying deployment and support of high-performance fabrics. The ClearLink diagnostic port is series of automated diagnostic tests that can be run to assess the health of links or identify issues on links exhibiting issues.

FEC and BBC recovery

IBM provides forward error correction (FEC) and buffer-to-buffer credit (BBC) recovery, which are important technologies for high-speed Fibre Channel communications. Both are used on the FC/FICON side of the SAN42B-R switch.

For more information about Fabric Vision technology, refer to: http://www.ibm.com/systems/storage/san/b-type/fabricvision

Fabric Vision extension use cases

The following use cases demonstrate various applications of the Fabric Vision extension. This is not an exhaustive list

Determine trouble location

Isolate IP network issues from storage network or storage device issues. Figure 9 shows a routed architecture. Routed architectures usually involve edge fabrics, which is why they are routed. FCR protects the edge fabrics from IP network and WAN anomalies. Depending on the edge fabric design and deployment, there is a greater possibility that HoLB may occur as a normal part of flow control. This example is a worst-case scenario showing a more complex network. This example pertains to simpler architectures as well, such as architectures that do not involve edge fabrics or FCR. The concept is to monitor flows traversing the trunk over the IP network. Users want to monitor interfabric flows through the backbone to make sure that traffic is not experiencing less than ideal conditions. MAPS monitors and alerts about WAN conditions using FCIP group policy based on comprehensive multilayer metrics. VE_Ports, tunnels, circuits and QoS are monitored for irregularities. Flow Vision monitors individual replication session data and I/O rates, plus RTT and jitter (latency variance).

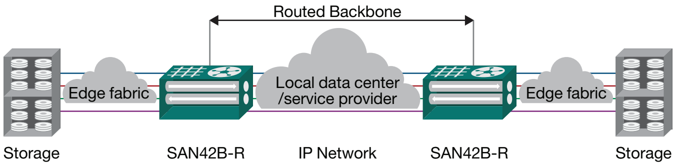


Figure 9. Routed extension architecture

Advance warning of replication under stress and effective troubleshooting is critical, given the short time and the impact to operations.

Is the IP network providing its SLA? Users receive an alert from MAPS in the event that one of the following occurs:

- Circuit retransmits: Retransmits grow above a certain rate.
- Circuit state change: A circuit goes down because the KATOV has occurred a given number of times within a period.
- Circuit round trip time (RTT): RTT has increased well beyond the typical operating value.
- Circuit jitter: Jitter has increased well beyond the typical operating value.

Any of these is an indicator of a problem in the IP network that will negatively affect the performance and impede the goal of safe RDR. The information provided by the IBM SAN42B-R switch can be used to enforce SLAs between storage administration and network administration and service providers.

Determine maximum stress

The potential drop of a replication session due to exceeded threshold indicates maximum stress during the monitoring interval. If the delta set cycle time is five or ten seconds (Figure 10 shows the cycle time set to 30 seconds), the expectation is to complete the transmission of that delta set across the extension network within that period of time. If that is not happening reliably, there are a few paths to resolution. First, is there enough bandwidth for the amount of data that needs to be sent? Use Flow Vision to monitor the replication flows and determine if they are fully using the bandwidth that is available. If they are, then there is not enough bandwidth. If they are not, the next step is to determine if the IP network is providing its SLA (refer to the previous case). If the bandwidth is not being fully consumed, and the IP network is clean, investigate the storage array for possible issues.

In Figure 10, the delta set cycle time is set to 30 seconds. Cycle times will not be less than 30 seconds. The expectation is that complete data transfer of the delta set will take place within the 30-second interval. As shown, this is not happening regularly. In many instances the amount of time to transfer the data set exceeds 30 seconds and may take as long as 43 seconds. A determination must be made if this is due to a network problem or if there is just too much data to transmit for the bandwidth and compression available.

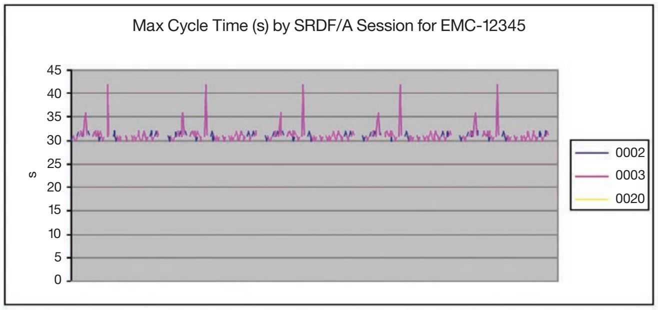


Figure 10. Data delta set cycle time to completion

Does RDR appear to be stressed, while the IP network administrators claim it should not be? Use these tools to gain visibility into your RDR environment:

- Use Flow Vision to visualize the RDR flows and determine if they are consuming all the available bandwidth.
- Use Flow Vision to monitor individual replication session data and I/O rates.
- Use Flow Vision to monitor RTT and jitter. If this is excessive, it can cause droop across the WAN connection. Protocol optimization such as FastWrite may be required, if applicable.
- Use MAPS to alert on events that indicate degraded IP network states, which elongate delta set transmission times and hinder recovery from delta set extension or backlog journaling when the network is running behind.

Configuration validation and IP network assessment

Setting up an RDR network is a daunting task. The technology is sophisticated and uses a significant number of moving parts. At fruition, when the user is at the validation point in the setup process, it can be difficult without the proper tools to make sure everything is functioning properly.

Once everything is installed and configured, it is time for validation testing. You can do end-to-end testing of the storage array, the storage network, the extension platform and the IP network.

The SAN42B-R switch provides tools to simplify this process considerably. These tools include ping and traceroute at a rudimentary level. The PMTU discovery tool can be used to verify the IP network MTU or to determine the MTU in cases where it is not already known. For a comprehensive test of the IP network, run the WAN tool for any desired period of time. For example, run the WAN tool for the duration of a typical work day. Generate line-rate extension TCP flows and gather IP network statistics for analysis and SLA evaluation. Upon completion, the WAN tool will provide information about the IP network, including maximum throughput, congestion, loss percentage, out-of-order datagrams, latency and other network conditions.

Summary

Storage administrators are facing challenges with infrastructure they do not manage or control, specifically the IP network, across which many of their applications pass. Because storage administrators often have little familiarity with the IP network, this can prolong time to resolution when there are operational problems. This is exacerbated when support and multiple vendors are involved. The process requires a fast and efficient way for storage administrators to pinpoint where a problem may be occurring.

But how can you more quickly pinpoint problems? Storage administrators may not have expertise in managing extension networks since storage arrays have no visibility into extension networks. The IBM System Storage SAN42B-R extension switch provides unique tools for quickly resolving hard-to-diagnose issues. Fabric Vision offers MAPS, Flow Vision and the WAN tool, which are all enterprise-class tools. MAPS has the ability to detect either sudden changes in the network or errors that present themselves slowly over time. Monitoring is easy to set up and customize with pre-defined policies and rules that incorporate thresholds determined through over more than a decade of experience. When a monitored network characteristic crosses marginal or critical thresholds, various alerts and actions are available. Beyond the automated monitoring, alerts and actions provided by MAPS, another valuable tool for visualizing flows is available, called Flow Vision.

Flow Vision offers the Flow Monitor tool. Flow Monitor can discover flows automatically or can be configured manually. Upon monitoring a particular flow, MAPS is integrated into Flow Monitor so that thresholds of interest can be set with corresponding alerts and actions. Another tool that Flow Vision provides is called Flow Generator. Flow Generator can generate test flows originating at the FC ASIC level and traverse b-type extension. The characteristics of the flows are either preconfigured or learned from the application.

WO-TCP is an advanced TCP stack with the ability to treat each individual flow autonomously and eliminate HoLB. WO-TCP also provides a special functionality for generating UDP-like traffic with extension headers. This permits the WAN tool to perform accurate testing of network conditions while the IP network sees actual extension traffic flows.
A couple of cases were discussed in this brief: One was how to determine where trouble may be originating. Another was how to determine stress on array replication applications. Overall, the IBM SAN42B-R extension switch comes with sophisticated tools that enable you to distinguish trouble with the network from storage array application problems. These effective tools facilitate more efficient support calls and faster problem resolution.

For more information

To learn more about the benefits of Fabric Vision technology for disaster recovery, please contact your IBM representative or IBM Business Partner, or visit the following website:
ibm.com/systems/storage/san/b-type/fabricvision

Additionally, IBM Global Financing can help you acquire the IT solutions that your business needs in the most cost-effective and strategic way possible. For credit-qualified clients we can customize an IT financing solution to suit your business requirements, enable effective cash management, and improve your total cost of ownership. IBM Global Financing is your smartest choice to fund critical IT investments and propel your business forward. For more information, visit:
ibm.com/financing

Route 100
Somers, NY 10589

Produced in the United States of America
November 2015

Content used by permission of Brocade Communications, Inc.

This document is current as of the initial date of publication and may be changed by IBM at any time. Not all offerings are available in every country in which IBM operates.

It is the user’s responsibility to evaluate and verify the operation of any other products or programs with IBM products and programs.

THE INFORMATION IN THIS DOCUMENT IS PROVIDED “AS IS” WITHOUT ANY WARRANTY, EXPRESS OR IMPLIED, INCLUDING WITHOUT ANY WARRANTIES OF MERCHANTABILITY, FITNESS FOR A PARTICULAR PURPOSE AND ANY WARRANTY OR CONDITION OF NON-INFRINGEMENT. IBM products are warranted according to the terms and conditions of the agreements under which they are provided.