SDC 20

BY Developers FOR Developers

Storage Developer Conference
September 22-23, 2020

Smart Fabrics: Building Self-healing Fibre Channel Networks

Brandon Hoff, Principal Architect, Broadcom

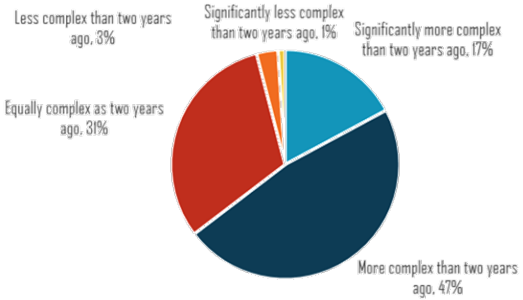
Rupin Mohan
Director R&D, CTO (SAN)
Hewlett Packard Enterprise

Agenda

- IT Trends in the Enterprise
- Fabric Notifications
- Congestion Management - *in action*
- Fabric Assisted MPIO
- OS Support
- Summary

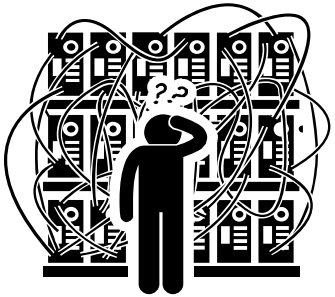
IT Trends Driving the Need for Autonomous SANs

Complexity of IT Environment



64% of organizations view their IT environments as more complex than they were two years ago [1]

Efficiency & Performance



58% of organizations plan to improve efficiency & performance by investing in monitoring and automation infrastructure management tools [2]

Resource Constraints

62% of new IT positions are being filled by generalists [1]

Source:
1. ESG Master Survey Results, *2019 Data Storage Trends*, November 2019.
2- Computer Weekly and TechTarget, 2020 IT Priorities Survey , March 2020.

SAN Automation – Enabling Deployment & Managing SAN's Simple

Cloud Orchestration Platforms

Industry Tools

Fibre Channel SAN Automation

Python Libraries

REST API

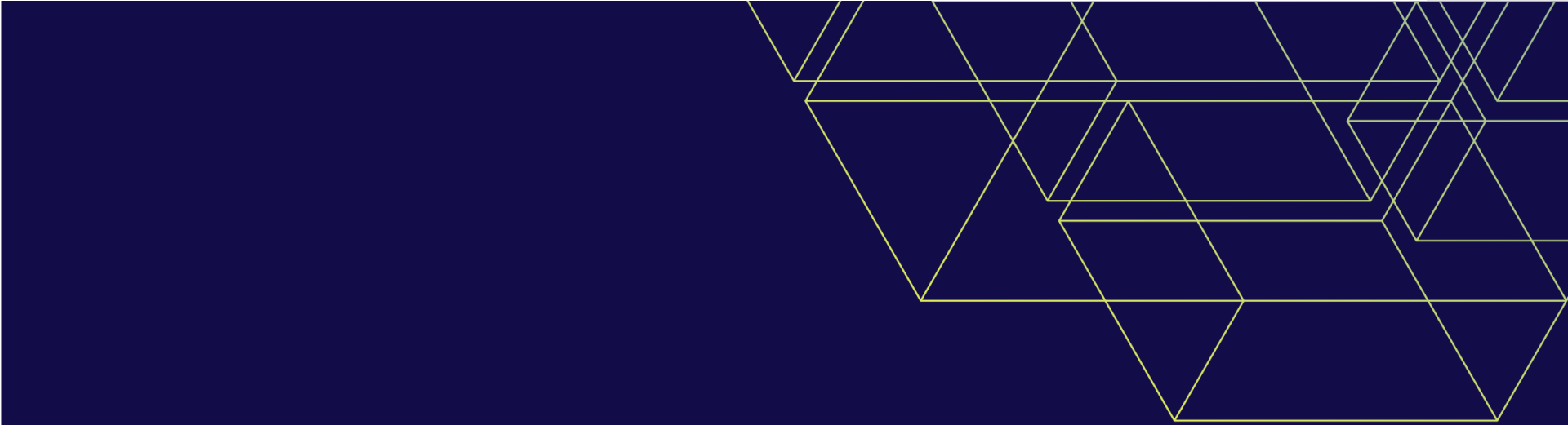
Fabric Notifications, Peer Zoning, Target Driven Peer Zoning

Fabric Notifications

Fabric Notification Summary

FPIN Type Fabric Performance Impact Notification	What it does
Congestion Notification (FPIN-CN)	Notifies the port that is causing congestion that it is causing congestion.
Link Integrity (FPIN-LI)	Notifies the server port that the link it is connected to is 'sick but not dead.'
Peer Notification (FPIN-PN)	When a FPIN-CN is sent to the server causing congestion, a Peer Notification is sent to all of the other ports in the zone.
Delivery Notification (FPIN-DN)	Delivery Notification - Best example is if a switch drops a frame due to egress hold time expiring (220ms).
Signals	Instead of using ELS, in Gen 7 ASICs providing congestion notifications at the hardware level. In essence, congestion signals look like rapid FPIN-CNs.

The details are in FC-FS-6 (Congestion Signal) and FC-LS-5 (EDC, RDF, and FPIN ELSs).

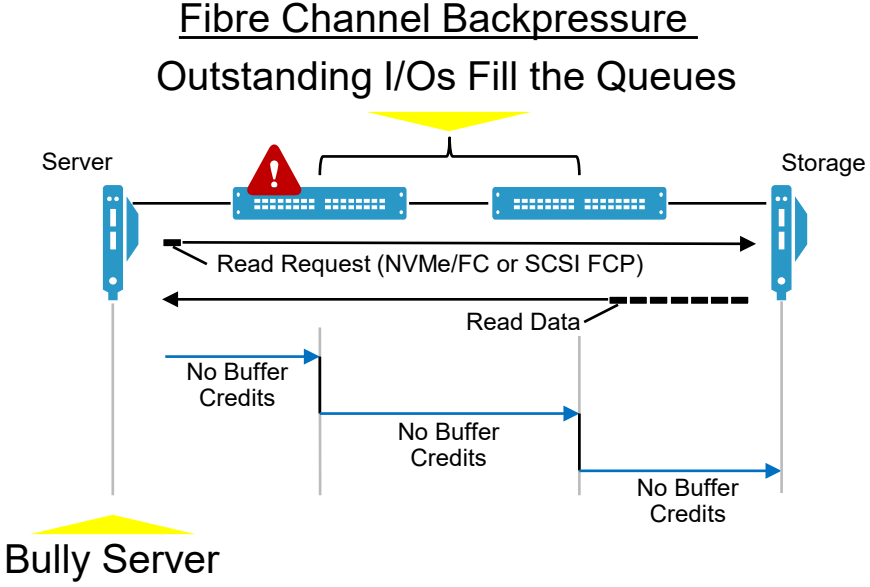


Congestion Management (FPIN-CN)

Congestion Spreading

Workload grows until it is too big for the hardware footprint

- Server asks for too much data
- Storage Array sends data to the server
- Server can't ingest the data fast enough
  - Out of CPU cycles
  - Out of memory
  - Out of PCIe bandwidth
  - Out of HBA bandwidth
- Outstanding I/Os from the Bully Server consume the hardware resources of the SAN
- The Bully Server causes a significant performance impact on the other servers connected to the SAN
- Alerts are created, but difficult to mitigate



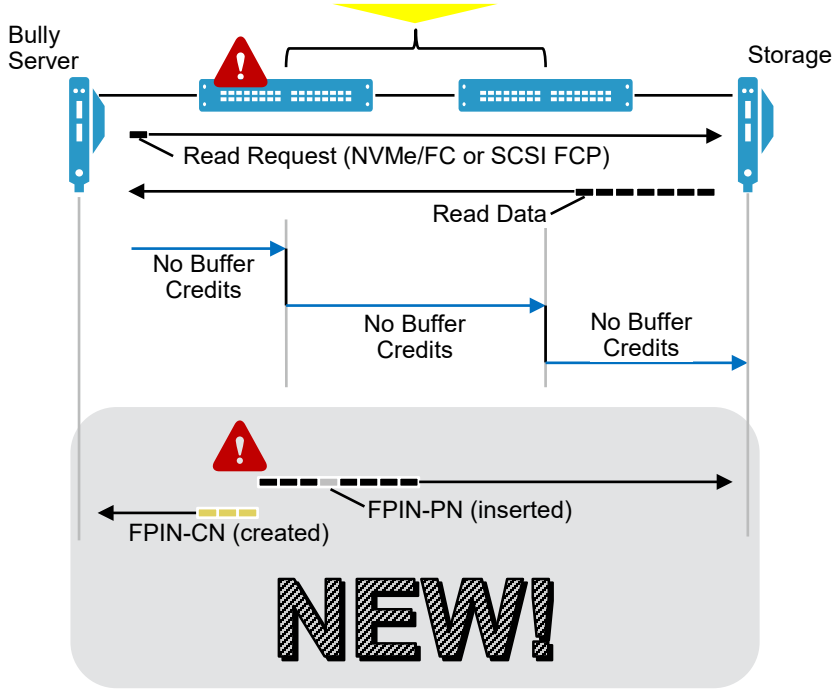
Congestion Notifications

Communicates Fabric detected congestion to the end-points

- Switch detects congestion
- Switch sends a FPIN-CN (Congestion Notification) to the bully server
- Switch sends a FPIN-PN (Peer Notification) to other end-ports in the Fabric
- Congestion mitigation is handled by the end-point
  - Automatic congestion remediation is now possible
- **Goal: Optimize hardware utilization for the SAN for all workloads**

New! Fabric Notifications for Congestion

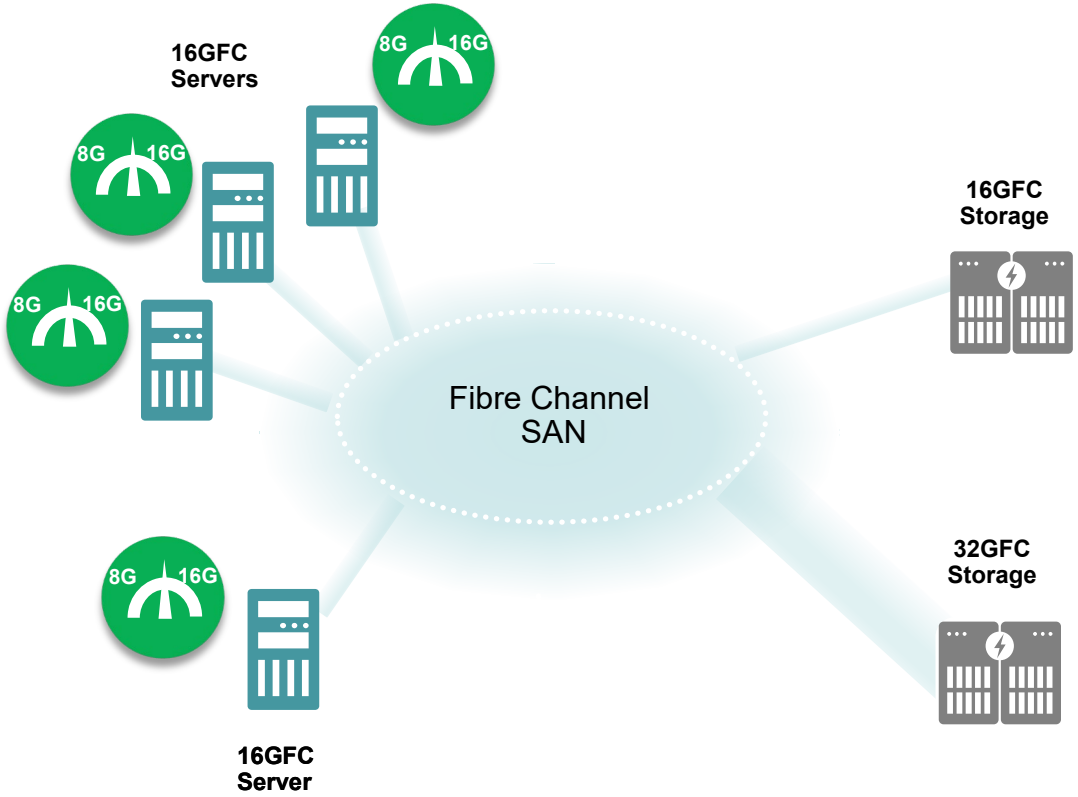
Optimize Hardware Utilization



Fibre Channel (FPIN-CN)

Normal Operations

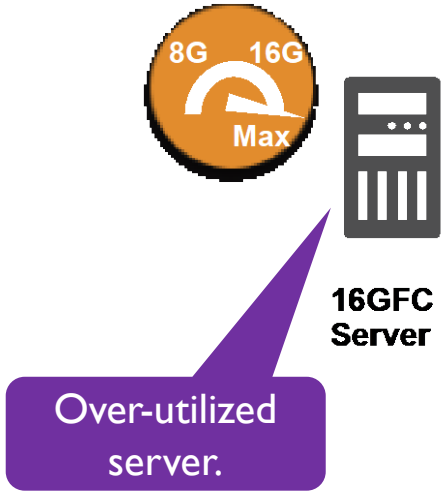
Under normal operations, the SAN operates very efficiently. No frame drops, plenty of hardware resources to service all I/O requests.



The Problem: Overutilized Server

Servers and workloads trend to

- An over-utilized server is running at
  - 100% CPU utilization and/or,
  - 100% PCIe bus utilization and/or,
  - 100% Memory utilization and/or,
  - 100% HBA bandwidth utilization
- Over time, all servers can move into this state
  - Someone moved too many VMs onto a server
  - The workload on the server outgrew the hardware

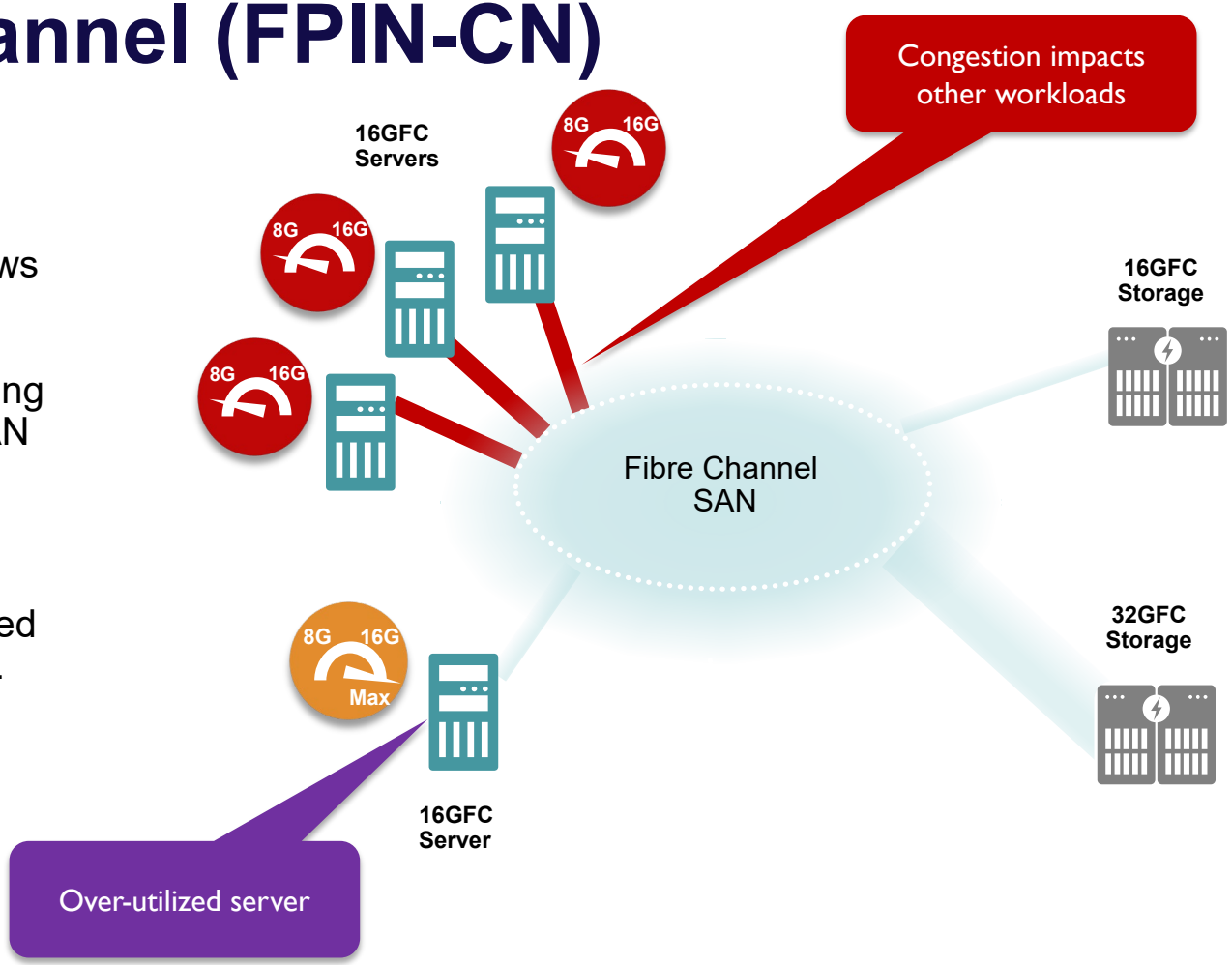


Fibre Channel (FPIN-CN)

Oversubscription

The workload grows larger than the hardware can handle. Outstanding I/Os consume SAN resources.

Performance is negatively impacted across the SAN.

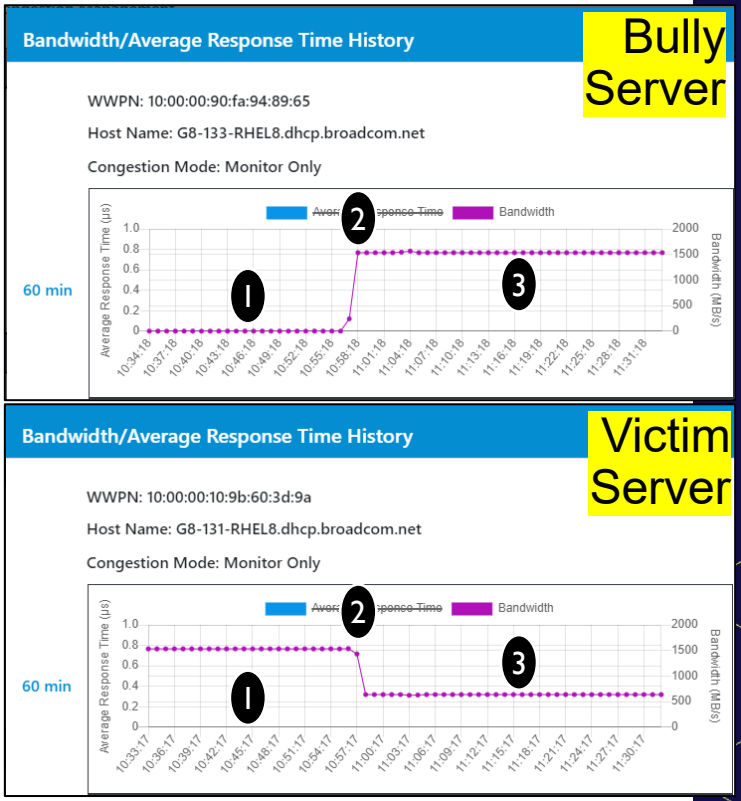


Real World Results of Congestion

Bully Server causes a 50% performance impact on Victim Servers

1. The Victim Server is running great
2. The Bully Server is turned on
   - Runs at full line rate
   - Looks like it is running great, but it isn't
3. The Victim Server's performance is cut in half
   - Users start filing trouble tickets as their applications run slow
   - To the administrator, it isn't clear where the root cause is
   - Once the administrator finds the root cause, it is difficult to remediate

➡ New! Fabric Notifications

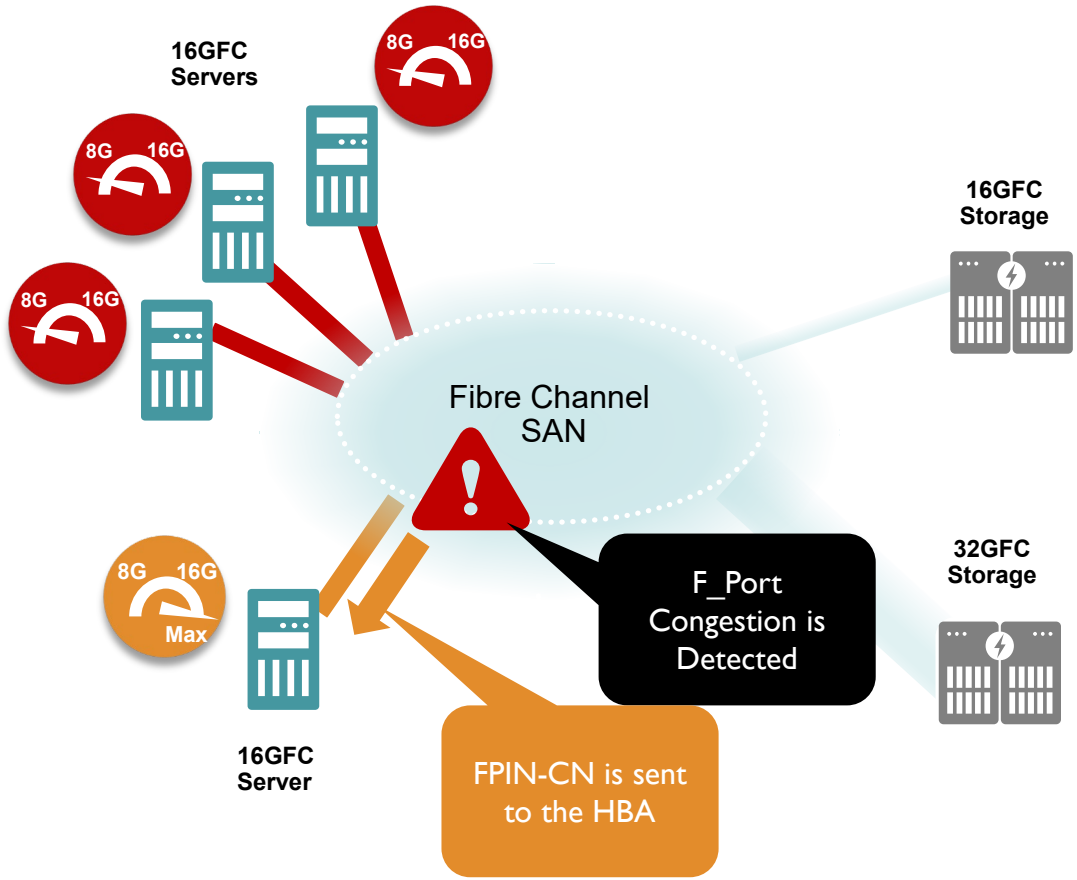


Source: Broadcom Customer Demo Lab

Congestion-free SAN (FPIN-CN)

How collaboration works

The Fabric detects F_Port congestion & sends FPIN to the HBA

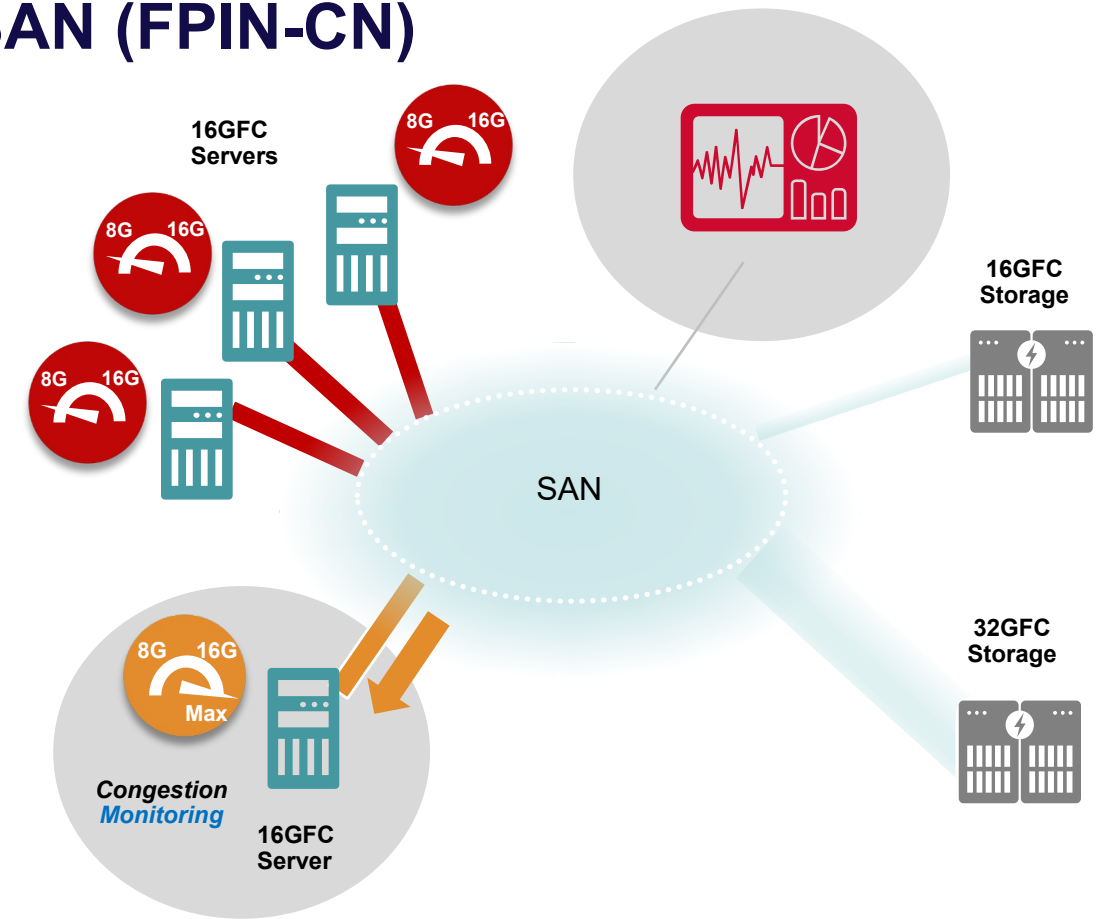


Congestion-free SAN (FPIN-CN)

How collaboration works

The administrator observes congestion data in their SAN management tool.

The administrator turns on congestion management.

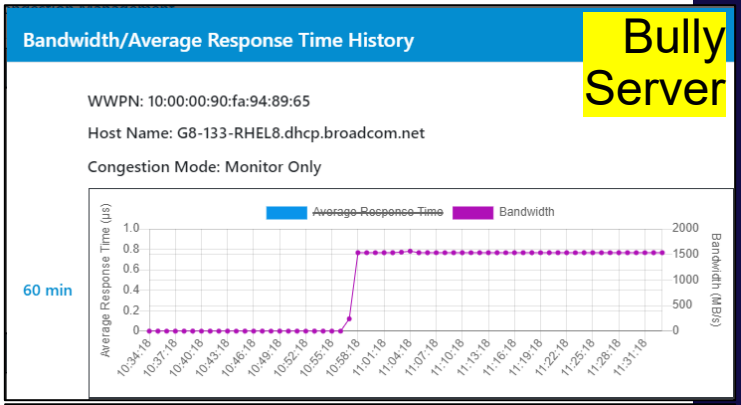


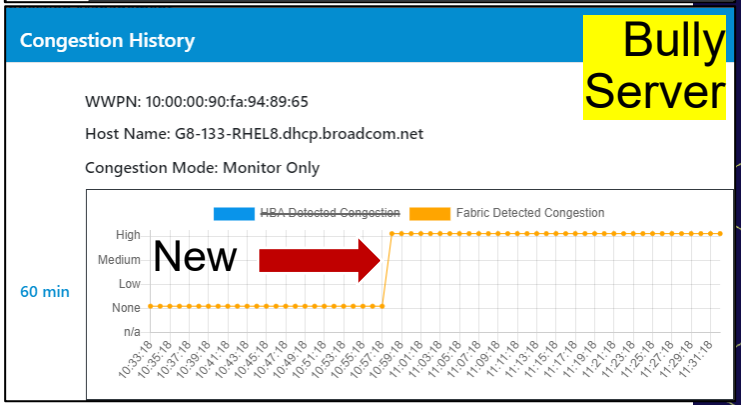
Fabric Notifications

FPIN-CNs tells the host that it's asking for too much data

- Basically, "The Bully's eyes are bigger than his stomach."
- The best place to detect congestion is in the Fabric
- The best place to mitigate congestion is at the end-point
- Now the Fabric sends notifications to the server with FPIN-CNs

Next, mitigation!





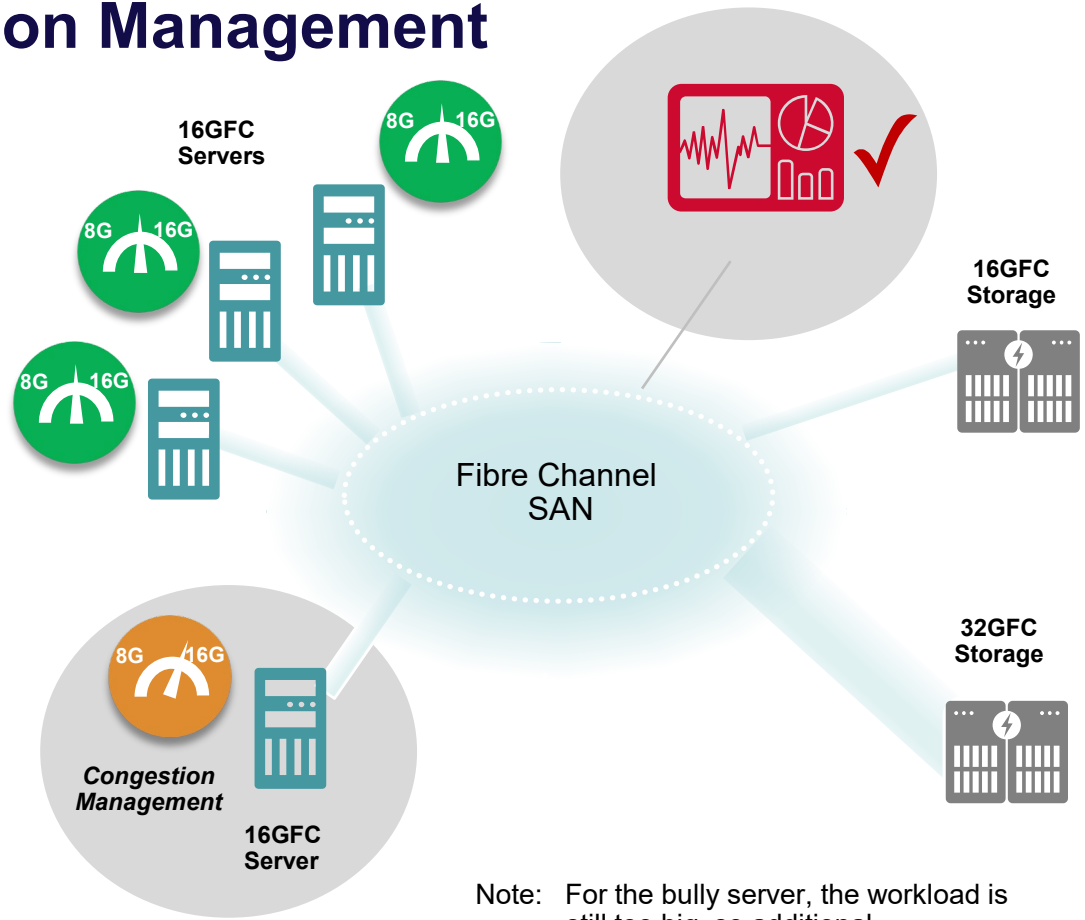
Source: Broadcom Customer Demo Lab

Adaptive Congestion Management

Congestion-free SAN

The administrator turns on congestion management on the Bully Server.

Congestion is mitigated.



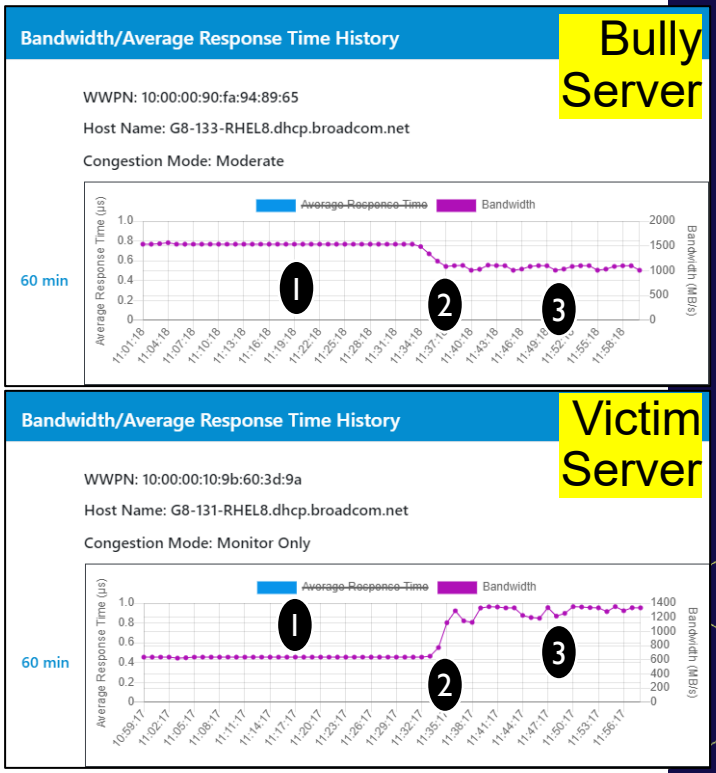
Note: For the bully server, the workload is still too big, so additional administrative attention is required.

Real World Results

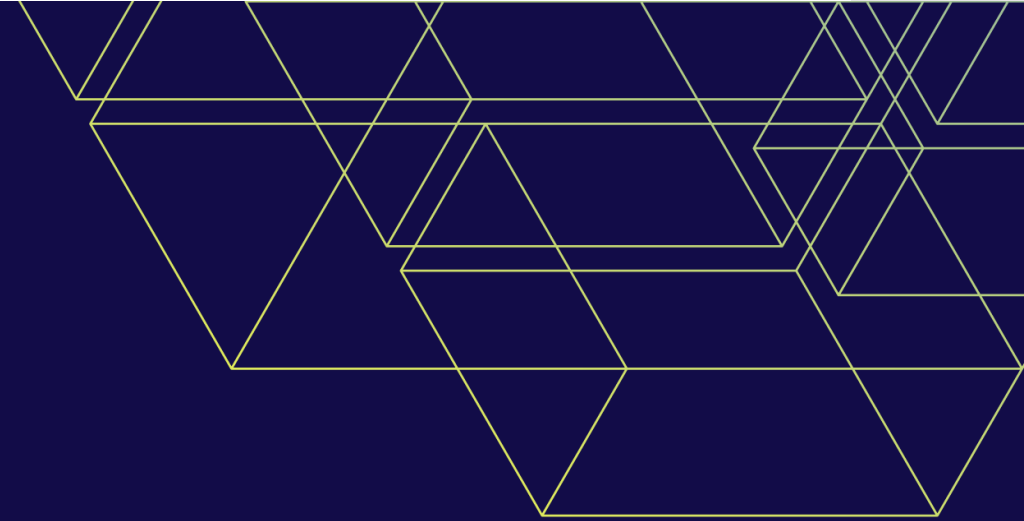
Optimize hardware utilization for all workloads connected to it.

1. The network is running in a congested state
2. Adaptive congestion management is enabled
   - The Bully server is paced
3. The Victim Server's performance is restored

Bonus: Latency of the Bully server is reduced by 2 orders of magnitude.



Source: Broadcom Customer Demo Lab



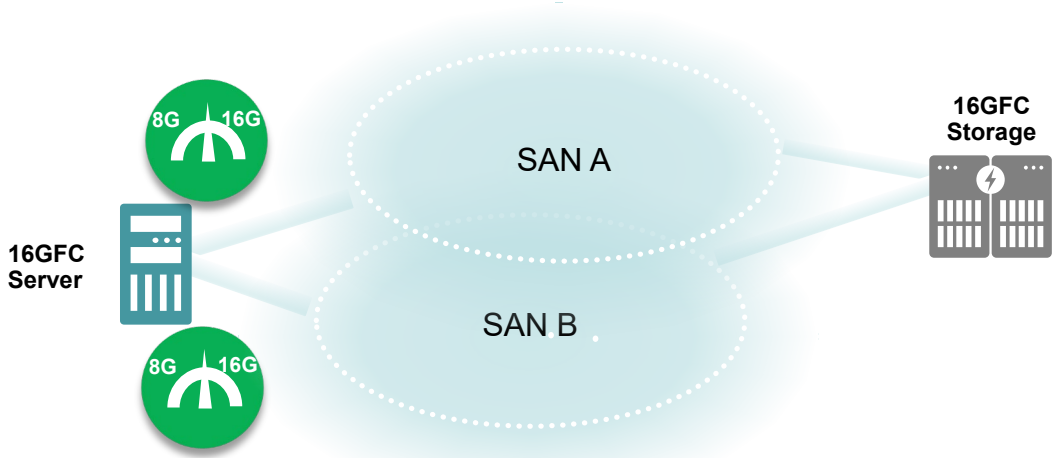
Fabric Assisted MPIO (FPIN-LI)

Fibre Channel (FPIN-LI)

Normal Operations

Under normal operations, with the Fibre Channel links working, data is balanced between the different paths to the storage system.

Over time, links may not work as well as they should.

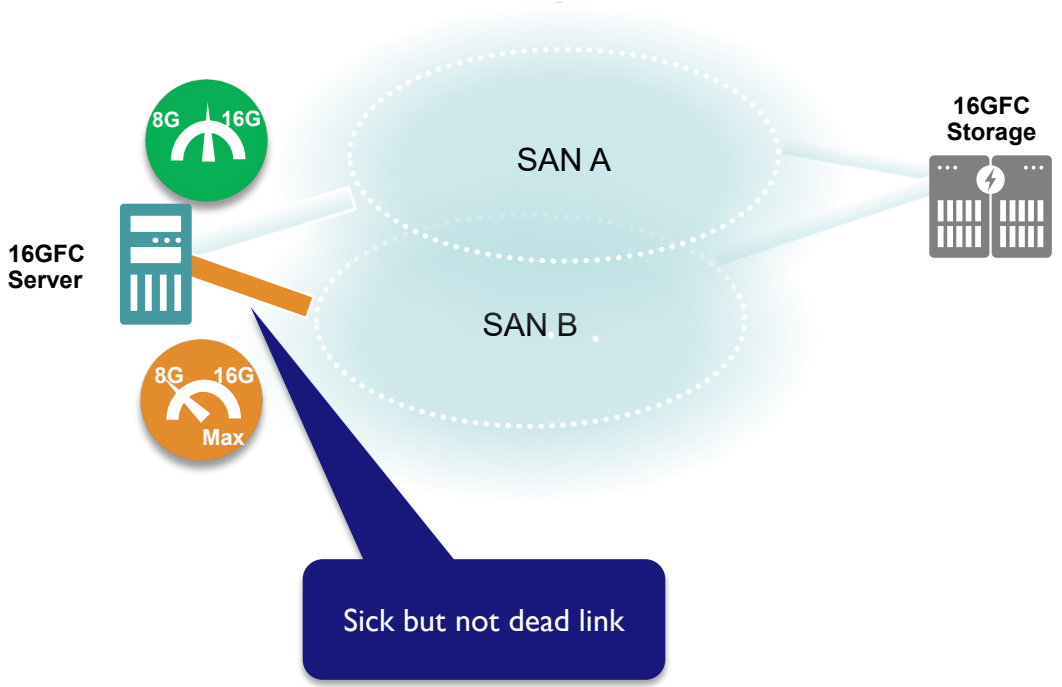


Fibre Channel (FPIN-LI)

The impact of bad links

When links are sick, but not dead, IOs going over the bad link reduce the data available to the application.

Quick detection and action is essential.

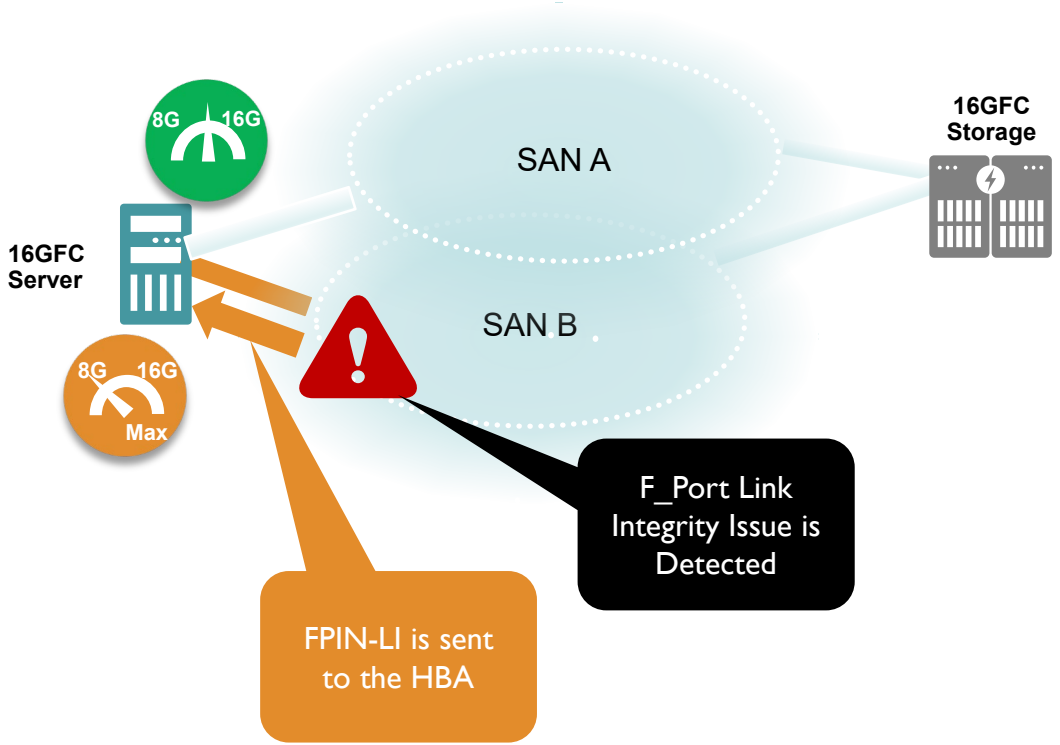


Fibre Channel (FPIN-LI)

Fabric Assisted Multipathing

FPIN-LI is sent from the Fabric to the HBA.

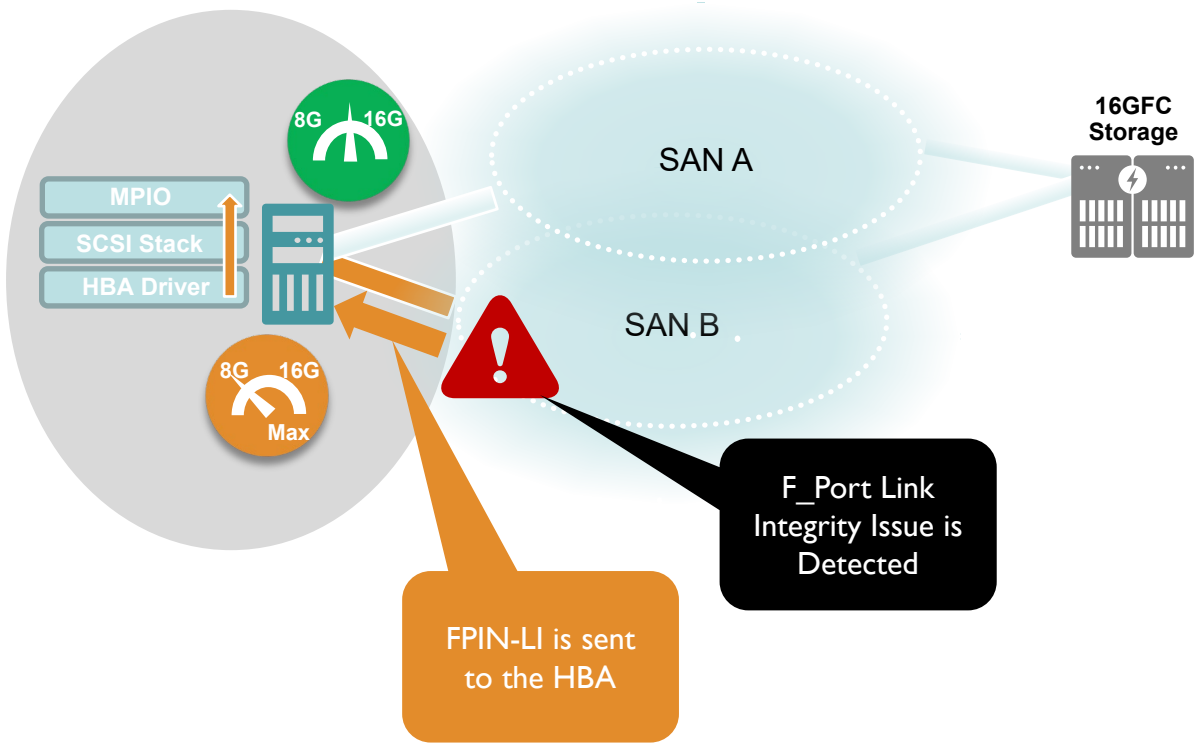
FPIN-LI is then received and processed by the HBA



Fibre Channel (FPIN-LI)

Fabric Detected Link Integrity Issues

The HBA passes the FPIN up the stack to the multipathing software.

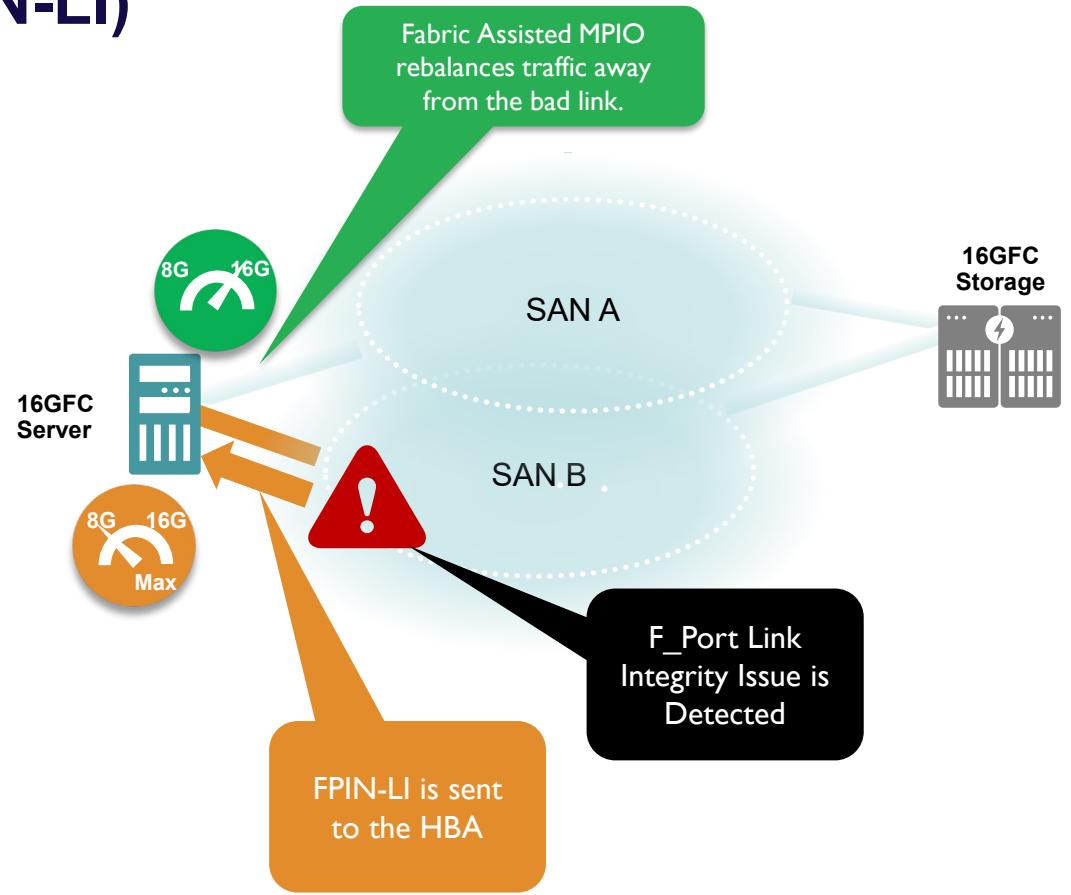


Fibre Channel (FPIN-LI)

Normal Operations

Benefit: Identify and address issues much quicker. The multipathing software makes more intelligent decisions based on data from the Fabric.

Performance problems are minimized.



OS Details and Vendor Support

Ecosystem Support Details

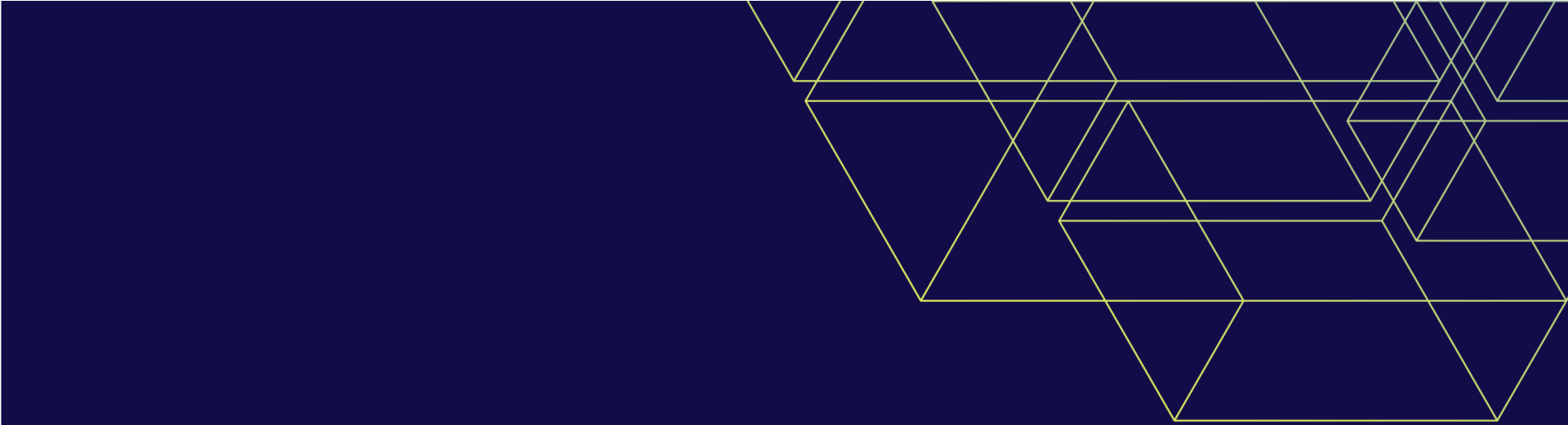
- Fabric Notifications are developed by T11
- Switch support available today
- HBA support available today
  - Windows, VMware, RedHat, SUSE
  - All server OEMs

Contact your vendor for more details!

Fibre Channel
The Autonomous SAN

- Fabric Notifications is a new platform to deliver better networking solutions
- Adaptive solutions are available today with this technology
- Get involved: Opportunity to build new solutions on the Fibre Channel platform

Questions?



Please take a moment
to rate this session.

Your feedback matters to us.