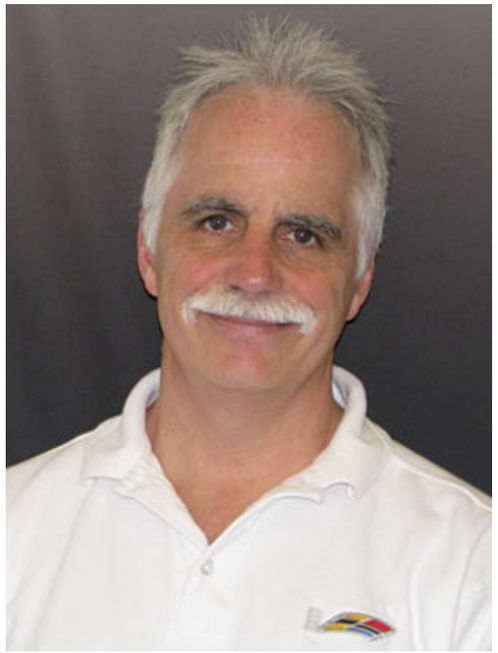
Fibre Channel Performance: Congestion, Slow Drain, and Over Utilization, Oh My!

Live Webcast
February 6, 2018
10:00 am PT

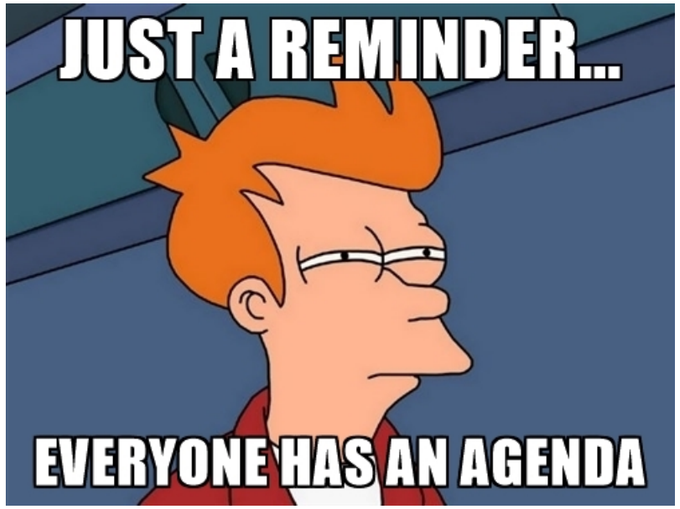
Today's Presenters



Ed Mazurek
Cisco

Earl Apellanes
Broadcom Limited

David J Rodgers
Teledyne LeCroy



- How Fibre Channel Achieves Lossless Data Delivery
- Congestion: Causes, Indications, and Meanings
- How to prevent, Identify, and resolve performance problems

How Fibre Channel Achieves Lossless Data Delivery

Fibre Channel – Lossless data delivery

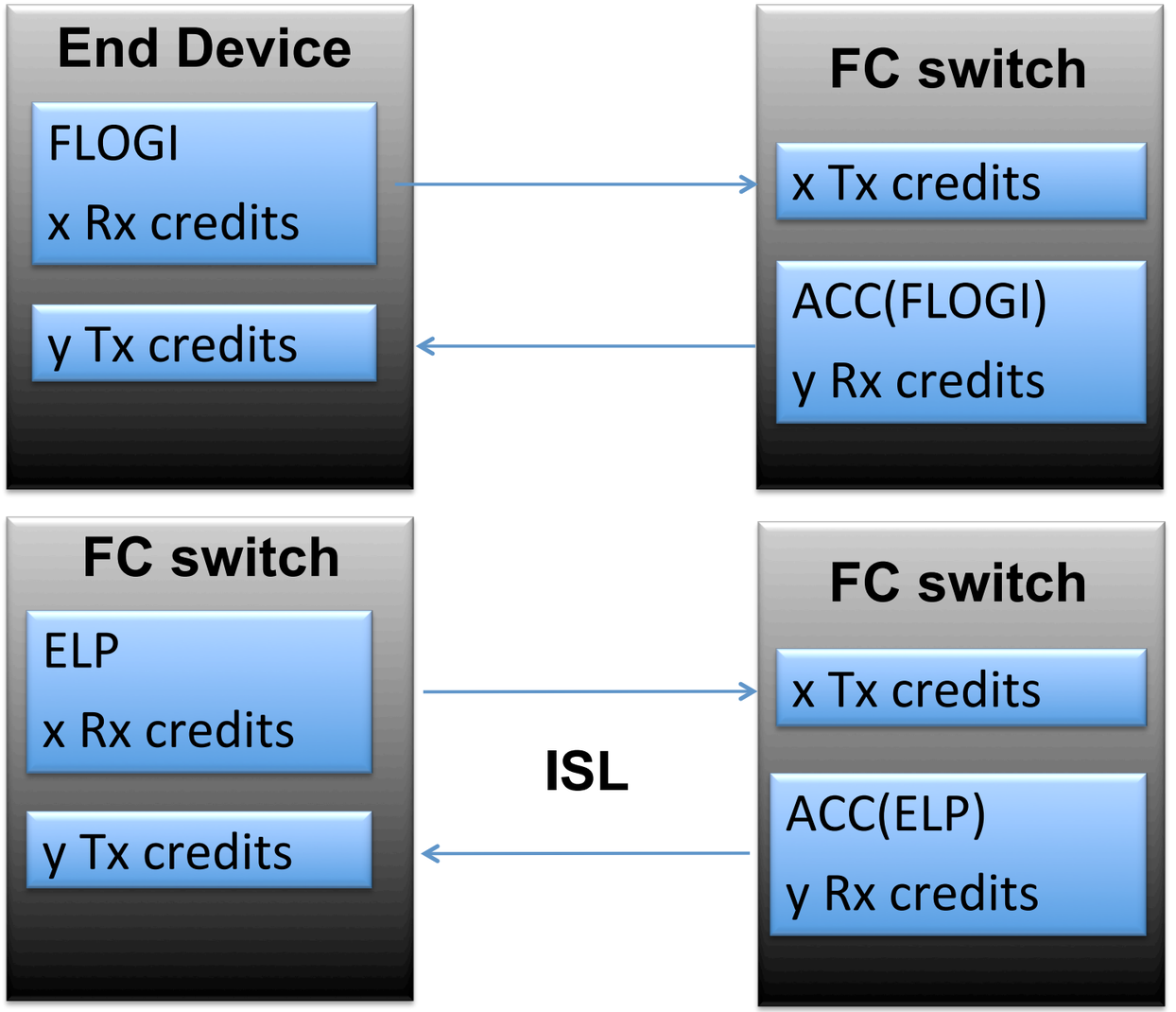
FC works on a credit mechanism

- Frames are only transmitted when it is known that the receiver has buffer space
- For each frame sent, an R_Rdy (B2B Credit) should be returned
- R_Rdys can only be returned once the frame that has previously occupied that buffer location has been handled
- Each side informs the other side of the number of buffer credits it has
  - F ports - In the Fabric Login(FLOGI)
  - E ports – In the Exchange Link Parameters(ELP)

Fibre Channel – Lossless data delivery

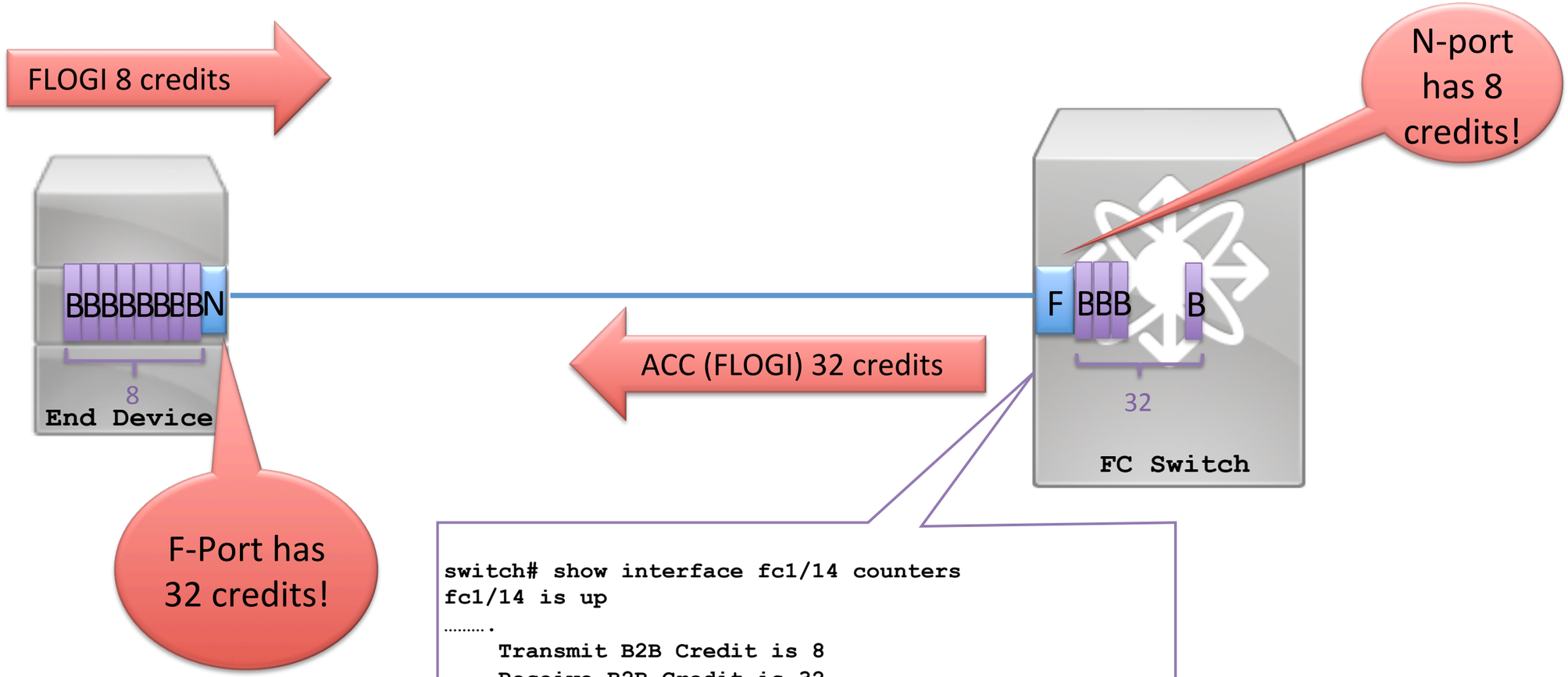
B2B credits / Credits remaining

- Buffer to Buffer credits or B2B credits are the agreed upon buffer space on each side of a FC link
  - Occurs on FLOGI and ACC(FLOGI)
  - Occurs on ELP and ACC(ELP)
- B2B credit remaining is the count of FC frames that still can be sent by each side of a FC link
- Credits are returned by R_Rdy FC ordered set



Fibre Channel – Lossless data delivery

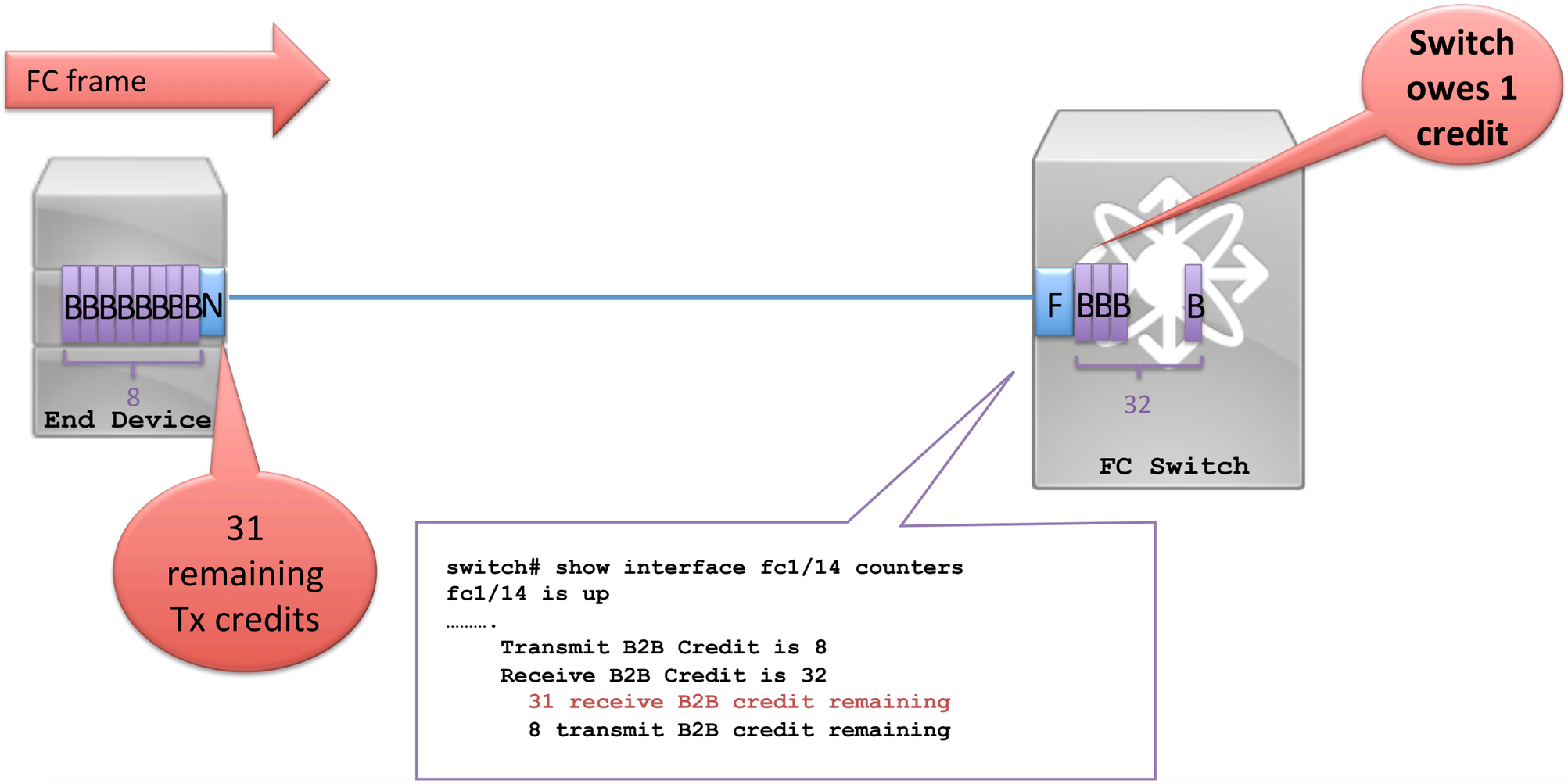
N-Port Login



Fibre Channel – Lossless data delivery

Frame flow

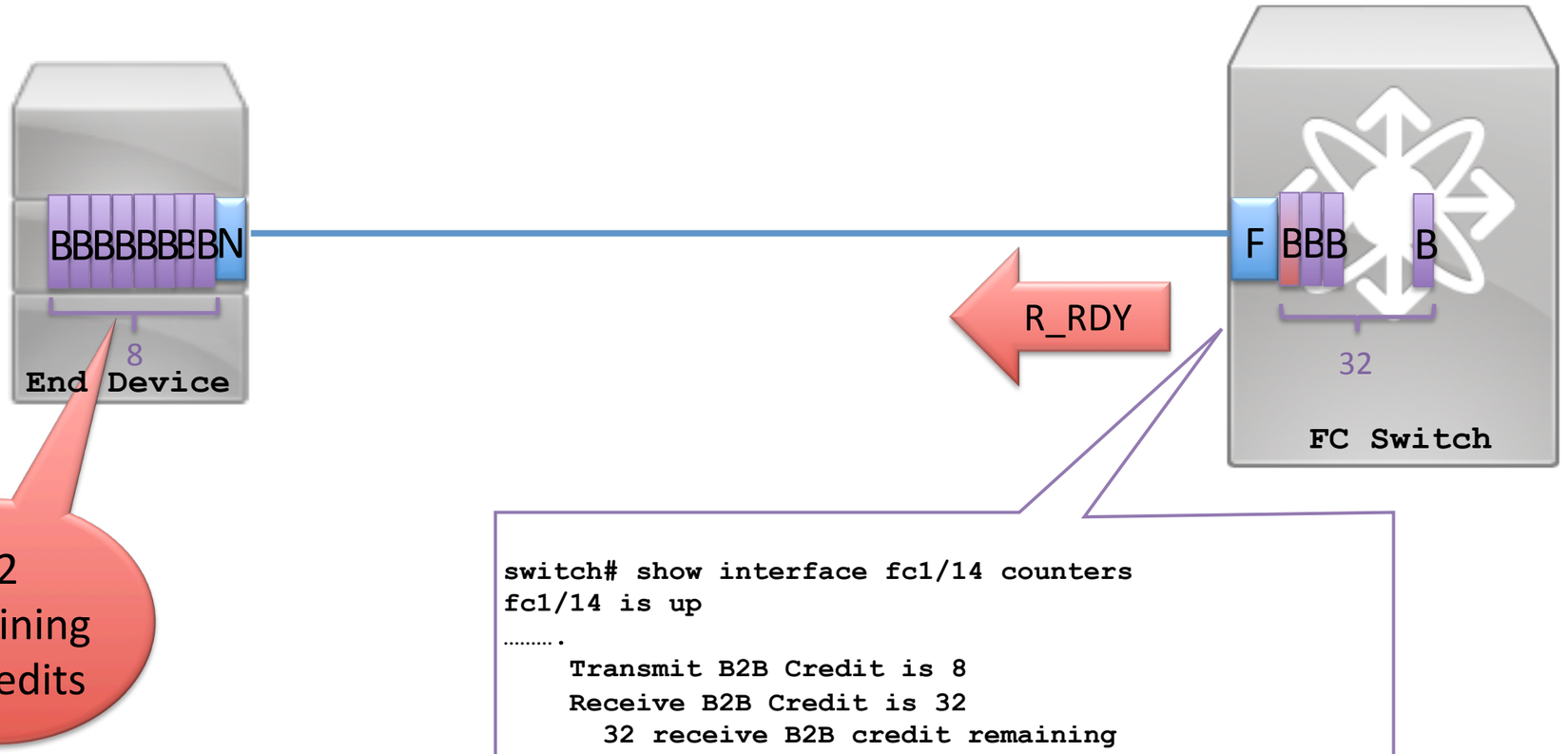
- End device checks remaining Tx credits
- Since it is > 0 decrement and send frame



Fibre Channel – Lossless data delivery

Frame flow

- End device checks remaining Tx credits
- Since it is > 0 decrement and send frame

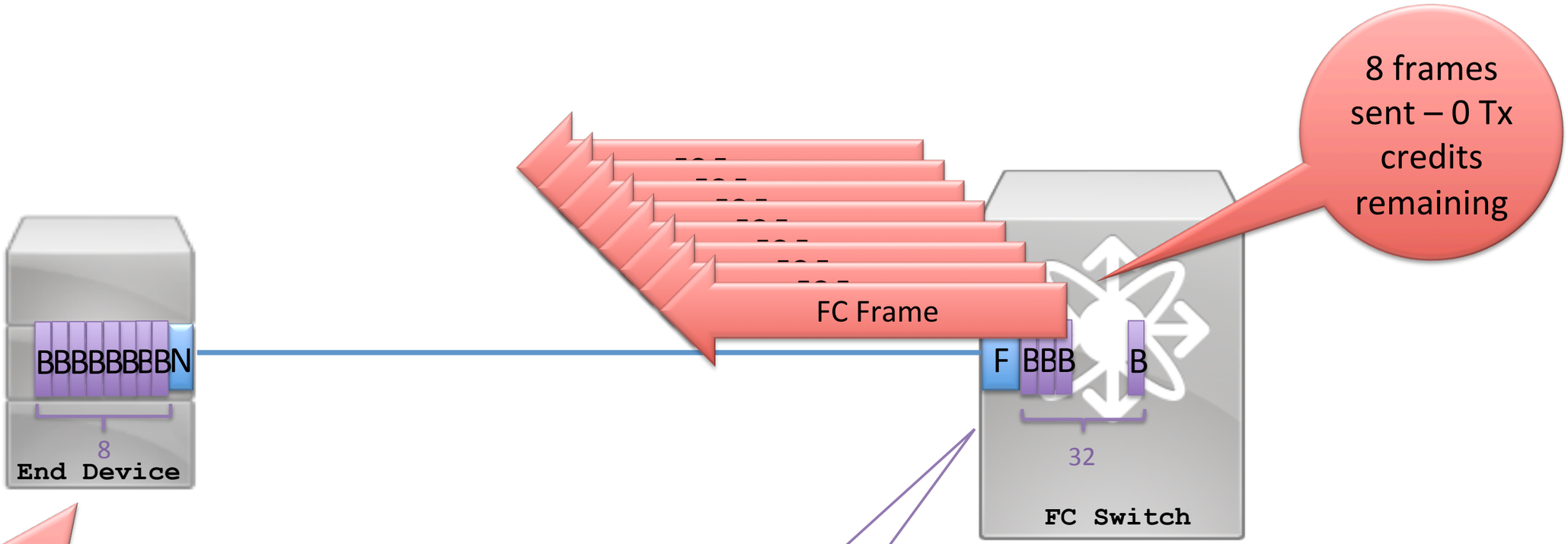


- Switch receives and processes frame
- Once processed R_RDY is sent

Fibre Channel – Lossless data delivery

Frame flow

- End device stops returning R_RDYs



Since the end device is not returning R_RDYs quickly it is a "slow drain" device

```
switch# show interface fc1/14 counters
fc1/14 is up
……..
     Transmit B2B Credit is 8
     Receive B2B Credit is 32
       32 receive B2B credit remaining
       0 transmit B2B credit remaining
```

Congestion: Causes, Indications, and Meanings

Three Main Causes of Congestion

- Credit Stalled Device/Slow-Drain
  - Abnormal or unexpected device behavior
  - Device induced credit latency
    - I.e. “Slow Drain,” “Slow Draining” device, or “Slow Drainer”
    - Sender’s Tx credits hit 0 forcing a stop to any transmissions until a credit is received
- Lost Credit
  - Physical errors
  - Credits and frames are not sent reliably resulting in a loss of credits over time
- Oversubscription/Overutilization
  - Bandwidth mismatch
    - Speed mismatch (e.g. 16G to 4G)
    - Fan-in mismatch (e.g. 8-ports to 1-port)
    - Device simply requesting more data than it can consume at its given link rate

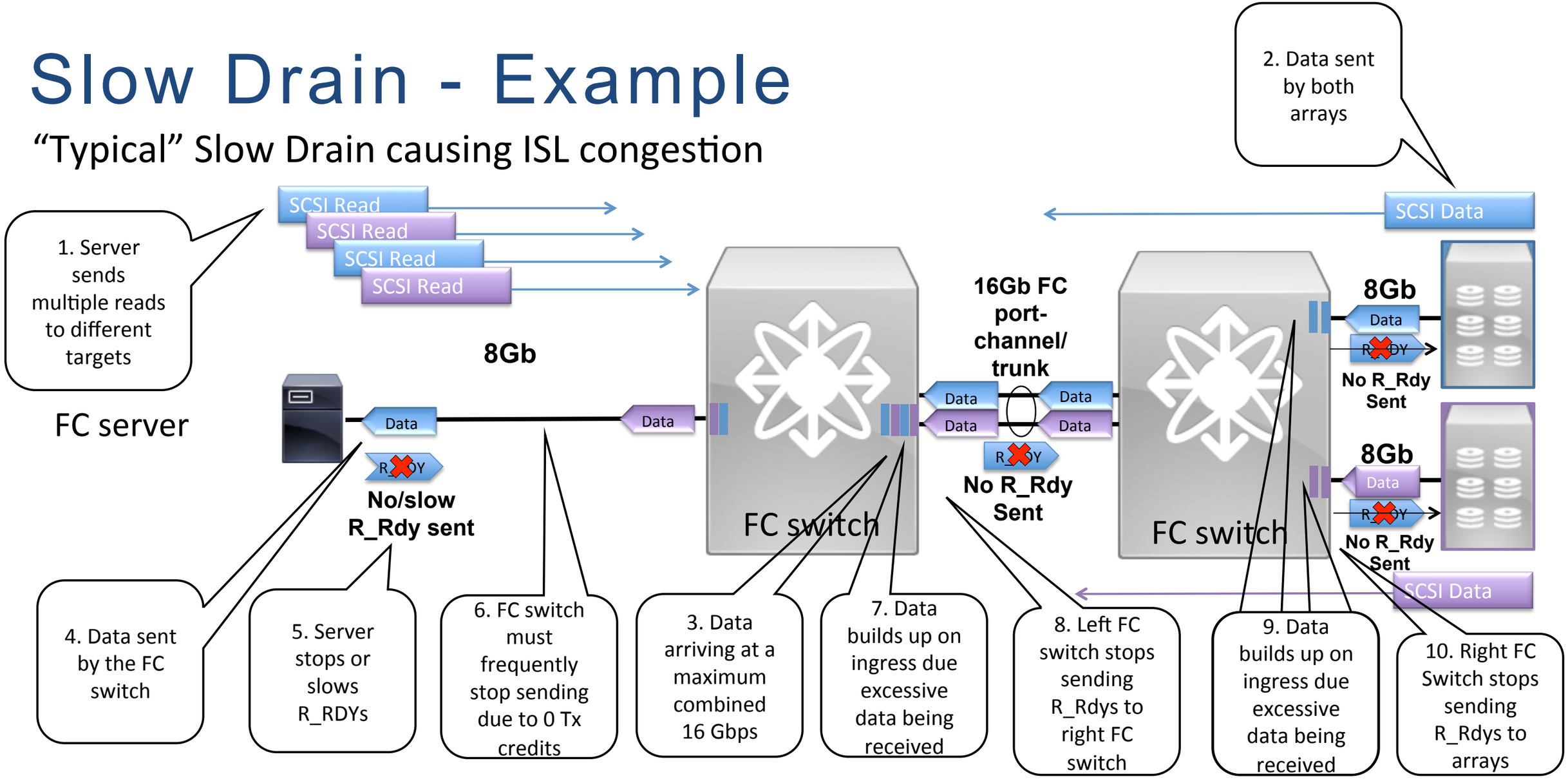
Credit Stalled/Slow Drain

- A “slow drain” device is a device that delays returning B2B credits (R_RDYs)
- FC switch quickly reaches 0 Tx credits remaining and is unable to send
- This causes data to build up in the FC switch
- FC switch then must withhold R_RDYs from adjacent end device or FC switch.
- This causes the congestion to work its way back toward the source of the frames

Mild/Some Latency	Moderate	Severe
Fabric Tools Provide Alerts eg: time at zero credit.	Poor Application Performance/Some SCSI Errors	Application Failure/ Link Failures

Slow Drain - Example

"Typical" Slow Drain causing ISL congestion



Both arrays and all devices utilizing ISLs are affected!

Lost Credit

- Typical fabric symptoms are credit latency, queue latency, frame loss, or link reset
- As the problem persists, frame loss occurs at the port or upstream ports and can eventually lead to Link Resets(LR) that occur when there is no credit for more than 1-2 seconds
- Identified by LRs increasing on the port
- Lost credits typically affect several flows in the fabric due to the significant congestion spreading effect
- These are physical errors on the medium: Port Counters will increment
- Causes for Lost Credit are typically transmission errors such as ITW, CRC, or other signal related problems – Check all physical components in connection (fiber, SFPs, patch panels, HBA)

Oversubscription/Over Utilization

- Over Utilization is when more data has been solicited than can be received at the link rate
- Hosts solicit data via SCSI Read commands
- Storage arrays solicit data via XFR_RDYs
- R_RDYs are returned such that the FC switch can always transmit
  - The link reaches line rate for a significant period of time
  - This is different from slow drain since there is no R_RDY delay
- As more data is requested it “backs up” in the fabric
- This causes congestion which looks identical to “slow drain”

- This is mostly a host issue although it could occur on target!

Oversubscription/Over Utilization

- Each SCSI Read specifies a Data Length in bytes
- Typical Read Data Lengths are 0x8000 to 0x40000 bytes
- Time to transfer bytes at various link speeds

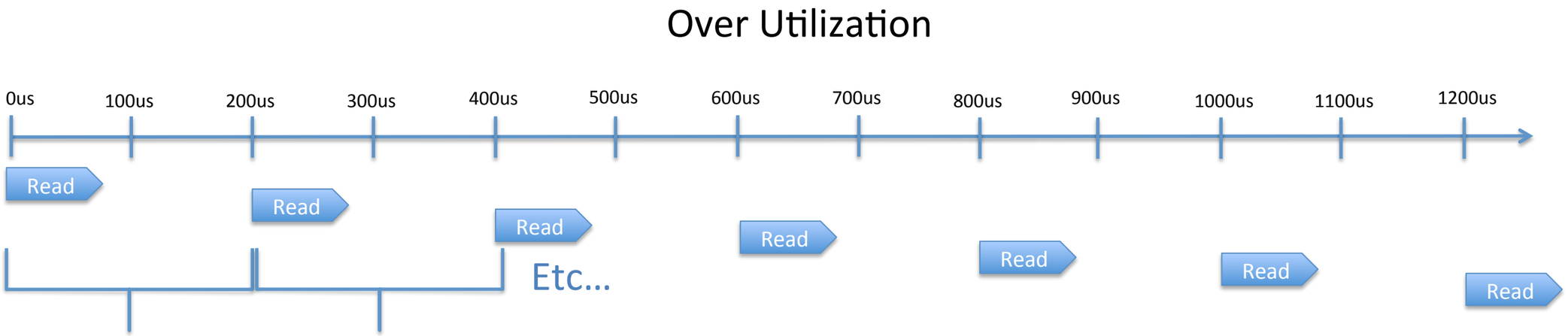
FC Link Speed	4Gbps	8Gbps	16Gbps	32Gbps
0x8,000 bytes	19us	10us	5us	3us
0x20,000 bytes	308us	154us	77us	39us
0x40,000 bytes	616us	308us	154us	77us

- If Reads are generated more frequently than the data solicited can be transmitted it will cause over utilization

Oversubscription/Over Utilization

Example:

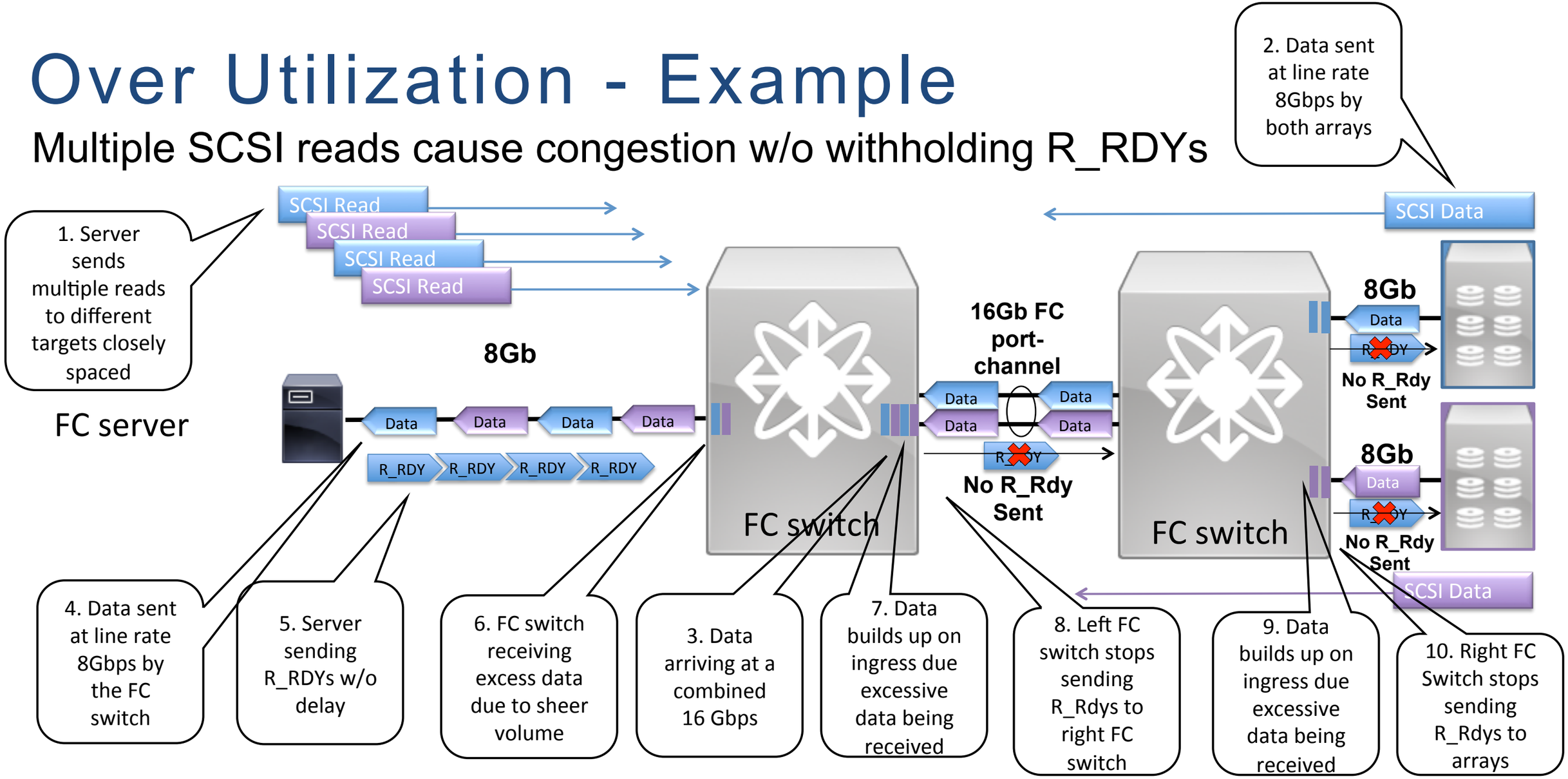
- Speed: 8Gbps
- SCSI Read Data Length: 0x40000 (1/4 MB)
- Time to transfer 0x40000 bytes @ 8Gbps: 308us



Reads spaced out 200us but will generate data for 308us

Over Utilization - Example

Multiple SCSI reads cause congestion w/o withholding R_RDYs



This isn't strictly "slow drain" but the effects are exactly the same!

How to Prevent, Identify, and Resolve Performance Problems

Summary: Three Main Causes

- Credit Stalled Device/Slow-Drain
  - Abnormal or unexpected device behavior
  - Device induced credit latency
    - I.e. “Slow Drain,” “Slow Draining” device, or “Slow Drainer”
    - Sender’s Tx credits hit 0 forcing a stop to any transmissions until a credit is received

- Lost Credit
  - Physical errors
  - Credits and frames are not sent reliably resulting in a loss of credits over time

- Oversubscription/Overutilization
  - Bandwidth mismatch
    - Speed mismatch (e.g. 16G to 4G)
    - Fan-in mismatch (e.g. 8-ports to 1-port)
    - Device simply requesting more data than it can consume at its given link rate

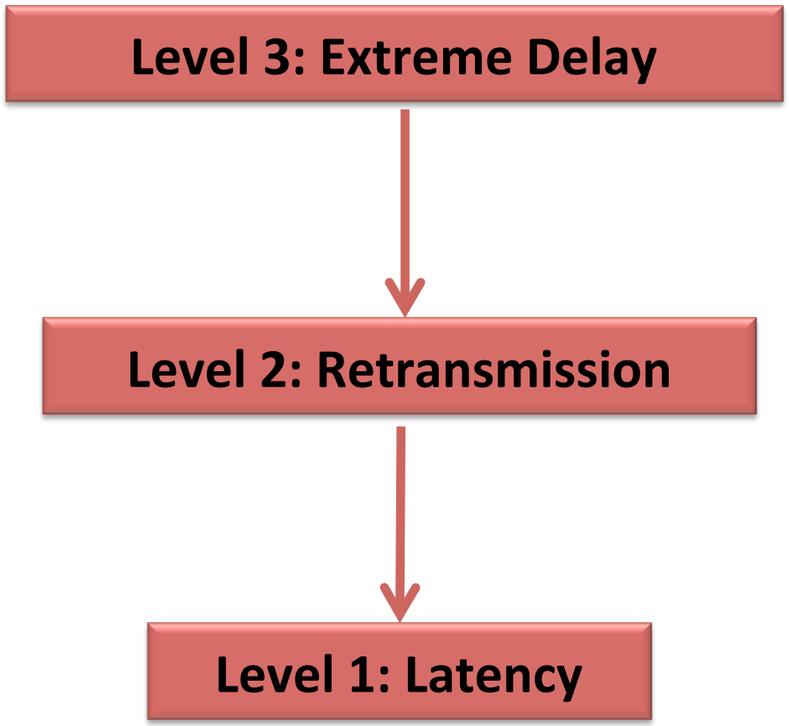
Summary: Prevent, Identify, Resolve

- Credit Stalled Device/Slow-Drain/Lost Credit
  - Utilize SAN Fabric Tools
    - Brocade Tools: Network Advisor, Dashboards, MAPS, Fabric Vision: Flow Vision, IO Insight, Fabric Performance Impact, FEC, SDDQ, Buffer Credit Recovery, Port Fencing, and ClearLink Diagnostics
    - Cisco Tools: no-credit-drop, port-monitor portguard, slowport-monitor, show tech-support slowdrain, congestion-isolation, DCNM Slow Drain Analysis
- Oversubscription/Overutilization
  - Tricky to identify
  - Use monitoring tools to identify periods of high Tx Utilization
  - Time correlate to slow drain indications
  - SAN Design Considerations
    - Reduce load on host
    - Increase speed of HBA
    - Add additional HBAs
    - Reduce the number of targets the host is zoned with
    - Implement array side rate limiting (array vendor specific)

Troubleshooting Slow Drain

Methodology

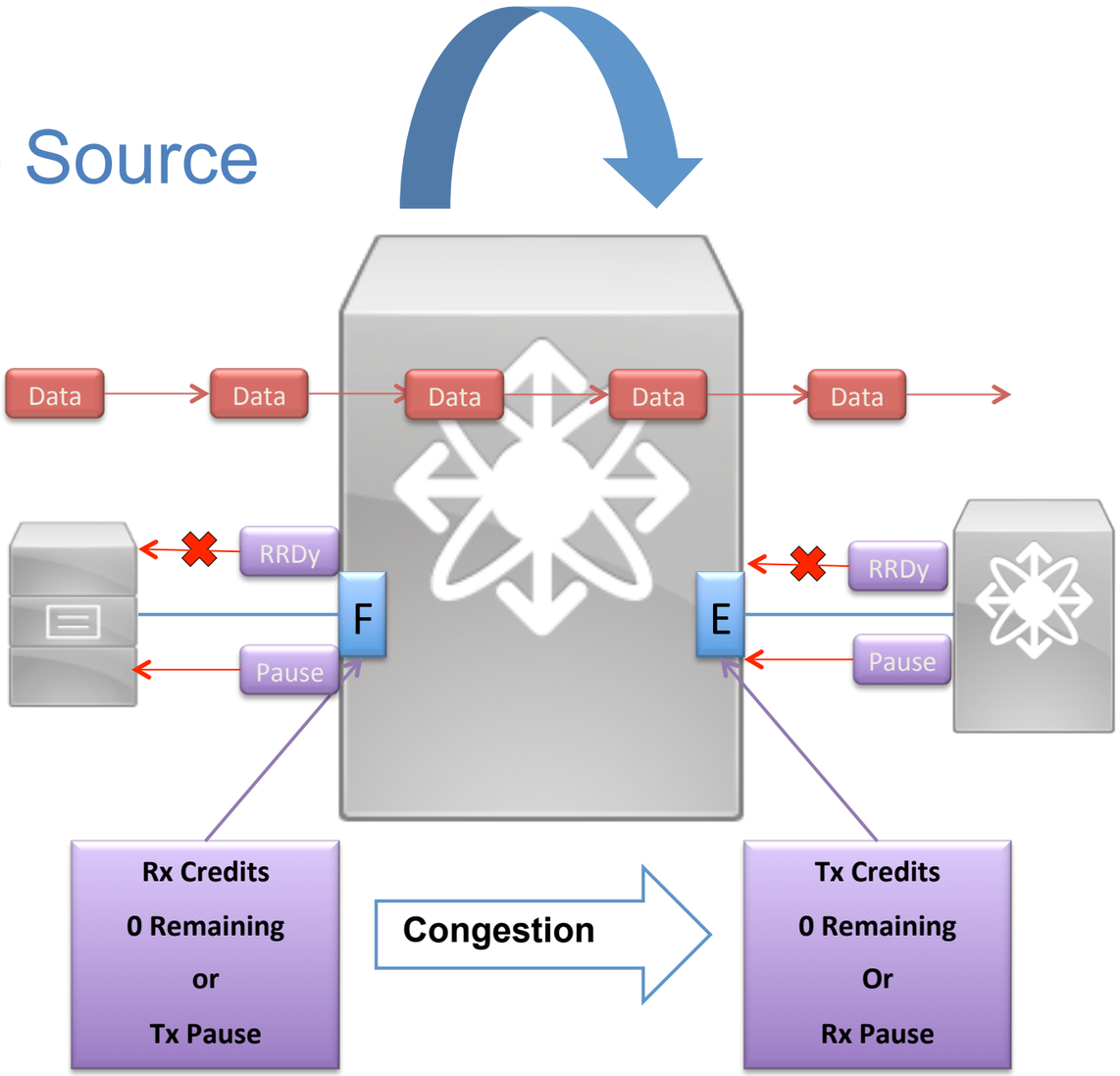
- We recommend troubleshooting slow drain in the following order:



Identifying Slow Drain

Methodology – Follow Congestion to Source

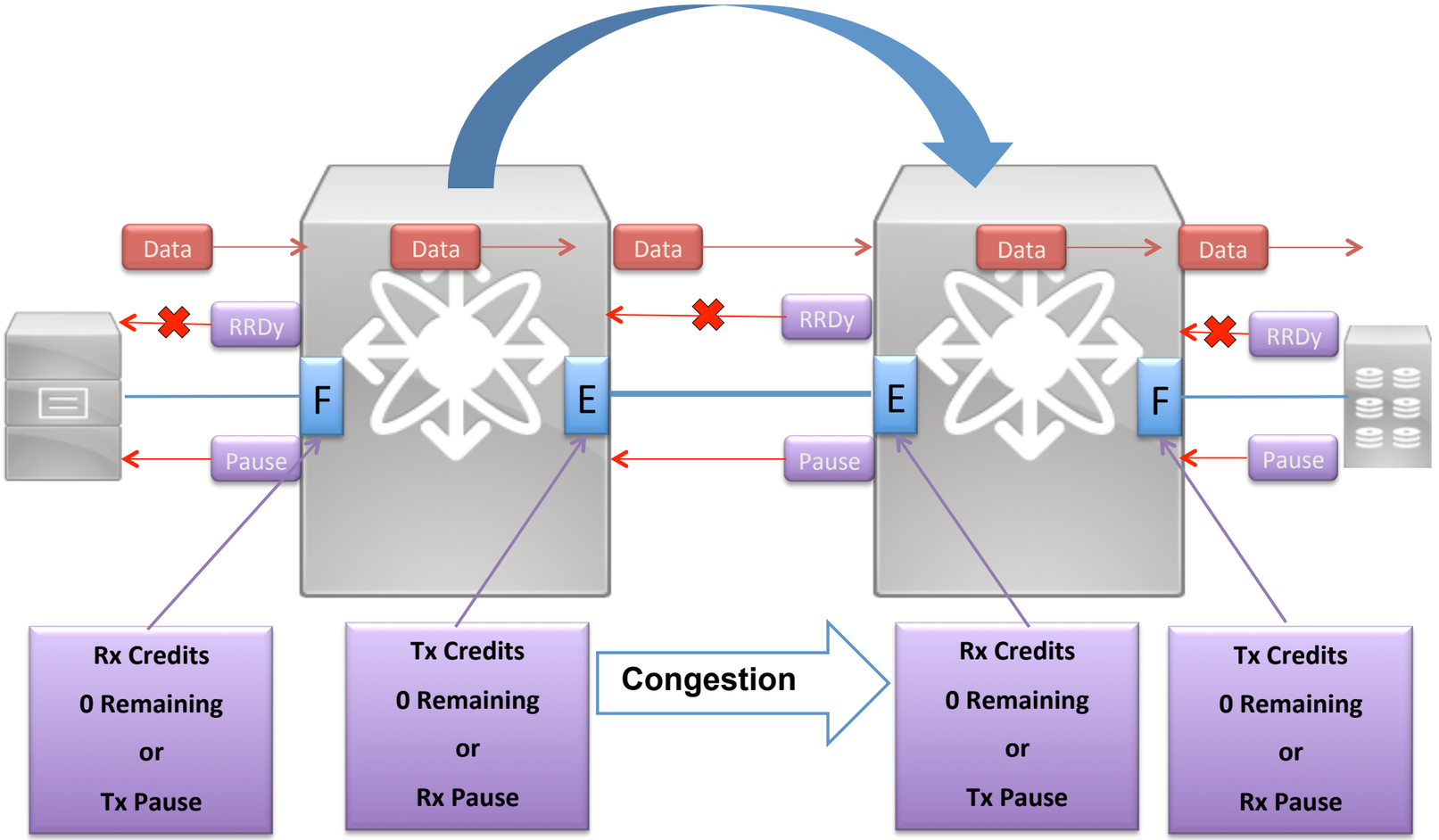
- If Rx congestion then find ports communicating with this port that have Tx congestion
  - Zoning defines which devices communicate with this port
  - Understand topology
- If port communicating with port showing Rx congestion is FCIP
  - Check for TCP retransmits
  - Check for overutilization of FCIP



Identifying Slow Drain

Methodology - Follow Congestion to Source

- If Tx congestion found
  - If F port then device attached is slow drain device
  - If E port then go to adjacent switch and continue troubleshooting
  - Continue to track through the fabric until destination F-port is discovered



Summary

- Fibre Channel protocol is extremely reliable and robust
- If problems are encountered follow congestion to the source
- Multiple vendor specific tools are available to identify, mitigate
- Almost always the real solution lies in fixing/updating the end device causing the problem

For more information, watch the FCIA webinar
"**Fibre Channel Fundamentals**" at:
https://www.brighttalk.com/webcast/14967/255009":

Q&A

Our Next FCIA Webcast:

Fibre Channel Cabling

April 19, 2018

10:00 am PT

Register at:

https://www.brighttalk.com/webcast/14967/303881

After this Webcast

- Please rate this event – we value your feedback
- We will post a Q&A blog at http://fibrechannel.org/ with answers to all the questions we received today
- Follow us on Twitter @FCIAnews for updates on future FCIA webcasts
- Visit our library of FCIA on-demand webcasts at http://fibrechannel.org/webcasts/ to learn about:
  - Fibre Channel Fundamentals
  - FC-NVMe
  - Long Distance Fibre Channel
  - Fibre Channel Speedmap
  - FCIP (Extension): Data Protection and Business Continuity

Thank You