SAN and Data Transport Technology Evaluation at the NASA Goddard Space Flight Center (GSFC)

Hoot Thompson
Patuxent Technology Partners, LLC
+1-703-250-3754
hoot@ptpnow.com
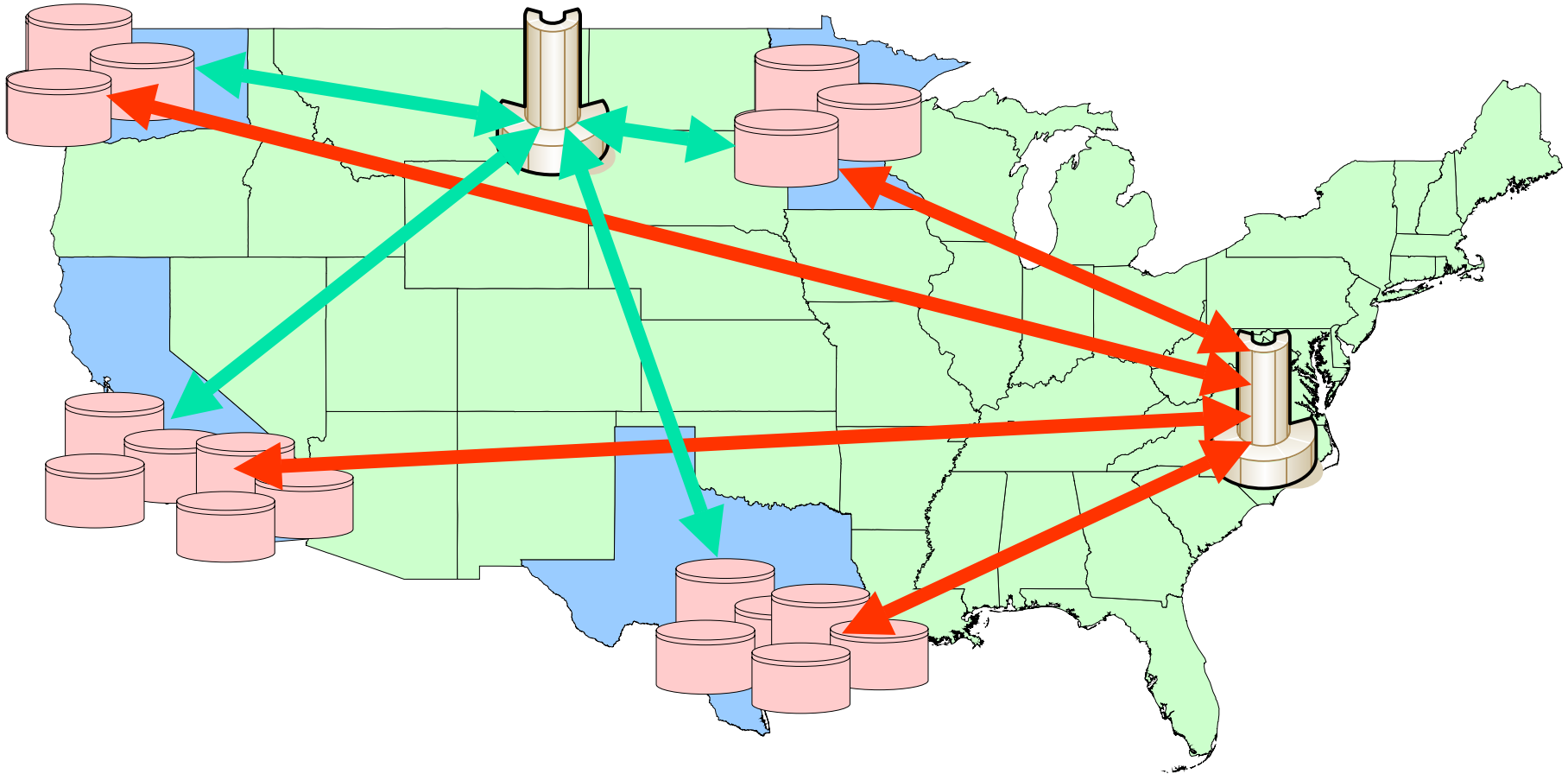
GSFC Objective

- Technology exploration focused on FC over IP
  - Emphasis on using readily available networks
  - Avoid "dark fibre" dependency
- First steps towards a SAN grid
  - Tools to locate data of interest
  - Distributed data stores processed directly by users
    - No copies of data
    - Secure, unconstrained, high bandwidth, shared

Notional SAN Grid

High bandwidth access to shared data

GSFC SAN Pilot
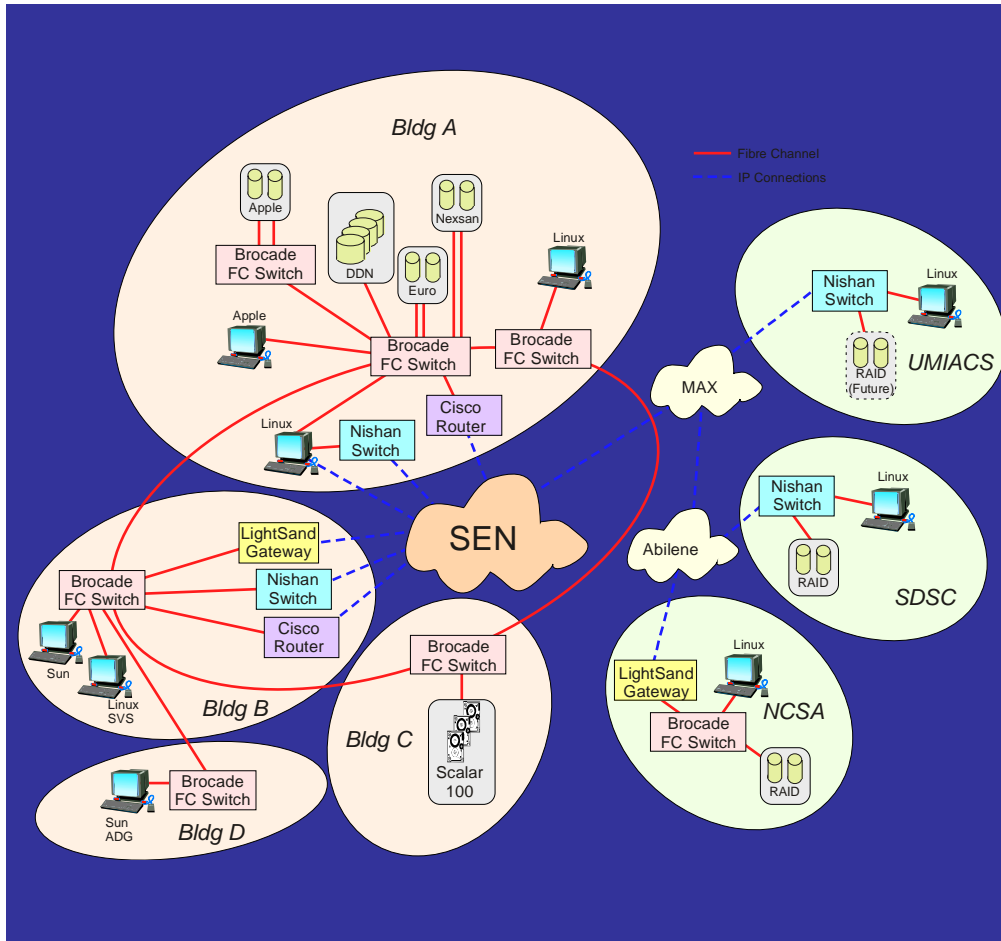
Sites
- GSFC Campus
- UMIACS (U of MD)
- SDSC
- NCSA

Key products
- Nishan Series 3000
- LightSand i-8100
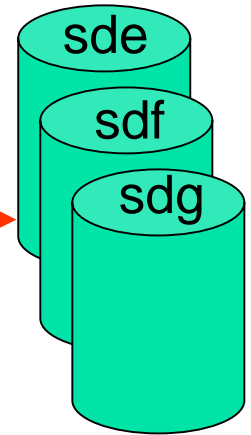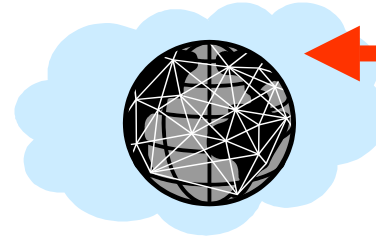- Cisco
- Brocade
- ADIC

Networks
- SEN: GSFC
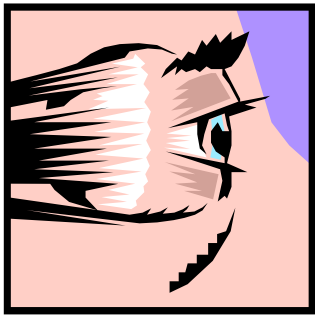- MAX: UMIACS
- CALREN-2: SDSC
- Abilene: Internet2 backbone

4

FC Over IP: Products In The Spotlight

- Nishan IPS 3000
  - McData acquisition
  - iFCP protocol
  - Key features
    - FastWrite™
    - Compression
    - Variable MTU
- LightSand i-8000
  - Formerly SandCastle
  - FCIP protocol
  - UDP as transport
  - Autonomous regions

Product Core Capability

Normal SCSI Exchange
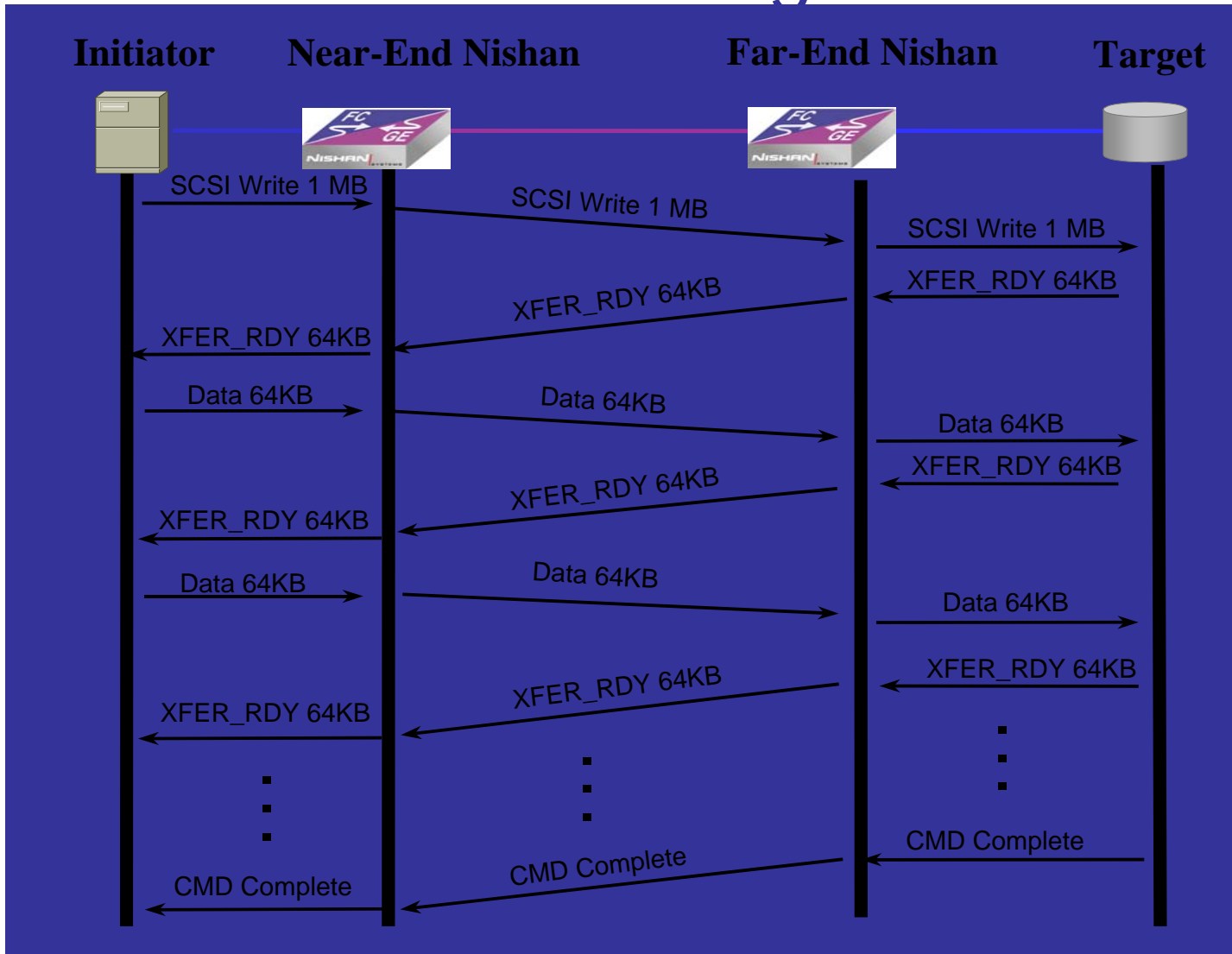
Fast Write SCSI Exchange

Initiator   **Near-End Nishan**   **Far-End Nishan**   **Target**

SCSI Write 1MB

SCSI Write 1MB

XFER_RDY 1MB

SCSI Write 1MB

Data 1MB

Data 1MB

XFER_RDY 64KB

Data 64KB

XFER_RDY 64KB

Data 64KB

XFER_RDY 64KB

Data 64KB

CMD Complete

CMD Complete

CMD Complete

8
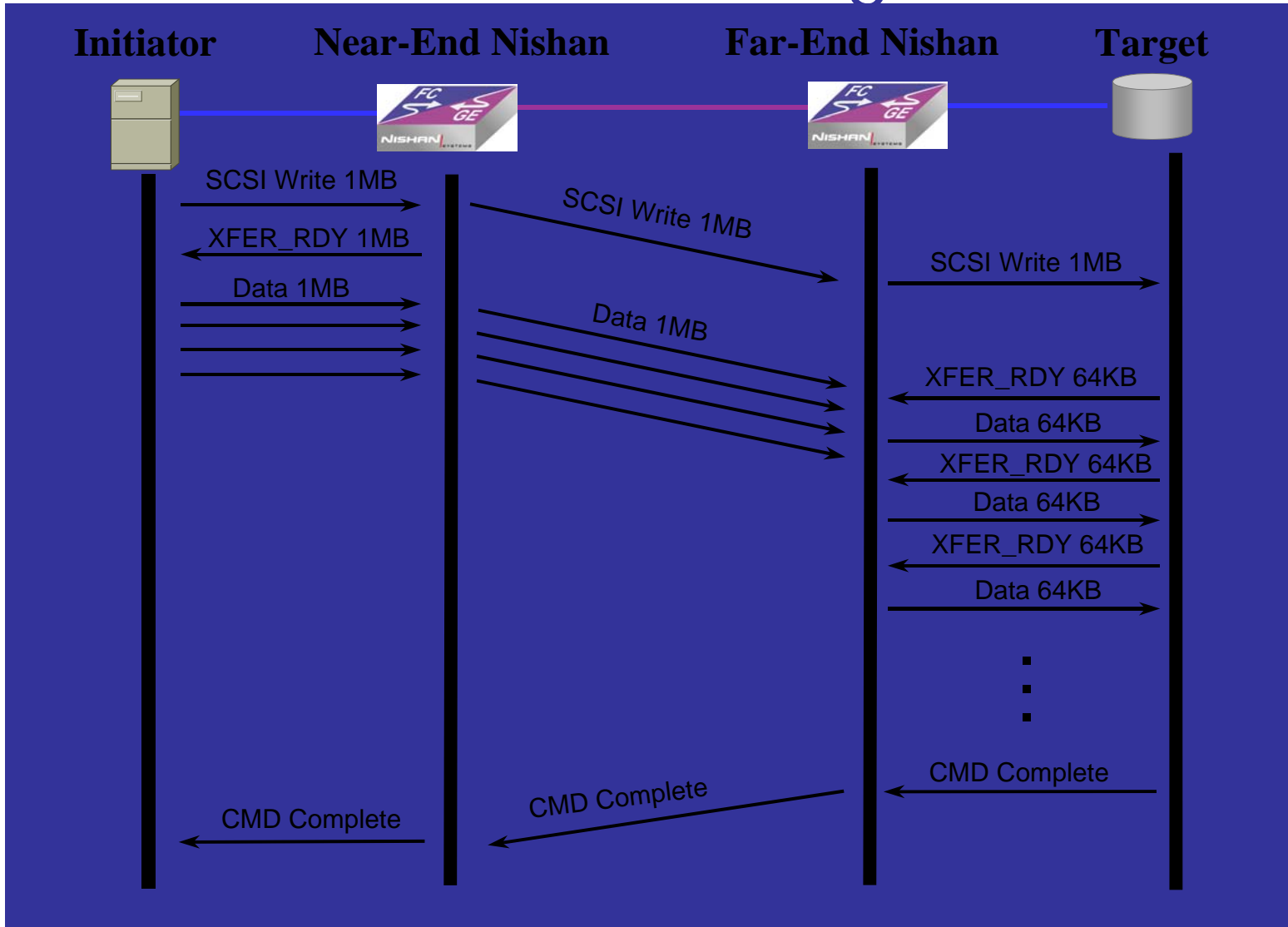
LightSand Autonomous Regions

Test Approach And Methods

- Focus on infrastructure technologies required for SAN grid
  - High bandwidth data transport using existing IP networks
    - Shared
    - Bandwidth contention
  - Shared file systems
- Evaluate technology in controlled, GSFC environment
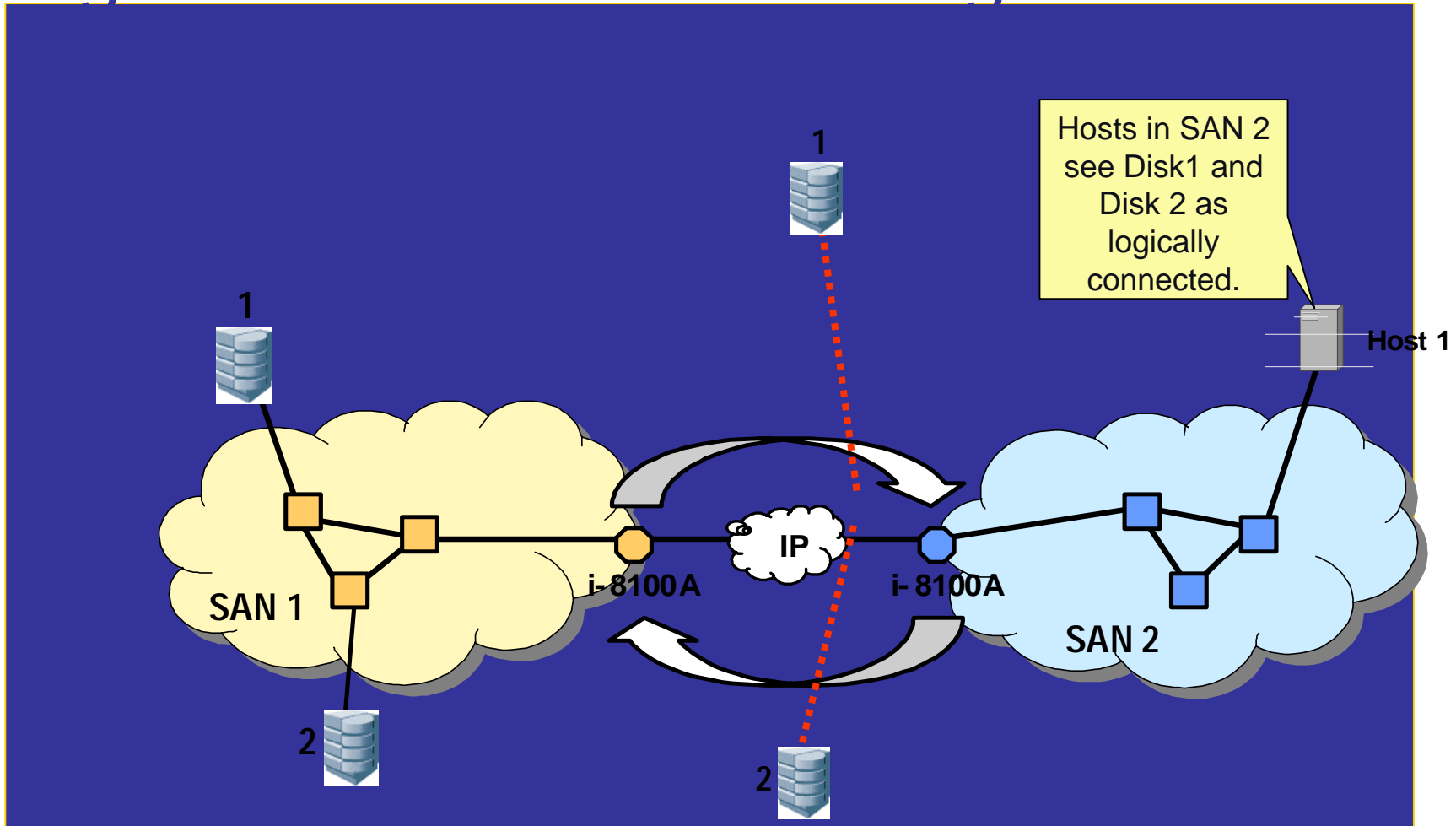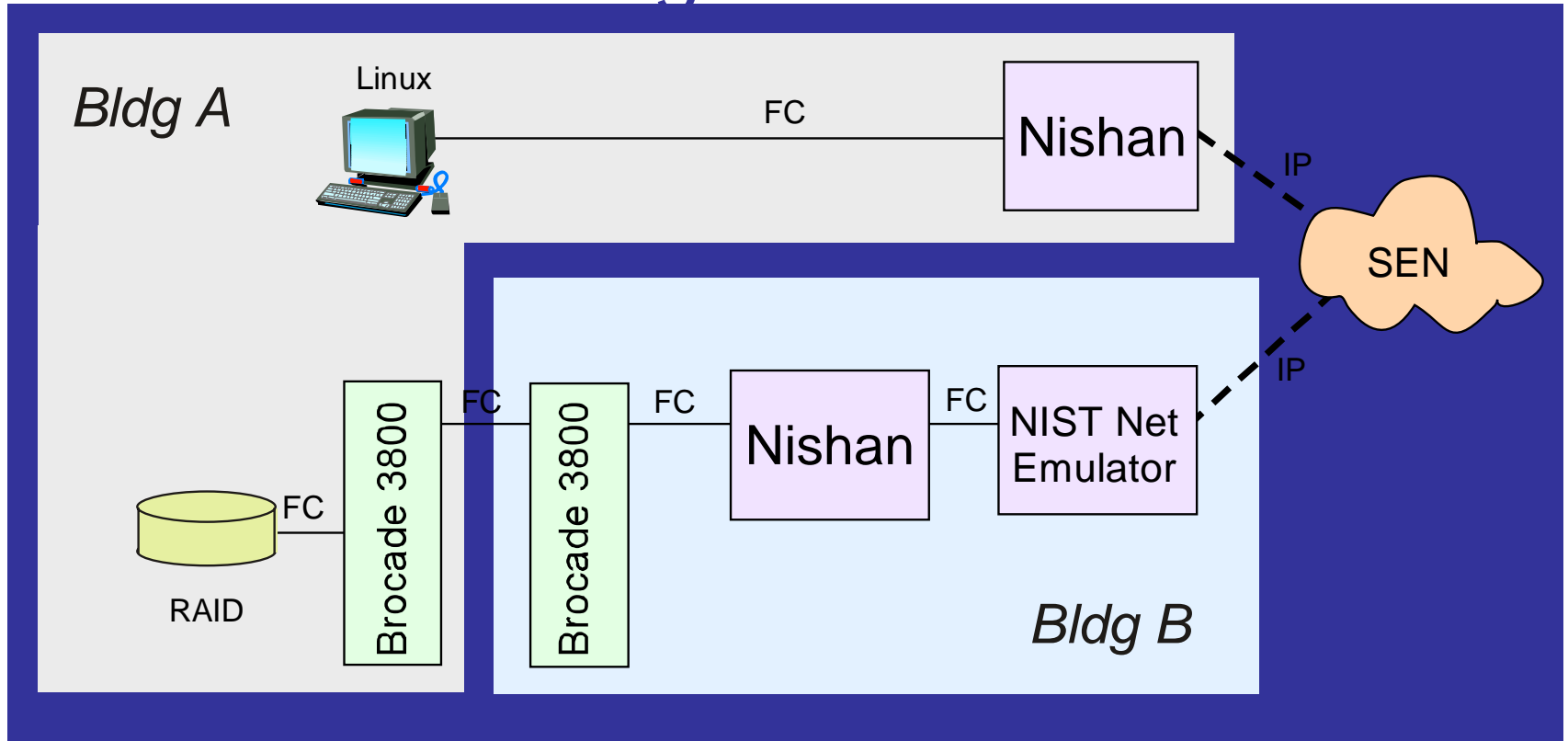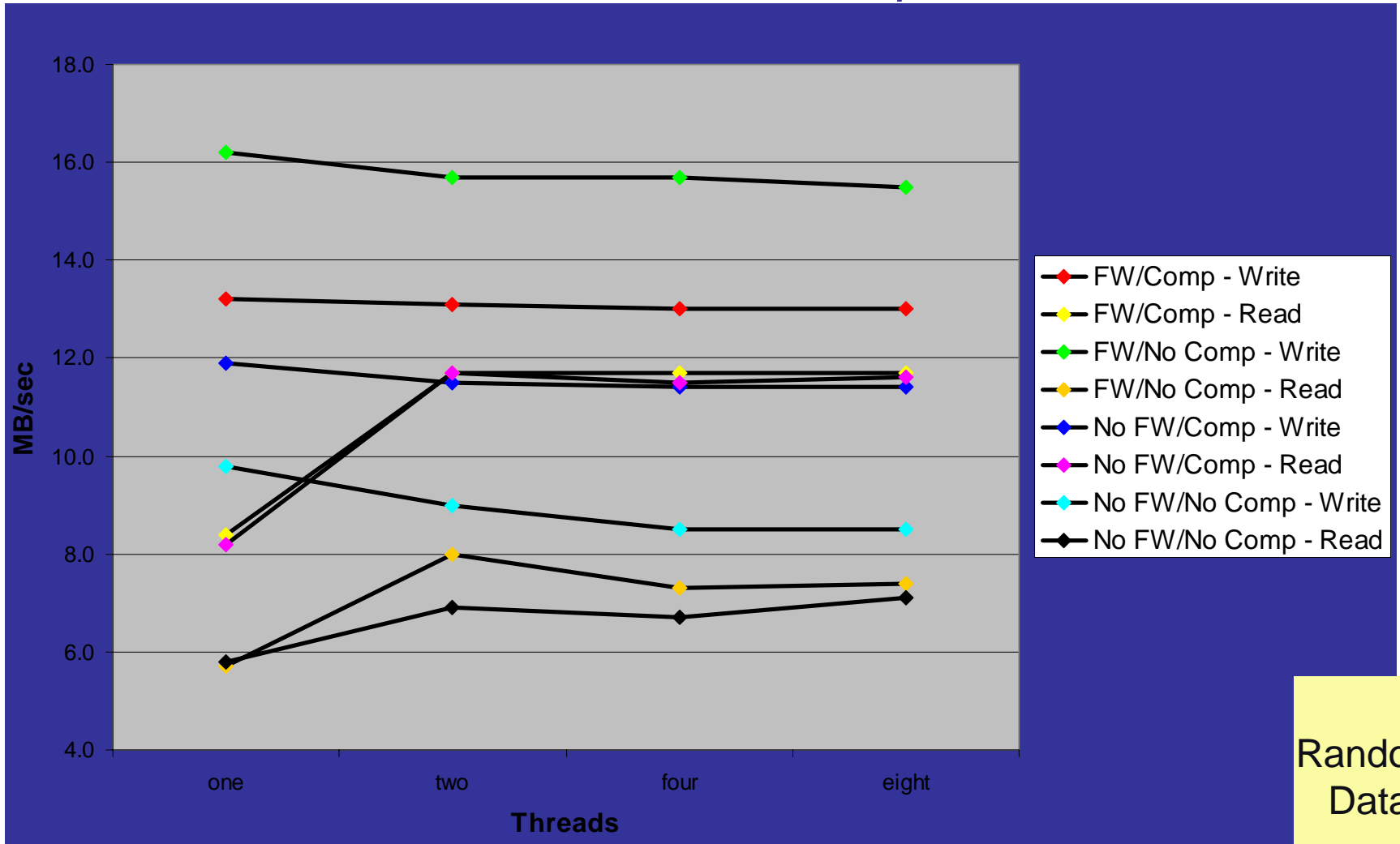 - Clean, congestion free network
  - Range of simulated round trip times (rtt)
- Connect to third-party sites with similar functionality interests
  - UMIACS at College Park, MD (~0msec rtt)
  - NCSA at Urban-Champagne, IL (~30msec rtt)
  - SDSC at San Diego, CA (~70msec)
- Native file systems and shared file system
  - Linux ext2
  - ADIC StorNext File System (SNFS)
  - SGI CXFS
- Standardized benchmarks emphasizing throughput performance
  - Large file sizes (multiple GBs), large block sizes (1MB)
  - lmdd for quick tests
  - IOzone for stress tests – multithreaded, modified for random data

GSFC Test Configuration

GSFC Results: 35msec rtt, MTU=1500

Legend:
- FW/Comp - Write
- FW/Comp - Read
- FW/No Comp - Write
- FW/No Comp - Read
- No FW/Comp - Write
- No FW/Comp - Read
- No FW/No Comp - Write
- No FW/No Comp - Read

Y-axis: MB/sec (4.0 to 18.0)

X-axis: Threads (one, two, four, eight)

Random Data

GSFC Results: 35msec rtt, MTU=4096

Legend:
- FW/Comp - Write
- FW/Comp - Read
- FW/No Comp - Write
- FW/No Comp - Read
- No FW/Comp - Write
- No FW/Comp - Read
- No FW/No Comp - Write
- No FW/No Comp - Read
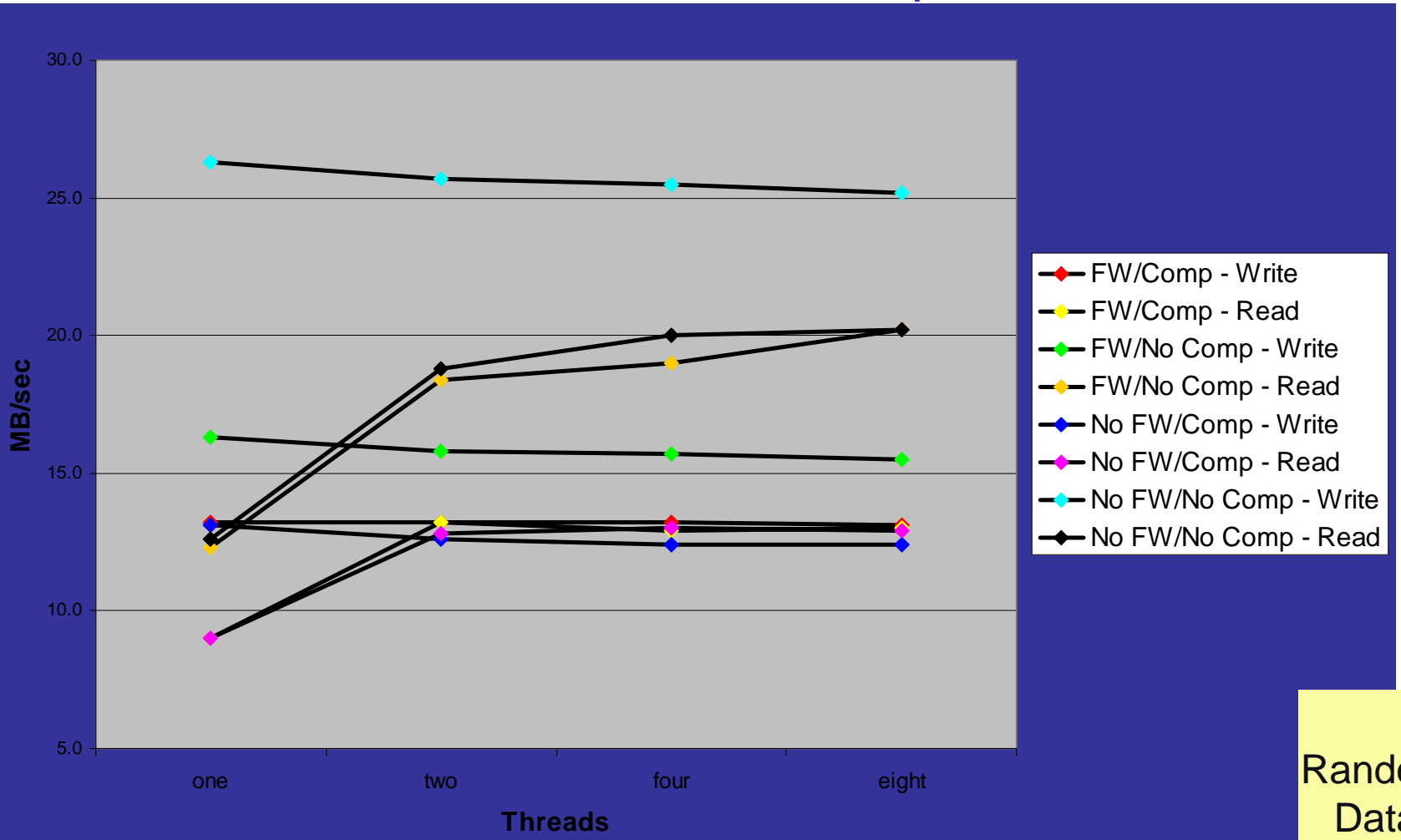
Random Data

UMIACS Test Configuration
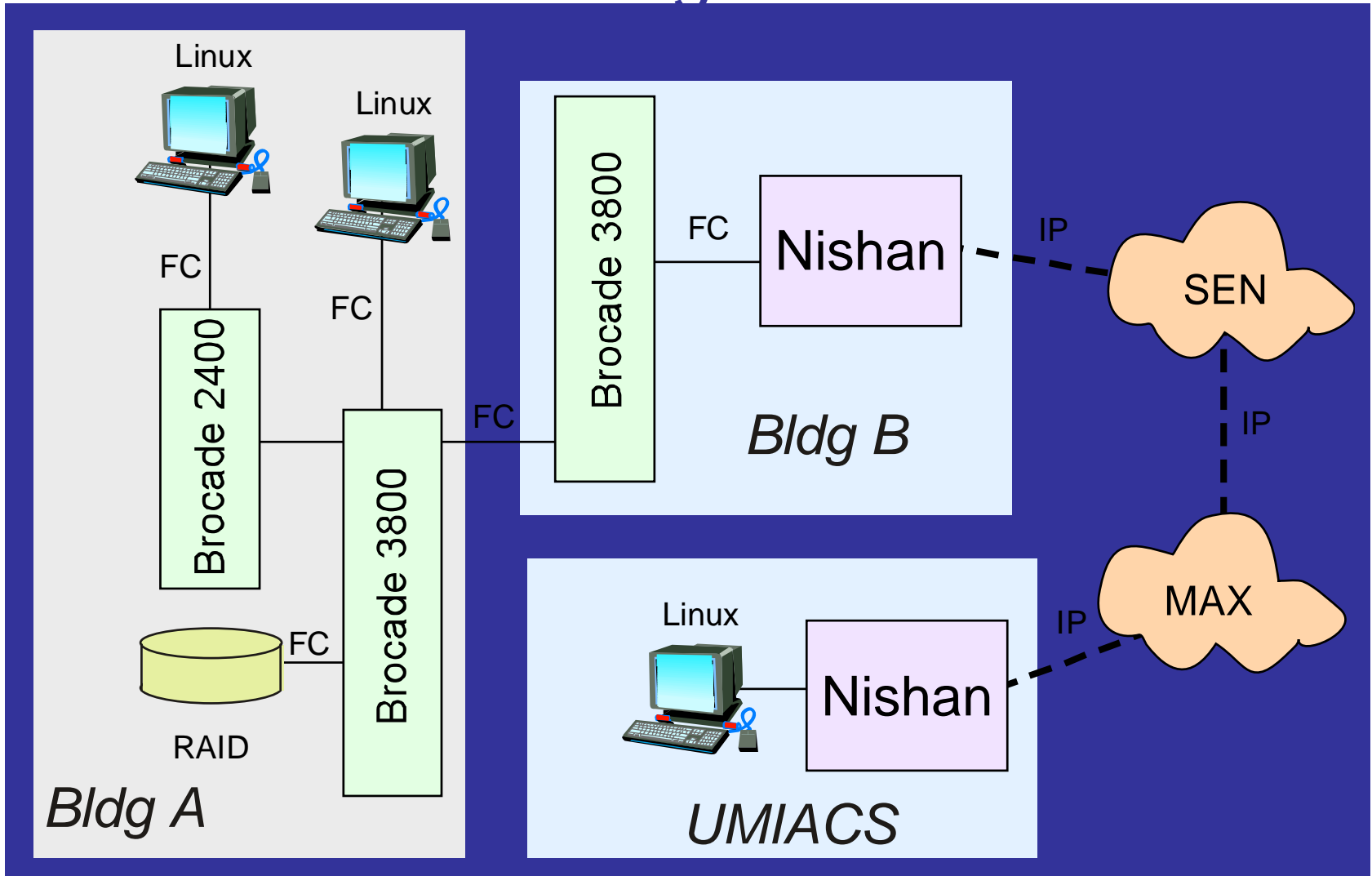
UMIACS Results: ~0msec rtt, MTU=1500

Threads	FastWrite, Compression		No FastWrite, No Compression	
	Write	Read	Write	Read
one	12.8	9.5	38.1	14.1
two	12.9	11.7	47.3	19.8
four	12.8	11.6	28.9	20.6
four	12.8	11.6	59.8	25.8

Random Data

SDSC Test Configuration

SDSC Results: rtt=70msec, MTU=1500

Legend:
- FW/Comp - Write
- FW/Comp - Read
- FW/No Comp - Write
- FW/No Comp - Read
- No FW/Comp - Write
- No FW/Comp - Read
- No FW/No Comp - Write
- No FW/No Comp - Read
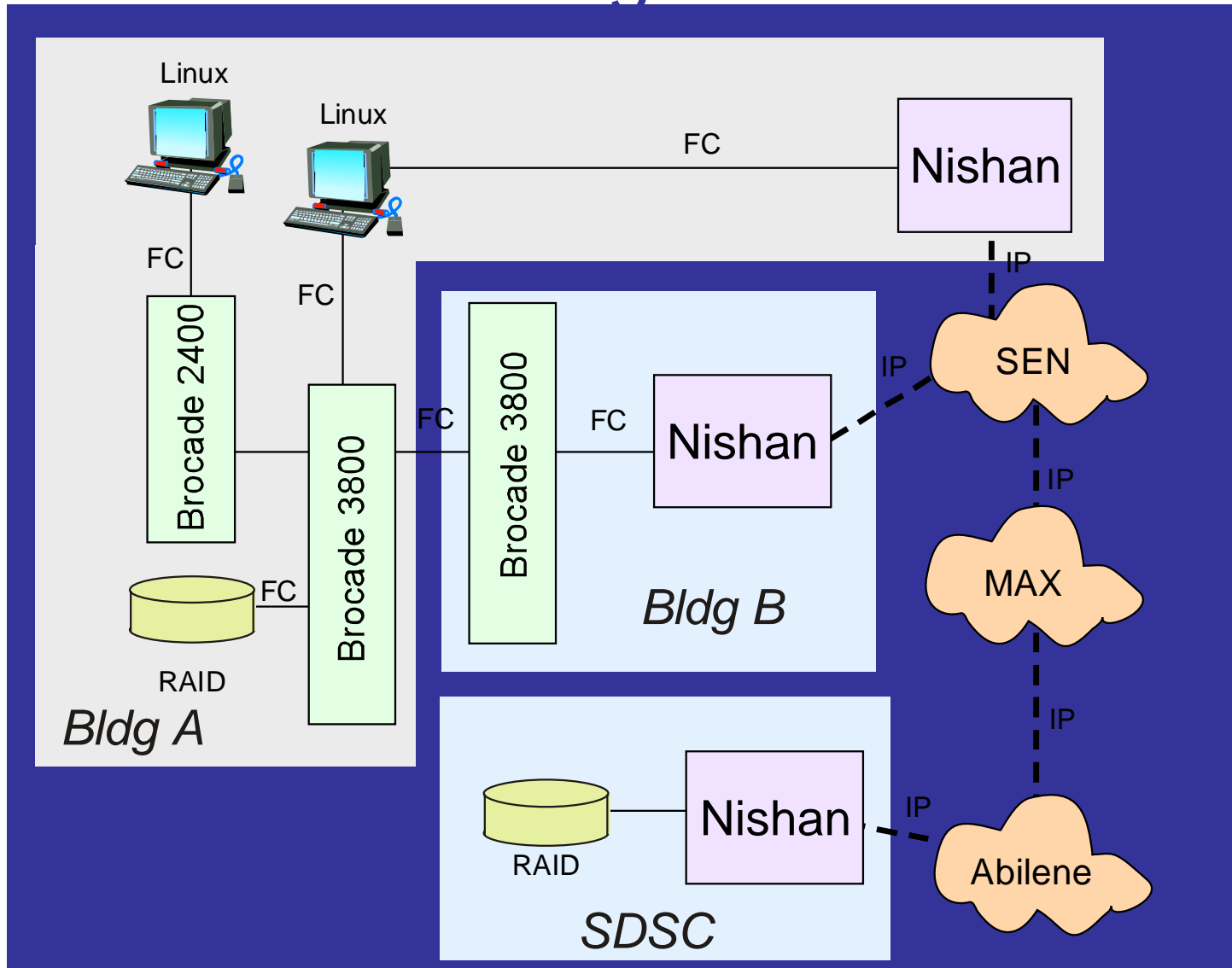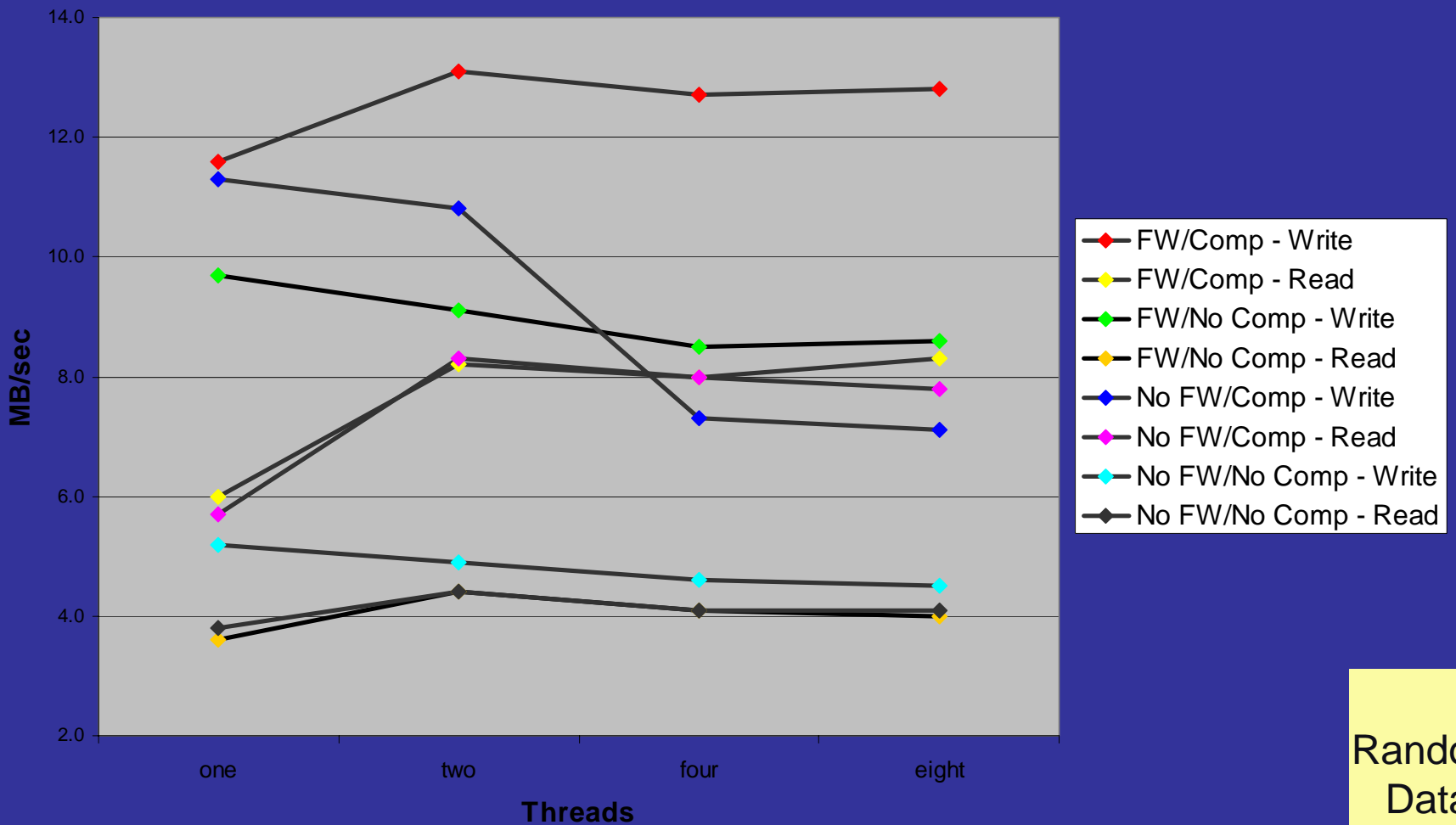
Random Data

SDSC Results: rtt=70msec, MTU=1500

Legend:
- FW/Comp - Write
- FW/Comp - Read
- FW/No Comp - Write
- FW/No Comp - Read
- No FW/Comp - Write
- No FW/Comp - Read
- No FW/No Comp - Write
- No FW/No Comp - Read
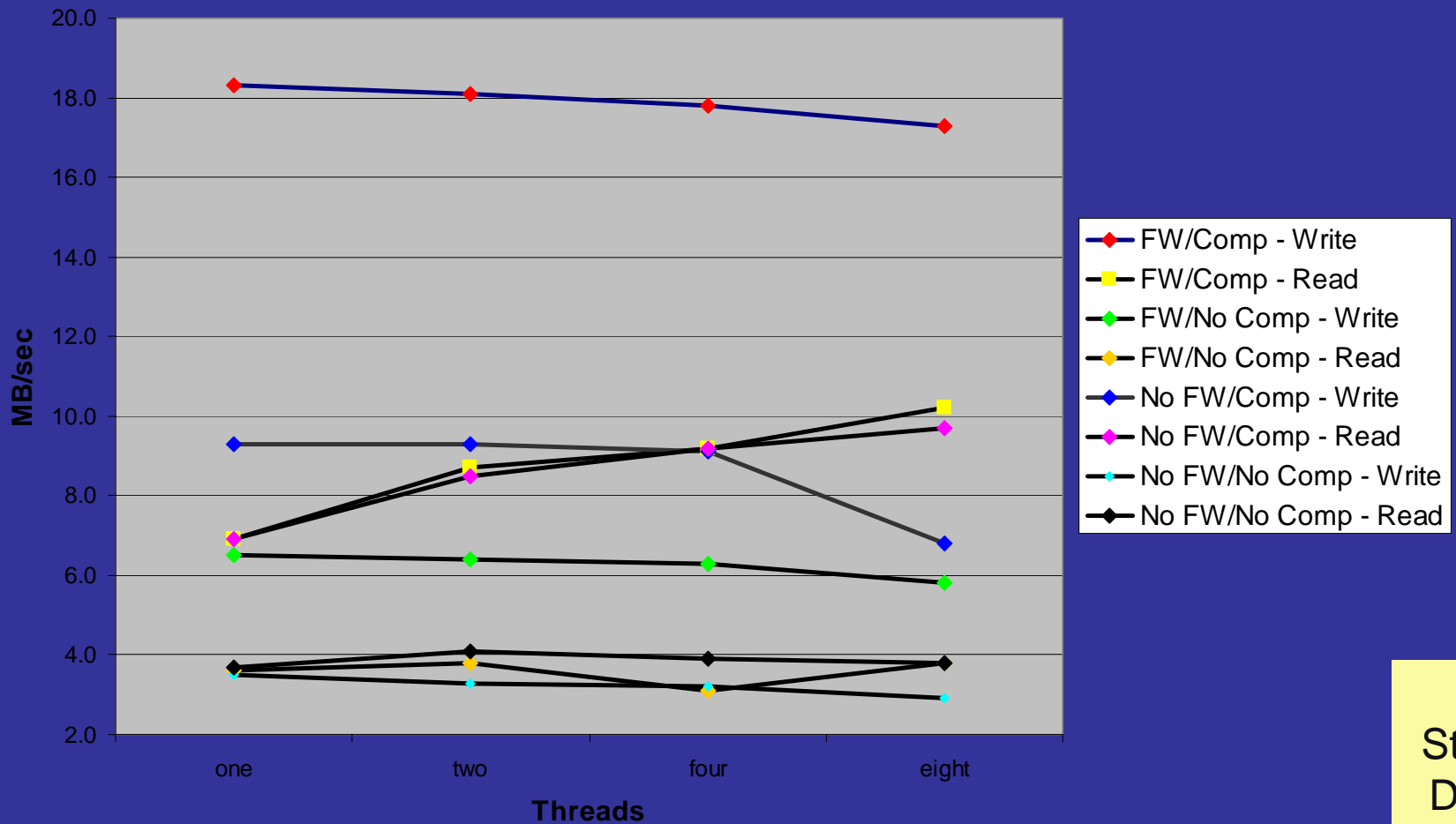
Static Data

GSFC To SDSC Comparisons

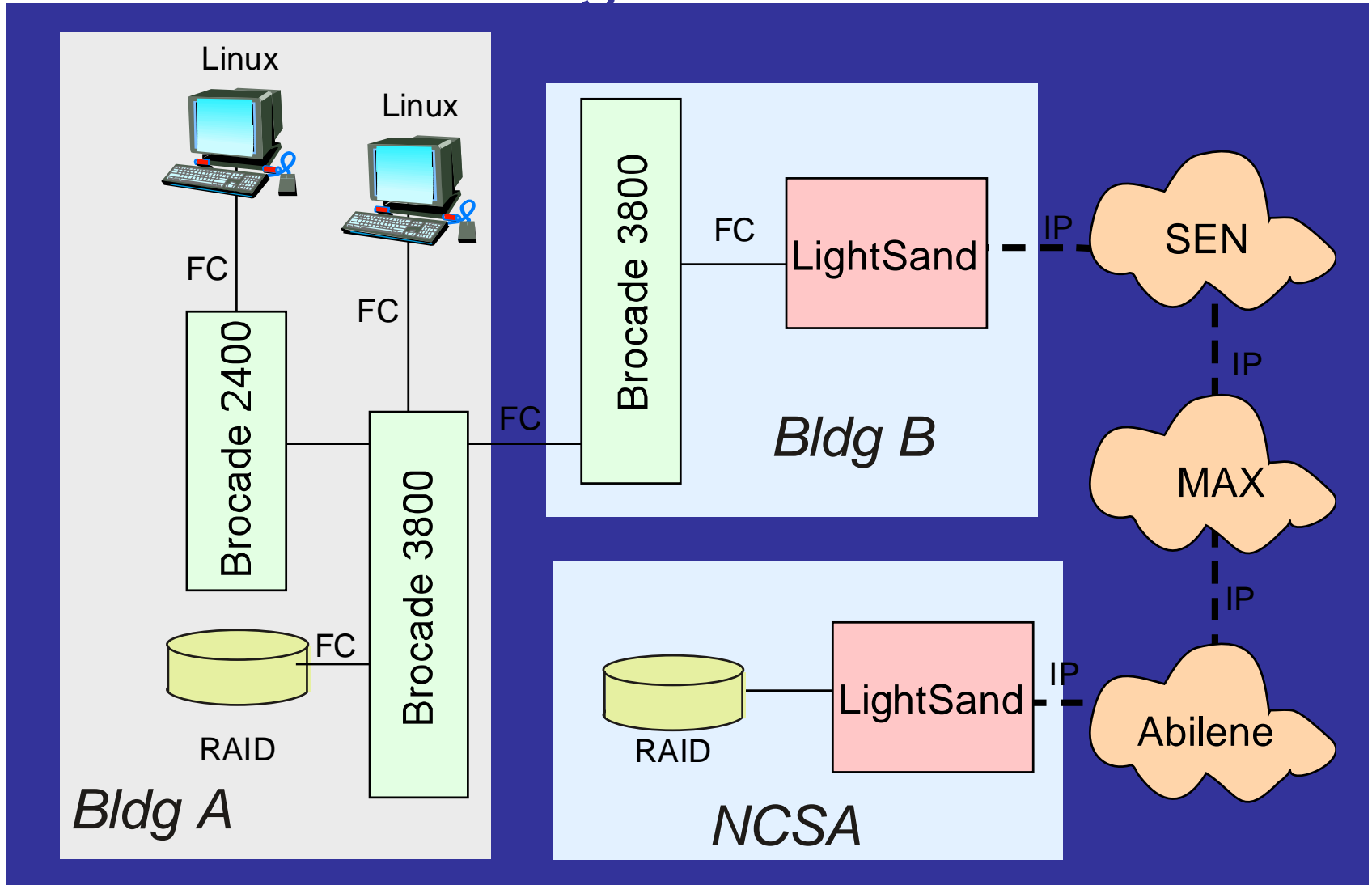| Threads	GSFC => GSFC rtt delay => 70msec MTU => 4096		GSFC => SDSC rtt actual => 70msec MTU => 1500	
	Write	Read	Write	Read
one	13.1	5.6	11.6	6.0
two	13.1	11.5	13.1	8.2
four	13.1	12.5	12.7	8.0

Random Data

NCSA Test Configuration

NCSA Results: rtt=30msec

Threads	Write	Read
one	37.0	12.1
two	37.5	28.9
four	37.3	35.6
four	37.3	36.2

Random Data

Summary

IP Device	Pros	Cons
General	• Perform as advertised. • Operationally fairly intuitive. • Both GUI and CLI management options. • Administrator defined level of SAN merging/isolation.	• Minimal security. • No ssh. • No CLI standard • Redundant, conflicting naming conventions. • Proprietary, same vendor product required at both ends of the WAN connection. • High skill level to configure, etc., multiple talents involved. • Incompatibilities, version issues, etc. reminiscent of the early days of FC.
Nishan 3000	• Built in performance graphs. • Good statistical info.	• Passwords in clear text.
LightSand i-8100	• Companion applications that provide data analysis.	• IP routes cleared by reboots. • Difficult to save and compare configurations.

Operational Impressions

General

  Set up and configuration relatively straightforward
    GUI and CLI options appreciated
    Application specific tuning a likelihood
    Expanded zone definition
    Management still Windows® centric
  Site security policies/firewalls require navigation
    ICMP off at NASA
    No SNMP at UMIACS
    Port 111 at NCSA
  Skill set to maintain and operate needs to be considered
Nishan
  Interesting interaction between compression and FastWrite
LightSand
  UDP can be throttled to not overrun TCP traffic

Future Testing

Full GE and jumbo frames between all the sites
Shared file system
   SNFS
   CXFS
Real applications
   GSFC's Scientific Visualization studio (SVS)

   Landsat Data Continuity Mission (LDCM) Grid Prototype (LGP)
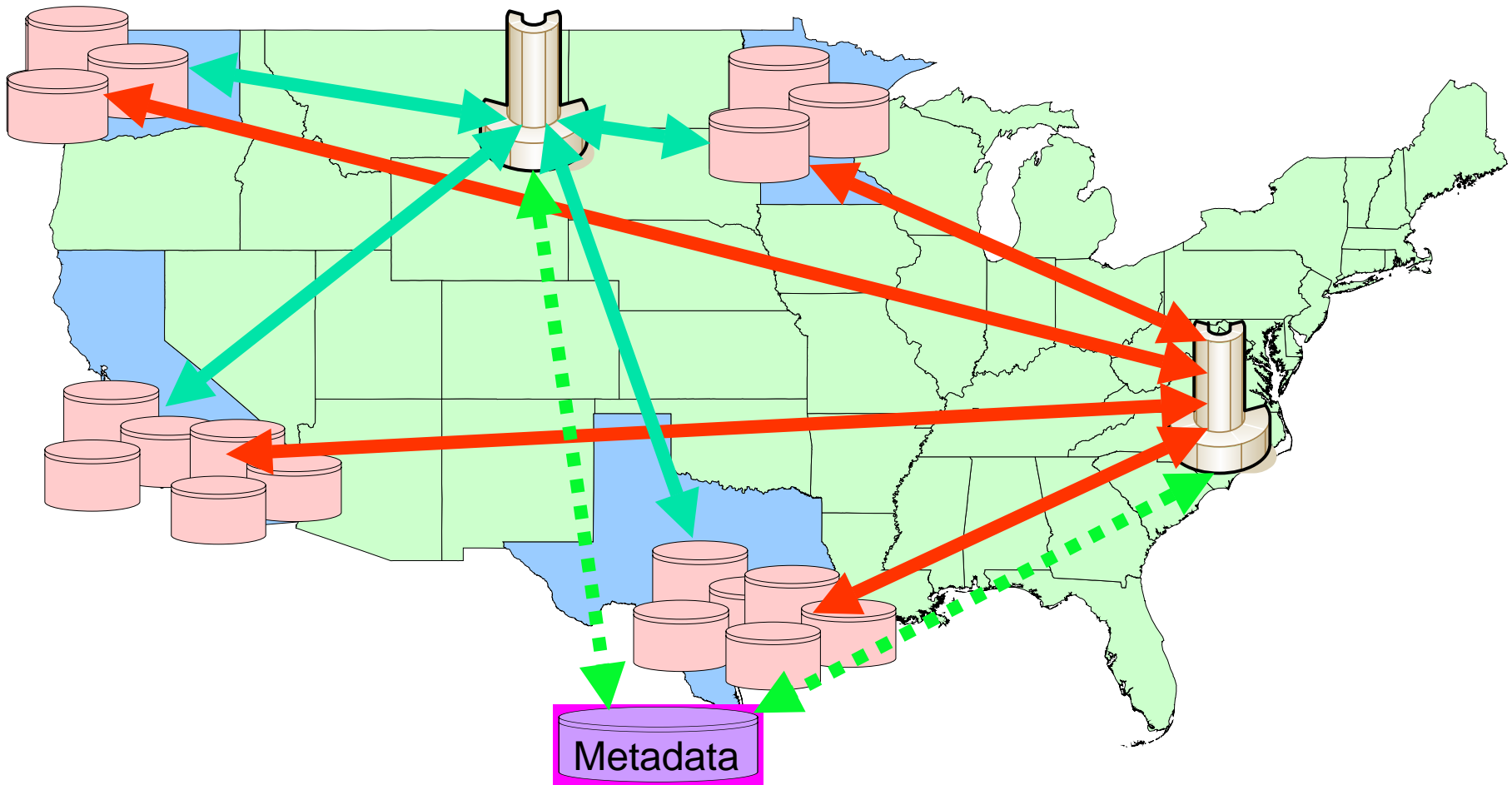More comparative testing – NIST NET and actual
   Interplay of rtt, congestion, packet loss and the basic protocols
   Analysis of protocol strengths and weaknesses
   Comparison to measured wire speed and theoretical peak

Notional SAN Grid – Shared Storage

Metadata

Acknowledgements

- Bill Fink – GSFC
- Paul Lang – GSFC
- Wei-Li Liu – GSFC
- Aruna Muppalla – GSFC
- Bryan Bannister – SDSC
- Nathaniel Mendoza – SDSC
- Chad Kerner – NCSA
- Fritz McCall – UMIACS
- Vendor community