Enhancing Data Transfer Performance with DTNs and the Science DMZ

Ken Miller
ken@es.net
ESnet / Lawrence Berkeley National Laboratory

Workshop on Science DMZ and P4-DPDK Online
August 6-8, 2024

National Science Foundation Award #2328479

Common Pitfalls

- Legacy transfer tools, scripts, and workflow vs. Globus Ecosystem
- A data, infrastructure, and users are the same.
  - One campus research data transfer can be as much traffic as all campus users Netflix data. So how many researchers do you support?
- High performance expectations with default configurations.
  - Just get it running vs. performance measurement expectations
  - Standard MTU vs Jumbo frames
- Research data should be BGP configured to route first on ESnet, Internet2, or your local regional network, then commodity internet.
- R&E networks are build for data movement
  - Express lane for research data on purpose built data networks
- Uptime and Availability are measured but performance is not.
- Not testing IPv4 and IPv6 the same.

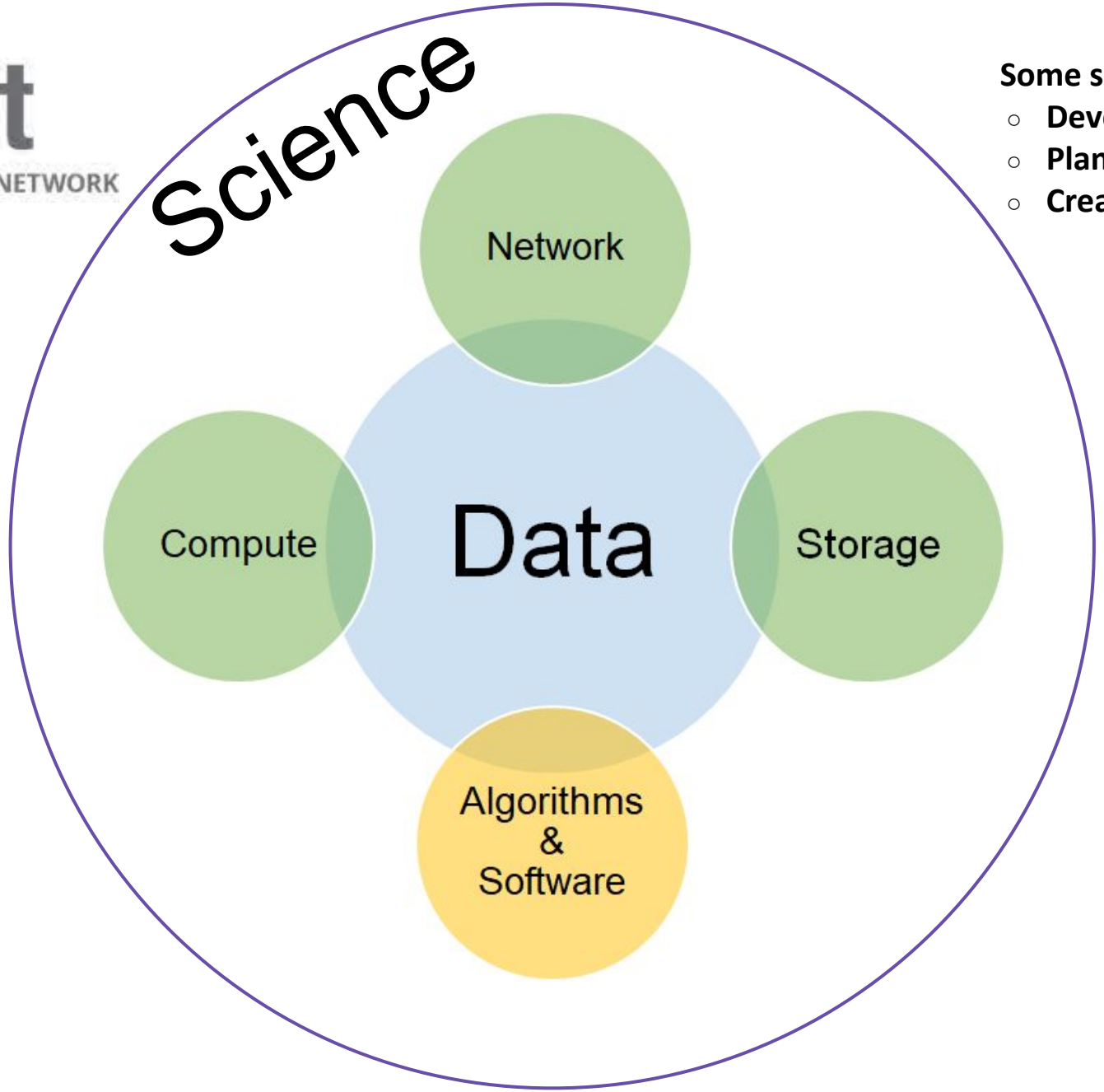
EPOC
Engagement and Performance Operations Center

Network as ~~Infrastructure~~ *Instrument*

Connectivity is the first step – ***usability*** must follow

EPOC
Engagement and Performance Operations Center

ESnet
ENERGY SCIENCES NETWORK



Some specific issues for networks are

- **Development of services**
- **Planning capacity growth**
- **Creation of collaborations**

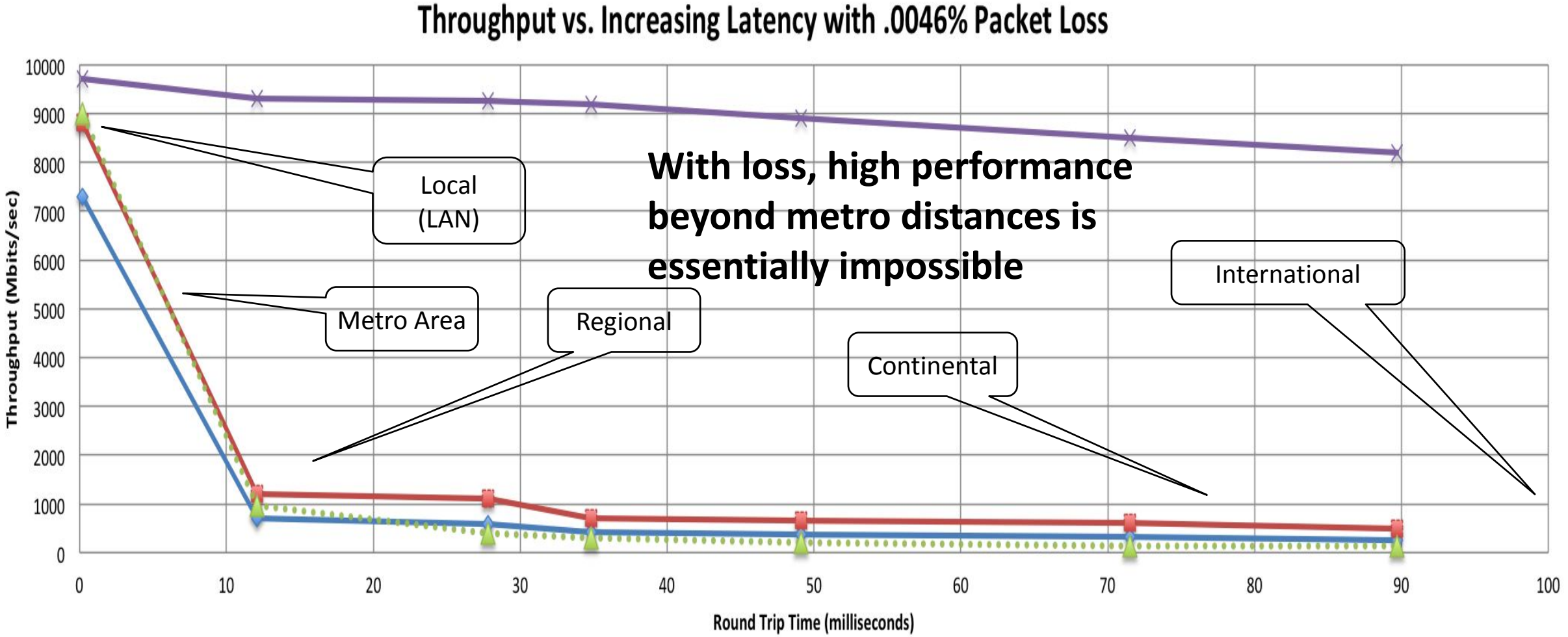
EPOC
Engagement and Performance Operations Center

The

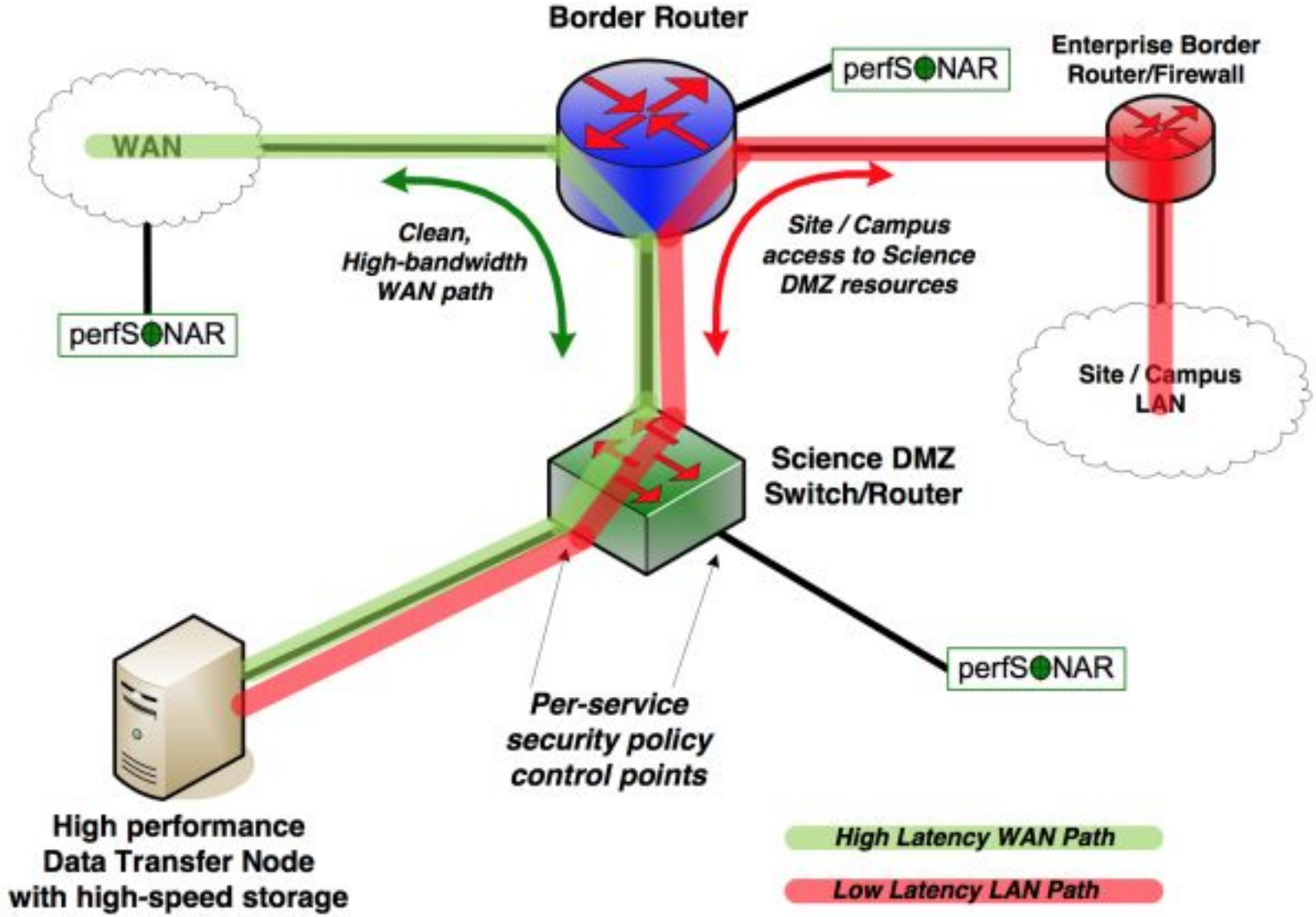
Consi

- "Fri
  -
  -
  -
  -
- Dec
  -
  -
- Per
  -
- Eng

A small amount of packet loss makes a huge difference in TCP performance

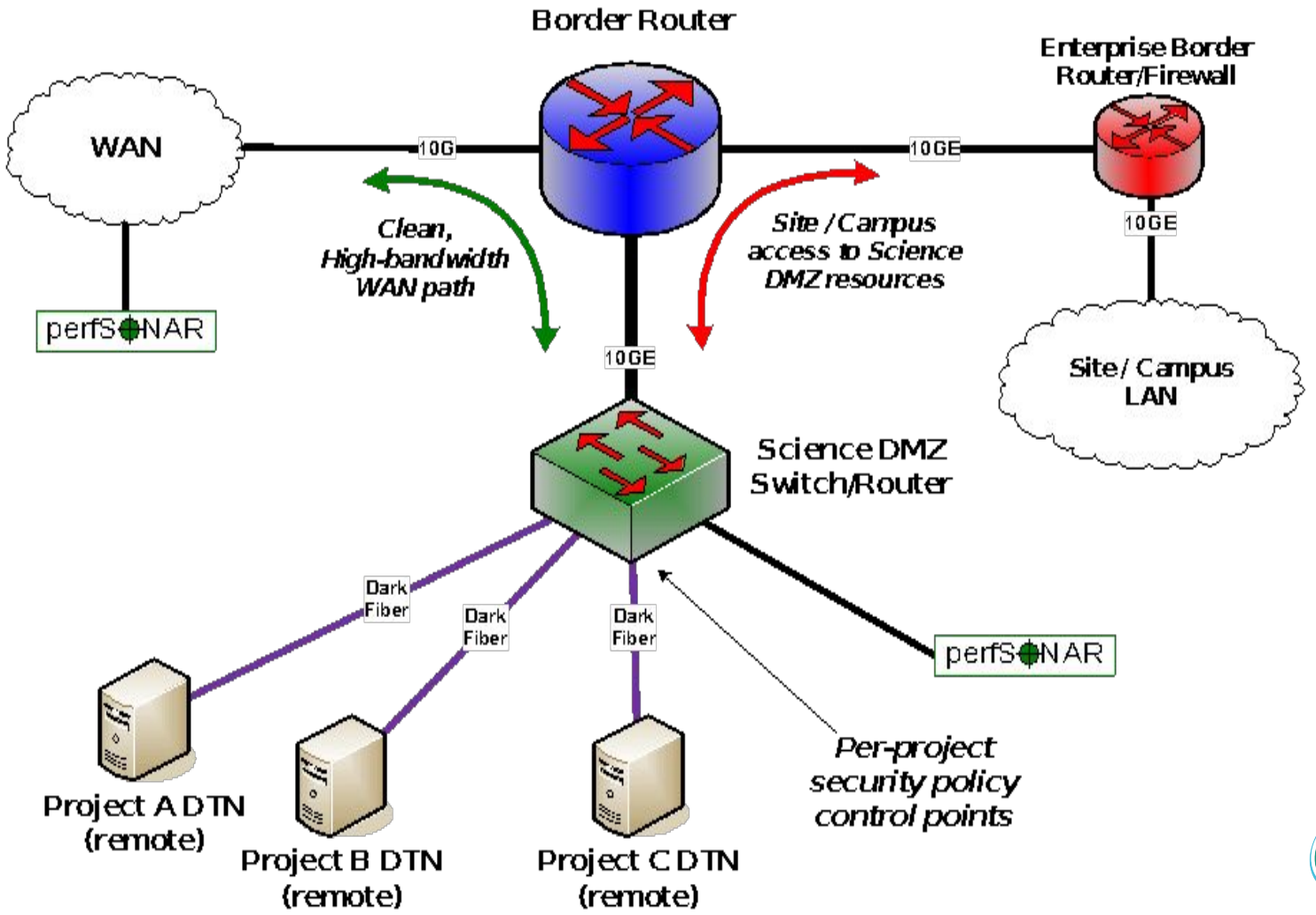


A better approach: simple Science DMZ



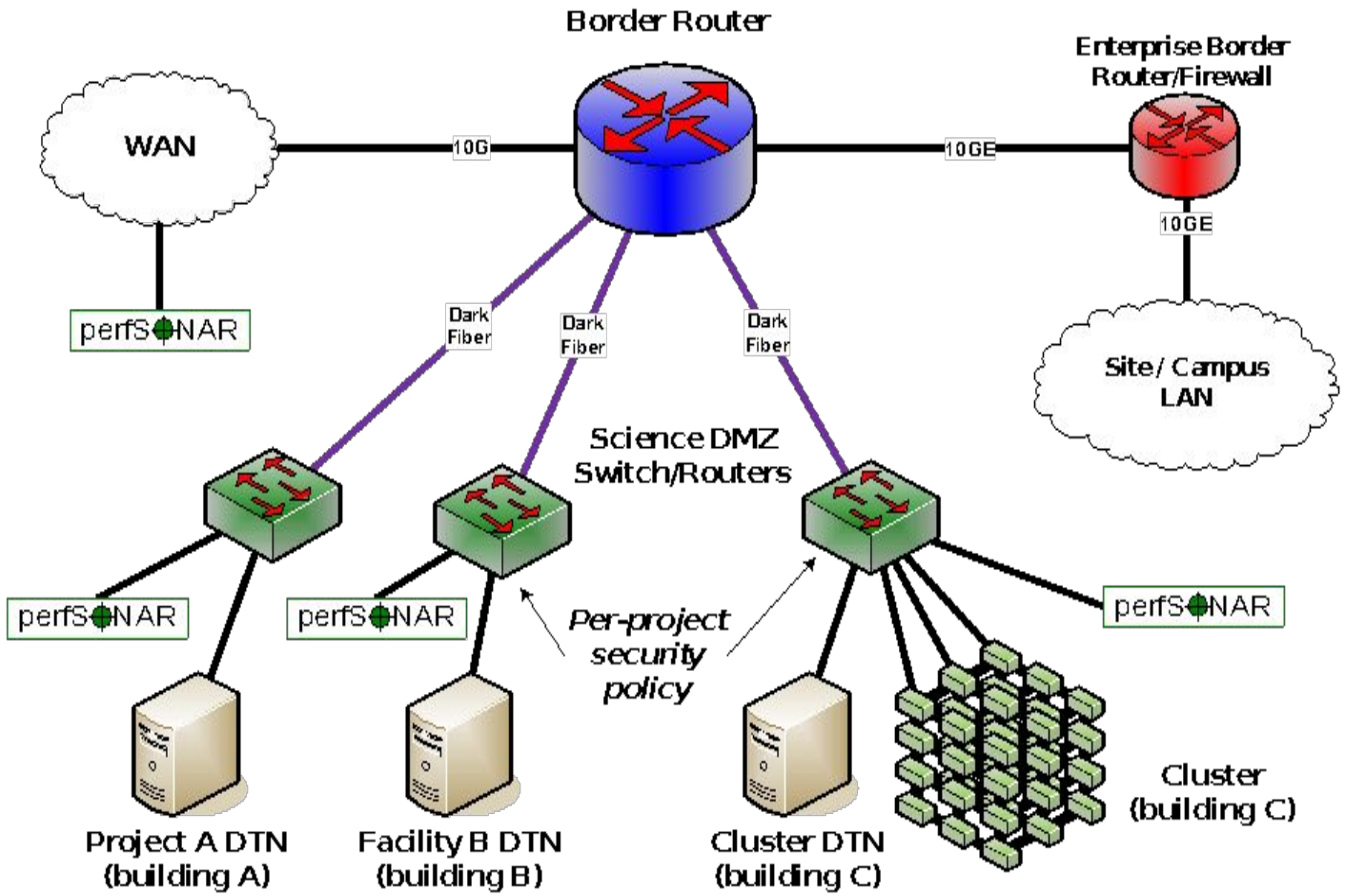
EPOC
Engagement and Performance Operations Center

Distributed Science DMZ – Dark Fiber



EPOC
Engagement and Performance Operations Center

Multiple Science DMZs – Dark Fiber to Dedicated Switches



EPOC
Engagement and Performance Operations Center

Equipment – Routers and Switches

- Requirements for Science DMZ gear are different than the enterprise
  - No need to go for the kitchen sink list of services
  - A Science DMZ box only needs to do a few things, but do them well
  - Support for the latest LAN integration magic with your Windows Active Directory environment is probably not super-important
  - **A clean architecture is important**
    - How fast can a single flow go?
    - Are there any components that go slower than interface wire speed?
- There is a temptation to go cheap
  - Hey, it only needs to do a few things, right?
  - You typically don’t get what you don’t pay for
    - (You sometimes don’t get what you pay for either)

EPOC
Engagement and Performance Operations Center

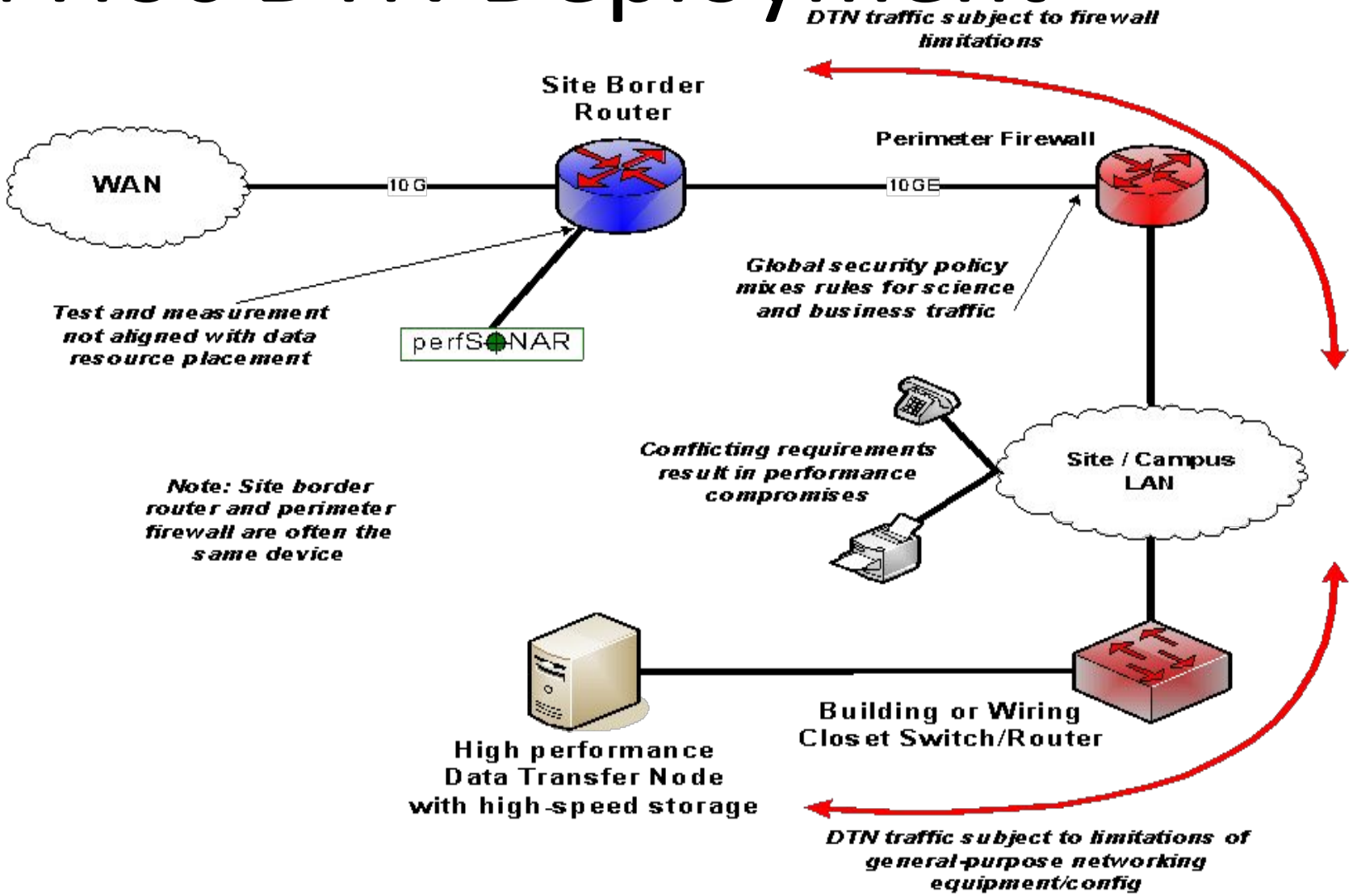
Legacy Method: Ad Hoc DTN Deployment

- This is often what gets tried first
- Data transfer node deployed where the owner has space
  - This is often the easiest thing to do at the time
  - Straightforward to turn on, hard to achieve performance

- If lucky, perfSONAR is at the border
  - This is a good start
  - Need a second one next to the DTN
- Entire LAN path has to be sized for data flows
- Entire LAN path is part of any troubleshooting exercise
- This usually fails to provide the necessary performance.

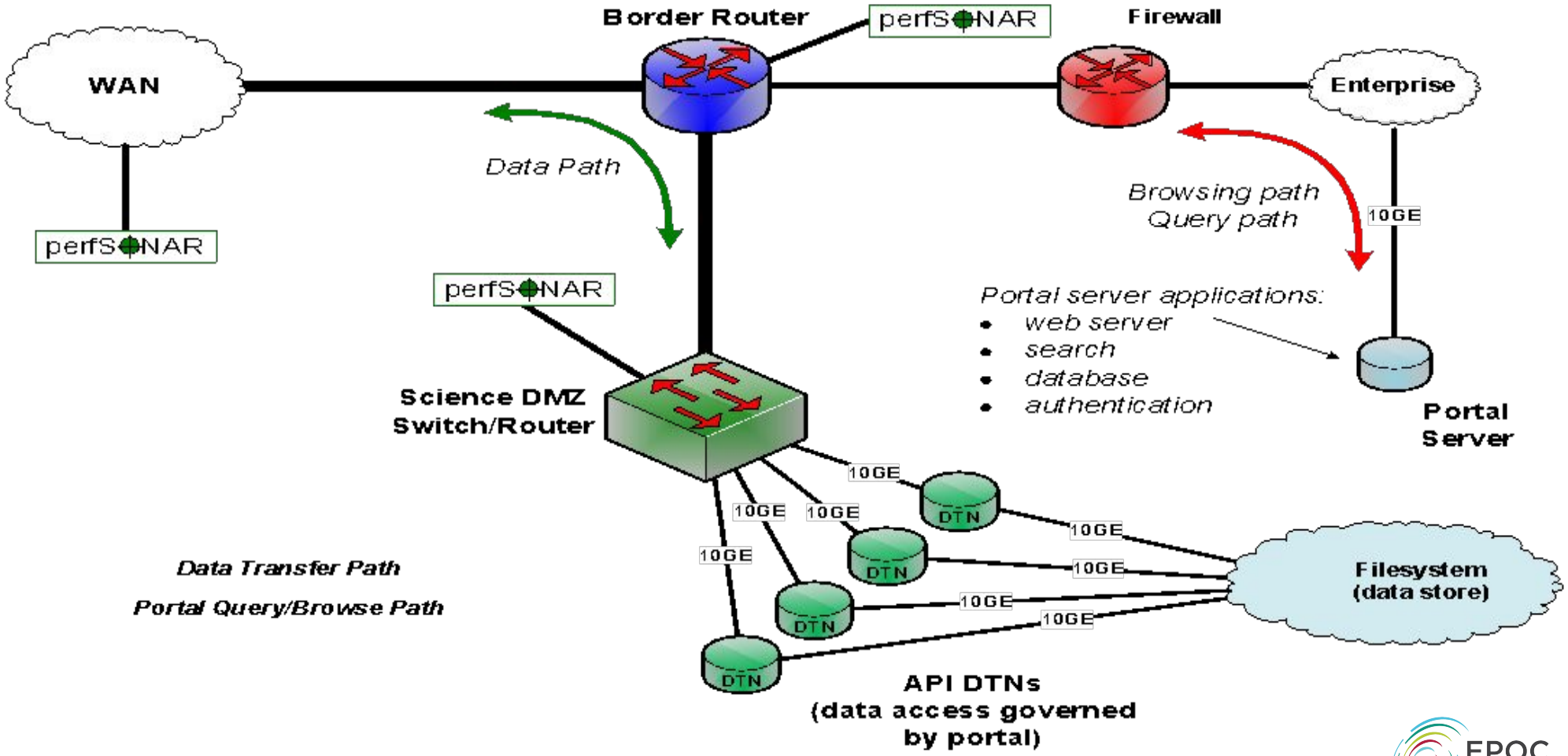
EPOC
Engagement and Performance Operations Center

Ad Hoc DTN Deployment



EPOC
Engagement and Performance Operations Center

Next-Generation Portal Leverages Science DMZ, DTN pool, Central Data Store



https://peerj.com/articles/cs-144/
https://docs.globus.org/guides/recipes/modern-research-data-portal/

EPOC
Engagement and Performance Operations Center

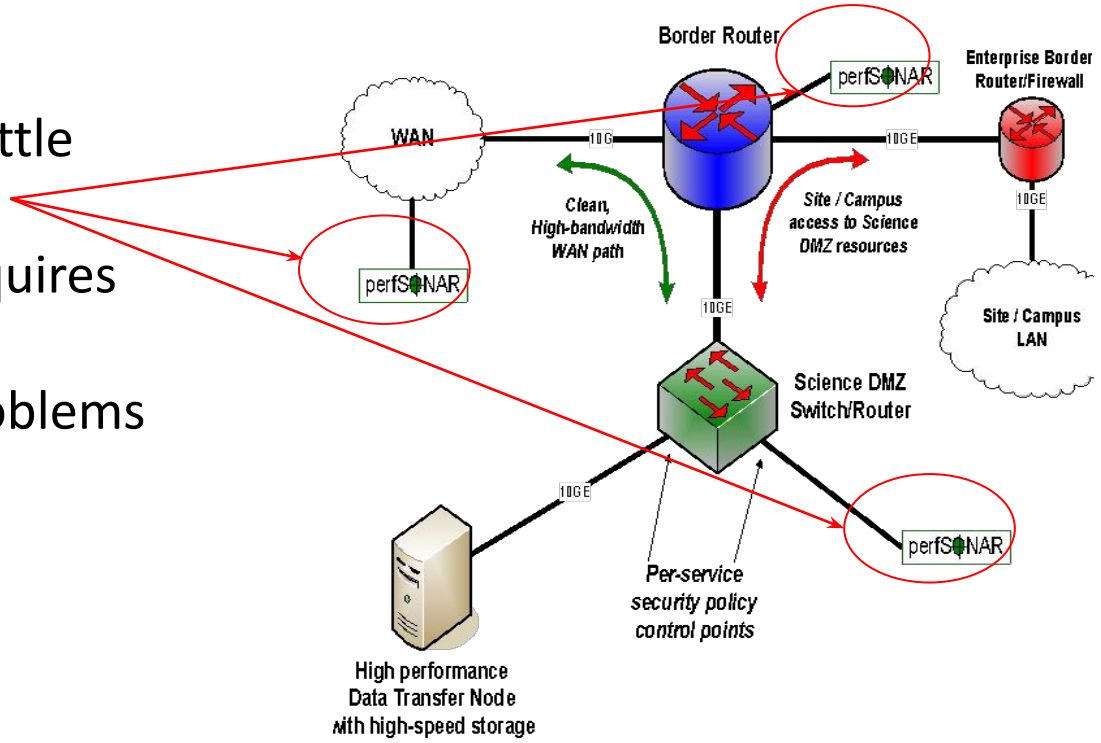
Test and Measurement – Keeping the Network Clean

- The wide area network, the Science DMZ, and all its systems can be functioning perfectly
- Eventually something is going to break
  - Networks and systems are built with many, many components
  - Sometimes things just break – this is why we buy support contracts
- Other problems arise as well – bugs, mistakes, whatever
- We must be able to find and fix problems when they occur
- Why is this so important? Because we use TCP!

EPOC
Engagement and Performance Operations Center

perfSONAR

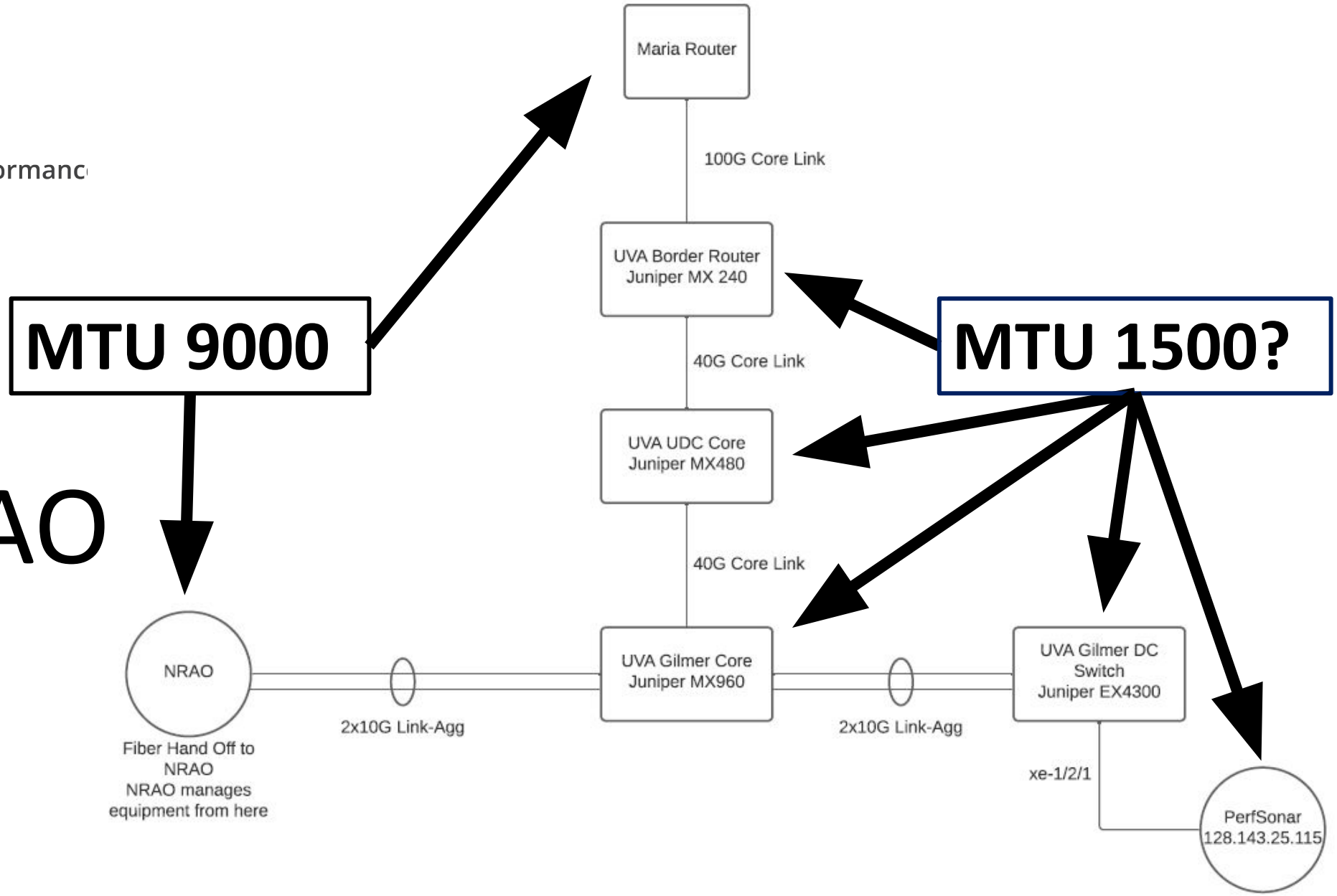
- Network diagrams throughout these materials have little perfSONAR boxes everywhere
  - The reason for this is that consistent behavior requires correctness
  - Correctness requires the ability to find and fix problems
    - ***You can't fix what you can't find***
    - ***You can't find what you can't see***
    - ***perfSONAR lets you see***
- Especially important when deploying high performance services
  - If there is a problem with the infrastructure, need to fix it
  - If the problem is not with your stuff, need to prove it
    - Many players in an end to end path
    - Ability to show correct behavior aids in problem localization



EPOC
Engagement and Performance Operations Center

EPOC
Engagement and Performanc
Operations Center

UVA/NRAO Network




MakeAGIF.com

Yeah, yeah, but what about performance??

Before a 1TB transfer would take ~243 days:
pscheduler task throughput --source cpt-chpc-10g.perfsonar.ac.za --dest perfsonar-10.cv.nrao.edu
Summary

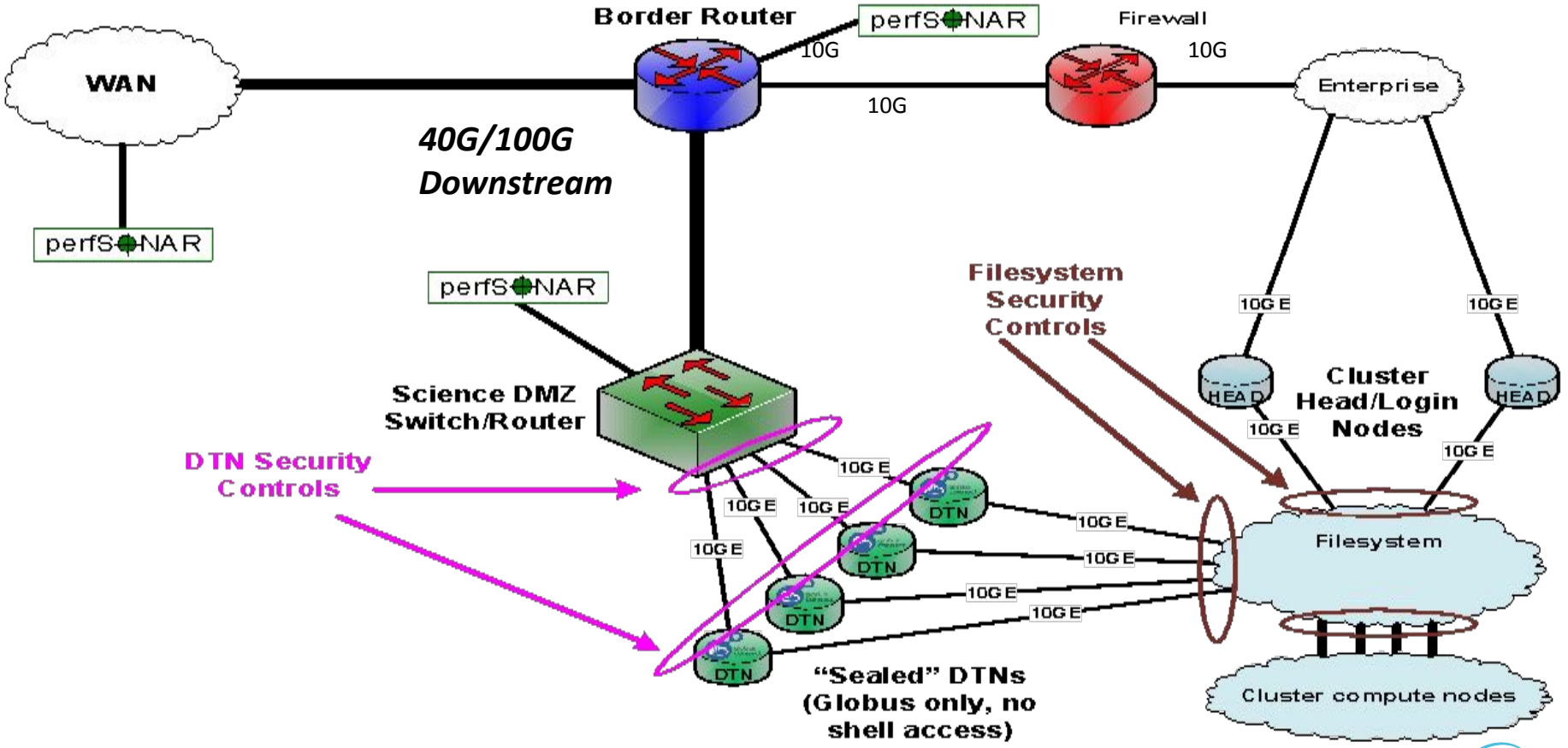
Interval	Throughput	Retransmits	Receiver Throughput
0.0 - 10.0	380.37 Kbps	58	108.18 Kbps

After a 1TB transfer would take ~49 minutes:
pscheduler task throughput -t 30 --source cpt-chpc-10g.perfsonar.ac.za --dest perfsonar-10.cv.nrao.edu
Summary

Interval	Throughput	Retransmits	Receiver Throughput
0.0 - 30.0	2.67 Gbps	0	2.62 Gbps

EPOC
Engagement and Performance Operations Center

Using Globus to test Data Mobility performance



EPOC
Engagement and Performance Operations Center

Software – Data Transfer

- Using the right data transfer tool is ***STILL*** very important
- Sample Results: Berkeley, CA to Argonne, IL (near Chicago) RTT = 53 ms, network capacity = 10Gbps.

Tool	Throughput
scp, rsync	330 Mbps
wget, Globus, FDT, 1 stream	6 Gbps
Globus and FDT, 4 streams	8 Gbps (disk limited)

- Notes
  - scp is 24x slower than Globus on this path!!
  - to get more than 1 Gbps (125 MB/s) disk to disk requires RAID array.
  - Assume host TCP buffers are set correctly for the RTT

EPOC
Engagement and Performance Operations Center

Data Transfer Performance and Expectations

Data set size	1 Minute	5 Minutes	20 Minutes	1 Hour
10PB	1,333.33 Tbps	266.67 Tbps	66.67 Tbps	22.22 Tbps
1PB	133.33 Tbps	26.67 Tbps	6.67 Tbps	2.22 Tbps
100TB	13.33 Tbps	2.67 Tbps	666.67 Gbps	222.22 Gbps
10TB > 100Gbps	1.33 Tbps	266.67 Gbps	66.67 Gbps	22.22 Gbps
1TB	133.33 Gbps	26.67 Gbps	6.67 Gbps	2.22 Gbps
100GB 100Gbps	13.33 Gbps	2.67 Gbps	666.67 Mbps	222.22 Mbps
10GB < 10Gbps	1.33 Gbps	266.67 Mbps	66.67 Mbps	22.22 Mbps
1GB	133.33 Mbps	26.67 Mbps	6.67 Mbps	2.22 Mbps
100MB < 100Mbps	13.33 Mbps	2.67 Mbps	0.67 Mbps	0.22 Mbps
	Time to transfer			

This table available at:

http://fasterdata.es.net/fasterdata-home/requirements-and-expectations/

EPOC
Engagement and Performance Operations Center

Data Transfer Scorecard with Rates by Audience

Host Transfer Rates	⅛ PetaScale (Minimum)	⅓ PetaScale	½ PetaScale		PetaScale: 1 PB/wk	PetaScale: 1 PB/day
	10G Capable DTN				10xG, 25G, 40G, 100G DTNs	
Data Transfer Rate/Volume (Researcher)	1 TB/hr	2 TB/hr	3 TB/hr		5.95 TB/hr	41.67 TB/hr
Network Transfer Rate (Network Admin)	2.22 Gb/s	4.44 Gb/s	6.67 Gb/s		13.23 Gb/s	92.59 Gb/s
Storage Transfer Rate (Sys/Storage Admin)	277.78 MB/s	555.54 MB/s	833.33 MB/s		1.65 GB/s	11.57 GB/s

A benchmark table is provided to gauge data architecture performance, which can vary depending on number of files, folders, size of files, distance between sites, CI performance (network, server, disk/filesystem), as well as data transfer tool.

EPOC
Engagement and Performance Operations Center

To Reiterate:

- Data movement is hard to get right.
- Globus transfer can overcome some network issues due to parallel transfers, but you still need a clean network to get
- Lots of moving parts in data movement -
  - Software, Servers, Networks, and People
  - Check your network and system MTU settings
  - Verify your routes
- Testing will reveal that it may not be ideal
- Testing will also motivate you to make it ideal
- Shared experience around the community –
  - Lift all the boats, share all the knowledge, etc.

EPOC
Engagement and Performance Operations Center

Questions?

- EPOC Helpdesk (send in anything you want):
  - epoc@tacc.utexas.edu
  - For NSF, NIH, NOAA, USDA, etc..
- For DOE Science Engagement
  - engage@es.net

EPOC
Engagement and Performance Operations Center

Enhancing Data Transfer Performance with DTNs and the Science DMZ

Ken Miller
ken@es.net
ESnet / Lawrence Berkeley National Laboratory

Workshop on Science DMZ and P4-DPDK
Online
August 6-8, 2024

National Science Foundation Award #2328479