EPOC
Engagement and Performance Operations Center

UNIVERSITY OF South Carolina

What's a Science DEE EM ZEE?

Jason Zurawski
LBNL / ESnet
zurawski@es.net

Jorge Crichigno, Elie Kfoury
University of South Carolina
jcrichigno@cec.sc.edu, ekfoury@email.sc.edu

2022 NSF CC* PI Workshop
Minneapolis MN,
Sept 19th, 2022

ESnet
ENERGY SCIENCES NETWORK

INDIANA UNIVERSITY

National Science Foundation Award #1826994

Outline

- ***Introduction***
- Solution Space
- Putting it Together
- Conclusions / QA

UNIVERSITY OF South Carolina

EPOC Engagement and Performance Operations Center

Measurable Outcomes

- How to prepare for research use of technology on a campus?
  - "Build CI" – but is this all?
  - ***Improve scientific outcomes in some measurable way***
- Things to consider (pre, during, post):
  - Has it/will it all work as expected? (e.g. more than plugging in wires)
  - Will we ***all*** be satisfied at the end? (researchers are the survey population, not just the IT org …)
  - How do you know when you/we are done? Are you ever done?
- Think of this set of content as a reset – we don't want to build IT for the sake of building IT
  - Tie things back to the user/use cases, and be sensible about the design, installation, and operation

UNIVERSITY OF South Carolina

EPOC Engagement and Performance Operations Center

Where do we go from here?

1. WHO and WHY are the most important questions
   - Know the users, know the use cases. Those will guide any technical solution
2. WHAT and HOW go together:
   - Figuring out the technology that will help without causing non-productive disruptions
   - Being able to sensibly design, implement, and operate

*We **have** to address item 1, then we will dive into item 2*

UNIVERSITY OF South Carolina

EPOC Engagement and Performance Operations Center

Outline

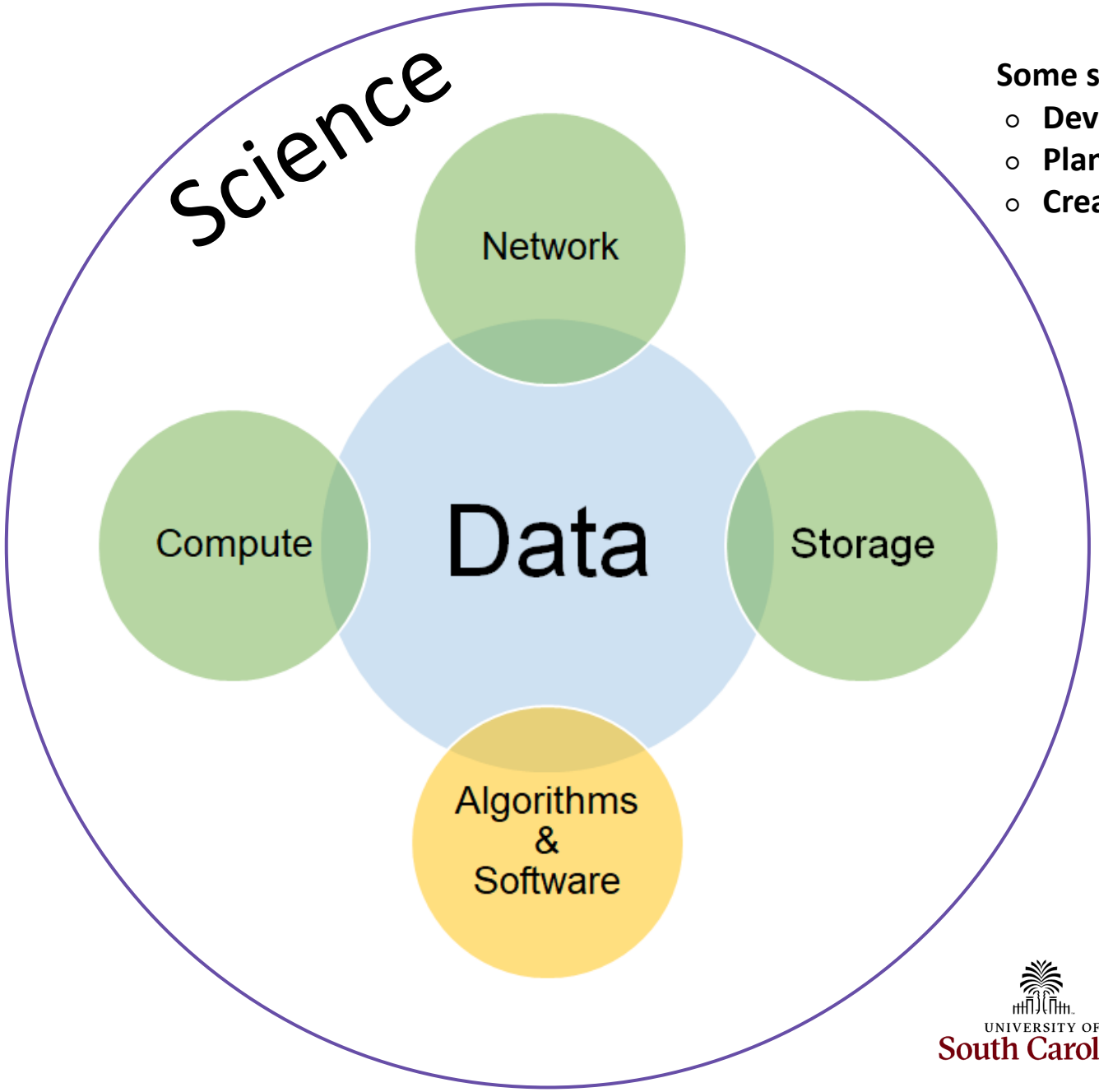
- Introduction
- ***Solution Space***
  1. ***Understanding the Solution Space (Users, Use Cases, Long Term Impacts)***
  2. Preliminaries (e.g. Network Protocols 101)
  3. Architecture & Design
  4. Security and Policy
- Putting it Together
- Conclusions / QA

UNIVERSITY OF South Carolina

EPOC Engagement and Performance Operations Center

Common Theme / New Mindset

- We aren't building a "Network Architecture", we want a "Data Architecture"
  - A lot of the items that will be thrown at you transcend the traditional network space.
- To get there:
  - Understand the data pipeline for your target user/use case – cradle to retirement home
  - This implies all the things:
    - Creation
    - Usage
    - Transfer/Share
    - Curation



Some specific issues for networks are

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

UNIVERSITY OF South Carolina

EPOC Engagement and Performance Operations Center

Common Theme / New Mindset

- What you build must be
  - ***Usable*** – if this becomes a ‘walled garden’, what’s the point? Make it such that people can be easily onboarded and integrated.
  - ***Defensible*** – it is not, nor should it be, the wild west. Control the users and use cases, but don’t impact the usage.
  - ***Scalable*** – as demand grows. Think cornfields and baseball diamonds.
  - ***a institutional capability / source of pride*** – this is something that will draw more users / research dollars if created/marketed/operated correctly. Treat it as such.

UNIVERSITY OF South Carolina

EPOC Engagement and Performance Operations Center

Outline

- Introduction
- ***Solution Space***
  1. Understanding the Solution Space (Users, Use Cases, Long Term Impacts)
  2. ***Preliminaries (e.g. Network Protocols 101)***
  3. Architecture & Design
  4. Security and Policy
- Putting it Together
- Conclusions / QA

UNIVERSITY OF South Carolina

EPOC Engagement and Performance Operations Center

Data Movement / TCP Background

- The data mobility performance requirements for data intensive science are beyond what can typically be achieved using traditional methods
  - Default host configurations (TCP, filesystems, NICs)
  - Converged network architectures designed for commodity traffic
  - Conventional security tools and policies
  - Legacy data transfer tools (e.g. SCP)
  - Wait-for-trouble-ticket operational models for network performance

UNIVERSITY OF South Carolina

EPOC Engagement and Performance Operations Center

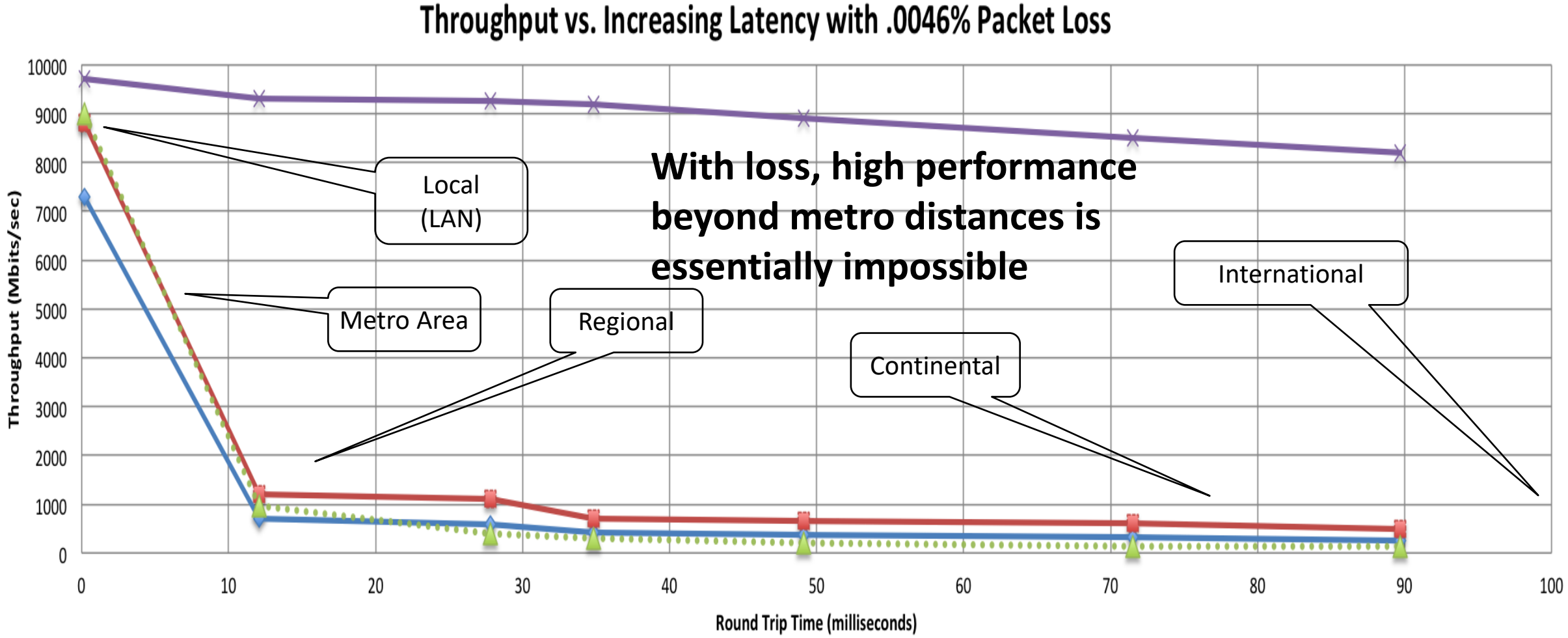
TCP – Ubiquitous and Fragile

- Networks provide connectivity between hosts – how do hosts see the network?
  - From an application's perspective, the interface to "the other end" is a socket
  - Communication is between applications – mostly over TCP
  - ***Congestion*** dictates performance – back off when danger is sensed to preserve/protect resources
- TCP – the fragile workhorse
  - TCP is (for very good reasons) timid – ***packet loss*** is interpreted as congestion
  - Packet loss in conjunction with latency is a performance killer
  - Like it or not, TCP is used for the vast majority of data transfer applications (more than 95% of ESnet traffic is TCP)

UNIVERSITY OF South Carolina

EPOC Engagement and Performance Operations Center

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



Data Movement / TCP Background

- The Science DMZ model describes a performance-based approach
  - Dedicated infrastructure for wide-area data transfer
    - Well-configured data transfer hosts with modern tools
    - Capable network devices
    - High-performance data path which does not traverse commodity LAN
  - Proactive operational models that enable performance
    - Well-deployed test and measurement tools (perfSONAR)
    - Periodic testing to locate issues instead of waiting for users to complain
  - Security posture well-matched to high-performance science applications

UNIVERSITY OF South Carolina

EPOC Engagement and Performance Operations Center

The

Consi

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

Outline

- Introduction
- ***Solution Space***
  1. Understanding the Solution Space (Users, Use Cases, Long Term Impacts)
  2. Preliminaries (e.g. Network Protocols 101)
  3. ***Architecture & Design***
  4. Security and Policy
- Putting it Together
- Conclusions / QA

UNIVERSITY OF South Carolina

EPOC Engagement and Performance Operations Center

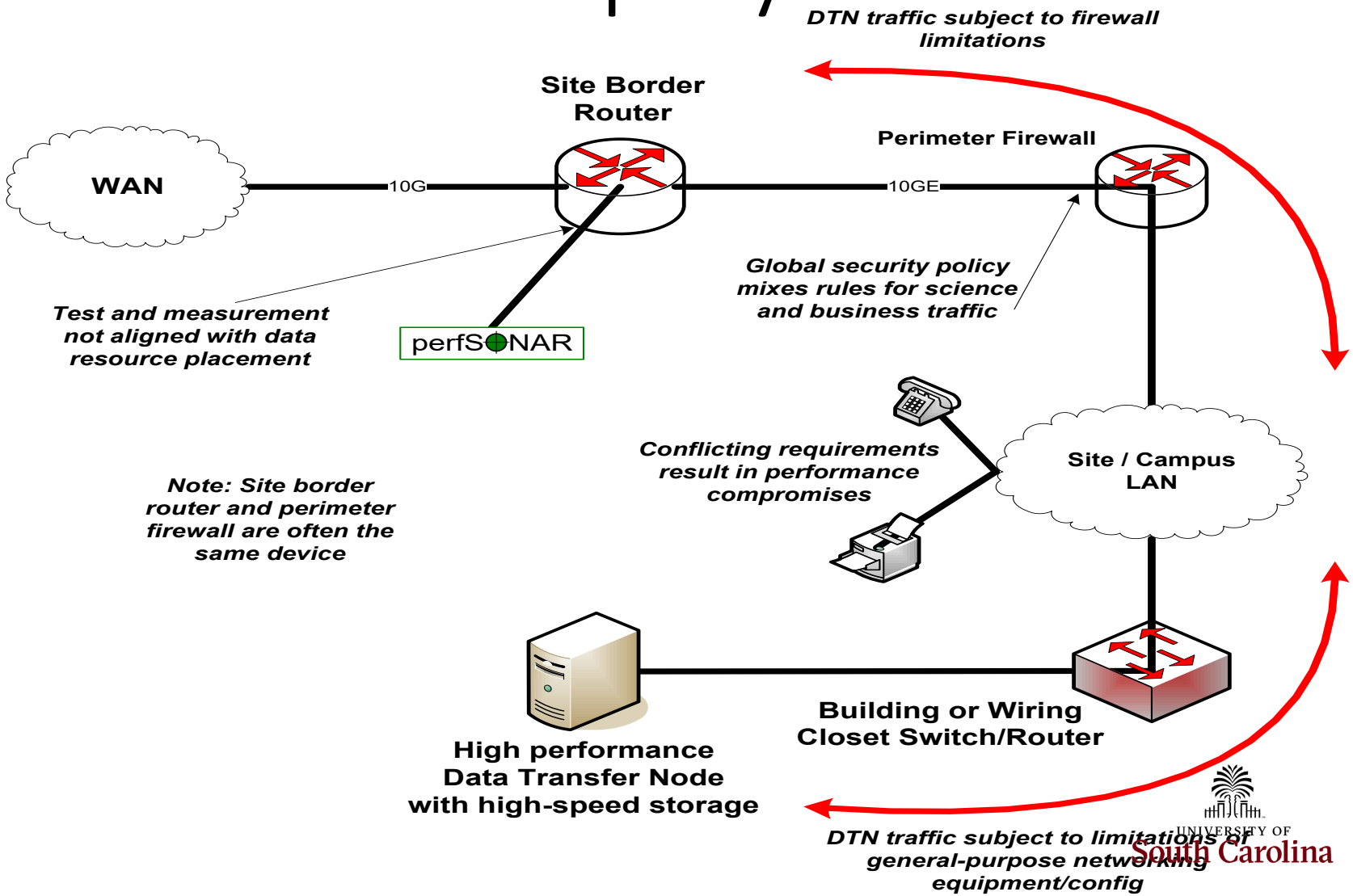
Science DMZ Takes Many Forms

- There are a lot of ways to combine these things – it all depends on what you need to do
  - Small installation for a project or two
  - Facility inside a larger institution
  - Institutional capability serving multiple departments/divisions
  - Science capability that consumes a majority of the infrastructure
- Some of these are straightforward, others are less obvious
- Key point of concentration: eliminate sources of packet loss / packet friction

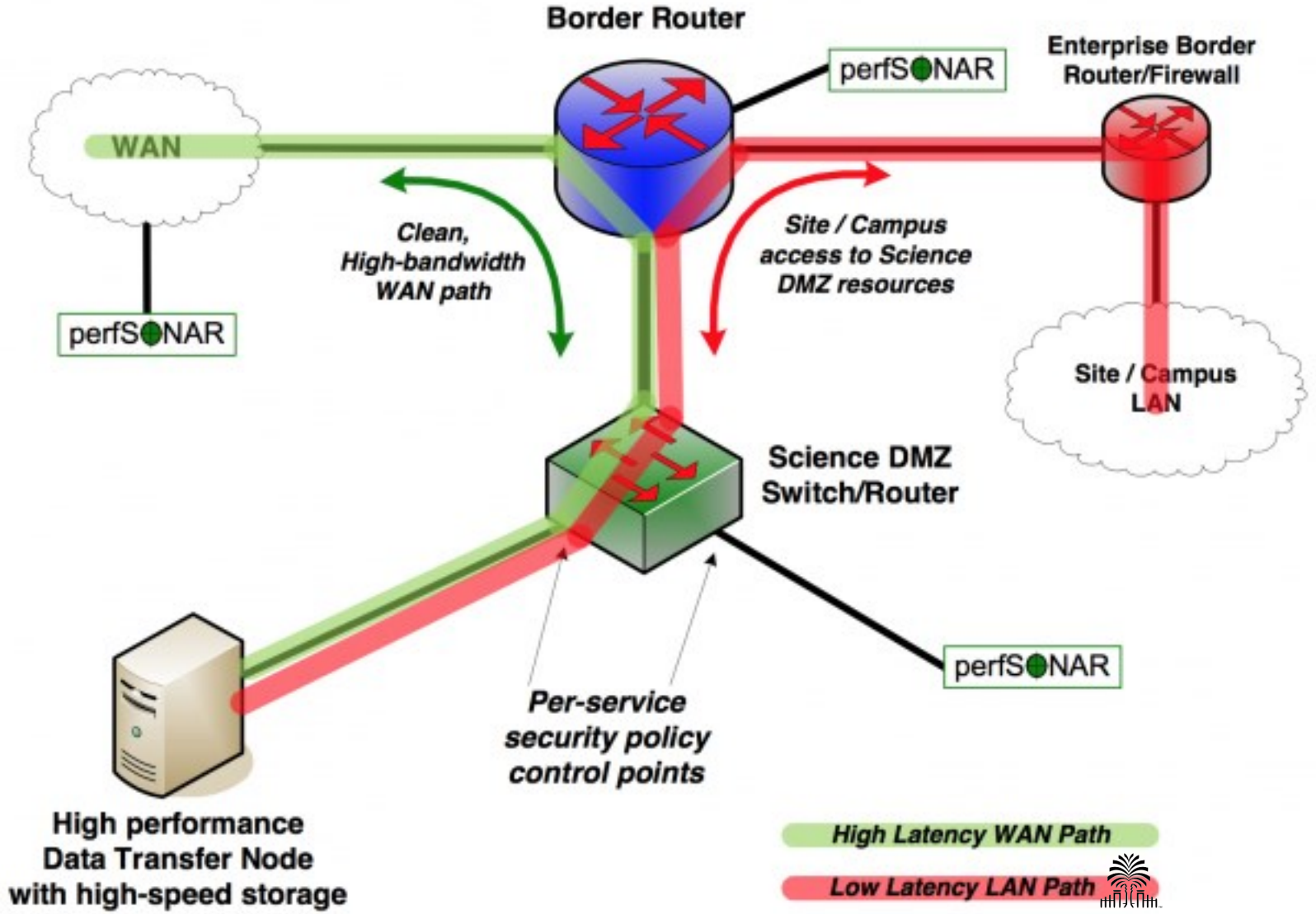
UNIVERSITY OF South Carolina

EPOC Engagement and Performance Operations Center

Ad Hoc DTN Deployment



A better approach: simple Science DMZ

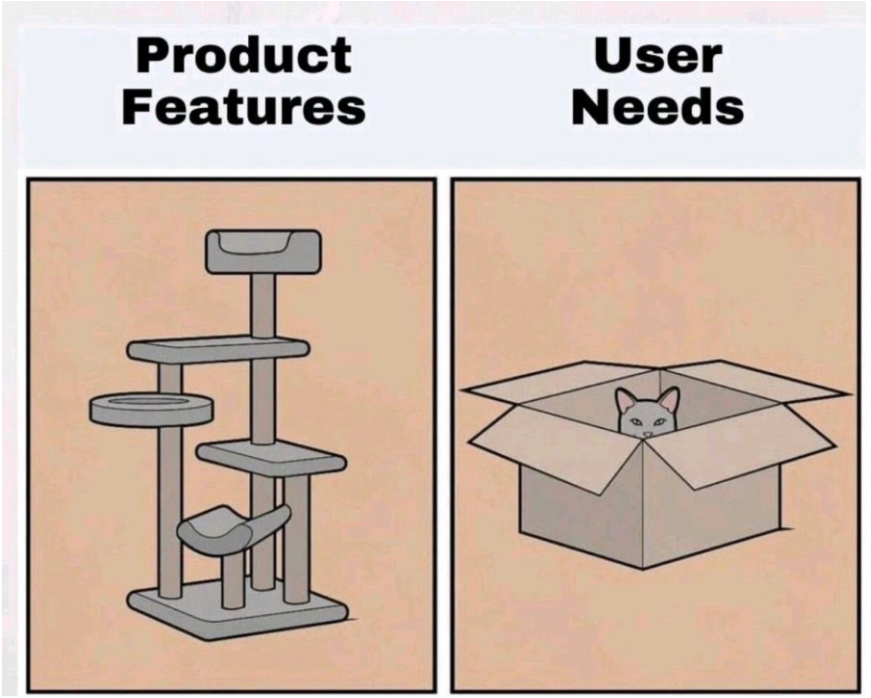


UNIVERSITY OF South Carolina

EPOC Engagement and Performance Operations Center

Design Thoughts – E.g. where does the capability live?

- There are two main schools of thought when going after the CC* design area:
  1. Create an enclave for the science use cases designed specifically to facilitate them. Move things in 1:1
  2. Drop in replacement of a network such that we can capture a bunch of use cases with a big net (*and not dig in too deeply*)

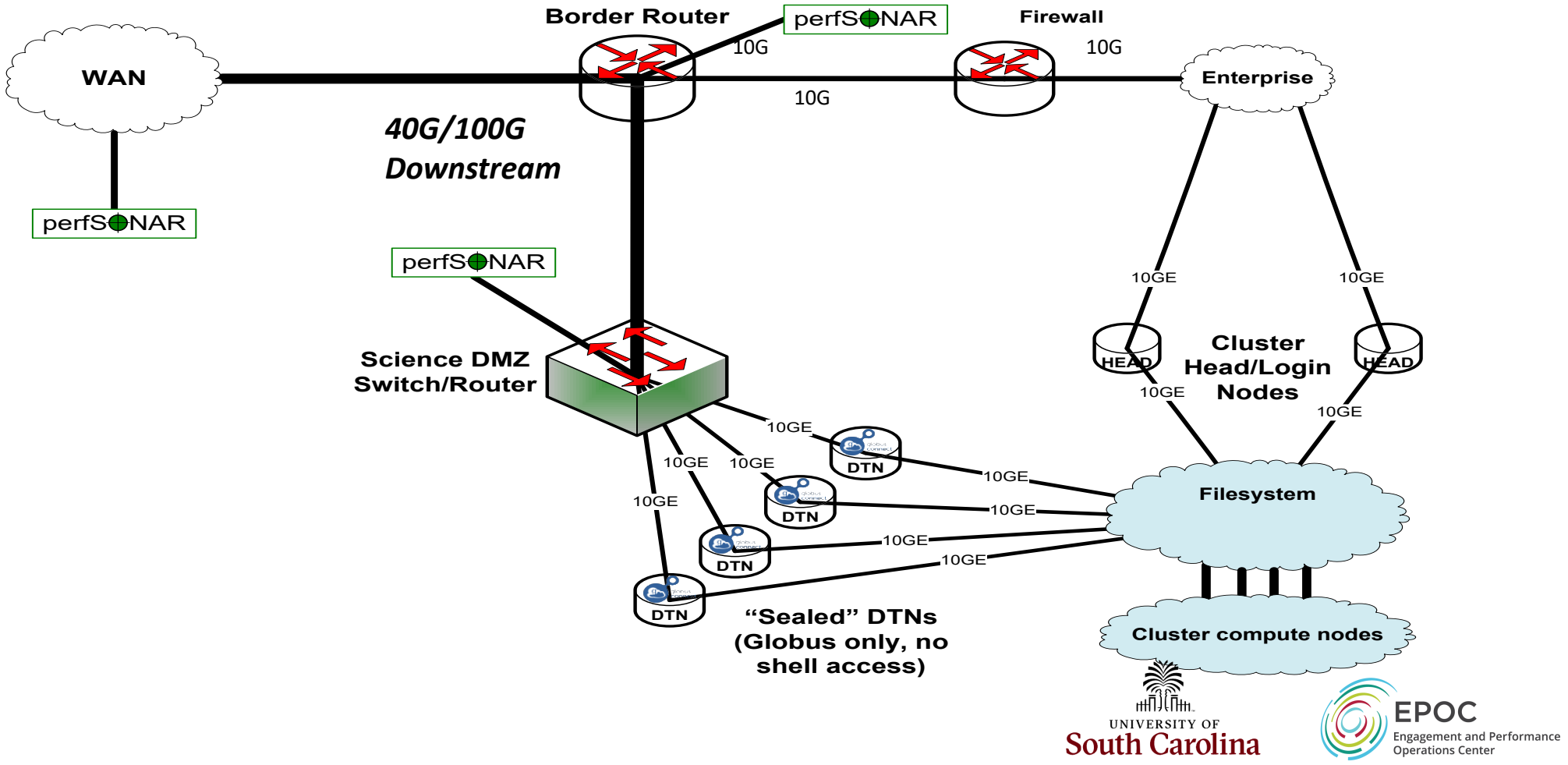


- A strong proposal adopts a stance closer to 1 – its important to know use cases. Its also important to justify expense of equipment to a need
- Number 2 is easier to accomplish (e.g. design, potentially to operate) since its just a 'bigger/better' network that could lift all boats.

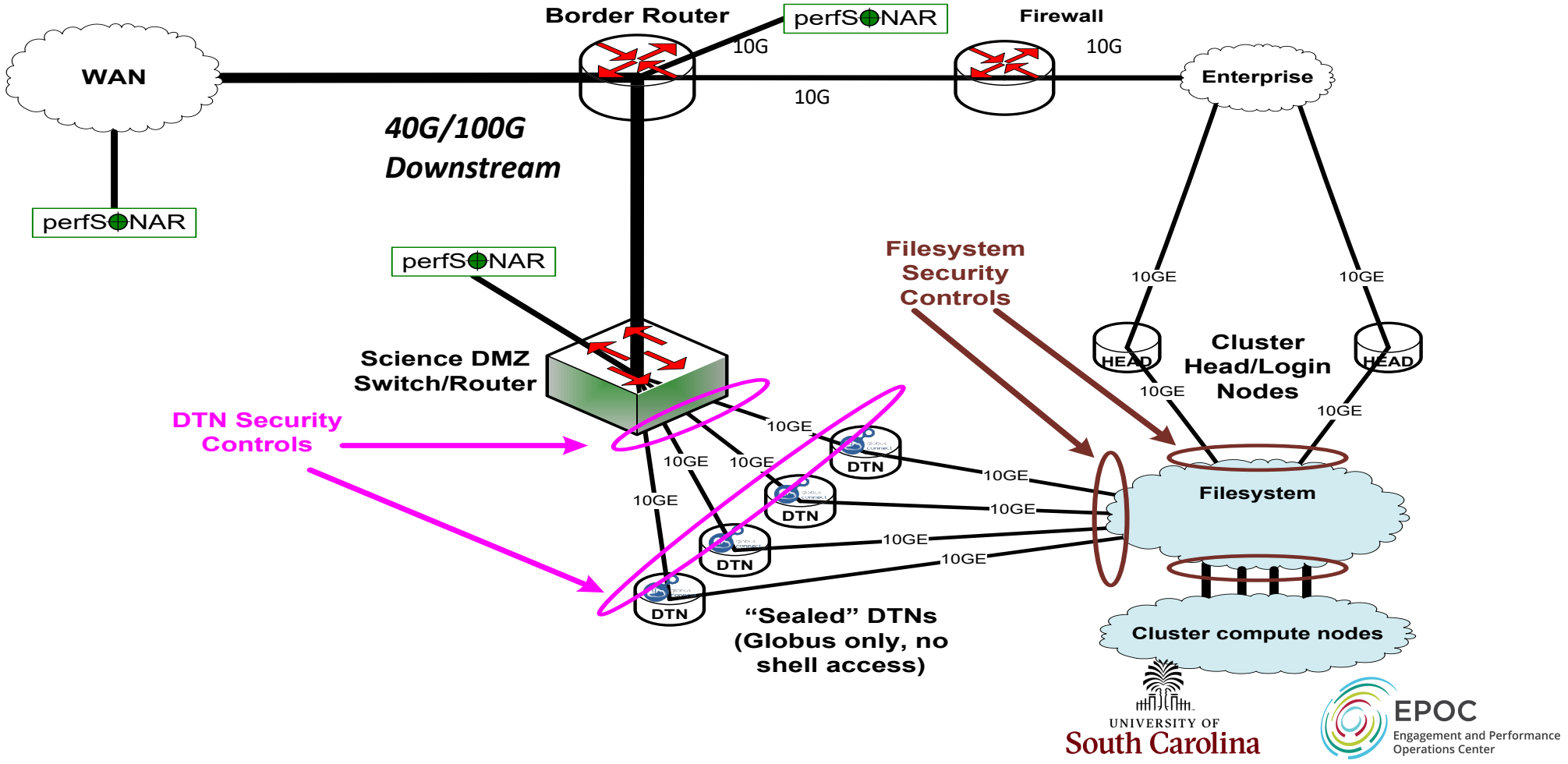
UNIVERSITY OF South Carolina

EPOC Engagement and Performance Operations Center

Solution Space – Data Mobility



Solution Space – Data Mobility



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)

UNIVERSITY OF South Carolina

EPOC Engagement and Performance Operations Center

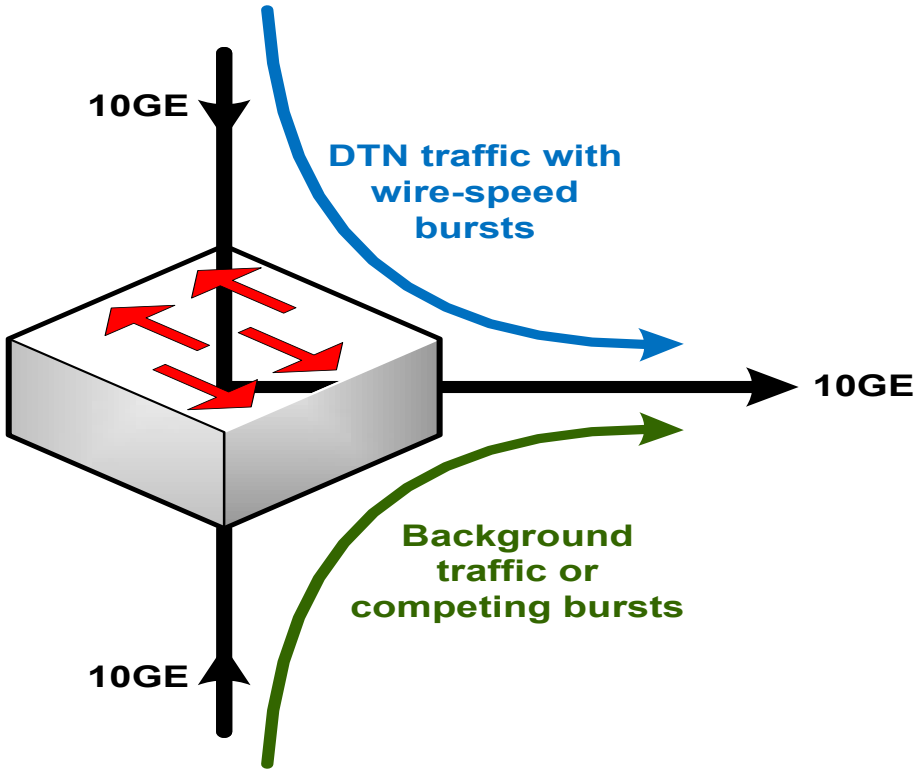
Common Circumstance: Multiple Ingress Data Flows, Common Egress

Hosts will typically send packets at the speed of their interface (1G, 10G, etc.)

- Instantaneous rate, not average rate
- If TCP has window available and data to send, host sends until there is either no data or no window

Hosts moving big data (e.g. DTNs) can send large bursts of back-to-back packets

- This is true even if the average rate as measured over seconds is slower (e.g. 4Gbps)
- On microsecond time scales, there is often congestion
- Router or switch must queue packets or drop them



UNIVERSITY OF South Carolina

EPOC Engagement and Performance Operations Center

All About That Buffer (No Cut Through)

- ***The Bandwidth Delay Product***
  - The amount of "in flight" data for a TCP connection (BDP = bandwidth * round trip time)
- Example: 10Gb/s cross country, ~100ms
  - 10,000,000,000 b/s * .1 s = 1,000,000,000 bits
  - 1,000,000,000 / 8 = 125,000,000 bytes
  - 125,000,000 bytes / (1024*1024) ~ ***125MB***
- Ignore the math aspect: its making sure there is memory to catch and send packets
  - *At **ALL** hops*
  - As the speed increases, there are more packets.
  - If there is not memory, we drop them, and that makes TCP react, and the user sad.

UNIVERSITY OF South Carolina

EPOC Engagement and Performance Operations Center

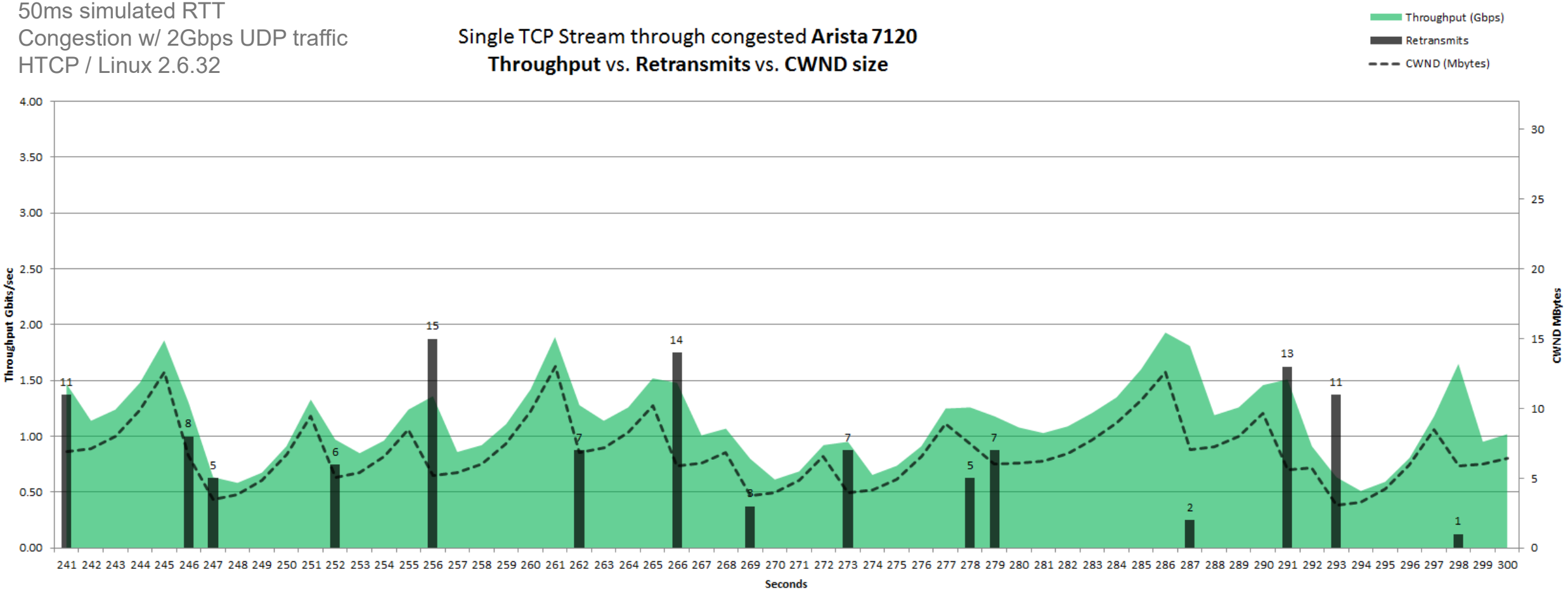
All About That Buffer (No Cut Through)

- Buffering isn't as important on the LAN (this is why you are normally pressured to buy 'cut through' devices)
  - Change the math to make the Latency 1ms and the expectation 10Gbps = ***1.25MB***
  - 'Cut through' and low latency switches are designed for the data center, and can handle typical data center loads that don't require buffering (e.g. same to same speeds, destinations within the broadcast domain)

- Buffering ****MATTERS**** for WAN Transfers
  - Placing something with inadequate buffering in the path reduces the buffer for the entire path. E.g. if you have an expectation of 10Gbps over 100ms – don't place a 12MB buffer anywhere in there – your reality is now ~10x less than it was before (e.g. 10Gbps @ 10ms or 1Gbps @ 100ms)

TCP's Congestion Control

50ms simulated RTT
Congestion w/ 2Gbps UDP traffic
HTCP / Linux 2.6.32



Slide from Michael Smitasin, LBLnet

Outline

- Introduction
- ***Solution Space***
    1. Understanding the Solution Space (Users, Use Cases, Long Term Impacts)
    2. Preliminaries (e.g. Network Protocols 101)
    3. Architecture & Design
    4. ***Security and Policy***
- Putting it Together
- Conclusions / QA

UNIVERSITY OF South Carolina

EPOC Engagement and Performance Operations Center

Science DMZ Security

- **Goal**: Disentangle security policy and enforcement for science flows from enterprise / business systems
- **Rationale**
  - Science data traffic is simple from a security perspective
  - Narrow application set on Science DMZ
    - Data transfer, data streaming packages
    - No printers, document readers, web browsers, building control systems, financial databases, staff desktops, etc.
  - Security controls that are typically implemented to protect business resources ***routinely*** cause performance problems
- **Separation allows each to be optimized**

UNIVERSITY OF South Carolina

EPOC Engagement and Performance Operations Center

Science DMZ as Security Architecture

- Allows for better segmentation of risks, more granular application of controls to those segmented risks.
  - Limit risk profile for high-performance data transfer applications
  - Apply specific controls to data transfer hosts
  - Avoid including unnecessary risks, unnecessary controls
- Remove degrees of freedom – focus only on what is necessary
  - Easier to secure
  - Easier to achieve performance
  - Easier to troubleshoot

UNIVERSITY OF South Carolina

EPOC Engagement and Performance Operations Center

Network Segmentation

- Think about residence hall networks, business application networks, and the networks that are primarily in research areas:
  - The risk profiles are clearly different
  - It makes sense to segment along these lines
- Your institution may already be doing this for things like HIPAA and PCI-DSS. Why? *Because of the controls!*
- The Science DMZ follows the same concept, from a security perspective.
- Using a Science DMZ to segment research traffic (especially traffic from specialized research instruments) can actually *improve* campus security posture.

UNIVERSITY OF South Carolina

EPOC Engagement and Performance Operations Center

Outline

- Introduction
- Solution Space
- ***Putting it Together***
- Conclusions / QA

Collaboration Within The Organization

- All stakeholders should collaborate on Science DMZ design, policy, and enforcement
- The security people have to be on board
  - Remember: security people already have political cover – it's called the firewall
  - If a host gets compromised, the security officer can say they did their due diligence because there was a firewall in place
  - If the deployment of a Science DMZ is going to jeopardize the job of the security officer, expect pushback
- The Science DMZ is a strategic asset, and should be understood by the strategic thinkers in the organization
  - Changes in security models
  - Changes in operational models
  - Enhanced ability to compete for funding
  - Increased institutional capability – greater science output

UNIVERSITY OF South Carolina

EPOC Engagement and Performance Operations Center

Sensible Usage Policies

- Everything we designed today was meant to be ‘used’ by the scientific community. What does that mean?
  - Do you need to allow them to plug in their old windows laptop directly to the border router?
  - Remember all that work you did to understand use cases? Lets revisit that instead:
- Data mobility (in a nutshell):
  1. I need research data
     a) I can create it or I can retrieve it
  2. Once I have it, I shall analyze it
     a) That could be local, it may not be
  3. After analysis I must do something with it
     a) Maybe I delete it? Who am I kidding … I want to hold it forever!

Sensible Usage Policies

- Define access methods
  - E.g. shared DTN that plugs into storage vs. someone's laptop
  - Tools that can be used
  - People who get accounts
  - Consider: http://fasterdata.es.net/science-dmz/science-dmz-users/
- Define AUP
  - What can/should/will be sent across the infrastructure
  - What happens when something bad occurs
  - How often the AUP is reviewed
  - Consider: http://fasterdata.es.net/science-dmz/usage-policy/
- Define monitoring/measurement/security expectations
  - Let the pros monitor/keep things up to date
  - Let the users just use
- ***BE TRANSPARENT***

UNIVERSITY OF South Carolina

EPOC Engagement and Performance Operations Center

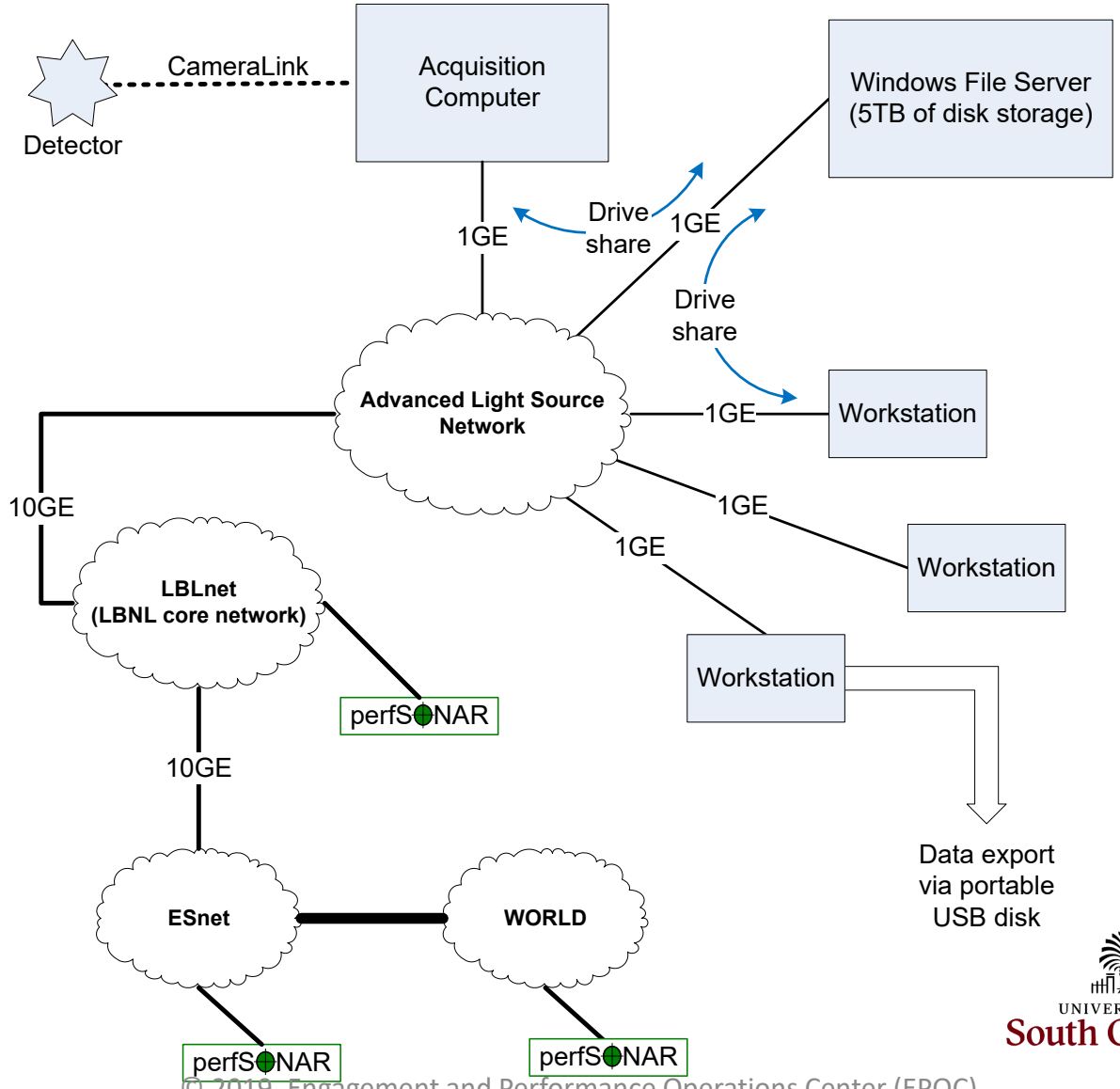
Putting it All Together

- Know the users, know the use cases
- Data Architecture – e.g. support the ingress and egress of information at all layers
  - Network Gear
  - Data Transfer Software / Hardware
  - ***Measurement / Monitoring [LATER TODAY!]***
  - Security Infrastructure
- Facilitate Usage
- Have a good story for long term usage/onboarding/expansion/maintenance

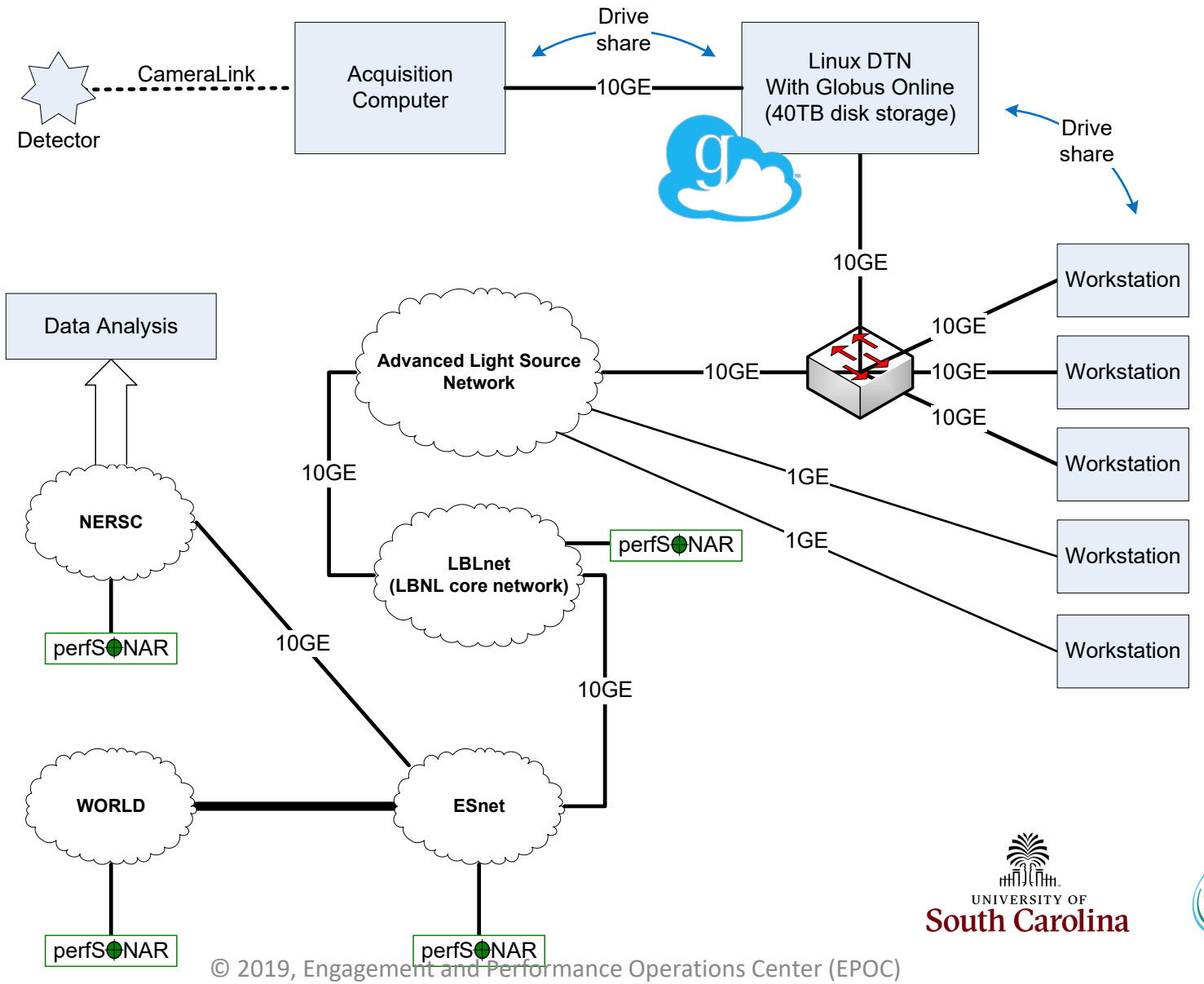
UNIVERSITY OF South Carolina

EPOC Engagement and Performance Operations Center

Science Workflow Consultation



Improved Workflow Infrastructure



EPOC
Engagement and Performance Operations Center

UNIVERSITY OF South Carolina

What's a Science DEE EM ZEE?

Jason Zurawski
LBNL / ESnet
zurawski@es.net

Jorge Crichigno, Elie Kfoury
University of South Carolina
jcrichigno@cec.sc.edu, ekfoury@email.sc.edu

2022 NSF CC* PI Workshop
Minneapolis MN,
Sept 19th, 2022

ESnet
ENERGY SCIENCES NETWORK

INDIANA UNIVERSITY

National Science Foundation Award #1826994