Science DMZ Security

Ken Miller, Jason Zurawski
ken@es.net, zurawski@es.net
ESnet / Lawrence Berkeley National Laboratory

Materials Cyberinfrastructure for Research Data Management Workshop
Princeton, NJ
May 23-24, 2023

National Science Foundation Award #2328479

Outline

- ***Buffering Discussion***
- Science DMZ Security
- Organizational Collaboration
- Performance Through Firewalls
- Questions/Conclusions

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

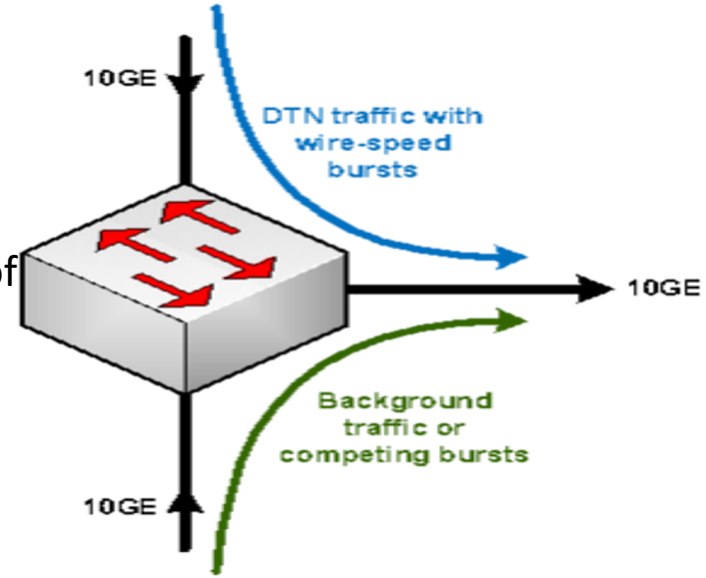
Common Circumstance: Multiple Ingress Data Flows, Common Egress

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



EPOC
Engagement and Performance Operations Center

Some Stuff We Think Is Important

- Deep interface queues (e.g. ***buffer***)
  - Output queue or VOQ – doesn't matter
  - What TCP sees is what matters – fan-in is *not* your friend
  - No, this isn't buffer bloat
- Good counters
  - We like the ability to reliably count *every* packet associated with a particular flow, address pair, etc
    - Very helpful for debugging packet loss
    - Must not affect performance (just count it, don't punt it)
    - sflow support if possible
  - If the box is going to drop a packet, it should increment a counter somewhere indicating that it dropped the packet
    - Magic vendor permissions and hidden commands should not be necessary
    - Some boxes just lie – run away!
- Single-flow performance should be wire-speed

EPOC
Engagement and Performance Operations Center

All About That Buffer (No Cut Through)

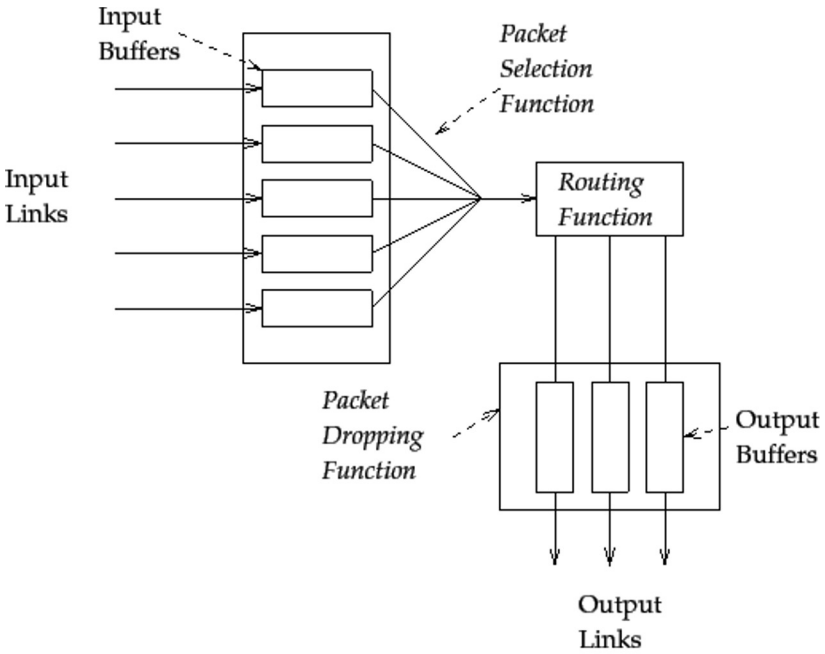


Figure 1: Basic Router Architecture

EPOC
Engagement and Performance Operations Center

All About That Buffer (No Cut Through)

- Data arrives from multiple sources

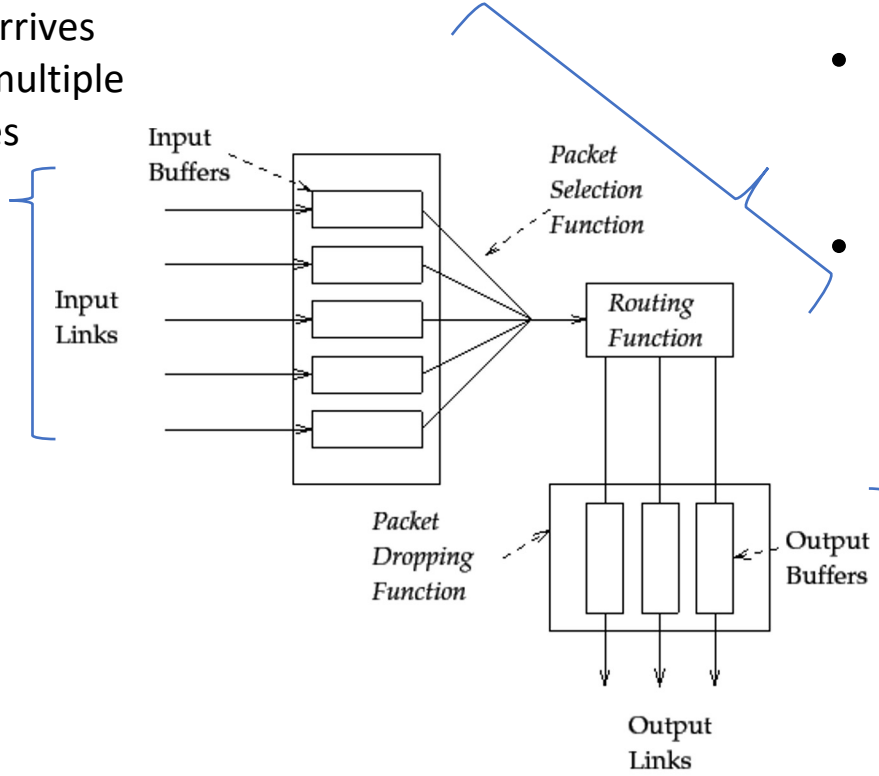


Figure 1: Basic Router Architecture

- Buffers have a finite amount of memory
  - Some have this per interface
  - Others may have access to a shared memory region with other interfaces
- The processing engine will:
  - Extract each packet/frame from the queues
  - Pull off header information to see where the destination should be
  - Move the packet/frame to the correct output queue
- Additional delay is possible as the queues physically write the packet to the transport medium (e.g. optical interface, copper interface)

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.

EPOC
Engagement and Performance Operations Center

All About That Buffer (No Cut Through)

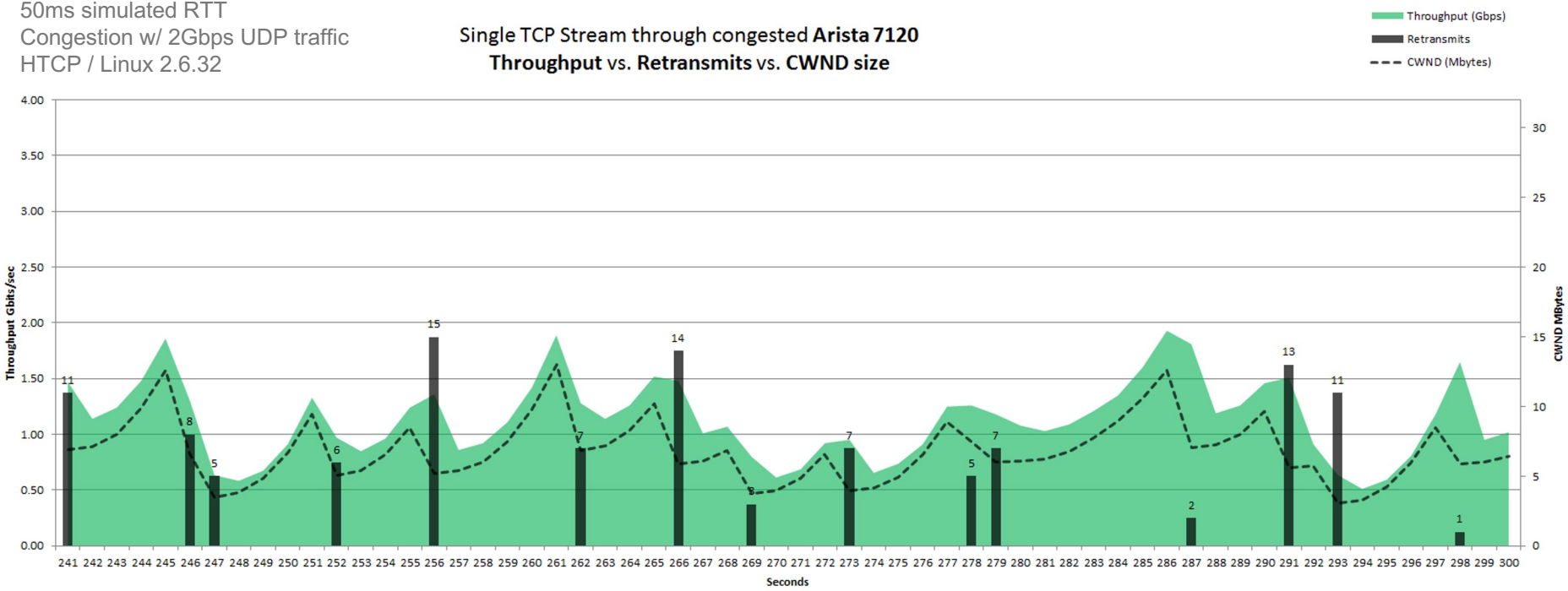
- Buffering isn't as important on the LAN (this is why you are normally pressured to buy 'cut through' devices)
  - Change the math to make the Latency 1ms and the expectation 10Gbps = ***1.25MB***
  - 'Cut through' and low latency switches are designed for the data center, and can handle typical data 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)

EPOC
Engagement and Performance Operations Center

TCP's Congestion Control

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



Slide from Michael Smitasin, LBLnet

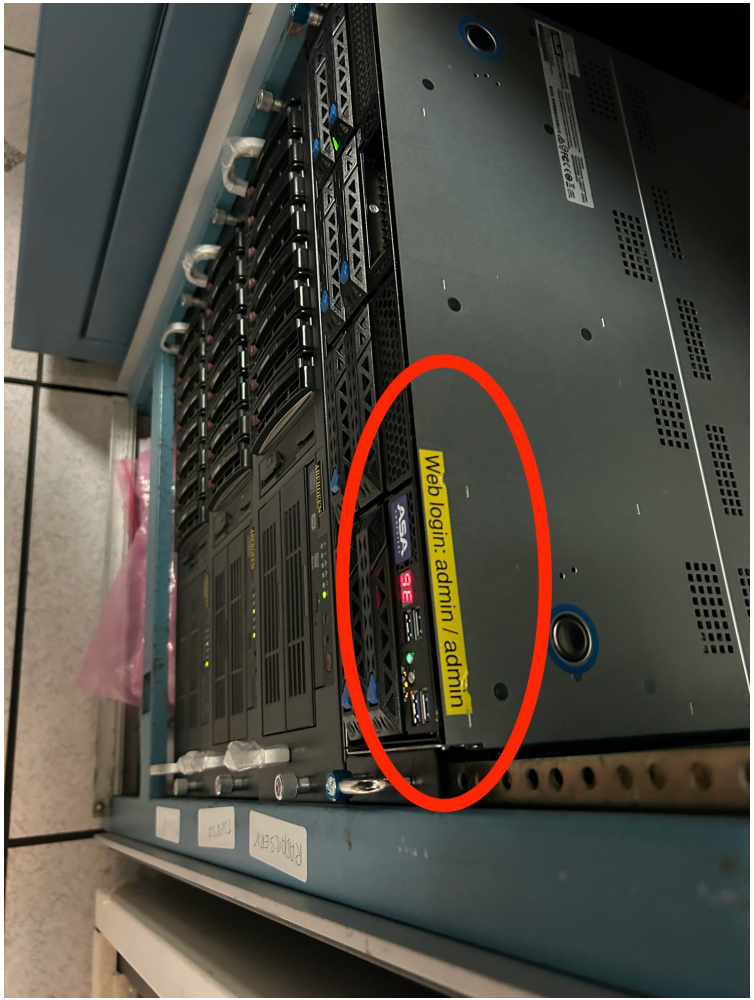
Outline

- Buffering Discussion
- ***Science DMZ Security***
- Organizational Collaboration
- Performance Through Firewalls
- Questions/Conclusions

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**



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

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.

EPOC
Engagement and Performance Operations Center

Outline

- Buffering Discussion
- Science DMZ Security
- ***Organizational Collaboration***
- Performance Through Firewalls
- Questions/Conclusions

EPOC
Engagement and Performance Operations Center

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

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? Let's 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***

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
  - Security Infrastructure
- Facilitate Usage
- Have a good story for long term usage/onboarding/expansion/maintenance

EPOC
Engagement and Performance Operations Center

Outline

- Buffering Discussion
- Science DMZ Security
- Organizational Collaboration
- ***Performance Through Firewalls***
- Questions/Conclusions

EPOC
Engagement and Performance Operations Center

Science DMZ Placement Outside the Enterprise Firewall

- Why? For performance reasons
  - Specifically: **Science DMZ traffic does not traverse the firewall data plane**
  - This has nothing to do with whether packet filtering is part of the security enforcement toolkit
- Lots of heartburn over this, especially from the perspective of a conventional firewall manager
  - Organizational policy directives can **mandate** firewalls
  - Firewalls are designed to protect converged enterprise networks
  - Why would you put critical assets outside the firewall??
- **The answer:** Firewalls are typically a poor fit for high-performance science applications

EPOC
Engagement and Performance Operations Center

Typical Firewall Internals

- Composed of a set of processors which inspect traffic in parallel
  - Traffic distributed among processors so all traffic for a particular connection goes to the same processor
  - Simplifies state management
  - Parallelization scales deep analysis
- Excellent fit for enterprise traffic profile
  - High connection count, low per-connection data rate
  - Complex protocols with embedded threats
- Each processor is a fraction of firewall link speed
  - Significant limitation for data-intensive science applications
  - Overload causes packet loss – performance crashes

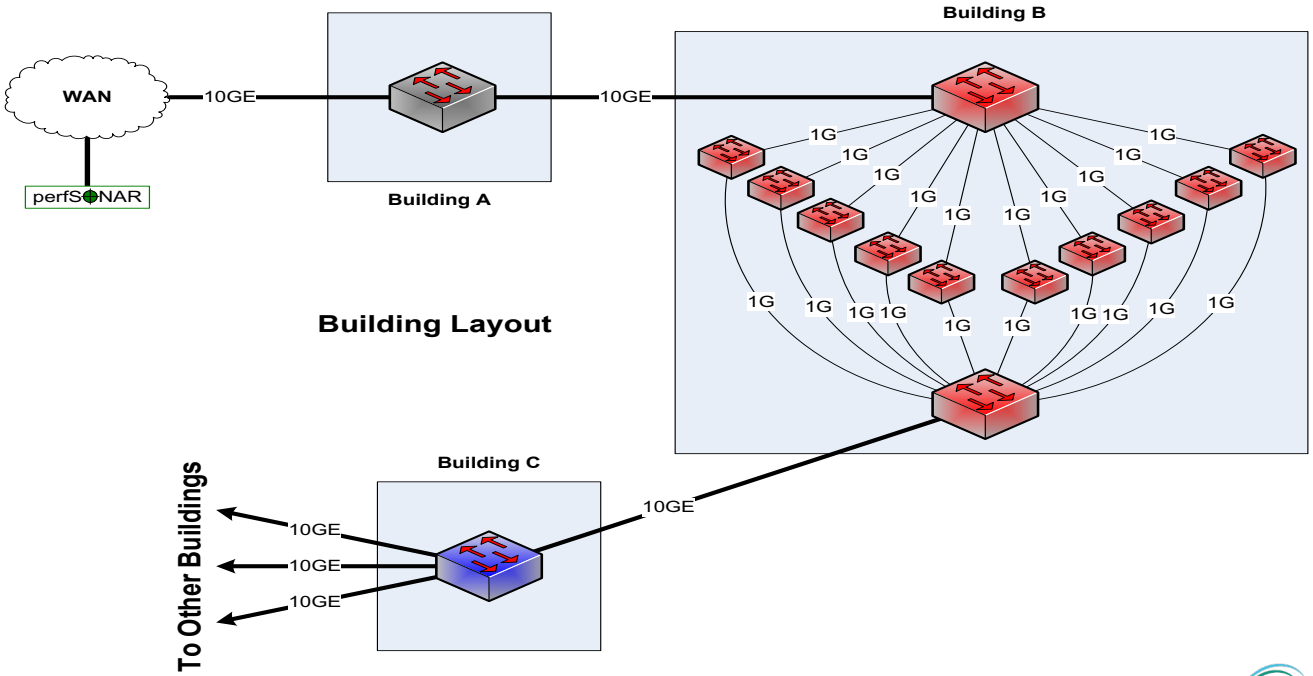
EPOC
Engagement and Performance Operations Center

Thought Experiment

- We’re going to do a thought experiment
- Consider a network between three buildings – A, B, and C
  - This is supposedly a 10Gbps network end to end (look at the links on the buildings)
  - Building A houses the border router – not much goes on there except the external connectivity
  - Lots of work happens in building B – so much so that the processing is done with multiple processors to spread the load in an affordable way, and aggregate the results after
  - Building C is where we branch out to other buildings
- Every link between buildings is 10Gbps – this is a 10Gbps network, right???

EPOC
Engagement and Performance Operations Center

Notional 10G Network Between Buildings



EPOC
Engagement and Performance Operations Center

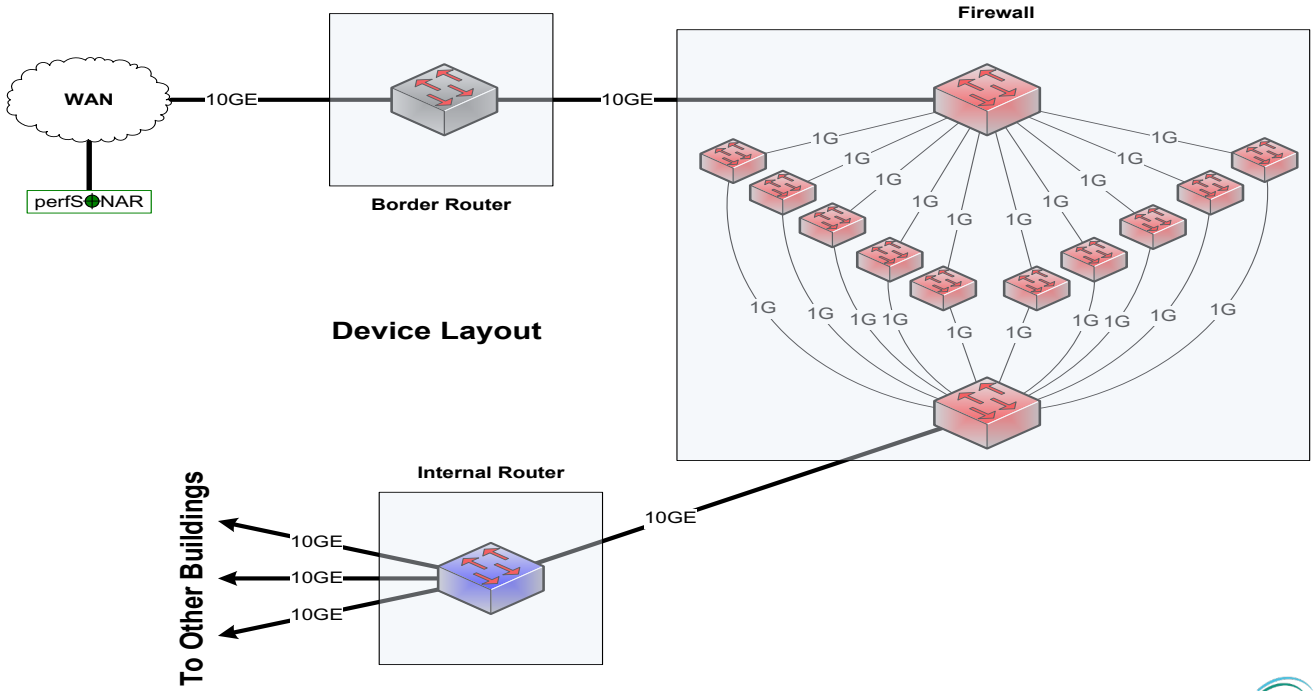
Clearly Not A 10Gbps Network

- If you look at the inside of Building B, it is obvious from a network engineering perspective that this is not a 10Gbps network
  - Clearly the maximum per-flow data rate is 1Gbps, not 10Gbps
  - However, if you convert the buildings into network elements while keeping their internals intact, you get routers and firewalls
  - What firewall did the organization buy? What's inside it?
  - Those little 1G "switches" are firewall processors

- This parallel firewall architecture has been in use for years
  - Slower processors are cheaper
  - Typically fine for a commodity traffic load
  - Therefore, this design is cost competitive and common

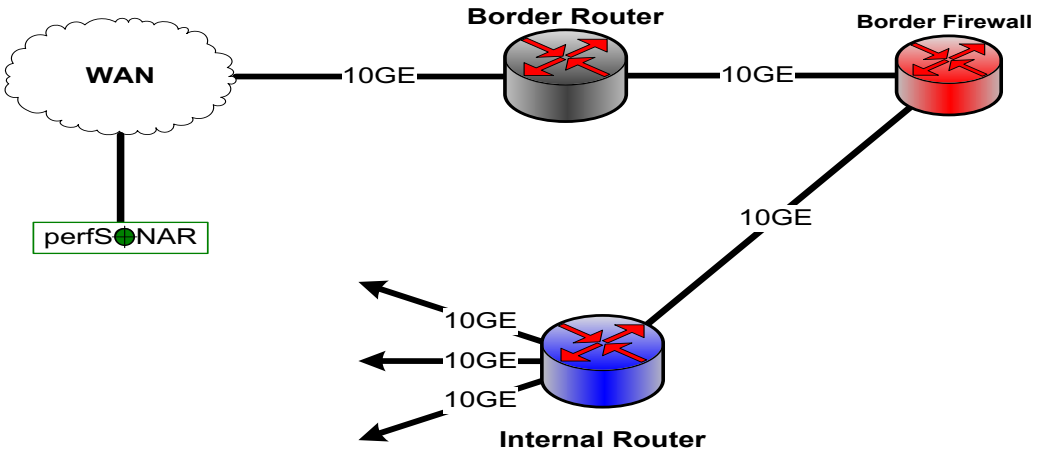
EPOC
Engagement and Performance Operations Center

Notional 10G Network Between Devices



EPOC
Engagement and Performance Operations Center

Notional Network Logical Diagram



EPOC
Engagement and Performance Operations Center

Firewall Capabilities and Science Traffic

- Firewalls have a lot of sophistication in an enterprise setting
  - Application layer protocol analysis (HTTP, POP, MSRPC, etc.)
  - Built-in VPN servers
  - User awareness
- Data-intensive science flows don't match this profile
  - Common case – data on filesystem A needs to be on filesystem Z
    - Data transfer tool verifies credentials over an encrypted channel
    - Then open a socket or set of sockets, and send data until done (1TB, 10TB, 100TB, …)
  - One workflow can use 10% to 50% or more of a 10G network link
- **Do we have to use a firewall?**

EPOC
Engagement and Performance Operations Center

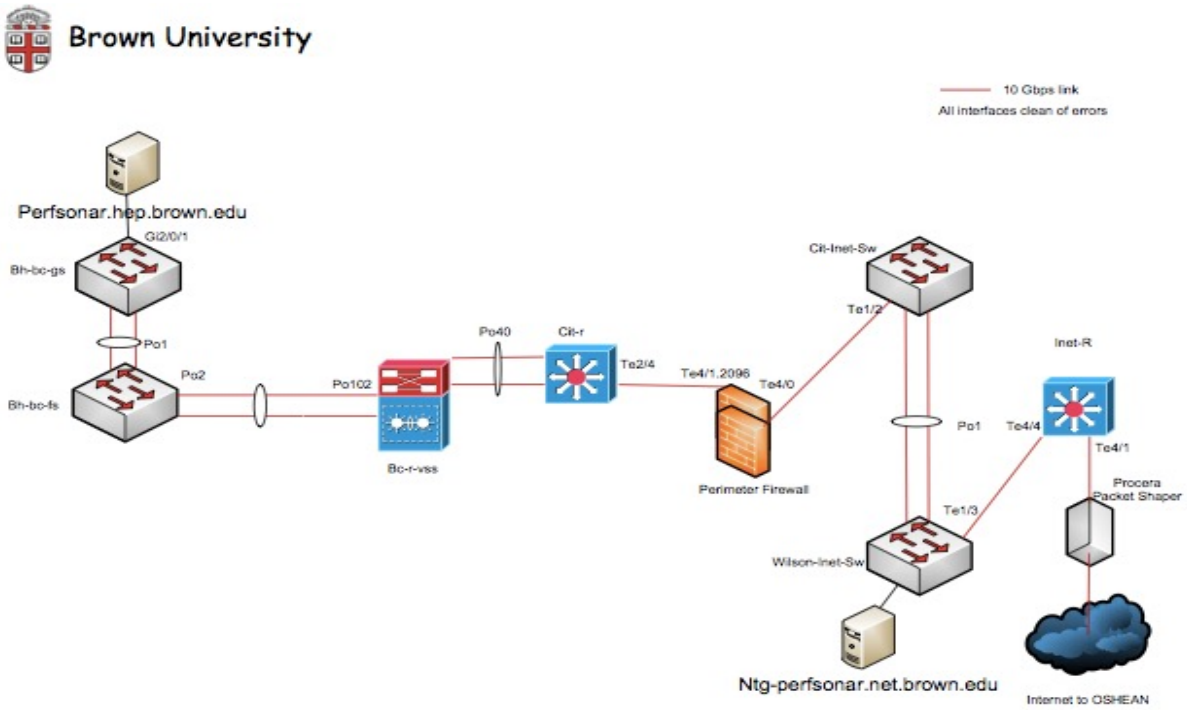
Firewalls as Router Access Control Lists

- When you ask a firewall administrator to allow data transfers through the firewall, what do they ask for?
  - IP address of your host
  - IP address of the remote host
  - Port range
  - **That looks like an ACL to me – I can do that on the router!**

- Firewalls make expensive, low-performance ACL filters compared to the ACL capabilities are typically built into the router
- Router ACLs do not drop traffic permitted by policy, while enterprise firewalls can (and often do)

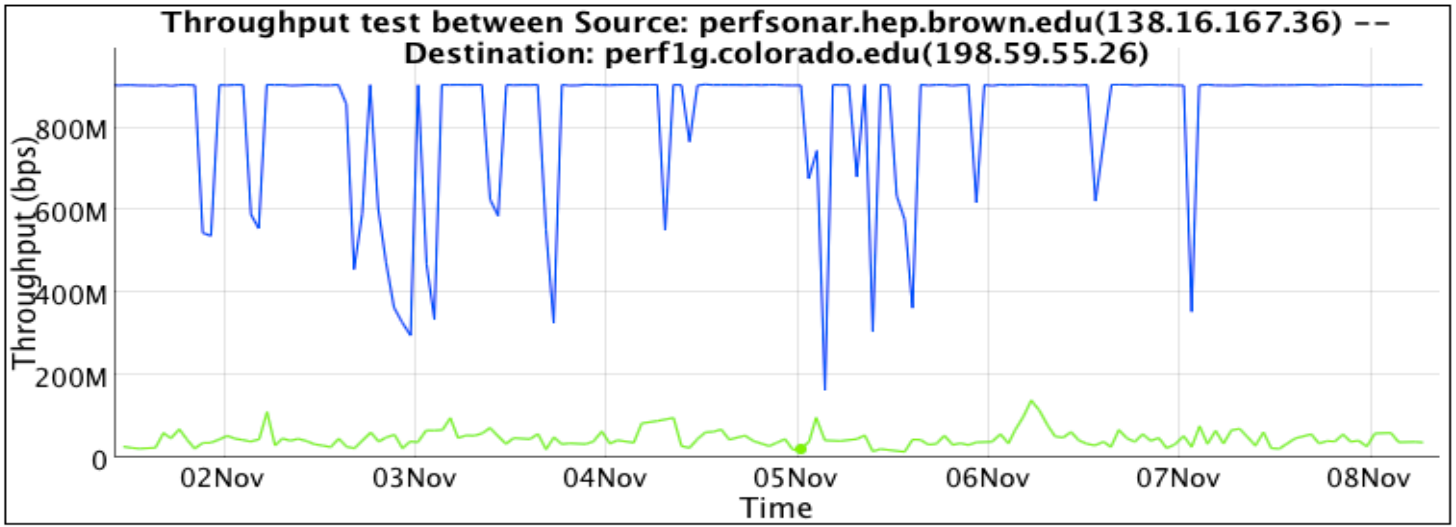
EPOC
Engagement and Performance Operations Center

Firewall Performance Example from Brown University



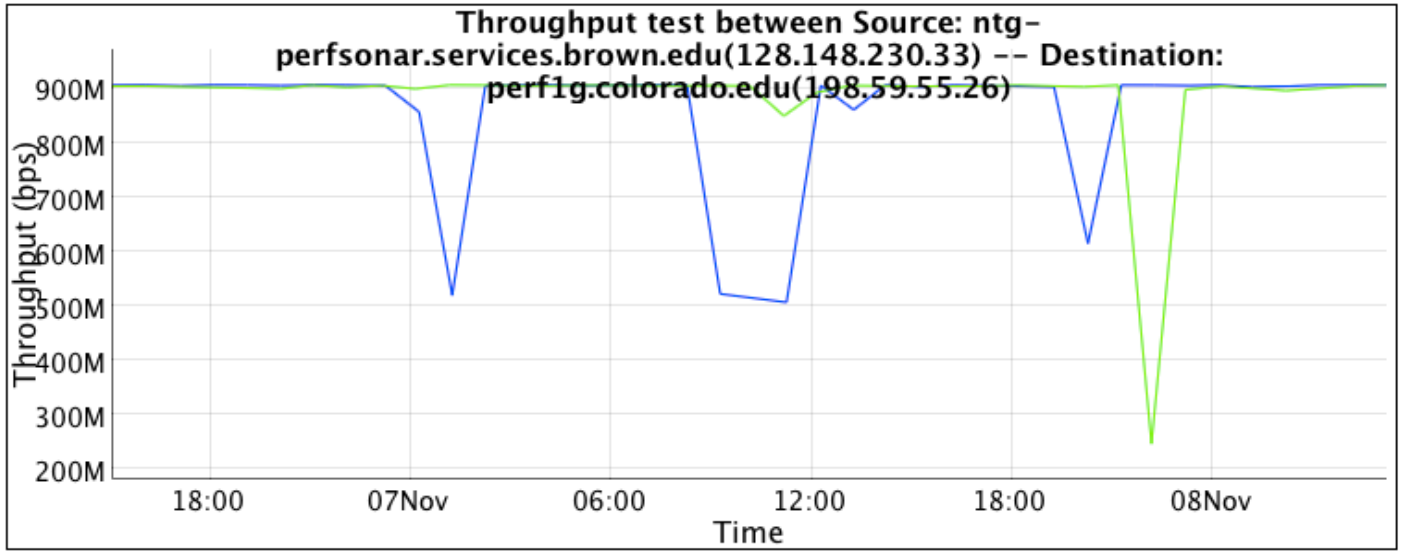
Brown University Example

- Results to host behind the firewall:



Brown University Example

- In front of the firewall:



TCP Dynamics

- Want more proof – lets look at a measurement tool through the firewall.
  - Measurement tools emulate a well behaved application
- ‘Outbound’, not filtered:

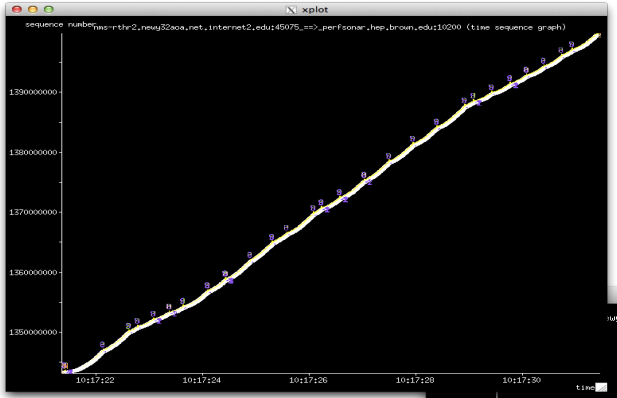
```
nuttcp -T 10 -i 1 -p 10200 bwctl.newy.net.internet2.edu
   92.3750 MB /   1.00 sec =  774.3069 Mbps     0 retrans
  111.8750 MB /   1.00 sec =  938.2879 Mbps     0 retrans
  111.8750 MB /   1.00 sec =  938.3019 Mbps     0 retrans
  111.7500 MB /   1.00 sec =  938.1606 Mbps     0 retrans
  111.8750 MB /   1.00 sec =  938.3198 Mbps     0 retrans
  111.8750 MB /   1.00 sec =  938.2653 Mbps     0 retrans
  111.8750 MB /   1.00 sec =  938.1931 Mbps     0 retrans
  111.9375 MB /   1.00 sec =  938.4808 Mbps     0 retrans
  111.6875 MB /   1.00 sec =  937.6941 Mbps     0 retrans
  111.8750 MB /   1.00 sec =  938.3610 Mbps     0 retrans

 1107.9867 MB /  10.13 sec =  917.2914 Mbps 13 %TX 11 %RX 0 retrans
8.38 msRTT
```

TCP Dynamics Through Firewall

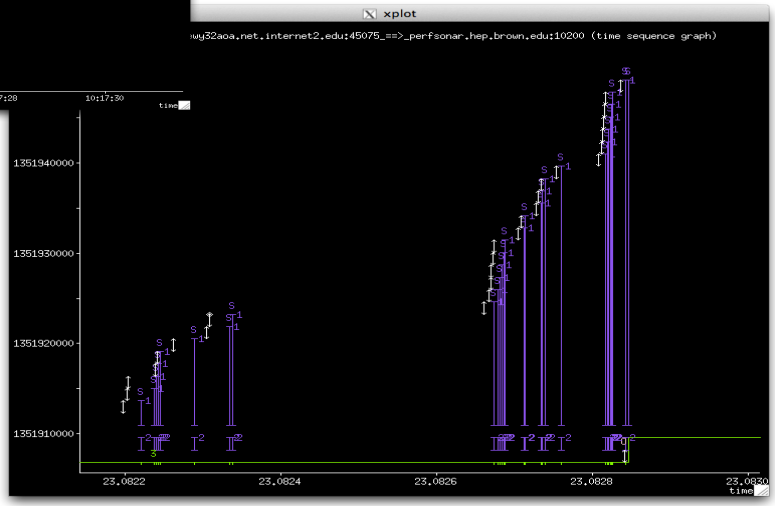
- ‘Inbound’, filtered:

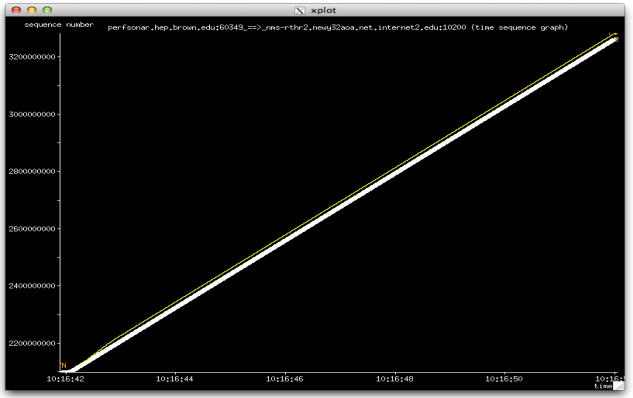
```
nuttcp -r -T 10 -i 1 -p 10200 bwctl.newy.net.internet2.edu
    4.5625 MB /   1.00 sec =   38.1995 Mbps    13 retrans
    4.8750 MB /   1.00 sec =   40.8956 Mbps     4 retrans
    4.8750 MB /   1.00 sec =   40.8954 Mbps     6 retrans
    6.4375 MB /   1.00 sec =   54.0024 Mbps     9 retrans
    5.7500 MB /   1.00 sec =   48.2310 Mbps     8 retrans
    5.8750 MB /   1.00 sec =   49.2880 Mbps     5 retrans
    6.3125 MB /   1.00 sec =   52.9006 Mbps     3 retrans
    5.3125 MB /   1.00 sec =   44.5653 Mbps     7 retrans
    4.3125 MB /   1.00 sec =   36.2108 Mbps     7 retrans
    5.1875 MB /   1.00 sec =   43.5186 Mbps     8 retrans

   53.7519 MB /  10.07 sec =   44.7577 Mbps 0 %TX 1 %RX 70 retrans 8.29 msRTT
```

tcptrace output: with and without a firewall



firewall





No firewall

Outline

- Buffering Discussion
- Science DMZ Security
- Organizational Collaboration
- Performance Through Firewalls
- ***Questions/Conclusions***

EPOC
Engagement and Performance Operations Center

Questions?

Transfer Performance problems? EPOC is here to help!

- epoc@tacc.utexas.edu
- https://epoc.global/

NSF Award: *2328479*

EPOC
Engagement and Performance Operations Center

Science DMZ Security

Ken Miller, Jason Zurawski
ken@es.net, zurawski@es.net
ESnet / Lawrence Berkeley National Laboratory

Materials Cyberinfrastructure for Research Data Management Workshop
Princeton, NJ
May 23-24, 2023

National Science Foundation Award #2328479