The Science DMZ: A Network Design Pattern for Data-Intensive Science

Jason Zurawski – zurawski@es.net
Science Engagement Engineer, ESnet
Lawrence Berkeley National Laboratory

New Mexico Technology in Education (NMTIE)
November 19th, 2014
Overview

• ESnet Overview
• Science DMZ Motivation and Introduction
• Science DMZ Architecture
• Network Monitoring
• Data Transfer Nodes & Applications
• Science DMZ Security
• User Engagement
• Wrap Up
ESnet at a Glance

- High-speed national network, optimized for DOE science missions:
  - connecting 40 labs, plants and facilities with >100 networks
  - $32.6M in FY14, 42FTE
  - older than commercial Internet, growing twice as fast

- $62M ARRA grant for 100G upgrade:
  - transition to new era of optical networking
  - world’s first 100G network at continental scale

- Culture of urgency:
  - 4 awards in past 3 years
  - R&D100 Award in FY13
  - “5 out of 5” for customer satisfaction in last review
  - *Dedicated staff to support the mission of science*

The Office of Science supports:
- 27,000 Ph.D.s, graduate students, undergraduates, engineers, and technicians
- 26,000 users of open-access facilities
- 300 leading academic institutions
- 17 DOE laboratories

Universities
DOE laboratories
ESnet Accepted Traffic: Jan 1990 - Aug 2014 (Log Scale)

Projected volume for Aug 2015: 75.0 PB

Actual volume for Aug 2014: 15.9 PB
Network as Infrastructure Instrument

**Vision**: Scientific progress will be completely unconstrained by the physical location of instruments, people, computational resources, or data.
Overview

• ESnet Overview
• Science DMZ Motivation and Introduction
• Science DMZ Architecture
• Network Monitoring
• Data Transfer Nodes & Applications
• Science DMZ Security
• User Engagement
• Wrap Up
Motivation

• Networks are an essential part of data-intensive science
  – Connect data sources to data analysis
  – Connect collaborators to each other
  – Enable machine-consumable interfaces to data and analysis resources (e.g. portals), automation, scale

• Performance is critical
  – Exponential data growth
  – Constant human factors
  – Data movement and data analysis must keep up

• Effective use of wide area (long-haul) networks by scientists has historically been difficult
Traditional “Big Science”
Big Science Now Comes in Small Packages

ALS Beamlines
May 2013

Magnetic Microscopy  7.3.1
Diagnostic Beamline  7.2
Surface, Materials Science (MAESTRO)  7.0.1
Calibration, Optics Testing, Spectroscopy  6.3.2
Magnetic Spectroscopy, Materials Science  6.3.1
X-Ray Microscopy  6.1.2
Femtosecond and Picosecond Phenomena  6.0.2
Femtosecond and Picosecond Phenomena  6.0.1
Polymer STXM  5.3.2
Research and Development  5.3.1
Macromolecular Crystallography  5.0.3
Macromolecular Crystallography  5.0.2
Macromolecular Crystallography  5.0.1
Macromolecular Crystallography  4.2.2
Magnetic Spectroscopy  4.0.2
MERLIN  4.0.3
Commercial LIGA  3.2.1
Diagnostic Beamline  3.1
National Center for X-Ray Tomography  2.1.2
IR Spectromicroscopy  1.4.4
IR Spectromicroscopy  1.4.3

KEY
--- Operational --- Under Construction

Insertion Device Beamlines
Bend Magnet Beamlines
Superbend Beamlines

7.3.3 SAXS/WAXS
8.0.1 Surface, Materials Science
8.2.1 Macromolecular Crystallography
8.2.2 Macromolecular Crystallography
8.3.1 Macromolecular Crystallography
8.3.2 Tomography
9.0.2 Chemical Dynamics
9.0.1 Diffraction Imaging
9.3.1 AMO, Materials Science
9.3.2 Chemical, Materials Science
10.0.1 Correlated Materials, AMO
10.3.1 X-Ray Fluorescence
10.3.2 MicroXAS
11.0.1 PEEM3
11.0.2 Molecular Environmental Science
11.3.1 Small-Molecule Crystallography
11.3.2 EUV Lithography, Mask Inspection
12.0.1 EUV Lithography, Photoemission
12.0.2 Coherent Soft X-Ray Science
12.2.2 High Pressure
12.3.1 Macromolecular Crystallography/Bio-SAXS
12.3.2 X-Ray Microdiffraction
Understanding Data Trends

- Small collaboration scale, e.g. light and neutron sources
- Medium collaboration scale, e.g. HPC codes
- Large collaboration scale, e.g. LHC

A few large collaborations have internal software and networking organizations.
Data Mobility in a Given Time Interval

<table>
<thead>
<tr>
<th>Data set size</th>
<th>1 Minute</th>
<th>5 Minutes</th>
<th>20 Minutes</th>
<th>1 Hour</th>
</tr>
</thead>
<tbody>
<tr>
<td>10PB</td>
<td>1,333.33 Tbps</td>
<td>266.67 Tb/s</td>
<td>66.67 Tb/s</td>
<td>22.22 Tbps</td>
</tr>
<tr>
<td>1PB</td>
<td>133.33 Tbps</td>
<td>26.67 Tb/s</td>
<td>6.67 Tb/s</td>
<td>2.22 Tbps</td>
</tr>
<tr>
<td>100TB</td>
<td>13.33 Tbps</td>
<td>2.67 Tb/s</td>
<td>0.66 Tb/s</td>
<td>0.22 Tb/s</td>
</tr>
<tr>
<td>10TB</td>
<td>1.33 Tbps</td>
<td>266.67 Gbps</td>
<td>66.67 Gbps</td>
<td>22.22 Gbps</td>
</tr>
<tr>
<td>1TB</td>
<td>133.33 Gbps</td>
<td>26.67 Gbps</td>
<td>6.67 Gbps</td>
<td>2.22 Gbps</td>
</tr>
<tr>
<td>100GB</td>
<td>13.33 Gbps</td>
<td>2.67 Gbps</td>
<td>0.66 Gbps</td>
<td>0.22 Gbps</td>
</tr>
<tr>
<td>10GB</td>
<td>1.33 Gbps</td>
<td>266.67 Mbps</td>
<td>66.67 Mbps</td>
<td>22.22 Mbps</td>
</tr>
<tr>
<td>1GB</td>
<td>133.33 Mbps</td>
<td>26.67 Mbps</td>
<td>6.67 Mbps</td>
<td>2.22 Mbps</td>
</tr>
<tr>
<td>100MB</td>
<td>13.33 Mbps</td>
<td>2.67 Mbps</td>
<td>0.66 Mbps</td>
<td>0.22 Mbps</td>
</tr>
</tbody>
</table>

This table available at:

http://fasterdata.es.net/fasterdata-home/requirements-and-expectations/
The Central Role of the Network

• The very structure of modern science assumes science networks exist: high performance, feature rich, global scope

• What is “The Network” anyway?
  – “The Network” is the set of devices and applications involved in the use of a remote resource
    • This is not about supercomputer interconnects
    • This is about data flow from experiment to analysis, between facilities, etc.
  – User interfaces for “The Network” – portal, data transfer tool, workflow engine
  – Therefore, servers and applications must also be considered

• What is important? Ordered list:
  1. Correctness
  2. Consistency
  3. Performance
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

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

Throughput vs. Increasing Latency with .0046% Packet Loss

With loss, high performance beyond metro distances is essentially impossible

Measured (TCP Reno)  Measured (HTCP)  Theoretical (TCP Reno)  Measured (no loss)

Local (LAN)  Metro Area  Regional  Continental  International

© 2014, Energy Sciences Network
Working With TCP In Practice

• Far easier to support TCP than to fix TCP
  – People have been trying to fix TCP for years – limited success
  – Like it or not we’re stuck with TCP in the general case

• Pragmatically speaking, we must accommodate TCP
  – Sufficient bandwidth to avoid congestion
  – Zero packet loss
  – Verifiable infrastructure
    • Networks are complex
    • Must be able to locate problems quickly
    • Small footprint is a huge win – small number of devices so that problem isolation is tractable
Putting A Solution Together

• Effective support for TCP-based data transfer
  – Design for correct, consistent, high-performance operation
  – Design for ease of troubleshooting

• Easy adoption is critical
  – Large laboratories and universities have extensive IT deployments
  – Drastic change is prohibitively difficult

• Cybersecurity – defensible without compromising performance

• Borrow ideas from traditional network security
  – Traditional DMZ
    • Separate enclave at network perimeter ("Demilitarized Zone")
    • Specific location for external-facing services
    • Clean separation from internal network
  – Do the same thing for science – *Science DMZ*
The Science DMZ Superfecta

**Engagement**
- Partnerships
- Education & Consulting
- Resources & Knowledgebase

**Data Transfer Node**
- High performance
- Configured for data transfer
- Proper tools

**perfSONAR**
- Enables fault isolation
- Verify correct operation
- Widely deployed in ESnet and other networks, as well as sites and facilities

**Science DMZ**
- Dedicated location for DTN
- Proper security
- Easy to deploy - no need to redesign the whole network

**Network Architecture**

**Dedicated Systems for Data Transfer**

**Performance Testing & Measurement**

**Engagement with Network Users**
Overview

• ESnet Overview
• Science DMZ Motivation and Introduction
• Science DMZ Architecture
• Network Monitoring
• Data Transfer Nodes & Applications
• Science DMZ Security
• User Engagement
• Wrap Up
Abstract or Prototype Deployment

• Add-on to existing network infrastructure
  – All that is required is a port on the border router
  – Small footprint, pre-production commitment

• Easy to experiment with components and technologies
  – DTN prototyping
  – perfSONAR testing

• Limited scope makes security policy exceptions easy
  – Only allow traffic from partners
  – Add-on to production infrastructure – lower risk
Science DMZ Design Pattern (Abstract)

Border Router

Clean, High-bandwidth WAN path

Site / Campus access to Science DMZ resources

PerfSONAR

High performance Data Transfer Node with high-speed storage

Per-service security policy control points
Local And Wide Area Data Flows

- **Border Router**
  - Clean, High-bandwidth WAN path
  - Site / Campus access to Science DMZ resources

- **Enterprise Border Router/Firewall**

- **Science DMZ Switch/Router**
  - Per-service security policy control points

- **High performance Data Transfer Node**
  - with high-speed storage

- **High Latency WAN Path**
- **Low Latency LAN Path**
Support For Multiple Projects

- Science DMZ architecture allows multiple projects to put DTNs in place
  - Modular architecture
  - Centralized location for data servers
- This may or may not work well depending on institutional politics
  - Issues such as physical security can make this a non-starter
  - On the other hand, some shops already have service models in place
- On balance, this can provide a cost savings – it depends
  - Central support for data servers vs. carrying data flows
  - How far do the data flows have to go?
Multiple Projects

- **Clean, High-bandwidth WAN path**
- **Site / Campus access to Science DMZ resources**
- **Per-project security policy control points**
Supercomputer Center Deployment

• High-performance networking is assumed in this environment
  – Data flows between systems, between systems and storage, wide area, etc.
  – Global filesystem often ties resources together
    • Portions of this may not run over Ethernet (e.g. IB)
    • Implications for Data Transfer Nodes
• “Science DMZ” may not look like a discrete entity here
  – By the time you get through interconnecting all the resources, you end up with most of the network in the Science DMZ
  – This is as it should be – the point is appropriate deployment of tools, configuration, policy control, etc.
• Office networks can look like an afterthought, but they aren’t
  – Deployed with appropriate security controls
  – Office infrastructure need not be sized for science traffic
Supercomputer Center

- WAN
  - perfSONAR

- Border Router
  - Core Switch/Router
  - perfSONAR

- Firewall
  - Offices

- Front end switch

- Front end nodes
  - Data Transfer Nodes
  - perfSONAR

- Supercomputer

- Parallel Filesystem
Supercomputer Center Data Path

- Supercomputer
- Border Router
- Firewall
-perfSONAR
- Core Switch/Router
- Offices
- perfSONAR
- Front end switch
- Data Transfer Nodes
- High Latency WAN Path
- Low Latency LAN Path
- High Latency VC Path

WAN
Virtual Circuit
Routed

Parallel Filesystem
Development Environment

• One thing that often happens is that an early power user of the Science DMZ is the network engineering group that builds it
  – Service prototyping
  – Deployment of test applications for other user groups to demonstrate value

• The production Science DMZ is just that – production
  – Once users are on it, you can’t take it down to try something new
  – Stuff that works tends to attract workload

• *Take-home message: plan for multiple Science DMZs from the beginning – at the very least you’re going to need one for yourself*

• The Science DMZ model easily accommodates this
Science DMZ – Flexible Design Pattern

• The Science DMZ design pattern is highly adaptable to research
• Deploying a research Science DMZ is straightforward
  – The basic elements are the same
    • Capable infrastructure designed for the task
    • Test and measurement to verify correct operation
    • Security policy well-matched to the environment, application set is strictly limited to reduce risk
  – Connect the research DMZ to other resources as appropriate
• The same ideas apply to supporting an SDN effort
  – Test/research areas for development
  – Transition to production as technology matures and need dictates
  – One possible trajectory follows…
Science DMZ – Separate SDN Connection

Border Router

High performance routed path

Site / Campus access to Science DMZ resources

WAN

Enterprise Border Router/Firewall

SDN

Production DTN

Research DTN

Per-service security policy control points

Science DMZ Connections

PerfSONAR

SDN

Production Switch/Router

PerfSONAR

Science DMZ Switch/Router

PerfSONAR
Science DMZ – Production SDN Connection

- **WAN**
  - **perfSONAR**

- **Border Router**
  - **High performance routed path**
  - **perfSONAR**
  - **Site / Campus access to Science DMZ resources**

- **Enterprise Border Router/Firewall**
  - **Site / Campus LAN**

- **Production SDN Switch/Router**
  - **Science DMZ Connections**

- **Research Science DMZ Switch/Router**
  - **Research DTN**
  - **perfSONAR**

- **Production DTN**

**Per-service security policy control points**
Common Threads

- Two common threads exist in all these examples
- Accommodation of TCP
  - Wide area portion of data transfers traverses purpose-built path
  - High performance devices that don’t drop packets
- Ability to test and verify
  - When problems arise (and they always will), they can be solved if the infrastructure is built correctly
  - Small device count makes it easier to find issues
  - Multiple test and measurement hosts provide multiple views of the data path
    - perfSONAR nodes at the site and in the WAN
    - perfSONAR nodes at the remote site
Multiple Ingress 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
Router and Switch Output Queues

- Interface output queue allows the router or switch to avoid causing packet loss in cases of momentary congestion.

- In network devices, queue depth (or ‘buffer’) is often a function of cost:
  - Cheap, fixed-config LAN switches (especially in the 10G space) have inadequate buffering. Imagine a 10G ‘data center’ switch as the guilty party.
  - Cut-through or low-latency Ethernet switches typically have inadequate buffering (the whole point is to avoid queuing!)

- Expensive, chassis-based devices are more likely to have deep enough queues:
  - Juniper MX and Alcatel-Lucent 7750 used in ESnet backbone.
  - Other vendors make such devices as well - details are important.
  - Thx to Jim: [http://people.ucsc.edu/~warner/buffer.html](http://people.ucsc.edu/~warner/buffer.html)

- This expense is one driver for the Science DMZ architecture – only deploy the expensive features where necessary.
Output Queue Drops – Common Locations

- Site Border Router
  - 10GE
  - Site Core Switch/Router
  - 10GE
  - Inbound data path

- Department Core Switch
  - Department uplink to site core constrained by budget or legacy equipment
  - 1GE
  - Outbound data path

- Cluster data transfer node
  - 10GE

- Department cluster switch
  - 1GE, 1GE

- 32+ cluster nodes

- Wiring closet switch
  - 1GE
  - 1GE, 1GE

- Workstations

- Common location of output queue drops for traffic inbound from the WAN

- Common locations of output queue drops for traffic outbound toward the WAN

WAN

© 2014, Energy Sciences Network
Overview

- ESnet Overview
- Science DMZ Motivation and Introduction
- Science DMZ Architecture
- Network Monitoring
- Data Transfer Nodes & Applications
- Science DMZ Security
- User Engagement
- Wrap Up
Performance Monitoring

• Everything may function perfectly when it is deployed
• Eventually something is going to break
  – Networks and systems are complex
  – Bugs, mistakes, ...
  – Sometimes things just break – this is why we buy support contracts
• Must be able to find and fix problems when they occur
• Must be able to find problems in other networks (your network may be fine, but someone else’s problem can impact your users)

• TCP was intentionally designed to hide all transmission errors from the user:
  – “As long as the TCPs continue to function properly and the internet system does not become completely partitioned, no transmission errors will affect the users.” (From RFC793, 1981)
Soft Network Failures – Hidden Problems

• Hard failures are well-understood
  – Link down, system crash, software crash
  – Traditional network/system monitoring tools designed to quickly find hard failures

• Soft failures result in degraded capability
  – Connectivity exists
  – Performance impacted
  – Typically something in the path is functioning, but not well

• Soft failures are hard to detect with traditional methods
  – No obvious single event
  – Sometimes no indication at all of any errors

• Independent testing is the only way to reliably find soft failures
Sample Soft Failures

Gradual failure of optical line card

Rebooted router with full route table

---

normal performance

degradating performance

repair

one month
Testing Infrastructure – perfSONAR

• perfSONAR is:
  – A widely-deployed test and measurement infrastructure
    • ESnet, Internet2, US regional networks, international networks
    • Laboratories, supercomputer centers, universities
  – A suite of test and measurement tools
  – A collaboration that builds and maintains the toolkit

• By installing perfSONAR, a site can leverage over 1100 test servers deployed around the world

• perfSONAR is ideal for finding soft failures
  – Alert to existence of problems
  – Fault isolation
  – Verification of correct operation
perfSONAR Deployment Footprint

pS Performance Toolkit Deployments

October 2012 Through August 2014

© 2014, Energy Sciences Network
perfSONAR Dashboard:
http://ps-dashboard.es.net

ESnet - ESnet to ESnet Packet Loss Testing

- Green: Loss rate is <= 0.001
- Yellow: Loss rate is >= 0.001
- Red: Loss rate is >= 0.1
- Orange: Unable to retrieve data
- Grey: Check has not yet run
Overview

• ESnet Overview
• Science DMZ Motivation and Introduction
• Science DMZ Architecture
• Network Monitoring
• Data Transfer Nodes & Applications
• Science DMZ Security
• User Engagement
• Wrap Up
Dedicated Systems – Data Transfer Node

• The DTN is dedicated to data transfer

• Set up specifically for high-performance data movement
  – System internals (BIOS, firmware, interrupts, etc.)
  – Network stack
  – Storage (global filesystem, Fibrechannel, local RAID, etc.)
  – High performance tools
  – No extraneous software

• Limitation of scope and function is powerful
  – No conflicts with configuration for other tasks
  – Small application set makes cybersecurity easier
Data Transfer Tools For DTNs

• Parallelism is important
  – It is often easier to achieve a given performance level with four parallel connections than one connection
  – Several tools offer parallel transfers, including Globus/GridFTP

• Latency interaction is critical
  – Wide area data transfers have much higher latency than LAN transfers
  – Many tools and protocols assume a LAN

• Workflow integration is important

• Key tools: Globus Online, HPN-SSH
Data Transfer Tool Comparison

- In addition to the network, using the right data transfer tool is critical.
- Data transfer test from Berkeley, CA to Argonne, IL (near Chicago). RTT = 53 ms, network capacity = 10Gbps.

<table>
<thead>
<tr>
<th>Tool</th>
<th>Throughput</th>
</tr>
</thead>
<tbody>
<tr>
<td>scp</td>
<td>140 Mbps</td>
</tr>
<tr>
<td>HPN patched scp</td>
<td>1.2 Gbps</td>
</tr>
<tr>
<td>ftp</td>
<td>1.4 Gbps</td>
</tr>
<tr>
<td>GridFTP, 4 streams</td>
<td>5.4 Gbps</td>
</tr>
<tr>
<td>GridFTP, 8 streams</td>
<td>6.6 Gbps</td>
</tr>
</tbody>
</table>

Note that to get more than 1 Gbps (125 MB/s) disk to disk requires properly engineered storage (RAID, parallel filesystem, etc.)
Overview

• ESnet Overview
• Science DMZ Motivation and Introduction
• Science DMZ Architecture
• Network Monitoring
• Data Transfer Nodes & Applications
• Science DMZ Security
• User Engagement
• Wrap Up
Science DMZ Security

• Goal – disentangle security policy and enforcement for science flows from security for 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 often cause performance problems

• Separation allows each to be optimized
Performance Is A Core Requirement

- Core information security principles
  - Confidentiality, Integrity, Availability (CIA)
  - Often, CIA and risk mitigation result in poor performance

- In data-intensive science, performance is an additional core mission requirement: CIA $\rightarrow$ PICA
  - CIA principles are important, but if performance is compromised the science mission fails
  - Not about “how much” security you have, but how the security is implemented
  - Need a way to appropriately secure systems without performance compromises
Placement Outside the Firewall

• The Science DMZ resources are placed outside the enterprise firewall for performance reasons
  – The meaning of this is specific – *Science DMZ traffic does not traverse the firewall data plane*
  – Packet filtering is fine – just don’t do it with a firewall

• Lots of heartburn over this, especially from the perspective of a conventional firewall manager
  – Lots of organizational policy directives mandating firewalls
  – Firewalls are designed to protect converged enterprise networks
  – Why would you put critical assets outside the firewall???

• The answer is that firewalls are typically a poor fit for high-performance science applications
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 typically 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?
Firewalls As Access 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!*

• No special config for advanced protocol analysis – just address/port

• Router ACLs are better than firewalls at address/port filtering
  – ACL capabilities are typically built into the router
  – Router ACLs typically do not drop traffic permitted by policy
Security Without Firewalls

• Data intensive science traffic interacts poorly with firewalls
• Does this mean we ignore security? **NO!**
  – We **must** protect our systems
  – We just need to find a way to do security that does not prevent us from getting the science done

**Key point – security policies and mechanisms that protect the Science DMZ should be implemented so that they do not compromise performance**

• Traffic permitted by policy should not experience performance impact as a result of the application of policy
Firewall Performance Example

- Observed performance, via perfSONAR, through a firewall:
  - Almost 20 times slower through the firewall
- Observed performance, via perfSONAR, bypassing the firewall:
  - Huge improvement without the firewall
If Not Firewalls, Then What?

• Intrusion Detection Systems (IDS)
  – Bro is high-performance and battle-tested
    • Bro protects several high-performance national assets
    • Bro can be scaled with clustering: [http://www.bro-ids.org/documentation/cluster.html](http://www.bro-ids.org/documentation/cluster.html)
  – Other IDS solutions are available also

• Netflow and IPFIX can provide intelligence, but not filtering

• Openflow and SDN
  – Using Openflow to control access to a network-based service seems pretty obvious
  – This could significantly reduce the attack surface for any authenticated network service
  – This would only work if the Openflow device had a robust data plane
If Not Firewalls, Then What? (2)

• Aggressive access lists
  – More useful with project-specific DTNs
  – If the purpose of the DTN is to exchange data with a small set of remote collaborators, the ACL is pretty easy to write
  – Large-scale data distribution servers are hard to handle this way (but then, the firewall ruleset for such a service would be pretty open too)

• Limitation of the application set
  – One of the reasons to limit the application set in the Science DMZ is to make it easier to protect
  – Keep desktop applications off the DTN (and watch for them anyway using logging, netflow, etc – take violations seriously)
  – This requires collaboration between people – networking, security, systems, and scientists
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
Overview

• ESnet Overview
• Science DMZ Motivation and Introduction
• Science DMZ Architecture
• Network Monitoring
• Data Transfer Nodes & Applications
• Science DMZ Security
• User Engagement
• Wrap Up
Challenges to Network Adoption

- Causes of performance issues are complicated for users.
- Lack of communication and collaboration between the CIO’s office and researchers on campus.
- Lack of IT experience within a science collaboration.
- User’s performance expectations are low (“The network is too slow”, “I tried it and it didn’t work”).
- Cultural change is hard (“we’ve always shipped disks!”).
- Scientists want to do science not IT support.
Requirements Reviews

http://www.es.net/about/science-requirements/network-requirements-reviews/

The purpose of these reviews is to accurately characterize the near-term, medium-term and long-term network requirements of the science conducted by each program office.

The reviews attempt to bring about a network-centric understanding of the science process used by the researchers and scientists, to derive network requirements.

*We have found this to be an effective method for determining network requirements for ESnet's customer base.*
How do we know what our scientists need?

- Each Program Office has a dedicated requirements review every three years
- Two workshops per year, attendees chosen by science programs
- Discussion centered on science case studies
  - Instruments and Facilities – the “hardware”
  - Process of Science – science workflow
  - Collaborators
  - Challenges
- Network requirements derived from science case studies + discussions
- Reports contain requirements analysis, case study text, outlook
2013 BER Sample Findings:

"EMSL frequently needs to ship physical copies of media to users when data sizes exceed a few GB. More often than not, this is due to lack of bandwidth or storage resources at the user's home institution."

<table>
<thead>
<tr>
<th>Science Instruments and Facilities</th>
<th>Process of Science</th>
<th>Data Set Size</th>
<th>LAN Transfer Time Needed</th>
<th>WAN Transfer Time Needed</th>
</tr>
</thead>
<tbody>
<tr>
<td>Near Term (0-2 years)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Broad suite of scientific instruments</td>
<td>• Primarily on-site access to instruments</td>
<td>• Data volume 6 TB/day</td>
<td>• 5 TB/day continuous, 24x7</td>
<td>• 200 GB/month at 1 GB/sec</td>
</tr>
<tr>
<td>• Chinook 160 Tflop supercomputer</td>
<td>• Remote access to Chinook computer, Aurora archive, and remote instrument operation</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Quiet wing with hires chemical</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5+ years</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Next-generation mass spectrometer</td>
<td>access and analysis</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Next-generation electron microscopy</td>
<td>• MyEMSL with search and first set of workflow and analysis capabilities</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5+ years</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>• HPCS-5 HPC system(s)</td>
<td>• Strong integration of data across capabilities</td>
<td>• Data volume 100 TB/day</td>
<td>• 200 TB/day continuous, 24x7</td>
<td>• 3 PB/month at 10 GB/sec to user's home institutions</td>
</tr>
<tr>
<td>• Next-next generation electron microscopy</td>
<td>• Comprehensive problem-solving environment on top of MyEMSL</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Enhanced imaging instruments</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
|”
Overview

- ESnet Overview
- Science DMZ Motivation and Introduction
- Science DMZ Architecture
- Network Monitoring
- Data Transfer Nodes & Applications
- On the Topic of Security
- User Engagement
- Wrap Up
Wrapup

• The Science DMZ design pattern provides a flexible model for supporting high-performance data transfers and workflows

• Key elements:
  – Accommodation of TCP
    • Sufficient bandwidth to avoid congestion
    • Loss-free IP service
  – Location – near the site perimeter if possible
  – Test and measurement
  – Dedicated systems
  – Appropriate security

• Support for advanced capabilities (e.g. SDN) is much easier with a Science DMZ
The Science DMZ in 1 Slide

Consists of **three key components**, all required:

- “Friction free” network path
  - Highly capable network devices (wire-speed, deep queues)
  - Virtual circuit connectivity option
  - Security policy and enforcement specific to science workflows
  - Located at or near site perimeter if possible
- Dedicated, high-performance Data Transfer Nodes (DTNs)
  - Hardware, operating system, libraries all optimized for transfer
  - Includes optimized data transfer tools such as Globus Online and GridFTP
- Performance measurement/test node
  - perfSONAR
- Engagement with end users

Details at [http://fasterdata.es.net/science-dmz/](http://fasterdata.es.net/science-dmz/)
Links

- ESnet fasterdata knowledge base
  • http://fasterdata.es.net/
- Science DMZ paper
  • http://www.es.net/assets/pubs_presos/sc13sciDMZ-final.pdf
- Science DMZ email list
  • https://gab.es.net/mailman/listinfo/sciencedmz
- perfSONAR
  • http://fasterdata.es.net/performance-testing/perfsonar/
  • http://www.perfsonar.net
Thanks!

Jason Zurawski – zurawski@es.net
Science Engagement Engineer, ESnet
Lawrence Berkeley National Laboratory

New Mexico Technology in Education (NMTIE)
November 19th, 2014