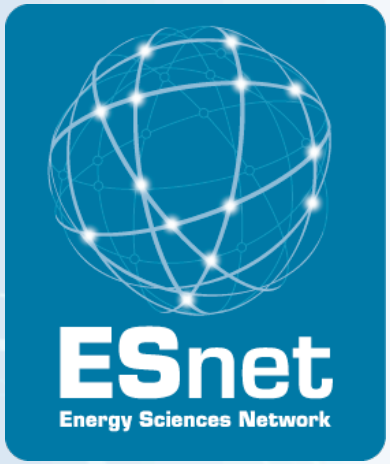


The Science DMZ

Eli Dart, Network Engineer

ESnet Engineering Group

Winter 2011 Joint Techs

Clemson, SC

February 1, 2011

Overview

Science Needs

- Data tsunami, new science processes
- Current problems
- Consequences

What is important and how to accomplish it

- Important aspects of the network for users
- Science DMZ architecture

Security Concerns

Recap

Resources

Science Needs – the "Data Tsunami"

"Data Tsunami" is a nice cliché

- Dramatic and catchy
- It's been used many times before
- That's because it's a good way of describing the issue succinctly

Data volume is causing many disciplines to re-think their strategies, collaboration structures, etc. – Examples:

- Genomics
- Materials science (e.g. light source users)
- Medicine

Most disciplines do not have deep internal networking expertise – they need our help

New Science Processes

Data Tsunami is not the whole story

New Science processes rely on networking for success

- Remote instrument control
- Real-time analysis for experiment health-check, collaboration, etc.
- Increasing scope of collaboration
  - Fewer instruments, more collaborators → increased need for collaboration tools, data sharing, etc.
  - Networks are the fabric the holds modern science together

Baseline performance requirements for wide-area networking are increasing dramatically – scientific productivity in many disciplines is gated on this

Example – Light Source Users

Current model – scientists take data at light source, data sets are relatively small (scale of ~1TB) – use USB hard drives today

Changes – data rates

- Next generation of instruments produce 10x to 100x data
- Instruments with data rates of 250MB/sec being deployed
- Physical transport on portable media will shortly become unworkable for many collaborations that have relied on this method for many years

Changes – science process

- Experiment automation bringing added efficiency, but requires experiment health checks
- Near-real-time data analysis needed
- Beamline computational resources inadequate – need supercomputer access
- Reliable, high speed data transfer to supercomputer centers necessary

Current Problems

The network serves many scientists poorly or not at all

- HEP is an exception (e.g. LHC)
- However, HEP is a bad model for judging other disciplines
  - HEP is large, sophisticated, experienced
  - HEP has its own internal software and networking organizations
- Most disciplines don't have the sophistication and capabilities of HEP
- Most disciplines simply can't use the network for data transfer or visualization, so the scientists have stopped trying – they "know" it can't be done

Since users "know" the network can't work for them, they don't open trouble tickets

Current Problems – Technical Causes

Poorly implemented foundations

- Many networks still have packet loss (incorrect config, cheap hardware, poor design, etc)
- System defaults are wrong – hosts still need to be tuned

Wrong tool for the job

- Use of SSH-based tools is common
- SSH has built-in performance limitations

Security blocks scientists at every turn

- Tools are blocked at the network layer or disallowed by policy
- Firewalls cause poor performance
- High-performance tools are often incompatible with system access technologies (e.g. SSH)

Consequences of Inaction

Science will proceed with or without networks

Networks will decrease in relevance for most scientific disciplines and the institutions that support science unless networks can be made useful

- Lower return on investment in scientific infrastructure
- Longer time to discovery
- Reduction of technological and scientific leadership for USA

If the high-performance networks built for science are not useful – if the scientists can't use them – then we have built a facility that is of no value to science

- Productivity/collaboration benefits from networking not realized
- Network-enabled modes of discovery unavailable
- For many disciplines, THIS IS THE CURRENT STATE

All Is Not Lost

Bad outcomes are not certain

There are successes – they just need to be generalized

Leadership by the networking community is required

- The old model of providing a toolkit and expecting scientists to learn networking has demonstrably failed
- The networking community must provide useful services and useful documentation for those services

Remember – *Most users are not networking experts, and it is unreasonable to expect them to become experts*

What is Important?

Ease of use by non-expert users

- Scientists are scientists – most are not network experts
- From the perspective of users, “the network” is the aggregate of applications, computers, storage systems, campus networks, regional networks, long-haul networks that participate in WAN data movement and similar tasks.
- In order for “the network” to be a useful tool, this complex assembly of components must somehow present a sane interface to non-experts

Zero packet loss between data movers

- High performance is impossible in the presence of packet loss
- This is different from commodity networking

Test and measurement for troubleshooting and repair

- Even though it works today, it will break tomorrow or the next day
- This is infrastructure – provisions must be made for maintenance and repair

Accomplishing the Important Things

Ease of use – reduce complexity

- Reduce number of devices in the path (fewer things to troubleshoot, configure, etc – more on this in a minute)
- Dedicated infrastructure for data movement

Zero packet loss

- Again, reduce device count
- Use appropriate network devices (e.g. with deep output queues) – this means eliminating LAN devices from WAN data path

Test and measurement

- Need well-defined location for well-configured test and measurement gear (e.g. perfSONAR)
- Locate test and measurement devices near data movers

Security – decouple science and business security policy and control points

Traditional DMZ

DMZ – “Demilitarized Zone”

- Network segment near the site perimeter with different security policy
- Commonly used architectural element for deploying WAN-facing services (e.g. email, DNS, web)

Traffic for WAN-facing services does not traverse the LAN

- WAN flows are isolated from LAN traffic
- Infrastructure for WAN services is specifically configured for WAN

Separation of security policy improves both LAN and WAN

- No conflation of security policy between LAN hosts and WAN services
- DMZ hosts provide specific services
- LAN hosts must traverse the same ACLs as WAN hosts to access DMZ

The Science DMZ

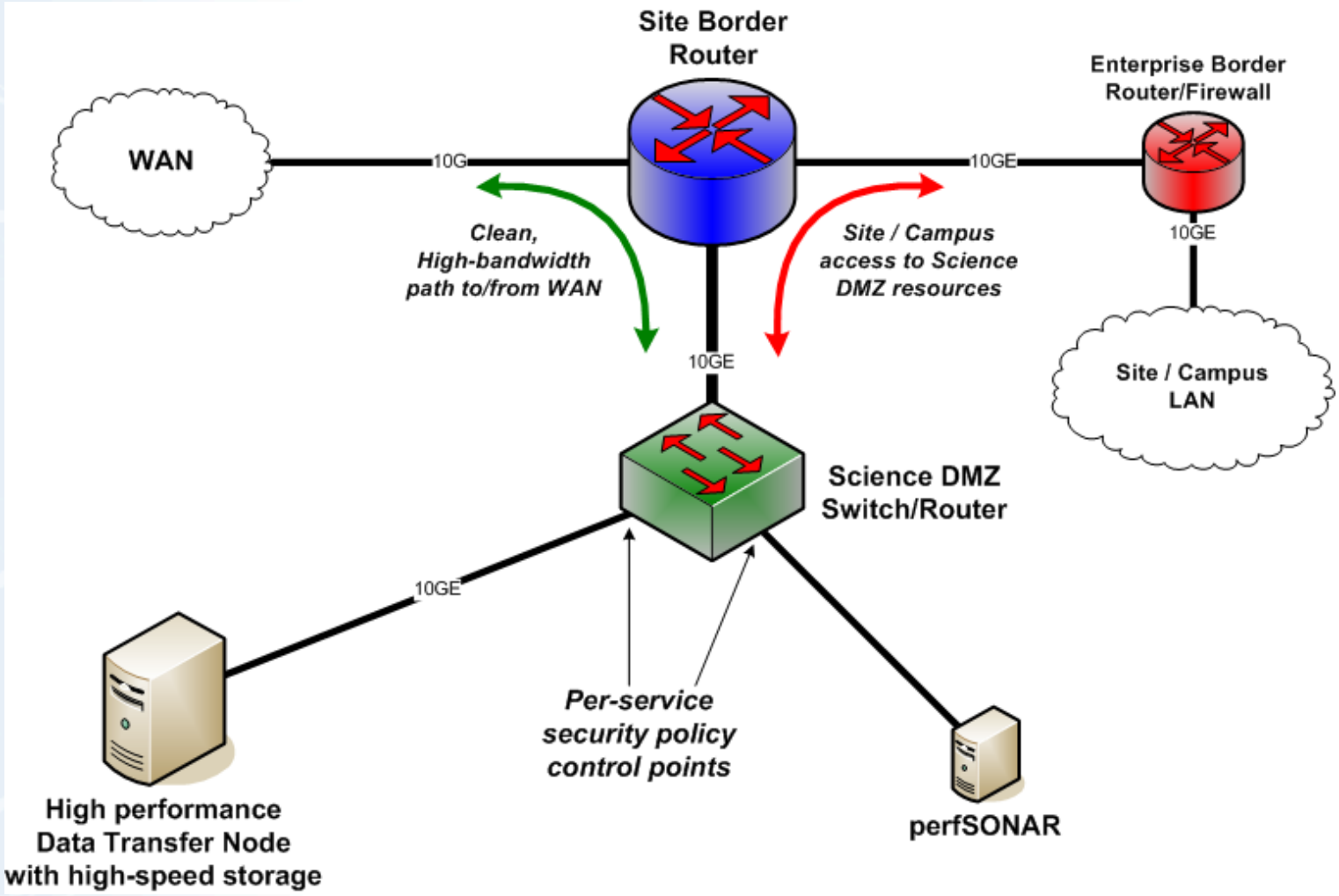
Science DMZ – a well-configured location for high-performance WAN-facing science services

- Dedicated, high-performance data movers
- Highly capable network devices (wire-speed, deep queues)
- Virtual circuit infrastructure
- perfSONAR

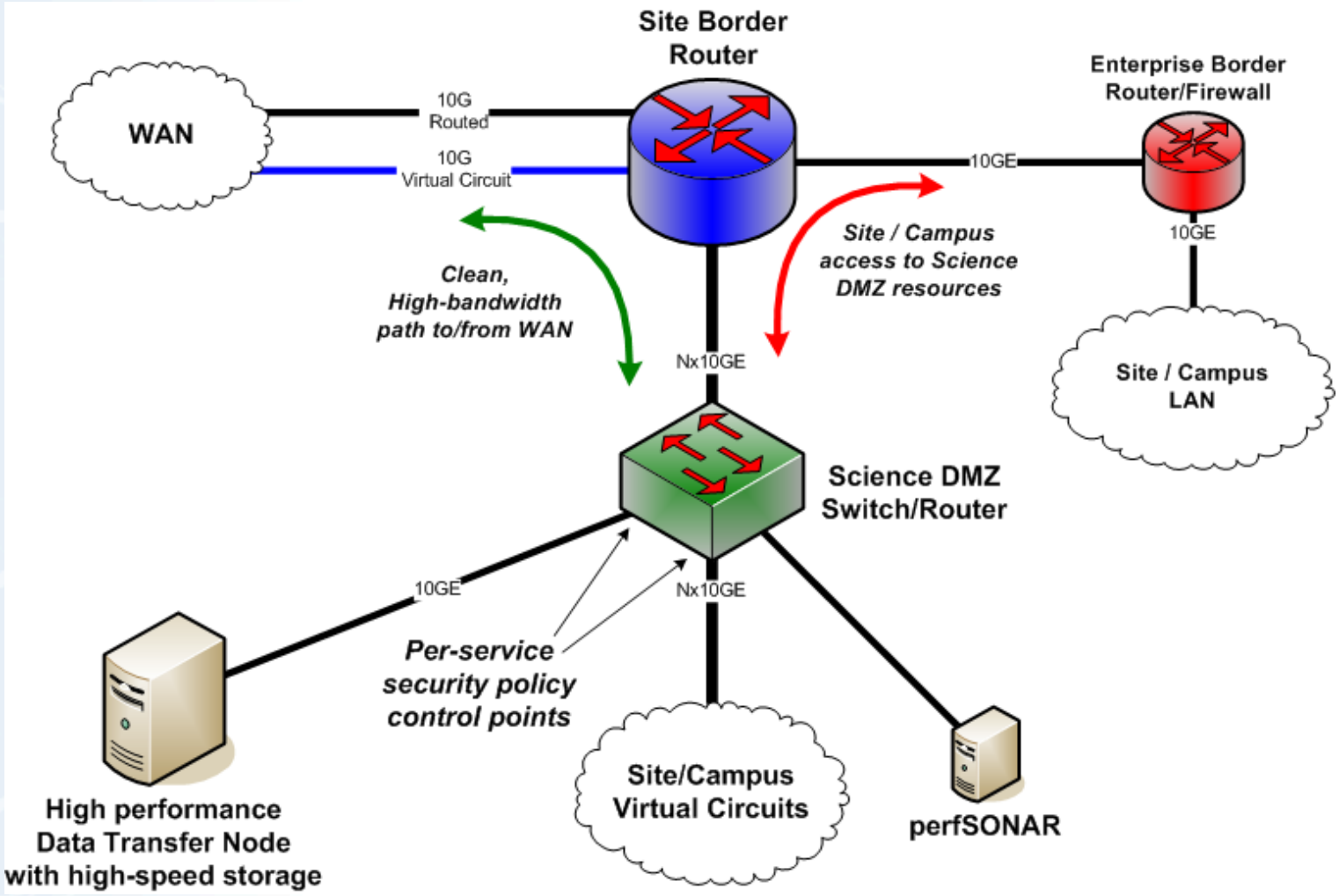
DYNES project is a specific example of this general architectural theme

- Dedicated infrastructure for data movers
- Virtual circuit termination
- Many high-bandwidth science sites have moved to this architecture already as a matter of necessity – DYNES is expanding on this because it works

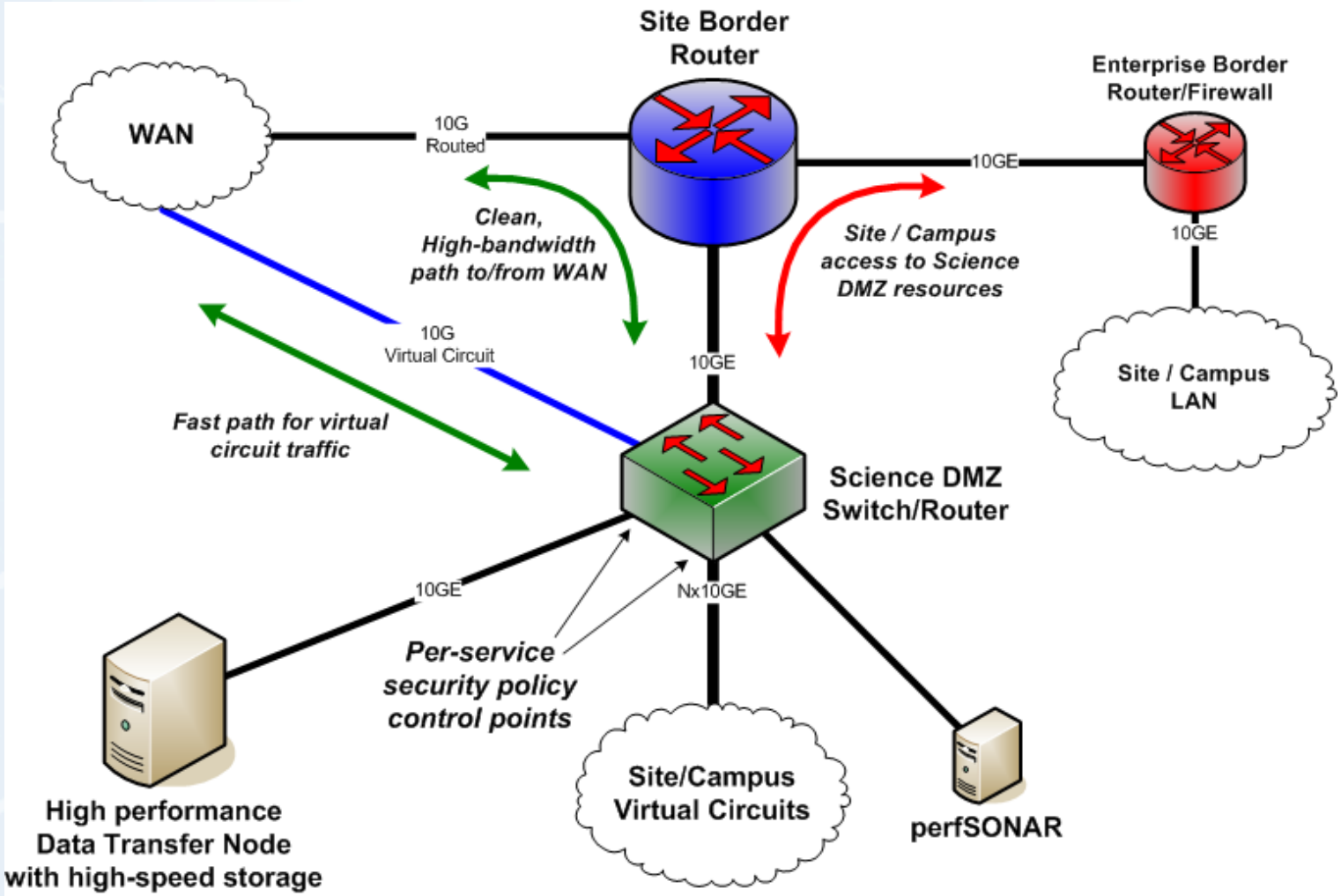
Science DMZ - Simple



Science DMZ - Advanced



Science DMZ – Separate Circuit Connection



Science DMZ Features and Components

Direct connection to site perimeter

- LAN devices eliminated from the path
  - LAN infrastructure need not be sized for science flows (reduced hardware costs)
  - LAN infrastructure need not be configured to support science flows (e.g. deep output queues can conflict with VOIP requirements)
  - LAN infrastructure need not implement features necessary for virtual circuits (reduced costs, reduced complexity)
- Security policy for science data movers is not conflated with policy for business systems, wireless devices, printers, VOIP, etc

Science DMZ Features and Components

Dedicated infrastructure for science applications

- LAN devices are not part of the troubleshooting mix
- Dedicated devices are easier to configure properly and maintain
- Data Transfer Node for high-performance data movement

Test and measurement deployed on same network infrastructure as science resources

The Data Transfer Node (DTN)

Dedicated, high-performance host for long-distance data transfer

High performance disk, for example:

- High-speed local RAID
- Fibrechannel attachment to SAN, if available
- Lustre or GPFS filesystem mount (e.g. when deployed at supercomputer center)

High-speed network connection (1G or 10G)

- Connected to Science DMZ
- Separate security policy from business traffic

Multiple sites and facilities are deploying DTNs (Supercomputer centers, labs, experiments, etc)

Significant performance gains from DTN deployment

Science DMZ Security

Goal – disentangle security policy and enforcement for science flows from that of business systems

Rationale

- Science flows are relatively 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, staff desktops, etc.
- Security controls that are typically implemented to protect business resources often cause performance problems

Science DMZ Security Implementation

Security policy for Science DMZ can typically be implemented with router filters/ACLs

- No need for an expensive firewall (and LAN firewall doesn't have to have WAN capabilities such as large buffers)
- Modern routers can filter packets at wire speed
- Routers do not rewrite packet headers and track state like firewalls do → entire classes of bugs/issues eliminated

Security policy for Science DMZ does not affect business systems (e.g. inbound ports for support of parallel data transfers)

- Desktops, printers, etc. not on Science DMZ
- Science DMZ systems are not subject to business security policies or policy enforcement devices
- This is sane – appropriate policies and controls are applied in both cases

Science DMZ Benefits

LAN infrastructure need not carry wide area science traffic

- Science traffic has different characteristics than business traffic
- Deep output queues, dedicated interfaces, etc. are expensive
- Accurate counters, per-filter counters, etc. are expensive

LAN transfers are much easier than WAN transfers

- Internal transfer of data from the site/campus LAN to/from the local Science DMZ will be much easier to debug, and is much more tolerant of the loss typically found in LANs
- LAN losses do not affect WAN transfers

Science DMZ Benefits

Separation of security policy

- Science DMZ security policy need not protect desktops, printers, etc
- Router filters do not have the problems firewalls have

Dedicated hosts

- Data Transfer Node is not some desktop, interactive node, cluster head-end, etc. with a bunch of other stuff running on it
- Fewer hosts to tune for WAN transfers
- Configuration is stable, CPU and I/O resources dedicated to moving data

Science DMZ scales with science need

- As science data intensity increases, Science DMZ can be upgraded without burdening entire LAN
- High-bandwidth instruments, connections to facilities, etc. can be added
- We know these needs are coming soon – it is time to prepare

Network Performance Knowledge Base

http://fasterdata.es.net/

Host tuning information:

- http://fasterdata.es.net/fasterdata/host-tuning/

Data transfer tools (including SCP/SFTP issues):

- http://fasterdata.es.net/fasterdata/data-transfer-tools/

Data Transfer Node, including sample hardware config:

- http://fasterdata.es.net/fasterdata/data-transfer-node/

Questions?

Thanks!

Questions?

Thanks!