The Science DMZ: Toward the Future

Eli Dart, Network Engineer
ESnet Science Engagement
Lawrence Berkeley National Laboratory

Internet2 Technology Exchange 2015
Cleveland, OH
October 5, 2015
Outline

• Context – Science DMZ in the community
• Building on the Science DMZ platform – examples:
  – Regional structures
  – Data portals
  – Cluster/HPC environments

• This talk assumes prior exposure to the Science DMZ
  – If this is new to you, come see me – happy to chat about it
  – Or check out the fasterdata knowledgebase:
    • http://fasterdata.es.net/science-dmz/
Context: Science DMZ Adoption

• DOE National Laboratories
  – HPC centers, LHC sites, experimental facilities
  – Both large and small sites
• NSF CC* programs have funded many Science DMZs
  – Significant investments across the US university complex
  – Big shoutout to the NSF – these programs are critically important
• Other US agencies
  – NIH
  – USDA Agricultural Research Service
• International
  – Australia https://www.rdsi.edu.au/dashnet
  – Brazil
  – UK
Strategic Impacts

• What does this mean?
  – We are in the midst of a significant cyberinfrastructure upgrade
  – Enterprise networks need not be unduly perturbed 😊

• Significantly enhanced capabilities compared to 3 years ago
  – Terabyte-scale data movement is much easier
  – Petabyte-scale data movement possible outside the LHC experiments
    • ~3.1Gbps = 1PB/month
    • ~14Gbps = 1PB/week
  – Widely-deployed tools are much better (e.g. Globus)

• Metcalfe’s Law of Network Utility
  – Value of Science DMZ proportional to the number of DMZs
    • $n^2$ or $n(\log n)$ doesn’t matter – the effect is real
  – Cyberinfrastructure value increases as we all upgrade
Next Steps – Building On The Science DMZ

- Enhanced cyberinfrastructure substrate now exists
  - Wide area networks (ESnet, GEANT, Internet2, Regionals)
  - Science DMZs connected to those networks
  - DTNs in the Science DMZs

- What does the scientist see?
  - Scientist sees a science application
    - Data transfer
    - Data portal
    - Data analysis
  - Science applications are the user interface to networks and DMZs

- The underlying cyberinfrastructure components (networks, Science DMZs, DTNs, etc.) are part of the instrument of discovery

- Large-scale data-intensive science requires that we build larger structures on top of those components
Pacific Research Platform - Interconnecting DMZs

- Pacific Research Platform is an effort to integrate the Science DMZs at the major West Coast universities, and extend to other institutions as well
  - Collaborating universities
  - National laboratories
  - National facilities
- Leverages investments in cyberinfrastructure, and builds upon them
- Larry Smarr (PI) with co-PIs:
  - Camille Crittenden, UC Berkeley
  - Tom DeFanti, Qualcomm Institute
  - Phil Papadopoulos, San Diego Supercomputer Center
  - Frank Wuerthwein, UC San Diego
- Interconnecting big facilities and institutional capabilities to accelerate scientific discovery
Science Data Portals

• Large repositories of scientific data
  – Climate data
  – Sky surveys (astronomy, cosmology)
  – Many others
  – Data search, browsing, access

• Many scientific data portals were designed 15+ years ago
  – Single-web-server design
  – Data browse/search, data access, user awareness all in a single system
  – All the data goes through the portal server
    • In many cases by design
    • E.g. embargo before publication (enforce access control)
Legacy Portal Design

• Very difficult to improve performance without architectural change
  – Software components all tangled together
  – Difficult to put the whole portal in a Science DMZ because of security
  – Even if you could put it in a DMZ, many components aren’t scalable
• What does architectural change mean?
Example of Architectural Change – CDN

• Let’s look at what Content Delivery Networks did for web applications

• CDNs are a well-deployed design pattern
  – Akamai and friends
  – Entire industry in CDNs
  – Assumed part of today’s Internet architecture

• What does a CDN do?
  – Store static content in a separate location from dynamic content
    • Complexity isn’t in the static content – it’s in the application dynamics
    • Web applications are complex, full-featured, and slow
      – Databases, user awareness, etc.
      – Lots of integrated pieces
    • Data service for static content is simple by comparison
  – Separation of application and data service allows each to be optimized
Classical Web Server Model

- Web browser fetches pages from web server
  - All content stored on the web server
  - Web applications run on the web server
    - Web server may call out to local database
    - Fundamentally all processing is local to the web server
  - Web server sends data to client browser over the network
- Perceived client performance changes with network conditions
  - Several problems in the general case
  - Latency increases time to page render
  - Packet loss + latency causes problems for large static objects
Solution: Place Large Static Objects Near Client

• CDN provides static content “close” to client
  – Latency goes down
    • Time to page render goes down
    • Static content performance goes up
  – Load on web server goes down (no need to serve static content)
  – Web server still manages complex behavior
    • Local reasoning / fast changes for application owner
• Significant win for web application performance
Client Simply Sees Increased Performance

- Client doesn’t see the CDN as a separate thing
  - Web content is all still viewed in a browser
    - Browser fetches what the page tells it to fetch
    - Different content comes from different places
    - User doesn’t know/care
- CDNs provide an architectural solution to a performance problem
  - Not brute-force
  - Work smarter, not harder
Architectural Examination of Data Portals

• Common data portal functions (most portals have these)
  – Search/query/discovery
  – Data download method for data access
  – GUI for browsing by humans
  – API for machine access – ideally incorporates search/query + download

• Performance pain is primarily in the data download piece
  – Rapid increase in data scale eclipsed legacy software stack capabilities
  – Portal servers often stuck in enterprise network

• Can we “disassemble” the portal and put the pieces back together better?
  – Use Science DMZ as a platform for the data piece
  – Avoid placing complex software in the Science DMZ
Legacy Portal Design
Next-Generation Portal Leverages Science DMZ

Portal server applications:
- web server
- search
- database
- authentication

Data Transfer Path
Portal Query/Browse Path

“Sealed” DTNs
(data access governed by portal)
Security Controls – DMZ, DTN, Filesystem

- DMZ
- DTN
- Filesystem

**Portal server applications:**
- web server
- search
- database
- authentication

**Border Router**

**Data Path**

**Firewall**

**Enterprise**

**WAN**

**Science DMZ Switch/Router**

**DTN Security Controls**

**Data Transfer Path**

**Portal Query/Browse Path**

**Filesystem**

**Filesystem (data store)**

**“Sealed” DTNs**
(data access governed by portal)

**ESnet**
Future HPC/Cluster Environments

• Common cluster architectural elements
  – Head/Login nodes
    • Primary user access
    • SSH typically required for access (for security reasons)
    • Job submission tools
    • Small-scale test jobs
  – Compute nodes
    • Run user jobs
    • Typically not accessible from outside
  – Central filesystem
    • Input data sets
    • Results of simulation/analysis
    • Mounted and available on compute nodes, login nodes, etc.
• Science DMZ can address some limitations of legacy designs
Abstract Cluster
Limitations of Older Design – no DTNs

• Abstract legacy cluster has limited data capabilities
• All data transfers must traverse Head/Login nodes
  – Firewall in the path
  – Configuration for data transfer tools conflated with cluster configuration
  – User interactive jobs keep CPUs busy
• Solution – add DTNs
  – Connect DTNs to Science DMZ
  – Mount central filesystem on DTNs
  – Only permit data transfer tools on DTNs
• Note: we’re going beyond the single-PI DTN
  – Moving toward institutional, multi-user capabilities
Future Cluster With Science DMZ Data Access

- **WAN**
- **Border Router**
- **Firewall**
- **Science DMZ Switch/Router**
- **Cluster Head/Login Nodes**
- **Filesistem**
- **Cluster compute nodes**

**Key Points:**
- "Sealed" DTNs (Globus only, no shell access)
- **perfSONAR**

**Network Elements:**
- **10GE** connections

**Note:**
- 10/5/15
Future Cluster Data Paths

Data Transfer Path
User Login/Shell Access Path
Compute Data Access Path
Future Cluster Security Controls

- WAN
- Border Router
- Firewall
- Enterprise
- Science DMZ Switch/Router
- PerfSONAR
- DTN Security Controls
- "Sealed" DTNs (Globus only, no shell access)
- Filesystem Security Controls
- Cluster Head/Login Nodes
- Filesystem
- Cluster compute nodes
Future Cluster With DTNs – Full Picture

Data Transfer Path
User Login/Shell Access Path
Compute Data Access Path

“Sealed” DTNs (Globus only, no shell access)
Let’s Talk About Risk

• What part of the cluster poses the greatest security risk?
  • Lowest risk: DTNs
    – No user shell access
    – No user programmatic filesystem access
    – Data transfer application (e.g. Globus) is the only method of interaction
  • More risky: Compute Nodes
    – Users can run arbitrary code
    – Programmatic access to filesystem
  • Greatest risk: Head/Login nodes
    – Users have shell access (via SSH – the firewall can’t see!)
    – Users have compiler access, can build/run arbitrary code
    – Programmatic access to filesystem
Moving Up The Stack

• Leverage networks and Science DMZs to build higher-level science services
• There are a lot of ways to do it
  – Traditional in HPC facilities
    • Shell access or sealed – whatever you want to build
    • User front-end to large-scale filesystems
  – Portal DTNs
    • Serve data based on portal context, but separate from portal services
    • Authentication for DTNs tied back to the portal
  – Regional integration, e.g. PRP
    • Common tools among collaborating universities
    • Working together to ensure performance and interoperability
• Think big – leverage ideas from other fields
  – Example: decomposition of Web applications using CDNs
  – How can we incorporate similar transformations into science applications?
Links and Lists

– 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
  • Send mail to sympa@lists.lbl.gov with subject "subscribe esnet-sciencedmz"
– perfSONAR
  • http://fasterdata.es.net/performance-testing/perfsonar/
  • http://www.perfsonar.net
– Globus
  • https://www.globus.org/
Thanks!

Eli Dart dart@es.net
Energy Sciences Network (ESnet)
Lawrence Berkeley National Laboratory

http://fasterdata.es.net/
http://my.es.net/
http://www.es.net/
Extra Slides
DTN Cluster Detail

WAN

Border Router

Science DMZ Switch/Router

Firewall

Enterprise

Configure as DTN Cluster

“Sealed” DTNs (Globus only, no shell access)

Cluster Head/Login Nodes

Filesistem

Cluster compute nodes

ESnet
DTN Cluster Design

• Configure all four DTNs as a single Globus endpoint
  – Globus has docs on how to do this

• Recent options for increased performance
  – Use additional parallel connections
  – Distribute transfers across multiple DTNs (Globus I/O Nodes)
  – Critical – only do this when all DTNs in the endpoint mount the same shared filesystem

• Use the Globus CLI command endpoint-modify
  – Use the --network-use option
  – Adjusts concurrency and parallelism
  – More info at globus.org (http://dev.globus.org/cli/reference/endpoint-modify/)
Security Footprint of a Globus Transfer

<table>
<thead>
<tr>
<th>Source Address</th>
<th>Source Port</th>
<th>Destination Address</th>
<th>Destination Port</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lab1 DTN</td>
<td>TCP 50000-51000</td>
<td>Lab2 DTN</td>
<td>TCP 50000-51000</td>
</tr>
<tr>
<td>Lab1 DTN</td>
<td>TCP 443, 2811, 7512</td>
<td>Globus Cloud</td>
<td>TCP unprivileged</td>
</tr>
<tr>
<td>Lab2 DTN</td>
<td>TCP 443, 2811, 7512</td>
<td>Lab1 DTN</td>
<td>TCP 50000-51000</td>
</tr>
<tr>
<td>Lab2 DTN</td>
<td>TCP 443, 2811, 7512</td>
<td>Globus Cloud</td>
<td>TCP unprivileged</td>
</tr>
<tr>
<td>Globus Cloud</td>
<td>TCP unprivileged</td>
<td>Lab1 DTN</td>
<td>TCP 443, 2811, 7512</td>
</tr>
</tbody>
</table>

Lab1 DTN security filters

Lab2 DTN security filters
Security Footprint of a Globus DTN

<table>
<thead>
<tr>
<th>Src Address</th>
<th>Src Port</th>
<th>Dst Address</th>
<th>Dst Port</th>
</tr>
</thead>
<tbody>
<tr>
<td>Local DTN</td>
<td>TCP 50000-51000</td>
<td>World</td>
<td>TCP 50000-51000</td>
</tr>
<tr>
<td>Local DTN</td>
<td>TCP 443, 2811, 7512</td>
<td>Globus Cloud</td>
<td>TCP unprivileged</td>
</tr>
<tr>
<td>World</td>
<td>TCP 50000-51000</td>
<td>Local DTN</td>
<td>TCP 50000-51000</td>
</tr>
<tr>
<td>Globus Cloud</td>
<td>TCP unprivileged</td>
<td>Local DTN</td>
<td>TCP 443, 2811, 7512</td>
</tr>
</tbody>
</table>

TCP ports 443, 2811, 7512

DTN security filters

Orchestration