Enabling large data transfers with the Science DMZ

Karl Newell, Network Engineer
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/)

*Science DMZ* is a trademark of The Energy Sciences Network (ESnet)

3 – ESnet Science Engagement (engage@es.net) - 7/16/15
**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 1,333.33 Tbps</td>
<td>266.67 Tbps</td>
<td>66.67 Tbps</td>
<td>22.22 Tbps</td>
<td></td>
</tr>
<tr>
<td>1PB 133.33 Tbps</td>
<td>26.67 Tbps</td>
<td>6.67 Tbps</td>
<td>2.22 Tbps</td>
<td></td>
</tr>
<tr>
<td>100TB &gt; 100Gbps 1.33 Tbps</td>
<td>266.67 Gbps</td>
<td>66.67 Gbps</td>
<td>22.22 Gbps</td>
<td></td>
</tr>
<tr>
<td>10TB 133.33 Gbps</td>
<td>266.67 Gbps</td>
<td>66.67 Gbps</td>
<td>22.22 Gbps</td>
<td></td>
</tr>
<tr>
<td>1TB &gt; 100Gbps 133.33 Gbps</td>
<td>266.67 Gbps</td>
<td>66.67 Gbps</td>
<td>22.22 Gbps</td>
<td></td>
</tr>
<tr>
<td>100GB &gt; 100Gbps 133.33 Gbps</td>
<td>266.67 Gbps</td>
<td>66.67 Gbps</td>
<td>22.22 Gbps</td>
<td></td>
</tr>
<tr>
<td>10GB 133.33 Gbps</td>
<td>266.67 Gbps</td>
<td>66.67 Gbps</td>
<td>22.22 Gbps</td>
<td></td>
</tr>
<tr>
<td>1GB &gt; 100Mbps 133.33 Mbps</td>
<td>266.67 Mbps</td>
<td>66.67 Mbps</td>
<td>22.22 Mbps</td>
<td></td>
</tr>
<tr>
<td>100MB &gt; 100Mbps 133.33 Mbps</td>
<td>266.67 Mbps</td>
<td>66.67 Mbps</td>
<td>22.22 Mbps</td>
<td></td>
</tr>
</tbody>
</table>

This table available at: [http://fasterdata.es.net/fasterdata-home/requirements-and-expectations/](http://fasterdata.es.net/fasterdata-home/requirements-and-expectations/)
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
Notional 10G Network Between Devices
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/)

---

*Science DMZ* is a trademark of The Energy Sciences Network (ESnet)
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 requires properly engineered storage (RAID, parallel filesystem).
• Hardware specs from fasterdata.es.net
  • Dual Xeon E5-2667v2 (8-core)
  • 64GB PC3-12800
  • 2 Mellanox ConnectX-3
    • Currently both used as 10G Ethernet
    • Transitioning backend to IB
  • RHEL 7
Data Transfer Nodes

- Using NFS to mount research data (DDN Gridscaler)
  - 2 Gbps bottleneck
- Transitioning to QDR InfiniBand and GPFS
The table below shows the performance of data transfer using the `iperf3` tool. The table includes the ID, interval, transfer size, bandwidth, and retransmissions for both the sender and receiver.

### Table 1: Data Transfer Performance

<table>
<thead>
<tr>
<th>ID</th>
<th>Interval</th>
<th>Transfer</th>
<th>Bandwidth</th>
<th>Retr</th>
<th>Role</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>0.00-10.00 sec</td>
<td>11.3 GBytes</td>
<td>9.68 Gbits/sec</td>
<td>0</td>
<td>sender</td>
</tr>
<tr>
<td>4</td>
<td>0.00-10.00 sec</td>
<td>11.3 GBytes</td>
<td>9.68 Gbits/sec</td>
<td>0</td>
<td>receiver</td>
</tr>
</tbody>
</table>

### Table 2: Additional Data Transfer Performance

<table>
<thead>
<tr>
<th>ID</th>
<th>Interval</th>
<th>Transfer</th>
<th>Bandwidth</th>
<th>Retr</th>
<th>Role</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>0.00-10.00 sec</td>
<td>2.98 GBytes</td>
<td>2.56 Gbits/sec</td>
<td>0</td>
<td>sender</td>
</tr>
<tr>
<td>4</td>
<td>0.00-10.00 sec</td>
<td>2.98 GBytes</td>
<td>2.56 Gbits/sec</td>
<td>0</td>
<td>receiver</td>
</tr>
</tbody>
</table>

```bash
server: iperf3 -s
client: iperf3 -c {server}

server: iperf3 -s -f filename
client: iperf3 -c {server} -f filename
```
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/)

* Science DMZ is a trademark of The Energy Sciences Network (ESnet)
Discover MTU issues after upgrading to 100GE
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/)

---

* Science DMZ is a trademark of The Energy Sciences Network (ESnet)
Campus Outreach

- Research Computing Governance Committee (RCGC) Network Subcommittee
  - Research faculty
  - UITS network and research staff
Security

- Router ACLs
- Host hardening
  - Well defined traffic patterns
- IDS