Distributed Virtual Network Operations Center (DVNOCC) - Towards Federated & Customer-focused Cyberinfrastructure

Harika Tandra, Software Engineer
GLORIAD

(presentation based on slides prepared by Greg Cole, Principal Investigator, GLORIAD project)
What is GLORIAD?

- A cooperative R&E network ringing the northern hemisphere linking scientists, educators and students in Russia, USA, China, Korea, Netherlands, Canada, the Nordic countries – and soon India, Egypt, Singapore – and others with specialized network services; co-funded, co-managed by all international partners.

- Variously sized circuits/services around the northern hemisphere.

- Hybrid circuit-(L1/L2) and packet-switched services(L3).

- Collaborative International Program to Develop/Deploy advanced Cyberinfrastructure and applications between partnering countries (and others) as effort to expand science, education and cultural cooperation and exchange.
GLORIAD mission

- Connecting the unconnected
- Better informing science and education community (and general public) about global opportunities for collaboration
- Promoting decentralized, distributed, transparent and open approach to global R&E networking
Addresses need for all levels of cyberinfrastructure operators (and users) to collaborate on decentralized, distributed and reliable operations of links and services.

Focus on customer-based performance.

Large development effort on part of Chinese, Dutch, Korean, Nordic and US (and we hope, soon, other national) GLORIAD teams.
DVNOC Contd..

- Web based application
- Developed using Flash/Flex platform
- Current version: [http://viz.gloriad.org/dvnoc/dvnoc.html](http://viz.gloriad.org/dvnoc/dvnoc.html)
DVNOC - GLORIAD Earth Tab

Title: Grazing and Iron Controls of Diatom Blooms in the Arabian Sea

The Arabian Sea is important in global C and N budgets because of its high rates of annual primary production, its extensive zone of oxygen depletion and denitrification, and its expected strong response to global warming via ocean-atmosphere feedback to monsoon winds and upwelling intensity, especially the Oman Upwelling, driven by the SW Monsoon. In 2007, one of the investigators found that a large region of the central and southern Arabian Sea was Fe-limited during the SW Monsoon, consistent with 3D is powered by Papervision3D
We’re trying to shift towards “customer-based performance” in all areas of cyberinfrastructure deployment.
"Needle" chart i.e., a blue needle (topped by a black marker) illustrates one flow.

3-D plot of throughput, loss & RTT using flow data from US to CSTNET over a 24hr period on GLORIAD network.

Wednesday, February 9, 2011
Problem: network operators have insufficient knowledge of nor relationship with each other (local/campus, regional, national, global operators) (and R&E customers less so)

Solution: encourage common view towards customer-based performance, lead effort towards community-developed shared performance measurement instrumentation and tools for joint engineering management (dvNOC)

(we will realize many other benefits from this community-building exercise)
Emphasis on Customer Performance

- We wish to know of individual customer-based performance problems before customer can call.
- We’re developing statistically important base of information about where there are weaknesses in our global/regional/regional/local networks.
- Based primarily (at moment) on measurements of packet retransmits.
Automated system to debug under-performing flows in wide area networks
Throughput vs Loss (contd..)

- We can see that the decrease in rate is steeper with the increase in loss than the increase in RTT.

- Half the loss rate gives throughput increase of $\approx 41\%$.

3-D plot of throughput derived from loss & RTT using Mathis formula.

X-axis: %loss  
Y-axis: RTT (ms)  
Z-axis: throughput in bits/sec
Hybrid monitoring/data collection system

1. Passive monitoring sub-system: Filters network flow data to identify under-performing flows
2. Active monitoring sub-system: Collects performance statistics of individual routers

**All the IPs are anonymized in the following slides**
Passive monitoring sub-system: Flow filter

- % retransmissions per bytes transferred > 0.01
- Bytes transferred > 5 MB
- Frequency > 4 hours. Same (ip_s, ip_d) pair is not labeled as under-performing for the minimum time period set by the frequency parameter
Filter the netflow records to identify underperforming flows

<table>
<thead>
<tr>
<th>ip_src</th>
<th>ip_dst</th>
<th>MB</th>
<th>%rtpct</th>
<th>starttime</th>
<th>endtime</th>
</tr>
</thead>
<tbody>
<tr>
<td>xx.xx.77.70</td>
<td>xx.xx.138.244</td>
<td>0.2137</td>
<td>0</td>
<td>2009-10-18 20:53:43</td>
<td>2009-10-18 20:58:50</td>
</tr>
<tr>
<td>xx.xx.16.49</td>
<td>xx.xx.4.71</td>
<td>0.2101</td>
<td>0</td>
<td>2009-10-18 20:53:42</td>
<td>2009-10-18 20:58:51</td>
</tr>
</tbody>
</table>

MB - MBytes transferred, %rtpct - Percentage retransmissions per byte
Active monitoring sub-system

For each under-performing flow identified, MTR runs are triggered to source and destination IPs

Triggered in near-real-time to the flow detected. Thus, test packets are triggered in network conditions similar to those seen by the real traffic

Combining the two gives approximate end-to-end performance
## Data collected

Result of MTR runs to source and destination of an under-performing flow

<table>
<thead>
<tr>
<th>ip_s</th>
<th>ip_d</th>
<th>MBytes</th>
<th>rtpct</th>
<th>starttime</th>
<th>endtime</th>
<th>keyid</th>
</tr>
</thead>
<tbody>
<tr>
<td>xx.3.226</td>
<td>xx.244.210</td>
<td>7.31</td>
<td>0.987</td>
<td>2009-10-18 20:56:50</td>
<td>2009-10-18 20:58:51</td>
<td>28995</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>serial_n</th>
<th>node_ip</th>
<th>loss_pct</th>
<th>packets_s</th>
<th>avg_rtt</th>
<th>best_rtt</th>
<th>wrst_rtt</th>
<th>target_ip</th>
<th>masterkeyid</th>
<th>target_lbl</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>192.31.99.97</td>
<td>0</td>
<td>50</td>
<td>4.7</td>
<td>0.4</td>
<td>12.8</td>
<td>xx.244.210</td>
<td>28995</td>
<td>Destination</td>
</tr>
<tr>
<td>2</td>
<td>192.31.99.146</td>
<td>0</td>
<td>50</td>
<td>2.3</td>
<td>1.3</td>
<td>17.7</td>
<td>xx.244.210</td>
<td>28995</td>
<td>Destination</td>
</tr>
<tr>
<td>3</td>
<td>216.24.186.5</td>
<td>0</td>
<td>50</td>
<td>28.8</td>
<td>27.2</td>
<td>49.3</td>
<td>xx.244.210</td>
<td>28995</td>
<td>Destination</td>
</tr>
<tr>
<td>4</td>
<td>192.43.217.137</td>
<td>0</td>
<td>50</td>
<td>28.6</td>
<td>26.3</td>
<td>62.8</td>
<td>xx.244.210</td>
<td>28995</td>
<td>Destination</td>
</tr>
<tr>
<td>5</td>
<td>192.43.217.114</td>
<td>0</td>
<td>50</td>
<td>27.2</td>
<td>27.1</td>
<td>27.7</td>
<td>xx.244.210</td>
<td>28995</td>
<td>Destination</td>
</tr>
<tr>
<td>6</td>
<td>128.117.243.75</td>
<td>0</td>
<td>50</td>
<td>27.8</td>
<td>27.2</td>
<td>41.7</td>
<td>xx.244.210</td>
<td>28995</td>
<td>Destination</td>
</tr>
<tr>
<td>7</td>
<td>???</td>
<td>100</td>
<td>50</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>xx.244.210</td>
<td>Destination</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>serial_n</th>
<th>node_ip</th>
<th>loss_pct</th>
<th>packets_s</th>
<th>avg_rtt</th>
<th>best_rtt</th>
<th>wrst_rtt</th>
<th>target_ip</th>
<th>masterkeyid</th>
<th>target_lbl</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>192.31.99.97</td>
<td>0</td>
<td>50</td>
<td>5.3</td>
<td>0.4</td>
<td>13.5</td>
<td>xx.3.226</td>
<td>28995</td>
<td>Source</td>
</tr>
<tr>
<td>2</td>
<td>192.31.99.166</td>
<td>0</td>
<td>50</td>
<td>189.9</td>
<td>189.8</td>
<td>196.3</td>
<td>xx.3.226</td>
<td>28995</td>
<td>Source</td>
</tr>
<tr>
<td>3</td>
<td>159.226.254.165</td>
<td>0</td>
<td>50</td>
<td>190.2</td>
<td>189.9</td>
<td>203.6</td>
<td>xx.3.226</td>
<td>28995</td>
<td>Source</td>
</tr>
<tr>
<td>4</td>
<td>159.226.254.253</td>
<td>0</td>
<td>50</td>
<td>228.8</td>
<td>228.8</td>
<td>229.1</td>
<td>xx.3.226</td>
<td>28995</td>
<td>Source</td>
</tr>
<tr>
<td>5</td>
<td>159.226.254.29</td>
<td>0</td>
<td>50</td>
<td>230.8</td>
<td>228.9</td>
<td>317.8</td>
<td>xx.3.226</td>
<td>28995</td>
<td>Source</td>
</tr>
<tr>
<td>6</td>
<td>159.226.254.190</td>
<td>2</td>
<td>50</td>
<td>229.6</td>
<td>229</td>
<td>254.1</td>
<td>xx.3.226</td>
<td>28995</td>
<td>Source</td>
</tr>
<tr>
<td>7</td>
<td>159.226.254.170</td>
<td>4</td>
<td>50</td>
<td>229.4</td>
<td>229.2</td>
<td>229.8</td>
<td>xx.3.226</td>
<td>28995</td>
<td>Source</td>
</tr>
<tr>
<td>8</td>
<td>159.226.46.230</td>
<td>2</td>
<td>50</td>
<td>229.6</td>
<td>229.4</td>
<td>230.3</td>
<td>xx.3.226</td>
<td>28995</td>
<td>Source</td>
</tr>
<tr>
<td>9</td>
<td>???</td>
<td>100</td>
<td>50</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>xx.3.226</td>
<td>Source</td>
</tr>
</tbody>
</table>
Network graphs show individual router behavior cutting across several MTR runs, to different target IPs. Thus, giving a snapshot of network router topology seen by the under-performing flows.
Example network graphs for a few end hosts in U.S.
A faulty node

$r_2$ is defined as a faulty node if

probability of loss ($l_i/t_i$) is high and is uniformly distributed across all its branches

$l_i = \# \text{ of runs via } r_2 \text{ to } r_i$ seeing loss

$t_i = \text{total \# of runs via } r_2 \text{ to } r_i$
Network Graph analysis

- Developed cost functions to learn the probability of each node being faulty
- Supervised pattern classification algorithms are used to learn the accuracy of the cost functions
Example network graphs for a few end hosts in China

Representation:
- Graph node - router in paths discovered by MTR.
- Rect. node - the end host.
- Node label -
  - 1st line - value of cost function
  - 2nd line - IP (anonymized)
  - 3rd line - Avg. % packet loss at the node.
- Color map ranges from Yellow through orange to red.
  - this graph is color mapped based on the 'Avg. % packet loss' value.
- Edges labels: ‘A-B’ where
  - A => Total number of mtr runs through the parent to child node.
  - B => Number of runs in which there was non-zero packet loss.
- Gray nodes are nodes which saw no packet loss.
Network-monitoring data collection
Packeteer box at Chicago

- Passively monitors traffic to/from GLORIAD router in Chicago
- Exports extended Netflow records
- Bytes retransmitted
- Application classification

**Replacing** Packeteer with open source monitoring box

- Commercial box
- Limited to 1G line speed
GOALS

(Network utilization and performance measurement box - running at least at 10G line speed)

(Emit extended netflow records including retransmissions, application classification and more)

HARDWARE

(Dell PowerEdge R410 Server - 8 core intel processor)

(10GE Intel Fiber Card)
Nprobe is open source software developed by Luca Deri (http://www.ntop.org/nProbe.html)

Development effort is in progress with help of Luca Deri and CSTNet (GLORIAD-China partners)

Current version exports retransmissions data

Next steps: Better application classification
Integrating data from other tools
Currently deployed at Seattle, Chicago and Singapore

Soon nodes will be installed in Amsterdam and Hong Kong

Looking for ways to integrate/visualize perfsonar data in DVNOC
Conclusion

- Common platform to share network operations, utilization, performance and security data
- Addresses “disconnect” between all the different levels of network operators
Thank you.