The CAIDA bandwidth estimation testbed

Margaret Murray
CAIDA/SDSC/UCSD

9 Dec 2003
Acknowledgements

- Nevil Brownlee: NeTraMet config
- Johnny Chang: data collection
- Tony Lee, Tuan Le: autotest.pl, autoplot.pl
- Jiri Navratil, Ravi Prasad, Vinay Ribeiro: remote testbed users
- Grant Duvall, Nate Mendoza: router config
- Kevin Walsh: CalNGI, NPRL access
  - Spirent SmartBits 6000 with SmartFlow software
  - Foundry Big Iron router

- Cisco: GSR12008 router
- Juniper: M20 router
- Endace: gigE DAG card for passive monitoring with NeTraMet
- Department of Energy SciDAC grant DE-FC02-01ER25466
The e2e tool grail...

- Users want to push a button to:
  - select the best available e2e path
  - optimize application utilization of bandwidth
- Sponsors expect growth in usage and need for better sharing/planning

...yet many existing high-speed backbones appear to be lightly loaded, and many applications can not take advantage of fatter pipes...
# Current e2e Tools

<table>
<thead>
<tr>
<th>Tool Class</th>
<th>Tool</th>
<th>Authors</th>
<th>Methodology</th>
<th>Tool</th>
<th>Authors</th>
<th>Methodology</th>
</tr>
</thead>
<tbody>
<tr>
<td>Per-hop Capacity</td>
<td>clink</td>
<td>Downey</td>
<td>VPS</td>
<td>pathchar</td>
<td>Jacobson</td>
<td>VPS</td>
</tr>
<tr>
<td></td>
<td>pchar</td>
<td>Mah</td>
<td>VPS</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>End-to-End Capacity</td>
<td>bprobe</td>
<td>Carte</td>
<td>pkt pair</td>
<td>pathrate</td>
<td>Dovrolis-Prasad</td>
<td>pkt pairs,train</td>
</tr>
<tr>
<td></td>
<td>nettimer</td>
<td>Lai</td>
<td>pkt pairs</td>
<td>sprobe</td>
<td>Saroiu</td>
<td>pkt pairs</td>
</tr>
<tr>
<td>End-to-End Available</td>
<td>ABw</td>
<td>Navratil</td>
<td>unknown</td>
<td>netest</td>
<td>Jin</td>
<td>unknown</td>
</tr>
<tr>
<td>Bandwidth</td>
<td>cprobe</td>
<td>Carter</td>
<td>pkt trains</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>IGI</td>
<td>Hu</td>
<td>SLoPs</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bulk Transfer Capacity</td>
<td>cap</td>
<td>Allman</td>
<td>emulate TCP tput</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>treno</td>
<td>Mathis</td>
<td>std TCP tput</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Achievable TCP Throughput</td>
<td>iperf</td>
<td>NLANR</td>
<td>TCP connect</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Netperf</td>
<td>NLANR</td>
<td>TCP connect</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>tcp</td>
<td>Muuss</td>
<td>TCP connect</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Why is it so hard to measure e2e performance?

• Are apps tuned for high-speed paths?
• Router slow paths, load balancing
• MTU mismatch
• NIC interrupt coalescence
• Host OS variations
• Router OS variations
• the list goes on...
Why use a bwest testbed?

• Use reproducible test conditions
• Test against saturated links
• Test “black box” e2e tools against same scenarios
  – Identify conditions where tools work well
  – Give developers an environment for refining their tools
Why use CAIDA/SDSC testbed?

• Take advantage of CAIDA and CalNGI vendor contacts and equipment
• Integrate with CAIDA passive monitoring technology (NeTraMet, CoralReef)
• Possibilities in the future: connect testbed to specific networks (via SDSC)
OC48/gigE 4hop configuration

SmartBits traffic gen
Cisco router
Juniper M20 router

end host
regen tap
passive monitor
Foundry router
e2e path
two loops
outer loop
OC48 link
gigE link
SmartFlow Application
Spirent Communication Smartbits 6050B for cross-traffic generation
## Working with Traffic (real vs. generated)

### Real Traffic

<table>
<thead>
<tr>
<th>Pro:</th>
<th>Con:</th>
</tr>
</thead>
<tbody>
<tr>
<td>• no question of validity</td>
<td>• hard to find, store and use traces of interest</td>
</tr>
</tbody>
</table>

### Generated (Simulated) Traffic

<table>
<thead>
<tr>
<th>Pro:</th>
<th>Con:</th>
</tr>
</thead>
<tbody>
<tr>
<td>• easily reproducible scenarios</td>
<td>• how realistic is traffic?</td>
</tr>
</tbody>
</table>
Comparison of Packet IATs (CCDF: 1 min. samples, % vs. IAT usec)

- **Red** = BB OC48 (Left: 246M, 9%) (Right: 686M, 27%)
- **Blue** = UCSD gigE (Left: 40M, 4%) (Right: 240M, 24%)
- **Pink** = bwest lab (Left: 593M, 59%) (Right: 589M, 59%)

9 Dec 2003
Comparison of Pkt Sizes
( CCDF: 1 min samples, % vs. PktSize b )

- **Red** = BB OC48  (Left: 246M, 9%)  (Right: 686M, 27%)
- **Blue** = UCSD gigE  (Left: 40M, 4%)  (Right: 240M, 24%)
- **Pink** = bwest lab  (Left: 593M, 59%)  (Right: 589M, 59%)

9 Dec 2003  BEst ISMA CAIDA/SDSC
Cross-traffic

• Tool developers told us that generated traffic needs enough packet dispersion to allow dynamics to manifest: We do that.
• SmartFlow traffic has appropriate protocol headers, but does NOT emulate TCP congestion control.
• While most current networks do not implement jumbo MTUs, they will, and we can generate them.
100M 3hop test

host -> Juniper -> Foundry -> host
("to" direction; likely has deeper queues)

Bandwidth Estimation Tool Measurement (Mbps) vs. Cross-traffic Load (%)
100M 3hop test surprise!

**Diagram:**

- **host -> Foundry -> Juniper -> host**
- "from" direction: likely has shallower queues

**Graph:**

- **Bandwidth Estimation Tool Measurement (Mbps)**
- **Cross-traffic Load (%)**

Legend:
- netest
- treno
- iperf
- Expected measurement
Hypothesis:

router queue depth affects e2e tools

• Juniper M20
  • 9M

...we plan to run single router tests to further investigate why

• Foundry Big Iron
  • 1M
OC48/gigE 4hop “to” direction

- Three tools give decreasing measurements as cross-traffic increases.
  - netest
  - pathchirp
  - pathload

- Note: pathchirp and pathload return multiple results
OC48/gigE 4hop “from” direction

- Different results in opposite direction for
  - netest
  - pathchirp
- Hypothesis: due to shallower router queues
- Note multiple results for pathchirp and pathload
gigE/OC48: 4hop test again
only pathload returns similar results in both directions

• Juniper M20
  - 9M
  ...we plan to run single router tests to further investigate this phenomenon.

• Foundry BigIron
  - 1M

9 Dec 2003
gigE/OC48: 4hop test

...more investigation needed to understand tool behavior

- **Juniper M20**
  - 9M

- **Foundry Big Iron**
  - 1M
OC48/gigE 4hop configuration

- SmartBits traffic gen
- Juniper M20 router
- Cisco router
- end host
- regen
tap
- passive monitor
- Foundry router

Cisco 12000
NetOptics Regen Tap
NetTraMet
Juniper
Foundry Big Iron

SmartFlow Application
Spirent Communication Smartbits 6000B for cross-traffic generation

Outer loop
two loops
e2e path
gigE link
OC48 link
Remote Access to Testbed

- ssh wednesday.sdsc.edu, then
  - ssh gomez
  - ssh fester
- VNC control of SmartFlow
  - several config files available, or
  - create your own
- /usr/local/bwest/
  - autotest.pl can run and time one or all tools
  - autoplot.pl can plot IAT and Packet Size distribution
How can the CAIDA bwest testbed address e2e problems?

• generate traffic; saturate high-speed paths
• experiment with different MTU settings
• evaluate different NICs
• try different end host OS configs
• evaluate impact of different routers
• isolate router slow paths, load balancing
• the list goes on...
For more information...

• Contact Marg
  - marg@caida.org
  - 858 534-8928

• Scheduling
  - first come, first served, or ???

• Support
  - DOE grant ends Aug 2004