Performance Metrics for Multilayer Switches

Session 2802
Objective

To understand the how to measure a device’s network ability to protect application networking resource performance and scalability.

Agenda

- Definition of Terms
- Performance Test Tools
- Testing Methods and Metrics
- Example—Buffering and Architecture
- Example—Flow Lookup and Architecture
- Example—WRR and Architecture
- Example—SLB and Architecture
**Definition of Terms**

- **Device under test (DUT)**—A single stand-alone or modular unit which receives frames on one or more of its interfaces and then forwards them to one or more interfaces according to the addressing information contained in the frame.

**Definition of Terms (Cont.)**

- **System Under Test (SUT)**—The collective set of network devices to which stimulus is offered as a single entity and response measured.
**Definition of Terms (Cont.)**

- **Queuing**—buffer mechanisms used to control congestion by temporarily storing packets in memory.

- **Latency (RFC 1944)**—Time in microseconds which a DUT takes to transmit one packet with same size packet at X rate.
Definition of Terms (Cont.)

• Latency (RFC 1944)—Time in microseconds which a DUT takes to transmit one packet with the same size packet at X rate

Definition of Terms (Cont.)

• Latency (RFC 2285)—Time in microseconds which a DUT takes to transmit all packets with sequence and latency tags
Definition of Terms (Cont.)

- **Latency (RFC 2285)**—Time in microseconds which a DUT takes to transmit all packets with sequence and latency tags.

- **Throughput**—Highest rate in packets-per-second through a DUT which frame loss does not occur.
Definition of Terms (Cont.)

- **Frame loss**—Percentage of lost frames when running at said frame size often confused with throughput.

Definition of Terms (Cont.)

- **Back-to-back burst**—The minimum interval between bursts which the DUT can process with no frame loss.
Definition of Terms (Cont.)

**Depth of inspection**—Amount of packet information processed to make a switching decision

128-->1518+ Layer 7
- 48 bytes Layer 4
- 32 bytes Layer 3
- 12 bytes Layer 2

Agenda

- Definition of Terms
- Performance Test Tools
- Testing Methods and Metrics
- Example—Buffering and Architecture
- Example—Flow Lookup and Architecture
- Example—WRR and Architecture
- Example—SLB and Architecture
Application Testing

- Only true application test
- User “feel” valuable but intangible
- No standard metrics
- Not wire speed

Appliance Testing

- End-to-End Latency
- Multi-user
- Multi-network
  - GPS SYNC
- Great for L3 & L2
- OK for L4
Why Ping-Pong is all Wrong

- Single user
- Single network
- Single Sample
- Asynchronous Routing Issues
- Minimal accuracy
- \( A_1 + A_2 + B + C + D + E = RTT \)

Multi-user
- Multi-network
- Multi-Sample
- End-to-End Latency
- GPS timing
- \( A_1 + B + C + D = RTT_1 \)
- \( D + E + A_2 = RTT_2 \)

*Netcom’s SmartISP™*
Verify your Service

- Service Level Agreements
- CIR-Committed Information rate

With Respect to time
- Latency & Deviation
- Throughput & Deviation

Performance Test Tools

PC Test Tool
- Application to Application Measurement
- A Few Scripts
- Metrics Unknown
- Metrics Unknown
- Metrics Unknown
- Not Wire Speed
- No Accuracy
- No Network Measurement
- Only Round Trip Times

Appliance
- Device to Device Measurement
- Scales to Many Users
- Setup Timing Etc
- Latency, Jitter Etc.
- No OS Delay
- Thru’Put, Loss, Etc.
- Wire Speed Nanosec Accuracy
- Full Visibility of Network Elements
- One Way Times, Accurate GPS Timing
Performance Test Tools Links

**PC Test Tool**
- LanMark Pro™
- SmartWindows™
- SmartFlow™
- SmartTCP™

**Appliance**
- etestinglabs
- WebBench™
- NetBench™
- TCLGui™

Application vs. Performance Test Tool

- **Application Tool**
  - Not Scalable
  - Real World

- **Performance Tool**
  - Scalable
  - Not Real World

Combine and Prosper
### Types of Tests

#### Type of Test  Typical Tests

1. **Performance**
   - Tests for delays in sending data or setting up calls
   - Tests for variation in delay and effect on applications
   - Tests for user application responsiveness
   - Tests use of theoretical bandwidth

2. **Capacity**
   - Tests number of addresses/flows supported

3. **Correct Operation**
   - Tests data integrity and data loss/ RFC conformance
   - Tests of device/network functionality under load

4. **Resilience**
   - Network ability to recover from errors or changes
   - Map to Service Level Agreement

5. **Security**
   - Tests loss of security under load
   - Tests loss of security based on port scanners
   - Tests loss of security based on new attacks

### Agenda

- Definition of Terms
- Performance Test Tools
- Testing Methods and Metrics
- Example—Buffering and Architecture
- Example—Flow Lookup and Architecture
- Example—WRR and Architecture
- Example—SLB and Architecture
Resource Constraints and Bottlenecks

Application Characteristics

- Buffering, queuing, and latency
- Interface and pipe sizes, speed
- Control plane, IP flow speed (if flow-based)
- Memory, IP flow capacities, buffering capacities, control plane and data plane

Relational Perflow Metrics

Any One Change Affects the Other
Relational Layer 2 Metrics

- MAC Addresses
- Frame Size
- Packets-Per-Second
- Latency

Relational Layer 3 Metrics

- IP Addresses
- IP Flows
- IP Flow Setup Rate
- Latency
- Sequence
- Layer 2

© 2000, Cisco Systems, Inc.
Relational Layer 4 Metrics

- TCP/UDP Flow Setup Rate
- Latency
- Sequence
- TCP/UDP Session Capacity

+ Layer 3

Relational Layer 7 QoS

- Application Recognition Rate
- Latency
- Sequence
- Application Definitions

+ Layer 4
Relational Layer 7 Termination

Latency Sequence

Terminated Sessions

Layer 7QoS

Application Termination Rate

Agenda

• Definition of Terms
• Performance Test Tools
• Testing Methods and Metrics
• Example—Buffering and Architecture
• Example—Flow Lookup and Architecture
• Example—WRR and Architecture
• Example—SLB and Architecture
Buffering and Switching Architectures

- Congestion management
- Switching decision

Congestion Management

- Required when multiple ports are contending for the same port
- Congestion occurs during short bursts
Closet/Server Buffering

Remember that all networks have hot spots of congestion
Output ports only transmit one frame-at-a-time

High-Speed Buffering

1 Gigabit Ethernet port at 100% fills a 500 KB buffer in:
4.1 ms with 1500 Byte MTU
5.5 ms with 64 Byte MTU
Instantaneous Over Subscription

Even if bandwidth of two input ports are ½ the speed of the output port congestion will occur!

Dynamic Buffer Queuing

- Each buffer element fixed in small increments (for example, 64 bytes each)
- Allows for efficient use of buffers
- More expensive
Fixed Buffer Queuing

- Buffer element is a fixed size
- Inefficient use of buffers
- Less controller overhead by 31 times—easier to make a faster switch!

Dynamic vs. Fixed Buffering

**Dynamic**
- Ethernet—Frame Size
- Voice
- Data Throughput

**Fixed**
- ATM—Cell Size
- Voice
- Data Requires SAR

Large—Small Buffers Available

Only Small Buffers Available
Buffering Tests 2:1

**Backbone Test**

- Two 1000 Mbps—one 1000 Mbps
- Burst 100–3000 frames for 2:1 gigabit Ethernet ports at 5–50% utilization

---

Buffering Capacity

**Packets Buffered**

- 64 Byte
- 512 Byte
- 1518 Byte

---

Dynamic 500K Buffer
64 byte Elements

Fixed 500K Buffer 2
Kbyte Elements
Buffering Capacity

Buffering Test 2:1

Device Under Test

Output Port at 100% Utilization 96 nsec IFG
Backplane to Port “Down Speed” reduces IFG
Input ports at 50% Utilization 768 nsec IFG (Interframe-Gap)*

*Note: Input ports should vary 5–50% per trial thus output utilization 10–100%
Buffering Test 2:1

Packet Blaster

Output Port at 100% Utilization 96 nsec IFG (Inter-frame-Gap)

Burst 3000 packets with 96 nsec IFG (Inter-frame-Gap)

Fixed Queuing Latency Distribution 2:1

Number of Packets

- 10% Input
- 20% Input
- 30% Input
- 40% Input
- 50% Input

Latency Distribution

<=10uSec

<=100uSec

0

1000000

2000000

3000000

4000000

5000000

6000000
Buffering Tests 40:1

**Closet Test**
- 40 100 Mbps—One 1000 Mbps
- Burst 10–30 frames for 20:1 100 Mbps Ethernet ports at 1–5% utilization

**Test Results**—64 bytes 10 ppb

![Bar chart](chart.png)

- Red: Dynamic Buffering
- Orange: Fixed Buffers

% Throughput

Number of Ports

- 15
- 20
- 25
- 30
- 35
- 40
Test Results—
500 bytes 10 ppb

Test Results—
1518 bytes 10 ppb
Test Results—64 bytes 24 ppb

Number of Ports

Test Results—500 bytes 24 ppb

Number of Ports
Test Results—
1518 bytes 24 ppb

Agenda

- Definition of Terms
- Performance Test Tools
- Testing Methods and Metrics
- Example—Buffering and Switch Architecture
- Example—Flow Lookup and Architecture
- Example—WRR and Architecture
- Example—SLB and Architecture
Flow Lookup and Switch Architectures

- Congestion management
- Switching decision

Route Caching (Flow-Based Switching)

- First packet processed switched by route processor
- Flow cache enabled (centrally and/or locally)
**CPU-Control Plane**

- Run routing protocols
- Maintain routing tables
- Check for CLI commands
- Accounting counters
- ICMP
- Queue packets, etc...

**CPU-Control Plane Issues**

- Control Plane must manage flow setup and system management
- What happens to Spanning Tree, VRRP and routing updates if the CPU is busy processing flows?

---

**Cisco Express Forwarding (FIB-Based Switching)**

- No process switching necessary
- FIB based on routing table entries, not traffic flows
- FIB can be kept central or distributed
IP-Host Flow Test

1000 IP Flows

Input Port Cache

IP Destination 4 bytes

Mask 4 bytes

Output Port ID

1000 IP Flows

Device Under Test

1.1.X.X
255.255.0.0

1.2.X.X
255.255.0.0

System Under Test

1.1.X.X
255.255.0.0

1.2.X.X
255.255.0.0

1.3.X.X
255.255.0.0

1.4.X.X

IP-Recursive Flow Test

1000 IP Flows

Input Port Cache

IP Destination 4 bytes

Mask 4 bytes

Output Port ID

1000 IP Flows

Device Under Test

Subnet-Mask 255.255.0.0

System Under Test

1.1.X.X
1.2.X.X
1.3.X.X
1.4.X.X
Flow Test Results—Loss

Packets Transmitted

Number of Frames Received out of 14000

Flow Test Results—Jitter

Latency in Microseconds

Maximum Latency  Minimum Latency
## Flow Test Results—Raw using Smart Windows

<table>
<thead>
<tr>
<th>L3Tests 1</th>
<th>Port</th>
<th>VTE</th>
<th>Sequence</th>
<th>TxTime</th>
<th>RxTime</th>
<th>DUT-A</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>632079</td>
<td>640709</td>
<td>863.0</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>2</td>
<td>0</td>
<td>632149</td>
<td>642458</td>
<td>1030.9</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>3</td>
<td>0</td>
<td>632220</td>
<td>644057</td>
<td>1183.7</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>4</td>
<td>0</td>
<td>632290</td>
<td>645658</td>
<td>1336.8</td>
</tr>
<tr>
<td>5</td>
<td>1</td>
<td>5</td>
<td>0</td>
<td>632361</td>
<td>647272</td>
<td>1491.1</td>
</tr>
<tr>
<td>6</td>
<td>1</td>
<td>6</td>
<td>0</td>
<td>632431</td>
<td>648887</td>
<td>1645.6</td>
</tr>
<tr>
<td>7</td>
<td>1</td>
<td>7</td>
<td>0</td>
<td>632501</td>
<td>650607</td>
<td>1810.6</td>
</tr>
<tr>
<td>8</td>
<td>1</td>
<td>8</td>
<td>0</td>
<td>632572</td>
<td>652236</td>
<td>1966.4</td>
</tr>
<tr>
<td>9</td>
<td>1</td>
<td>9</td>
<td>0</td>
<td>632642</td>
<td>653978</td>
<td>2133.6</td>
</tr>
<tr>
<td>10</td>
<td>1</td>
<td>10</td>
<td>0</td>
<td>632713</td>
<td>655607</td>
<td>2289.4</td>
</tr>
<tr>
<td>11</td>
<td>1</td>
<td>11</td>
<td>0</td>
<td>632783</td>
<td>657360</td>
<td>2457.7</td>
</tr>
<tr>
<td>12</td>
<td>1</td>
<td>12</td>
<td>0</td>
<td>632853</td>
<td>658982</td>
<td>2612.9</td>
</tr>
<tr>
<td>13</td>
<td>1</td>
<td>13</td>
<td>0</td>
<td>632924</td>
<td>660781</td>
<td>2785.7</td>
</tr>
<tr>
<td>14</td>
<td>1</td>
<td>14</td>
<td>0</td>
<td>632994</td>
<td>662465</td>
<td>2947.1</td>
</tr>
</tbody>
</table>

## Flow Test Results—Setup Time

![Graph showing Flow Test Results—Setup Time]
Agenda

- Definition of Terms
- Performance Test Tools
- Testing Methods and Metrics
- Example—Buffering and Switch Architecture
- Example—Flow Lookup and Architecture
- Example—WRR and Architecture
- Example—SLB and Architecture

IP Weighted-Round Robin Queues Test

TX'r
- 4-1000 IP Flows
- ToS 00
- ToS 01
- ToS 10
- ToS 11

RX'r

Device Under Test
- 1.1.X.X 255.255.0.0
- 1.2.X.X 255.255.0.0

Input Port Cache and Buffer

IP Lookup and/or Classification

Queue Depth

WRR Scheduler
IP WRR Baseline
No Congestion

[Graph showing MicroSeconds vs. Time]

IP WRR
With Congestion

[Graph showing MicroSeconds vs. Time with annotations]

- Scheduling Latency in MicroSeconds
- Scheduling Interval in Number of Packets
Agenda

- Definition of Terms
- Performance Test Tools
- Testing Methods and Metrics
- Example—Buffering and Switch Architecture
- Example—Flow Lookup and Architecture
- Example—WRR and Architecture
- Example—SLB and Architecture

Centralized L4 Switching

- More memory = More connections
- Centralized State
- Centralized Persistence
- Physical port independent
Centralized L7 Switching

- One place to keep persistent information
  - Cookies
  - URL
  - SSL
- Works with and without interface caching
- Interface “punts” to CPU
- Cache examples:
  - Vito, C6k (Dispatch) IOS
  - SSLB & ASLB, Arrowpoint

Layer 4 Load Balancing

48 bytes Layer 4

MAC   IP   TCP/UDP

SYN   SYN ACK   ACK   HTTP Get
HTTP Get   HTTP Res.1
FIN

CPU

L4 Lookup

SYN   SYN ACK   ACK   HTTP Get
HTTP Get   HTTP Res.1
FIN
Layer 7 Load Balancing
HTTP 1.0 Single HTTP Get Per-Session

Layer 7 Load Balancing
HTTP 1.1 Multi HTTP Get Per-Session
Layer 4 Testing-SmartTCP™

Layer 4 Testing-SmartTCP™

3-Packet Connection Setup Time - Number of Connections

Test Completes as a BURST of connections NOT continuous
Layer 4 Testing-SmartTCP™

### Available SmartCards

<table>
<thead>
<tr>
<th>Card ID</th>
<th>Speed</th>
<th>Duplex</th>
<th>MAC Address</th>
<th>IP Address</th>
<th>Gateway</th>
<th>Response Delay (sec)</th>
<th>Auto-Negotiation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Client Port 1</td>
<td>100M</td>
<td>Full</td>
<td>00 00 00 00 00 00 01</td>
<td>172.020.040.001</td>
<td>000.000.000.000</td>
<td>0.96</td>
<td>Off</td>
</tr>
<tr>
<td>Client Port 2</td>
<td>100M</td>
<td>Full</td>
<td>00 00 00 00 00 00 02</td>
<td>172.020.040.011</td>
<td>000.000.000.000</td>
<td>0.96</td>
<td>Off</td>
</tr>
<tr>
<td>Client Port 3</td>
<td>100M</td>
<td>Full</td>
<td>00 00 00 00 00 00 03</td>
<td>172.020.040.021</td>
<td>000.000.000.000</td>
<td>0.96</td>
<td>Off</td>
</tr>
<tr>
<td>Client Port 4</td>
<td>100M</td>
<td>Full</td>
<td>00 00 00 00 00 00 04</td>
<td>172.020.040.031</td>
<td>000.000.000.000</td>
<td>0.96</td>
<td>Off</td>
</tr>
<tr>
<td>Server Port 5</td>
<td>100M</td>
<td>Full</td>
<td>00 00 00 00 00 00 05</td>
<td>172.020.040.020</td>
<td>000.000.000.000</td>
<td>0.96</td>
<td>Off</td>
</tr>
<tr>
<td>Server Port 6</td>
<td>100M</td>
<td>Full</td>
<td>00 00 00 00 00 00 06</td>
<td>172.020.040.020</td>
<td>000.000.000.000</td>
<td>0.96</td>
<td>Off</td>
</tr>
<tr>
<td>Server Port 7</td>
<td>100M</td>
<td>Full</td>
<td>00 00 00 00 00 00 07</td>
<td>172.020.040.020</td>
<td>000.000.000.000</td>
<td>0.96</td>
<td>Off</td>
</tr>
<tr>
<td>Server Port 8</td>
<td>100M</td>
<td>Full</td>
<td>00 00 00 00 00 00 08</td>
<td>172.020.040.020</td>
<td>000.000.000.000</td>
<td>0.96</td>
<td>Off</td>
</tr>
</tbody>
</table>

Layer 4 Testing-SmartTCP™

TCP Streams... | OK | Cancel
Layer 4 Testing-SmartTCP™

RATE x (NUMBER OF CLIENT CARDS) = TOTAL NUMBER OF CONNECTIONS PER SECOND

Layer 4 Testing (no data)

MAC IP TCP/UDP
48 bytes Layer 4

SYN → SYN
SYN ACK ← SYN ACK
ACK → ACK
FIN → FIN
FIN ACK ← FIN ACK
ACK → ACK

6-Packet Connection Setup Time - Number of Connections
Layer 4 Testing

- 2000 CPS per Client-Server Pair
- IP Destination 4 bytes
- TCP Port Sequence# Cookie URLID
- Output Port ID
- Cache
- Input Port
- Device Under Test
- 2000 CPS per Client-Server Pair

Layer 7 Testing 1.0 vs. 1.1

- MAC
- IP
- TCP/UDP
- HTTP GET WWW.CISCO.COM
- URL Lookup
- CPU
- SYN
- SYN ACK
- ACK
- HTTP Get1
- HTTP Res1
- HTTP Get2
- HTTP Res2
- FIN
- TCP Termination
- SYN
- SYN ACK
- ACK
- HTTP Get1
- HTTP Res1
- HTTP Get2
- HTTP Res2
- FIN
Layer 7 Testing 1.0 vs. 1.1

- Device Lookup Table
  - IP Destination: 4 bytes
  - Mask: 4 bytes
  - TCP Port Sequence#: Cookie
  - URLID

Device Under Test

Layer 7 Testing 1.0 vs. 1.1

- **WebStone**
- **WebBench** - Ziffdavis (400cps)
  - Best functionality tool
- **Tarantula** - Cisco/Arrowpoint (4000cps)
  - High Performance
- **Socrates**
  - Very Easy to setup
  - Good functionality tool
Performance Metrics for Multilayer Switches

Session 2802

Please Complete Your Evaluation Form

Session 2802