Gigabit Ethernet Design

Laura Jeanne Knapp
Network Consultant
1-919-254-8801
laura@lauraknapp.com
www.lauraknapp.com

Tom Hadley
Network Consultant
1-919-301-3052
tmhadley@us.ibm.com
Switched and Shared Ethernet

Shared Ethernet

Switched Ethernet
Ethernet Topology

- Gigabit Switch
- 100 Mbps Switch
- 10 Mbps Switch
- 10/100 Mbps Switch
- 10/100 Mbps Shared
- 10 Mbps Shared
**GbE Switch Performance Characteristics**

**Port filter rate**
Rate switch can receive, table look up/update, forward decision
Frames/second/port
Measured using worst case (minimum length frames)

**Forwarding rate**
Rate switch can actually forward frames
Frames/second/port
Less than or equal to the port filter rate

**Wire speed**
Filter/forward rate would need to be 1.5 million frames/second/port
Must make frame forwarding decision in under 700 ns
Workgroup switch less likely to see sustained wire speed traffic
Campus switch aggregates traffic and needs wire speed
GbE Switch Blocking Characteristics

Connection networks (telephone network)
- Blocking occurs when the offered load exceeds designed capacity
- Non-blocking means the capacity is greater than the worst case offered traffic

Connectionless networks (Ethernet network)
- A non-blocking switch’s capacity is greater than the sum of all the ports
- A blocking switch’s capacity is less than the sum of all the ports

Under-subscribing bandwidth
- In practice the internal capacity must be greater
- Prevents excess queuing delays or queuing overflow
- Common practice is to have a 2:1 ratio

This does not mean you won’t get congestion: a port may be blocked due to inadequate buffers, however, your switch has enough capacity to move the data
GbE A System Perspective

Disk I/O Bus

Application
Protocol
Device Driver

Per BUS

Disk I/O Bus

Application
Protocol
Device Driver

Per BUS

Campus Network

Networking - Connecting people to information through technology
# Gbe Ethernet and the Desktop

<table>
<thead>
<tr>
<th>Hardware</th>
<th>Average User</th>
<th>Performance User</th>
<th>Servers</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intel</td>
<td>Intel, RS/6000, SPARCstation.</td>
<td>Intel, SP2, 390 SPARCserver,</td>
<td>Intel, RS/6000, SPARCserver,</td>
</tr>
<tr>
<td></td>
<td>Intel, RS/6000, SPARCstation.</td>
<td>MIPS/SGI, ALPHA</td>
<td>MIPS/SGI, AS/400 ALPHA, RS/6000</td>
</tr>
<tr>
<td>I/O Bus</td>
<td>ISA/EISA/PCI</td>
<td>PCI, SBus, SGI</td>
<td>PCI, SBus, SGI</td>
</tr>
<tr>
<td>OS</td>
<td>Windows</td>
<td>Windows NT, Unix, Solaris, SunOS, Linux, AIX</td>
<td>Windows NT, OS/400 Netware, Unix, Solaris, SunOS, OS/390, AIX</td>
</tr>
<tr>
<td>Data Rate</td>
<td>10 Mbps</td>
<td>100 Mbps</td>
<td>100 Mbps - 1Gbps</td>
</tr>
<tr>
<td>Media</td>
<td>UTP</td>
<td>UTP cat 5</td>
<td>UTP cat 5 or fiber</td>
</tr>
</tbody>
</table>

Networking - Connecting people to information through technology
Single-threaded OS (windows) cannot take advantage of full duplex operation

Single-interrupt-per-frame drivers (NDIS) incur high overhead as network waits on OS to prepare frame

Data-link only NIC drivers cannot take advantage of performance enhancements
Gbe Desktop Operating System Changes

Move all checksums (CRC) to hardware (IP, TCP, UDP)
Offloads CPU
Packets with invalid checksum never reach the CPU

Newer bus interface logic allows interrupts to be avoided
Costly to implement
Suitable for servers and service machines

Use virtual-to-physical memory mapping
Relieves CPU processing
Extra NIC and operating system logic
Many disk drives today transfer at less than 50 Mbps

Disk I/O may only be 10 Mbps (reason for SAN - Storage Area Networks)

Speed dependent on physical organization of data

A head seek of 10ms is 1 million byte-times on a GbE

Look for interleaved drives and RAM disk caching

Server may have high performance disk subsystem, but what about the client?
GbE and the Protocol Stack

- CPU processing speed
- Efficiency of protocol being used
- Efficiency of protocol implementation on OS
- Operating system design (context switching, memory allocation, etc)
- Demands of other applications (multitasking, multithreading)
- Memory availability through pipelining
**GbE : System CPU Bus and NICs**

**CPU Bus**
- A 32 bit 33 MHz PCI bus can theoretically transfer 1 Gbps
- Contrived single long burst
- Practical transfer rate is 500-700 Mbps
- Also remember this is shared with other peripherals

**NICs**
- Direct Memory Access (DMA) is necessary
- FIFO overflow can be prevented using SRAM (more expensive)
- NICs should transfer multiple frames without interrupting host software
Upgrading Switch to Server Links

Focus on switched GbE
Make sure server hardware, OS, protocol stack and applications can support these speeds
Upgrading Switch to Switch links

While many 100 Mbps switches support a 1 Gbps uplink, check the switch for buffer availability and non-blocking capability with the Gbps addition.

Add 1 Gbps uplink to existing 100 Mbps switches
Upgrading Backbone to Gigabit Ethernet

Remove links between 100 Mbps switches

Add 1 Gbps links and connect to 1 Gbps switch
Redundant GbE Ethernet Backbone

Spanning Tree Algorithm still used to prevent loops
Recovery time for alternate link activation is unbearable in a GbE network
Need better mechanisms to have high availability GbE networks