

# G igit E themet and L inux: A P erform ance S tudy



Jeff Bain  
HP - Automated Storage

R&D Firmware Engineer  
jeff\_bain@hp.com

700 71st Ave.  
Greeley, CO 80634



## Some questions you might ask...

OUTLINE

- What's Gigabit Ethernet?
- What's in it for me?
- Will my Linux system really talk at 1000 Mbits/sec?
- What kind of speed should I really expect?
- How can I measure my system's speed?
- Ugh! Why is it so slow?
- How do I make my gigabit Linux system faster?
- What should I look for in the future?
- Okay I'm sold: Where do I get it?
- How much does it cost?

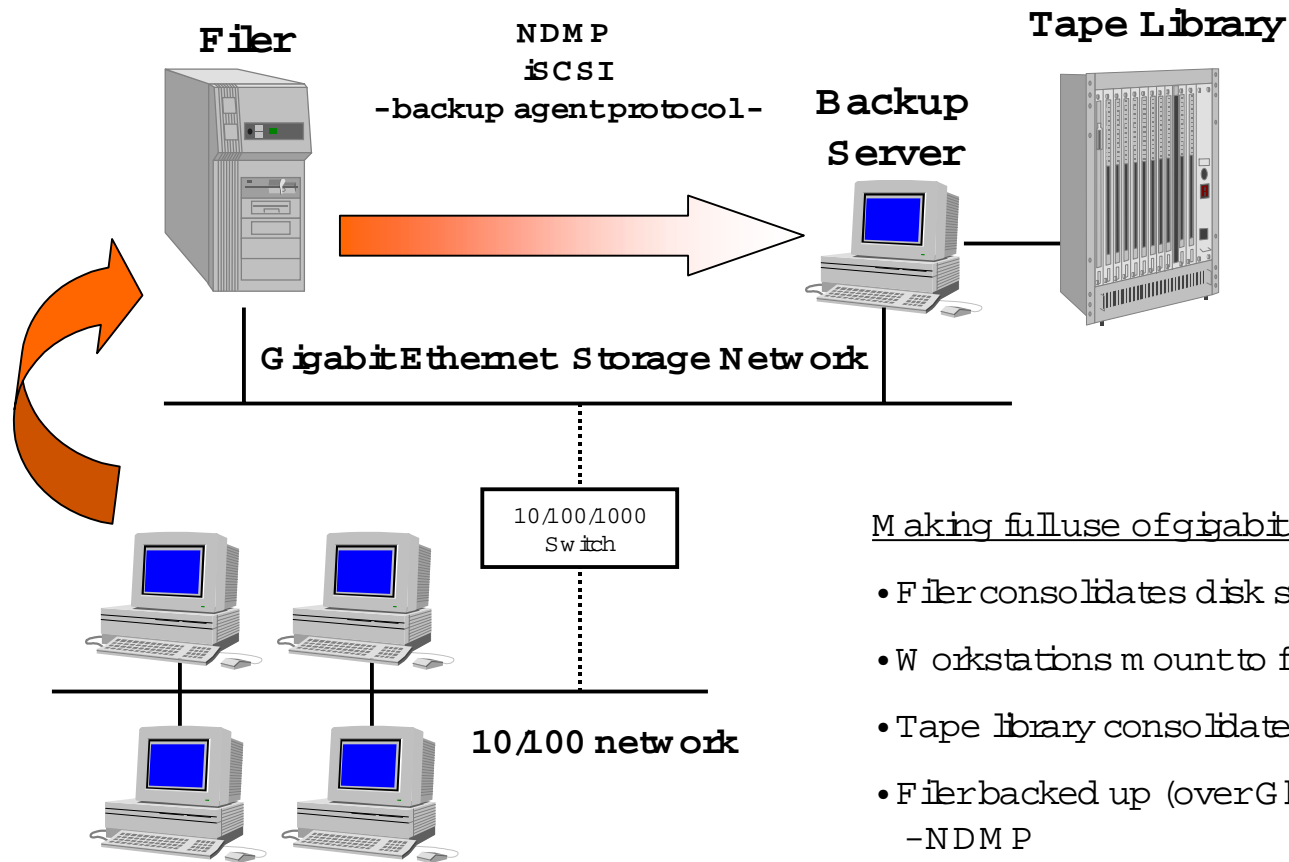
## W hat's G igit E themet?

- Fully backward compatible ethernet link
- Connects to your current 10Base or 100Base LAN (with switch support)
- Supports a line speed of 1000 M bits/sec
- ... equals 125 M Bytes/sec
- ... equals 450 G Bytes/hour
- Supports copper RJ-45 cables ("Base-TX") or optical cables ("Base-SX")

## What's in it for me?

- Foremail, file sharing, etc., *probably not a bit*
- *But*, for back-office and data center work:
  - Faster "remote mirroring" of disk volumes
  - Faster, shorter backups ([see next slide](#))
- Versus Fibre Channel:
  - It's Ethernet. "you're soaking in it right now!"
  - Works with your current infrastructure and cabling
  - No interoperability problems: mature technology
  - Minimal capital investment

# What's in it for me?



## Making full use of gigabit speeds [For Backup]

- Filer consolidates disk storage
- Workstations mount to filer
- Tape library consolidates backup
- Filer backed up (over GbE) using:
  - NDMP
  - Future: iSCSI
  - or backup agent protocol

## What's in it for me?

- Another Example [Disk Mirroring]:

How long does it take to "remote mirror" a  
20GB disk volume?

- at 10 Mbits/sec = **4 hours, 26 minutes**
- at 100 Mbits/sec = **26 minutes**
- at 1000 Mbits/sec = **3 minutes** \*

\* = "well, almost" ... see next slide

## Will my Linux system really talk at 1000 Mbits/sec?

- If (and only if):
  - the motherboard is really fast:
    - minimum CPU speed: 1GHz
    - minimum PCIbus: 64 bit, 33MHz
  - the hard disk is really fast:
    - minimum read speed: 20 MBytes/sec
    - minimum internalbus: Ultra2 SCSI
- Returning to our example, you might expect:  
20GB disk-to-disk mirror over a gigabit LAN to take:  
**15 minutes**, compared to 26 minutes for 100Base-T

## What kind of speed should I really expect?

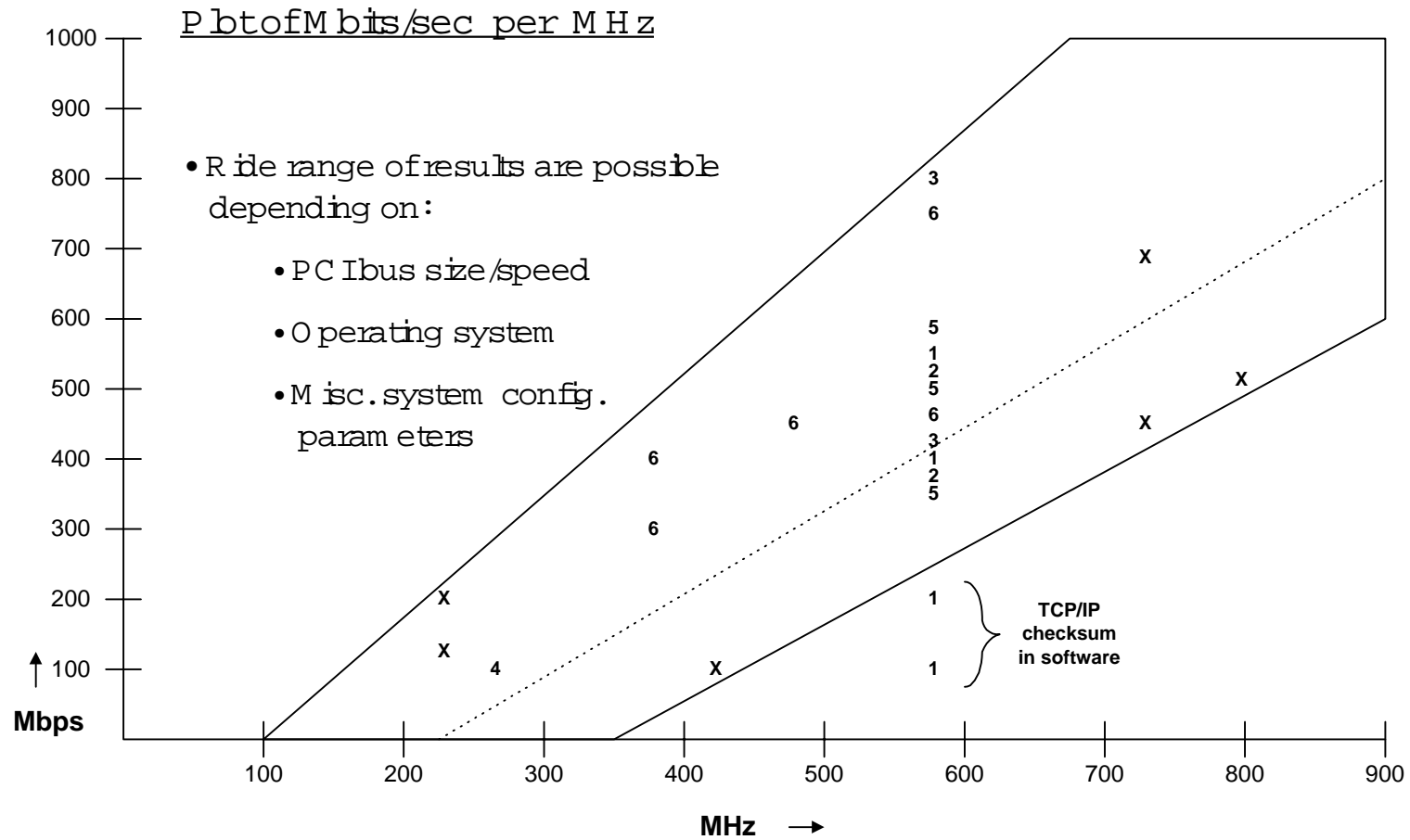
- Ethernet speed mainly depends on:
  - CPU speed (in MHz)
  - PCI bus size and speed
  - and to a lesser degree:
    - Instruction cache size
    - RAM speed
- Realworld disk-to-disk or disk-to-tape speed depends on:
  - Hard disk read and write speed
  - NFS/CIFS file system speed (slow!)
  - Raw tape streaming speed
  - Data compressibility, for tape



## What kind of speed should I really expect?

- Raw Data Rule of Thumb (for a 32 bit, 33MHz PCI bus):
  - (1 M-bit per MHz of CPU speed) \* 0.55
- Example:
  - an 800MHz system could be expected to handle 440 Mbits/sec maximum (at 100% CPU utilization)
- Example:
  - a 233MHz system could be expected to handle 128 Mbits/sec of data at 100% CPU utilization

# What kind of speed should I really expect?



Data comes from : University studies , Trade periodicals , HP research



# Gigabit Ethernet and Linux

## OUTLINE

- What's Gigabit Ethernet?
- What's in it for me?
- Will my Linux system really talk at 1000 Mbits/sec?
- What kind of speed should I really expect?
- >>> **Break for Questions** <<<
- Ugh! Why is it so slow?
- How do I make my gigabit Linux system faster?
- What should I look for in the future?
- Okay I'm sold: Where do I get it?
- How much does it cost?

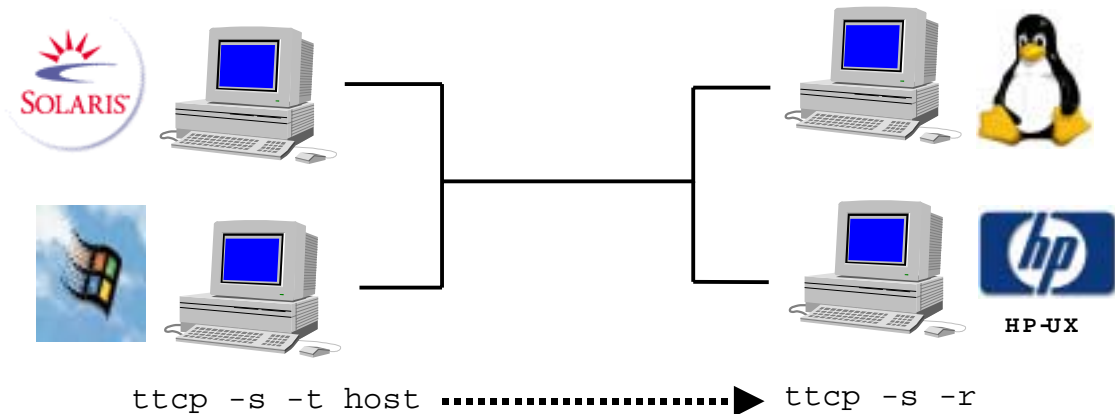


## How can I measure my system's speed?

- Know what you want to measure:
  - Raw memory-to-memory ethernet speed
  - Disk-to-disk speed over the network
  - Disk-to-tape speed over the network
- Use some easy, free, portable tools:
  - **ttcp**
    - for any device-to-device network speed tests
  - **gkrellm**
    - to measure CPU utilization while tests run
- For raw memory-to-memory tests,  
Don't use `ftp`, `rcp`, `tar`, or other file transfer utilities:  
These have to dig through the file system, which makes things slow

# How can I measure my system's speed?

- **ttcp**: a basic socket send/receive program



```
ttcp: buflen=8192, nbuf=2048, align=16384/0, port=5001 tcp
ttcp: socket
ttcp: accept from 15.38.73.188
ttcp: 16777216 bytes in 0.63 real seconds = 25.52 MB/sec
ttcp: 2050 I/O calls, msec/call = 0.31, calls/sec = 3269.71
ttcp: 0.0user 0.2sys 0:00real 42% 0i+32d 14maxrss 0+0pf 919+563csw
```

# How can I measure my system's speed?

- **gkrellm** : a run-time performance monitor



← CPU utilization (%)

← Disk speed (blocks or I/O per sec)

← Network card speed (M Bytes/sec)

# How can I measure my system's speed?

- Where to get the tools:

**ttcp** (for any UNIX or Windows NT, and as source code)

<http://jes.home.cern.ch/jes/www/gige/acenic.html>

<http://www.pcausa.com/Utilities/pcttcp.htm>

**gkrellm** (for any Linux distributions, and as source code)

<http://web.wtnet/~billw/gkrellm/gkrellm.html>

## Ugh! Why is it so slow?

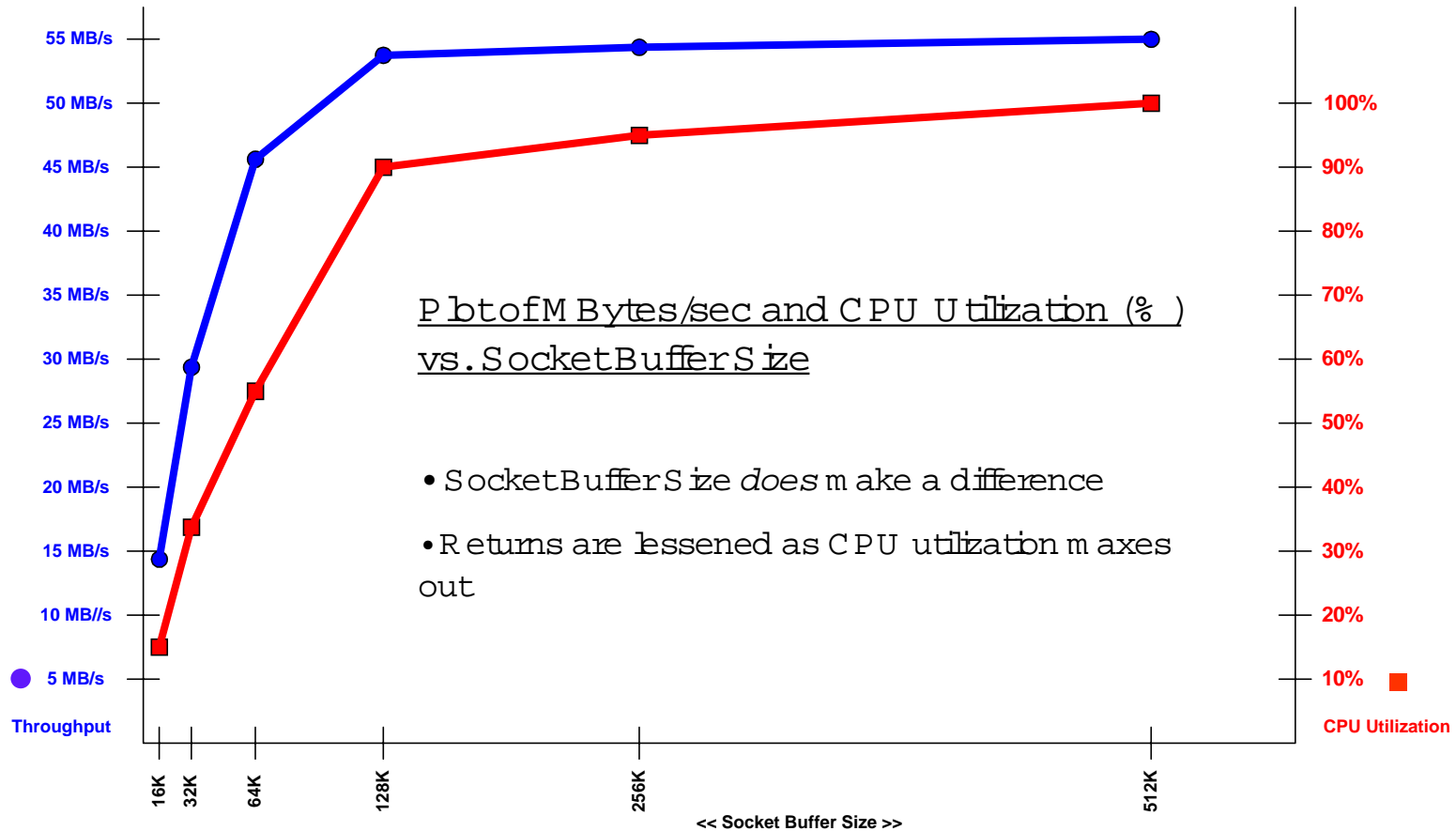
- Your CPU is maxed out
  - Handling TCP/IP transfers requires CPU horsepower
- Your PC Bus is too small or too slow
  - A 64-bit bus eliminates 50% of the CPU work vs. a 32-bit bus
  - A 66 MHz bus can speed things up, but only if your CPU can handle it
- Your socket buffer size is too small
  - See next slide for how to change this
- Your 10/100/1000 switch is routing gigabit traffic over a 10/100 infrastructure



# How do I make my gigabit Linux system faster?

- Hardware
  - Upgrade motherboard: 1 GHz
  - Make sure gigabit card is in a 64 bit, 66 MHz slot
- OS configuration
  - Increase your socket buffer size:
    - edit `/proc/sys/net/core/rmem_max`
    - Change the value in this file to 524284
    - repeat for `/proc/sys/net/core/wmem_max`
    - Note: applications must take advantage of these settings!
    - Note: `rmem_max` often resets to 65536 at start-up:
      - Can be modified handled by `profile` and other script

# What kind of speed should I really expect?



Data comes from :HP research, sending M Bytes of data between two 800MHz servers

## What should I look for in the future?

- Increased hardware acceleration
  - GbE NICs already doing some of this
  - TOE: TCP/IP Offload Engines, on NICs
    - Relieves server CPU of all TCP/IP leg-work
- Leaner network stack
  - 2.4 kernel already moving in that direction
- Leaner file transfer (and block transfer) protocols
  - NDMP: Network Data Management Protocol
  - iSCSI: SCSI over Ethernet

# Where do I get it? How much does it cost?

- NIC Vendors
  - 3Com, HP, Intel, SysKonnect, NetGear
- NIC Pricing
  - \$150 - \$900, copper or optical
- GbE Switch Vendors
  - Cisco, HP, 3Com, D-Link, many others
- GbE Switch Pricing
  - \$1400 - \$6000+ (for advanced routing switches)
- Drivers
  - Most GbE NICs are based on Alteon's ACEnic chipset
  - Opensource ACEnic driver: ships with many distributions

# Gigabit Ethernet and Linux

## OUTLINE

- What's Gigabit Ethernet?
- What's in it for me?
- Will my Linux system really talk at 1000 Mbits/sec?
- What kind of speed should I really expect?
- How can I measure my system's speed?
- Ugh! Why is it so slow?
- How do I make my gigabit Linux system faster?
- What should I look for in the future?
- Okay I'm sold: Where do I get it?
- How much does it cost?

>>> Other Questions? <<<



# G igit E themet and L inux: A P erform ance S tudy



Jeff Bain  
HP - Automated Storage  
jeff\_bain@hp.com



# THANKS !

