Quality of Service on Gigabit Ethernet for Event Builder

Y.Yasu(1), Y.Nagasaka(2), H.Hasegawa(3), A.Manabe(1), M.Nomachi(4), H.Fujii(1), Y.Watase(1)

(1)High Energy Accelerator Research Organization (KEK)
    Oho 1-1, Tsukuba, Ibaraki 305 Japan
(2)Department of Electrical Engineering
    Nagasaki Institute of Applied Science (NIAS)
    536 Aba-machi, Nagasaki 851-01 Nagasaki Japan
(3)Faculty of Science
    Shinshu University
    Asahi 3-1-1, Matsumoto, 390-8621 Nagano Japan
(4)Graduate School of Science
    Osaka University
    Machikaneyama 1-1, Toyanaka-city, 560-0043 Osaka Japan

Contents

Gigabit Ethernet and Quality of Service(QoS)
        for Event Builder

Basic performance of Gigabit Ethernet

QoS in Linux

QoS performance on Gigabit Ethernet

Gigabit Ethernet and Quality of Service for Event Builder

Congestion Avoidance
  by Global Traffic Control(Fermilab, KEK)
        Special Hardware with Barrel shifter etc.
  by traffic shaping for each node(RD31/ CERN)
        ATM including QoS(Constant Bit Rate)

How can we use Gigabit Ethernet with TCP/IP ?

Basic performance of Gigabit Ethernet

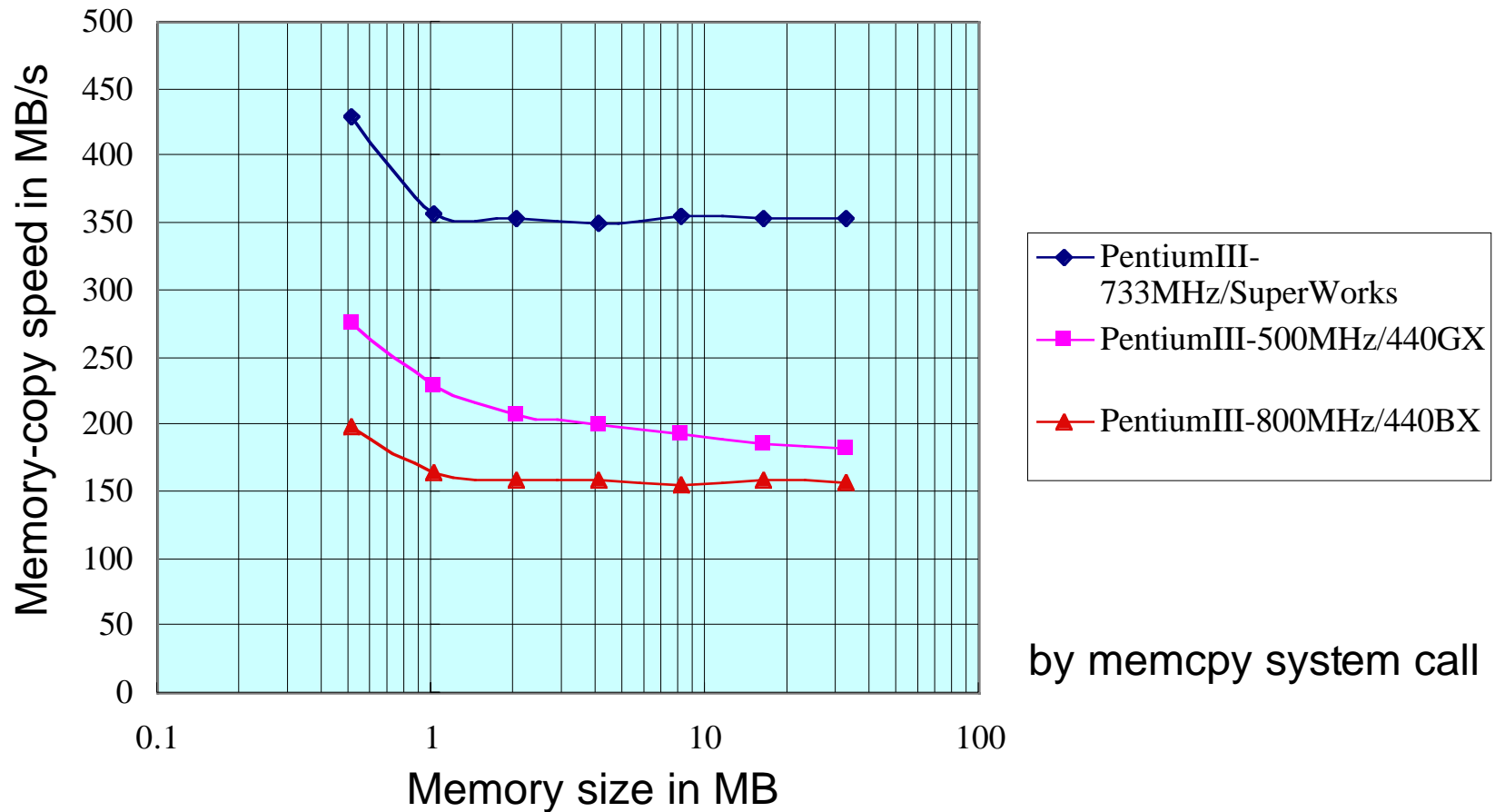Hardware configuration and setup

Performance measurement of memory copy

Tuning network driver
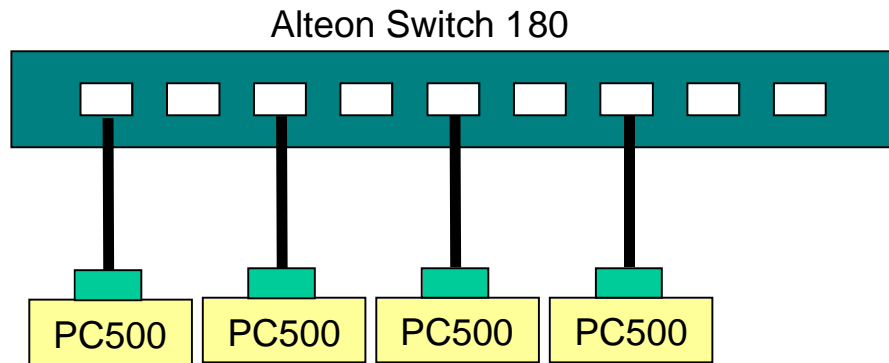
Performance measurement of Gigabit Ethernet

Hardware configuration

System	PC500	PC733	PC800
CPU type	PentiumIII	PentiumIII	PentiumIII
CPU Freq.	500MHz	733MHz	800MHz
Cache	512KB	256KB	256KB
Chipset	440GX	SuperWorks ServerSet III LE	440BX
Bus speed	100MHz	133MHz	100MHz
Memory type	SDRAM	SDRAM	SDRAM
Memory size	256	128	256
PCI(bus width)	32-bit	64-bit/32-bit	32-bit
OS version	2.2.14-12	2.2.14-12	2.2.14-5
gcc version	egcs-2.91.66	egcs-2.91.66	egcs-2.91.66
NIC	AceNIC(1MB)	AceNIC(1MB)	AceNIC(1MB)
Network Driver version	v0.47	v0.47	v0.47

Performance measurement of memory copy

by memcpy system call

Hardware setup

Alteon Switch 180

PC733   PC800

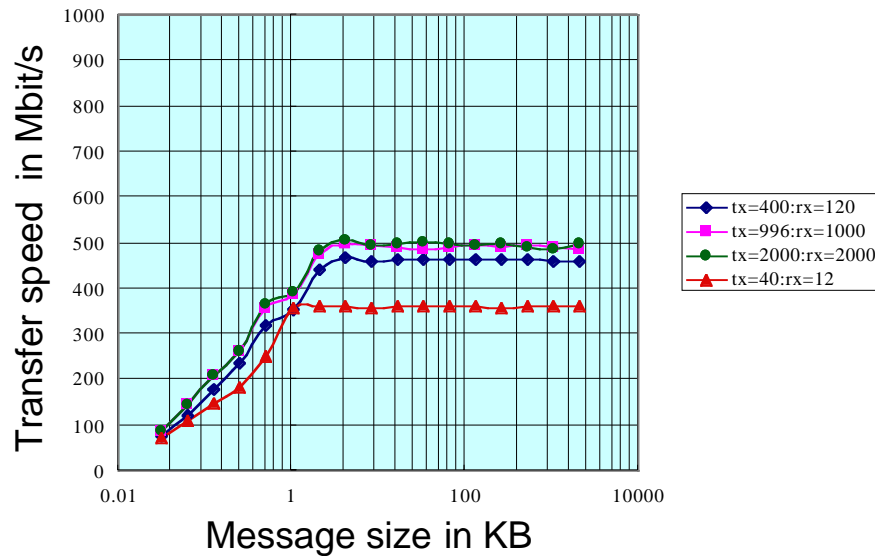PC500   PC500   PC500   PC500

PC733   PC800

For measurement of QoS performance
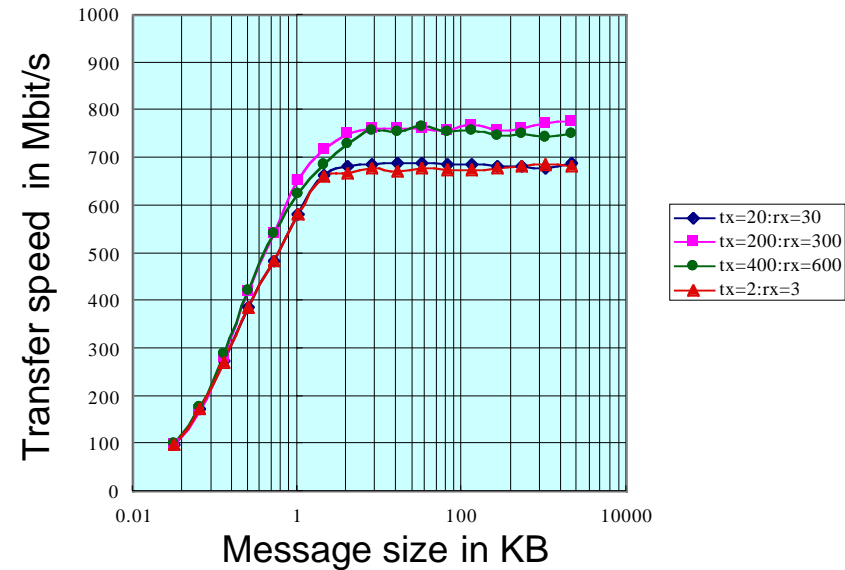and measurement of basic performance
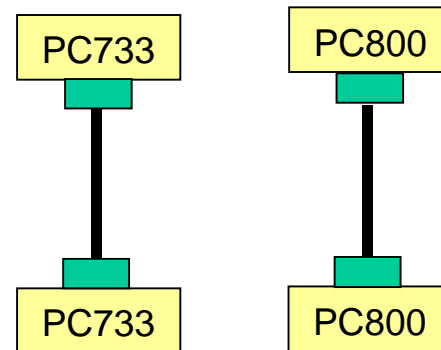
For measurement of
basic performance

Tuning network driver

Normal frame

Jumbo frame

Tuned parameters of Network driver :
/proc/sys/net/core/rmem_max 4000000
/proc/sys/net/core/wmem_max 4000000
TCP buffer size : 524288 bytes
Nagle algorithm is enable.

Normal frame : tx_coal=996 : rx_coal=1000
Jumbo frame   : tx_coal=200 : rx_coal=300
Other parameters : default
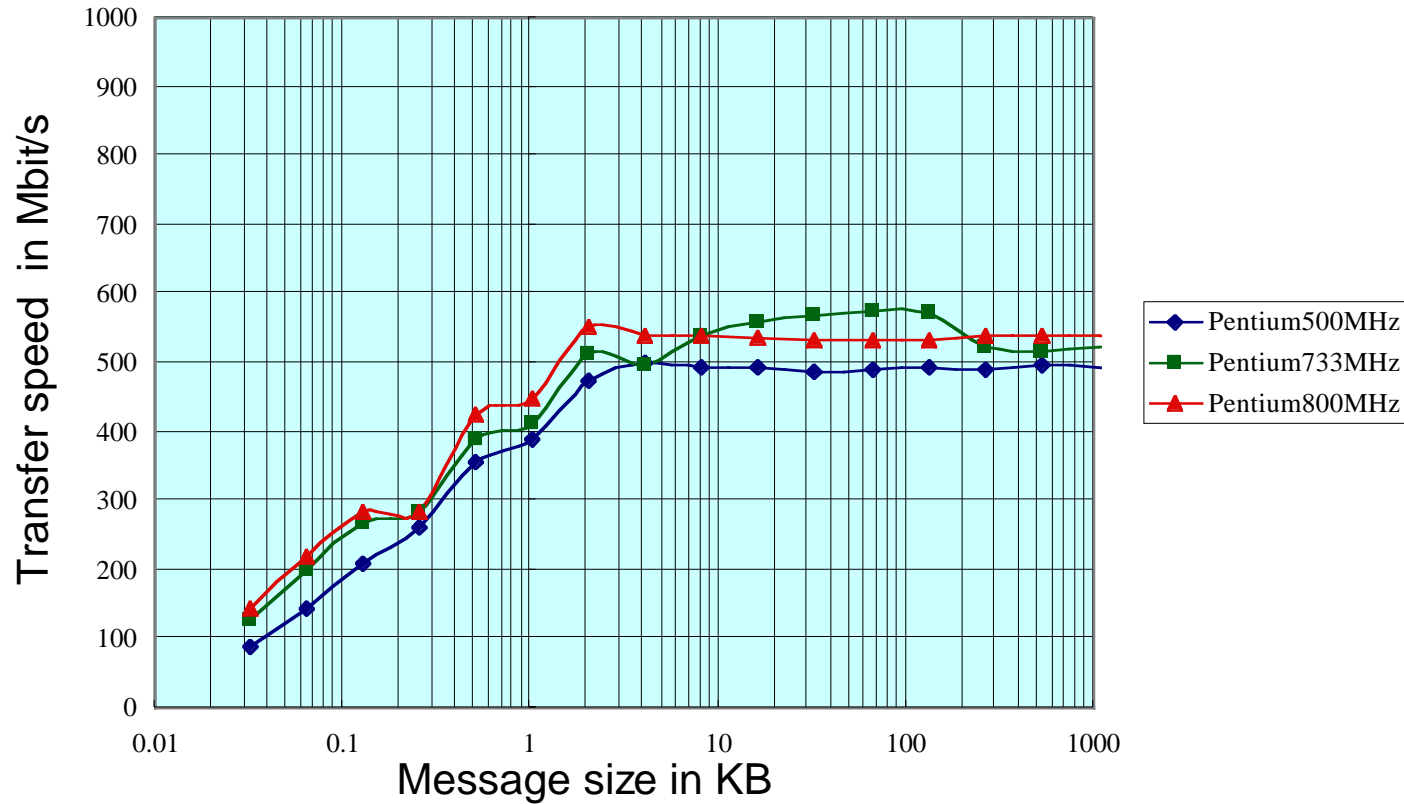Used systems : PC500s
Measuring tool : Netperf

Performance measurement of Gigabit Ethernet

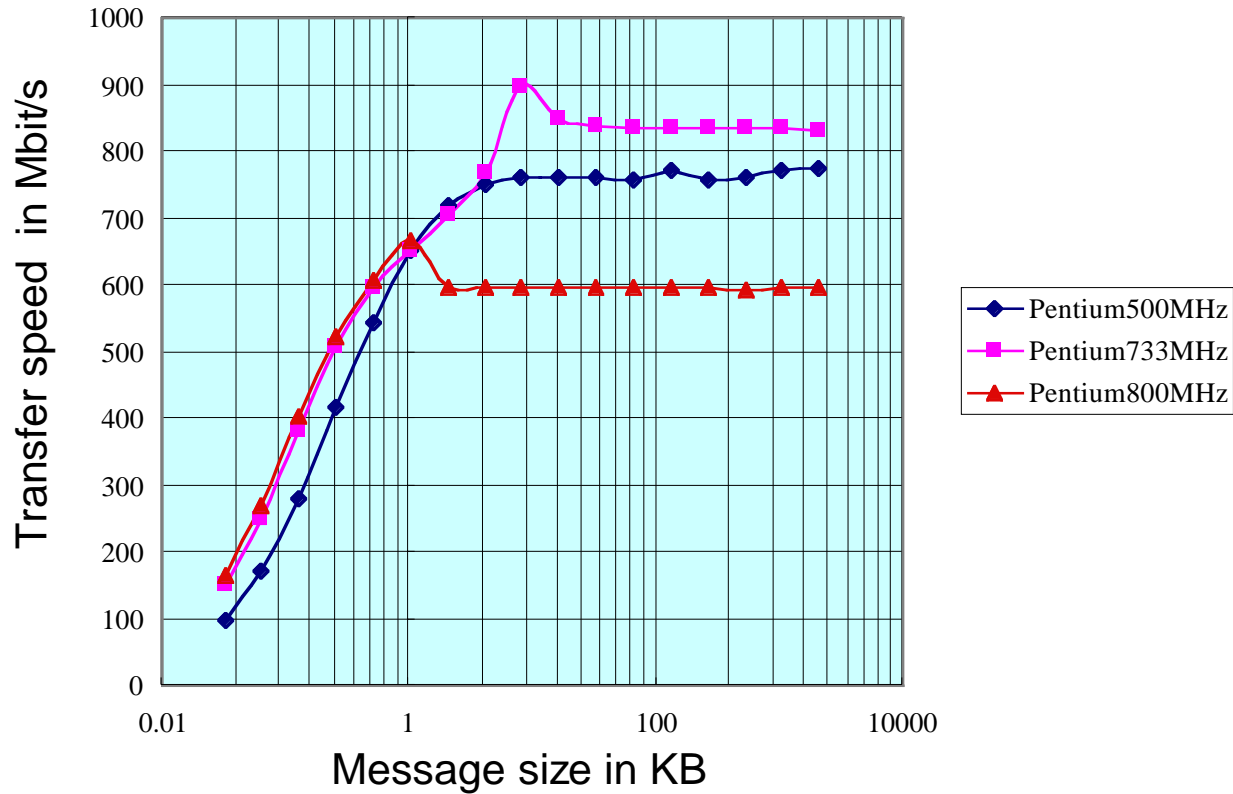Transfer speed in normal frame

Transfer speed in jumbo frame

Bus width dependency(64-bit PCI vs. 32-bit PCI)

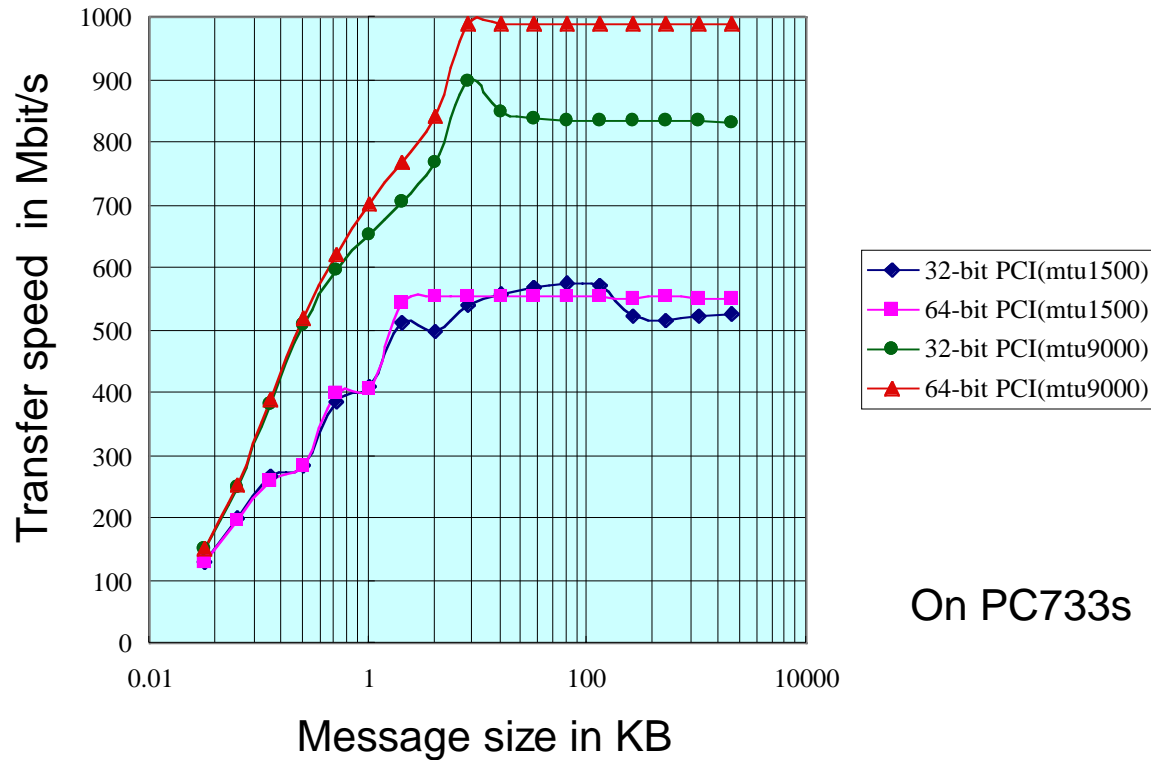# Transfer speed in normal frame

Transfer speed does not depend on CPU speed in normal frame

Transfer speed in jumbo frame

Transfer speed depends on Chipset performance in jumbo frame.

Bus width dependency(64-bit PCI vs. 32-bit PCI)

On PC733s

Bus width does not affect transfer speed in normal frame.
64-bit PCI improves transfer speed in jumbo frame, up to 990Mbit/s.

QoS in Linux

- Class-Based Queueing
- Token Bucket Filter
- Traffic control in TCP/IP
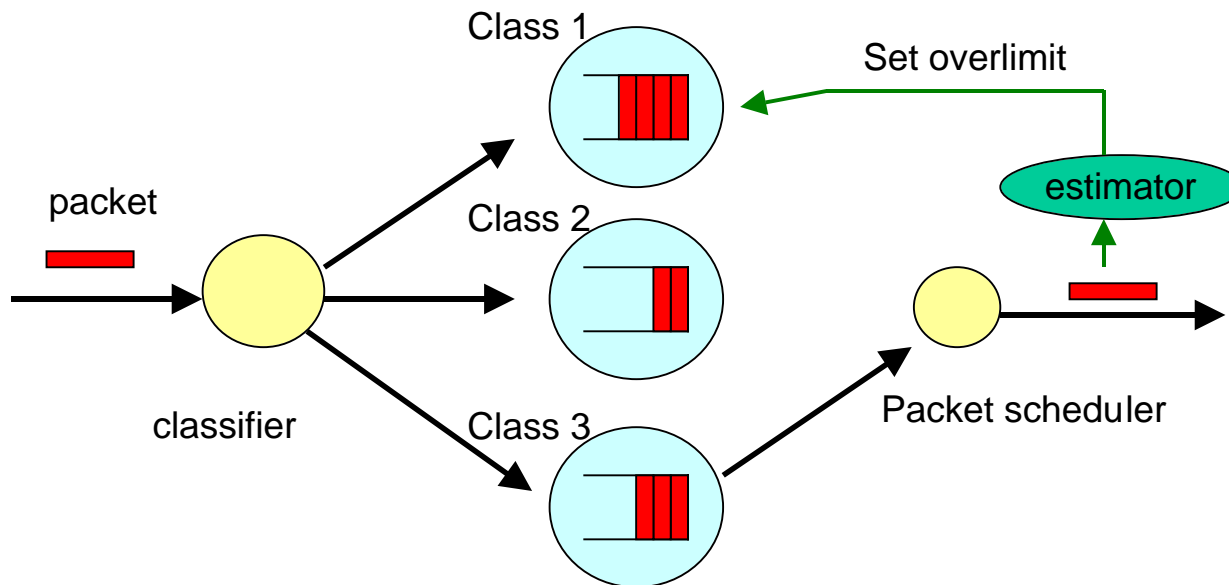- Queueing discipline for Linux

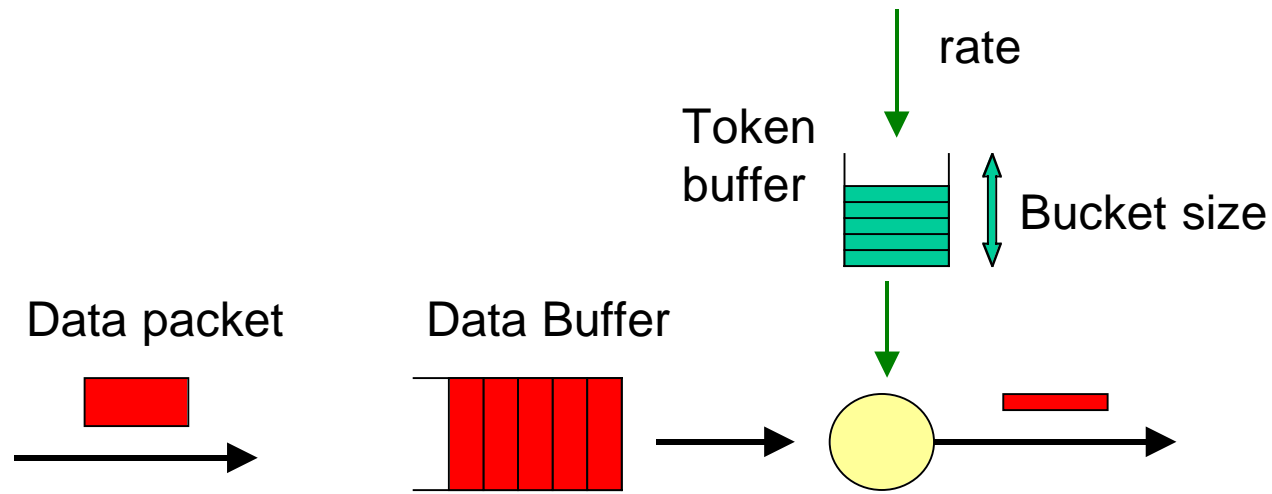## QoS performance on Gigabit Ethernet

1x1 system

3x1 system

1x3 system

Class-Based Queueing (CBQ)

CBQ can classify incoming packets into multiple classes.
CBQ can share and limit the transfer rate.

Token Bucket Filter (TBF)

TBF can limit the transfer rate according to the token rate.

QoS in Linux

QoS in Linux kernel 2.2.x and Iproute2 for Linux
supplies the following functions to control IP packets;
For IP packets of inbound and outbound
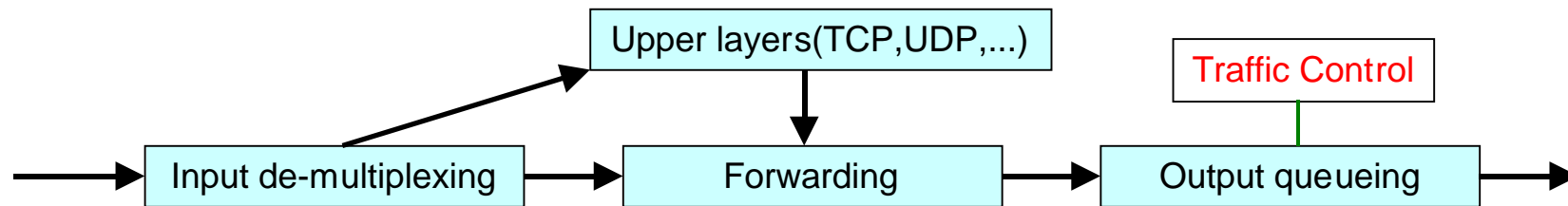Classifying
Prioritizing
Sharing
Limiting

QoS in Linux kernel 2.2.x : in kernel standard distribution
Iproute2 : ftp://ftp.inr.ac.ru/ip-routing/iproute2-current.tar.gz
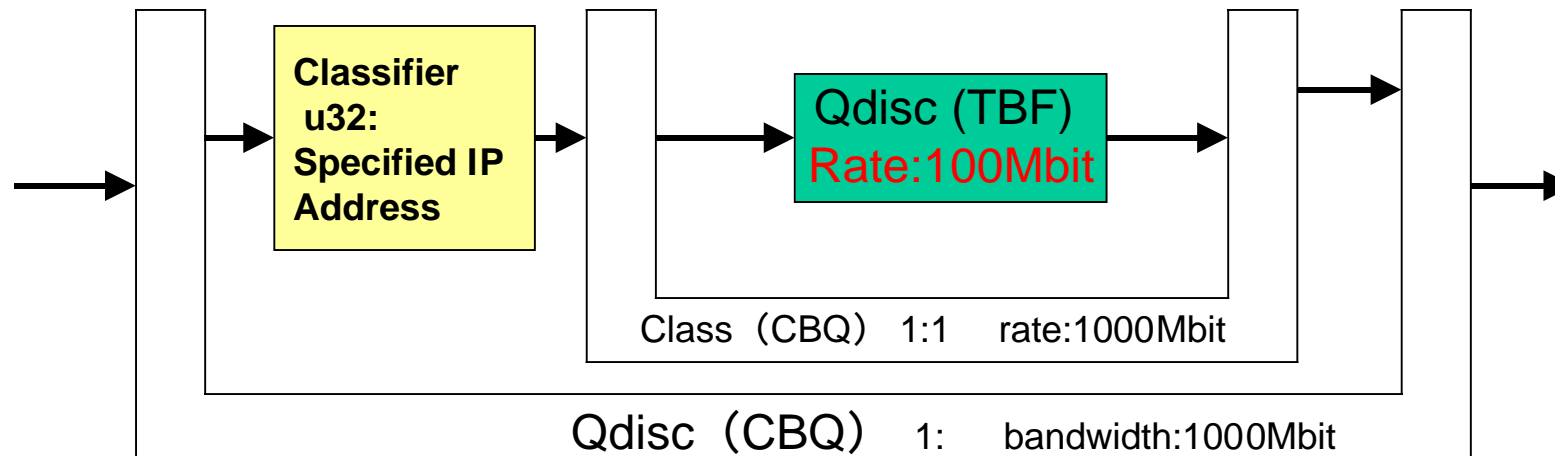**Iproute2** supports a typical command called tc.

Can we use the QoS such as Class-Based Queueing and
 Token Bucket Filter for shaping traffic on Event Builder?

Traffic control in TCP/IP

Traffic control is done only at output queueing.

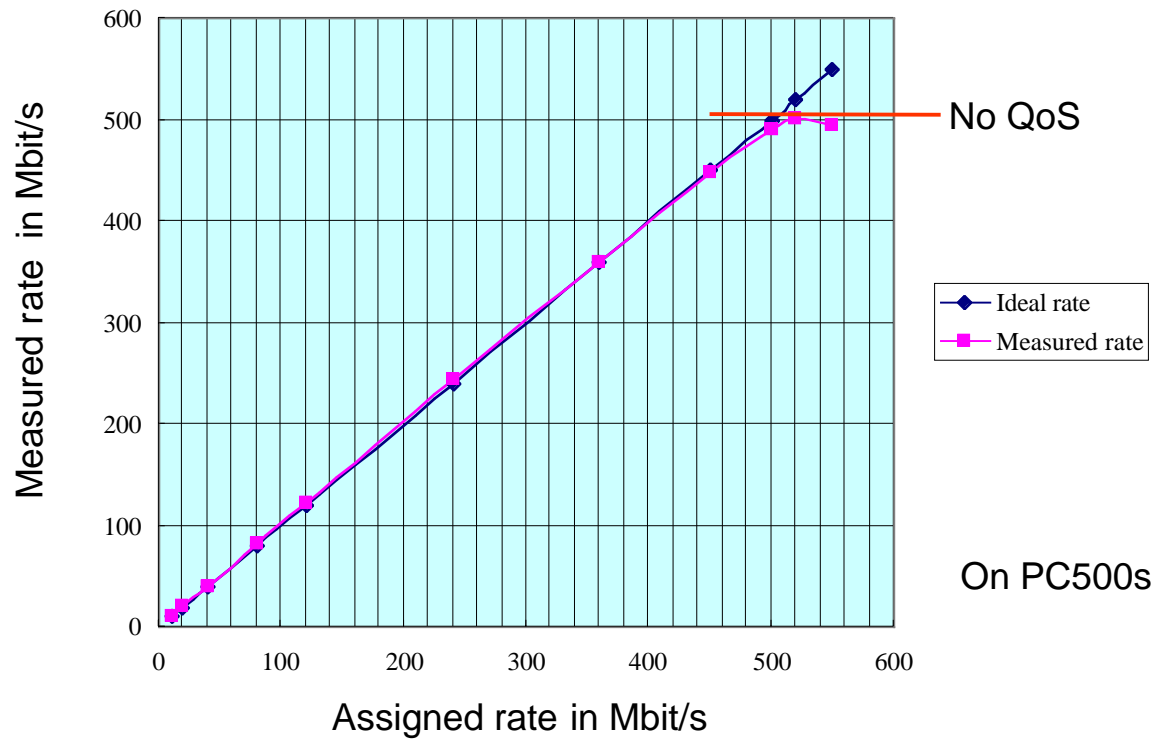# Queueing discipline for Linux

U32 is a classifier.
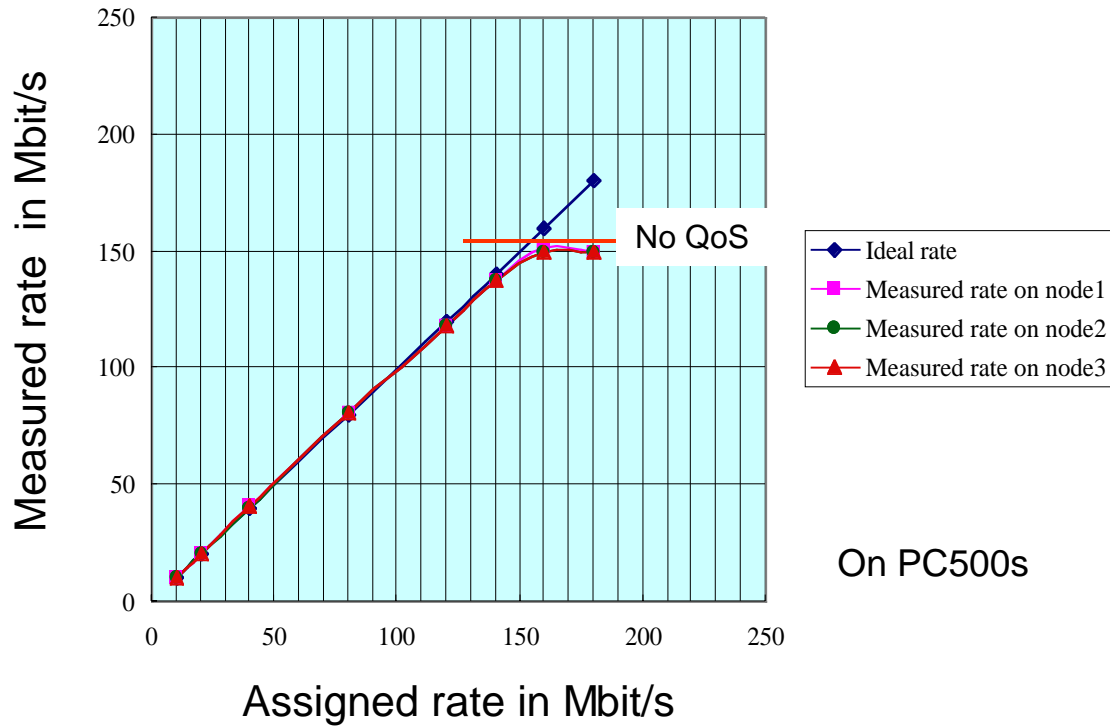TBF limits the transfer rate to 100Mbit/s at this example.
The configuration was used at 1x1 system for the evaluation.
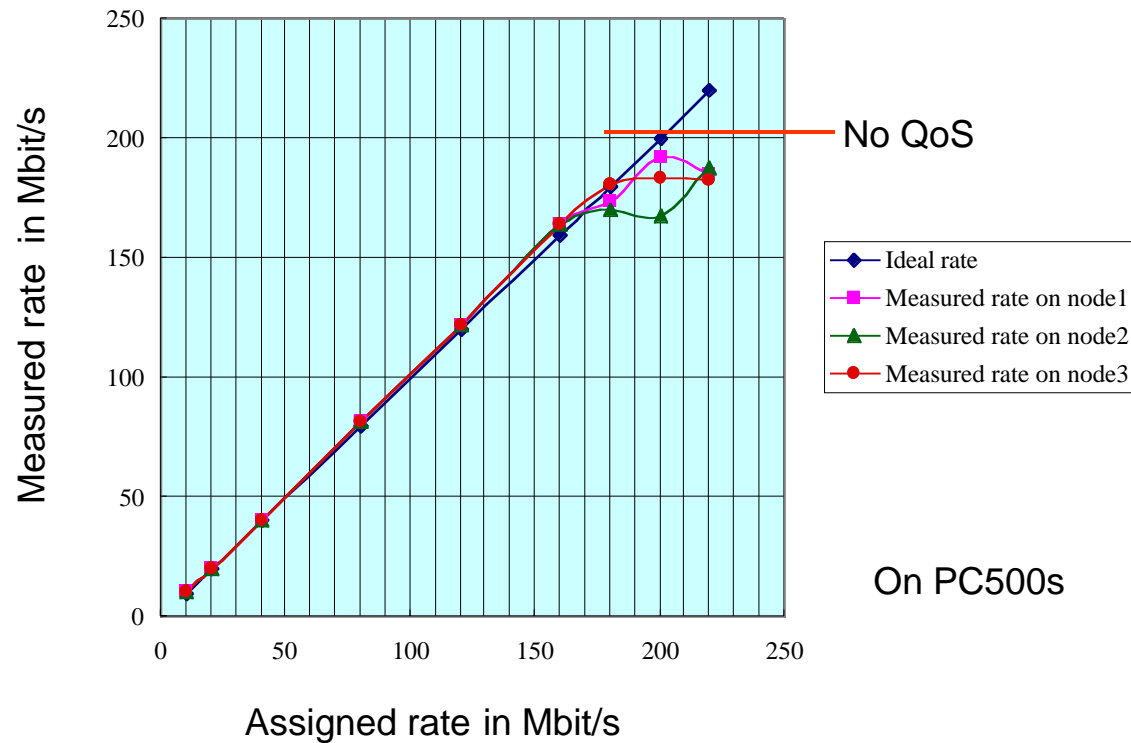
QoS performance on 1x1 system

Bandwidth assignment worked well.
The overhead of QoS was very small and nearly 1 %.

QoS performance on 3x1 system

Bandwidth assignment worked well

QoS performance on 1x3 system

Bandwidth assignment worked well at small rate, but did not work at full rate.
Packet sharing performance without QoS was better than that with QoS.

Conclusion

* Traffic management of event data flow is necessary for event builder.

* IP-based QoS with Gigabit Ethernet on PC/Linux was investigated.

* The transfer speed did not depend on CPU speed in normal frame.

* High performance chipset and 64-bit PCI improved
        the transfer speed in jumbo frame, up to 990Mbit/s.

* Assignment of transfer rate was possible by using Linux/TC command,
   but it is not clear that the QoS on Gigabit Ethernet is effective for
   event builder. Thus, more investigation is necessary.