Performance Evaluation of IEEE 802.1Qbu: Experimental and Simulation Results

Wen-Kang Jia
Department of Computer Science
National Chiao Tung University
Hsinchu, Taiwan
wkchia@cs.nctu.edu.tw

Gen-Hen Liu
Department of Computer Science
National Chiao Tung University
Hsinchu, Taiwan
gordon.cm99g@nctu.edu.tw

Yaw-Chung Chen
Department of Computer Science
National Chiao Tung University
Hsinchu, Taiwan
ycchen@cs.nctu.edu.tw

***Abstract*—In this paper, we apply IEEE 802.1Qbu Preemptive Priority Frame Forwarding mechanism to the IEEE 802.3z Gigabit Ethernet networks, and compare its performance with non-frame-preemption queuing schemes for various traffic patterns. A feasible implementation of the proposed solutions is addressed. The FPF effectively reduces the forwarding delay as well as jitter, so as to enhance response time of real-time services in GbE networks. Through simulations, we show that the FPF achieves superior and precise QoS control compared to existing non-preemptive-based priority scheduling schemes.**

***Keywords*—preemptive priority; frame preemption; Gigabit Ethernet (GbE); 802.1Qbu; 802.3z**

I. INTRODUCTION

Nowadays, Local Area Network (LAN) technology has made more popular on almost every industry, and Ethernet is by far the most commonly used LAN technology. Since Ethernet and its descendant have become a major technology for high speed networks from distribution, edge to core of all types. Ethernet is the most popular data link technology that covers MANs and WANs. As the new applications in e-commerce, high-quality streaming media, emergency communication, circuit emulation services [1], network management, and dependable remote sensing and control over the LAN are ever increasing, massive demand for QoS becomes a must and administrators of LANs have to seek and adopt new technologies.

Since the delay is a critical QoS factor, frame scheduling over an Ethernet link often becomes a primary challenge to meeting each endpoint's delay requirement. However, most current packet switched networks such as Ethernet do not support Preemptive Priority Frame Forwarding mechanism. This is because once a frame is being transmitted; it could not be stopped in such a baseband communication link. A conventional priority queue scheme is restricted by physical capacity and specifications of Ethernet MAC and PHY layers; hence, most networking applications deal with the non-preemptive level of priority queue management methods as regards services. In order to accommodate delay- and jitter-guaranteed transmission, the Ethernet networks should support technology such as preemptive-based priority instead of non-preemptive priority. Thus, delay- and jitter-guaranteed traffics, when transported over Ethernet networks, should maintain a constant (or as short as possible) latency. The technology will help in the increasing of Ethernet and solve the current QoS problems.

This study attempts to employ the priority-based frame preemption mechanism in the IEEE 802.3z Gigabit Ethernet (GbE) [2] networks to minimize delay for *time-critical traffic (TCT)* with minimize interference for *non-time-critical traffic (NTCT)*. This scheme of FPF has been subsumed into the IEEE 802.1Qbu [4] PAR. The proposed technique is developed with Ethernet to satisfy the growing needs of real-time applications; it also can be applied to all kinds of Ethernet such as 10 Gigabit Ethernet (10GbE) [5]. More opportunities are regarding delay-guaranteed services to benefit from these technologies.

II. 802.1QBU PREEMPTIVE PRIORITY FRAME FORWARDING

Preemptive and non-preemptive are the two basic approaches to priority queuing [13]. The basic approach of Ethernet is the latter. With non-preemptive priority queuing, if a system receives a *time-critical frame (TCF)* while processing a *non-time-critical frame (NTCF)*, it inserts the TCF to the beginning of queue. Once the current NTCF is transmitted, it will insure the TCF being processed as soon as possible. The system resumes processing the next NTCF after the TCF is processed if there is no more TCF in the queue. In this scheme, TCFs are processed in a minimum wait (*O (MTU)*) upon arrival so they experience as smaller delay and delay variation (jitter) as possible under system generating capacity at light load. In GbE with jumbo frame and/or Frame Burst Mode (FBM) supported [14], the MTU could be larger than 9k bytes, which mean it may have a forwarding delay larger than 80μs in one hop.

The increased delay in a NPPQ scheme could become a problem. All high priority frames (HPFs) such as voice, video, system alarms, failure indication, etc. may have to endure a longer delay. For example, the delay in 10M Ethernet may be large so that the QoS of delay-sensitive services is affected, and the advantages of PPQ would be more apparent.

The need to protect time-critical data via QoS mechanisms in networks has escalated over the past few years. In QoS enabled network, the pragmatic tradeoffs between TCT and

978-1-4799-0537-9/13/$31.00 ©2013 IEEE

NTCT must be considered in its design. In the perspective of TCT, the delay is raised by interference from NTCT. IEEE 802.1Qbu presents balanceable design recommendations to address these needs on link layer, which minimizes delay for TCT while minimizing the effect on NTCT.

In original 802.1Qbu, it defined a class of service for time-critical frames that request the transmitter in a QoS enabled network to suspend the transmission of NTCF, and allow for one or more TCFs to be transmitted. When the TCFs have been transmitted, it resumes the transmission of the preempted frame. This process is illustrated in Figure 1. Since a non-time-critical frame could be preempted multiple times even in nested model, from 802.3's perspective, eventually the transmitted frame size could exceed the MTU in 802.3's specification, such as a jumbo frame that is larger than 1544 bytes and less than 9K bytes. We called it a preempted burst.

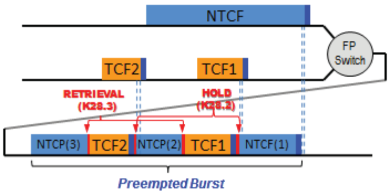


Fig. 1. The operation of 802.1Qbu priority-based frame preemption.

With 802.1Qbu, if a traffic aggregation system receives a TCF while processing a NTCF, it stops processing the latter and begins processing the former. The system resumes processing the NTCF after the TCF is processed. In this scheme, TCFs are processed immediately upon arrival so they experience much smaller delay and delay variation.

To implement 802.1Qbu preemptive priority queuing, both sender and receiver must enable 802.1Qbu support. There are two existing approaches: Firstly, the receiver must extract and identify high-priority Ethernet frames on input queues. Frame priority is used to determine the order in which frames are submitted to the network in appropriate periods. Then, based on the priority and flow rate, the QoS frame scheduler determines the transmission schedule of each output queue and resolves competition between queued frames that simultaneously request accessing the network.

To ensure preferential treatment throughout the network, TCFs must be marked so that the MAC control can detect the priority and handle the frames appropriately. Typically, Ethernet frames are marked with an IEEE 802.1p [17] priority for prioritization by the MAC layer. We employ IEEE 802.1p to provide the mechanism for implementing Classes of Service (CoS) in the MAC layer. The IEEE 802.1p defines eight different CoS which are usually expressed through a 3-bit priority field within a tag added to the standard Ethernet header. These tags indicate a TCF preempting in which a NTCF will be transmitted in the next transmission opportunity in the proposed 802.1Qbu MAC.

In order to identify a specific restriction fragment using special control symbols in each preempted burst. These special symbols consist of bit sequences that have not been used by the standard Ethernet coding. The receiver uses the sequences to identify the position of the high priority frames and extract them. We employed two reserved control symbols: the HOLD (K28.2) and RETRIEVAL (K28.3) to indicate a TCF which can be preemptively inserted into an NTCF. The Ethernet 8b/10b encoder and decoder circuits are implemented in the Physical Coding Sublayer (PCS) [3] of the IEEE 802.3 stack, and located in the upper PHY layer and engaged with the MAC layer. Once the receiver recognizes the HOLD symbol as an escape signal during a normal-priority transmission, the normal-priority receive procedure will be paused (the first segment), and a high priority receive procedure will be started until the receiver recognizes a RETRIEVAL symbol and resumes the interrupted NTCF (the second segment). Since IEEE 802.1Bbu allows a NTCF to be interrupted multiple times until it reaches preempted burst limitation, the aforementioned operation will be repeated as many times as needed during a NTCF transmission period.

III. SIMULATION AND RESULTS

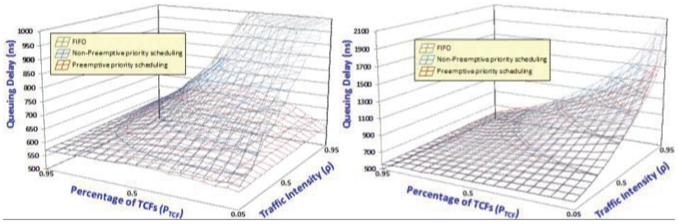
In this section, the IEEE 802.1Qbu priority-based frame preemption for GbE is evaluated. We compare 3 types of scheduling schemes: 1) FIFO; 2) strict and non-preemptive priority scheduling (NPPS); 3) strict preemptive priority scheduling (PPS). The assessment has been evaluated based on two attributes, average frame forwarding delay and jitter performance. We established frame preemption model using EstiNet 8.0 simulation tool.

The TCFs were transmitted from four GbE ingress links to an Ethernet switch. Simultaneously, NTCFs come from same four GbE ingress links was also transmitted as background traffic. We measure the performance indexes from the only egress link. The MTU of the measured background traffic was 1,518 bytes according to IEEE 802.3 standards, and its traffic patterns were in statistical characteristic by real frame length distribution in Internet; it is trimodal in GbE links. The traffic intensity was set to ρ of the GbE line speed, and the time-critical traffic was set to P_{TCF} of the ρ including overhead. In other words, the non-time-critical traffic was set to $P_{TCF}= (1-P_{NTCF})$ of the entire traffic. Based on the above setting, the average delays, the average jitters of TCFs and NTCFs were measured. FIFO is not considered because it has no priority capability and cannot support high priority frames such as VoIP.

A. Forwarding Delay vs Traffic Intensities

Firstly, we consider the mean forwarding delay in various traffic intensities and background NTCT as well as its difference among different scheduling schemes. The propagation, transmission and processing delay are ignored. We collected mean delay from 50K tested frames under simulation, and we varied the traffic intensity ratio (ρ) from 0.05 to 0.95 in increments of 0.05, as well as the percentage of TCFs (P_{TCF}), which was set from 0.05 to 0.95 in increments of 0.05. We have chosen the number of ingress port I to16; and the traffic pattern of TCF and NTCF are both with the real Internet characteristic.

We found that the forwarding delay curves of TCF and NTCF are totally different from each other's as shown in Figure 2. The forwarding delay should be kept as short as possible in both sides, the mean delay of TCFs feature a descending order in all simulation situations as follows: FIFO>NPPS>PPS (with 802.1Qbu) on average, and the delay of NTCFs feature a descending order as NPPS>FIFO>PPS (with 802.1Qbu) on average. Compared with NPPS, our simulation results showed that the performance of PPS (with 802.1Qbu) is obviously better than NPPS not only in high-priority TCT, but also in normal-priority NTCT. Under heavy background load, the mean forwarding delay of TCT in PPS can be maintained smaller than 700ns, and the delay of NTCT in PPS can be maintained no worse than FIFO (without queue management) and NPPQ scheme. Compared to NPPS, PPS mechanism could effectively control the increase of the forwarding delay of TCF under extreme cases such as in full load (P_{TCF}=ρ≈0.95). The forwarding delay initially increased and then decreased with the increase of P_{TCF}, the peak delay has been found to occur regularly in both P_{TCF} and P_{NTCF} are 50% with full ρ.



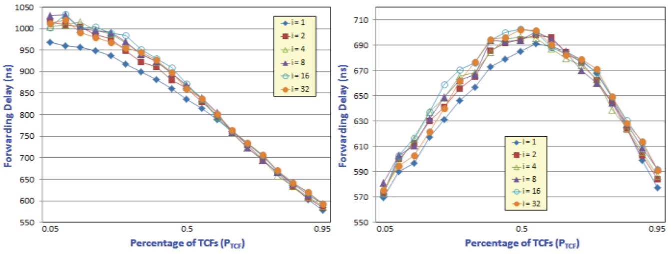
(a)

(b)

Fig. 2. Mean forwarding delay of (a)TCFs and (b)NTCFs with various schedulers, ρ and P_{TCF}. (*i*=16, traffic pattern=Real Internet traffic)

B. Forwarding Delay vs Number of Ingress Ports

Secondly, we also concern that the number of ingress ports of 802.1Qbu switch can affect forwarding delay, and compare the results in various P_{TCF} with that of NPPS's. We setup a simulation in which increasing the number of ingress ports, and increasing the P_{TCF} was ordered from 5% to 95%, where the traffic intensity was set to 95% of the wired speed of GbE (ρ=0.95), and the number of ingress ports varies as 1, 2, 4, 8, 16 and 32 (*i*=1~32).

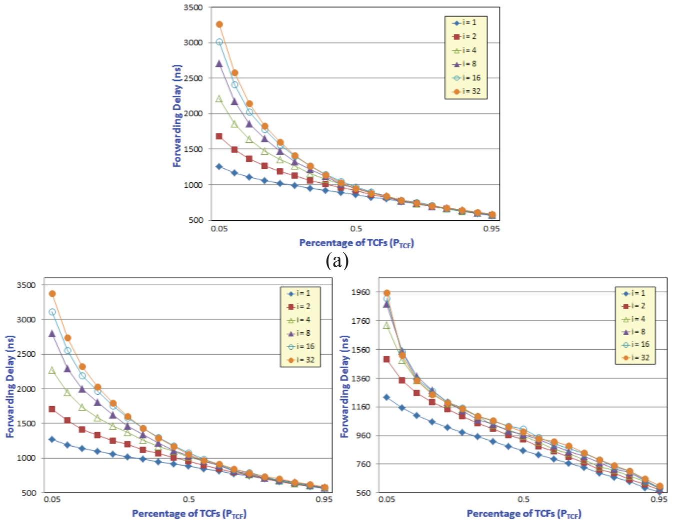


(a)

(b)

Fig. 3. Mean forwarding delay of TCFs with various number of ingress ports and P_{TCF}. (scheduler=(a)NPPS, (b)PPS ρ=0.95, traffic pattern=Real Internet traffic)

The simulation results shows that the average forwarding delay of 1) TCF w/NPPS; 2) TCF w/PPS; 3) TCP and NTCF w/FIFO; 4) NTCF w/NPPS; 5) NTCF w/PPS in Figures 3 and 4 respectively. In case of the FIFO, NPPS and PPS, the delay of NTCF are basically raised according to the number of ingress ports; but in case of TCF, The relationships between delay and number of ingress ports are not significant except that *i*=1. This is because while the ingress port is unique, the mutual preemption between TCF and NTCF will not occur frequently.



(a)

(b)

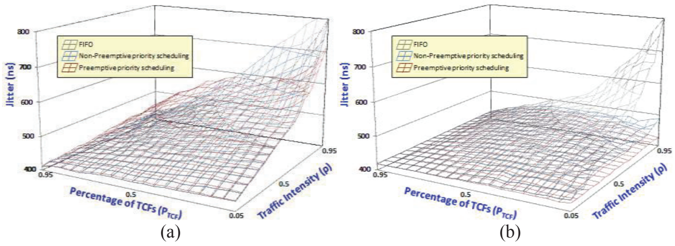
(c)

Fig. 4. Mean forwarding delay of NTCFs with various number of ingress ports and P_{TCF}. (scheduler=(a) FIFO, (b)NPPS and (c)PPS, ρ=0.95, traffic pattern=Real Internet traffic)

C. Jitter

We also concern the jitter of PPS with 802.1Qbu in various ρ and P_{TCF} in the system, and compare the results with that of NPPS's. In this case jitter is the difference in packet delay from the 802.1Qbu ingress ports to egress port, from one packet to the next. We setup a simulation with all conditions same as above. Figure 5 shows the results obtained for the mean jitter of TCF and NTCF in several conditions.

We found that jitter of TCF will be raised according to the background load P_{NTCF}. Compare to NPPS, the maximum jitter in PPS could also be precisely estimated as between 400~460ns depending on the ρ and P_{NTCF}, the jitter of TCF will not exceed 460ns under heavy load with PPS. If the 802.1Qbu switch has multiple ingress ports including multiple TCF, the jitter of TCF is expected higher, this part is left as future work.



(a)

(b)

Fig. 5. Mean jitter of (a)TCFs and (b)NTCFs with various schedulers, ρ and P_{TCF}. (*i*=4, traffic pattern=Real Internet traffic)

IV. PROS AND CONS

Several issues must be clarified as follows:

1) ***Too long frame:*** although jumbo frame is supported in GbE and its latter revisions, the TCFs and their successors will rapidly consume the packet length budget of an Ethernet jumbo frame. In cut-through method, it will be hard to early determine whether a TCF was also a jumbo frame because generating a too long frame violates the IEEE 802.3 standards. That will cause a trouble in preemptive priority encapsulation.

2) ***Starvation of NPFs:*** When the percentage of TCFs reach a threshold, the NTCFs will be starved. An appropriate admission control may avoid the situation effectively in ordinary condition, but the network planer should prevent the possible TCT from being oversubscribed. Moreover, the repeated insertion based on PPS and 802.1Qbu needs to be discussed.

3) ***Flow-control of NPFs:*** As mentioned above, the opposite of starvation is buffer overflow. When NTCFs and their successors experience a long waiting due to consecutive TCFs during busy period, the input buffers in the switch may be exhausted. Therefore, it should have a flow-control mechanism to stop further transmission of superordinate stations. We know that 802.3 technologies have a 802.1Qbb *Priority-based Flow Control (PFC)* [18] mechanism in MAC layer. Hence, such method should be employed to solve the problem.

4) ***Bad HPFs estimate:*** In cut-through method, bad frames such as that with FCS error, alignment error, runt, and short frames may be distributed to subordinate switches. These frames may show up in the queue. If a bad frame is also in high-priority (time-critical), it may raise an issue whether to directly drop the frame and initiate a retransmission, for which the handling procedure needs to be verified.

5) ***Capacity Negotiation:*** Since 802.1Qbu and non-802.1Qbu MAC are both used during migration period, we need a capacity negotiation mechanism such as 802.1ab [19] to coordinate various MAC to prevent the incompatabiity problem. The most direct attack on this problem appears to be extending the current Ethernet standards; we are expecting the realization for that.

V. CONCLUSIONS

In this paper, we introduce IEEE 802.1Qbu frame preemption in MAC layer of GbE networks, which is able to provide an absolute hop-to-hop transmission delay guarantee; further, we have particularly addressed the details of implementation obstacle and the enhancement of performance in an end-to-end GbE environment.

Comparing our preemptive priority scheme with original non preemptive-based priority scheduling through system simulations, the performance of the switch which supports PPS with 802.1Qbu is improved significantly. Both the average delay and the jitter are reduced several fold.

Ethernet was now a scalable transport technology, both in speed and distance. Today, Ethernet is extended to enable telecommunications network providers to provide services to customers and to utilize Ethernet technology both in their core and access networks. The 802.1Qbu is the key technology for wide-area Ethernet services widely deployed especially in next-generation Ethetnet technologies such as 40GbE and 100GbE [20]. Coordinating with higher layers QoS protocols and the bandwidth reservation function of packet networks, the 802.1Qbu-based priority scheduling can provide the end-to-end carrier-grade SLA transmission service in a DiffServ based IP network, even the Internet backbone and next-generation technologies such as NGN (Next Generation Network) or FMC (Fixed Mobile Convergence).

REFERENCES

[1] http://www.MetroEthernetForum.org/

[2] IEEE Standard 802.3z/D5, “Media Access Control (MAC) Parameters, Physical Layer, Repeater and Management Parameters for 1000 Mb/s operation,” June 1998

[3] 1000BASE-X Physical Coding Sub layer (PCS) and Physical Medium Attachment (PMA) tutorial. http://www.iol.unh.edu/services/ testing/ge/training/.

[4] IEEE PAR 802.1Qbu, “amendment for frame preemption,” Jan. 2012.

[5] IEEE Standard 802.3ae, “Media Access Control (MAC) parameters, physical layers, and management parameters for 10 Gb/s operation,” June 2002.

[6] T. Skeie, S. Johannessen, and Ø. Holmeide. Introducing End-to-End Priority across Switched Ethernet, Industrial Ethernet Book (IEB), 14:10-16, Mar. 2003.

[7] H. Hoang and M. Jonsson, “Switched real-time ethernet in industrial applications - deadline partitioning,” in proc. 9th Asia-Pacific Conference on Communications (APCC'03), September 2003.

[8] Powerlink. *Ethernet Powerlink Standardization Group*. http://www.ethernet-powerlink.org.

[9] IEC “*Digital data communications for measurement and control – Profiles.* for ISO/IEC 802-3 based communication networks in real time applications”, MT9 / WG1. 2003.

[10] C. Ferner and R. Vetter, “*An Integrated Framework for Implementing Quality of Network Concepts,”* Journal of Network and Systems Management. 10(4). 439-455. Dec. 2002.

[11] Y. Yamada, S. Ohta and H. Uematsu, “Hardware-Based Precise Time Synchronization on Gb/s Ethernet Enhanced with Preemptive Priority,” IEICE Transactions on Communications, vol.89, no.B(3), pp. 683-689, 2006.

[12] X. Tu, J. Li, X. J, X. Li, and X. Wang, “An insertion mode scheduler for improving QoS of real-time services,” International Conference on Communications, Circuits and Systems 2007, pp. 569-570, 2007.

[13] Y. Shimokawa and Y. Shiobara, “Real-Time Ethernet for Industrial Applications,” *Proc. Conf. IEEE Industrial Electronics Soc. (IECON)*, pp. 829-834, 1985.

[14] W. K. Jia and Y. C. Chen, “Performance Evaluation of Ethernet Frame Burst Mode in EPON Downstream Link”, *ETRI Journal*, Special Issue: NGN Transport and Control Technologies, vol. 30, no. 2, pp.290-300. Apr. 2008.

[15] M.-F. Homg, W.-T. Lee, K.-R. Lee, and Y.-H. Kuo, “An adaptive approach to weighted fair queue with QoS enhanced on IP network,” in Proc. IEEE TENCN 2001, vol.1, pp.181-186. Aug 2001.

[16] J. Heisswolf, R. König, and E. Jüirgen Becker,”A Scalable NoC Router Design Providing QoS Support Using Weighted Round Robin Scheduling,” in Proc. IEEE 10th International Symposium on Parallel and Distributed Processing with Applications (ISPA 2012),”pp.625-632. Jul. 2012.

[17] IEEE Standard 802.1p, “LAN Layer 2 QoS/CoS Protocol for Traffic Prioritization,” Sep. 1998.

[18] IEEE Standard 802.1Qbb-2011, “Priority-based Flow Control,” Jul. 2011.

[19] IEEE Standard 802.1ab-2009, “Station and Media Access Control Connectivity Discovery (LLDP),” Jun. 2009.

[20] IEEE P802.3ba 40Gb/s and 100Gb/s Ethernet Task Force. http://www.ieee802.org/3/ba/index.html