Method to Measure Packet Processing Time of Internet Hosts with Specialized Packet-Capture Line Card

Khondaker M. Salehin, Roberto Rojas-Cessa, and Sotirios G. Ziavras

***Abstract*—The packet processing time (PPT) of a host is the time elapsed between the arrival of a packet at the data-link layer and the time the packet is processed at the application layer of the TCP/IP protocol stack. To measure the PPT of a host, stamping the times when these two events occur is needed. However, popular network interface cards (NICs) do not provide such time stamping facility. The use of a packet-capture line card (PCL) may be considered, but the clocks of the PCL and the host under measurement may need to be synchronized, and this is complex as the clock of the host's operating system runs at lower speed than that of the PCL and the PCL may be unable to forward packets to the application layer on real time. In this letter, we propose a scheme to measure the PPT of a host using the Internet Control Message Protocol (ICMP) echo request and reply packets, and a PCL in the same network segment. The proposed scheme does not require clock synchronization between the host under measurement and the PCL. We tested the scheme on different hosts under different transmission speeds, using integrated and extended NICs.**

***Index Terms*—Packet processing time, packet capture hardware, clock resolution, local area network**

I. Introduction

PACKET processing time (PPT) of a host (i.e., workstation) is the time elapsed between the arrival of a packet in the host's input queue of the network interface card, NIC, (i.e., the data-link layer of the TCP/IP protocol stack) and the time the packet is processed at the application layer [1], [2]. As link rates increase faster than processing speeds [2], the role of PPT becomes more significant in the measurement of different network parameters.

One-way delay (OWD) in a local area network (LAN) is an example of a parameter that PPT may impact significantly [3]. Figure 1 illustrates the OWD of a packet P over an end-to-end path, between two end hosts, the source (src) and the destination (dst) hosts. The figure shows the different layers of the TCP/IP protocol stack that P traverses at both end hosts, as defined in RFC 2679 [1]. The transmission time (t_t) and propagation time (t_p) of P take place at the physical layer, the queuing delay (t_q) takes place at the network layers, and the time stamping of the packet creation at src (PPT_{src}) and packet receiving at dst (PPT_{dst}) take place at the application layer of the end hosts. The actual OWD experienced by P from src to dst is:

Khondake M. Salehin and Roberto Rojas-Cessa are with Networking Research Laboratory, ECE Department, New Jersey Institute of Technology, Newark, NJ, 07032. {kms29,rojas}@njit.edu.

Sotirios G. Ziavras is with Computer Architecture & Parallel Processing Laboratory, ECE Department, New Jersey Institute of Technology, Newark, NJ, 07032. E-mail: ziavras@njit.edu.

$$OWD = PPT_{src} + t_t + t_q + t_p + PPT_{dst} \quad (1)$$

However, because of the slow transmission speeds of legacy networking systems, PPT has been considered so far negligible (i.e., $PPT_{src} = PPT_{dst} \simeq 0$). As data rates increase, the contribution of PPT increases.

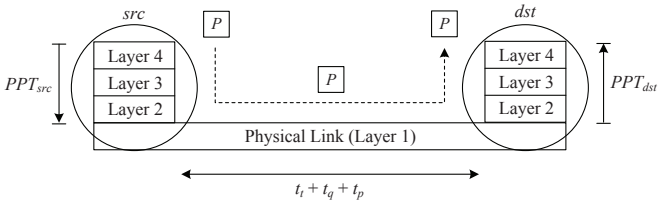


Fig. 1. End-to-end traveling path of packet P between two directly connected hosts.

The error in the measurement of OWD in high speed LANs can be large if PPTs are neglected. For example, the measurement of OWD of 1500- and 40-byte packets between the end hosts with $PPT_{src} = PPT_{dst} = 2$ μs and an average level of queuing delay, $t_q = 40$ μs [4], on a 100-Mb/s link would have an error of 2.5 and 8.5%, respectively. In these calculations, error = $|\frac{(OWD-OWD')}{OWD}| \times 100$ %, $OWD = OWD' + PPT_{src} + PPT_{dst}$, $OWD' = t_t + t_q + t_p$, and $t_p = 0.5$ μs, considering the maximum transmission length (100 m) of a Fast-Ethernet cable and 2×10^8 m/sec of propagation speed in optical fiber [5]. This error increases up to 52% when queuing delay is relieved ($t_q \simeq 0$ μs [4]) for a 40-byte packet. In a similar scenario, the error of OWD on a 1-Gb/s link can be up to 14% (as $t_p = 25$ μs for a 5-km optical cable in Gigabit Ethernet). Therefore, PPT must be considered for an accurate measurement of OWD in LAN.

Similarly, knowledge of the PPT of servers used in financial-trading datacenters would increase customer confidence as OWD is estimated with high accuracy [6].

In a wide area network (WAN), high-resolution OWD measurement can be used to increase accuracy in IP geolocation [7]. In IP geolocation, each microsecond of propagation delay varies the estimated geographic distance by 200 m between two end hosts connected over optical links.

The measurement of the PPT of a host can be complex because the host must record the time a packet arrives at the data-link layer and the time the application layer processes the packet (here, the time stamping performed at the application layer is considered to be the packet-processing event). However, time stamping at the data-link layer is not readily

available in popular and deployed NICs. PPT measurement can be performed by placing a specialized packet-capture line card (PCL) in the same network segment of the host under measurement. However, PCLs have time stamping resolution in the nanosecond range [8], and their use require time synchronization with the host's clock. This is difficult to accomplish as operating systems of a host may provide up to microsecond resolution [9].

In this letter, we present a scheme to measure the PPT of a host using a PCL in the host. We believe this is the first scheme to measure the PPT of a host. The proposed approach does not require synchronization between the host under measurement and the PCL. We present an experimental evaluation of the proposed scheme.

II. Related Work

There is no existing scheme to measure the PPT of an end host to the best of our knowledge, but the measurement of PPT of a router has been considered of interest. A previous work measured PPT of hardware routers in an end-to-end path by instrumenting their input and output links with packet-capture card given that the method has physical access to the routers under test [4]. Here, the PPT of a router is defined as the time interval between the departure of a packet from the ingress queue of the input link and the arrival of the same at the egress queue of the output link of the router; therefore, the actual value is equivalent to two times PPT plus the packet-switching latency through the router's switching fabric.

Another study extended the above mentioned method to measure PPT of software routers by instrumenting the routers with dedicated software processes (i.e., kernel functions) that capture the ongoing traffic between the input and output links both at the data and application layers [10].

Beside PPT of routers, a scheme, called fast-path/slow-path discriminator (*fsd*), was proposed to measure packet generation time of routers using ICMP packets [11]. In *fsd*, the source host sends two different types of probing packets, a direct probe and a hop-limited probe, toward the destination host of a multiple-hop path, consisting of n nodes, for estimating OWD between the end hosts to measure the packet generation time of routers, e.g., node i, where $2 \leq i \leq n-1$, in the path. The direct probe is a specially crafted ICMP echo reply packet with a Time-to-live (TTL) value to enable reaching the destination host through node i, which is the router under test. The hop-limited probe is a specially crafted ICMP echo reply packet spoofed with the destination's IP address as its source address and a TTL value that expires at node i. The hop-limited probe forces node i to generate an ICMP Time Exceeded (TE) packet and sends it to the destination (because of the spoofed source address) so that the OWD of the end-to-end path can be measured. Because the OWD measured by the hop-limited packet is overestimated by the packet generation time of TE at node i, the packet generation time of node i is estimated from the difference between the OWDs of the direct and hop-limited probes over the path.

III. Proposed PPT Measurement Scheme

We propose a scheme to measure PPT that does not require synchronization between the PCL and host under measurement. Figure 2 shows the experimental setup to measure PPT of dst, which is directly connected to src through an Ethernet link. A sniffer (*hsf*), a workstation equipped with a two-port Endace DAG 7.5G2 card [8] as PCL, captures the packets transmitted between src and dst by connecting the two ports of *hsf* to the Ethernet link using a wire tap. In this experimental setup, the propagation times of the sniffed packets are considered negligible because the distance between src and dst is 2 m.

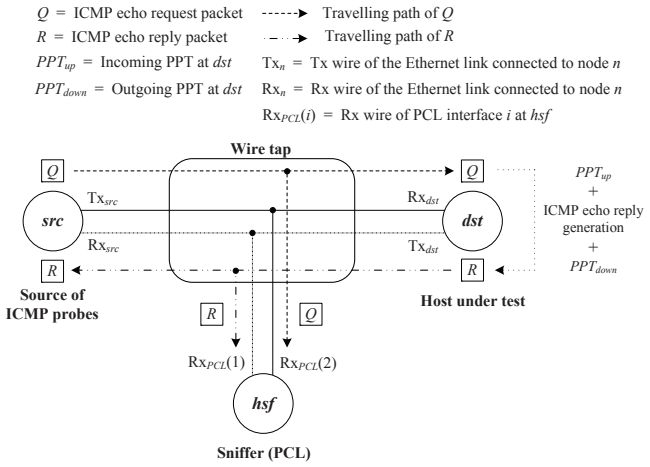


Fig. 2. Experimental setup to measure PPT of dst.

To measure the PPT, src sends an Internet Control Message Protocol (ICMP) echo request packet, Q, to trigger an ICMP echo reply packet, R, at dst. *hsf* captures the exchanged ICMP echo packets and time stamps them at the data-link layer. dst also time stamps the ICMP echo packets, however, at the application layer. The proposed scheme has the measurement uncertainty of the system with the lowest clock resolution, or dst.

Figure 3 presents the time line of the events that take place between the exchange of packets Q and R at dst. Here, the gap measured by hsf, $t_7 - t_1$, includes the receiving time of Q, $t_2 - t_1$, the PPT experienced by Q on its travel up through the TCP/IP protocol stack, $t_3 - t_2$ or PPT_{up}, the time taken by dst to generate R, $t_5 - t_3$, and the PPT experienced by R on its travel down through the TCP/IP protocol stack, $t_7 - t_5$ or PPT_{down}, at dst. The gap measured by dst at the application layer, $t_5 - t_3$, includes the actual time needed to generate R, $t_4 - t_3$, plus the system-call latency for time stamping R, $t_5 - t_4$. Therefore, at dst:

$$(t_7 - t_1) - (t_2 - t_1) - (t_5 - t_3) = PPT_{up} + PPT_{down} \quad (2)$$

If the intervals $t_7 - t_1$, $t_2 - t_1$, and $t_5 - t_3$ are known, PPT_{up}, PPT_{down}, and PPT at dst can be determined from (2) assuming $PPT_{up} = PPT_{down} = PPT$. Here, the assumption may not hold in some load scenarios, but in our experiments dst receives no other traffic nor processes with additional load besides the packet capture.

IV. Experimental Results

We measured PPT of three hosts, a Dell Dimension 3000 (D3000), a Dell Inspiron I531S (I531S), and a Dell Optiplex 790 (DO790) workstations, to evaluate the proposed scheme.

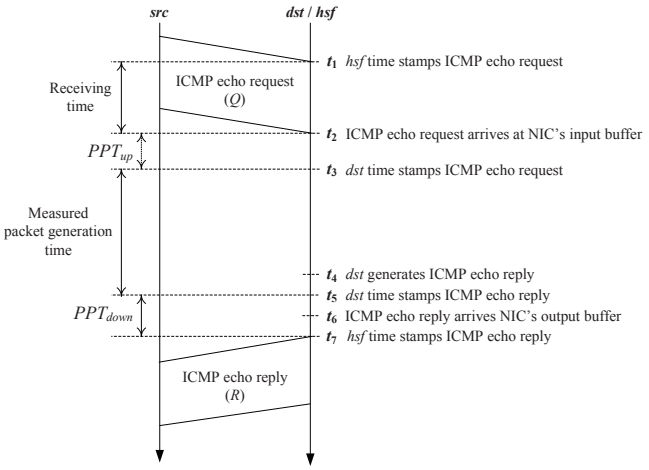


Fig. 3. Timeline of ICMP echo request and reply packets at dst.

Table I shows the specifications of the workstations: 1. CPU (speed), 2. RAM size, 3. RAM speed (data width), 4. PCI bus speed, and 5. Linux kernel version.

TABLE I
HOSTS' SPECIFICATIONS

	Dell Dimension 3000	Dell Inspiron I531S	Dell Optiplex 790
1	Intel Pentium 4 (3 GHz)	AMD Athlon 64 X2 (1 GHz)	Intel Core i3 (3.3 GHz)
2	512 MB	1024 MB	8148 MB
3	400 MHz (64 bits)	667 MHz (64 bits)	1333 MHz (64 bits)
4	266 MB/s	133 MB/s	4 GB/s
5	2.6.18	2.6.18	2.6.35

A. Measured PPTs

We performed PPT measurements on the D3000 and I531S workstations using their integrated Fast-Ethernet NICs, Intel Corp. 82562EZ and nVidia Corp. MCP61, respectively. We tested each workstation under 10- and 100-Mb/s transmission speeds using 500 ICMP echo packets and repeated the test 10 times. Each ICMP echo packet consists of the default frame length of 110 bytes.

Table II shows the summary of the measured PPTs (the PPT column) and the standard deviations (the std column) of the D3000 and I531S workstations. Table II shows that the PPTs of the D3000 workstation are 21 and 14 μs under 10- and 100-Mb/s transmission speeds, respectively. For the I531S workstation, these values are 16 and 7 μs, respectively. The standard deviations of the measured PPTs on both workstations are smaller than 1 μs. The small standard deviation of the measured PPTs show the stability of the proposed scheme.

TABLE II
MEASURED PPTS USING INTEGRATED NICS

dst	Link capacity (Mb/s)	$t_7 - t_1$ (μs)	$t_2 - t_1$ (μs)	$t_5 - t_3$ (μs)	$2 \times PPT$ (μs)	PPT (μs)	std (μs)
D3000	10	151	88	22	41	21	0.31
D3000	100	57	9	22	27	14	0.42
I531S	10	161	88	41	32	16	0.31
I531S	100	63	9	41	13	7	0.42

We performed PPT measurements on the I531S and DO790 workstations using an extended Gigabit-Ethernet NIC, Marvell Tech. 88E8053 PCI-E, and the same number of ICMP echo packets, as used in the previous set of experiments. Table III shows the summary of the measured PPTs (the PPT column) and the standard deviations (the std column) of the above workstations. Here, the PPTs of the I531S workstation with the extended NIC are 71 μs under 10-Mb/s transmission speed, and 63 μs under 100- and 1000-Mb/s transmission speeds. The measured PPTs on the DO790 workstation are 99, 91, and 90 μs under 10-, 100-, and 1000-Mb/s transmission speeds, respectively. The standard deviations of the measured PPTs on both workstations, under each transmission speed, are smaller than 1 μs.

TABLE III
MEASURED PPTS USING EXTENDED NICS

dst	Link capacity (Mb/s)	$t_7 - t_1$ (μs)	$t_2 - t_1$ (μs)	$t_5 - t_3$ (μs)	$2 \times PPT$ (μs)	PPT (μs)	std (μs)
I531S	10	265	88	35	142	71	0.42
I531S	100	170	9	35	126	63	0.48
I531S	1000	162	88	35	126	63	0.70
DO790	10	301	88	17	197	99	0.52
DO790	100	208	9	17	182	91	0.16
DO790	1000	198	1	16	180	90	0.72

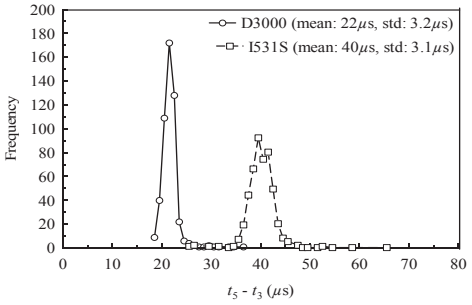
Tables II and III show that the PPTs measured on each workstation using the integrated and extended NICs under 10-Mb/s transmission speed is about 8 μs larger than that under 100-Mb/s transmission speed. The variation in the measured PPTs for these two speeds could be the minimum idle time period required after receiving a packet at the NIC, called Interframe Gap (IFG), as defined by the Ethernet standard [12]. IFG under 10- and 100-Mb/s transmission speeds are 9.6 and 0.96 μs, respectively; therefore, there is a difference of 8.64 μs in IFG, which is close to the variation of the PPTs measured under these two speeds. The measured PPTs on the I531S and DO790 workstations under 100- and 1000-Mb/s transmission speeds are similar because the IFG under the latter speed is 0.096 μs.

The PPTs measured on the I531S workstation using the integrated and extended NICs show that the type of NIC plays a major role in determining the PPT of a host, in addition to the transmission speed. As Tables II and III show, the PPTs measured on the I531S workstation using the extended NIC is around 50 μs larger than that using the integrated NIC.

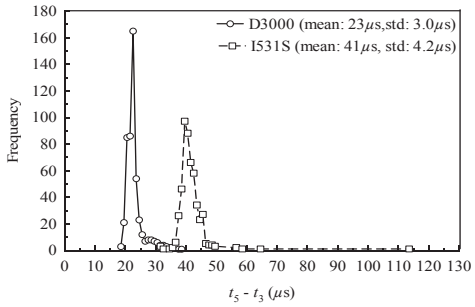
B. Quality of Intermediate Variables

1) ICMP Packet Generation Time: Figure 4 shows the sampled distributions of the ICMP packet generation times ($t_5 - t_3$) of each workstation under a) 10- and b) 100-Mb/s transmission speeds for 500 ICMP packets using the integrated NIC. According to the figure the mean packet generation time measured on the D3000 workstation under 10- and 100-Mb/s transmission speeds are 22 and 23 μs, respectively. On the I531S workstation under above stated transmission speeds, the mean packet generation times are 40 and 41 μs, respectively.

Packet generation times measured on the I531S and DO790 workstations using the extended NIC under 10-, 100-, 1000-Mb/s transmission speeds also have similar (i.e., unimodal and tight) distributions as those in Figure 4. However, the packet generation times according to Tables II and III show that the DO790 workstation has the smallest packet generation time (i.e., 17 μs) among all workstations irrespective of the transmission speed because this workstation has the highest

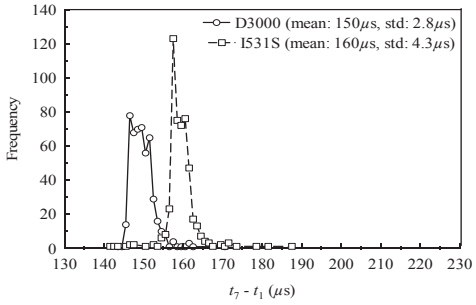


(a) 10 Mb/s

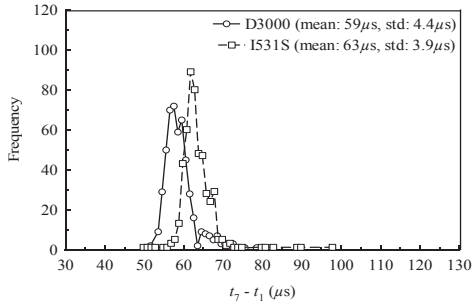


(b) 100 Mb/s

Fig. 4. ICMP packet generation time ($t_5 - t_3$) using the integrated NIC under (a) 10- and (b) 100-Mb/s transmission speeds.



(a) 10 Mb/s



(b) 100 Mb/s

Fig. 5. Time stamping interval between Q and R ($t_7 - t_1$) recorded by hst using the PCL on the workstations using the integrated NIC under (a) 10- and (b) 100-Mb/s transmission speeds.

CPU, RAM, and bus speeds, as shown in Table I. On the other hand, the I531S workstation has different packet generation times (22 and 35 μs) when the PPT is measured using integrated and extended NICs.

2) PCL Time Stamping: Figure 5 shows sampled distributions of the time stamping intervals between Q and R ($t_7 - t_1$), measured by hsf in the integrated NIC based experiments under a) 10- and b) 100-Mb/s transmission speeds. The figure shows that the distributions of the measured PCL time stamping for both transmission speeds are unimodal and tight where the means on the D3000 and I531S workstations are 150 and 59 μs and 160 and 63 μs, respectively. The measured PCL time stamping intervals in the extended NIC based experiments also have unimodal distributions, with different means, like the above stated experiments.

3) Probing Load: We monitored the CPU utilization of each workstation under 10, 100, and 1000-Mb/s transmission speeds while measuring PPT using the proposed scheme. We found that the capture of the probes adds a small CPU load, between 3 and 11%, on the workstations whose PPT is under measurement.

V. Conclusions

We proposed a scheme to measure the PPT of a host (i.e., workstation) using a specialized packet-capture line card in a LAN setup. To measure PPT, we send an ICMP echo request packet to trigger an ICMP echo reply packet at the host under test, and collect time stamps at the data-link and application layers using the clocks of the packet-capture line card and the host, respectively. The scheme does not require synchronization between the host and the packet-capture line card. We tested the proposed scheme on two different hosts connected to the network with integrated and extended NICs running at 10, 100, and 1000 Mb/s. The experimental results show that our scheme can measure PPT of the hosts consistently, and without clock synchronization.

References

[1] G. Almes, S. Kalidindi, and M. Zekauskas. RFC 2679 – A one-way delay metric for IPPM. [Online]. Available: http://www.ietf.org/rfc/rfc2679.txt.

[2] R. Prasad, M. Jain, and C. Dovrolis, "Effects of interrupt coalescence on network measurements," in *Proc. of PAM*, France, 2004, pp. 247-256.

[3] A. Hernandez and E. Magana, "One-way delay measurement and characterization," in *Proc. of ICNS*, Athens, Greece, 2007, pp. 1-6.

[4] K. Papagiannaki, S. Moon, C. Fraleigh, P. Thiran, and C. Diot, "Measurement and analysis of single-hop delay on an IP backbone network," *IEEE JSAC*, vol. 21, no. 6, pp. 908-921, 2003.

[5] R. Percacci and A. Vespignani, "Scale-free behavior of the Internet global performance," *The European Physical Journal B - Condensed Matter*, vol. 32, no. 4, pp. 411-414, 2003.

[6] M. Lee, N. Duffield, and R. Kompella, "Not all microseconds are equal: Fine-grained per-flow measurementswith reference latency interpolation," in *Proc. of ACM SIGCOMM*, Dehli, India, 2010, pp. 27-38.

[7] B. Gueye, A. Ziviani, M. Crovella, and S. Fdida, "Constraint-based geolocation of Internet hosts," *IEEE/ACM ToN*, vol. 14, no. 6, pp. 1219-1232, 2006.

[8] Endace DAG 7.5G2 datasheet. [Online]. Available: http://www.endace.com/assets/files/resources/END_Datash eet_DAG7.5G2 3.0.pdf.

[9] L. De Vito, S. Rapuano, and L. Tomaciello, "One-Way Delay Measurement: State of the Art," *IEEE TIM*, vol. 57, no. 12, pp. 2742-2750, 2008.

[10] L. Angrisani, G. Ventre, L. Peluso, and A. Tedesco, "Measurement of Processing and Queuing Delays Introduced by an Open-Source Router in a Single-Hop Network," *IEEE TIM*, vol. 55, no. 4, pp. 1065–1076, 2006.

[11] R. Govindan and V. Paxson, "Estimating Router ICMP Generation Delays," in Proc. *PAM*, CO, USA, 2002, pp. 1-8.

[12] R. Mandeville and J. Perser. RFC 2889 – Benchmarking methodology for LAN switching devices. [Online]. Available: http://www.ietf.org/rfc/rfc2889.txt.