Scheme for Measuring Queueing Delay of a Router using Probe-Gap Model: The Single-Hop Case

Khondaker M. Salehin and Roberto Rojas-Cessa

Abstract—In this letter, we propose a probe-gap scheme to measure queueing delay of a router in a single-hop path. The proposed scheme is based on estimating the change of an intra-probe gap between a pair of probing packets after they traverse the router under test. The proposed scheme can be applied to online routers and is sensitive to small queueing delays. Our simulation results show that this scheme consistently achieves high measurement accuracy under different levels of queueing delays.

Index Terms—Queueing delay, active measurement, packet pair, intra-probe gap, single hop, clock synchronization.

I. INTRODUCTION

The delay a packet experiences in a router consists of 1) processing delay, 2) queueing delay, and 3) transmission delay. The processing delay is the time a router takes to forward a packet from its input queue to its output queue, including destination lookup (IP lookup [1]), packet-header updates, etc. The queueing delay is the time that a packet waits in a router after arriving at the router queue until it starts transmission through to the output link. Queueing delay is a dynamic network parameter which is proportional to the number (or transmission delays) of packets that are already waiting in the output queue at a given instance of time.

Figure 1 shows examples of two different queueing scenarios at a router, node i. Node i has multiple input links (L1, ..., Ln) and one output link (Lo), with link capacities l1, ..., lo, respectively. The processing time of node i to forward a packet from the input queues (Q1, ..., Qn) to the output queue (Qo) is constant. Figure 1(a) shows that node i has only one traffic flow between Ln and Lo (Fn), where the flow rate is smaller than or equal to l<sub>o</sub>; therefore, packet P<sub>1</sub> from F<sub>n</sub> does not experience additional waiting time at Q<sub>o</sub> before transmission through L<sub>o</sub>. In Figure 1(b), node i carries two traffic flows, one between L<sub>1</sub> and L<sub>o</sub> (F<sub>1</sub>) and the other between L<sub>n</sub> and L<sub>o</sub> (F<sub>n</sub>), where the combined traffic rate of F<sub>1</sub> and F<sub>n</sub> is larger than l<sub>o</sub>. In this scenario, Q<sub>o</sub> receives P<sub>1</sub> from F<sub>n</sub> while packet C<sub>1</sub> from F<sub>1</sub> is about to be transmitted through L<sub>o</sub>; therefore, P<sub>1</sub> will experience a queueing delay equivalent to the transmission time of C<sub>1</sub> over L<sub>o</sub>, or \( \frac{s_c}{l_o} \), where \( s_c \) is the size of C<sub>1</sub>. Measurement of queueing delay of routers is key for characterizing the level of congestion in a network [2]. Precise knowledge of queueing delay allows a service provider to determine and maintain Quality-of-service (QoS) and service-level agreements, respectively, for various delay-sensitive applications in the network. For example, accurate one-way delay (OWD) measurements of end-to-end paths is necessary to attain high accuracy in triangulation-based IP geolocation [3], [4]. Because queueing delays increase OWD of a path (e.g., 10 µs of queueing delay at a router may vary the estimated geographic distance by 2 km, considering the end-to-end delay vs. physical distance mapping over a optical-fiber link [4]), it has the potential to degrade the geolocation accuracy.

Queueing-delay measurement is complex because it may require Internet control message protocol (ICMP) packets [5], [6], infrastructural support from the Internet [6], [7], clock synchronization [8], and complex processing of the sampled data set [7], [9] to obtain high accuracy. In this letter, we propose an active scheme to measure queueing delay of a router using a pair of user datagram protocol (UDP) packets that does not require infrastructural support or clock synchronization. The scheme is simple and it measures queueing delays by determining the change in the intra-probe gap of packet pairs (i.e., the separation between the last bits of the packets in each pair) at the router under test. The scheme is designed to measure queueing delay in a single-hop scenario. It is sensitive to small queueing delays and can be applied to online routers.

II. RELATED WORK

Existing schemes for measuring queueing delay can be divided into two methods: a) active and b) passive. Active schemes proactively inject probing packets in the path under test for measuring queueing delay of an end-to-end path, e.g., Pathchar [5], cing [6], network tomography based schemes [7], [10], etc. Pathchar measures link capacity and queueing delay of an end-to-end path using linear-regression model on the round-trip times (RTTs) of ICMP packets. It is considered network intrusive (i.e., the probing packets themselves can affect the existing traffic flow(s) in the path under test) and requires long measurement time [11]. Cing measures queueing delay using pairs of ICMP packets. Even though this scheme is simple, its feasibility of deployment and measurement accuracy heavily depends on collecting timestamps for every
probing packet from the routers in the path under test [6]. Network-tomography based schemes are complex and they use unicast or multicast packets for measuring queueing delay. These schemes require large infrastructural support from networks as they can measure queueing delay only if multiple end-to-end paths share links amongst them [7], [10].

Passive schemes do not generate probing packets, they rather utilize the existing data traffic in the router under test for queueing-delay measurement, e.g., instrumentation-based [8], [12], [13] and analysis-based [9] schemes. Instrumentation-based schemes measure queueing delay of a router by instrumenting its input and output links using specialized packet-capture cards [8] or kernel processes [12], [13], and a packet monitor [8], [12]. These schemes may require a reliable timing source (e.g., global-positioning system) and synchronization between the packet-capture cards [8]. The analysis-based scheme applies a stochastic model on the captured data traffic for characterizing the queueing delay of a router. However, this scheme is complex and requires prior knowledge of the traffic load, number of traffic flows, packet loss, etc., of the router under test [9].

III. PROPOSED SCHEME

The proposed scheme uses a packet-pair structure consisting of a small heading packet ($P_h$) and a large trailing packet ($P_t$) to measure queueing delay of a router from intra-probe gap measurement between the last bits of $P_h$ and $P_t$, as shown in Figure 2. The proposed scheme uses UDP packets in its probing structure. Therefore, its deployment feasibility in the Internet is high. Because the scheme determines queueing delay from the change in the intra-probe gap, it does not require clock synchronization during measurement.

![Figure 2. A packet-pair structure consisting of a small heading packet, $P_h$, and a large trailing packet, $P_t$.](image)

The basic principle of the proposed scheme is that a non-zero dispersion gap between $P_h$ and $P_t$ (i.e., the separation between the last bit of $P_h$ and the first bit of $P_t$, as Figure 2 shows) may decrease in the presence of cross-traffic packet(s) ahead of $P_h$. This phenomenon would reflect the size of queue in a route with a change (i.e., decrement) in the intra-probe gap, transmission time of $P_t$ plus dispersion gap, between the probing packets. When a small $P_h$ ahead of a large $P_t$ is used in the packet pair, a non-zero dispersion gap can be created at the output link of a router under test [14]--[16], in order to determine a change in the intra-probe gap between the probing packets when there is queueing. For example, in node $i$, if the transmission time of $P_h$ on the output link ($L_{i+1}$) is smaller than the transmission time of $P_t$ on the input link ($L_i$), the probing packets experience a non-zero dispersion $\delta_i$, or:

$$\delta_i = \frac{s_t}{l_t} - \frac{s_h}{l_h}$$  

where $\delta_i > 0$, $s_h$ and $s_t$ are the sizes of $P_h$ and $P_t$, respectively, such that $s_h \leq s_t$, and $l_i$ and $l_{i+1}$ are the link capacities of $L_i$ and $L_{i+1}$, respectively. If the queueing delays of $P_h$ and $P_t$ at node $i$ are $t_{Q(h)}$ and $t_{Q(t)}$, respectively, the queueing delay at node $i$ can be determined from the change in the intra-probe gap ($\Delta$) as:

$$\Delta = \frac{s_h}{t_{i+1}} + \delta_i - \left[ \frac{s_t}{t_{i+1}} + \delta_i \right] - \left( t_{Q(h)} - t_{Q(t)} \right)$$  

where $\left( \frac{s_h}{t_{i+1}} + \delta_i \right)$ is the intra-probe gap between $P_h$ and $P_t$, $\left( \frac{s_t}{t_{i+1}} + \delta_i \right) - \left( t_{Q(h)} - t_{Q(t)} \right)$ is the affected intra-probe gap due to queueing delay at node $i$, $\delta_i \geq t_{Q(h)}$, and $t_{Q(t)}$ is expected to be 0.

Figure 3 illustrates the intra-probe gaps between $P_h$ and $P_t$ with and without cross traffic at node $i$. In Figure 3(a), $P_h$ and $P_t$ arrive at node $i$ from $L_i$ back to back (i.e., with a zero-dispersion gap) when the node does not carry any other traffic; however, the probing packets experience a non-zero $\delta_i$ in the intra-probe gap ($G_i$) on $L_{i+1}$ because of the difference between their transmission times on $L_i$ and $L_{i+1}$, respectively. Figure 3(b) shows the effect of cross traffic at node $i$ on $G_i$ at $L_{i+1}$, which produces a smaller than the expected intra-probe gap. A magnified view of node $i$, Figure 3(c), shows that interference by a cross-traffic packet, $C$, ahead of $P_h$ at the output queue, $Q_{i+1}$, delays the transmission of $P_h$ through $L_{i+1}$, and this additional delay decreases the expected $G_i$ on the output link to reflect the queueing delay at node $i$.

![Figure 3. Intra-probe gap between $P_h$ and $P_t$ (a) without and (b) with cross-traffic interference at node $i$, and (c) magnified view of node $i$ when a cross-traffic packet, $C$, is queued before $P_h$ at the output queue $Q_{i+1}$.](image)

To measure queueing delay, the proposed scheme uses a pre-defined dispersion gap between $P_h$ and $P_t$ considering that the input- and output-link capacities of a router under test is known. Link capacities of a router can be known a priori through physical access or through estimation using existing capacity-measurement schemes [5], [11], [14], [15], [17]. The detailed steps of the proposed scheme to measure queueing delay at node $i$ are presented below:

1. Identify the link capacities $l_i$ and $l_{i+1}$ of node $i$.
2. Determine the packet sizes $s_h$ and $s_t$ such that (1) holds.
3. Estimate the expected intra-probe gap between \( P_h \) and \( P_t \) at node \( i \), \( E[G_i] \), as:

\[
E[G_i] = \frac{s_t}{l_i+1} + \delta_i \tag{3}
\]

4. Send \( P_h \) and \( P_t \) using \( s_h \) and \( s_t \), respectively, with a zero-dispersion gap between the probing packets to node \( i \) from \( L_i \).

5. Time stamp \( P_h \) and \( P_t \) after receiving them on \( L_{i+1} \) and measure the intra-probe gap \( M[G_i] \), where \( M[G_i] \geq \frac{s_t}{l_i+1} \).

6. Determine the change in the intra-probe gap in reference to \( E[G_i] \) as:

\[
\Delta = E[G_i] - M[G_i], \tag{4}
\]

such that \( \Delta > 0 \) as it corresponds to the decrement of the intra-probe gap, whereas \( \Delta \leq 0 \) corresponds to either no change or increment in the intra-probe gap between \( P_h \) and \( P_t \).

7. Discard \( \Delta \) if the corresponding \( M[G_i] = \frac{s_t}{l_i+1} \) since this condition suggests that the queueing delay at node \( i \) could be larger than \( \delta_i \) that is used in step 3.

8. Repeat steps 4 to 6 for \( n > 1 \) times and determine \( \Delta \) in each iteration.

9. Calculate the mean and standard deviation of the sampled set, \( \mu_\Delta \) and \( \sigma \), respectively, to provide a range of estimated queueing delay at node \( i \) as:

\[
\text{Range of queueing delay} = \mu_\Delta \pm \sigma \tag{5}
\]

IV. Simulation Results

We evaluated the performance of the proposed scheme in a simulation environment using ns-2 [18]. Figure 4 shows the simulation setup of a single-hop topology consisting of a router (node 1) connected between \( src \) (node 0) and \( dst \) (node 2). The capacities of the input and output links (\( L_1 \) and \( L_2 \), respectively) of node 1 are 100 Mb/s. The single-hop path carries four cross-traffic flows between node 1 and node 2 over \( L_2 \). The traffic flows are used to build queueing delay at node 1, which is of interest in our evaluation. In this simulation, we did not evaluate the performance of our scheme against other existing active schemes, e.g., cing, because they either use ICMP packets or require infrastructural support from the networks, which are not feasible to emulate in a simulation environment.

![Fig. 4. A single-hop path with n cross-traffic flows, where n = 4.](image)

Table I shows the expected dispersion gaps between \( P_h \) and \( P_t \) (the last column) when different \( s_t \) sizes are used with \( s_h = 64 \) bytes considering the link capacities (i.e., 100 Mb/s) at node 1. We used 64-bytes packet in \( P_h \) as it is the smallest frame size in the Ethernet standard. Because the maximum decrease in the length of the intra-probe gap between \( P_h \) and \( P_t \) is determined by the expected dispersion between \( P_h \) and \( P_t \), the values in the last column of Table I correspond to the upper bounds of the queueing delays that can be measured at node 1 using \( s_h = 64 \) bytes and \( s_t = \{800, 1000, 1200, 1500\} \) bytes for \( P_h \) and \( P_t \). We measured three different queueing delays, i.e., 11.21 \( \mu s \) (6.86 \( \mu s \) ± 4.35 \( \mu s \)), 17.36 \( \mu s \) (10.16 \( \mu s \) ± 7.2 \( \mu s \)), and 35.04 \( \mu s \) (21.44 \( \mu s \) ± 13.6 \( \mu s \)), at node 1 to evaluate the performance of the proposed scheme that are within the upper bounds of the queueing delays, as stated in Table I. The three different queueing delays are generated at node 1 by using 64-, 100-, 200-byte packets in each cross-traffic flow, respectively. The queue-length pattern (the first 100 samples for every 5 \( \mu s \) time interval) at node 1 for the above stated queueing delays are shown in Figure 5.

<table>
<thead>
<tr>
<th>Probing-packet sizes</th>
<th>Link capacity</th>
<th>Dispersion gap</th>
</tr>
</thead>
<tbody>
<tr>
<td>( s_h ) (bytes)</td>
<td>( s_t ) (bytes)</td>
<td>( l_1 ) (Mb/s)</td>
</tr>
<tr>
<td>64</td>
<td>800</td>
<td>100</td>
</tr>
<tr>
<td>64</td>
<td>1000</td>
<td>100</td>
</tr>
<tr>
<td>64</td>
<td>1200</td>
<td>100</td>
</tr>
<tr>
<td>64</td>
<td>1500</td>
<td>100</td>
</tr>
</tbody>
</table>

We evaluate the performance of the proposed scheme by sending 1000 pairs of probing packets between \( src \) and \( dst \), where each pair is separated by a maximum of 100 \( \mu s \). Figure 6 shows the simulation results of the proposed scheme when the upper bounds of the generated queueing delays at node 1 are: (a) 11.21 \( \mu s \) (6.86 \( \mu s \) ± 4.35 \( \mu s \)), (b) 17.36 \( \mu s \) (10.16 \( \mu s \) ± 7.2 \( \mu s \)), and (c) 35.04 \( \mu s \) (21.44 \( \mu s \) ± 13.6 \( \mu s \)). In Figure 6, the hollow boxes and solid circles are the mean of the actual queueing delays and measured queueing delays by the proposed scheme, respectively. The standard deviation of the actual and measured queueing delays are shown with the dashed and solid whiskers, respectively. We repeated the above stated evaluation of the scheme using different number of packet pairs, e.g., 30, 500, etc.; however, the 1000-pair test provided us the best possible results.

The simulation results in Figure 6 show that the proposed scheme achieves high accuracy for the tested queueing delays. For example, the range of the measured queueing delay, illustrated by the upper and lower whiskers from each mean value, overlap with range of the actual queueing delays at the router under test in every instance. We also calculated the error in the measured queueing delay of every case presented in Figure 6, where the error is the ratio between the length of the range of measured queueing delay that is outside the range of the actual value and the total length of the range of measured queueing delay. In majority of the cases, the accuracy of measurement is high (i.e., 100%). Otherwise, the errors are 14 and 50% when \( s_t = 800 \) and 1500 bytes in Figure 6(a), respectively, and 27, 18, and 4% when \( s_t = 800, 1000, \) and 1500 bytes in Figure 6(b), respectively. Considering the dynamic changes of the queueing delays and the errors in the measured values, the proposed scheme consistently achieves high measurement accuracy for different level of queueing delays. The measured values in Figure 6 also show that when the sizes of the probing packets create a larger dispersion in the packet pair, the queueing delay can be measured with high accuracy.
The scheme is simple and can be applied to online routers. It achieves high measurement accuracy. Show that the proposed scheme is sensitive to small queueing delays at the router under test; therefore, it does not require infrastructure support and clock synchronization for measurement. The scheme is simple and can be applied to online routers. Simulation results for three different levels of queueing delays show that the proposed scheme is sensitive to small queueing delay and achieves high measurement accuracy.

V. CONCLUSIONS

We proposed an active scheme to measure queueing delay of routers in a single-hop scenario using a packet-pair model. The proposed scheme uses UDP packets in pairs and identifies the change in the intra-probe gap for measuring queueing delay at the router under test; therefore, it does not require infrastructure support and clock synchronization for measurement. The scheme is simple and can be applied to online routers. Simulation results for three different levels of queueing delays show that the proposed scheme is sensitive to small queueing delay and achieves high measurement accuracy.

REFERENCES