Workgroup: Transport Working Group
Internet-Draft:
draft-morton-tsvwg-interflow-intraflow-
delays-00
Published: 17 May 2021
Intended Status: Informational
Expires: 18 November 2021
Authors: J. Morton   P. Heist

Interflow vs Intraflow Delays

Abstract

   Much current literature discusses queuing delays, and the effects of
   different queue disciplines, active queue management algorithms, and
   congestion control measures on these delays. This draft highlights
   an important distinction between different types of delay, which may
   be helpful to practitioners and theoreticians alike.

Status of This Memo

Copyright Notice

Table of Contents

1. Introduction

Throughput, packet loss ratio, and latency are the three most prominent performance characteristics of Internet paths. Of these, throughput has always been the most heavily marketed to consumers, possibly because it is the only metric from this group in which bigger numbers are better. Packet loss is also closely managed by network engineers, and is mostly kept to usefully low levels in practice, probably because excessive packet loss tends to cripple the throughput of typical congestion-controlled traffic. However, while latency has great practical importance to many Internet applications, it is rarely given the attention it needs for proper management.

One consequence of this neglect is the phenomenon of bufferbloat. Any given Internet path has a natural baseline delay, which is a consequence of the speed of information propagation in the physical media, plus processing delays in network nodes that connect link segments together, plus (for some link types) additional delays associated with shared media negotiation. To this baseline, we must add the delay caused by packets waiting in a queue behind other packets, which occurs if the link is busy. If the queue is permitted to grow too much, these additional queuing delays can become very noticeable to the user, and may even affect the reliability of Internet protocols.

This document does not discuss in detail the many and varied means of controlling latency that are currently or might someday become available. Instead the characteristics of this delay are discussed, including the distinction between "inter-flow induced delay" and "intra-flow induced delay". Typically these two types of delay, despite their similar names, have different effects and may be controlled by different queue mechanisms. Simple queues, however, do not attempt to distinguish them.

To improve the likelihood of distinguishing the names, the terms BFID (Between-Flow Induced Delay) and WFID (Within-Flow Induced

Delay) will be used as synonyms for inter-flow and intra-flow delays, respectively.

2. **Baseline Path Delay (BPD) and Baseline Round-Trip Time (BRTT)**

Definition: The delay on a one-way path or round-trip due entirely to link characteristics and unavoidable processing delays.

For the avoidance of doubt, the word "unavoidable" in this definition refers to the agency of the traffic traversing the path in question, and not to that of the network operators or equipment manufacturers involved.

The speed of light is a fundamental limitation on information transmission velocity, and thus on the minimum latency of a geographically long Internet path. On radio-based links, this limit is approached closely; in optical fibre or copper wires, the transmission velocity is somewhat slower. When avian carriers [RFC1149] are involved, the transmission velocity necessarily falls below the speed of sound. In practice, an allowance of one millisecond round-trip delay per 100km is usually appropriate.

When a packet is received by a network node, it must be directed into a processing buffer for at least long enough to determine in which direction it should be sent next. Since the necessary information is typically in the packet header, this may sometimes be less time than is necessary to receive the entire packet, in which case the head of the packet may be sent onward while the tail is still being received. In other cases, the node may receive the packet in whole before making a processing decision, and may even aggregate the packet with others for efficiency of dispatch. This efficiency in throughput or power consumption may be achieved at the expense of processing delay.

Some link types have significant overhead associated with initiating a transmission, and/or utilise a shared medium into which only one or a small number of stations (out of a larger possible total) may transmit simultaneously. Similar characteristics may also be exhibited by power-saving measures on portable devices. These may result in significant and/or variable delays in forwarding over these links, which cannot be avoided by altering characteristics of the traffic itself.

In practice, an Internet packet can be sent around the world in about 300 milliseconds with current technology. The round-trip latency between Eastern Europe and Western North America is presently about 160 milliseconds. A "typical" Internet round-trip delay can be taken to be 80 milliseconds, though more localised paths are significantly quicker in this respect. Within a LAN or a

datacentre, the baseline delay will often be less than one
millisecond.

Whenever two or more packets require sending over the same link
within the time required to send either one of them, link contention
exists and must be resolved. This generally involves either placing
packets into a queue or discarding them. These practices are not
within the definition of "baseline" delays, but influence "induced"
delays as below.

3. **Between-Flow Induced Delay (BFID)**

    Definition: The delay which the presence and volume of one flow
    induces in traffic belonging to another flow.

    When packets are held in a queue awaiting delivery, the order in
    which these packets are dequeued is significant for managing delay.
    The most common strategy to date is to employ a simple FIFO queue.
    This means that all traffic traversing the same link at about the
    same time experience the same amount of queue delay. It also means
    that a single flow occupying a large part of the queue induces a
    large delay to all other flows sharing that queue, even if without
    the presence of that single flow there would be no need for queuing
    at all. This is the essence of BFID.

    Large BFIDs can be avoided by discriminating flows with high queue
    occupancy from those with little or no queue occupancy, and queuing
    them separately. One effective method of doing so, that is, placing
    every flow in its own FIFO and serving them in deficit-round-robin
    order, is described in detail by [RFC8290]; this "flow-isolating"
    mechanism reduces the maximum BFID to the serialisation time of one
    full-size packet from each active flow, and can be implemented with
    or without the use of Active Queue Management. It is also feasible
    to merely categorise flows into queue occupancy bands and use a
    separate FIFO only for each band; this renders the BFID experienced
    by each flow proportionate to the BFID it produces.

    BFID can also be reduced in a simple FIFO by implementing Active
    Queue Management. This is because in a simple FIFO, BFID and WFID
    have the same cause and extent, so reducing WFID also reduces BFID.
    The extent to which BFID can be reduced by this method is limited
    compared to dedicated methods, and a significant amount of delay
    variation typically remains, but this is significantly better than
    allowing a large, uncontrolled BFID to exist.

    Capacity-seeking flows with little latency sensitivity are
    particularly prone to produce BFID, while latency-sensitive flows
    that typically use little capacity are particularly affected by
    receiving BFID.

4. **Within-Flow Induced Delay (WFID)**

   Definition: The delay which the presence and volume of one flow
   induces in traffic belonging to itself.

   Regardless of the order in which packets are delivered from a queue,
   if more than one packet belonging to a given flow is held in a
   queue, one of them induces delay to the other by occupying
   transmission capacity ahead of it. In general this WFID is
   calculable as the product of the packet delivery rate of that flow
   and the packet occupancy in the queue of that flow.

   In congestion-controlled flows, one typical cause of WFID is that
   the flow's congestion window exceeds the baseline Bandwidth-Delay
   Product (BDP) of the flow's path, and the queue in question is the
   controlling bottleneck defining the Bandwidth factor. This is a
   natural result of capacity-seeking behaviour, where the congestion
   window is increased continuously until some explicit signal of
   capacity overload is detected. If the queue is large and does not
   implement Active Queue Management, WFIDs of many seconds are easily
   achieved and have been observed in practice.

   Another typical cause is that the sender emitted a short-term burst
   of packets, which subsequently collects in one or more downstream
   queues and is thereby spread out in time at the receiver. This cause
   also applies to non-congestion-controlled protocols that can have
   large datagram payloads. This form of WFID is usually harmless to
   the flow causing it, except that large bursts can exceed the
   capacity of a queue to absorb them, resulting in packet loss and the
   need for retransmission.

   In simple FIFOs, or where a flow-isolating mechanism is defeated by
   hash collisions or information hiding, the presence of WFID also
   implies the presence of an equal degree of BFID to any other flows
   sharing that queue. This implies a responsibility to try to minimise
   WFID, even when the flow causing it is not very sensitive to its
   effects (as is typical of capacity-seeking protocols). Buffer sizing
   guidelines (eg. typical BDP / sqrt(flows)) are among the simplest
   ways to limit WFID to tolerable levels.

   Active Queue Management (AQM) is the primary means of effectively
   controlling WFID without impairing the ability to absorb short-term
   bursts of traffic, by sending congestion signals to flows
   experiencing high queue occupancy. Early forms of AQM were only able
   to generate congestion signals by artificially inducing packet loss.
   ECN [RFC3168] introduced the ability to flag congestion on a packet
   without dropping it. AQM may be used alone as in [RFC8289], or in
   conjunction with flow-isolation mechanisms as in [RFC8290]. In the

latter case, both WFID and BFID are addressed individually by
natively appropriate mechanisms.

Some flows fail to respond to congestion signals applied by an AQM.
If these flows cause high degrees of WFID, it is reasonable and
probably wise to include a backstop mechanism to prevent them from
completely dominating the queue, by artificially inducing enough
packet loss (without using the ECN "flag" mechanism) to materially
reduce that flow's queue occupancy. If possible, this "queue
protection" mechanism should be specific to the offending flow(s),
such that it mostly avoids dropping packets from appropriately
responsive or inoffensive flows. Without these features, an
unresponsive flow could seriously impair the quality of service of
other flows, either by producing a lot of BFID, or by causing an
overzealous AQM to drop the wrong packets.

5. Latency Sensitivity of Traffic

Some protocols and applications are more sensitive to latency, and
variations in delay, than others. Variations in delay are often
referred to as "jitter", which is the origin of the term "jitter
buffer" commonly used in some types of application.

If the response time for a DNS request exceeds 2 seconds, a timeout
occurs and the request may be retried or an error reported to the
application. Since DNS is a critical support protocol for many
Internet applications, the degree of BFID should be kept well below
2 seconds in all foreseeable cases. DNS timeouts are a significant
cause of user-visible application failure, often resulting in manual
retries and user frustration. If DNS stops working, "the Internet is
down".

Congestion-controlled reliable transports, such as TCP, can have
difficulty recovering from occasional packet loss efficiently if the
effective RTT is high, which can be caused by excessive WFID. The
recovery process may be visible to the user in the form of a "stall"
in the progress of a download or rendering of a Web page, since data
received beyond the lost packet(s) cannot be delivered to the
application until the lost packet's retransmission is successfully
received. The duration of the stall is proportional to the effective
RTT, so keeping WFID low can maintain reasonably smooth perceived
application performance even in the face of packet loss and
recovery. Implementing AQM with ECN can also eliminate packet loss
entirely, if the underlying path is sufficiently reliable.

NTP assumes that delay is approximately symmetric on each path. In
the case of BPD, that is usually true except in certain highly
asymmetric routing scenarios. The assumption is violated, however,
in the case where BFID persists for an extended period of time that

exceeds NTP's built-in filter against it. Even quite small degrees of BFID can distort NTP synchronisation.

VoIP and videoconferencing protocols can usually tolerate a surprisingly high BRTT, often more than the human users communicating over them. To accommodate delay variations caused by inherent link characteristics, BFID and WFID, they require jitter buffers. The round-trip latency presented to the users is the sum of the BRTT and the jitter buffers in both directions, so the jitter buffers are tuned at runtime to be only as large as necessary to accommodate observed delay variations. Since these protocols usually don't produce much WFID, protecting them from BFID to the greatest extent practical will noticeably improve perceived call quality.

Multiplayer games are among the most latency-sensitive applications visible to consumers. The effective RTT determines how quickly it is possible for each player to perceive situations in the game and transmit responses to them. In very fast-paced games, every millisecond is considered a valuable competitive edge, and experienced players become highly sensitive to even minor glitches caused by network disturbances. In slower-paced games, there is slightly more tolerance, but a significant "lag spike" at an inopportune moment will still be noticed. Crucially, a defeat caused by such a glitch is far more difficult for a player to accept than one caused by his own mistakes or an opponent's genuinely superior performance. Accordingly, this class of application requires strictly minimising both BRTT and BFID, even at the expense of throughput, and should not be routed over links with significant inherent delay variation characteristics.

6. Security Considerations

This is an informational document and raises no security considerations.

7. IANA Considerations

There are no IANA considerations.

8. Informative References

[RFC1149] Waitzman, D., "Standard for the transmission of IP datagrams on avian carriers", RFC 1149, DOI 10.17487/ RFC1149, April 1990, <https://www.rfc-editor.org/info/ rfc1149>.

[RFC3168] Ramakrishnan, K., Floyd, S., and D. Black, "The Addition of Explicit Congestion Notification (ECN) to IP", RFC 3168, DOI 10.17487/RFC3168, September 2001, <https:// www.rfc-editor.org/info/rfc3168>.

[RFC8289]   Nichols, K., Jacobson, V., McGregor, A., Ed., and J.
            Iyengar, Ed., "Controlled Delay Active Queue Management",
            RFC 8289, DOI 10.17487/RFC8289, January 2018, <https://
            www.rfc-editor.org/info/rfc8289>.

[RFC8290]   Hoeiland-Joergensen, T., McKenney, P., Taht, D., Gettys,
            J., and E. Dumazet, "The Flow Queue CoDel Packet
            Scheduler and Active Queue Management Algorithm", RFC
            8290, DOI 10.17487/RFC8290, January 2018, <https://
            www.rfc-editor.org/info/rfc8290>.

Authors' Addresses

Jonathan Morton
Kokkonranta 21
FI-31520 Pitkajarvi
Finland

Phone: +358 44 927 2377
Email: chromatix99@gmail.com

Peter G. Heist
Redacted
463 11 Liberec 30
Czech Republic

Email: pete@heistp.net