MPTCP
Internet Draft
Intended status: Informational
Expires: Dec 18, 2019

F. Song
H. Zhang
Beijing Jiaotong University
H. Chan
A. Wei
Huawei Technologies

June 19, 2019

One Way Latency Considerations for MPTCP
draft-song-mptcp-owl-06

Status of this Memo

This Internet-Draft is submitted in full conformance with the provisions of BCP 78 and BCP 79.

Internet-Drafts are working documents of the Internet Engineering Task Force (IETF). Note that other groups may also distribute working documents as Internet-Drafts. The list of current Internet-Drafts is at http://datatracker.ietf.org/drafts/current/.

Internet-Drafts are draft documents valid for a maximum of six months and may be updated, replaced, or obsoleted by other documents at any time. It is inappropriate to use Internet-Drafts as reference material or to cite them other than as "work in progress."

This Internet-Draft will expire on Dec 18, 2019.

Abstract

This document discusses the use of One Way Latency (OWL) for enhancing multipath TCP (MPTCP). Several usages of OWL, such as retransmission policy and crucial data scheduling are analyzed. Two kinds of OWL measurement approaches are also provided and compared. More explorations related with OWL will be contribute to the performance of MPTCP.

Table of Contents

1. Introduction

The terminal hosts and the intermediate devices in the Internet both have been equipped with more and more physical networkinterfaces basically. The efficiency of interfaces, which had been widely used in packet forwarding at the terminal hosts, had been confirmed and utilized [RFC6419]. In addition,in order to aggregate more bandwidths,reduce packet delays, and provide better service, the increased capacity provided by multiple paths created by multiple interfaces is used. Unlike traditional TCP [RFC0793], many transport layer protocols, such as MPTCP [RFC6182] [RFC6824] enable the terminal hosts to concurrently transfer data on top of multiple paths to greatly increase the overall throughput.

Round-trip time (RTT) is commonly used in congestion control and loss recovery mechanism for data transmission. However,the key issue with data transmission is simply the delay in data transmission along the path that does not include the return. The uplink and downlink delays between two peers can be very different. RTT can be

easily influenced by the latency in the oppsite direction along a path, which does not accurately reflect the delay in data transmission along the path. Therefore, it is recommended to use One Way Latency (OWL) to describe the exact latency from sending data to receiving time data.

It is explained in this document that the full use of One Way Latency (OWL) in the transmission process can further improve the performance of the current practice of MPTCP. It may be asymmetric of the OWL components in the forward and reverse direction of a RTT so that it can provide a better measure to the user such as for congestion control even with the regular TCP. It will be more benefits when there are multiple paths to choose.

The necessary considerations of OWL in MPTCP has been discussed in this document. The structure of this document is as follows: First, the use cases of several OWLs in MPTCP are analyzed. Second, two OWL measurements are listed and compared. Finally, the precautions related to security and IANA are given.

The application programmers whose products may benefit from MPTCP will be potential target audiences for this document significantly. This document also demonstrates the necessary information for MPTCP developers to implement the new version of the API in the TCP/IP network stack.

2. Conventions and Terminology

The key words "MUST", "MUST NOT", "GLUIRED", "SHALL","SHALL NOT", "SHOULD", "SHOULD NOT", "RECOMMENDED", "MAY", and "OPTIONAL" in this document are to be interpreted as described in [RFC2119].

One Way Latency (OWL): the propagation delay between a sender and a receiver from the time a signal is sent to the time the signal is received.

3. Potential Usages of OWL in MPTCP

There are many OWL use cases when the sender and receiver enable MPTCP. Although, only 5 use cases are illustrated in this document, more explorations are still needed.

3.1. Crucial Data Scheduling

During the transfer process, it is usually necessary to send some basic data to the destination immediately. Examples of such data include multimedia key frames and high priority appearance communication blocks. No one can guarantee the order of arrival by using an RTT with multiple paths alone.

The data rate in any given link can be asymmetric. In addition, the delay in a given direction can vary depending on the number of packet queues. Therefore, the same as the opposite direction illustrated in Figure 1, the positive delay in the path is not important.

```
     --------  OWL(s-to-c,path1)=16ms  <--------
   /                                            \
  |    ----->  OWL(c-to-s,path1)= 5ms    -----    |
  |  /           RTT(path1)=21ms              \   |
  | |                                          |  |
+------+                                    +------+
|      |-----> OWL(c-to-s,path2)= 8ms   -----|      |
|Client|                                    |Server|
|      |-----  OWL(s-to-c,path2)= 8ms  <-----|      |
+------+          RTT(path2)=16ms           +------+
  | |                                          |  |
  |  \                                        /   |
  |    ----->  OWL(c-to-s,path3)=10ms    -----    |
   \                                            /
     --------  OWL(s-to-c,path3)= 8ms  <--------
                   RTT(path3)=18ms
```

Figure 1. Example with 3 paths between the client and the server with OWL as indicated in the figure. RTT information alone would indicate to the client that the fastest path to the server is path 2, followed by path 3, and then followed by path 1. Although path 2 is the fastest, whereas OWL indicates to the client that the fastest path to the server is path 1, followed by path 2, and then followed by path 3.

For critical data transfers, the sender can easily select a faster path by using OWL measurements (forward delay). In addition, the confirmation of these critical data can be sent on the path with minimal reverse delay. If the duplex communication mode is adopted, the piggybacking is also useful.

3.2. Congestion control

Congestion in a given direction does not necessarily imply the congestion in the reverse direction.

```
     --------  No congestion (path 1)  <--------
   /                                            \
  |    ----->  Congestion    (path 1)  -----     |
  |  /                                     \    |
  | |                                       |   |
+------+                                 +------+
|Client|                                 |Server|
+------+                                 +------+
  | |                                       |   |
  |  \                                     /    |
  |    ----->  No congestion (path 2)   -----   |
   \                                            /
     --------  Congestion    (path 2)  <--------
```

Figure 2. Example of a congestion situation with 2 paths between the client and the server. There is congestion from client to server along both path 1 and path 2. RTT information alone will indicate congestion in both paths, whereas OWL information will show the client that path 2 is the more lightly loaded path to get to the server.

We can use OWL instead of RTT to better describe network congestion in a given direction. Especially when congestion can be the case in a one-way path, congestion in the path from the client to the server is different from congestion in the path from the server to the client. The delay of interest for data transmission along a path cannot be reflected by the RTT accurately. For MPTCP, the client needs to choose a more lightly loaded path to send packets [RFC6356]. Instead of comparing the RTT among different paths, it should use OWL to compare among the paths.

Current version of MPTCP includes different kinds of congestion control mechanisms [RFC6356]. The network congestion situation in a single direction could be better described by reasonably utilizing OWL.

3.3. Packet Retransmission

Continuous Multi-Path Transport (CMT) improves throughput by simultaneously transmitting new data from the source to the target host through multiple paths. However, when the packet is identified as lost by three repeated acknowledgments or timeouts, the sender needs to select the appropriate retransmission path. Outstanding packets on multiple paths may reach to the destination disorderly and trigger Receive Buffer Blocking (RBB) problem (Figure 3), which will further affect the transmission performance, due to the popular mechanisms of sequence control in reliable transport protocols.

```
             Packetwith octets sequence #  0- 499(lost)
        ---> Packetwith octets sequence #1000-1499(rcvd) ------
    /        Packetwith octets sequence #2000-2499(rcvd)       \
   |                                                            |
  +------+                                                 +--------+
  |Sender|                                                 |Receiver|
  +------+                                                 +--------+
   |                                                           |
    \        Packetwith octets sequence # 501- 999(lost)       /
      -----> Packetwith octets sequence #1501-1999(lost) -----
             Packetwith octets sequence #2501-2999(lost)
```

Figure 3. Example of Receive Buffer Blocking: The packet containing octets 0-499 is lost. On the other hand the packets containing Octets 500-999, 1000-1499, 1500-1999, 2000-2499, and 2500-2999 have all been received. The octets 500-2999 are then all buffered at the receiver, and are blocked by the missing octets 0-499.

By using the results of the OWL measurements, the sender can quickly determine the specific path of the positive minimum delay. Once the receiver gets the most needed packet, the RBB can be released and submitted to the upper layer.

3.4. Bandwidth Estimation

Understanding bandwidth conditions such as packet scheduling and load balancing is critical. OWL can be integrated with bandwidth estimation methods without disrupting the regular transmission of packets.

3.5. Shared Bottleneck Detection

Fairness is essential when MPTCP and normal TCP coexist in the same network. The sender can treat the OWL measurement as a sample process for shared bottleneck detection, and the sender adjusts the amount of packets on multiple paths accordingly.

4. OWL Measurements in TCP

The timestamp option in TCP [RFC7323] may be invoked to estimate latency. The time (TSval) of sending the data is provided in the option when sending data. The receiver acknowledges the receipt of this data by echoing this time (TSecr). And also the time (TSval) of sending this acknowledgment is provided. Although there are two problems, these differences in time when validating data from the sender can help estimate the OWL from the sender to the receiver.

First, there may be a delay from the time the recipient of the data is received until the time the acknowledgment is sent. Then, the above value can be the upper limit of the OWL.

Second, the clock between the sender and the receiver may not be synchronized. OWL can only be displayed in different paths by the above measures when the clock is synchronized. The comparison of OWL between different paths is limited to showing the OWL difference between them without clock synchronization.

Two kinds of OWL measurement approaches are available: absolute value measurement and relative value measurement.

In order to obtain the absolute value of OWL, the primary condition of measurement is clock synchronization. End hosts can calibrate the local clock with the remote NTP server by using network time protocol (NTP) [RFC5905]. The additional information or optional capabilities can even be added via extension fields in the standard NTP header [RFC7822]. The calibration accuracy can reach to the millisecond level in less congested situations. The obvious burden here is to persuade the end hosts to initialize the NTP option.

In some cases it is more than enough to get the relative value of OWL to build an application on it. For example, the sender may only care which path has the minimum forwarding delay when retransmission is required. When estimating bandwidth, the difference in forward latency in all available paths is required, ie incremental latency.

Both sides could obtain therelative value of OWL by exchanging with correspondent end host the local timestamps of receiving and sending the packets.

Overhead is an additional protocol requirement and synchronization accuracy, while OWL's absolute value measurement is more convenient for the application. Instead, relative values are not needed to worry about accuracy, and overhead is to add timestamps to the original protocol stack.

5. Security Considerations

This document does not contain any safety precautions. However, the future application of OWL in MPTCP definitely needs to establish related mechanisms to improve security.

6. IANA Considerations

This document presents no IANA considerations.

7. References

7.1. Normative References

[RFC2119] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", BCP 14, RFC 2119, DOI 10.17487/RFC2119, March 1997, <http://www.rfc-editor.org/info/rfc2119>.

7.2. Informative Reference

[RFC0793] Postel, J., "Transmission Control Protocol", STD 7, RFC 793, DOI 10.17487/RFC0793, September 1981, <http://www.rfc-editor.org/info/rfc793>.

[RFC5905] Mills, D., Martin, J., Ed., Burbank, J., and W. Kasch, "Network Time Protocol Version 4: Protocol and Algorithms Specification", RFC 5905, DOI 10.17487/RFC5905, June 2010, <http://www.rfc-editor.org/info/rfc5905>.

[RFC6182] Ford, A., Raiciu, C., Handley, M., Barre, S., and J. Iyengar, "Architectural Guidelines for Multipath TCP Development", RFC 6182, DOI 10.17487/RFC6182, March 2011, <http://www.rfc-editor.org/info/rfc6182>.

[RFC6356] Raiciu, C., Handley, M., and D. Wischik, "Coupled Congestion Control for Multipath Transport Protocols", RFC 6356, DOI 10.17487/RFC6356, October 2011, <http://www.rfc-editor.org/info/rfc6356>.

[RFC6419] Wasserman, M. and P. Seite, "Current Practices for Multiple-Interface Hosts", RFC 6419, DOI 10.17487/RFC6419, November 2011, <http://www.rfc-editor.org/info/rfc6419>.

[RFC6824] Ford, A., Raiciu, C., Handley, M., and O. Bonaventure, "TCP Extensions for Multipath Operation with Multiple Addresses", RFC 6824, DOI 10.17487/RFC6824, January 2013, <http://www.rfc-editor.org/info/rfc6824>.

[RFC7323] Borman, D., Braden, B., Jacobson, V., and R. Scheffenegger, Ed., "TCP Extensions for High Performance", RFC 7323, DOI 10.17487/RFC7323, September 2014, <http://www.rfc-editor.org/info/rfc7323>.

[RFC7822] Mizrahi, T. and D. Mayer, "Network Time Protocol Version 4 (NTPv4) Extension Fields", RFC 7822, DOI 10.17487/RFC7822, March 2016, <http://www.rfc-editor.org/info/rfc7822>.

Authors' Addresses

Fei Song
Beijing Jiaotong University
Beijing, 100044
P.R. China

Email: fsong@bjtu.edu.cn

Hongke Zhang
Beijing Jiaotong University
Beijing, 100044
P.R. China

Email: hkzhang@bjtu.edu.cn

H Anthony Chan
Huawei Technologies
5340 Legacy Dr. Building 3
Plano, TX 75024
USA

Email: h.a.chan@ieee.org

Anni Wei
Huawei Technologies
Xin-Xi Rd. No. 3, Haidian District
Beijing, 100095
P.R. China

Email: weiannig@huawei.com