Internet Engineering Task Force
Internet-Draft
Intended status: Informational
Expires: 17 March 2022

Q.W. Wu
H.L. Li
Q.Z. Zhang
J.L. Liu
Tsinghua University
13 September 2021

Predictable Multipath TCP (MPTCP) extension
draft-qian-mptcp-predict-01

Abstract

Multipath TCP (MPTCP) adds the capability of using multiple paths to a regular TCP session, which is quite suitable for integrated network scenarios where multiple paths almost always exist. However, link handover and link on-off switching happen frequently in network systems that integrate different systems (especially the systems that continually move), which defeats the quality of MPTCP connections. Information about the link handover and on-off switching in above-mentioned scenarios can be predicted in advance, but MPTCP is not capable of utilizing the prediction information. This document suggests MPTCP be extended with the capacity of obtaining and utilizing the prediction information. Furthermore, the document describes one possible way to enhance MPTCP with prediction information, which proposes a modified MPTCP scheduler utilizing link on-off prediction information.

Copyright Notice

Table of Contents

1. Introduction

Multipath TCP (MPTCP) adds the capability of using multiple paths to a regular TCP session [RFC793]. The motivations for this extension include increasing throughput, overall resource utilization, and resilience to network failure, and these motivations are discussed, along with high-level design decisions, as part of the multipath TCP architecture [RFC6182].

In integrated networks that integrate different systems (especially the systems that continually move), like Integrated Satellite-Territorial Network (ISTN) and aviation aircrafts network that simultaneously using satellite communications and ATG (Air to Ground) Broadband, multiple paths between different hosts almost always exist and therefore it is quite suitable to apply MPTCP in such scenarios to increase throughput and resilience to network failure. However, link handover and link on-off switching happen frequently in those scenarios and they severely decrease the quality of MPTCP transmission connections.

On the other hand, some information about the link handover and on-off switching in above-mentioned scenarios can be predicted in advance according to prior knowledge including satellite orbit, preset route of airlines, the fixed position of the ATG ground stations and so on. But MPTCP is not capable of obtaining or utilizing the prediction information, which leads to a contradiction between the existence of prediction information and the disability of utilizing the information in MPTCP.

The main idea of this document is to suggest that MPTCP be extended with the capacity of obtaining and utilizing the prediction information. This document first describes how link handover and on-off switching weaken MPTCP and how some information about the link changes can be predicted. Finally, this document shows an example of utilizing prediction information in MPTCP.

The target readers of this document are protocol designers and network researchers who work on MPTCP.

2. Terminology

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

This document uses the MPTCP terminology introduced in [I-D.ietf-mptcp-rfc6824bis] [RFC6824].

RTT: Round-Trip Time;

ISTN: Integrated Satellite-Territorial Network;

QoS: Quality of Service;

LEO: Low Earth Orbit;

GEO: Geostationary Orbit;

ATG: Air To Ground.

3. Problem: How do handover and on-off switching weaken MPTCP?

In integrated networks, multiple and heterogeneous paths almost always exist between different hosts. However, link handover and on-off switching happen frequently in these networks. This part describes how link handover and on-off switching weaken MPTCP performance.

3.1. How Link Handover weakens MPTCP

QoS including bandwidth, RTT of paths exist in different network systems are quite different from each other. In integrated networks, link handover between different systems often happen. For example, when the satellite link of an aircraft handovers from a LEO satellite to a GEO satellite, the end-to-end delay will increase from tens of milliseconds (ms) to more than 500 ms. If the size of congestion window doesn't increase with the end-to-end delay, it will limit and decrease the transmission rate of the sender. What's more, the large delay results in a slow congestion window increase, which means the sender will stay at low transmission rate for a long period. On the other hand, MPTCP scheduler usually schedules packets to different paths based on the paths' RTT, therefore the end-to-end delay changes will also result in more out-of-order packets at the receiver.

3.2. How Link On-off switching weakens MPTCP

On-off switching of links also happen in integrated networks. For example, when an aviation aircraft connects to Internet with satellite communications and ATG (Air to Ground) Broadband link simultaneously, the ATG link may switch between on and off states as the aircraft flies, due to the discontinuity of ATG Broadband resulted from complicated topography (like mountains and lakes). As a transport layer protocol, the only way that MPTCP detects disconnection and reconnection of links is to analyze the packets that the receiver replied based on timers. Therefore, it usually takes MPTCP a long time to detects link disconnection and reconnection. When a link suddenly disconnects, the unsent data scheduled to that link will not be transmitted in time and it will take MPTCP scheduler a long time to find the link off and retransmit the data, which results in out-of-order packets at the receiver. What's more, if the sender can't receive ACKs for quite a long time, the send window of all the sub-flows will get exhausted. When a link reconnects, it will also take MPTCP a period to find the link on and schedule some data to the link.

4. Prediction Information

4.1. Predictability

In emerging integrated networks, some new nodes including LEO satellites, GEO satellites, aviation aircrafts, ground base stations, ships on the ocean and high-speed railways are involved. They are different: some of them, like ground base stations, are at fixed positions on the earth while others keep moving at expectable speed on their pre-defined orbits or routines on earth, on ocean or in the space. These characteristics mean that the relative positions between them are predictable and furthermore, the state of the network links between them are predictable.

4.2. Predictable Information

For link handover, predictable information include the time handover happens, link bandwidth and end-to-end delay changes after handover and so on. For on-off switching, predictable information include the time link disconnection and reconnection happens, link bandwidth and end-to-end delay when link reconnects and so on.

5. An example: Scheduler Utilizing Prediction Information

This part shows an example of utilizing prediction information to enhance MPTCP under a link on-off switching scenario.

5.1. Scenario

It’s suitable for Aviation aircrafts to connect to Internet with MPTCP, where ATG Broadband and Satellite Communications are simultaneously available. However, the complicated topography (like mountains and lakes) leads to the discontinuity of ATG Broadband, and therefore defeats the quality of aircraft network service (Refer to Section 3.2). On the other hand, the on-off connection state of the ATG Broadband access link of the aircrafts can be predicted from the preset route of airlines and the position of the ATG ground stations in advance of the conversion of link connection states.

5.2. Modification to Scheduler: How to utilize Prediction Information

The default scheduling principle of MPTCP scheduler is "min-RTT" principle, which schedules each packet to the path with minimum RTT within the paths whose send window is not full. The modified version is called "MPTCP-P", which introduces prediction information to calculate a modified round trip time called "RTT-P" and the calculating details are as follows:

The prediction information that MPTCP-P requests include a) the time when the next disconnection of the ATG link happens, b) the time when the next reconnection of the ATG link happens, which are respectively denoted as time_down and time_up. The current time is denoted as time_now. The RTT calculated by default method of TCP is denoted as RTT_Orignal.

When the ATG link of the aircraft is connected (i.e. time_now < time_down):

RTT-P = RTT_Orignal;

When the ATG link of the aircraft is disconnected (i.e. time_down < time_now < time_up):

RTT-P = RTT_Orignal + time_up - time_now;

MPTCP-P schedules packets in "min-RTT-P" principle, which schedules each packet to the path with minimum RTT-P within the paths whose send window is not full.

5.3. Enhancement

Without prediction information about link on-off switching, original MPTCP detects link connection state based on received packets and timers. Therefore, it takes time for MPTCP to find link disconnected or reconnected and then accordingly schedule packets to different paths.

With prediction information about link on-off switching, MPTCP-P could a) reduce the packets to be allocated to a path and cancel timers for the packets sent through a path when the predictable link of the path is going to disconnect, b) allocate packets to a reconnecting path at the moment it gets reconnected according to the prediction information.

5.4. Corresponding extension for path management

In order to keep the temporarily disconnected subflow valid in case it is in a quite long disconnected period, it might be necessary to a) setting it into backup state [I-D.ietf-mptcp-rfc6824bis] [RFC6824] or b) inform the other end the link on-off state through newly extended MPTCP option. What's more, some extension to TCP might be needed to keep the corresponding paths valid.

5.5. Supplemental suggestions for scenario deployment

As it's hardly feasible for a user's device(e.g. smartphone, laptop, etc.) to directly connect to satellites or ATG network, a better choice is to connect to the Internet with Satellite Communication and ATG simultaneously with the air-borne antennas and run MPTCP at aircraft-level. We also suggest adopting PEP(TCP-split proxy)[RFC3135] at the aircraft-level, so that a user's device can directly connect to the aircraft (e.g. through on-board Wi-Fi) and then enjoy Internet services through the proxy. Another two considerations for adopting PEP are : a) it's a common way to adopt TCP-split proxy at the transition point between satellite and non-satellite links to improve TCP performance, especially for GEO satellite communication, b) a common user could connect to the aircraft with his/her unmodified device, which does not necessarily support MPTCP(which is largely true) and enjoy the benefits of MPTCP (and MPTCP-P suggested in this draft) through the proxy.

6. Acknowledgements

7. IANA Considerations

This memo includes no request to IANA.

8. Security Considerations

9. Informative References

[I-D.ietf-mptcp-rfc6824bis]
Ford, A., Raiciu, C., Handley, M., Bonaventure, O., and C. Paasch, "TCP Extensions for Multipath Operation with Multiple Addresses", June 2019, <https://datatracker.ietf.org/doc/html/draft-ietf-mptcp-rfc6824bis-18>.

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

[RFC3135] Border, J., Kojo, M., Griner, J., Montenegro, G., and Z. Shelby, "Performance Enhancing Proxies Intended to Mitigate Link-Related Degradations", RFC 3135, DOI 10.17487/RFC3135, June 2001, <https://www.rfc-editor.org/info/rfc3135>.

[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, <https://www.rfc-editor.org/info/rfc6182>.

[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, <https://www.rfc-editor.org/info/rfc6824>.

[RFC793] Information Sciences Institute,University of Southern California, "Transmission Control Protocol", September 1981, <https://datatracker.ietf.org/doc/rfc793/>.

Authors' Addresses

Qian Wu
Tsinghua University
Institute for Network Sciences and Cyberspace, Tsinghua University
Beijing
100084
China

Email: wuqian@cernet.edu.cn

Hewu Li
Tsinghua University
Institute for Network Sciences and Cyberspace, Tsinghua University
Beijing
100084
China

Email: lihewu@cernet.edu.cn

Qi Zhang
Tsinghua University
Institute for Network Sciences and Cyberspace, Tsinghua University
Beijing
100084
China

Email: qi-zhang19@mails.tsinghua.edu.cn

Jun Liu
Tsinghua University
Institute for Network Sciences and Cyberspace, Tsinghua University
Beijing
100084
China

Email: juneliu@tsinghua.edu.cn