MPTCP and Product Support Overview

Document ID: 116519

Contributed by Jay Young and Daniel Wing, Cisco TAC Engineers.
Sep 17, 2013

Contents

Introduction

This document provides an overview of Multipath TCP (MPTCP), its impact on flow inspection, and the Cisco products that are and are not affected by it.

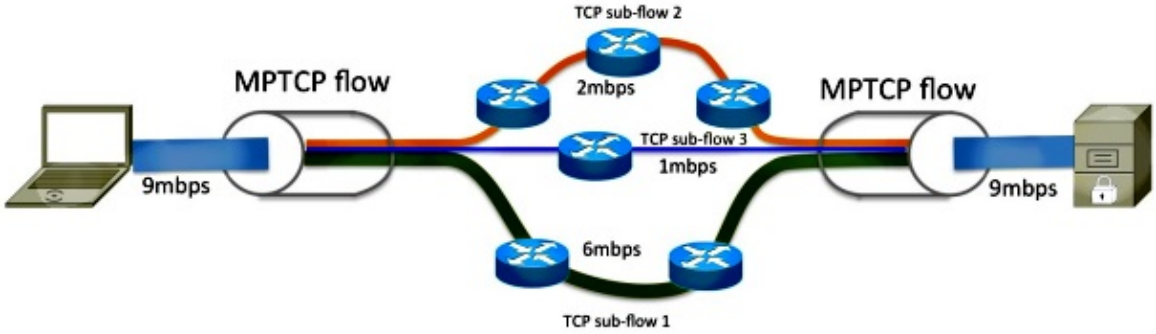
MPTCP Overview

Background Information

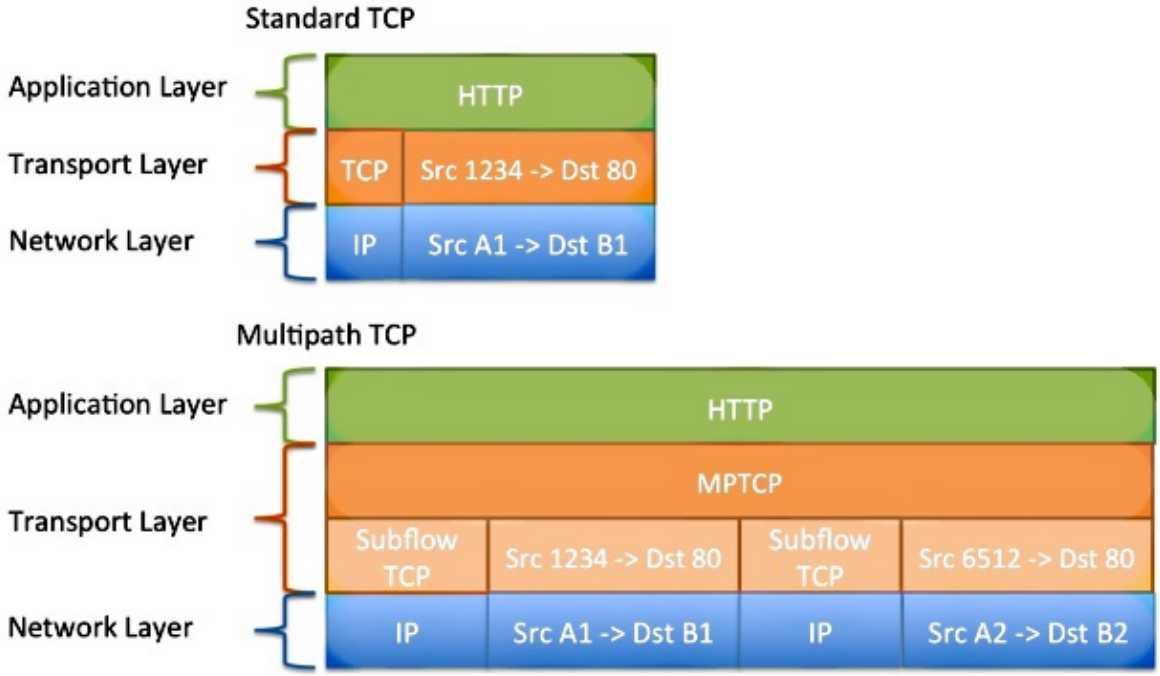
Hosts connected to the Internet or within a data center environment are often connected by multiple paths. However, when TCP is used for data transport, communication is restricted to a single network path. It is possible that some paths between the two hosts are congested, whereas alternate paths are underutilized. A more efficient use of network resources is possible if these multiple paths are used concurrently. In addition, the use of multiple connections enhances the user experience, because it provides higher throughput and improved resilience against network failures.

MPTCP is a set of extensions to regular TCP that enables a single data flow to be separated and carried across multiple connections. Refer to RFC6824: TCP Extensions for Multipath Operation with Multiple Addresses for more information.

As shown in this diagram, MPTCP is able to separate the 9mbps flow into three different sub−flows on the sender node, which is subsequently aggregated back into the original data flow on the receiving node.



The data that enters the MPTCP connection acts exactly as it does through a regular TCP connection; the transmitted data has guaranteed and in−order delivery. Since MPTCP adjusts the network stack and operates within the transport layer, it is used transparently by the application.

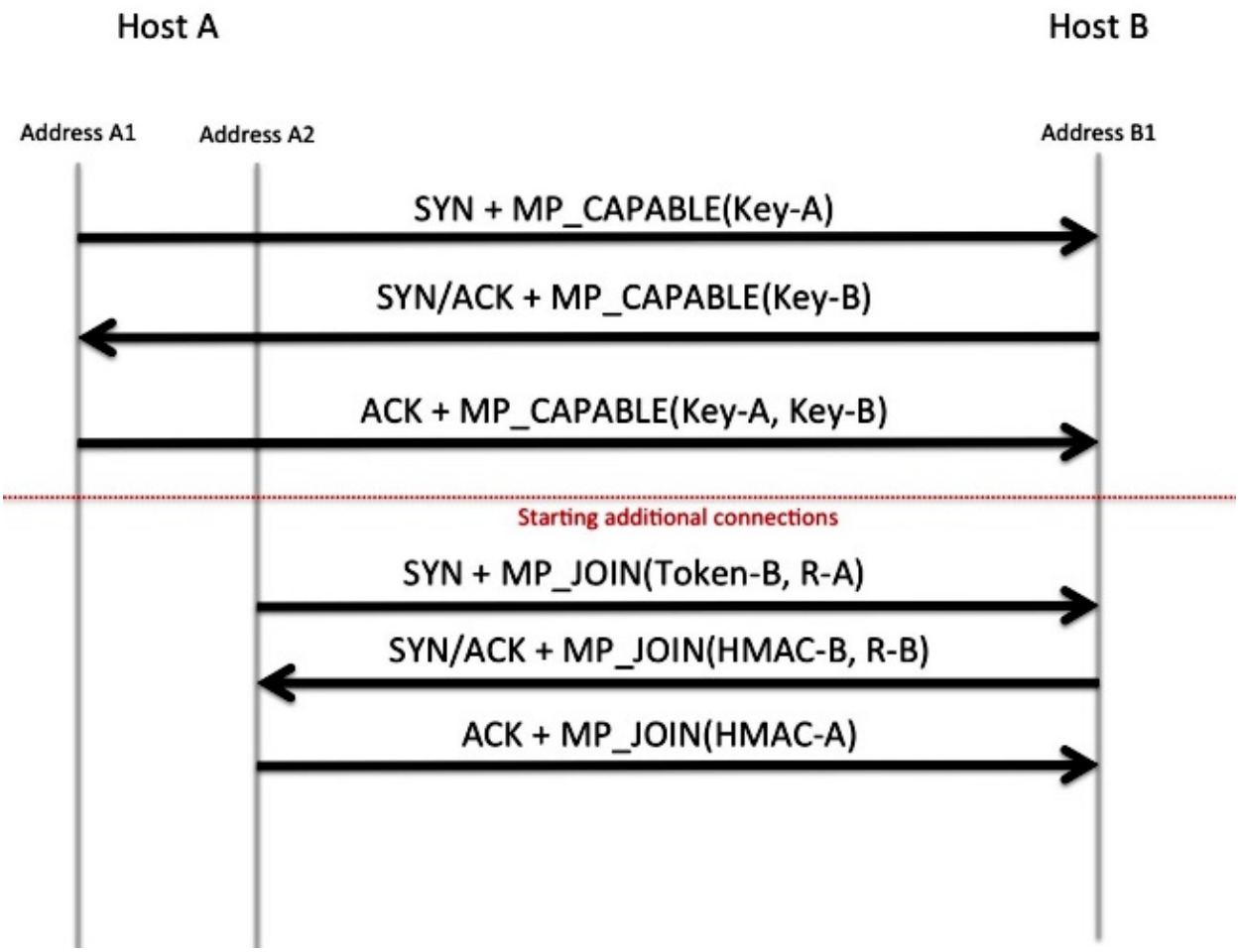


Session Establishment

MPTCP uses TCP options in order to negotiate and orchestrate the separation and reassembly of data over the multiple sub−flows. ***TCP option 30*** is reserved by the Internet Assigned Numbers Authority (IANA) for exclusive use by MPTCP. Refer to Transmission Control Protocol (TCP) Parameters for more information. In the establishment of a regular TCP session, a ***MP_CAPABLE*** option is included in the initial synchronize (SYN) packet. If the responder supports and chooses to negotiate MPTCP, it also responds with the ***MP_CAPABLE*** option in the SYN−acknowledge (ACK) packet. The keys exchanged within this handshake are used in the future in order to authenticate the joining and removal of other TCP sessions into this MPTCP flow.

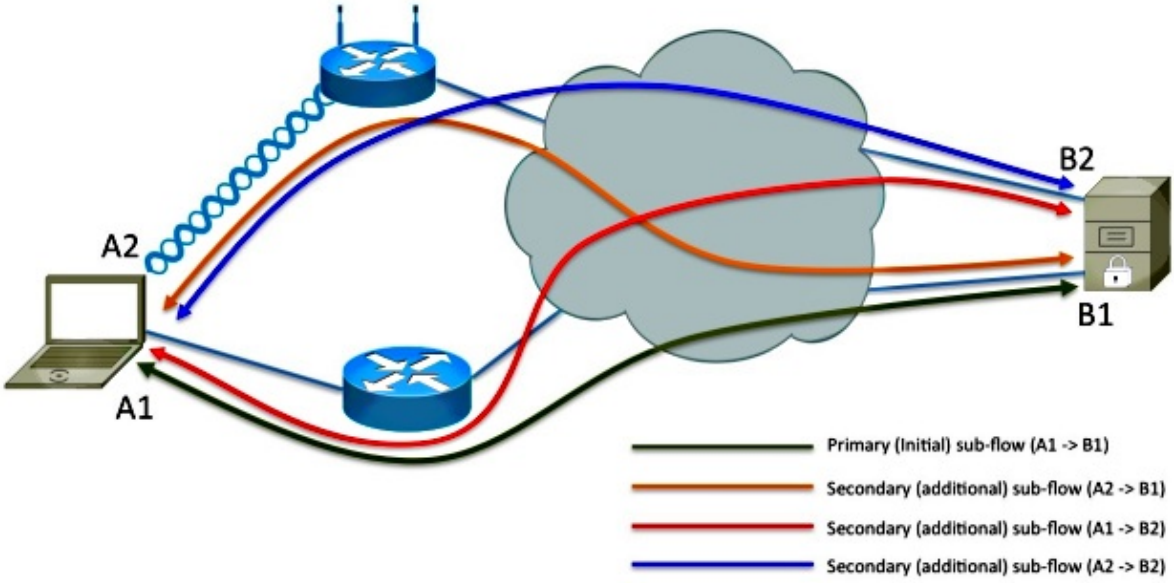
Join Additional Sub−flows

When deemed necessary, ***Host−A*** might initiate additional sub−flows sourced from a different interface or address to ***Host−B***. As with the initial sub−flow, TCP options are used in order to indicate the desire to merge this sub−flow with the other sub−flow. The keys that are exchanged within the initial sub−flow establishment (along with a hashing algorithm) are used by ***Host−B*** in order to confirm that the join request is indeed sent by ***Host−A***. The secondary sub−flow 4−tuple (source IP, destination IP, source Port, and destination Port) is different than that of the primary sub−flow; this flow might take a different path through the network.



Add Address

Host−A has multiple interfaces, and it is possible that ***Host−B*** has multiple network connections. ***Host−B*** learns about addresses A1 and A2 implicitly as a result of ***Host−A*** sourcing sub−flows from each of its addresses destined to B1. It is possible that ***Host−B*** advertises its additional address (B2) to ***Host−A*** so that other sub−flows are made to B2. This is completed via the ***TCP option 30***. As shown in this diagram, ***Host−B*** advertises its secondary address (B2) to ***Host−A***, and two additional sub−flows are created. Because MPTCP operates above the Network layer of the Open System Interconnection (OSI) stack, the IP addresses advertised can be IPv4, IPv6, or both. It is possible that some of the sub−flows are transported by IPv4 simultaneously as other sub−flows are transported by IPv6.



Segmentation, Multipath, and Reassembly

A data stream given to MPTCP by the application must be segmented and distributed across the multiple sub−flows by the sender. It then must be reassembled into the single data stream before it is delivered back to the application.

MPTCP inspects the performance and latency of each sub−flow, and dynamically adjusts the distribution of data in order to gain the highest aggregate throughput. During data transfer, the TCP header option includes information about the MPTCP sequence/acknowledgement numbers, the current sub−flow sequence/acknowledgement number, and a checksum.

Impact on Flow Inspection

Many security devices might zero−out or replace unknown TCP headers with a No Option (NOOP) value. If the network device does this to the TCP SYN packet on the initial sub−flow, the ***MP_CAPABLE*** advertisement is removed. As a result, it appears to the server that the client does not support MPTCP, and it reverts to normal TCP operation.

If the header is preserved and MPTCP is able to establish multiple sub−flows, in−line packet analysis by network devices might not function reliably. This is because only portions of the data flow are carried over to each sub−flow. The effect of protocol inspection upon MPTCP might vary from nothing to full disruption of service. The effect varies based on what and how much data is inspected. Packet analysis might include firewall Application Layer Gateway (ALG or fixup), Network Address Translation (NAT) ALG, Application Visibility and Control (AVC), or Network Based Application Recognition (NBAR). If application inspection is required in your environment, it is recommended that clearing of ***TCP option 30*** is enabled.

If the flow cannot be inspected due to encryption or if the protocol is unknown, then the inline device should have no impact on the MPTCP flow.

Cisco Products Impacted by MPTCP

These products are impacted by MPTCP:

- Adaptive Security Appliance (ASA)

- Cisco ASA Next−Generation Firewall Services
- Intrusion Prevention System (IPS)
- Cisco IOS®
- Application Control Engine (ACE)
- Cloud Web Security (ScanSafe)

Each product is described in detail in subsequent sections of this document.

ASA

TCP Operations

By default, the Cisco ASA firewall replaces unsupported TCP options, which include the ***MPTCP option 30***, with the NOOP option (option 1). In order to permit the MPTCP option, use this configuration:

1. Define the policy in order to allow ***TCP option 30*** (used by MPTCP) through the device:

```
tcp−map my−mptcp
   tcp−options range 30 30 allow
```

2. Define the traffic selection:

```
class−map my−tcpnorm
   match any
```

3. Define a map from traffic to action:

```
policy−map my−policy−map
   class my−tcpnorm
       set connection advanced−options my−mptcp
```

4. Activate it on the box or per−interface:

```
service−policy my−policy−map global
```

Protocol Inspection

The ASA supports inspection of many protocols. The effect that the inspection engine might have on the application varies. It is recommended that, if inspection is required, the TCP−map described previously is NOT applied.

Cisco ASA Next−Generation Firewall Services

TCP Operations

CX Versions 9.1.2.42 and later permit the MPTCP option, if the ASA is configured in order to allow it. CX software versions prior to Version 9.1.2.42 remove the MPTCP option.

Inline Secure Sockets Layer (SSL) Decryption

CX, when configured for SSL decryption, acts as a full−session proxy. As a result, it impersonates the SSL server when it communicates to the client, and also impersonates the client when it communicates to the SSL server. Since the CX terminates the two TCP connections, it does not use MPTCP, and operations revert to normal TCP operations.

IPS

The Cisco IPS product alerts on signature ***1306/0*** when the ***TCP option 30*** is seen in the TCP SYN. Because the signature belongs in the normalizer class, it does nothing in promisicious mode. It is possible to edit the signature in order to remove option 30 from the trigger conditions. This prevents the signature from matching.

Cisco IOS Firewall

Context−Based Access Control (CBAC)

CBAC does not remove the TCP headers from the TCP stream. MPTCP builds a connection through the firewall.

Zone−Based Firewall (ZBFW)

ZBF does not remove the TCP headers from the TCP stream. MPTCP builds a connection through the firewall.

ACE

By default, the ACE device strips TCP options from the TCP connections. The MPTCP connection falls back to regular TCP operations.

The ACE device might be configured in order to allow the TCP options via the ***tcp−options*** command, as described in the Configuring How the ACE Handles TCP Options section of the Security Guide vA5(1.0), Cisco ACE Application Control Engine. However, this is not always recommended, because the secondary sub−flows might be balanced to different real−servers, and the join fails.

Cloud Web Security (ScanSafe)

Cloud Web Security acts as a Layer−7 proxy. Whether you use the Web Security Appliance (WSA) or Cisco ScanSafe Towers, the client terminates its TCP connection at the proxy. Because the servers do not support the MPTCP options, the connection acts as a regular TCP connection.

The workaround to address the problem is to configure the ScanSafe connector in order to prevent a redirect of MPTCP flows to ScanSafe. When MPTCP flows are bypassed, the ScanSafe connector retrieves the content directly from the originally−requested web server without contacting ScanSafe. For more information, refer to the ***Bypassing Scanning on the IS*** section on page 12 of the Cisco ISR Web Security with Cisco ScanSafe Solution Guide.

Cisco Products not Impacted by MPTCP

Generally, any device that does not inspect TCP flows or Layer−7 information does not alter the TCP headers, and as a result should be transparent to MPTCP. These devices might include:

- Cisco 1000 Series Aggregation Services Routers (ASRs), ISR 4451−X, Cloud Services Router (CSR) (Cisco IOS−XE products)
- Cisco 5000 Series ASRs (Starent)
- Wide Area Application Services (WAAS)
- Carrier−Grade NAT (CGN) (Carrier−Grade Services Engine (CGSE) blade in Carrier Routing System (CRS)−1)
- All Ethernet switch products

- All router products (unless firewall or NAT functionality is enabled; see the affected products section earlier in the document for more details)

Updated: Sep 17, 2013 Document ID: 116519