

# IP路径MTU发现和DLSw

## 目录

[简介](#)

[开始使用前](#)

[规则](#)

[先决条件](#)

[使用的组件](#)

[背景信息](#)

[与PMTD的DLSw](#)

[DLSW的正在验证的PMTD](#)

[相关信息](#)

## 简介

IBM协议组，DLSw、STUN和BSTUN建立从一个路由器的一条IP会话管道到另一个。TCP是常用的作为在路由器之间的传输方法由于其信度。本文在TCP的能力提供信息动态地发现在会话管道能使用，最小化分段并且最大化效率的最大的MTU。

## 开始使用前

### 规则

有关文档规则的详细信息，请参阅 [Cisco 技术提示规则](#)。

### 先决条件

本文档没有任何特定的前提条件。

### 使用的组件

本文档不限于特定的软件和硬件版本。

本文档中的信息都是基于特定实验室环境中的设备创建的。本文档中使用的所有设备最初均采用原始（默认）配置。如果您是在真实网络上操作，请确保您在使用任何命令前已经了解其潜在影响。

## 背景信息

路径MTU发现(PMTD)，正如RFC 1191所描述，指定IP数据包的默认字节大小是576。帧的IP和TCP部分占用留下536个字节的40个字节作为数据负载。此空间叫作最大分段尺寸或MSS。RFC1191的部分3.1指定一个更加大的MSS能协商，并且这正确地是什么发出`ip tcp path-mtu-`

**discovery**命令在Cisco路由器执行。当此命令配置时，并且TCP会话开始，在路由器外面的SYN数据包包含指定一个更加大的MSS的TCP选项。此更加大的MSS是出站接口的MTU减40个字节的。如果出站接口的MTU是1500个字节，通告的MSS是1460。如果出站接口有更加大的MTU，例如，有4096字节的MTU帧中继，MSS将是减40字节的4096个字节的IP信息和显示在**show tcp**命令输出中(最大数据段是4056个字节)。

配置在路由器的PMTD在存在没有任何效果TCP会话已经建立从/至路由器。PMTD介绍到11.3.5T IOS级别，并且在IOS后续版本，变为可选命令。默认情况下在IOS 11.3(5)T之前，它打开。另外，PMTD自动，当IP地址在相同子网时，指示他们直接地附加在同一媒体。

两路由器必须配置为了PMTD能适当地工作。当两路由器配置时，从一个路由器的SYN到其他包含通告更加高的MSS的可选TCP值。返回的SYN然后通告更高的MSS值。因此，他们能接受一个更加大的MSS的两边通告对其他。如果仅一个路由器，路由器1，有**ip tcp path-mtu-discovery**命令已配置的，将通告更加大的MSS，并且因而，Router2能发送到路由器1 1460字节帧。Router2不会通告更加大的MSS，并且路由器1因而锁定到发送默认值。

## 与PMTD的DLSw

在IBM套件，DLSw、STUN和BSTUN可以被分派任务传送在—TCP会话的很多数据从路由器到路由器。实现PMTD，特别是考虑可以是重要和非常有利的默认情况下它启用11.2和前期IOS级别。默认情况下根据RFC，最大的帧是576个字节，减IP/TCP封装的40个字节。DLSw使用另外16个字节封装。传输，使用默认MSS的实际数据，是520个字节。DLSw也有功能传送两不同的逻辑链路控制2 (LLC2)数据包到—TCP帧。如果两个工作站每发送LLC2帧，DLSw能运载两LLC2帧对一帧的DLSw远端对等体。由有一个更加大的MSS，TCP驱动程序能适应此piggybacking模式。下列是说明值的三个主要方案**path-mtu-discovery**命令。

### 方案1 -不需要的开销

SDLC设备为265或521字节的数据maxdata在每帧的通常将配置。如果值是521和3174发送到路由器1—521个字节SDLC帧，路由器1将做两TCP帧传输此到DLSW对等`路由器2。第一帧将包含520字节的数据、16字节的DLSw信息和40字节的总共576个字节的IP信息。第二数据包将包含1字节的数据、16字节的DLSw信息和40字节的IP信息。当PMTD是过去常常和假设1500字节MTU获得一个1460个MSS时，路由器1由Router2通知Router2能接收1460字节的数据。路由器1将发送全部521字节的SDLC数据对在—数据包的Router2有16字节的DLSw信息和40字节的IP信息。通过使用PMTD，因为DLSw是进程交换的事件，处理的CPU利用率此一个SDLC帧被对分了。另外，Router2不必等待第二数据包装配LLC2帧。使用，启用的PMTD，Router2接收整个数据包，并且能从数据包然后取消IP和DLSW信息和赶快发送它到3745。

### 方案2 -从无序信息包的迪莱

在此方案中，有两IP网云可用与相等的量度为负载平衡或冗余。不启用PMTD能严重地减慢DLSw。没有启用的PMTD，路由器1必须装配521-byte帧到与520字节的数据和第三的两TCP数据包与1字节的数据。如果第一数据包横断顶部IP网云，在第二数据包以后将到达的有重大的可能性，如果第二数据包通过更低，相等代价IP网云发送。这生成什么是公认的无序信息包。内在的在IP/TCP协议的功能是能力管理此问题。无序信息包在等待整个数据流的内存存储到达然后被重新召集。无序信息包不是不常见，然而，应该做所有尝试最小化他们作为此事件利用内存和CPU资源。很多无序能导致明显的延迟在TCP级别。如果layer3/DLSw会话延迟，则LLC2/SDLC会话转入此DLSw随后将遭受。如果PMTD在此方案启用，单个521-byte帧在任一IP网云的一TCP帧发送。接收路由器仅需要缓冲区和解除封装—TCP帧。

PMTD没有对最大的帧通告的终端站的关系到终端站在SNA环境。这在路由信息字段(RIF)包括Largest Frame (LF) token-ring。PMTD严格指明可以被封装到—TCP帧的相当数量数据。LLC2和

SDLC没有功能分段的信息包，然而，IP/TCP。一大SNA帧可以被分段作为它被封装到TCP。此数据被解封装在远程DLSw路由器和再被提交作为不可成片断的SNA数据。

### 方案3 -更加快速的LLC2连接和吞吐量

在此方案中，3174和 workstation 有会话通过3745个TIC到大型机，如果两个设备发送为主机注定的数据，它是可能的TCP能放两LLC2帧到一数据包。没有PMTD，如果从两帧的组合数据是521字节或更加极大的，这不是可能的。在这种情况下，TCP软件将需要分开发送每数据包。例如，如果3174和 workstation 发送这些数据包帧在大约同一时间和中的每一有400字节的数据，路由器接收并且缓冲每帧。路由器必须当前封装这400个字节数据流中的每一个到发射的独立的TCP信息包对对等体。

当PMTD启用和假设MSS 1460，路由器收到并且缓冲两LLC2数据包。当前能封装两个到一数据包。此一TCP数据包将包含40字节的IP信息，16字节的配对第一个DLSw的电路的DLSw信息，400字节的数据、另外16字节的数据配对第二个DLSw的电路的和其他400字节的数据。此特定的方案使用两个设备和因而，两个DLSw电路。PMTD允许DLSw效率更高扩展到DLSw电路大数字。许多集线器网络要求数百远程站点，中的每一用一两个SNA设备，并列到连接的中心站点路由器与提供存取对于主机应用程序的OSA或FEP。PMTD给TCP和DLSw能力扩展到更加大的需求无在利用路由器CPU和内存资源以及提供快速传输。

**注意：**有在12.2(5)T找到在延迟12.1(5)T和解决的软件Bug PMTD哪里没有在虚拟专用网络(VPN)通道工作。此软件缺陷是[CSCdt49552](#) (仅限注册用户)。

## DLSW的正在验证的PMTD

发出show tcp命令。

```
havoc#show tcp Stand-alone TCP connection to host 10.1.1.1 Connection state is ESTAB, I/O
status: 1, unread input bytes: 0 Local host: 30.1.1.1, Local port: 11044 Foreign host: 10.1.1.1,
Foreign port: 2065 Enqueued packets for retransmit: 0, input: 0 mis-ordered: 0 (0 bytes) TCP
driver queue size 0, flow controlled FALSE Event Timers (current time is 0xA18A78): Timer Starts
Wakeup Next Retrans 3 0 0x0 TimeWait 0 0 0x0 AckHold 0 0 0x0 SendWnd 0 0 0x0 KeepAlive 0 0 0x0
GiveUp 2 0 0x0 PmtuAger 0 0 0x0 DeadWait 0 0 0x0 iss: 3215333571 snduna: 3215334045 sndnxt:
3215334045 sndwnd: 20007 irs: 3541505479 rcvnx: 3541505480 rcvwnd: 20480 delrcvwnd: 0 SRTT: 99
ms, RTTO: 1539 ms, RTV: 1440 ms, KRTT: 0 ms minRTT: 24 ms, maxRTT: 300 ms, ACK hold: 200 ms
Flags: higher precedence, retransmission timeout Datagrams (max data segment is 536 bytes):
Rcvd: 30 (out of order: 0), with data: 0, total data bytes: 0 Sent: 4 (retransmit: 0,
fastretransmit: 0), with data: 2, total data bytes: 473
```

因为其中一个端口在TCP会话上是2065，此显示识别作为DLSw TCP会话。在输出的底部附近是最大数据段是536个字节。此值表明10.1.1.1的远程DLSw对等`路由器没有ip tcp path-mtu-discovery命令已配置的。536字节值已经占40字节的在IP帧的IP信息。此536字节值不占将被添加到运载SNA流量的TCP数据包的16字节的DLSw信息。

使用ip tcp path-mtu-discovery命令已配置的，最大数据段当前是1460。另外，show tcp命令输出指示路径MTU有能力在最大数据段语句前。出站接口有1500个字节MTU。MTU等于减40字节的1500字节的IP信息是1460个字节。DLSw将使用另外16个字节。所以，1444 LLC2或SDLC字节帧在一TCP帧可以传送。

```
havoc#show tcp Stand-alone TCP connection to host 10.1.1.1 Connection state is ESTAB, I/O
status: 1, unread input bytes: 0 Local host: 30.1.1.1, Local port: 11045 Foreign host: 10.1.1.1,
Foreign port: 2065 Enqueued packets for retransmit: 0, input: 0 mis-ordered: 0 (0 bytes) TCP
driver queue size 0, flow controlled FALSE Event Timers (current time is 0xA6DA58): Timer Starts
Wakeup Next Retrans 4 0 0x0 TimeWait 0 0 0x0 AckHold 1 0 0x0 SendWnd 0 0 0x0 KeepAlive 0 0 0x0
GiveUp 3 0 0x0 PmtuAger 0 0 0x0 DeadWait 0 0 0x0 iss: 3423657490 snduna: 3423657976 sndnxt:
```

3423657976 sndwnd: 19995 irs: 649085675 rcvnext: 649085688 rcvwnd: 20468 delrcvwnd: 12 SRTT: 124 ms, RTTO: 1405 ms, RTV: 1281 ms, KRTT: 0 ms minRTT: 24 ms, maxRTT: 300 ms, ACK hold: 200 ms  
Flags: higher precedence, retransmission timeout, path mtu capable Datagrams (max data segment is 1460 bytes): Rcvd: 5 (out of order: 0), with data: 1, total data bytes: 12 Sent: 6 (retransmit: 0, fastretransmit: 0), with data: 3, total data bytes: 485

## [相关信息](#)

- [Compatible Systems技术说明 : VPN的IP分段和MTU路径发现](#)
- [技术支持 - Cisco Systems](#)