

IP Path MTU Discovery und DLSw

Inhalt

[Einführung](#)

[Bevor Sie beginnen](#)

[Konventionen](#)

[Voraussetzungen](#)

[Verwendete Komponenten](#)

[Hintergrundinformationen](#)

[DLSw mit PMTD](#)

[Verifizieren von PMTD für DLSW](#)

[Zugehörige Informationen](#)

Einführung

Die IBM-Protokoll-Suite, DLSw, STUN und BSTUN erstellen ein IP-Sitzungspipp von einem Router zum anderen. TCP wird aufgrund seiner Zuverlässigkeit häufig als Transportmethode zwischen Routern verwendet. Dieses Dokument enthält Informationen zur Fähigkeit von TCP, die größte MTU, die auf der Sitzungspope verwendet werden kann, dynamisch zu ermitteln. Dadurch wird die Fragmentierung minimiert und die Effizienz maximiert.

Bevor Sie beginnen

Konventionen

Weitere Informationen zu Dokumentkonventionen finden Sie in den [Cisco Technical Tips Conventions](#).

Voraussetzungen

Für dieses Dokument bestehen keine besonderen Voraussetzungen.

Verwendete Komponenten

Dieses Dokument ist nicht auf bestimmte Software- und Hardwareversionen beschränkt.

Die in diesem Dokument enthaltenen Informationen wurden aus Geräten in einer bestimmten Laborumgebung erstellt. Alle in diesem Dokument verwendeten Geräte haben mit einer leeren (Standard-)Konfiguration begonnen. Wenn Sie in einem Live-Netzwerk arbeiten, stellen Sie sicher, dass Sie die potenziellen Auswirkungen eines Befehls verstehen, bevor Sie es verwenden.

Hintergrundinformationen

Die Path MTU Discovery (PMTD), wie in RFC 1191 beschrieben, legt fest, dass die standardmäßige Bytegröße eines IP-Pakets 576 beträgt. Die IP- und TCP-Bereiche des Frames belegen 40 Byte, wobei 536 Byte als Datennutzlast verbleiben. Dieses Leerzeichen wird als maximale Segmentgröße oder MSS bezeichnet. In Abschnitt 3.1 von RFC1191 wird festgelegt, dass ein größeres MSS ausgehandelt werden kann. Dies ist genau der Punkt, an dem der Befehl **ip tcp path-mtu-discovery** in einem Cisco Router ausgeführt wird. Wenn dieser Befehl konfiguriert und eine TCP-Sitzung gestartet wird, enthält das SYN-Paket außerhalb des Routers eine TCP-Option, die eine größere MSS angibt. Diese größere MSS ist die MTU der ausgehenden Schnittstelle minus 40 Byte. Wenn die MTU der ausgehenden Schnittstelle 1.500 Byte beträgt, wird eine MSS von 1.460 angekündigt. Wenn die ausgehende Schnittstelle eine größere MTU aufweist, z. B. Frame Relay mit einer MTU von 4096 Byte, beträgt die MSS 4096 Byte abzüglich 40 Byte IP-Informationen und wird in der Ausgabe des **Befehls show tcp** angezeigt (max. Datensegment beträgt 4056 Byte).

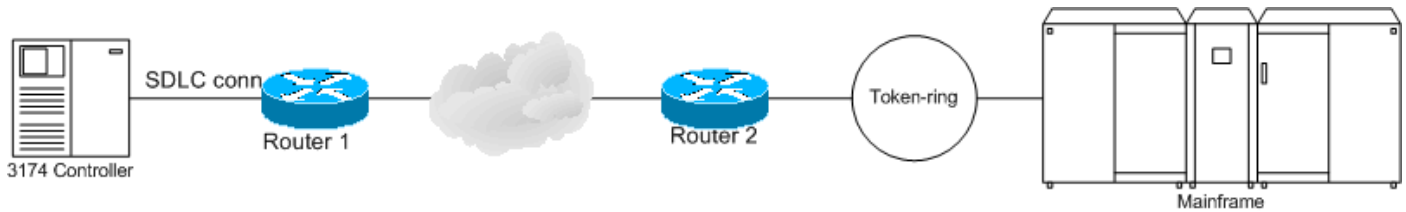
Die Konfiguration von PMTD auf dem Router hat keine Auswirkungen auf vorhandene TCP-Sitzungen, die bereits vom/zum Router eingerichtet wurden. PMTD wurde auf der IOS-Ebene 11.3.5T eingeführt, und in späteren Versionen von IOS wurde es zu einem optionalen Befehl. Vor IOS 11.3(5)T war es standardmäßig aktiviert. Darüber hinaus wird die PMTD automatisch ausgeführt, wenn sich die IP-Adressen im gleichen Subnetz befinden. Dies bedeutet, dass sie direkt auf dem gleichen Medium angeschlossen sind.

Beide Router müssen konfiguriert werden, damit PMTD ordnungsgemäß funktioniert. Wenn beide Router konfiguriert sind, enthält das SYN von einem Router zum anderen den optionalen TCP-Wert, der die höhere MSS angibt. Das zurückgebende SYN gibt dann den höheren MSS-Wert bekannt. So kündigen beide Seiten an die anderen, können sie eine größere MSS akzeptieren. Wenn der Befehl **ip tcp path-mtu-discovery** nur für Router 1 konfiguriert ist, wird der größere MSS angekündigt, und Router 2 kann einen 1460-Byte-Frame an Router 1 senden. Router 2 kündigt die größere MSS nie an, und Router 1 ist daher an das Senden der Standardwerte gebunden.

DLSw mit PMTD

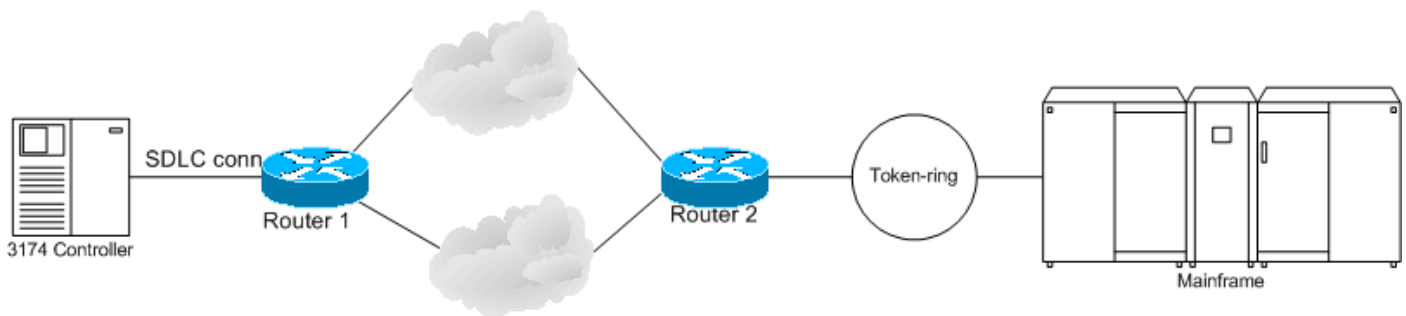
In der IBM-Suite können DLSw, STUN und BSTUN beauftragt werden, große Datenmengen über eine TCP-Sitzung von einem Router zum anderen zu übertragen. Die Implementierung von PMTD kann wichtig und äußerst vorteilhaft sein. Dies gilt insbesondere für die Tatsache, dass die PMTD in den 11.2- und früheren IOS-Levels standardmäßig aktiviert wurde. Laut RFC beträgt der größte Frame standardmäßig 576 Byte, abzüglich 40 Byte für die IP/TCP-Kapselung. DLSw verwendet weitere 16 Byte für die Kapselung. Die tatsächlichen Daten, die mithilfe der Standard-MSS übertragen werden, sind 520 Byte. DLSw kann außerdem zwei verschiedene LLC2-Pakete (Logical Link Control 2) in einem TCP-Frame übertragen. Wenn zwei Workstations jeweils einen LLC2-Frame senden, kann DLSw beide LLC2-Frames in einem Frame zum DLSw-Remote-Peer übertragen. Durch eine größere MSS können die TCP-Treiber dieses piggybackschema aufnehmen. Die folgenden drei Hauptszenarien veranschaulichen den Wert des Befehls **path-mtu-discovery**.

Szenario 1 - Unerwünschter Overhead



SDLC-Geräte werden in der Regel für maximal 265 oder 521 Byte Daten in jedem Frame konfiguriert. Wenn der Wert 521 ist und der 3174 einen 521-Byte-SDLC-Frame an Router 1 sendet, erstellt Router 1 zwei TCP-Frames, um diesen an den DLSw-Peer-Router 2 zu übertragen. Der erste Frame enthält 520 Byte Daten, 16 Byte DLSw-Informationen und 40 Byte IP-Informationen für insgesamt 576 Byte. Das zweite Paket enthält 1 Byte Daten, 16 Byte DLSw-Informationen und 40 Byte IP-Informationen. Wenn PMTD verwendet wird und eine MTU von 1.500 Byte für die Bereitstellung einer MSS von 1.460 angenommen wird, wurde Router 1 von Router 2 darauf hingewiesen, dass Router 2 1.460 Byte Daten empfangen kann. Router 1 sendet alle 521 Byte SDLC-Daten in einem Paket mit 16 Byte DLSw-Informationen und 40 Byte IP-Informationen an Router 2. Da DLSw ein prozessgesteuertes Ereignis ist, wurde mithilfe von PMTD die CPU-Auslastung zur Verarbeitung dieses SDLC-Frames halbiert. Darüber hinaus muss Router 2 nicht warten, bis das zweite Paket den LLC2-Frame zusammenfügt. Wenn PMTD aktiviert ist, empfängt Router 2 das gesamte Paket und kann anschließend die IP- und DLSw-Informationen aus dem Paket entfernen und unverzüglich an den 3745 senden.

Szenario 2 - Verzögerung bei nicht bestellten Paketen

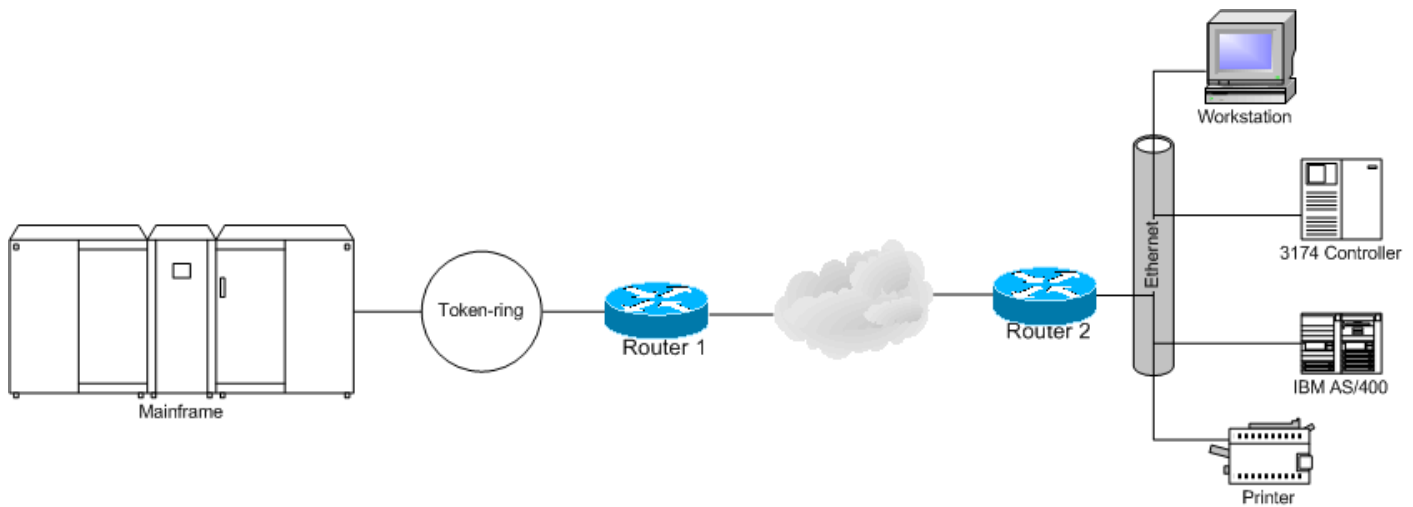


In diesem Szenario stehen zwei IP-Clouds mit gleichen Kennzahlen für Lastenausgleich oder Redundanz zur Verfügung. Wenn PMTD nicht aktiviert wird, kann DLSw stark verlangsamt werden. Ohne PMTD muss Router 1 den Frame mit 521 Byte in zwei TCP-Pakete zusammenfassen - eines mit 520 Byte Daten und das zweite mit 1 Byte Daten. Wenn das erste Paket die obere IP-Cloud durchläuft, besteht eine erhebliche Wahrscheinlichkeit, dass es nach dem zweiten Paket eingeht, wenn das zweite Paket über die IP-Cloud mit niedrigeren Kosten gesendet wird. Dadurch wird ein so genanntes Out-of-Order-Paket generiert. Die Funktion des IP/TCP-Protokolls beinhaltet die Fähigkeit, dieses Problem zu verwalten. Out-of-Order-Pakete werden im Arbeitsspeicher gespeichert, bis der gesamte Stream eingeht und dann wieder assembliert wird. Out-of-Order-Pakete sind nicht ungewöhnlich. Es sollten jedoch alle Versuche unternommen werden, diese zu minimieren, da dieses Ereignis Arbeitsspeicher- und CPU-Ressourcen nutzt. Eine große Anzahl von Out-of-Order-Bestellungen kann auf TCP-Ebene zu erheblichen Verzögerungen führen. Wenn sich die Layer3/DLSw-Sitzung verzögert, leidet die LLC2/SDLC-Sitzung, die über diesen DLSw übertragen wird. Wenn in diesem Szenario PMTD aktiviert ist, wird ein einzelner Frame mit 521 Byte in einem TCP-Frame über eine der beiden IP-Clouds gesendet. Der empfangende Router muss nur einen TCP-Frame puffern und entkapseln.

PMTD steht in SNA-Umgebungen nicht in Zusammenhang mit der größten per Frame gemeldeten Endstation zur Endstation. Dazu gehört der Größte Frame (LF) im Feld "Routing Information" (RIF) auf Token-Ring. PMTD legt die Datenmenge fest, die in einen TCP-Frame gekapselt werden kann. LLC2 und SDLC verfügen nicht über die Fragment-Pakete für die Funktionen, IP/TCP

jedoch nicht. Ein großer SNA-Frame kann segmentiert werden, da er in TCP gekapselt ist. Diese Daten werden am Remote-DLSw-Router entkapselt und erneut als nicht fragmentierte SNA-Daten angezeigt.

Szenario 3: Schnellere LLC2-Verbindungen und -Durchsatz



In diesem Szenario führen der 3174 und die Workstation Sitzungen durch das 3745 TIC zum Mainframe. Wenn beide Geräte Daten an den Host senden, ist es möglich, dass TCP beide LLC2-Frames in einem Paket ablegen kann. Ohne PMTD ist dies nicht möglich, wenn die kombinierten Daten der beiden Frames 521 Byte oder mehr betragen. In diesem Fall muss die TCP-Software jedes Paket separat senden. Wenn beispielsweise der 3174 und die Workstation einen Frame ungefähr zur gleichen Zeit senden und jedes dieser Pakete 400 Byte an Daten enthält, empfängt und puffert der Router jeden Frame. Der Router muss nun jeden dieser 400-Byte-Datenströme in separate TCP-Pakete einkapseln, um sie zum Peer zu übertragen.

Wenn PMTD aktiviert ist und eine MSS von 1460 angenommen wird, empfängt und puffert der Router die beiden LLC2-Pakete. Sie kann nun beide Pakete in einem Paket einkapseln. Dieses ein TCP-Paket enthält 40 Byte IP-Informationen, 16 Byte DLSw-Informationen für das erste DLSw-Schaltkreis-Pairing, 400 Byte Daten, weitere 16 Byte Daten für das zweite DLSw-Schaltungs-Pairing und die anderen 400 Byte Daten. In diesem speziellen Szenario werden zwei Geräte und somit zwei DLSw-Schaltkreise verwendet. Mit PMTD kann DLSw effizienter auf eine höhere Anzahl von DLSw-Schaltungen skaliert werden. Viele Spoke-Hub-Netzwerke erfordern Hunderte von Remote-Standorten, von denen jedes über ein oder zwei SNA-Geräte verfügt. Dabei erfolgt der Peering über einen Router am zentralen Standort, der mit einem OSA oder FEP verbunden ist und Zugriff auf die Host-Anwendungen bietet. Mit PMTD können TCP und DLSw auf größere Anforderungen skaliert werden, ohne dass CPU- und Speicherressourcen des Routers überlastet werden müssen. Darüber hinaus wird ein schnellerer Transport ermöglicht.

Hinweis: Ende 12.1(5)T wurde ein Softwarefehler gefunden, der in 12.2(5)T behoben wurde, wenn PMTD nicht über einen VPN-Tunnel (Virtual Private Network) arbeitete. Dieser Softwarefehler ist [CSCdt49552](#) (nur [registrierte](#) Kunden).

[Verifizieren von PMTD für DLSW](#)

Geben Sie den Befehl `show tcp` ein.

```
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 | Wakeups | 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  rcvnxt: 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
```

Diese Anzeige wird als DLSw-TCP-Sitzung identifiziert, da einer der Ports in der TCP-Sitzung 2065 ist. Unten in der Ausgabe befindet sich ein max. Datensegment von 536 Byte. Dieser Wert gibt an, dass der Remote-DLSw-Peer-Router von 10.1.1.1 nicht über den Befehl **ip tcp path-mtu-discovery** verfügt. Der 536-Byte-Wert berücksichtigt bereits die 40-Byte-IP-Informationen in einem IP-Frame. Dieser 536-Byte-Wert berücksichtigt nicht die 16-Byte-DLSw-Informationen, die einem TCP-Paket mit SNA-Datenverkehr hinzugefügt werden.

Wenn der Befehl **ip tcp path-mtu-discovery** konfiguriert ist, beträgt das maximale Datensegment jetzt 1460. Zusätzlich gibt die Ausgabe des Befehls **show tcp** den **path mtu-fähigen Befehl** unmittelbar vor der **max data segment**-Anweisung an. Die MTU der ausgehenden Schnittstelle beträgt 1.500 Byte. MTU entspricht 1.500 Byte abzüglich 40 Byte IP-Informationen 1.460 Byte. DLSw verwendet weitere 16 Byte. Daher können bis zu 1444 Byte Frame von LLC2 oder SDLC in einem TCP-Frame übertragen werden.

```
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 | Wakeups | 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 rcvnxt: 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

[Zugehörige Informationen](#)

- [Kompatible Systems - Technische Hinweise: IP-Fragmentierung und MTU-Pfaderkennung mit VPN](#)
- [Technischer Support - Cisco Systems](#)