

Descoberta de MTU de caminho IP e DLSw

Contents

[Introduction](#)

[Antes de Começar](#)

[Conventions](#)

[Prerequisites](#)

[Componentes Utilizados](#)

[Informações de Apoio](#)

[DLSw com PMTD](#)

[Verificação do PMTD para DLSW](#)

[Informações Relacionadas](#)

[Introduction](#)

O conjunto de protocolos IBM, DLSw, STUN e BSTUN estabelecem um pipe de sessão IP de um roteador para outro. O TCP é comumente usado como o método de transporte entre roteadores devido à sua confiabilidade. Este documento fornece informações sobre a capacidade do TCP de descobrir dinamicamente o maior MTU que pode ser usado no pipe da sessão, o que minimiza a fragmentação e maximiza a eficiência.

[Antes de Começar](#)

[Conventions](#)

Para obter mais informações sobre convenções de documento, consulte as [Convenções de dicas técnicas Cisco](#).

[Prerequisites](#)

Não existem requisitos específicos para este documento.

[Componentes Utilizados](#)

Este documento não se restringe a versões de software e hardware específicas.

As informações neste documento foram criadas a partir de dispositivos em um ambiente de laboratório específico. All of the devices used in this document started with a cleared (default) configuration. Se você estiver trabalhando em uma rede ativa, certifique-se de que entende o impacto potencial de qualquer comando antes de utilizá-lo.

Informações de Apoio

Path MTU Discovery (PMTD), conforme descrito no RFC 1191, especifica que o tamanho padrão de byte de um pacote IP é 576. As partes IP e TCP do quadro ocupam 40 bytes, deixando 536 bytes como payload de dados. Esse espaço é conhecido como o tamanho máximo do segmento ou MSS. A seção 3.1 do RFC1191 especifica que um MSS maior pode ser negociado, e é exatamente isso que a emissão do comando **ip tcp path-mtu-discovery** faz em um roteador Cisco. Quando esse comando é configurado e uma sessão TCP é iniciada, o pacote SYN fora do roteador contém uma opção TCP especificando um MSS maior. Esse MSS maior é o MTU da interface de saída menos 40 bytes. Se o MTU da interface de saída for de 1.500 bytes, o MSS anunciado será de 1.460. Se a interface de saída tiver um MTU maior, por exemplo, Frame Relay com um MTU de 4.096 bytes, o MSS será de 4.096 bytes menos 40 bytes de informações de IP e será exibido na saída do comando **show tcp** (o segmento de dados máximo é de 4.056 bytes).

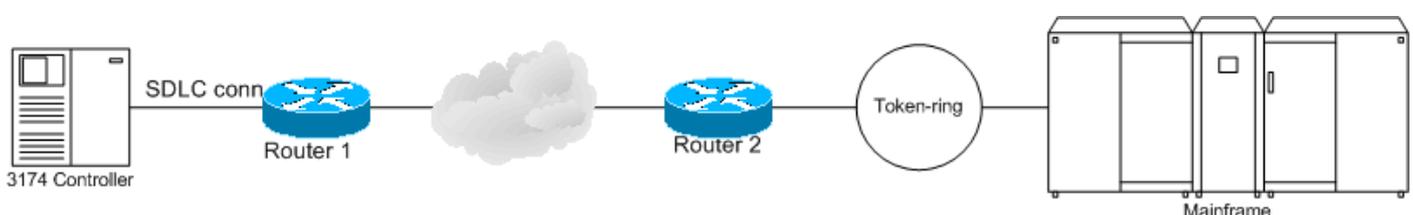
A configuração do PMTD no roteador não tem nenhum efeito nas sessões TCP existentes já estabelecidas a partir do/para o roteador. O PMTD foi introduzido no nível IOS 11.3.5T e, em versões subsequentes do IOS, tornou-se um comando opcional. Antes do IOS 11.3(5)T, ele estava ativado por padrão. Além disso, o PMTD é automático quando os endereços IP estão na mesma sub-rede, indicando que estão diretamente conectados na mesma mídia.

Ambos os roteadores devem ser configurados para que o PMTD funcione corretamente. Quando ambos os roteadores são configurados, o SYN de um roteador para o outro contém o valor TCP opcional anunciando o MSS mais alto. Em seguida, o SYN retornado anuncia o valor MSS mais alto. Assim, ambos os lados anunciam ao outro que podem aceitar um MSS maior. Se apenas um roteador, o Roteador 1, tiver o comando **ip tcp path-mtu-discovery** configurado, ele anunciará o MSS maior e, portanto, o Roteador 2 pode enviar ao Roteador 1 um quadro de 1460 bytes. O Roteador 2 nunca anunciará o MSS maior e, portanto, o Roteador 1 é bloqueado para enviar os valores padrão.

DLSw com PMTD

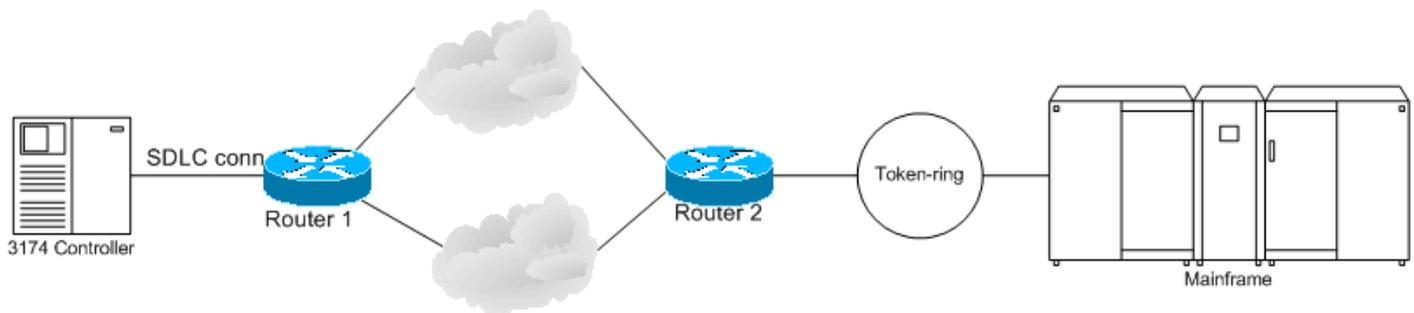
No conjunto IBM, DLSw, STUN e BSTUN podem ser incumbidos de transportar grandes quantidades de dados em uma sessão TCP de roteador para roteador. Pode ser importante e extremamente benéfico implementar o PMTD, especialmente considerando que ele foi ativado por padrão em níveis 11.2 e anteriores do IOS. Conforme o RFC, o maior quadro é de 576 bytes por padrão, menos 40 bytes para encapsulamento IP/TCP. O DLSw usa outros 16 bytes para encapsulamento. Os dados reais que estão sendo transportados, usando o MSS padrão, são 520 bytes. O DLSw também tem a capacidade de transportar dois pacotes diferentes do Logical Link Control 2 (LLC2) em um quadro TCP. Se duas estações de trabalho enviarem um quadro LLC2 cada, o DLSw poderá transportar ambos os quadros LLC2 para o peer remoto DLSw em um quadro. Com um MSS maior, os drivers TCP podem acomodar esse esquema de piggybacking. A seguir estão três cenários principais para ilustrar o valor do comando **path-mtu-discovery**.

Cenário 1 - Sobrecarga indesejada



Os dispositivos SDLC geralmente serão configurados para um máximo de dados de 265 ou 521 bytes de dados em cada quadro. Se o valor for 521 e o 3174 enviar ao Roteador 1 um quadro SDLC de 521 bytes, o Roteador 1 fará dois quadros TCP para transportá-lo para o Roteador 2 peer DLSw. O primeiro quadro conterá 520 bytes de dados, 16 bytes de informações de DLSw e 40 bytes de informações de IP para um total de 576 bytes. O segundo pacote conterá 1 byte de dados, 16 bytes de informações de DLSw e 40 bytes de informações de IP. Quando o PMTD é usado e presume-se que um MTU de 1.500 bytes obtenha um MSS de 1.460, o Roteador 1 foi informado pelo Roteador 2 que o Roteador 2 pode receber 1.460 bytes de dados. O roteador 1 enviará todos os 521 bytes de dados SDLC ao roteador 2 em um pacote com 16 bytes de informações DLSw e 40 bytes de informações de IP. Como o DLSw é um evento de processo comutado, usando o PMTD, a utilização da CPU para processar este único quadro SDLC foi reduzida para metade. Além disso, o Roteador 2 não precisa esperar que o segundo pacote monte o quadro LLC2. Com o PMTD ativado, o Roteador 2 recebe o pacote inteiro e pode, então, remover as informações de IP e DLSw do pacote e enviá-lo ao 3174 sem atrasos.

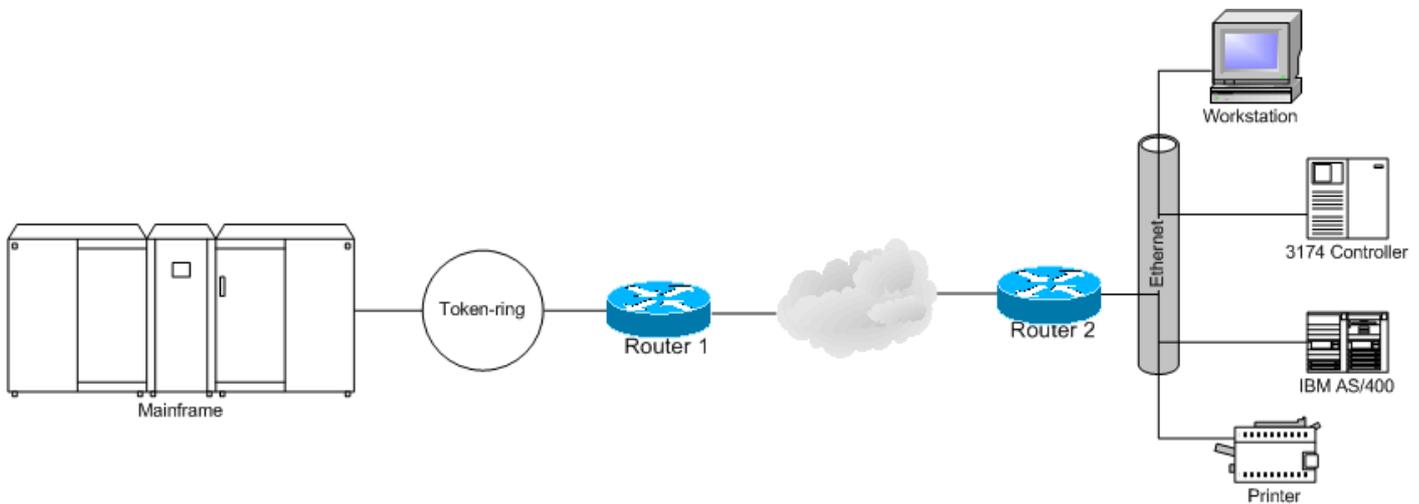
Cenário 2 - Atraso de pacotes fora de ordem



Neste cenário, há duas nuvens IP disponíveis com métricas iguais para balanceamento de carga ou redundância. Não habilitar o PMTD pode tornar o DLSw muito lento. Sem o PMTD ativado, o roteador 1 deve reunir o quadro de 521 bytes em dois pacotes TCP - um com 520 bytes de dados e o segundo com 1 byte de dados. Se o primeiro pacote atravessar a nuvem IP superior, há uma probabilidade significativa de ele chegar após o segundo pacote se o segundo pacote for enviado através da nuvem IP mais baixa e de custo igual. Isso gera o que é conhecido como pacote fora de ordem. Inerente na capacidade do protocolo IP/TCP é a capacidade de gerenciar esse problema. Pacotes fora de ordem são armazenados na memória aguardando que todo o fluxo chegue e, em seguida, sejam remontados. Pacotes fora de ordem não são incomuns, no entanto, todas as tentativas de minimizá-los devem ser feitas, pois esse evento utiliza recursos de memória e CPU. Uma grande quantidade de desperdícios pode causar um atraso significativo no nível do TCP. Se a sessão de camada 3/DLSw for atrasada, a sessão LLC2/SDLC que está sendo transportada sobre esse DLSw sofrerá subsequentemente. Se o PMTD estiver ativado nesse cenário, um único quadro de 521 bytes será enviado em um quadro TCP em qualquer nuvem IP. O roteador receptor precisa apenas de buffer e desencapsula um quadro TCP.

O PMTD não tem relação com o maior quadro anunciado de estação final para estação final em ambientes SNA. Isso inclui o maior quadro (LF) no campo Routing Information (RIF) em Token-Ring. O PMTD determina estritamente a quantidade de dados que pode ser encapsulada em um quadro TCP. LLC2 e SDLC não têm a capacidade de fragmentar pacotes, no entanto, o IP/TCP tem. Um grande quadro SNA pode ser segmentado à medida que é encapsulado em TCP. Esses dados são desencapsulados no roteador DLSw remoto e novamente apresentados como dados SNA não fragmentados.

Cenário 3 - Conectividade e throughput mais rápidos do LLC2



Neste cenário, o 3174 e a estação de trabalho têm sessões através do 3745 TIC para o Mainframe, se ambos os dispositivos enviam dados destinados ao host, é possível que o TCP possa colocar ambos os quadros LLC2 em um pacote. Sem o PMTD, isso não é possível se os dados combinados dos dois quadros tiverem 521 bytes ou mais. Nesse caso, o software TCP precisará enviar cada pacote separadamente. Por exemplo, se o 3174 e a estação de trabalho enviam um quadro aproximadamente ao mesmo tempo e cada um desses pacotes tem 400 bytes de dados, o roteador recebe e armazena em buffer cada quadro. O roteador agora deve encapsular cada um desses fluxos de dados de 400 bytes em pacotes TCP separados para transmissão ao peer.

Com o PMTD ativado e assumindo um MSS de 1460, o roteador recebe e coloca em buffer os dois pacotes LLC2. Agora, ele poderá encapsular ambos em um pacote. Esse pacote TCP conterá 40 bytes de informações de IP, 16 bytes de informações de DLSw para o primeiro emparelhamento de circuito DLSw, 400 bytes de dados, outros 16 bytes de dados para o segundo emparelhamento de circuito DLSw e os outros 400 bytes de dados. Esse cenário específico usa dois dispositivos e, portanto, dois circuitos DLSw. O PMTD permite que o DLSw escale para números mais altos de circuitos DLSw com mais eficiência. Muitas redes spoke-hub exigem centenas de locais remotos, cada um com um ou dois dispositivos SNA, fazendo ping em um roteador de local central que se conecta a um OSA ou FEP, fornecendo acesso aos aplicativos do host. O PMTD oferece ao TCP e ao DLSw a capacidade de escalar para requisitos maiores sem utilizar excessivamente a CPU do roteador e os recursos de memória, bem como fornecer um transporte mais rápido.

Observação: foi encontrado um bug de software no final da 12.1(5)T e resolvido na 12.2(5)T, em que o PMTD não estava trabalhando em um túnel VPN (Virtual Private Network). Este defeito de software é [CSCdt49552](#) (apenas clientes [registrados](#)).

Verificação do PMTD para DLSW

Emita o comando **show tcp**.

```
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

Essa exibição é identificada como uma sessão TCP DLSw porque uma das portas na sessão TCP é 2065. Perto da parte inferior da saída, o segmento de dados máximo é de 536 bytes. Esse valor indica que o roteador de peer DLSw remoto de 10.1.1.1 não tem o comando **ip tcp path-mtu-discovery** configurado. O valor de 536 bytes já é responsável pelos 40 bytes de informações de IP em um quadro IP. Esse valor de 536 bytes não representa os 16 bytes de informações de DLSw que seriam adicionadas a um pacote TCP que transporta tráfego SNA.

Com o comando **ip tcp path-mtu-discovery** configurado, o segmento de dados máximo agora é 1460. Além disso, a saída do comando **show tcp** indica que o **caminho mtu é capaz** imediatamente antes da instrução **máxima de segmento de dados**. A interface de saída tem um MTU de 1500 bytes. MTU é igual a 1.500 bytes menos 40 bytes de informações de IP é 1.460 bytes. O DLSw usará outros 16 bytes. Portanto, um quadro de até 1.444 bytes de LLC2 ou SDLC pode ser transmitido em um quadro TCP.

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

[Informações Relacionadas](#)

- [Notas técnicas de sistemas compatíveis: Fragmentação de IP e descoberta de caminho MTU com VPN](#)
- [Suporte Técnico - Cisco Systems](#)