Nota técnica de embalagem de OSPF, MTU e LSA

Contents

Introduction

Este documento descreve a interação de pacotes OSPF (Open Shortest Path First), MTU (Maximum Transition Unit), LSAs (Link State Advertisements) e LS (Link State Update Packets) no contexto do bug da Cisco ID CSCse01519.

Tamanho do pacote OSPF

Os links nos roteadores têm uma MTU. Os pacotes de saída, como os pacotes OSPF, não podem ser maiores que o MTU da interface.

Request for Comments (RFC) 2328 documenta a versão 2 do protocolo OSPF. O Apêndice A.1 do RFC 2328 descreve o Encapsulamento de pacotes OSPF desta maneira:

O OSPF é executado diretamente na camada de rede do Internet Protocol. Os pacotes OSPF são, portanto, encapsulados somente por cabeçalhos IP e de enlace de dados local.

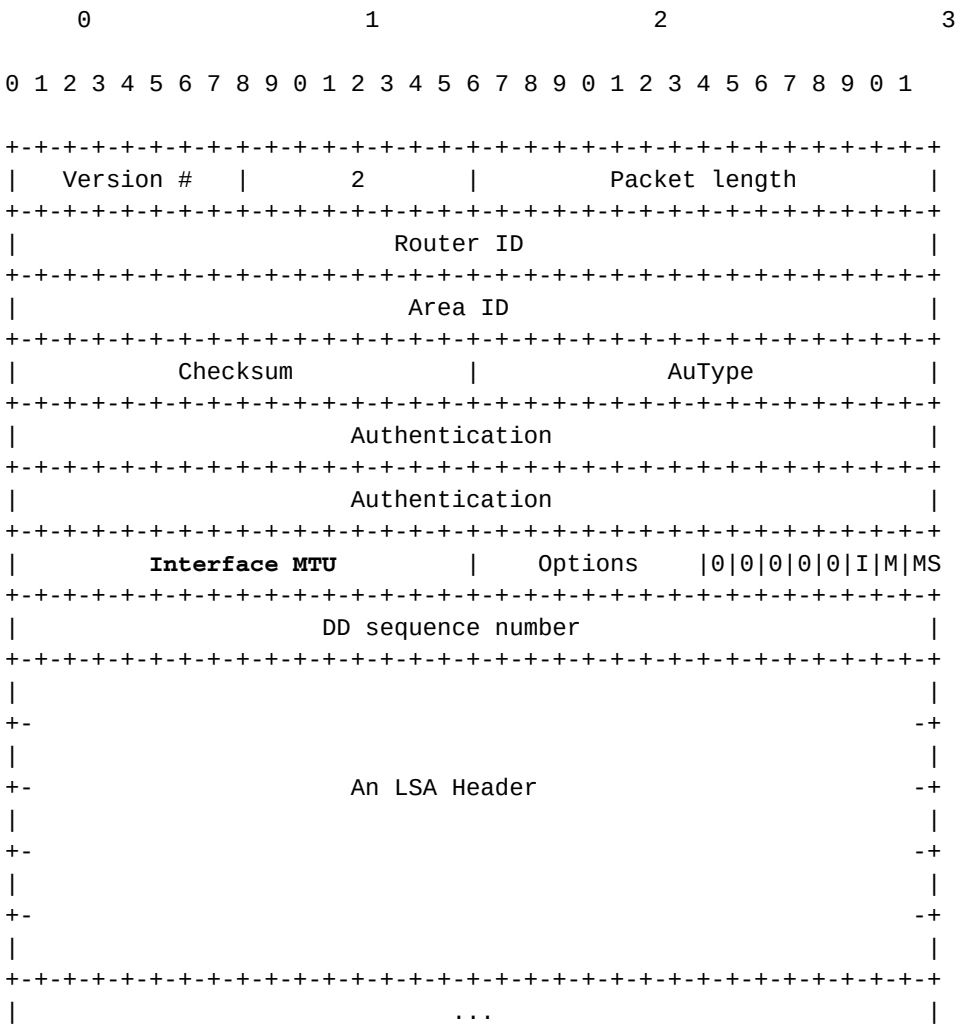
O OSPF não define uma maneira de fragmentar seus pacotes de protocolo e depende da fragmentação de IP ao transmitir pacotes maiores que o MTU da rede. Se necessário, o comprimento dos pacotes OSPF pode ser de até 65.535 bytes (incluindo o cabeçalho IP). Os tipos de pacotes OSPF que provavelmente serão grandes (Pacotes de Descrição de Banco de Dados, Solicitação de Estado de Link, Atualização de Link State e Pacotes de Confirmação de Link State) geralmente podem ser divididos em vários pacotes de protocolo separados, sem perda de funcionalidade. Recomenda-se que assim seja; A fragmentação de IP deve ser evitada sempre que possível.

Pode haver um ou mais LSAs em um pacote LS Update. Muitos LSAs em um pacote de

atualização de LS são conhecidos como empacotando LSAs em um pacote de atualização de LS.

MTU em pacote DBD

O pacote Database Description (DBD), também especificado no RFC 2328, descreve o conteúdo do banco de dados de link-state do OSPF:



Apêndice A.3.3. de RFC 2328 descreve a interface MTU como:

O tamanho em bytes do maior datagrama IP que pode ser enviado para fora da interface associada, sem fragmentação.

Os roteadores conectados a um link trocam seu valor de MTU de interface em pacotes DBD quando a adjacência OSPF é inicializada.

A seção 10.6 do RFC 2328 afirma:

Se o campo Interface MTU no pacote Database Description indicar um tamanho de datagrama IP maior do que o roteador pode aceitar na interface de recebimento sem fragmentação, o pacote Database Description será rejeitado.

Quando o comando **debug ip ospf adj** é usado, você pode ver a chegada desses pacotes DBD.

Neste exemplo, há uma incompatibilidade nos valores de MTU entre dois vizinhos OSPF. Este

roteador tem MTU 1600:

```
OSPF: Rcv DBD from 10.100.1.2 on GigabitEthernet0/1 seq 0x2124 opt 0x52 flag 0x2
len 1452 mtu 2000 state EXSTART
OSPF: Nbr 10.100.1.2 has larger interface MTU
```

O outro roteador OSPF tem a interface MTU 2000:

```
OSPF: Rcv DBD from 10.100.100.1 on GigabitEthernet0/1 seq 0x89E opt 0x52 flag 0x7
len 32 mtu 1600 state EXCHANGE
OSPF: Nbr 10.100.100.1 has smaller interface MTU
```

Os pacotes de DBD são retransmitidos continuamente até que a adjacência de OSPF seja finalmente destruída.

```
OSPF: Send DBD to 10.100.1.2 on GigabitEthernet0/1 seq 0x9E6 opt 0x52 flag 0x7
len 32
OSPF: Retransmitting DBD to 10.100.1.2 on GigabitEthernet0/1 [10]
OSPF: Send DBD to 10.100.1.2 on GigabitEthernet0/1 seq 0x9E6 opt 0x52 flag 0x7
len 32
OSPF: Retransmitting DBD to 10.100.1.2 on GigabitEthernet0/1 [11]
%OSPF-5-ADJCHG: Process 1, Nbr 10.100.1.2 on GigabitEthernet0/1 from EXSTART to
DOWN, Neighbor Down: Too many retransmissions
```

Comportamento do OSPF e empacotamento de LSAs em um pacote de atualização de LS

Antes do bug da Cisco ID CSCse01519

Antes do bug da Cisco ID CSCse01519, o OSPF no software Cisco IOS® construiu pacotes OSPF não maiores do que 1500 bytes, independentemente do MTU da interface. Portanto, se o MTU da interface fosse maior que 1.500 bytes, o OSPF ainda estaria empacotado somente até 1.500 bytes em um pacote OSPF. Isso foi um pouco ineficiente porque o OSPF poderia enviar pacotes maiores no link e alcançar maior throughput.

> **Note**: Houve uma exceção para este cenário. Se um LSA tivesse mais de 1500 bytes, o OSPF construiu esse pacote, independentemente do tamanho, porque o OSPF não pode fragmentar um LSA. A pilha IP do roteador então fragmentou o pacote para se ajustar à MTU da interface de saída. Isso normalmente ocorreu quando um roteador OSPF tinha muitos links e o roteador LSA se tornou maior que o link MTU.

Da mesma forma, se o MTU da interface de saída fosse menor que 1500 bytes, o processo OSPF ainda construiu ou empacotou pacotes OSPF com até 1500 bytes, e a pilha IP do roteador fragmentou o pacote em pacotes IP menores para ajustar o MTU do link de saída. Isso normalmente ocorreu com um túnel IPSec entre dois roteadores que estavam executando OSPF. A sobrecarga adicional dos bytes de encapsulamento do túnel levou a um MTU menor que 1.500 bytes. O OSPF construiu pacotes OSPF de até 1500 bytes e os pacotes foram fragmentados antes que o roteador os transmitisse. Isso foi uma ineficiência adicional.

Após o bug da Cisco ID CSCse01519

Após o bug da Cisco ID CSCse01519, o OSPF no IOS pode empacotar pacotes OSPF para ter mais de 1500 bytes. Isso ocorre se o MTU da interface de saída for maior que 1.500 bytes. As transmissões são mais eficientes porque mais informações podem ser embaladas em um pacote maior. Em outras palavras, se um roteador OSPF precisar transmitir muitos LSAs externos a um vizinho OSPF, ele poderá empacotar mais LSAs externos em um pacote de atualização de LS se esse roteador executar o IOS com o bug da Cisco ID CSCse01519 implementado.

O bug da Cisco ID CSCse01519 também permite que o OSPF crie pacotes menores que 1500 bytes. Em alguns cenários, o MTU entre dois vizinhos OSPF é menor que 1.500 bytes. No exemplo anterior com um túnel IPSec, o OSPF transmite pacotes OSPF menores que 1500 bytes e evita a fragmentação de IP; novamente, a exceção é o caso de um LSA maior que o MTU da interface.

ID de bug Cisco CSCse01519

Ao atualizar um roteador OSPF, você pode descobrir um problema de OSPF MTU causado pelo bug da Cisco ID CSCse01519.

Overview

Muitas redes têm vizinhos OSPF conectados através de uma rede comutada de Camada 2 (L2) ou de uma rede de transporte, composta por um serviço VPN de L2 ou uma rede SDH/SONET (Synchronous Digital Hierarchy/Synchronous Optical Network). Essas redes de transporte podem ter configurações de MTU diferentes dos roteadores que estão executando OSPF.

Embora a configuração MTU deva estar correta em todos os roteadores e deva refletir o MTU verdadeiro, muitas vezes há erros que passam despercebidos.

Esta é uma rede de exemplo com dois roteadores que estão executando OSPF. O roteador 1 (R1) e o roteador 2 (R2) estão conectados por meio de um switch L2.

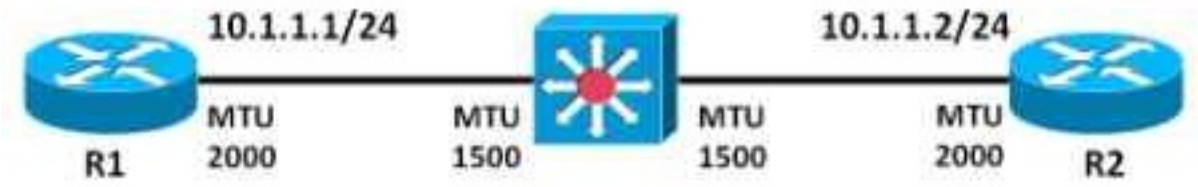


Figure 1 : Example network

Neste exemplo, os roteadores têm interfaces GigabitEthernet com um MTU definido como 2000.

O MTU do switch L2 tem apenas 1500 bytes.

Se o tamanho do tráfego de dados nunca for maior que 1500 bytes, você poderá usar o IOS sem o bug da Cisco ID CSCse01519 porque os pacotes OSPF nunca são maiores que 1500 bytes. No entanto, se houver um LSA com 1800 bytes, por exemplo, o processo OSPF em R1 ou R2 cria um pacote de atualização de LS maior que 1500 bytes e o transmite, mas o pacote é descartado pelo switch L2 entre os roteadores.

Se o banco de dados OSPF em R2 tiver redes suficientes, os LSAs originados localmente serão tão grandes que um pacote de atualização de LS pode ser maior que o MTU da interface.

- Se essas redes forem originadas pelo comando de rede de cobertura, as redes aparecerão no roteador LSA de R2. O R2 cria um LSA de roteador maior que 2.000 bytes e o transmite, mas o IP o fragmenta em 2.000 bytes, o MTU da interface. No entanto, o switch L2 descarta esses pacotes. O OSPF retransmite esse pacote infinitamente, e o estado de adjacência do OSPF nunca está cheio. Portanto, o problema é imediatamente descoberto, mesmo quando você está executando o IOS sem o bug da Cisco ID CSCse01519.
- Se essas redes forem originadas pelo comando **redistribute connected**, as redes aparecerão em LSAs externos. O OSPF tenta empacotar LSAs externos em um pacote de atualização de LS com até 1500 bytes de tamanho. Nesse caso, como o MTU da interface é de 2.000 bytes, a adjacência de OSPF alcança o estado 'FULL'. A questão de uma MTU subjacente inadequada não é imediatamente descoberta. O problema será descoberto quando um roteador for atualizado para o IOS com o bug da Cisco ID CSCse01519.

Cenário

Suponha que ambos os roteadores executem uma versão do IOS sem o bug da Cisco ID CSCse01519.

Quando a adjacência de OSFP é criada, observe que R1 nunca recebe um pacote OSPF maior que 1.500 bytes, embora o MTU das interfaces seja 2.000.

Ative o comando **debug ip ospf packets**.

```
OSPF: rcv. v:2 t:1 l:48 rid:10.100.1.2
     aid:0.0.0.0 chk:72CF aut:0 auk: from GigabitEthernet0/1
...
OSPF: rcv. v:2 t:4 l:1468 rid:10.100.1.2
     aid:0.0.0.0 chk:8389 aut:0 auk: from GigabitEthernet0/1
OSPF: rcv. v:2 t:4 l:136 rid:10.100.1.2
...
```

Nesta saída de depuração, 'l:1468' é o comprimento do pacote OSPF, de modo que você pode ver que o maior pacote OSPF tinha 1468 bytes. 't:4' indica que o pacote OSPF é tipo 4, que é um pacote de Atualização de estado de link. Esta tabela da seção 4.3 do RFC 2328 define os diferentes tipos de pacotes OSPF:

Tipo	Nome do pacote	Função de protocolo
1	Saudação	Descobrir/manter vizinhos
2	Descrição do banco de dados	Resuma o conteúdo do banco de dados
3	Solicitação de estado de enlace	Download do banco de dados
4	Atualização do estado de enlace	Atualização do banco de dados

5 Link State Ack Confirmação de inundação

A adjacência OSPF alcança o estado 'FULL'.

```
R1#show ip ospf neighbor gigabitEthernet 0/1

Neighbor ID     Pri   State           Dead Time   Address         Interface
10.100.1.2        0   FULL/  -        00:00:34    10.1.1.2        GigabitEthernet0/1

R2#show ip ospf neighbor gigabitEthernet 0/1

Neighbor ID     Pri   State           Dead Time   Address         Interface
10.100.100.1      0   FULL/  -        00:00:34    10.1.1.1        GigabitEthernet0/1
```

Em seguida, atualize o IOS em R2 para uma versão do IOS com o bug da Cisco ID CSCse01519.

```
R2#show ip ospf neighbor gigabitEthernet 0/1

Neighbor ID     Pri   State           Dead Time   Address         Interface
10.100.100.1      0   LOADING/  -     00:00:33    10.1.1.1        GigabitEthernet0/1

R2#show ip ospf neighbor gigabitEthernet 0/1 detail
Neighbor 10.100.100.1, interface address 10.1.1.1
   In the area 0 via interface GigabitEthernet0/1
   Neighbor priority is 0, State is LOADING, 5 state changes
   DR is 0.0.0.0 BDR is 0.0.0.0
   Options is 0x12 in Hello (E-bit L-bit )
   Options is 0x52 in DBD (E-bit L-bit O-bit)
   LLS Options is 0x1 (LR)
   Dead timer due in 00:00:39
   Neighbor is up for 00:00:49
   Index 1/1, retransmission queue length 0, number of retransmission 0
   First 0x0(0)/0x0(0) Next 0x0(0)/0x0(0)
   Last retransmission scan length is 0, maximum is 0
   Last retransmission scan time is 0 msec, maximum is 0 msec
   Number of retransmissions for last link state request packet 9
   Poll due in 00:00:00

R2#show ip ospf neighbor gigabitEthernet 0/1 detail
Neighbor 10.100.100.1, interface address 10.1.1.1
   In the area 0 via interface GigabitEthernet0/1
   Neighbor priority is 0, State is LOADING, 5 state changes
   DR is 0.0.0.0 BDR is 0.0.0.0
   Options is 0x12 in Hello (E-bit L-bit )
   Options is 0x52 in DBD (E-bit L-bit O-bit)
   LLS Options is 0x1 (LR)
   Dead timer due in 00:00:33
   Neighbor is up for 00:02:06
   Index 1/1, retransmission queue length 0, number of retransmission 0
   First 0x0(0)/0x0(0) Next 0x0(0)/0x0(0)
   Last retransmission scan length is 0, maximum is 0
   Last retransmission scan time is 0 msec, maximum is 0 msec
   Number of retransmissions for last link state request packet 25
   Poll due in 00:00:03

%OSPF-5-ADJCHG: Process 1, Nbr 10.100.100.1 on GigabitEthernet0/1 from LOADING
to DOWN, Neighbor Down: Too many retransmissions
```

A adjacência de OSPF está presa no estado 'CARREGANDO' e não atinge o estado 'COMPLETO'. As retransmissões ocorrem até que o OSPF atinja seu limite de 25 retransmissões. O OSPF tenta estabelecer a adjacência novamente, o mesmo problema ocorre novamente e o

loop continua infinitamente.

Assim, a atualização em R2 descobre um problema anteriormente oculto: o MTU subjacente é menor do que o usado pelos roteadores OSPF.

Quando o switch altera MTU para 2000, um pacote OSPF maior que 1500 bytes ('l:1980') é transmitido sem problemas.

```
R1#
OSPF: rcv. v:2 t:3 l:1980 rid:10.100.1.2
     aid:0.0.0.0 chk:AC5B aut:0 auk: from GigabitEthernet0/1
```

Para verificar os problemas subjacentes de MTU, faça sempre ping no endereço IP do vizinho OSPF com um tamanho igual ao MTU e ao bit DF (não fragmentar) definido.

Para descobrir o valor do MTU subjacente, execute um ping e varra o tamanho. Conte o número de pontos de exclamação (!) na saída para determinar o MTU correto. Neste exemplo, a última resposta de eco do comando **ping** tem tamanho 1500 bytes.

```
R2#ping
Protocol [ip]:
Target IP address: 10.1.1.1
Repeat count [5]: 1
Datagram size [100]:
Timeout in seconds [2]:
Extended commands [n]: yes
Source address or interface:
Type of service [0]:
Set DF bit in IP header? [no]: yes
Validate reply data? [no]:
Data pattern [0xABCD]:
Loose, Strict, Record, Timestamp, Verbose[none]:
Sweep range of sizes [n]: yes
Sweep min size [36]: 1460
Sweep max size [18024]: 1540
Sweep interval [1]:
Type escape sequence to abort.
Sending 81, [1460..1540]-byte ICMP Echos to 10.1.1.1, timeout is 2 seconds:
Packet sent with the DF bit set
!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!!..............................
...........
Success rate is 49 percent (40/81), round-trip min/avg/max = 1/1/4 ms
```