

# OSPF、MTU和LSA打包技術說明

## 目錄

[簡介](#)

[OSPF資料包大小](#)

[DBD封包中的MTU](#)

[OSPF行為並將LSA打包到LS更新資料包中](#)

[使用思科錯誤ID CSCse01519之前](#)

[思科錯誤ID CSCse01519之後](#)

[思科錯誤ID CSCse01519](#)

[概觀](#)

[案例](#)

## 簡介

本檔案介紹開放最短路徑優先(OSPF)封包、最大過渡單元(MTU)、連結狀態通告(LSA)和連結狀態(LS)更新封包在思科錯誤ID [CSCse01519](#)情況下的互動。

## OSPF資料包大小

路由器上的鏈路具有MTU。傳出資料包 ( 如OSPF資料包 ) 不能大於介面MTU。

[請求註釋\(RFC\)2328文檔](#)了OSPF協定的第2版。RFC 2328的附錄A.1以這種方式描述OSPF資料包的封裝：

OSPF直接通過Internet協定的網路層運行。因此，OSPF資料包僅通過IP報頭和本地資料鏈路報頭進行封裝。

OSPF不定義將其協定資料包分段的方法，並且在傳輸大於網路MTU的資料包時取決於IP分段。如有必要，OSPF資料包的長度最多可達65,535位元組 ( 包括IP報頭 )。可能較大的OSPF資料包型別 ( 資料庫描述資料包、鏈路狀態請求、鏈路狀態更新和鏈路狀態確認資料包 ) 通常可以拆分為多個獨立的協定資料包，而不會丟失功能。建議這樣做；應儘可能避免IP分段。

LS更新資料包中可以有一個或多個LSA。一個LS更新資料包中的許多LSA稱為將LSA打包到LS更新資料包中。

## DBD封包中的MTU

資料庫說明(DBD)資料包 ( 也在RFC 2328中指定 ) 描述了OSPF鏈路狀態資料庫的內容：



```
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
```

## OSPF行為並將LSA打包到LS更新資料包中

### 使用思科錯誤ID CSCse01519之前

在Cisco錯誤ID [CSCse01519](#)之前，Cisco IOS<sup>®</sup>軟體中的OSPF會構建不大於1500位元組的OSPF資料包，而不考慮介面MTU。因此，如果介面MTU大於1500位元組，OSPF仍最多只將1500位元組打包到OSPF資料包中。這種效率有些低下，因為OSPF可以在鏈路上傳送較大的資料包並實現更大的吞吐量。

**附註：**這種情況有一個例外。如果一個LSA的長度超過1500個位元組，無論大小如何，OSPF都會構建該資料包，因為OSPF無法分段一個LSA。接著路由器的IP堆疊將封包分段，以適合傳出介面的MTU。這通常發生在OSPF路由器有多個鏈路，並且路由器LSA變得大於鏈路MTU時。

同樣，如果傳出介面的MTU小於1500位元組，OSPF程式仍會建立或壓縮1500位元組的OSPF封包，而路由器的IP堆疊會將封包分段為較小的IP封包，以適合傳出連結的MTU。這通常發生在運行OSPF的兩台路由器之間的IPSec隧道中。通道封裝位元組的額外負荷導致的MTU小於1500位元組。OSPF構建的OSPF資料包最多1500位元組，然後資料包在路由器傳輸之前被分段。這是額外的低效率。

### 思科錯誤ID CSCse01519之後

在Cisco錯誤ID [CSCse01519](#)之後，IOS中的OSPF可以將OSPF資料包封裝為大於1500位元組。如果傳出介面的MTU大於1500位元組，就會發生這種情況。由於可以將更多資訊打包到一個更大的資料包中，因此傳輸效率更高。換句話說，如果某個OSPF路由器需要將多個外部LSA傳輸到OSPF鄰居，則在該路由器運行具有Cisco錯誤ID CSCse01519的IOS時，該路由器可以將多個外部LSA打包到一個LS更新資料包中。

思科錯誤ID CSCse01519還允許OSPF構建小於1500位元組的資料包。在某些情況下，兩個OSPF鄰居之間的MTU小於1500位元組。在上一個使用IPSec通道的範例中，OSPF傳輸小於1500位元組的OSPF封包，並避免IP分段；同樣地，大於介面MTU的LSA也會例外。

## 思科錯誤ID CSCse01519

升級OSPF路由器時，可能會發現Cisco錯誤ID [CSCse01519](#)導致的OSPF MTU問題。

### 概觀

許多網路具有OSPF鄰居，這些鄰居通過第2層(L2)交換網路或傳輸網路連線，該網路由L2 VPN服務或同步數字階層/同步光纖網路(SDH/SONET)網路組成。這些傳輸網路的MTU設定可能與運行OSPF的路由器不同。

雖然所有路由器上的MTU設定都應該正確且應反映真實的MTU，但經常會有一些錯誤未被注意。

這是運行OSPF的兩台路由器的網路示例。路由器1(R1)和路由器2(R2)通過L2交換機連線。

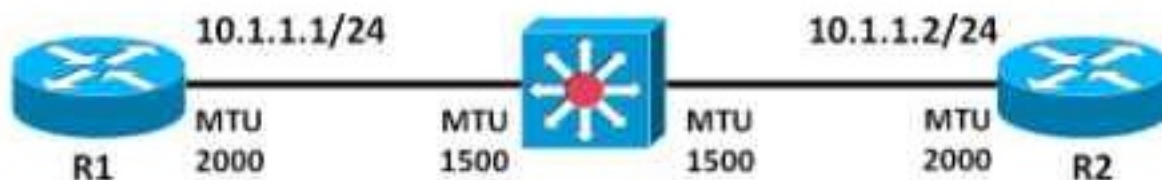


Figure 1 : Example network

在本例中，路由器的GigabitEthernet介面的MTU設定為2000。L2交換機的MTU只有1500位元組。

如果資料流量大小從不大於1500位元組，可以使用IOS而不使用思科錯誤ID [CSCse01519](#)，因為OSPF封包永遠不會大於1500位元組。例如，如果存在一個1800位元組的LSA，則R1或R2上的OSPF進程會構建一個大於1500位元組的LS更新資料包並進行傳輸，但該資料包會被路由器之間的L2交換機丟棄。

如果R2上的OSPF資料庫有足夠的網路，則本地產生的LSA非常大，因此LS更新資料包可能大於介面MTU。

- 如果這些網路是由covering network命令發起的，則這些網路會顯示在R2的路由器LSA中。R2構建的路由器LSA大於2000位元組並傳輸它，但IP將其分段為2000位元組，即介面MTU。但是，L2交換機丟棄這些資料包。然後，OSPF會無休止地重新傳輸此資料包，並且OSPF鄰接狀態永遠不會滿。因此，即使您執行的作業系統沒有思科錯誤ID CSCse01519，問題也會立即被發現。
- 如果這些網路由redistribute connected命令生成，則這些網路將顯示在外部LSA中。OSPF嘗試將外部LSA封裝到一個LS更新資料包中，該資料包的大小最高為1500位元組。在這種情況下，由於介面MTU為2000位元組，因此OSPF鄰接關係達到「FULL」狀態。不會立即發現基礎MTU不足的問題。將一台路由器升級為具有Cisco錯誤ID CSCse01519的IOS時，將發現此問題。

## 案例

假設兩台路由器都運行沒有Cisco錯誤ID [CSCse01519](#)的IOS版本。

建立OSPF鄰接關係時，請注意R1從未收到大於1500位元組的OSPF資料包，儘管介面的MTU為2000。

## 啟用debug ip ospf packets命令。

```
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
...
```

在此調試輸出中，「l:1468」是OSPF資料包的長度，因此您可以看到最大的OSPF資料包是1468位元組。「t:4」表示OSPF資料包為型別4，即鏈路狀態更新資料包。RFC 2328第4.3節中的下表定義了不同的OSPF資料包型別：

類型	資料包名稱	協定功能
1	Hello	發現/維護鄰居
2	資料庫說明	彙總資料庫內容
3	鏈路狀態請求	資料庫下載
4	鏈路狀態更新	資料庫更新
5	鏈路狀態Ack	泛洪確認

## OSPF鄰接關係達到「FULL」狀態。

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

Neighbor ID	Pri	State	Dead Time	Address	Interface
10.100.1.2	0	<b>FULL/</b> -	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	<b>FULL/</b> -	00:00:34	10.1.1.1	GigabitEthernet0/1

接下來，使用Cisco錯誤ID CSCse01519將R2上的IOS升級到IOS版本。

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

Neighbor ID	Pri	State	Dead Time	Address	Interface
10.100.100.1	0	<b>LOADING/</b> -	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
```

OSPF鄰接關係停滯在「LOADING」狀態，未達到「FULL」狀態。在OSPF達到25次重傳的限制之前會進行重傳。OSPF嘗試再次建立鄰接關係，相同的問題再次出現，環路無休止地繼續。

因此，R2上的升級發現了一個以前隱藏的問題：底層MTU小於OSPF路由器使用的MTU。

交換器將MTU變更為2000時，會順利傳輸大於1500位元組(「1:1980」)的OSPF封包。

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

若要檢查基本MTU問題，請一律對大小等於MTU和DF (不分段) 位集的OSPF鄰居IP位址執行ping。

為了發現底層MTU的值，請執行ping並掃描大小。對輸出中的感歎號(!)進行計數以確定正確的MTU。在本範例中，ping指令的最後一次回應回覆大小為1500位元組。

```
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
```