在Cisco IOS XR和BGP上验证路径MTU发现

目录

简介

背景信息

TCP PMTUD和TCP MSS

场景 — TCP PMTUD已禁用

- 使用默认MTU值
- 使用非默认MTU值 — 活动TCP对等体
- 使用非默认MTU值 — 被动TCP对等体
- 使用TCP选项 — XR活动
- 使用TCP选项 — XR被动
- 未直接连接的TCP对等体
- 未直接连接的TCP对等体 — 使用TCP选项(MD5)
- 未直接连接的TCP对等体 — 路径段的IP MTU较低

方案 — 启用TCP PMTUD

- 启用PMTUD
- PMTUD — 路径段的IP MTU较低
- PMTUD - TCP选项(MD5)
- PMTUD — 黑洞检测

简介

本文档介绍Cisco IOS® XR设备上的传输控制协议(TCP)路径最大传输单元(MTU)发现(PMTUD)。

背景信息

PMTUD机制尝试确定不需要在两台主机之间路径的任何位置分段的最大互联网协议(IP)数据包大小。建立的值是指定的路径MTU，并等于每一跳的MTU值的最小值。如果在传输信息时考虑路径MTU，则它允许您充分利用网络容量并避免分段和传输效率。通过使用边界网关协议(BGP)作为客户端协议，PMTUD机制和实施跨越了各种场景，逐渐揭示了PMTUD行为。

TCP PMTUD和TCP MSS

TCP利用PMTUD结果来影响本地最大分段大小(MSS)，这意味着它会动态地适应发现的路径MTU。因此，在继续使用PMTUD之前，您可以快速查看TCP最大分段大小(MSS)并了解其含义及其用途。

根据RFC879中的MSS原始定义：MSS选项的定义如下：在没有IP报头选项的IP数据报中传输没有TCP报头选项的TCP数据段中，此TCP选项的发送方可以接收的最大数据八位组数。

澄清某些方面并向实施者提供建议， RFC6691 突出显示必须如何计算MSS值：

当您计算要放入TCP MSS选项的值时，MTU值必须仅减小固定IP和TCP报头的大小，且不得减小以考虑任何可能的IP或TCP选项；相反，发送方必须缩短TCP数据长度，以说明其发送的数据包中包含的任何IP或TCP选项。

有关更详细的MSS定义，请参阅Cisco ASR 9000系列路由器IOS XR版本7.9.x的路由配置指南:

MSS是计算机或通信设备在单个未分段的TCP数据段中可以接收的最大数据量。所有TCP会话都受单个数据包中可以传输的字节数的限制；此限制为MSS。在将数据包向下传递到IP层之前，TCP会将数据包分成传输队列中的块。

TCP MSS值取决于接口的MTU，这是在一个实例中协议可以传输的最大数据长度。最大TCP数据包长度由源设备上出站接口的MTU和目的设备在TCP设置过程中通告的MSS共同决定。MSS越接近MTU，BGP消息的传输就越有效。每个数据流方向都可以使用不同的MSS值。

那么，对于给定TCP会话上的MSS，TCP必须考虑的值是什么？如何计算？

对于根据RFC879确定的默认值，您有：主机必须发送大于576个二进制八位数的数据报，除非它们明确知道目的主机准备接受更大的数据报。TCP最大数据段大小为IP最大数据报大小减四十个。

默认IP最大数据报大小为576。

默认TCP最大数据段大小为536。

这将考虑576字节的IP MTU值。但是，如果忽略实际IP MTU值，则可以将TCP MSS计算总结如下：

- 活动对等体 — 计算并发送带有SYN数据包的初始MSS。

```
MSS = IPMTU - sizeof(minimum TCPHDR) - sizeof(minimum IPHDR)
Where,
sizeof(minimum TCPHDR) = 20 bytes.
sizeof(minimum IPHDR) = 20 bytes.
```

- 被动对等体 — 计算初始MSS，与从主动对等体收到的MSS进行比较，并使用这些MSS值中的较低者发送SYN、ACK。

```
MIN[IPMTU - sizeof(minimum TCPHDR) - sizeof(minimum IPHDR) , Received MSS value]
Where,
sizeof(minimum TCPHDR) = 20 bytes.
sizeof(minimum IPHDR) = 20 bytes.

Received MSS value = MSS value received with Active Peer TCP SYN.
```

没有关于MSS选项值的协商。每个节点确定自己的值，并在TCP会话建立时通告相同的值。显然，如果可以从PMTUD导出用于MSS计算的IP MTU值，则MSS值可以适用于给定路径MTU的最有效值。思科IOS XR行为具有一些有关MSS计算和PMTUD角色的具体信息，概述如下。

默认情况下，思科IOS XR上禁用PMTUD:

- 根据以下内容，本地初始MSS计算会考虑IP MTU:
  - 如果是直连对等体 — 请考虑出口接口IP MTU。
  - 如果是非直连对等体 — 请考虑包含1280字节的IP MTU。
  - MSS值受配置的TCP选项的影响。

在Cisco IOS XR上启用PMTUD时：

- 根据以下内容，本地初始MSS计算会考虑IP MTU:
  - 不考虑直接/非直接连接对等体 — 考虑出口接口IP MTU。
  - MSS值受配置的TCP选项的影响。

关于PMTUD机制和实施还有更多细节需要考虑，本文档将通过下表中总结的实际示例来介绍这些细节。此表还显示所考虑的每个场景的主动TCP对等体和被动TCP对等体IP MTU以及所选MSS值。

PMTUD	Scenarios	ACTIVE IP MTU	PASSIVE IP MTU	MSS
Disabled	Using default MTU values	1500	1500	1460
	Using non-default MTU value – Active TCP peer	4460	1500	1460
	Using non-default MTU value – Passive TCP peer	1500	4460	1460
	Using TCP Options (MD5) – XR Active	1500	1500	1436
	Using TCP Options (MD5) – XR Passive	1500	1500	1460
	TCP peers not directly connected	1500	1500	1240
	TCP peers not directly connected – Using TCP Options (MD5)	1500	1500	1216
Enabled	Enabling TCP PMTUD	1500	1500	1460
	PMTUD in action – Path segment has lower MTU	1500	1500	1460
	PMTUD in action – TCP Options (MD5)	1500	1500	1436

场景 — TCP PMTUD已禁用

使用默认MTU值

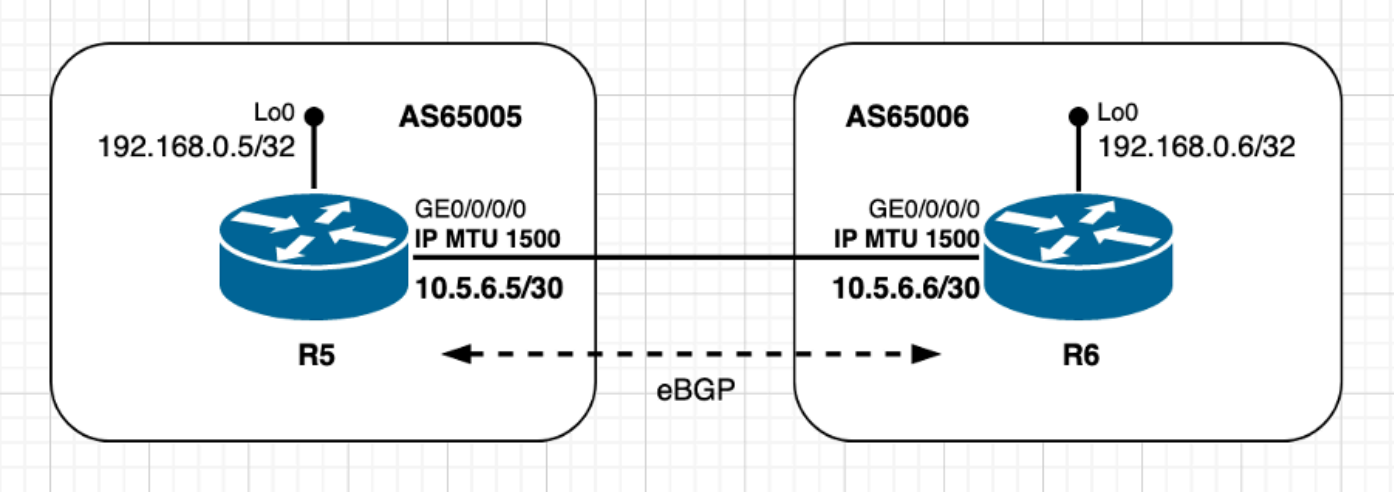


图2.1.使用默认MTU值

在映像2.1中显示的eBGP对等体管理TCP连接的情况下，这意味着，它会充当活动角色，并在目标端口179上启动与R5的TCP会话。对等体是直接连接的，并且都在各自的接口上使用默认IP MTU值。根据本文档开头部分共享的信息，此方案中的MSS计算可概述如下：

- 两个节点都使用默认IP MTU 1500字节
- 默认情况下禁用TCP路径MTU发现
- TCP对等体直接连接
  - R6管理BGP连接
  - R6发送MSS为1460字节的SYN
    - 1500（接口IP MTU）— 20(minTCP_H)- 20(minIP_H)
  - R5发送SYN、MSS为1460字节的ACK
    - 发送[Received MSS;本地初始MSS]
    - 接收的MSS 1460字节；本地初始MSS 1460字节
  - 两个对等体上均使用最低MSS值

TCP会话详细信息，如R6 - ACTIVE所示：

<#root>

```
! - As seen on R6 - ACTIVE

RP/0/0/CPU0:R6#show interfaces gigabitEthernet 0/0/0/0
Fri Jan  8 09:35:48.553 UTC
GigabitEthernet0/0/0/0 is up, line protocol is up
Interface state transitions: 1
Hardware is GigabitEthernet, address is fa16.3e85.3dc2 (bia fa16.3e85.3dc2)
Internet address is 10.5.6.6/30
```

MTU 1514 bytes

```
, BW 1000000 Kbit (Max: 1000000 Kbit)
<snip>

RP/0/0/CPU0:R6#show tcp brief
Fri Jan  8 09:36:22.491 UTC
PCB      VRF-ID     Recv-Q Send-Q Local Address          Foreign Address        State
<snip>
```

```
0x121649fc 0x60000000      0      0  10.5.6.6:24454         10.5.6.5:179           ESTAB
<snip>

RP/0/0/CPU0:R6#show tcp detail pcb 0x121649fc
Fri Jan  8 09:37:00.888 UTC
================================================================
Connection state is ESTAB, I/O status: 0, socket status: 0
Established at Fri Jan  8 09:28:28 2021

PCB 0x121649fc, SO 0x121561b8, TCPCB 0x12156f64, vrfid 0x60000000,
Pak Prio: Medium, TOS: 192, TTL: 1, Hash index: 78
Local host: 10.5.6.6, Local port: 24454 (Local App PID: 1011918)
Foreign host: 10.5.6.5, Foreign port: 179

Current send queue size in bytes: 0 (max 24576)
Current receive queue size in bytes: 0 (max 32768)  mis-ordered: 0 bytes
Current receive queue size in packets: 0 (max 0)

Timer          Starts    Wakeups         Next(msec)
Retrans            13          1                  0
SendWnd             0          0                  0
TimeWait            0          0                  0
AckHold            10          2                  0
KeepAlive           1          0                  0
PmtuAger            0          0                  0
GiveUp              0          0                  0
Throttle            0          0                  0

iss: 3757770712  snduna: 3757770960  sndnxt: 3757770960
sndmax: 3757770960  sndwnd: 32574       sndcwnd: 4380
irs: 1072103647  rcvnxt: 1072103895  rcvwnd: 32593   rcvadv: 1072136488

SRTT: 155 ms,  RTTO: 540 ms,  RTV: 385 ms,  KRTT: 0 ms
minRTT: 9 ms,  maxRTT: 229 ms

ACK hold time: 200 ms, Keepalive time: 0 sec, SYN waittime: 30 sec
Giveup time: 0 ms, Retransmission retries: 0, Retransmit forever: FALSE
Connect retries remaining: 30, connect retry interval: 50 secs

State flags: none
Feature flags: Win Scale, Nagle
Request flags: Win Scale


Datagrams (in bytes): MSS 1460, peer MSS 1460, min MSS 1460, max MSS 1460



Window scales: rcv 0, snd 0, request rcv 0, request snd 0
Timestamp option: recent 0, recent age 0, last ACK sent 0
Sack blocks {start, end}: none
Sack holes {start, end, dups, rxmit}: none

Socket options: SO_REUSEADDR, SO_REUSEPORT, SO_NBIO
Socket states: SS_ISCONNECTED, SS_PRIV
Socket receive buffer states: SB_DEL_WAKEUP
Socket send buffer states: SB_DEL_WAKEUP
Socket receive buffer: Low/High watermark 1/32768
Socket send buffer   : Low/High watermark 2048/24576, Notify threshold 0

PDU information:
#PDU's in buffer: 0
FIB Lookup Cache:  IFH: 0x40  PD ctx: size: 0  data:
```

```
Num Labels: 0  Label Stack:

RP/0/0/CPU0:R6
```

TCP会话详细信息（如R5上所示）— PASSIVE:

<#root>

```
! - As seen on R5 - PASSIVE

RP/0/0/CPU0:R5#show interfaces gigabitEthernet 0/0/0/0
Fri Jan  8 09:33:04.564 UTC
GigabitEthernet0/0/0/0 is up, line protocol is up
Interface state transitions: 1
Hardware is GigabitEthernet, address is fa16.3ead.518f (bia fa16.3ead.518f)
Internet address is 10.5.6.5/30
```

MTU 1514 bytes

```
, BW 1000000 Kbit (Max: 1000000 Kbit)
<snip>

RP/0/0/CPU0:R5#show tcp brief
Fri Jan  8 09:33:53.221 UTC
PCB      VRF-ID     Recv-Q Send-Q Local Address          Foreign Address        State
<snip>
0x12155884 0x60000000      0      0  10.5.6.5:179           10.5.6.6:24454         ESTAB
<snip>

RP/0/0/CPU0:R5#show tcp detail pcb 0x12155884
Fri Jan  8 09:34:47.317 UTC
================================================================
Connection state is ESTAB, I/O status: 0, socket status: 0
Established at Fri Jan  8 09:28:29 2021

PCB 0x12155884, SO 0x1215568c, TCPCB 0x12155a54, vrfid 0x60000000,
Pak Prio: Medium, TOS: 192, TTL: 1, Hash index: 78
Local host: 10.5.6.5, Local port: 179 (Local App PID: 1044686)
Foreign host: 10.5.6.6, Foreign port: 24454

Current send queue size in bytes: 0 (max 24576)
Current receive queue size in bytes: 0 (max 32768)  mis-ordered: 0 bytes
Current receive queue size in packets: 0 (max 0)

Timer          Starts    Wakeups         Next(msec)
Retrans             9          0                0
SendWnd             0          0                0
TimeWait            0          0                0
AckHold             9          7                0
KeepAlive           1          0                0
PmtuAger            0          0                0
GiveUp              0          0                0
Throttle            0          0                0

iss: 1072103647  snduna: 1072103857  sndnxt: 1072103857
sndmax: 1072103857  sndwnd: 32631       sndcwnd: 4380
irs: 3757770712  rcvnxt: 3757770922  rcvwnd: 32612   rcvadv: 3757803534

SRTT: 47 ms,  RTTO: 300 ms,  RTV: 170 ms,  KRTT: 0 ms
```

```
minRTT: 19 ms,  maxRTT: 219 ms

ACK hold time: 200 ms, Keepalive time: 0 sec, SYN waittime: 30 sec
Giveup time: 0 ms, Retransmission retries: 0, Retransmit forever: FALSE
Connect retries remaining: 0, connect retry interval: 0 secs

State flags: none
Feature flags: Win Scale, Nagle
Request flags: Win Scale

Datagrams (in bytes): MSS 1460, peer MSS 1460, min MSS 1460, max MSS 1460

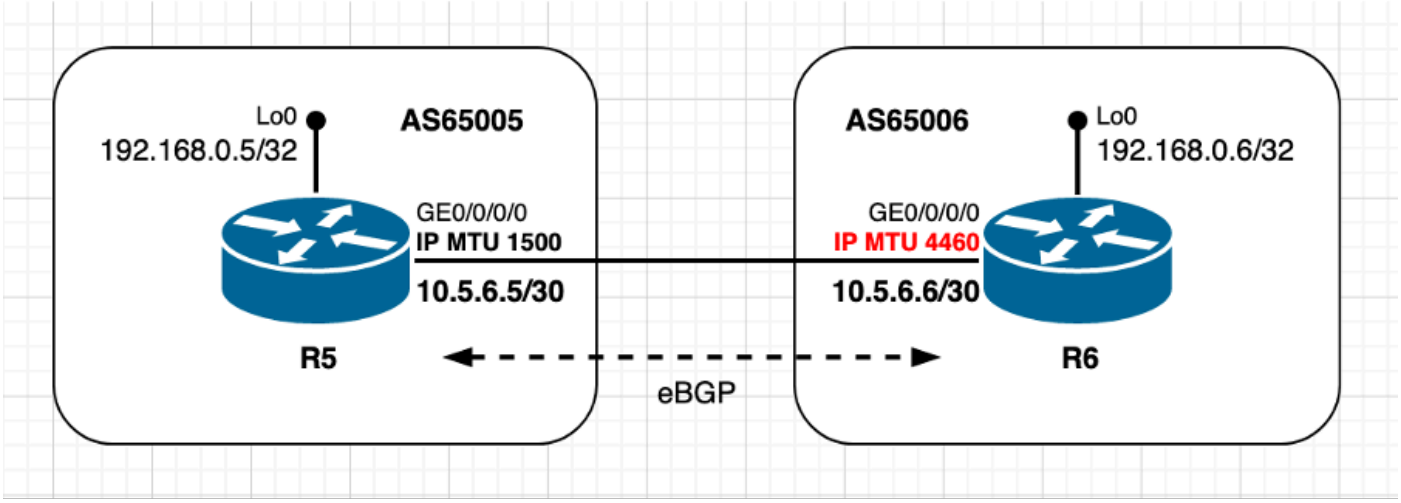

Window scales: rcv 0, snd 0, request rcv 0, request snd 0
Timestamp option: recent 0, recent age 0, last ACK sent 0
Sack blocks {start, end}: none
Sack holes {start, end, dups, rxmit}: none

Socket options: SO_REUSEADDR, SO_REUSEPORT, SO_NBIO
Socket states: SS_ISCONNECTED, SS_PRIV
Socket receive buffer states: SB_DEL_WAKEUP
Socket send buffer states: SB_DEL_WAKEUP
Socket receive buffer: Low/High watermark 1/32768
Socket send buffer    : Low/High watermark 2048/24576, Notify threshold 0

PDU information:
#PDU's in buffer: 0
FIB Lookup Cache:  IFH: 0x40  PD ctx: size: 0  data:
Num Labels: 0  Label Stack:

RP/0/0/CPU0:R5#
```

使用非默认MTU值 — 活动TCP对等体



映像2.2 — 活动对等体使用非默认MTU值

此场景基本上与上一个场景相同，唯一的区别是活动TCP对等体R6现在使用非默认IP MTU值。请注意被动TCP对等体R5是如何对MSS值进行初始计算和决策的。此场景中的TCP MSS计算可以总结如下：

- R6使用非默认IP MTU 4460字节
- 默认情况下禁用TCP路径MTU发现
- TCP对等体直接连接
  - R6管理BGP连接
  - R6发送MSS为4420字节的SYN
    - 4460（接口IP MTU） — 20(minTCP_H)- 20(minIP_H)
  - R5发送SYN、ACK，MSS为1460字节
    - 发送[Received MSS;本地初始MSS]
    - 接收的MSS 4420字节；本地初始MSS 1460字节
  - 两个对等体上均使用最低MSS值

源自R6的TCP SYN:

<#root>

```
! – TCP SYN sourced from R6

140    1598.150521    10.5.6.6       10.5.6.5       TCP    62     35502 → 179 [SYN] Seq=0 Win=16384 Len=0

MSS=4420

 WS=1

Frame 140: 62 bytes on wire (496 bits), 62 bytes captured (496 bits) on interface 0
Ethernet II, Src: fa:16:3e:85:3d:c2 (fa:16:3e:85:3d:c2), Dst: fa:16:3e:ad:51:8f (fa:16:3e:ad:51:8f)
Internet Protocol Version 4, Src: 10.5.6.6, Dst: 10.5.6.5
Transmission Control Protocol, Src Port: 35502, Dst Port: 179, Seq: 0, Len: 0
    Source Port: 35502
    Destination Port: 179
    [Stream index: 6]
    [TCP Segment Len: 0]
    Sequence number: 0    (relative sequence number)
    Acknowledgment number: 0
    Header Length: 28 bytes
    Flags: 0x002 (SYN)
    Window size value: 16384
    [Calculated window size: 16384]
    Checksum: 0x219d [unverified]
    [Checksum Status: Unverified]
    Urgent pointer: 0
    Options: (8 bytes), Maximum segment size, Window scale, End of Option List (EOL)
        Maximum segment size: 4420 bytes
            Kind: Maximum Segment Size (2)
            Length: 4

MSS Value: 4420

        Window scale: 0 (multiply by 1)
        End of Option List (EOL)
```

来自R5的TCP SYN、ACK:

<#root>

```
! - TCP SYN, ACK sourced from R5

141    1598.154866    10.5.6.5       10.5.6.6       TCP    62     179 → 35502 [SYN, ACK] Seq=0 Ack=1 Win=16

MSS=1460

 WS=1

Frame 141: 62 bytes on wire (496 bits), 62 bytes captured (496 bits) on interface 0
Ethernet II, Src: fa:16:3e:ad:51:8f (fa:16:3e:ad:51:8f), Dst: fa:16:3e:85:3d:c2 (fa:16:3e:85:3d:c2)
Internet Protocol Version 4, Src: 10.5.6.5, Dst: 10.5.6.6
Transmission Control Protocol, Src Port: 179, Dst Port: 35502, Seq: 0, Ack: 1, Len: 0
    Source Port: 179
    Destination Port: 35502
    [Stream index: 6]
    [TCP Segment Len: 0]
    Sequence number: 0    (relative sequence number)
    Acknowledgment number: 1    (relative ack number)
    Header Length: 28 bytes
    Flags: 0x012 (SYN, ACK)
    Window size value: 16384
    [Calculated window size: 16384]
    Checksum: 0xe2b4 [unverified]
    [Checksum Status: Unverified]
    Urgent pointer: 0
    Options: (8 bytes), Maximum segment size, Window scale, End of Option List (EOL)
        Maximum segment size: 1460 bytes
            Kind: Maximum Segment Size (2)
            Length: 4

MSS Value: 1460

        Window scale: 0 (multiply by 1)
        End of Option List (EOL)
```

TCP会话详细信息，如R6 - ACTIVE所示：

<#root>

```
! - as seen on R6 - Active

RP/0/0/CPU0:R6#show interfaces gigabitEthernet 0/0/0/0
Fri Jan  8 09:46:54.138 UTC
GigabitEthernet0/0/0/0 is up, line protocol is up
  Interface state transitions: 1
  Hardware is GigabitEthernet, address is fa16.3e85.3dc2 (bia fa16.3e85.3dc2)
  Internet address is 10.5.6.6/30

MTU 4474 bytes

, BW 1000000 Kbit (Max: 1000000 Kbit)
<snip>

RP/0/0/CPU0:R6#show tcp detail pcb 0x1215761c
Fri Jan  8 09:56:25.819 UTC
==============================================================
```

```
Connection state is ESTAB, I/O status: 0, socket status: 0
Established at Fri Jan  8 09:51:46 2021

PCB 0x1215761c, SO 0x12156f64, TCPCB 0x1216419c, vrfid 0x60000000,
Pak Prio: Medium, TOS: 192, TTL: 1, Hash index: 886
Local host: 10.5.6.6, Local port: 35502 (Local App PID: 1011918)
Foreign host: 10.5.6.5, Foreign port: 179

Current send queue size in bytes: 0 (max 24576)
Current receive queue size in bytes: 0 (max 32768)  mis-ordered: 0 bytes
Current receive queue size in packets: 0 (max 0)

Timer          Starts    Wakeups         Next(msec)
Retrans             9          1                 0
SendWnd             0          0                 0
TimeWait            0          0                 0
AckHold             6          5                 0
KeepAlive           1          0                 0
PmtuAger            0          0                 0
GiveUp              0          0                 0
Throttle            0          0                 0

   iss: 764231407   snduna: 764231579   sndnxt: 764231579
sndmax: 764231579   sndwnd: 32650       sndcwnd: 4380
   irs: 2712512697  rcvnxt: 2712512869  rcvwnd: 32669   rcvadv: 2712545538

SRTT: 31 ms,  RTTO: 300 ms,  RTV: 130 ms,  KRTT: 0 ms
minRTT: 9 ms,  maxRTT: 239 ms

ACK hold time: 200 ms, Keepalive time: 0 sec, SYN waittime: 30 sec
Giveup time: 0 ms, Retransmission retries: 0, Retransmit forever: FALSE
Connect retries remaining: 30, connect retry interval: 50 secs

State flags: none
Feature flags: Win Scale, Nagle
Request flags: Win Scale


Datagrams (in bytes): MSS 1460, peer MSS 1460, min MSS 4420, max MSS 4420



Window scales: rcv 0, snd 0, request rcv 0, request snd 0
Timestamp option: recent 0, recent age 0, last ACK sent 0
Sack blocks {start, end}: none
Sack holes {start, end, dups, rxmit}: none

Socket options: SO_REUSEADDR, SO_REUSEPORT, SO_NBIO
Socket states: SS_ISCONNECTED, SS_PRIV
Socket receive buffer states: SB_DEL_WAKEUP
Socket send buffer states: SB_DEL_WAKEUP
Socket receive buffer: Low/High watermark 1/32768
Socket send buffer   : Low/High watermark 2048/24576, Notify threshold 0

PDU information:
 #PDU's in buffer: 0
FIB Lookup Cache:  IFH: 0x40  PD ctx: size: 0  data:
  Num Labels: 0  Label Stack:

RP/0/0/CPU0:R6#
```

TCP会话详细信息（如R5上所示）— PASSIVE:

<#root>

```
! - as seen on R5 - Passive

RP/0/0/CPU0:R5#show tcp detail pcb 0x12155a98
Fri Jan  8 09:55:18.193 UTC
==============================================================
Connection state is ESTAB, I/O status: 0, socket status: 0
Established at Fri Jan  8 09:51:47 2021

PCB 0x12155a98, SO 0x12153ea0, TCPCB 0x12154e18, vrfid 0x60000000,
Pak Prio: Medium, TOS: 192, TTL: 1, Hash index: 886
Local host: 10.5.6.5, Local port: 179 (Local App PID: 1044686)
Foreign host: 10.5.6.6, Foreign port: 35502

Current send queue size in bytes: 0 (max 24576)
Current receive queue size in bytes: 0 (max 32768)  mis-ordered: 0 bytes
Current receive queue size in packets: 0 (max 0)

Timer          Starts    Wakeups         Next(msec)
Retrans             6          0                 0
SendWnd             0          0                 0
TimeWait            0          0                 0
AckHold             6          1                 0
KeepAlive           1          0                 0
PmtuAger            0          0                 0
GiveUp              0          0                 0
Throttle            0          0                 0

   iss: 2712512697  snduna: 2712512850  sndnxt: 2712512850
sndmax: 2712512850  sndwnd: 32688       sndcwnd: 4380
   irs: 764231407   rcvnxt: 764231560   rcvwnd: 32669   rcvadv: 764264229

SRTT: 107 ms,  RTTO: 538 ms,  RTV: 431 ms,  KRTT: 0 ms
minRTT: 29 ms,  maxRTT: 219 ms

ACK hold time: 200 ms, Keepalive time: 0 sec, SYN waittime: 30 sec
Giveup time: 0 ms, Retransmission retries: 0, Retransmit forever: FALSE
Connect retries remaining: 0, connect retry interval: 0 secs

State flags: none
Feature flags: Win Scale, Nagle
Request flags: Win Scale


Datagrams (in bytes): MSS 1460, peer MSS 4420, min MSS 1460, max MSS 1460

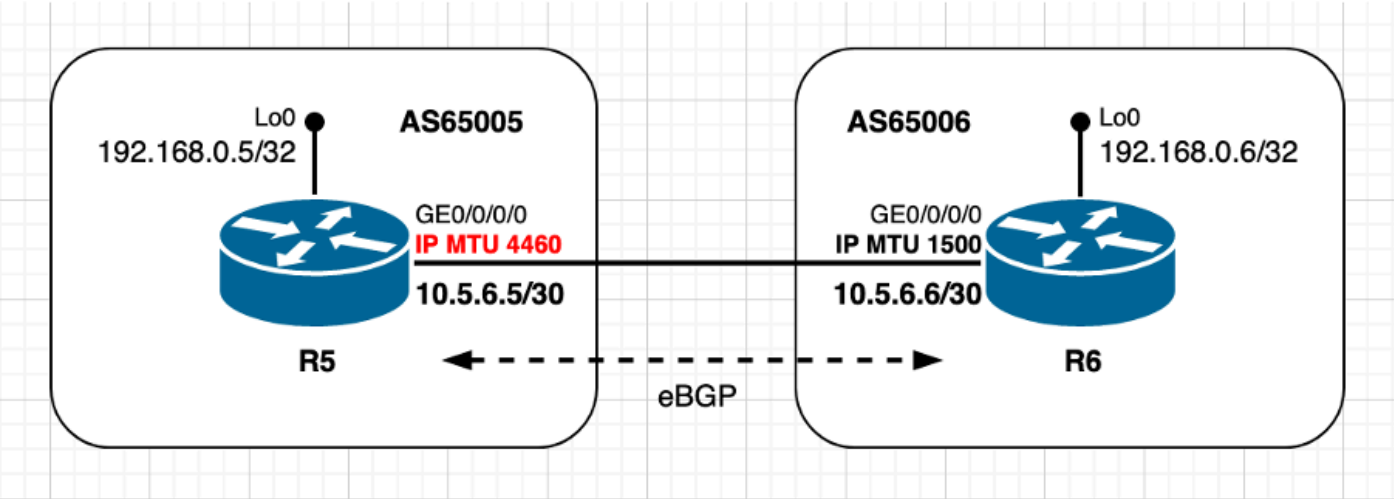


Window scales: rcv 0, snd 0, request rcv 0, request snd 0
Timestamp option: recent 0, recent age 0, last ACK sent 0
Sack blocks {start, end}: none
Sack holes {start, end, dups, rxmit}: none
Socket options: SO_REUSEADDR, SO_REUSEPORT, SO_NBIO
Socket states: SS_ISCONNECTED, SS_PRIV
Socket receive buffer states: SB_DEL_WAKEUP
Socket send buffer states: SB_DEL_WAKEUP
Socket receive buffer: Low/High watermark 1/32768
Socket send buffer   : Low/High watermark 2048/24576, Notify threshold 0
```

```
PDU information:
 #PDU's in buffer: 0
FIB Lookup Cache:  IFH: 0x40  PD ctx: size: 0  data:
  Num Labels: 0  Label Stack:

RP/0/0/CPU0:R5#
```

使用非默认MTU值 — 被动TCP对等体



映像2.3 — 被动对等体使用非默认MTU值。

仍然使用相同的eBGP方案，但现在使用非默认IP MTU配置被动TCP对等体R5，使用默认IP MTU值配置主动TCP对等体R6。与上一个场景一样，请注意被动对等体R5如何选择MSS值。此场景中的TCP MSS计算可概述如下：

- R5使用非默认IP MTU 4460字节
- 默认情况下禁用TCP路径MTU发现
- TCP对等体直接连接
  - R6管理BGP连接
  - R6发送MSS为1460字节的SYN
    - 1500（接口IP MTU） — 20(minTCP_H)- 20(minIP_H)
  - R5发送SYN、ACK，MSS为1460字节
    - 发送[Received MSS;本地初始MSS]
    - 接收的MSS 1460字节；本地初始MSS 4420字节
  - 两个对等体上均使用最低MSS值

源自R6的TCP SYN:

<#root>

```
! - TCP SYN sourced from R6

237    2696.666481    10.5.6.6       10.5.6.5       TCP    62     47007 → 179 [SYN] Seq=0 Win=16384 Len=0
```

MSS=1460

```
 WS=1

Frame 237: 62 bytes on wire (496 bits), 62 bytes captured (496 bits) on interface 0
Ethernet II, Src: fa:16:3e:85:3d:c2 (fa:16:3e:85:3d:c2), Dst: fa:16:3e:ad:51:8f (fa:16:3e:ad:51:8f)
Internet Protocol Version 4, Src: 10.5.6.6, Dst: 10.5.6.5
Transmission Control Protocol, Src Port: 47007, Dst Port: 179, Seq: 0, Len: 0
    Source Port: 47007
    Destination Port: 179
    [Stream index: 10]
    [TCP Segment Len: 0]
    Sequence number: 0    (relative sequence number)
    Acknowledgment number: 0
    Header Length: 28 bytes
    Flags: 0x002 (SYN)
    Window size value: 16384
    [Calculated window size: 16384]
    Checksum: 0x2025 [unverified]
    [Checksum Status: Unverified]
    Urgent pointer: 0
    Options: (8 bytes), Maximum segment size, Window scale, End of Option List (EOL)
        Maximum segment size: 1460 bytes
            Kind: Maximum Segment Size (2)
            Length: 4
```

MSS Value: 1460

```
        Window scale: 0 (multiply by 1)
        End of Option List (EOL)
```

来自R5的TCP SYN、ACK:

<#root>

```
! - TCP SYN, ACK sourced from R5

238    2696.702792    10.5.6.5       10.5.6.6       TCP    62     179 → 47007 [SYN, ACK] Seq=0 Ack=1 Win=16
```

MSS=1460

```
 WS=1

Frame 238: 62 bytes on wire (496 bits), 62 bytes captured (496 bits) on interface 0
Ethernet II, Src: fa:16:3e:ad:51:8f (fa:16:3e:ad:51:8f), Dst: fa:16:3e:85:3d:c2 (fa:16:3e:85:3d:c2)
Internet Protocol Version 4, Src: 10.5.6.5, Dst: 10.5.6.6
Transmission Control Protocol, Src Port: 179, Dst Port: 47007, Seq: 0, Ack: 1, Len: 0
    Source Port: 179
    Destination Port: 47007
    [Stream index: 10]
    [TCP Segment Len: 0]
    Sequence number: 0    (relative sequence number)
    Acknowledgment number: 1    (relative ack number)
    Header Length: 28 bytes
    Flags: 0x012 (SYN, ACK)
    Window size value: 16384
    [Calculated window size: 16384]
    Checksum: 0x7078 [unverified]
    [Checksum Status: Unverified]
```

```
    Urgent pointer: 0
    Options: (8 bytes), Maximum segment size, Window scale, End of Option List (EOL)
        Maximum segment size: 1460 bytes
            Kind: Maximum Segment Size (2)
            Length: 4
```

MSS Value: 1460

```
        Window scale: 0 (multiply by 1)
        End of Option List (EOL)
```

TCP会话详细信息，如R6 - ACTIVE所示：

<#root>

```
! - as seen on R6 - Active

RP/0/0/CPU0:R6#show tcp detail pcb 0x1215761c
Fri Jan  8 10:15:20.351 UTC
==============================================================
Connection state is ESTAB, I/O status: 0, socket status: 0
Established at Fri Jan  8 10:10:04 2021

PCB 0x1215761c, SO 0x12162aac, TCPCB 0x12156f64, vrfid 0x60000000,
Pak Prio: Medium, TOS: 192, TTL: 1, Hash index: 103
Local host: 10.5.6.6, Local port: 47007 (Local App PID: 1011918)
Foreign host: 10.5.6.5, Foreign port: 179

Current send queue size in bytes: 0 (max 24576)
Current receive queue size in bytes: 0 (max 32768)  mis-ordered: 0 bytes
Current receive queue size in packets: 0 (max 0)

Timer          Starts    Wakeups         Next(msec)
Retrans            10          1                 0
SendWnd             0          0                 0
TimeWait            0          0                 0
AckHold             7          5                 0
KeepAlive           1          0                 0
PmtuAger            0          0                 0
GiveUp              0          0                 0
Throttle            0          0                 0

   iss: 3949093168  snduna: 3949093359  sndnxt: 3949093359
sndmax: 3949093359  sndwnd: 32631       sndcwnd: 4380
   irs: 54439005    rcvnxt: 54439196    rcvwnd: 32650   rcvadv: 54471846

SRTT: 75 ms,  RTTO: 459 ms,  RTV: 384 ms,  KRTT: 0 ms
minRTT: 9 ms,  maxRTT: 239 ms

ACK hold time: 200 ms, Keepalive time: 0 sec, SYN waittime: 30 sec
Giveup time: 0 ms, Retransmission retries: 0, Retransmit forever: FALSE
Connect retries remaining: 30, connect retry interval: 50 secs

State flags: none
Feature flags: Win Scale, Nagle
Request flags: Win Scale
```

Datagrams (in bytes): MSS 1460, peer MSS 1460, min MSS 1460, max MSS 1460

```
Window scales: rcv 0, snd 0, request rcv 0, request snd 0
Timestamp option: recent 0, recent age 0, last ACK sent 0
Sack blocks {start, end}: none
Sack holes {start, end, dups, rxmit}: none
Socket options: SO_REUSEADDR, SO_REUSEPORT, SO_NBIO
Socket states: SS_ISCONNECTED, SS_PRIV
Socket receive buffer states: SB_DEL_WAKEUP
Socket send buffer states: SB_DEL_WAKEUP
Socket receive buffer: Low/High watermark 1/32768
Socket send buffer   : Low/High watermark 2048/24576, Notify threshold 0

PDU information:
 #PDU's in buffer: 0
FIB Lookup Cache:  IFH: 0x40  PD ctx: size: 0  data:
  Num Labels: 0  Label Stack:

RP/0/0/CPU0:R6#
```

TCP会话详细信息（如R5上所示）— PASSIVE:

<#root>

```
! - as seen on R5 - Passive

RP/0/0/CPU0:R5#show interfaces gigabitEthernet 0/0/0/0
Fri Jan  8 10:10:39.110 UTC
GigabitEthernet0/0/0/0 is up, line protocol is up
  Interface state transitions: 1
  Hardware is GigabitEthernet, address is fa16.3ead.518f (bia fa16.3ead.518f)
  Internet address is 10.5.6.5/30
```

MTU 4474 bytes

```
, BW 1000000 Kbit (Max: 1000000 Kbit)
<snip>

RP/0/0/CPU0:R5#show tcp detail pcb 0x121550fc
Fri Jan  8 10:14:20.105 UTC
==============================================================
Connection state is ESTAB, I/O status: 0, socket status: 0
Established at Fri Jan  8 10:10:05 2021

PCB 0x121550fc, SO 0x12154e18, TCPCB 0x12154304, vrfid 0x60000000,
Pak Prio: Medium, TOS: 192, TTL: 1, Hash index: 103
Local host: 10.5.6.5, Local port: 179 (Local App PID: 1044686)
Foreign host: 10.5.6.6, Foreign port: 47007

Current send queue size in bytes: 0 (max 24576)
Current receive queue size in bytes: 0 (max 32768)  mis-ordered: 0 bytes
Current receive queue size in packets: 0 (max 0)

Timer          Starts    Wakeups         Next(msec)
Retrans             7          0                 0
SendWnd             0          0                 0
```

```
TimeWait            0          0               0
AckHold             7          2               0
KeepAlive           1          0               0
PmtuAger            0          0               0
GiveUp              0          0               0
Throttle            0          0               0

   iss: 54439005    snduna: 54439177    sndnxt: 54439177
sndmax: 54439177    sndwnd: 32669       sndcwnd: 4380
   irs: 3949093168  rcvnxt: 3949093340  rcvwnd: 32650   rcvadv: 3949125990

SRTT: 117 ms,  RTTO: 570 ms,  RTV: 453 ms,  KRTT: 0 ms
minRTT: 19 ms,  maxRTT: 229 ms

ACK hold time: 200 ms, Keepalive time: 0 sec, SYN waittime: 30 sec
Giveup time: 0 ms, Retransmission retries: 0, Retransmit forever: FALSE
Connect retries remaining: 0, connect retry interval: 0 secs

State flags: none
Feature flags: Win Scale, Nagle
Request flags: Win Scale


Datagrams (in bytes): MSS 1460, peer MSS 1460, min MSS 4420, max MSS 4420



Window scales: rcv 0, snd 0, request rcv 0, request snd 0
Timestamp option: recent 0, recent age 0, last ACK sent 0
Sack blocks {start, end}: none
Sack holes {start, end, dups, rxmit}: none

Socket options: SO_REUSEADDR, SO_REUSEPORT, SO_NBIO
Socket states: SS_ISCONNECTED, SS_PRIV
Socket receive buffer states: SB_DEL_WAKEUP
Socket send buffer states: SB_DEL_WAKEUP
Socket receive buffer: Low/High watermark 1/32768
Socket send buffer   : Low/High watermark 2048/24576, Notify threshold 0

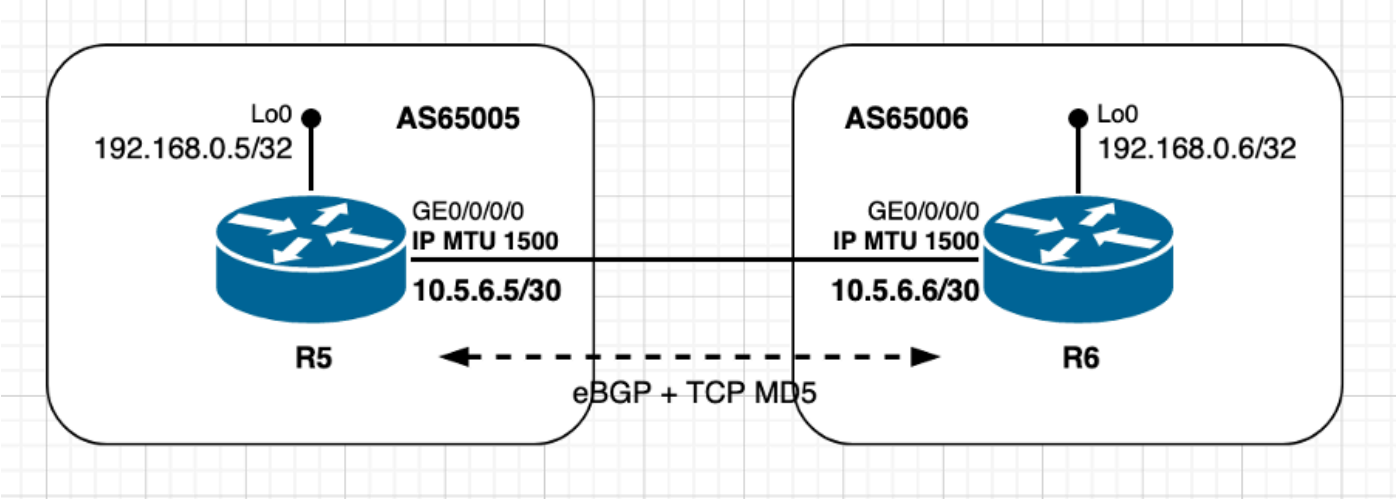
PDU information:
 #PDU's in buffer: 0
FIB Lookup Cache:  IFH: 0x40  PD ctx: size: 0  data:
  Num Labels: 0  Label Stack:

RP/0/0/CPU0:R5#
```

使用TCP选项 — XR活动

如本文档前面所述，使用TCP选项(例如TCP MD5、TCP selective-ack或TCP timestamps)会影响MSS计算，因为这些选项会导致在MSS计算中计入更多字节。

本部分以及下一部分旨在说明存在TCP选项时对等体进行的MSS计算。以TCP MD5身份验证选项为例。请参阅图2.4中的参考场景，如图所示。



图像2.4 — 使用TCP选项(MD5)- XR活动。

在这种情况下，两个对等体都使用默认IP MTU值，它们是直接连接的，并且对等体R6扮演TCP主动角色。因为已经共享TCP MD5身份验证帐户的配置和使用，因此会产生额外开销。此特定场景中的TCP MSS计算可概述如下：

- 两个节点都使用默认IP MTU 1500字节
- 默认情况下禁用TCP路径MTU发现
- TCP对等体直接连接
- 两个节点上都启用了TCP MD5身份验证
    - R6管理BGP连接
    - R6发送MSS为1436字节的SYN
        - 1500（接口IP MTU）— 20(minTCP_H)- 20(minIP_H)- 24字节（IOS XR TCP选项开销）
    - R5发送SYN、ACK，MSS为1436字节
        - 发送[Received MSS;本地初始MSS]
        - 接收的MSS 1436字节；本地初始MSS 1460字节
    - 两个对等体上均使用最低MSS值

从摘要中可以看到，Cisco IOS XR的行为方式并不严格符合RFC 879和RFC 6691，后者指出TCP选项不能计入MSS计算。

Cisco Bug ID CSCvf20166中进一步记录了Cisco IOS XR对tcp报头长度的额外因素的描述：

"(...)当XR启动BGP连接时，BGP首先创建套接字，然后设置套接字选项，包括MD5。这会使tcp选项报头长度= 24。因此初始MSS变为1500 - 40 - 24 = 1436。这会发送到对等设备，对等设备使用min(1436, 1460)= 1436.(..)

源自R6的TCP SYN:

```
<#root>

! - TCP SYN sourced from R6

430     5775.839420     10.5.6.6        10.5.6.5        TCP     82      24785 → 179 [SYN] Seq=0 Win=16384 Len=0
```

MSS=1436

WS=1

```
Frame 430: 82 bytes on wire (656 bits), 82 bytes captured (656 bits) on interface 0
Ethernet II, Src: fa:16:3e:85:3d:c2 (fa:16:3e:85:3d:c2), Dst: fa:16:3e:ad:51:8f (fa:16:3e:ad:51:8f)
Internet Protocol Version 4, Src: 10.5.6.6, Dst: 10.5.6.5
Transmission Control Protocol, Src Port: 24785, Dst Port: 179, Seq: 0, Len: 0
    Source Port: 24785
    Destination Port: 179
    [Stream index: 14]
    [TCP Segment Len: 0]
    Sequence number: 0    (relative sequence number)
    Acknowledgment number: 0
    Header Length: 48 bytes
    Flags: 0x002 (SYN)
    Window size value: 16384
    [Calculated window size: 16384]
    Checksum: 0xd62b [unverified]
    [Checksum Status: Unverified]
    Urgent pointer: 0
    Options: (28 bytes), Maximum segment size, Window scale, No-Operation (NOP),
```

TCP MD5 signature

```
, End of Option List (EOL)
        Maximum segment size: 1436 bytes
            Kind: Maximum Segment Size (2)
            Length: 4
```

MSS Value: 1436

```
        Window scale: 0 (multiply by 1)
        No-Operation (NOP)
        TCP MD5 signature
        End of Option List (EOL)
```

来自R5的TCP SYN、ACK:

<#root>

```
! – TCP SYN, ACK sourced from R5

431    5775.845744    10.5.6.5       10.5.6.6       TCP    82     179 → 24785 [SYN, ACK] Seq=0 Ack=1 Win=16
```

MSS=1436

WS=1

```
Frame 431: 82 bytes on wire (656 bits), 82 bytes captured (656 bits) on interface 0
Ethernet II, Src: fa:16:3e:ad:51:8f (fa:16:3e:ad:51:8f), Dst: fa:16:3e:85:3d:c2 (fa:16:3e:85:3d:c2)
Internet Protocol Version 4, Src: 10.5.6.5, Dst: 10.5.6.6
Transmission Control Protocol, Src Port: 179, Dst Port: 24785, Seq: 0, Ack: 1, Len: 0
    Source Port: 179
    Destination Port: 24785
    [Stream index: 14]
    [TCP Segment Len: 0]
    Sequence number: 0    (relative sequence number)
```

```
    Acknowledgment number: 1    (relative ack number)
    Header Length: 48 bytes
    Flags: 0x012 (SYN, ACK)
    Window size value: 16384
    [Calculated window size: 16384]
    Checksum: 0xe83d [unverified]
    [Checksum Status: Unverified]
    Urgent pointer: 0
    Options: (28 bytes), Maximum segment size, Window scale, No-Operation (NOP),
```

TCP MD5 signature

```
, End of Option List (EOL)
        Maximum segment size: 1436 bytes
            Kind: Maximum Segment Size (2)
            Length: 4
```

MSS Value: 1436

```
        Window scale: 0 (multiply by 1)
        No-Operation (NOP)
        TCP MD5 signature
        End of Option List (EOL)
```

TCP会话详细信息，如R6 - ACTIVE所示：

<#root>

```
! - as seen on R6 - Active

RP/0/0/CPU0:R6#show tcp detail pcb 0x1215761c
Fri Jan  8 11:14:13.599 UTC
================================================================
Connection state is ESTAB, I/O status: 0, socket status: 0
Established at Fri Jan  8 11:01:21 2021

PCB 0x1215761c, SO 0x1216419c, TCPCB 0x121649fc, vrfid 0x60000000,
Pak Prio: Medium, TOS: 192, TTL: 1, Hash index: 409
Local host: 10.5.6.6, Local port: 24785 (Local App PID: 1011918)
Foreign host: 10.5.6.5, Foreign port: 179

Current send queue size in bytes: 0 (max 24576)
Current receive queue size in bytes: 0 (max 32768)  mis-ordered: 0 bytes
Current receive queue size in packets: 0 (max 0)

Timer          Starts    Wakeups         Next(msec)
Retrans            17          1                0
SendWnd             0          0                0
TimeWait            0          0                0
AckHold            14         13                0
KeepAlive           1          0                0
PmtuAger            0          0                0
GiveUp              0          0                0
Throttle            0          0                0

   iss: 1379482495  snduna: 1379482819  sndnxt: 1379482819
sndmax: 1379482819  sndwnd: 32498       sndcwnd: 4308
```

```
   irs: 3750694052  rcvnxt: 3750694376  rcvwnd: 32517   rcvadv: 3750726893

SRTT: 55 ms,  RTTO: 300 ms,  RTV: 176 ms,  KRTT: 0 ms
minRTT: 9 ms,  maxRTT: 259 ms

ACK hold time: 200 ms, Keepalive time: 0 sec, SYN waittime: 30 sec
Giveup time: 0 ms, Retransmission retries: 0, Retransmit forever: FALSE
Connect retries remaining: 30, connect retry interval: 50 secs

State flags: none
Feature flags:
```

MD5

```
, Win Scale, Nagle
Request flags: Win Scale
```

Datagrams (in bytes): MSS 1436, peer MSS 1436, min MSS 1436, max MSS 1436

```
Window scales: rcv 0, snd 0, request rcv 0, request snd 0
Timestamp option: recent 0, recent age 0, last ACK sent 0
Sack blocks {start, end}: none
Sack holes {start, end, dups, rxmit}: none

Socket options: SO_REUSEADDR, SO_REUSEPORT, SO_NBIO
Socket states: SS_ISCONNECTED, SS_PRIV
Socket receive buffer states: SB_DEL_WAKEUP
Socket send buffer states: SB_DEL_WAKEUP
Socket receive buffer: Low/High watermark 1/32768
Socket send buffer   : Low/High watermark 2048/24576, Notify threshold 0

PDU information:
 #PDU's in buffer: 0
FIB Lookup Cache:  IFH: 0x40  PD ctx: size: 0  data:
  Num Labels: 0  Label Stack:

RP/0/0/CPU0:R6#
```

TCP会话详细信息（如R5上所示）— PASSIVE:

<#root>

```
! - as seen on R5 - Passive

RP/0/0/CPU0:R5#show tcp detail pcb 0x12155d04
Fri Jan  8 11:12:51.984 UTC
================================================================
Connection state is ESTAB, I/O status: 0, socket status: 0
Established at Fri Jan  8 11:01:22 2021

PCB 0x12155d04, SO 0x12154e18, TCPCB 0x12154304, vrfid 0x60000000,
Pak Prio: Medium, TOS: 192, TTL: 1, Hash index: 409
Local host: 10.5.6.5, Local port: 179 (Local App PID: 1044686)
Foreign host: 10.5.6.6, Foreign port: 24785

Current send queue size in bytes: 0 (max 24576)
```

```
Current receive queue size in bytes: 0 (max 32768)  mis-ordered: 0 bytes
Current receive queue size in packets: 0 (max 0)

Timer          Starts    Wakeups         Next(msec)
Retrans            14          0                0
SendWnd             0          0                0
TimeWait            0          0                0
AckHold            14          3                0
KeepAlive           1          0                0
PmtuAger            0          0                0
GiveUp              0          0                0
Throttle            0          0                0

iss: 3750694052  snduna: 3750694357  sndnxt: 3750694357
sndmax: 3750694357  sndwnd: 32536       sndcwnd: 4308
irs: 1379482495  rcvnxt: 1379482800  rcvwnd: 32517   rcvadv: 1379515317
SRTT: 181 ms,  RTTO: 443 ms,  RTV: 262 ms,  KRTT: 0 ms
minRTT: 29 ms,  maxRTT: 219 ms

ACK hold time: 200 ms, Keepalive time: 0 sec, SYN waittime: 30 sec
Giveup time: 0 ms, Retransmission retries: 0, Retransmit forever: FALSE
Connect retries remaining: 0, connect retry interval: 0 secs

State flags: none
Feature flags:
```

MD5

```
, Win Scale, Nagle
Request flags: Win Scale
```

Datagrams (in bytes): MSS 1436, peer MSS 1436, min MSS 1460, max MSS 1460

```
Window scales: rcv 0, snd 0, request rcv 0, request snd 0
Timestamp option: recent 0, recent age 0, last ACK sent 0
Sack blocks {start, end}: none
Sack holes {start, end, dups, rxmit}: none

Socket options: SO_REUSEADDR, SO_REUSEPORT, SO_NBIO
Socket states: SS_ISCONNECTED, SS_PRIV
Socket receive buffer states: SB_DEL_WAKEUP
Socket send buffer states: SB_DEL_WAKEUP
Socket receive buffer: Low/High watermark 1/32768
Socket send buffer   : Low/High watermark 2048/24576, Notify threshold 0

PDU information:
#PDU's in buffer: 0
FIB Lookup Cache:  IFH: 0x40  PD ctx: size: 0  data:
Num Labels: 0  Label Stack:

RP/0/0/CPU0:R5#
```

其他TCP选项也存在类似行为，这些选项在配置时会导致额外的开销并影响Cisco IOS XR中的MSS计算。请考虑配置TCP时间戳和TCP选择性确认选项时记录MSS计算的相同方案和这些示例。

TCP会话详细信息，如R6上所示 — ACTIVE — 已配置TCP选项时间戳和选择性确认选项：

<#root>

```
! - as seen on R6 – Active

! -- tcp timestamp configured


! -- 12 bytes of additional overhead

RP/0/0/CPU0:R6#show tcp detail pcb 0x1539c844
<snip>
Feature flags: Timestamp, Win Scale, Nagle
Request flags: Timestamp, Win Scale


Datagrams (in bytes): MSS 1448, peer MSS 1448, min MSS 1448, max MSS 1448


<snip>

! - as seen on R6 – Active

! -- tcp selective-ack configured


! -- 36 bytes of additional overhead

RP/0/0/CPU0:R6#show tcp detail pcb 0x1539df38
<snip>
Feature flags: Sack, Win Scale, Nagle
Request flags: Sack, Win Scale


Datagrams (in bytes): MSS 1424, peer MSS 1424, min MSS 1424, max MSS 1424


<snip>

! - as seen on R6 – Active

! -- tcp selective-ack and tcp timestamp configured


! -- 40 bytes of additional overhead

RP/0/0/CPU0:R6#show tcp detail pcb 0x1539e130
<snip>
State flags: none
Feature flags: Sack, Timestamp, Win Scale, Nagle
Request flags: Sack, Timestamp, Win Scale


Datagrams (in bytes): MSS 1420, peer MSS 1420, min MSS 1420, max MSS 1420


<snip>

! - as seen on R6 – Active

! -- MD5 and tcp selective-ack configured


! -- 36 bytes of additional overhead

RP/0/0/CPU0:R6#show tcp detail pcb 0x1539b3cc
```

```
<snip>
Feature flags: Sack, MD5, Win Scale, Nagle
Request flags: Sack, Win Scale


Datagrams (in bytes): MSS 1424, peer MSS 1424, min MSS 1424, max MSS 1424


<snip>

! - as seen on R6 – Active

! -- MD5 and tcp timestamp configured


! -- 36 bytes of additional overhead

RP/0/0/CPU0:R6#show tcp detail pcb 0x15397b4c
<snip>

Feature flags: MD5, Timestamp, Win Scale, Nagle
Request flags: Timestamp, Win Scale


Datagrams (in bytes): MSS 1424, peer MSS 1424, min MSS 1424, max MSS 1424


<snip>

! - as seen on R6 – Active

! -- MD5, tcp timestamp, and tcp selective-ack configured


! -- 40 bytes of additional overhead

RP/0/0/CPU0:R6#show tcp detail pcb 0x1539a4cc
<snip>
State flags: none
Feature flags: MD5, Timestamp, Win Scale, Nagle
Request flags: Timestamp, Win Scale


Datagrams (in bytes): MSS 1420, peer MSS 1420, min MSS 1420, max MSS 1420


<snip>
```

使用TCP选项 — XR被动

在前面的场景中，您可能已经注意到Cisco IOS XR节点在初始MSS计算方面处于被动角色时的独特行为。节点不考虑tcp选项报头长度。此场景旨在突出显示此不同行为，Cisco Bug ID也对此进行了描述：

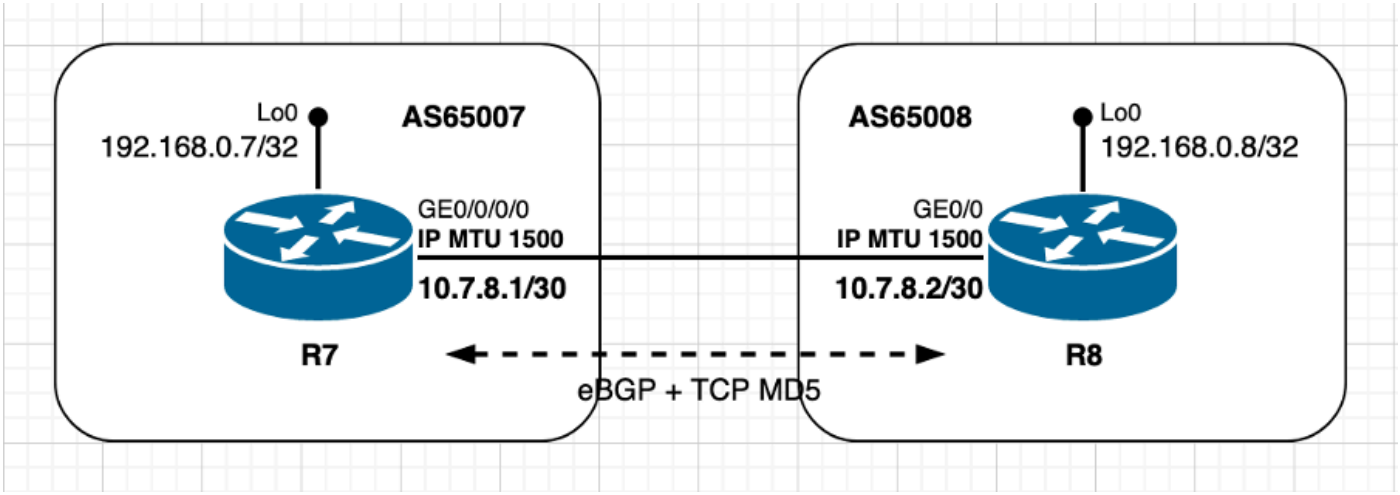
"(...) — 当对等体发起连接时，它发送初始MSS为1460。XR TCP创建套接字、pcb等，然后按指定顺序执行以下两个操作：

— 首先，它在减去tcp选项报头长度后计算初始MSS。这是“0”，因为MD5选项尚未从侦听套接字继承到此套接字。

— 然后，它会继承“MD5”和其他选项，这会使“选项报头字节长度”变为24。

因此，在本例中，XR TCP将初始MSS作为1460发送，因此供两者使用。(…)“

在此场景中，尽管活动TCP对等体R8是Cisco IOS节点，但此事实不会带来任何差异或场景要强调的具体内容。然而，有趣的是，请注意，与上一节场景所示的Cisco IOS XR不同，此处活动TCP对等体R8在初始MSS计算时不会考虑TCP选项。



图像2.5 — 使用TCP选项(MD5)- XR被动。

两个对等体都使用默认IP MTU值并直接连接。Cisco IOS对等体R8扮演主动角色。此场景中的TCP MSS计算可概述如下：

- 两个节点都使用默认IP MTU 1500字节
- 默认情况下，思科IOS XR R7上禁用TCP路径MTU发现
- 默认情况下，TCP路径MTU发现在Cisco IOS R8上启用
- TCP对等体直接连接
- 两个节点上都启用了TCP MD5身份验证
    - IOS R8管理BGP连接
    - IOS R8发送MSS为1460字节的SYN
        - 1500（接口IP MTU） — 20(minTCP_H)- 20(minIP_H)
    - IOS XR R7发送SYN、ACK，MSS为1460字节
        - 发送[Received MSS;本地初始MSS]
        - 接收的MSS 1460字节；本地初始MSS 1460字节
    - 两个对等体上均使用最低MSS值

源自R8的TCP SYN - Cisco IOS:

<#root>

```
! – TCP SYN sourced from R8

96      5.907127        10.7.8.2        10.7.8.1        TCP     78      52975 → 179 [SYN] Seq=0 Win=16384 Len=0

MSS=1460
```

```
Frame 96: 78 bytes on wire (624 bits), 78 bytes captured (624 bits) on interface 0
Ethernet II, Src: fa:16:3e:58:21:ba (fa:16:3e:58:21:ba), Dst: fa:16:3e:68:d9:e5 (fa:16:3e:68:d9:e5)
Internet Protocol Version 4, Src: 10.7.8.2, Dst: 10.7.8.1
Transmission Control Protocol, Src Port: 52975, Dst Port: 179, Seq: 0, Len: 0
    Source Port: 52975
    Destination Port: 179
    [Stream index: 3]
    [TCP Segment Len: 0]
    Sequence number: 0    (relative sequence number)
    Acknowledgment number: 0
    Header Length: 44 bytes
    Flags: 0x002 (SYN)
    Window size value: 16384
    [Calculated window size: 16384]
    Checksum: 0xb612 [unverified]
    [Checksum Status: Unverified]
    Urgent pointer: 0
    Options: (24 bytes), Maximum segment size,
```

TCP MD5 signature

```
, End of Option List (EOL)
        Maximum segment size: 1460 bytes
            Kind: Maximum Segment Size (2)
            Length: 4
```

MSS Value: 1460

```
        TCP MD5 signature
        End of Option List (EOL)
```

来自R7的TCP SYN、ACK - Cisco IOS XR:

<#root>

```
! – TCP SYN,ACK sourced from R7

97     0.003446       10.7.8.1       10.7.8.2       TCP    78     179 → 52975 [SYN, ACK] Seq=0 Ack=1 Win=16
```

MSS=1460

```
Frame 97: 78 bytes on wire (624 bits), 78 bytes captured (624 bits) on interface 0
Ethernet II, Src: fa:16:3e:68:d9:e5 (fa:16:3e:68:d9:e5), Dst: fa:16:3e:58:21:ba (fa:16:3e:58:21:ba)
Internet Protocol Version 4, Src: 10.7.8.1, Dst: 10.7.8.2
Transmission Control Protocol, Src Port: 179, Dst Port: 52975, Seq: 0, Ack: 1, Len: 0
    Source Port: 179
    Destination Port: 52975
    [Stream index: 3]
    [TCP Segment Len: 0]
    Sequence number: 0    (relative sequence number)
    Acknowledgment number: 1    (relative ack number)
    Header Length: 44 bytes
    Flags: 0x012 (SYN, ACK)
    Window size value: 16384
    [Calculated window size: 16384]
```

```
    Checksum: 0xfb47 [unverified]
    [Checksum Status: Unverified]
    Urgent pointer: 0
    Options: (24 bytes), Maximum segment size,
```

TCP MD5 signature

```
, End of Option List (EOL)
        Maximum segment size: 1460 bytes
            Kind: Maximum Segment Size (2)
            Length: 4
```

MSS Value: 1460

```
        TCP MD5 signature
        End of Option List (EOL)
```

TCP会话详细信息，如R8上所示 — Cisco IOS - ACTIVE:

<#root>

```
! - as seen from R8 - Cisco IOS

R8#show ip bgp neighbors
BGP neighbor is 10.7.8.1,  remote AS 65007, external link
  BGP version 4, remote router ID 192.168.0.7
  BGP state = Established, up for 00:06:12
  Last read 00:00:16, last write 00:00:16, hold time is 180, keepalive interval is 60 seconds
  Neighbor sessions:
    1 active, is not multisession capable (disabled)
  Neighbor capabilities:
    Route refresh: advertised and received(new)
    Four-octets ASN Capability: advertised and received
    Address family IPv4 Unicast: advertised and received
    Enhanced Refresh Capability: advertised
    Multisession Capability:
    Stateful switchover support enabled: NO for session 1
  Message statistics:
    InQ depth is 0
    OutQ depth is 0

                         Sent       Rcvd
    Opens:                  1          1
    Notifications:          0          0
    Updates:                1          1
    Keepalives:             7          7
    Route Refresh:          0          0
    Total:                  9          9
  Do log neighbor state changes (via global configuration)
  Default minimum time between advertisement runs is 30 seconds

 For address family: IPv4 Unicast
  Session: 10.7.8.1
  BGP table version 1, neighbor version 1/0
  Output queue size : 0
  Index 6, Advertise bit 0
  6 update-group member
```

```
  Slow-peer detection is disabled
  Slow-peer split-update-group dynamic is disabled
                                Sent       Rcvd
  Prefix activity:              ----       ----
    Prefixes Current:              0          0
    Prefixes Total:                0          0
    Implicit Withdraw:             0          0
    Explicit Withdraw:             0          0
    Used as bestpath:            n/a          0
    Used as multipath:           n/a          0
    Used as secondary:           n/a          0

                                  Outbound    Inbound
  Local Policy Denied Prefixes:   --------    -------
    Total:                               0          0
  Number of NLRIs in the update sent: max 0, min 0
  Last detected as dynamic slow peer: never
  Dynamic slow peer recovered: never
  Refresh Epoch: 1
  Last Sent Refresh Start-of-rib: never
  Last Sent Refresh End-of-rib: never
  Last Received Refresh Start-of-rib: never
  Last Received Refresh End-of-rib: never
                                     Sent       Rcvd
        Refresh activity:            ----       ----
          Refresh Start-of-RIB        0          0
          Refresh End-of-RIB          0          0

  Address tracking is enabled, the RIB does have a route to 10.7.8.1
  Connections established 6; dropped 5
  Last reset 00:06:18, due to BGP Notification received of session 1, Administrative Reset
  External BGP neighbor configured for connected checks (single-hop no-disable-connected-check)
  Interface associated: GigabitEthernet0/1 (peering address in same link)

 Transport(tcp) path-mtu-discovery is enabled


  Graceful-Restart is disabled
  SSO is disabled
Connection state is ESTAB, I/O status: 1, unread input bytes: 0
Connection is ECN Disabled, Mininum incoming TTL 0, Outgoing TTL 1
Local host: 10.7.8.2, Local port: 52975
Foreign host: 10.7.8.1, Foreign port: 179
Connection tableid (VRF): 0
Maximum output segment queue size: 50

Enqueued packets for retransmit: 0, input: 0  mis-ordered: 0 (0 bytes)

Event Timers (current time is 0x15DD97):
Timer          Starts    Wakeups            Next
Retrans            10          0             0x0
TimeWait            0          0             0x0
AckHold             9          5             0x0
SendWnd             0          0             0x0
KeepAlive           0          0             0x0
GiveUp              0          0             0x0
PmtuAger            1          0        0x195465
DeadWait            0          0             0x0
Linger              0          0             0x0
ProcessQ            0          0             0x0

iss: 1154289541  snduna: 1154289755  sndnxt: 1154289755
irs: 2149897425  rcvnxt: 2149897635
```

```
sndwnd:  32612  scale:      0  maxrcvwnd:  16384
rcvwnd:  16175  scale:      0  delrcvwnd:    209

SRTT: 737 ms, RTTO: 2506 ms, RTV: 1769 ms, KRTT: 0 ms
minRTT: 7 ms, maxRTT: 1000 ms, ACK hold: 200 ms
uptime: 372981 ms, Sent idletime: 16648 ms, Receive idletime: 16431 ms
Status Flags: active open
Option Flags: nagle, path mtu capable,
```

md5

```
IP Precedence value : 6
```

Datagrams (max data segment is 1460 bytes)

```
:
Rcvd: 18 (out of order: 0), with data: 8, total data bytes: 209
Sent: 16 (retransmit: 0, fastretransmit: 0, partialack: 0, Second Congestion: 0), with data: 9, total d

 Packets received in fast path: 0, fast processed: 0, slow path: 0
 fast lock acquisition failures: 0, slow path: 0
TCP Semaphore      0x0FBFA8A4  FREE

R8#
```

TCP会话详细信息，如R7所示 — Cisco IOS XR - PASSIVE:

<#root>

```
! - as seen from R7 - Cisco IOS XR

RP/0/0/CPU0:R7#show tcp detail pcb 0x12152e48
Wed Jan 13 13:03:43.363 UTC
------------------------------------------------------------
Connection state is ESTAB, I/O status: 0, socket status: 0
Established at Wed Jan 13 12:58:16 2021

PCB 0x12152e48, SO 0x1213c130, TCPCB 0x12156060, vrfid 0x60000000,
Pak Prio: Medium, TOS: 192, TTL: 1, Hash index: 947
Local host: 10.7.8.1, Local port: 179 (Local App PID: 983244)
Foreign host: 10.7.8.2, Foreign port: 52975

Current send queue size in bytes: 0 (max 24576)
Current receive queue size in bytes: 0 (max 32768)  mis-ordered: 0 bytes
Current receive queue size in packets: 0 (max 0)

Timer          Starts    Wakeups         Next(msec)
Retrans             8          0                 0
SendWnd             0          0                 0
TimeWait            0          0                 0
AckHold             8          7                 0
KeepAlive           1          0                 0
PmtuAger            0          0                 0
GiveUp              0          0                 0
Throttle            0          0                 0
```

```
   iss: 2149897425  snduna: 2149897616  sndnxt: 2149897616
sndmax: 2149897616  sndwnd: 16194       sndcwnd: 4380
   irs: 1154289541  rcvnxt: 1154289736  rcvwnd: 32631   rcvadv: 1154322367

SRTT: 125 ms,  RTTO: 552 ms,  RTV: 427 ms,  KRTT: 0 ms
minRTT: 19 ms,  maxRTT: 229 ms

ACK hold time: 200 ms, Keepalive time: 0 sec, SYN waittime: 30 sec
Giveup time: 0 ms, Retransmission retries: 0, Retransmit forever: FALSE
Connect retries remaining: 0, connect retry interval: 0 secs

State flags: none
Feature flags:
```

MD5

```
, Nagle
Request flags: none
```

Datagrams (in bytes): MSS 1460, peer MSS 1460, min MSS 1460, max MSS 1460

```
Window scales: rcv 0, snd 0, request rcv 0, request snd 0
Timestamp option: recent 0, recent age 0, last ACK sent 0
Sack blocks {start, end}: none
Sack holes {start, end, dups, rxmit}: none

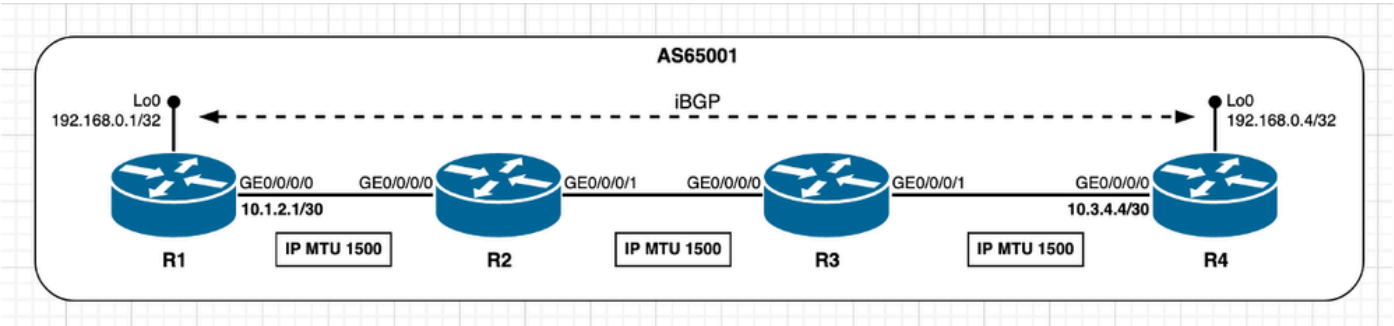
Socket options: SO_REUSEADDR, SO_REUSEPORT, SO_NBIO
Socket states: SS_ISCONNECTED, SS_PRIV
Socket receive buffer states: SB_DEL_WAKEUP
Socket send buffer states: SB_DEL_WAKEUP
Socket receive buffer: Low/High watermark 1/32768
Socket send buffer   : Low/High watermark 2048/24576, Notify threshold 0

PDU information:
 #PDU's in buffer: 0
FIB Lookup Cache:  IFH: 0x40  PD ctx: size: 0  data:
  Num Labels: 0  Label Stack:

RP/0/0/CPU0:R7#
```

未直接连接的TCP对等体

当对等体未直接连接时，TCP MSS初始计算的方式将更改，如本文档的介绍部分中所述。使用配置了默认IP MTU值的所有对等体的iBGP会话场景来演练MSS计算。



映像2.6 — 未直接连接的TCP对等体 — iBGP。

需要注意的重要方面是，当禁用TCP路径MTU发现且对等体未直接连接时，根据设计，Cisco IOS XR使用固定IP MTU值1280字节。

在上一张图片中，R4扮演活动角色并管理TCP连接，R4在目的端口179上打开与R1的TCP会话。两个节点在其接口上使用默认IP MTU值。此场景中的MSS计算可概述如下：

- 所有节点都使用默认IP MTU 1500字节
- 默认情况下禁用TCP路径MTU发现
- TCP对等体未直接连接
    - R4管理BGP连接
    - R4发送MSS为1240字节的SYN
        - 当对等体未直接连接且禁用TCP路径MTU发现时，不考虑接口MTU
        - 根据Cisco IOS XR设计，1280字节被视为TCP_DEFAULT_MTU
        - 1280(TCP_DEFAULT_MTU)- 20(minTCP_H)- 20(minIP_H)
    - R1发送SYN、MSS为1240字节的ACK
        - 发送[Received MSS;本地初始MSS]
        - 接收的MSS 1240字节；本地初始MSS 1240字节
    - 两个对等体上均使用最低MSS值

源自R4的TCP SYN:

<#root>

```
! - TCP SYN sourced from R4

194    434.274181     192.168.0.4 192.168.0.1 TCP    62     37740 → 179 [SYN] Seq=0 Win=16384 Len=0

MSS=1240

 WS=1

Frame 194: 62 bytes on wire (496 bits), 62 bytes captured (496 bits) on interface 0
Ethernet II, Src: fa:16:3e:d7:7e:f6 (fa:16:3e:d7:7e:f6), Dst: fa:16:3e:8f:8f:54 (fa:16:3e:8f:8f:54)
Internet Protocol Version 4, Src: 192.168.0.4, Dst: 192.168.0.1
Transmission Control Protocol, Src Port: 37740, Dst Port: 179, Seq: 0, Len: 0
    Source Port: 37740
    Destination Port: 179
    [Stream index: 7]
    [TCP Segment Len: 0]
    Sequence number: 0    (relative sequence number)
    Acknowledgment number: 0
    Header Length: 28 bytes
    Flags: 0x002 (SYN)
    Window size value: 16384
    [Calculated window size: 16384]
    Checksum: 0x8643 [unverified]
    [Checksum Status: Unverified]
    Urgent pointer: 0
    Options: (8 bytes), Maximum segment size, Window scale, End of Option List (EOL)
        Maximum segment size: 1240 bytes
            Kind: Maximum Segment Size (2)
            Length: 4
```

MSS Value: 1240

```
        Window scale: 0 (multiply by 1)
        End of Option List (EOL)
```

来自R1的TCP SYN、ACK:

<#root>

```
! - TCP SYN,ACK sourced from R1

195    434.277985     192.168.0.1 192.168.0.4 TCP    62     179 → 37740 [SYN, ACK] Seq=0 Ack=1 Win=16384 Le
```

MSS=1240

```
 WS=1

Frame 195: 62 bytes on wire (496 bits), 62 bytes captured (496 bits) on interface 0
Ethernet II, Src: fa:16:3e:8f:8f:54 (fa:16:3e:8f:8f:54), Dst: fa:16:3e:d7:7e:f6 (fa:16:3e:d7:7e:f6)
Internet Protocol Version 4, Src: 192.168.0.1, Dst: 192.168.0.4
Transmission Control Protocol, Src Port: 179, Dst Port: 37740, Seq: 0, Ack: 1, Len: 0
    Source Port: 179
    Destination Port: 37740
    [Stream index: 7]
    [TCP Segment Len: 0]
    Sequence number: 0    (relative sequence number)
    Acknowledgment number: 1    (relative ack number)
    Header Length: 28 bytes
    Flags: 0x012 (SYN, ACK)
    Window size value: 16384
    [Calculated window size: 16384]
    Checksum: 0xd8f7 [unverified]
    [Checksum Status: Unverified]
    Urgent pointer: 0
    Options: (8 bytes), Maximum segment size, Window scale, End of Option List (EOL)
        Maximum segment size: 1240 bytes
            Kind: Maximum Segment Size (2)
            Length: 4
```

MSS Value: 1240

```
        Window scale: 0 (multiply by 1)
        End of Option List (EOL)
```

TCP会话详细信息（如R4上所示）— ACTIVE:

<#root>

```
! - as seen on R4 - Active

RP/0/0/CPU0:R4#show tcp detail pcb 0x12154d3c
Fri Jan  8 12:32:41.096 UTC
================================================================
```

```
Connection state is ESTAB, I/O status: 0, socket status: 0
Established at Fri Jan  8 12:17:46 2021

PCB 0x12154d3c, SO 0x12154460, TCPCB 0x1215486c, vrfid 0x60000000,
Pak Prio: Medium, TOS: 192, TTL: 255, Hash index: 1577
Local host: 192.168.0.4, Local port: 37740 (Local App PID: 1052958)
Foreign host: 192.168.0.1, Foreign port: 179

Current send queue size in bytes: 0 (max 24576)
Current receive queue size in bytes: 0 (max 32768)  mis-ordered: 0 bytes
Current receive queue size in packets: 0 (max 0)

Timer          Starts    Wakeups         Next(msec)
Retrans            19          1                0
SendWnd             0          0                0
TimeWait            0          0                0
AckHold            16         15                0
KeepAlive           1          0                0
PmtuAger            0          0                0
GiveUp              0          0                0
Throttle            0          0                0

   iss: 2075436506  snduna: 2075436868  sndnxt: 2075436868
sndmax: 2075436868  sndwnd: 32460       sndcwnd: 3720
   irs: 4238127261  rcvnxt: 4238127623  rcvwnd: 32479   rcvadv: 4238160102

SRTT: 65 ms,  RTTO: 300 ms,  RTV: 40 ms,  KRTT: 0 ms
minRTT: 9 ms,  maxRTT: 229 ms

ACK hold time: 200 ms, Keepalive time: 0 sec, SYN waittime: 30 sec
Giveup time: 0 ms, Retransmission retries: 0, Retransmit forever: FALSE
Connect retries remaining: 30, connect retry interval: 30 secs

State flags: none
Feature flags: Win Scale, Nagle
Request flags: Win Scale


Datagrams (in bytes): MSS 1240, peer MSS 1240, min MSS 1240, max MSS 1240



Window scales: rcv 0, snd 0, request rcv 0, request snd 0
Timestamp option: recent 0, recent age 0, last ACK sent 0
Sack blocks {start, end}: none
Sack holes {start, end, dups, rxmit}: none

Socket options: SO_REUSEADDR, SO_REUSEPORT, SO_NBIO
Socket states: SS_ISCONNECTED, SS_PRIV
Socket receive buffer states: SB_DEL_WAKEUP
Socket send buffer states: SB_DEL_WAKEUP
Socket receive buffer: Low/High watermark 1/32768
Socket send buffer   : Low/High watermark 2048/24576, Notify threshold 0

PDU information:
 #PDU's in buffer: 0
FIB Lookup Cache:  IFH: 0x40  PD ctx: size: 0  data:
  Num Labels: 0  Label Stack:

RP/0/0/CPU0:R4#
```

R1上看到的TCP会话详细信息 — PASSIVE:

<#root>

```
! - as seen on R1 - Passive

RP/0/0/CPU0:R1#show tcp detail pcb 0x12155390
Fri Jan  8 12:23:52.041 UTC
================================================================
Connection state is ESTAB, I/O status: 0, socket status: 0
Established at Fri Jan  8 12:17:43 2021

PCB 0x12155390, SO 0x121573e4, TCPCB 0x12156948, vrfid 0x60000000,
Pak Prio: Medium, TOS: 192, TTL: 255, Hash index: 1577
Local host: 192.168.0.1, Local port: 179 (Local App PID: 983326)
Foreign host: 192.168.0.4, Foreign port: 37740

Current send queue size in bytes: 0 (max 24576)
Current receive queue size in bytes: 0 (max 32768)  mis-ordered: 0 bytes
Current receive queue size in packets: 0 (max 0)

Timer          Starts    Wakeups         Next(msec)
Retrans             9          0                0
SendWnd             0          0                0
TimeWait            0          0                0
AckHold             9          1                0
KeepAlive           1          0                0
PmtuAger            0          0                0
GiveUp              0          0                0
Throttle            0          0                0

iss: 4238127261  snduna: 4238127471  sndnxt: 4238127471
sndmax: 4238127471  sndwnd: 32631       sndcwnd: 3720
irs: 2075436506  rcvnxt: 2075436716  rcvwnd: 32612   rcvadv: 2075469328

SRTT: 144 ms,  RTTO: 578 ms,  RTV: 434 ms,  KRTT: 0 ms
minRTT: 19 ms,  maxRTT: 239 ms

ACK hold time: 200 ms, Keepalive time: 0 sec, SYN waittime: 30 sec
Giveup time: 0 ms, Retransmission retries: 0, Retransmit forever: FALSE
Connect retries remaining: 0, connect retry interval: 0 secs

State flags: none
Feature flags: Win Scale, Nagle
Request flags: Win Scale


Datagrams (in bytes): MSS 1240, peer MSS 1240, min MSS 1240, max MSS 1240



Window scales: rcv 0, snd 0, request rcv 0, request snd 0
Timestamp option: recent 0, recent age 0, last ACK sent 0
Sack blocks {start, end}: none
Sack holes {start, end, dups, rxmit}: none

Socket options: SO_REUSEADDR, SO_REUSEPORT, SO_NBIO
Socket states: SS_ISCONNECTED, SS_PRIV
Socket receive buffer states: SB_DEL_WAKEUP
Socket send buffer states: SB_DEL_WAKEUP
Socket receive buffer: Low/High watermark 1/32768
```

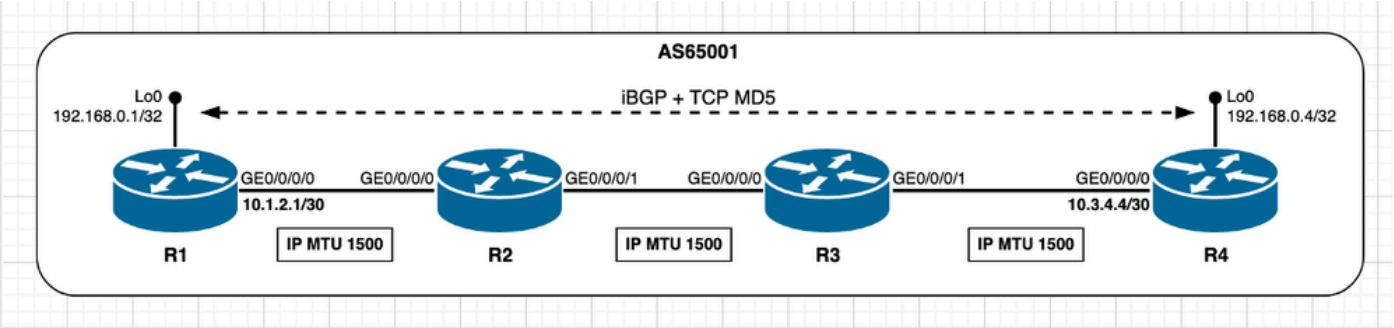
```
Socket send buffer   : Low/High watermark 2048/24576, Notify threshold 0

PDU information:
#PDU's in buffer: 0
FIB Lookup Cache:  IFH: 0x40  PD ctx: size: 0  data:
Num Labels: 0  Label Stack:

RP/0/0/CPU0:R1#
```

未直接连接的TCP对等体 — 使用TCP选项(MD5)

对于非直连对等体场景和使用TCP MD5身份验证的情况，与之前的测试案例或已经描述的场景没有根本性差异。如前所述，TCP MD5身份验证中，Cisco IOS XR考虑额外开销，初始MSS值反映相同。有关TCP选项对TCP MSS计算的影响的其他详细信息，请参阅前面的使用TCP选项 — XR主动和使用TCP选项 — XR被动部分。



映像2.7 — 未直接连接的TCP对等体 — iBGP + TCP MD5。

此场景中的TCP MSS计算可概述如下：

- 所有节点都使用默认IP MTU 1500字节
- 默认情况下禁用TCP路径MTU发现
- TCP对等体未直接连接
    - R4管理BGP连接
    - 目标R1未直接连接
    - R4发送MSS为1216字节的SYN
        - 当对等体未直接连接且禁用TCP路径MTU发现时，不考虑接口MTU
        - 根据设计，1280字节被视为TCP_DEFAULT_MTU
        - 1280(TCP_DEFAULT_MTU)- 20(minTCP_H)- 20(minIP_H)- 24字节（IOS XR TCP选项开销）
    - R1发送SYN、ACK，MSS为1216字节
        - 发送[Received MSS;本地初始MSS]
        - 接收的MSS 1216字节；本地初始MSS 1240字节
    - 两个对等体上均使用最低MSS值

源自R4的TCP SYN:

```
! - TCP SYN sourced from R4

3425   3.691042       192.168.0.4 192.168.0.1 TCP    82     42135 → 179 [SYN] Seq=0 Win=16384 Len=0
```

MSS=1216

```
 WS=1

Frame 3425: 82 bytes on wire (656 bits), 82 bytes captured (656 bits) on interface 0
Ethernet II, Src: fa:16:3e:d7:7e:f6 (fa:16:3e:d7:7e:f6), Dst: fa:16:3e:8f:8f:54 (fa:16:3e:8f:8f:54)
Internet Protocol Version 4, Src: 192.168.0.4, Dst: 192.168.0.1
Transmission Control Protocol, Src Port: 42135, Dst Port: 179, Seq: 0, Len: 0
    Source Port: 42135
    Destination Port: 179
    [Stream index: 10]
    [TCP Segment Len: 0]
    Sequence number: 0    (relative sequence number)
    Acknowledgment number: 0
    Header Length: 48 bytes
    Flags: 0x002 (SYN)
    Window size value: 16384
    [Calculated window size: 16384]
    Checksum: 0xc503 [unverified]
    [Checksum Status: Unverified]
    Urgent pointer: 0
    Options: (28 bytes), Maximum segment size, Window scale, No-Operation (NOP),
```

TCP MD5 signature

```
, End of Option List (EOL)
        Maximum segment size: 1216 bytes
            Kind: Maximum Segment Size (2)
            Length: 4
```

MSS Value: 1216

```
        Window scale: 0 (multiply by 1)
        No-Operation (NOP)
        TCP MD5 signature
        End of Option List (EOL)
```

来自R1的TCP SYN、ACK:

<#root>

```
! - TCP SYN,ACK sourced from R1

3426   0.004186       192.168.0.1 192.168.0.4 TCP    82     179 → 42135 [SYN, ACK] Seq=0 Ack=1 Win=16384 Le
```

MSS=1216

```
 WS=1

Frame 3426: 82 bytes on wire (656 bits), 82 bytes captured (656 bits) on interface 0
Ethernet II, Src: fa:16:3e:8f:8f:54 (fa:16:3e:8f:8f:54), Dst: fa:16:3e:d7:7e:f6 (fa:16:3e:d7:7e:f6)
Internet Protocol Version 4, Src: 192.168.0.1, Dst: 192.168.0.4
Transmission Control Protocol, Src Port: 179, Dst Port: 42135, Seq: 0, Ack: 1, Len: 0
    Source Port: 179
```

```
    Destination Port: 42135
    [Stream index: 10]
    [TCP Segment Len: 0]
    Sequence number: 0    (relative sequence number)
    Acknowledgment number: 1    (relative ack number)
    Header Length: 48 bytes
    Flags: 0x012 (SYN, ACK)
    Window size value: 16384
    [Calculated window size: 16384]
    Checksum: 0xbb05 [unverified]
    [Checksum Status: Unverified]
    Urgent pointer: 0
    Options: (28 bytes), Maximum segment size, Window scale, No-Operation (NOP),
```

TCP MD5 signature

```
, End of Option List (EOL)
        Maximum segment size: 1216 bytes
            Kind: Maximum Segment Size (2)
            Length: 4
```

MSS Value: 1216

```
        Window scale: 0 (multiply by 1)
        No-Operation (NOP)
        TCP MD5 signature
        End of Option List (EOL)
```

TCP会话详细信息（如R4上所示）— ACTIVE:

<#root>

```
! - as seen from R4 - Active

RP/0/0/CPU0:R4#show tcp detail pcb 0x12154490
Tue Jan 12 14:37:32.097 UTC
==============================================================
Connection state is ESTAB, I/O status: 0, socket status: 0
Established at Tue Jan 12 14:27:42 2021

PCB 0x12154490, SO 0x12155014, TCPCB 0x12155a84, vrfid 0x60000000,
Pak Prio: Medium, TOS: 192, TTL: 255, Hash index: 1876
Local host: 192.168.0.4, Local port: 42135 (Local App PID: 1052958)
Foreign host: 192.168.0.1, Foreign port: 179

Current send queue size in bytes: 0 (max 24576)
Current receive queue size in bytes: 0 (max 32768)  mis-ordered: 0 bytes
Current receive queue size in packets: 0 (max 0)

Timer          Starts    Wakeups         Next(msec)
Retrans            14          1                  0
SendWnd             0          0                  0
TimeWait            0          0                  0
AckHold            11          9                  0
KeepAlive           1          0                  0
PmtuAger            0          0                  0
GiveUp              0          0                  0
```

```
Throttle            0          0                0


   iss: 3124761989  snduna: 3124763317  sndnxt: 3124763317
sndmax: 3124763317  sndwnd: 32711       sndcwnd: 3648
   irs: 1090344992  rcvnxt: 1090346320  rcvwnd: 32730   rcvadv: 1090379050

SRTT: 28 ms,  RTTO: 300 ms,  RTV: 57 ms,  KRTT: 0 ms
minRTT: 9 ms,  maxRTT: 229 ms

ACK hold time: 200 ms, Keepalive time: 0 sec, SYN waittime: 30 sec
Giveup time: 0 ms, Retransmission retries: 0, Retransmit forever: FALSE
Connect retries remaining: 30, connect retry interval: 30 secs

State flags: none
Feature flags: MD5, Win Scale, Nagle
Request flags: Win Scale


Datagrams (in bytes): MSS 1216, peer MSS 1216, min MSS 1216, max MSS 1216



Window scales: rcv 0, snd 0, request rcv 0, request snd 0
Timestamp option: recent 0, recent age 0, last ACK sent 0
Sack blocks {start, end}: none
Sack holes {start, end, dups, rxmit}: none

Socket options: SO_REUSEADDR, SO_REUSEPORT, SO_NBIO
Socket states: SS_ISCONNECTED, SS_PRIV
Socket receive buffer states: SB_DEL_WAKEUP
Socket send buffer states: SB_DEL_WAKEUP
Socket receive buffer: Low/High watermark 1/32768
Socket send buffer   : Low/High watermark 2048/24576, Notify threshold 0

PDU information:
 #PDU's in buffer: 0
FIB Lookup Cache:  IFH: 0x40  PD ctx: size: 0  data:
  Num Labels: 0  Label Stack:

RP/0/0/CPU0:R4#
```

R1上看到的TCP会话详细信息 — PASSIVE:

<#root>

```
! - as seen from R1 - Passive

RP/0/0/CPU0:R1#show tcp detail pcb 0x12168df4
Tue Jan 12 14:36:38.860 UTC
==============================================================
Connection state is ESTAB, I/O status: 0, socket status: 0
Established at Tue Jan 12 14:27:32 2021

PCB 0x12168df4, SO 0x12156bf8, TCPCB 0x12157a44, vrfid 0x60000000,
Pak Prio: Medium, TOS: 192, TTL: 255, Hash index: 1876
Local host: 192.168.0.1, Local port: 179 (Local App PID: 983326)
Foreign host: 192.168.0.4, Foreign port: 42135
```

```
Current send queue size in bytes: 0 (max 24576)
Current receive queue size in bytes: 0 (max 32768)  mis-ordered: 0 bytes
Current receive queue size in packets: 0 (max 0)

Timer          Starts    Wakeups         Next(msec)
Retrans            12          0                0
SendWnd             0          0                0
TimeWait            0          0                0
AckHold            12          1                0
KeepAlive           1          0                0
PmtuAger            0          0                0
GiveUp              0          0                0
Throttle            0          0                0

   iss: 1090344992  snduna: 1090346320  sndnxt: 1090346320
sndmax: 1090346320  sndwnd: 32730       sndcwnd: 3648
   irs: 3124761989  rcvnxt: 3124763317  rcvwnd: 32711   rcvadv: 3124796028

SRTT: 150 ms,  RTTO: 558 ms,  RTV: 408 ms,  KRTT: 0 ms
minRTT: 19 ms,  maxRTT: 239 ms

ACK hold time: 200 ms, Keepalive time: 0 sec, SYN waittime: 30 sec
Giveup time: 0 ms, Retransmission retries: 0, Retransmit forever: FALSE
Connect retries remaining: 0, connect retry interval: 0 secs

State flags: none
Feature flags: MD5, Win Scale, Nagle
Request flags: Win Scale


Datagrams (in bytes): MSS 1216, peer MSS 1216, min MSS 1240, max MSS 1240



Window scales: rcv 0, snd 0, request rcv 0, request snd 0
Timestamp option: recent 0, recent age 0, last ACK sent 0
Sack blocks {start, end}: none
Sack holes {start, end, dups, rxmit}: none

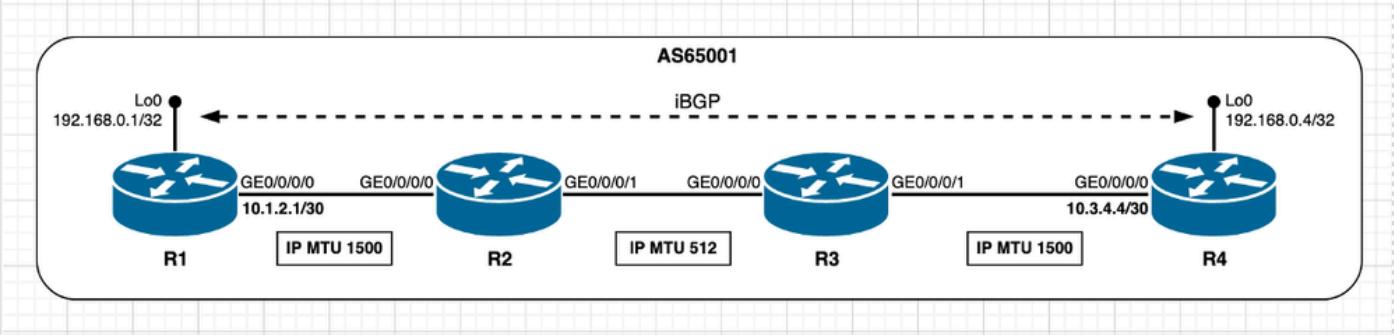
Socket options: SO_REUSEADDR, SO_REUSEPORT, SO_NBIO
Socket states: SS_ISCONNECTED, SS_PRIV
Socket receive buffer states: SB_DEL_WAKEUP
Socket send buffer states: SB_DEL_WAKEUP
Socket receive buffer: Low/High watermark 1/32768
Socket send buffer   : Low/High watermark 2048/24576, Notify threshold 0

PDU information:
 #PDU's in buffer: 0
FIB Lookup Cache:  IFH: 0x40  PD ctx: size: 0  data:
  Num Labels: 0  Label Stack:

RP/0/0/CPU0:R1#
```

未直接连接的TCP对等体 — 路径段的IP MTU较低

在下一个场景中，目的是观察并总结在默认情况下中间路径网段具有更低的IP MTU时会发生的情况，这意味着禁用TCP PMTUD。请参考此图。



映像2.8 - R2/R3路径段的IP MTU较低。

作为初始方案，认为BGP信息是最小的，也就是说，BGP对等体之间需要交换的任何内容都可以通过适合在512字节的最小路径MTU以下的IP数据包来完成。在此假设下，MSS计算按照TCP对等体未直接连接一节中所述的方式进行。R1和R4都选择1240字节的MSS值。

TCP会话详细信息（如R4上所示）— ACTIVE:

<#root>

```
! - as seen from R4 - Active

RP/0/0/CPU0:R4#show tcp detail pcb 0x15390fe8
==============================================================
Connection state is ESTAB, I/O status: 0, socket status: 0
Established at Wed May 12 12:09:48 2021

PCB 0x15390fe8, SO 0x15391a7c, TCPCB 0x15391368, vrfid 0x60000000,
Pak Prio: Medium, TOS: 192, TTL: 255, Hash index: 835
Local host: 192.168.0.4, Local port: 39046 (Local App PID: 1196319)
Foreign host: 192.168.0.1, Foreign port: 179
(Local App PID/instance/SPL_APP_ID: 1196319/1/0)

Current send queue size in bytes: 0 (max 24576)
Current receive queue size in bytes: 0 (max 32768)  mis-ordered: 0 bytes
Current receive queue size in packets: 0 (max 0)

Timer          Starts    Wakeups         Next(msec)
Retrans          1267          1                 0
SendWnd             0          0                 0
TimeWait            0          0                 0
AckHold          1280       1235                 0
KeepAlive           1          0                 0
PmtuAger            0          0                 0
GiveUp              0          0                 0
Throttle            0          0                 0

   iss: 1991226354  snduna: 1991250450  sndnxt: 1991250450
sndmax: 1991250450  sndwnd: 32578       sndcwnd: 2480
   irs: 4276699304  rcvnxt: 4276746737  rcvwnd: 31568   rcvadv: 4276778305

SRTT: 213 ms,  RTTO: 300 ms,  RTV: 54 ms,  KRTT: 0 ms
minRTT: 9 ms,  maxRTT: 269 ms

ACK hold time: 200 ms, Keepalive time: 0 sec, SYN waittime: 30 sec
Giveup time: 0 ms, Retransmission retries: 0, Retransmit forever: FALSE
Connect retries remaining: 10, connect retry interval: 30 secs
```

```
State flags: none
Feature flags: Win Scale, Nagle
Request flags: Win Scale


Datagrams (in bytes): MSS 1240, peer MSS 1240, min MSS 1240, max MSS 1240


<snip>
```

R1上看到的TCP会话详细信息 — PASSIVE:

<#root>

```
! - as seen from R1 - Passive

RP/0/0/CPU0:R1#show tcp detail pcb 0x15393770
==============================================================
Connection state is ESTAB, I/O status: 0, socket status: 0
Established at Wed May 12 12:09:46 2021

PCB 0x15393770, SO 0x15392224, TCPCB 0x153928cc, vrfid 0x60000000,
Pak Prio: Medium, TOS: 192, TTL: 255, Hash index: 835
Local host: 192.168.0.1, Local port: 179 (Local App PID: 1192224)
Foreign host: 192.168.0.4, Foreign port: 39046
(Local App PID/instance/SPL_APP_ID: 1192224/1/0)

Current send queue size in bytes: 0 (max 24576)
Current receive queue size in bytes: 0 (max 32768)  mis-ordered: 0 bytes
Current receive queue size in packets: 0 (max 0)

Timer          Starts    Wakeups         Next(msec)
Retrans          1280          0                0
SendWnd             0          0                0
TimeWait            0          0                0
AckHold          1264       1213                0
KeepAlive           1          0                0
PmtuAger            0          0                0
GiveUp              0          0                0
Throttle            0          0                0

   iss: 4276699304  snduna: 4276746718  sndnxt: 4276746718
sndmax: 4276746718  sndwnd: 31587       sndcwnd: 3720
   irs: 1991226354  rcvnxt: 1991250431  rcvwnd: 32597   rcvadv: 1991283028

SRTT: 202 ms,  RTTO: 355 ms,  RTV: 153 ms,  KRTT: 0 ms
minRTT: 9 ms,  maxRTT: 309 ms

ACK hold time: 200 ms, Keepalive time: 0 sec, SYN waittime: 30 sec
Giveup time: 0 ms, Retransmission retries: 0, Retransmit forever: FALSE
Connect retries remaining: 0, connect retry interval: 0 secs

State flags: none
Feature flags: Win Scale, Nagle
Request flags: Win Scale


Datagrams (in bytes): MSS 1240, peer MSS 1240, min MSS 1240, max MSS 1240
```

<snip>

现在建立BGP会话后，请考虑触发大小大于最小路径MTU 512字节的BGP更新消息。从输出中可以看到，Cisco IOS XR没有使用BGP更新消息设置df-bit，这意味着BGP信息的传输是以牺牲中间节点上的数据包分段为代价的。

由R1发起的BGP更新 — 被动：

<#root>

```
! - as seen from R1 - Passive - BGP UPDATE
! - Note Total Length of 1097 bytes higher than the IP MTU value of 512 bytes at R2-R3 path segment

23     3.450878       192.168.0.1 192.168.0.4 BGP    1111   UPDATE Message

Frame 23: 1111 bytes on wire (8888 bits), 1111 bytes captured (8888 bits) on interface 0
Ethernet II, Src: fa:16:3e:42:18:05 (fa:16:3e:42:18:05), Dst: fa:16:3e:5c:f1:80 (fa:16:3e:5c:f1:80)
Internet Protocol Version 4, Src: 192.168.0.1, Dst: 192.168.0.4
    0100 .... = Version: 4
    .... 0101 = Header Length: 20 bytes (5)
    Differentiated Services Field: 0xc0 (DSCP: CS6, ECN: Not-ECT)


Total Length: 1097


    Identification: 0x5841 (22593)
    Flags: 0x00
        0... .... = Reserved bit: Not set
        .0.. .... = Don't fragment: Not set
        ..0. .... = More fragments: Not set
    Fragment offset: 0
    Time to live: 255
    Protocol: TCP (6)
    Header checksum: 0x54a4 [validation disabled]
    [Header checksum status: Unverified]
    Source: 192.168.0.1
    Destination: 192.168.0.4
    [Source GeoIP: Unknown]
    [Destination GeoIP: Unknown]
Transmission Control Protocol, Src Port: 179, Dst Port: 39046, Seq: 20, Ack: 20, Len: 1057
Border Gateway Protocol - UPDATE Message
    Marker: ffffffffffffffffffffffffffffffff
    Length: 1057
    Type: UPDATE Message (2)
    Withdrawn Routes Length: 0
    Total Path Attribute Length: 1034
    Path attributes
        Path Attribute - MP_REACH_NLRI
        Path Attribute - ORIGIN: INCOMPLETE
        Path Attribute - AS_PATH: empty
        Path Attribute - MULTI_EXIT_DISC: 0
        Path Attribute - LOCAL_PREF: 100
```

从节点R1发出的BGP更新消息的分段发生在节点R2，这可以通过R2接口GE0/0/0/1执行的流量捕获来观察到。

节点R2上的IP分段：

<#root>

```
! - as seen from R2 - GE0/0/0/1
! - Node R2 fragments original packet in three distinct packets

4      1.334852       192.168.0.1 192.168.0.4 BGP    522    UPDATE Message
5      0.000289       192.168.0.1 192.168.0.4 IPv4   522    Fragmented IP protocol (proto=TCP 6, off=48
6      0.000122       192.168.0.1 192.168.0.4 IPv4   135    Fragmented IP protocol (proto=TCP 6, off=97

! - Captured frame details

Frame 4: 522 bytes on wire (4176 bits), 522 bytes captured (4176 bits) on interface 0
Ethernet II, Src: fa:16:3e:61:25:f0 (fa:16:3e:61:25:f0), Dst: fa:16:3e:23:ab:27 (fa:16:3e:23:ab:27)
Internet Protocol Version 4, Src: 192.168.0.1, Dst: 192.168.0.4
    0100 .... = Version: 4
    .... 0101 = Header Length: 20 bytes (5)
    Differentiated Services Field: 0xc0 (DSCP: CS6, ECN: Not-ECT)


Total Length: 508



Identification: 0x7b41 (31553)


    Flags: 0x01 (More Fragments)
        0... .... = Reserved bit: Not set
        .0.. .... = Don't fragment: Not set


..1. .... = More fragments: Set



Fragment offset: 0


    Time to live: 254
    Protocol: TCP (6)
    Header checksum: 0x14f1 [validation disabled]
    [Header checksum status: Unverified]
    Source: 192.168.0.1
    Destination: 192.168.0.4
    [Source GeoIP: Unknown]
    [Destination GeoIP: Unknown]
Transmission Control Protocol, Src Port: 179, Dst Port: 39046, Seq: 4276759681, Ack: 1991250830
Border Gateway Protocol - UPDATE Message
<snip>

Frame 5: 522 bytes on wire (4176 bits), 522 bytes captured (4176 bits) on interface 0
Ethernet II, Src: fa:16:3e:61:25:f0 (fa:16:3e:61:25:f0), Dst: fa:16:3e:23:ab:27 (fa:16:3e:23:ab:27)
Internet Protocol Version 4, Src: 192.168.0.1, Dst: 192.168.0.4
    0100 .... = Version: 4
    .... 0101 = Header Length: 20 bytes (5)
```

```
    Differentiated Services Field: 0xc0 (DSCP: CS6, ECN: Not-ECT)
```

Total Length: 508

Identification: 0x7b41 (31553)

```
    Flags: 0x01 (More Fragments)
        0... .... = Reserved bit: Not set
        .0.. .... = Don't fragment: Not set
```

..1. .... = More fragments: Set

Fragment offset: 488

```
    Time to live: 254
    Protocol: TCP (6)
    Header checksum: 0x14b4 [validation disabled]
    [Header checksum status: Unverified]
    Source: 192.168.0.1
    Destination: 192.168.0.4
    [Source GeoIP: Unknown]
    [Destination GeoIP: Unknown]
Data (488 bytes)
<snip>

Frame 6: 135 bytes on wire (1080 bits), 135 bytes captured (1080 bits) on interface 0
Ethernet II, Src: fa:16:3e:61:25:f0 (fa:16:3e:61:25:f0), Dst: fa:16:3e:23:ab:27 (fa:16:3e:23:ab:27)
Internet Protocol Version 4, Src: 192.168.0.1, Dst: 192.168.0.4
    0100 .... = Version: 4
    .... 0101 = Header Length: 20 bytes (5)
    Differentiated Services Field: 0xc0 (DSCP: CS6, ECN: Not-ECT)
```

Total Length: 121

Identification: 0x7b41 (31553)

```
    Flags: 0x00
        0... .... = Reserved bit: Not set
        .0.. .... = Don't fragment: Not set
```

..0. .... = More fragments: Not set

Fragment offset: 976

```
    Time to live: 254
    Protocol: TCP (6)
```

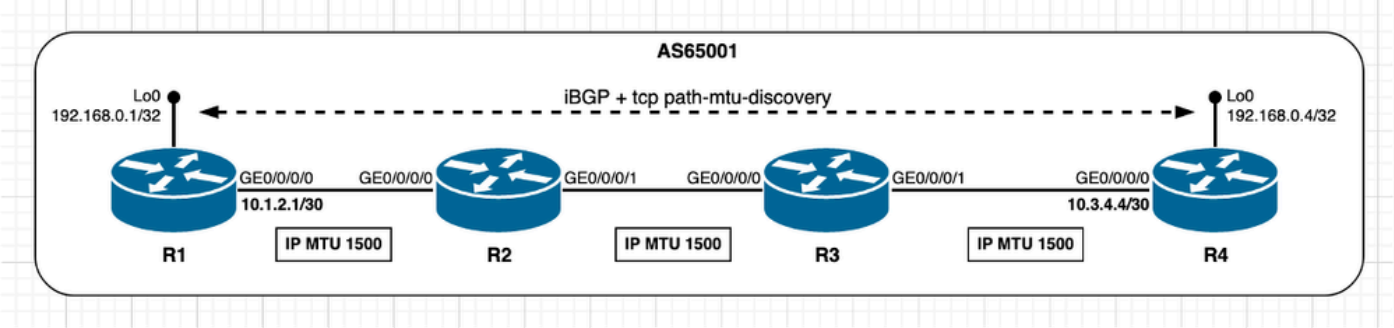
```
    Header checksum: 0x35fa [validation disabled]
    [Header checksum status: Unverified]
    Source: 192.168.0.1
    Destination: 192.168.0.4
    [Source GeoIP: Unknown]
    [Destination GeoIP: Unknown]
Data (101 bytes)
<snip>
```

方案 — 启用TCP PMTUD

启用PMTUD

启用PMTUD后，无论对等体是直接连接还是非直接连接，MSS初始计算始终会考虑出口接口IP MTU。

此场景提供有关启用PMTUD时的预期行为的见解。此处，Cisco IOS XR节点R4扮演活动角色，管理TCP连接，并在目标端口179上打开与Cisco IOS XR节点R1的TCP会话。两个节点在其接口上使用默认IP MTU值。



映像3.1 — 启用TCP PMTUD。

此场景中的MSS计算可概述如下：

- 所有节点都使用默认IP MTU 1500字节
- TCP路径MTU发现已启用
- TCP对等体未直接连接
    - R4管理BGP连接
    - R4发送MSS为1460字节的SYN
        - 1500（接口IP MTU） — 20(minTCP_H)- 20(minIP_H)
    - R1发送SYN、MSS为1460字节的ACK
        - 发送[Received MSS;本地初始MSS]
        - 接收的MSS 1460字节；本地初始MSS 1460字节
    - 两个对等体上均使用最低MSS值

为了突出启用PMTUD后引入的行为更改，接下来的输出说明了事件的顺序：

1. 在默认情况下禁用PMTUD时已建立TCP会话的初始状态；
2. 在TCP对等体R4和R1上配置并启用PMTUD;

3. TCP会话重新启动，MSS计算发生，并受TCP PMTUD的影响。

如在R4上看到的 — ACTIVE - TCP PMTUD已禁用（默认）：

<#root>

```
! - as seen on R4 - Active
! - TCP path mtu discovery disabled (default)
! - TCP session initial state

RP/0/0/CPU0:R4#show tcp detail pcb 0x121536c8
Fri Jan  8 16:06:30.237 UTC
------------------------------------------------------------
Connection state is ESTAB, I/O status: 0, socket status: 0
Established at Fri Jan  8 16:05:15 2021

PCB 0x121536c8, SO 0x12155370, TCPCB 0x12154f64, vrfid 0x60000000,
Pak Prio: Medium, TOS: 192, TTL: 255, Hash index: 376
Local host: 192.168.0.4, Local port: 20155 (Local App PID: 1052958)
Foreign host: 192.168.0.1, Foreign port: 179

Current send queue size in bytes: 0 (max 24576)
Current receive queue size in bytes: 0 (max 32768)  mis-ordered: 0 bytes
Current receive queue size in packets: 0 (max 0)

Timer          Starts    Wakeups         Next(msec)
Retrans             6          1                0
SendWnd             0          0                0
TimeWait            0          0                0
AckHold             3          2                0
KeepAlive           1          0                0
PmtuAger            0          0                0
GiveUp              0          0                0
Throttle            0          0                0

   iss: 357400981   snduna: 357401257   sndnxt: 357401257
sndmax: 357401257   sndwnd: 32546       sndcwnd: 3720
   irs: 524019443   rcvnxt: 524019719   rcvwnd: 32565   rcvadv: 524052284

SRTT: 72 ms,  RTTO: 416 ms,  RTV: 344 ms,  KRTT: 0 ms
minRTT: 19 ms,  maxRTT: 229 ms

ACK hold time: 200 ms, Keepalive time: 0 sec, SYN waittime: 30 sec
Giveup time: 0 ms, Retransmission retries: 0, Retransmit forever: FALSE
Connect retries remaining: 30, connect retry interval: 30 secs

State flags: none
Feature flags: Win Scale, Nagle
Request flags: Win Scale

Datagrams (in bytes): MSS 1240, peer MSS 1240, min MSS 1240, max MSS 1240

Window scales: rcv 0, snd 0, request rcv 0, request snd 0
Timestamp option: recent 0, recent age 0, last ACK sent 0
Sack blocks {start, end}: none
Sack holes {start, end, dups, rxmit}: none
```

```
Socket options: SO_REUSEADDR, SO_REUSEPORT, SO_NBIO
Socket states: SS_ISCONNECTED, SS_PRIV
Socket receive buffer states: SB_DEL_WAKEUP
Socket send buffer states: SB_DEL_WAKEUP
Socket receive buffer: Low/High watermark 1/32768
Socket send buffer   : Low/High watermark 2048/24576, Notify threshold 0

PDU information:
 #PDU's in buffer: 0
FIB Lookup Cache:  IFH: 0x40  PD ctx: size: 0  data:
  Num Labels: 0  Label Stack:

RP/0/0/CPU0:R4#
```

如在R1上看到的 — PASSIVE - TCP PMTUD已禁用（默认）：

<#root>

```
! - as seen on R1 - Passive
! - TCP path mtu discovery disabled (default)
! - TCP session initial state

RP/0/0/CPU0:R1#show tcp detail pcb 0x12157020
Fri Jan  8 16:05:52.868 UTC
================================================================
Connection state is ESTAB, I/O status: 0, socket status: 0
Established at Fri Jan  8 16:05:12 2021

PCB 0x12157020, SO 0x121565ac, TCPCB 0x121560ec, vrfid 0x60000000,
Pak Prio: Medium, TOS: 192, TTL: 255, Hash index: 376
Local host: 192.168.0.1, Local port: 179 (Local App PID: 983326)
Foreign host: 192.168.0.4, Foreign port: 20155

Current send queue size in bytes: 0 (max 24576)
Current receive queue size in bytes: 0 (max 32768)  mis-ordered: 0 bytes
Current receive queue size in packets: 0 (max 0)

Timer          Starts    Wakeups         Next(msec)
Retrans             3          0                 0
SendWnd             0          0                 0
TimeWait            0          0                 0
AckHold             3          1                 0
KeepAlive           1          0                 0
PmtuAger            0          0                 0
GiveUp              0          0                 0
Throttle            0          0                 0

   iss: 524019443   snduna: 524019700   sndnxt: 524019700
sndmax: 524019700   sndwnd: 32584       sndcwnd: 3720
   irs: 357400981   rcvnxt: 357401238   rcvwnd: 32565   rcvadv: 357433803

SRTT: 46 ms,  RTTO: 300 ms,  RTV: 249 ms,  KRTT: 0 ms
minRTT: 19 ms,  maxRTT: 239 ms

ACK hold time: 200 ms, Keepalive time: 0 sec, SYN waittime: 30 sec
Giveup time: 0 ms, Retransmission retries: 0, Retransmit forever: FALSE
Connect retries remaining: 0, connect retry interval: 0 secs
```

```
State flags: none
Feature flags: Win Scale, Nagle
Request flags: Win Scale


Datagrams (in bytes): MSS 1240, peer MSS 1240, min MSS 1240, max MSS 1240



Window scales: rcv 0, snd 0, request rcv 0, request snd 0
Timestamp option: recent 0, recent age 0, last ACK sent 0
Sack blocks {start, end}: none
Sack holes {start, end, dups, rxmit}: none

Socket options: SO_REUSEADDR, SO_REUSEPORT, SO_NBIO
Socket states: SS_ISCONNECTED, SS_PRIV
Socket receive buffer states: SB_DEL_WAKEUP
Socket send buffer states: SB_DEL_WAKEUP
Socket receive buffer: Low/High watermark 1/32768
Socket send buffer   : Low/High watermark 2048/24576, Notify threshold 0

PDU information:
 #PDU's in buffer: 0
FIB Lookup Cache:  IFH: 0x40  PD ctx: size: 0  data:
  Num Labels: 0  Label Stack:

RP/0/0/CPU0:R1#
```

如在R4上看到的 — ACTIVE - TCP PMTUD已启用：

<#root>

```
! - 'debug tcp pmtud' output on R4
! - tcp path mtu discovery enabled and uses default Path MTU aging timer (10 min / 600000 msec)

RP/0/0/CPU0:Jan  8 16:09:28.285 : tcp[399]: [t21] Try to enable path MTU discovery(neww age timer: 10 m
RP/0/0/CPU0:Jan  8 16:09:28.285 : tcp[399]: [t21] Path mtu is ON (age-timer: 10)

! - as seen on R4 - Active
! - TCP PMTUD is enabled

RP/0/0/CPU0:R4#show tcp detail pcb 0x121536c8
Fri Jan  8 16:11:00.138 UTC
================================================================
Connection state is ESTAB, I/O status: 0, socket status: 0
Established at Fri Jan  8 16:05:15 2021

PCB 0x121536c8, SO 0x12155370, TCPCB 0x12154f64, vrfid 0x60000000,
Pak Prio: Medium, TOS: 192, TTL: 255, Hash index: 376
Local host: 192.168.0.4, Local port: 20155 (Local App PID: 1052958)
Foreign host: 192.168.0.1, Foreign port: 179

Current send queue size in bytes: 0 (max 24576)
Current receive queue size in bytes: 0 (max 32768)  mis-ordered: 0 bytes
Current receive queue size in packets: 0 (max 0)

Timer          Starts    Wakeups         Next(msec)
Retrans            10          1                  0
SendWnd             0          0                  0
```

```
TimeWait            0          0               0
AckHold             7          4               0
KeepAlive           1          0               0

PmtuAger            1          0          508096


GiveUp              0          0               0
Throttle            0          0               0

   iss: 357400981   snduna: 357401333   sndnxt: 357401333
sndmax: 357401333   sndwnd: 32470       sndcwnd: 3720
   irs: 524019443   rcvnxt: 524019795   rcvwnd: 32489   rcvadv: 524052284

SRTT: 116 ms,  RTTO: 578 ms,  RTV: 462 ms,  KRTT: 0 ms
minRTT: 9 ms,  maxRTT: 229 ms

ACK hold time: 200 ms, Keepalive time: 0 sec, SYN waittime: 30 sec
Giveup time: 0 ms, Retransmission retries: 0, Retransmit forever: FALSE
Connect retries remaining: 30, connect retry interval: 30 secs

State flags: PMTU ager
Feature flags: Win Scale, Nagle,

Path MTU


Request flags: Win Scale


Datagrams (in bytes): MSS 1240, peer MSS 1240, min MSS 1240, max MSS 1240



Window scales: rcv 0, snd 0, request rcv 0, request snd 0
Timestamp option: recent 0, recent age 0, last ACK sent 0
Sack blocks {start, end}: none
Sack holes {start, end, dups, rxmit}: none

Socket options: SO_REUSEADDR, SO_REUSEPORT, SO_NBIO
Socket states: SS_ISCONNECTED, SS_PRIV
Socket receive buffer states: SB_DEL_WAKEUP
Socket send buffer states: SB_DEL_WAKEUP
Socket receive buffer: Low/High watermark 1/32768
Socket send buffer   : Low/High watermark 2048/24576, Notify threshold 0

PDU information:
 #PDU's in buffer: 0
FIB Lookup Cache:  IFH: 0x40  PD ctx: size: 0  data:
  Num Labels: 0  Label Stack:

RP/0/0/CPU0:R4#
```

如在R1上看到的 — 被动 — 启用TCP PMTUD:

<#root>

```
! - 'debug tcp pmtud' output on R1
! - tcp path mtu discovery is enabled and uses default Path MTU aging timer (10 min / 600000 msec)
```

```
RP/0/0/CPU0:Jan  8 16:09:25.214 : tcp[399]: [t21] Try to enable path MTU discovery(neww age timer: 10 m
RP/0/0/CPU0:Jan  8 16:09:25.214 : tcp[399]: [t21] Path mtu is ON (age-timer: 10)

! - as seen on R1 - Passive
! - TCP PMTUD is enabled

RP/0/0/CPU0:R1#show tcp detail pcb 0x12157020
Fri Jan  8 16:10:03.101 UTC
================================================================
Connection state is ESTAB, I/O status: 0, socket status: 0
Established at Fri Jan  8 16:05:12 2021

PCB 0x12157020, SO 0x121565ac, TCPCB 0x121560ec, vrfid 0x60000000,
Pak Prio: Medium, TOS: 192, TTL: 255, Hash index: 376
Local host: 192.168.0.1, Local port: 179 (Local App PID: 983326)
Foreign host: 192.168.0.4, Foreign port: 20155

Current send queue size in bytes: 0 (max 24576)
Current receive queue size in bytes: 0 (max 32768)  mis-ordered: 0 bytes
Current receive queue size in packets: 0 (max 0)

Timer          Starts    Wakeups         Next(msec)
Retrans             7          0                0
SendWnd             0          0                0
TimeWait            0          0                0
AckHold             7          4                0
KeepAlive           1          0                0

PmtuAger            1          0           562042


GiveUp              0          0                0
Throttle            0          0                0

   iss: 524019443   snduna: 524019776   sndnxt: 524019776
sndmax: 524019776   sndwnd: 32508       sndcwnd: 3720
   irs: 357400981   rcvnxt: 357401314   rcvwnd: 32489   rcvadv: 357433803

SRTT: 95 ms,  RTTO: 528 ms,  RTV: 433 ms,  KRTT: 0 ms
minRTT: 19 ms,  maxRTT: 239 ms

ACK hold time: 200 ms, Keepalive time: 0 sec, SYN waittime: 30 sec
Giveup time: 0 ms, Retransmission retries: 0, Retransmit forever: FALSE
Connect retries remaining: 0, connect retry interval: 0 secs

State flags: PMTU ager
Feature flags: Win Scale, Nagle,

Path MTU


Request flags: Win Scale


Datagrams (in bytes): MSS 1240, peer MSS 1240, min MSS 1240, max MSS 1240



Window scales: rcv 0, snd 0, request rcv 0, request snd 0
Timestamp option: recent 0, recent age 0, last ACK sent 0
Sack blocks {start, end}: none
Sack holes {start, end, dups, rxmit}: none

Socket options: SO_REUSEADDR, SO_REUSEPORT, SO_NBIO
```

```
Socket states: SS_ISCONNECTED, SS_PRIV
Socket receive buffer states: SB_DEL_WAKEUP
Socket send buffer states: SB_DEL_WAKEUP
Socket receive buffer: Low/High watermark 1/32768
Socket send buffer   : Low/High watermark 2048/24576, Notify threshold 0

PDU information:
 #PDU's in buffer: 0
FIB Lookup Cache:  IFH: 0x40  PD ctx: size: 0  data:
  Num Labels: 0  Label Stack:

RP/0/0/CPU0:R1#
```

注意PMTU老化器计时器行为：

```
! - Note PmtuAger timer initial value is 10min
! - but after initial interval expires then it expires every 2min
! - As seen from 'debug tcp pmtud' output
! - TCP PMTUD is enabled

RP/0/0/CPU0:Jan  8 16:09:25.214 : tcp[399]: [t21] Try to enable path MTU discovery(neww age timer: 10 m
RP/0/0/CPU0:Jan  8 16:09:25.214 : tcp[399]: [t21] Path mtu is ON (age-timer: 10)
RP/0/0/CPU0:Jan  8 16:19:25.233 : tcp[399]: [t21] PCB 0x12157020: Trying next higher MTU: 1240
RP/0/0/CPU0:Jan  8 16:21:25.245 : tcp[399]: [t21] PCB 0x12157020: Trying next higher MTU: 1240
RP/0/0/CPU0:Jan  8 16:23:25.256 : tcp[399]: [t21] PCB 0x12157020: Trying next higher MTU: 1240
```

如R4上所示 — ACTIVE - BGP会话重新启动 — TCP SYN:

<#root>

```
! - Once BGP session is cleared
! - TCP SYN sourced from R4 - Active
! - MSS calculation takes place and is influenced by TCP PMTUD

2734   4.810311       192.168.0.4 192.168.0.1 TCP    62     32077 → 179 [SYN] Seq=0 Win=16384 Len=0
```

MSS=1460

```
 WS=1

Frame 2734: 62 bytes on wire (496 bits), 62 bytes captured (496 bits) on interface 0
Ethernet II, Src: fa:16:3e:d7:7e:f6 (fa:16:3e:d7:7e:f6), Dst: fa:16:3e:8f:8f:54 (fa:16:3e:8f:8f:54)
Internet Protocol Version 4, Src: 192.168.0.4, Dst: 192.168.0.1
Transmission Control Protocol, Src Port: 32077, Dst Port: 179, Seq: 0, Len: 0
    Source Port: 32077
    Destination Port: 179
    [Stream index: 25]
    [TCP Segment Len: 0]
    Sequence number: 0    (relative sequence number)
    Acknowledgment number: 0
    Header Length: 28 bytes
    Flags: 0x002 (SYN)
    Window size value: 16384
    [Calculated window size: 16384]
```

```
    Checksum: 0x6398 [unverified]
    [Checksum Status: Unverified]
    Urgent pointer: 0
    Options: (8 bytes), Maximum segment size, Window scale, End of Option List (EOL)
        Maximum segment size: 1460 bytes
            Kind: Maximum Segment Size (2)
            Length: 4
```

MSS Value: 1460

```
        Window scale: 0 (multiply by 1)
        End of Option List (EOL)
```

如R1上所示 — PASSIVE - BGP Session restart - TCP SYN，ACK。

<#root>

```
! - Once BGP session is cleared
! - TCP SYN,ACK sourced from R1 - Passive
! - MSS calculation takes place and is influenced by TCP PMTUD

2735   0.003879       192.168.0.1 192.168.0.4 TCP    62     179 → 32077 [SYN, ACK] Seq=0 Ack=1 Win=16384 Le
```

MSS=1460

```
 WS=1

Frame 2735: 62 bytes on wire (496 bits), 62 bytes captured (496 bits) on interface 0
Ethernet II, Src: fa:16:3e:8f:8f:54 (fa:16:3e:8f:8f:54), Dst: fa:16:3e:d7:7e:f6 (fa:16:3e:d7:7e:f6)
Internet Protocol Version 4, Src: 192.168.0.1, Dst: 192.168.0.4
Transmission Control Protocol, Src Port: 179, Dst Port: 32077, Seq: 0, Ack: 1, Len: 0
    Source Port: 179
    Destination Port: 32077
    [Stream index: 25]
    [TCP Segment Len: 0]
    Sequence number: 0    (relative sequence number)
    Acknowledgment number: 1    (relative ack number)
    Header Length: 28 bytes
    Flags: 0x012 (SYN, ACK)
    Window size value: 16384
    [Calculated window size: 16384]
    Checksum: 0xbf77 [unverified]
    [Checksum Status: Unverified]
    Urgent pointer: 0
    Options: (8 bytes), Maximum segment size, Window scale, End of Option List (EOL)
        Maximum segment size: 1460 bytes
            Kind: Maximum Segment Size (2)
            Length: 4
```

MSS Value: 1460

```
        Window scale: 0 (multiply by 1)
        End of Option List (EOL)
```

TCP会话详细信息，如R4上所见 — 活动 — 在启用TCP PMTUD并清除BGP会话之后：

<#root>

```
! - BGP session re-established
! - as seen on R4 - Active

RP/0/0/CPU0:R4#show tcp detail pcb 0x121567f4
Fri Jan  8 16:45:13.928 UTC
==============================================================
Connection state is ESTAB, I/O status: 0, socket status: 0
Established at Fri Jan  8 16:41:49 2021

PCB 0x121567f4, SO 0x12154460, TCPCB 0x12156190, vrfid 0x60000000,
Pak Prio: Medium, TOS: 192, TTL: 255, Hash index: 10
Local host: 192.168.0.4, Local port: 32077 (Local App PID: 1052958)
Foreign host: 192.168.0.1, Foreign port: 179

Current send queue size in bytes: 0 (max 24576)
Current receive queue size in bytes: 0 (max 32768)  mis-ordered: 0 bytes
Current receive queue size in packets: 0 (max 0)

Timer          Starts    Wakeups         Next(msec)
Retrans             8          1                 0
SendWnd             0          0                 0
TimeWait            0          0                 0
AckHold             5          3                 0
KeepAlive           1          0                 0
PmtuAger            0          0                 0
GiveUp              0          0                 0
Throttle            0          0                 0

   iss: 1254100669  snduna: 1254100983  sndnxt: 1254100983
sndmax: 1254100983  sndwnd: 32508       sndcwnd: 4380
   irs: 839938559   rcvnxt: 839938873   rcvwnd: 32527   rcvadv: 839971400

SRTT: 79 ms,  RTTO: 485 ms,  RTV: 406 ms,  KRTT: 0 ms
minRTT: 9 ms,  maxRTT: 229 ms

ACK hold time: 200 ms, Keepalive time: 0 sec, SYN waittime: 30 sec
Giveup time: 0 ms, Retransmission retries: 0, Retransmit forever: FALSE
Connect retries remaining: 30, connect retry interval: 30 secs

State flags: none
Feature flags: Win Scale, Nagle,

Path MTU


Request flags: Win Scale


Datagrams (in bytes): MSS 1460, peer MSS 1460, min MSS 1460, max MSS 1460



Window scales: rcv 0, snd 0, request rcv 0, request snd 0
Timestamp option: recent 0, recent age 0, last ACK sent 0
Sack blocks {start, end}: none
Sack holes {start, end, dups, rxmit}: none
```

```
Socket options: SO_REUSEADDR, SO_REUSEPORT, SO_NBIO
Socket states: SS_ISCONNECTED, SS_PRIV
Socket receive buffer states: SB_DEL_WAKEUP
Socket send buffer states: SB_DEL_WAKEUP
Socket receive buffer: Low/High watermark 1/32768
Socket send buffer   : Low/High watermark 2048/24576, Notify threshold 0

PDU information:
 #PDU's in buffer: 0
FIB Lookup Cache:  IFH: 0x40  PD ctx: size: 0  data:
  Num Labels: 0  Label Stack:

RP/0/0/CPU0:R4#
```

在R1上看到的TCP会话详细信息 — 被动 — 在TCP PMTUD启用并清除BGP会话之后。

<#root>

```
! - BGP session re-established
! - as seen on R1 - Passive

RP/0/0/CPU0:R1#show tcp detail pcb 0x121558cc
Fri Jan  8 16:44:59.448 UTC
==============================================================
Connection state is ESTAB, I/O status: 0, socket status: 0
Established at Fri Jan  8 16:41:46 2021

PCB 0x121558cc, SO 0x121556d4, TCPCB 0x121575bc, vrfid 0x60000000,
Pak Prio: Medium, TOS: 192, TTL: 255, Hash index: 10
Local host: 192.168.0.1, Local port: 179 (Local App PID: 983326)
Foreign host: 192.168.0.4, Foreign port: 32077

Current send queue size in bytes: 0 (max 24576)
Current receive queue size in bytes: 0 (max 32768)  mis-ordered: 0 bytes
Current receive queue size in packets: 0 (max 0)

Timer          Starts    Wakeups         Next(msec)
Retrans             6          0                 0
SendWnd             0          0                 0
TimeWait            0          0                 0
AckHold             6          3                 0
KeepAlive           1          0                 0
PmtuAger            0          0                 0
GiveUp              0          0                 0
Throttle            0          0                 0

   iss: 839938559   snduna: 839938873   sndnxt: 839938873
sndmax: 839938873   sndwnd: 32527       sndcwnd: 4380
   irs: 1254100669  rcvnxt: 1254100983  rcvwnd: 32508   rcvadv: 1254133491

SRTT: 76 ms,  RTTO: 454 ms,  RTV: 378 ms,  KRTT: 0 ms
minRTT: 19 ms,  maxRTT: 219 ms

ACK hold time: 200 ms, Keepalive time: 0 sec, SYN waittime: 30 sec
Giveup time: 0 ms, Retransmission retries: 0, Retransmit forever: FALSE
Connect retries remaining: 0, connect retry interval: 0 secs

State flags: none
```

```
Feature flags: Win Scale, Nagle,
```

Path MTU

```
Request flags: Win Scale
```

Datagrams (in bytes): MSS 1460, peer MSS 1460, min MSS 1460, max MSS 1460

```
Window scales: rcv 0, snd 0, request rcv 0, request snd 0
Timestamp option: recent 0, recent age 0, last ACK sent 0
Sack blocks {start, end}: none
Sack holes {start, end, dups, rxmit}: none

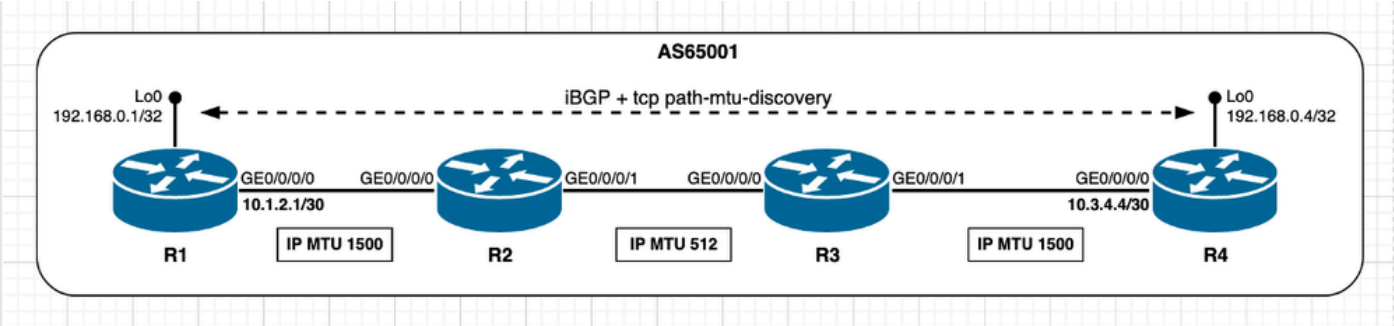
Socket options: SO_REUSEADDR, SO_REUSEPORT, SO_NBIO
Socket states: SS_ISCONNECTED, SS_PRIV
Socket receive buffer states: SB_DEL_WAKEUP
Socket send buffer states: SB_DEL_WAKEUP
Socket receive buffer: Low/High watermark 1/32768
Socket send buffer    : Low/High watermark 2048/24576, Notify threshold 0

PDU information:
 #PDU's in buffer: 0
FIB Lookup Cache:  IFH: 0x40  PD ctx: size: 0  data:
  Num Labels: 0  Label Stack:

RP/0/0/CPU0:R1#
```

PMTUD — 路径段的IP MTU较低

前一场景有助于了解启用PMTUD的初始TCP会话建立后会发生什么情况。此场景构建在顶部，有助于了解TCP PMTUD的工作方式及其对已建立的TCP会话的影响。



映像3.2 — 启用PMTUD，且路径段的IP MTU较低。

将上一个映像视为参考，假设BGP会话已建立，并且R1发送由大小大于512字节的IP数据包承载的BGP更新消息。启用PMTUD后，现在设置DF位（不分段）。因此，节点R2会丢弃IP数据包并发送ICMP（Internet控制消息协议）消息(目的地不可达 — 类型3;需要分段 — 代码4)返回R1。在节点R1收到ICMP消息后，会触发PMTUD并尝试建立路径最低的IP MTU。它使用一组明确定义的平台级别中的下一个较低值，即视为新的TCP会话MSS值。然后，TCP使用新的MSS值重新传输原始BGP更新，此过程会根据需要重复多次，直到ICMP消息(目的地不可达 — 类型3;不再接收需要分段 — 代码4)。这意味着直到所使用的MSS值使得发送的每个数据包都位于最低路径段IP MTU之下。

随着时间的流逝，由PmtuAger定时器控制的PMTUD在相反方向上走过平台级别，并将MSS提升回其最大值。在任何指定时间，如果ICMP消息(目的地不可达 — 类型3;需要分段 — 再次收到代码4)，然后PMTUD按之前所述操作。

接下来的输出将介绍刚才介绍的PMTUD行为，并从已建立的TCP会话的场景开始。此处，Cisco IOS XR节点R4扮演主动角色，因此可管理TCP连接，并在目标端口179上打开与R1的TCP会话。两个节点在其接口上使用默认IP MTU值。此方案中的初始MSS计算可概述如下：

- R2和R3节点之间的中间网段使用非默认IP MTU 512字节。
- R1和R4在其接口上使用默认MTU值。
- TCP路径MTU发现已启用。
- TCP对等体未直接连接。
    - R4管理BGP连接。
    - R4发送MSS为1460字节的SYN。
        - 1500（接口IP MTU） — 20(minTCP_H)- 20(minIP_H)。
    - R1发送SYN、ACK，MSS为1460字节。
        - 发送[Received MSS;本地初始MSS]。
        - 接收的MSS 1460字节；本地初始MSS 1460字节。
    - 两个对等体上均使用最低MSS值。

源自R4的TCP SYN:

<#root>

```
! - Initial TCP session establishment
! - TCP SYN sourced from R4

392    6.752774       192.168.0.4 192.168.0.1 TCP    62     32449 → 179 [SYN] Seq=0 Win=16384 Len=0

MSS=1460

 WS=1

Frame 392: 62 bytes on wire (496 bits), 62 bytes captured (496 bits) on interface 0
Ethernet II, Src: fa:16:3e:5c:f1:80 (fa:16:3e:5c:f1:80), Dst: fa:16:3e:42:18:05 (fa:16:3e:42:18:05)
Internet Protocol Version 4, Src: 192.168.0.4, Dst: 192.168.0.1
Transmission Control Protocol, Src Port: 32449, Dst Port: 179, Seq: 0, Len: 0
    Source Port: 32449
    Destination Port: 179
    [Stream index: 10]
    [TCP Segment Len: 0]
    Sequence number: 0    (relative sequence number)
    Acknowledgment number: 0
    Header Length: 28 bytes
    Flags: 0x002 (SYN)
    Window size value: 16384
    [Calculated window size: 16384]
    Checksum: 0x6858 [unverified]
    [Checksum Status: Unverified]
    Urgent pointer: 0
    Options: (8 bytes), Maximum segment size, Window scale, End of Option List (EOL)
        Maximum segment size: 1460 bytes
            Kind: Maximum Segment Size (2)
            Length: 4
```

MSS Value: 1460

```
        Window scale: 0 (multiply by 1)
        End of Option List (EOL)
```

来自R1的TCP SYN、ACK:

<#root>

```
! - Initial TCP session establishment
! - TCP SYN,ACK sourced from R1

393    0.003628      192.168.0.1 192.168.0.4 TCP    62     179 → 32449 [SYN, ACK] Seq=0 Ack=1 Win=16384 Len
```

MSS=1460

```
 WS=1

Frame 393: 62 bytes on wire (496 bits), 62 bytes captured (496 bits) on interface 0
Ethernet II, Src: fa:16:3e:42:18:05 (fa:16:3e:42:18:05), Dst: fa:16:3e:5c:f1:80 (fa:16:3e:5c:f1:80)
Internet Protocol Version 4, Src: 192.168.0.1, Dst: 192.168.0.4
Transmission Control Protocol, Src Port: 179, Dst Port: 32449, Seq: 0, Ack: 1, Len: 0
    Source Port: 179
    Destination Port: 32449
    [Stream index: 10]
    [TCP Segment Len: 0]
    Sequence number: 0    (relative sequence number)
    Acknowledgment number: 1    (relative ack number)
    Header Length: 28 bytes
    Flags: 0x012 (SYN, ACK)
    Window size value: 16384
    [Calculated window size: 16384]
    Checksum: 0x509e [unverified]
    [Checksum Status: Unverified]
    Urgent pointer: 0
    Options: (8 bytes), Maximum segment size, Window scale, End of Option List (EOL)
        Maximum segment size: 1460 bytes
            Kind: Maximum Segment Size (2)
            Length: 4
```

MSS Value: 1460

```
        Window scale: 0 (multiply by 1)
        End of Option List (EOL)
```

建立BGP会话后，节点R1发送BGP更新消息并接收ICMP消息(目标无法到达 — 类型3;需要分段 — 代码4)返回源自R2节点。

发生这种情况是因为传输BGP更新消息的IP数据包设置了DF位，并且R2/R3网段上使用的512字节的IP MTU小于1116字节的IP数据包大小。如前所述，接收ICMP消息会触发PMTUD。

在R1 ICMP上，收到类型3/代码4消息：

<#root>

```
! - as seen from R1 - Passive
! - After session is established R1 sends BGP Update message with IP length of 1116 Bytes
! - note IP Header Flags shows DF bit set

528    5.893055       192.168.0.1 192.168.0.4 BGP
```

1130

```
   UPDATE Message, KEEPALIVE Message

Frame 528: 1130 bytes on wire (9040 bits), 1130 bytes captured (9040 bits) on interface 0
Ethernet II, Src: fa:16:3e:42:18:05 (fa:16:3e:42:18:05), Dst: fa:16:3e:5c:f1:80 (fa:16:3e:5c:f1:80)
Internet Protocol Version 4, Src: 192.168.0.1, Dst: 192.168.0.4
    0100 .... = Version: 4
    .... 0101 = Header Length: 20 bytes (5)
    Differentiated Services Field: 0xc0 (DSCP: CS6, ECN: Not-ECT)
```

Total Length: 1116

```
    Identification: 0x8c37 (35895)
```

Flags: 0x02 (Don't Fragment)

```
    Fragment offset: 0
    Time to live: 255
    Protocol: TCP (6)
    Header checksum: 0xe09a [validation disabled]
    [Header checksum status: Unverified]
    Source: 192.168.0.1
    Destination: 192.168.0.4
    [Source GeoIP: Unknown]
    [Destination GeoIP: Unknown]
Transmission Control Protocol, Src Port: 179, Dst Port: 32449, Seq: 318, Ack: 251, Len: 1076
Border Gateway Protocol - UPDATE Message
Border Gateway Protocol - KEEPALIVE Message
<snip>

! - as seen from R1 - Passive
! - IP MTU on R2/R3 is lower than IP packet length and DF bit is set
! - R1 receives ICMP error message from R2
! - note R2 ICMP error message carries Next-Hop MTU
! - “The size in octets of the largest datagram that could be forwarded, along the path of
!   the original datagram, without being fragmented at this router. The size includes the
!   IP header and IP data, and does not include any lower-level headers.”

529    0.002423       10.2.3.1       192.168.0.1 ICMP   110
```

Destination unreachable (Fragmentation needed)

```
Frame 529: 110 bytes on wire (880 bits), 110 bytes captured (880 bits) on interface 0
Ethernet II, Src: fa:16:3e:5c:f1:80 (fa:16:3e:5c:f1:80), Dst: fa:16:3e:42:18:05 (fa:16:3e:42:18:05)
Internet Protocol Version 4, Src: 10.2.3.1, Dst: 192.168.0.1
    0100 .... = Version: 4
    .... 0101 = Header Length: 20 bytes (5)
    Differentiated Services Field: 0x00 (DSCP: CS0, ECN: Not-ECT)
    Total Length: 96
```

```
    Identification: 0x0001 (1)
    Flags: 0x00
    Fragment offset: 0
    Time to live: 255
    Protocol: ICMP (1)

    Header checksum: 0xac97 [validation disabled]
    [Header checksum status: Unverified]
    Source: 10.2.3.1
    Destination: 192.168.0.1
    [Source GeoIP: Unknown]
    [Destination GeoIP: Unknown]
Internet Control Message Protocol
    Type: 3 (Destination unreachable)

    Code: 4 (Fragmentation needed)

    Checksum: 0x2d52 [correct]
    [Checksum Status: Good]
    Length: 17
    [Length of original datagram: 68]
    Unused: 0011

  MTU of next hop: 512

    Internet Protocol Version 4, Src: 192.168.0.1, Dst: 192.168.0.4
        0100 .... = Version: 4
        .... 0101 = Header Length: 20 bytes (5)
        Differentiated Services Field: 0xc0 (DSCP: CS6, ECN: Not-ECT)
        Total Length: 1116
        Identification: 0x8c37 (35895)
        Flags: 0x02 (Don't Fragment)
        Fragment offset: 0
        Time to live: 254
        Protocol: TCP (6)
        Header checksum: 0xe19a [validation disabled]
        [Header checksum status: Unverified]
        Source: 192.168.0.1
        Destination: 192.168.0.4
        [Source GeoIP: Unknown]
        [Destination GeoIP: Unknown]
    Transmission Control Protocol, Src Port: 179, Dst Port: 32449, Seq: 2847698730, Ack: 2130367817
    Border Gateway Protocol - UPDATE Message
[Packet size limited during capture: IPv4 truncated]
```

在节点R1（由ICMP消息触发），TCP PMTUD尝试使用一组明确定义的平台(IP MTU)级别中的下一个较低值来建立端到端最低IP MTU。这些平台级别记录在RFC1191 — 路径MTU发现中。

```
MTU plateaus from RFC 1191
- values include both TCP and IP headers
65535
```

```
32000
17914
8166
4352
2002
1492
1006
508
296
68
```

但由于ICMP(目的地不可达 — 第3类；需要分段 — 代码4)节点R1收到的消息传递下一跳的MTU，然后如图所示，节点R1使用此值（在本例中为512字节），并调整TCP会话MSS值。请注意，原始TCP数据段的长度为1076个字节，因此重新传输原始TCP数据段需要三个数据包。

如在R1 - PASSIVE - PMTUD操作上所见：

<#root>

```
! - As seen from R1 - Passive
! - Hint is provided by ICMP unreachable message MTU of next-hop field: 512 bytes
! - R1 then considers this value and retransmits BGP Update split in three distinct packets
! - Sum of TCP length = 472 + 472 + 132 = 1076 bytes

530    0.007497       192.168.0.1 192.168.0.4 TCP    526    [TCP Out-Of-Order] 179 → 32449 [ACK] Seq=318 A

Len=472


532    0.015374       192.168.0.1 192.168.0.4 TCP    526    [TCP Retransmission] 179 → 32449 [ACK] Seq=79

Len=472


533    0.004129       192.168.0.1 192.168.0.4 TCP    186    [TCP Retransmission] 179 → 32449 [PSH, ACK] S

Len=132
```

如前所述，所有数据包经过一段时间传输后，PMTUD会按照PmtuAger计时器所定方向走过平台级别，并尝试根据当前场景将MSS提升至其最大值。

如R1上所见 — PMTUD跨定义的平台：

```
! - As seen from R1 - Passive - 'debug tcp pmtud' and 'debug icmp' active
! - TCP PMTUD is triggered once ICMP unreachable received

RP/0/0/CPU0:May 12 09:09:22.763 UTC: ipv4_io[266]: IPv4 ICMP: Received ICMP too big from 192.168.0.1 ab
RP/0/0/CPU0:May 12 09:09:22.763 UTC: ipv4_io[266]: ipv4_icmp_unreachable_rcvd ICMP unreach recvd: sendi
RP/0/0/CPU0:May 12 09:09:22.763 UTC: ipv4_io[266]: ip_icmp_lib_ipv4_receive: sending pak(0xb0c07d8f) to
RP/0/0/CPU0:May 12 09:09:22.763 UTC: tcp[399]: [t4] PCB 0x15393770: Process ICMP Dest-unreach (next hop

! - attempt new MSS 472 = MTU of next-hop(512) - TCP_H(20) - IP_H(20)
```

```
RP/0/0/CPU0:May 12 09:09:22.763 UTC: tcp[399]: [t4] PCB 0x15393770: Process ICMP Dest-unreach (next hop
RP/0/0/CPU0:May 12 09:09:22.763 UTC: tcp[399]: [t4] PCB 0x15393770: Try to use new MSS: 472
RP/0/0/CPU0:May 12 09:09:22.763 UTC: tcp[399]: [t4] PCB 0x15393770, New path MTU decided to use: 472  c

! - over time PMTUD attempts to raise MSS as per egress interface configured MTU

RP/0/0/CPU0:May 12 09:19:22.782 UTC: tcp[399]: [t23] PCB 0x15393770: Trying next higher MTU: 966
RP/0/0/CPU0:May 12 09:21:22.793 UTC: tcp[399]: [t23] PCB 0x15393770: Trying next higher MTU: 1452
RP/0/0/CPU0:May 12 09:23:22.805 UTC: tcp[399]: [t23] PCB 0x15393770: Trying next higher MTU: 1460
```

在这些输出上可以观察到最终状态。特别要注意节点R1显示的最小和最大MSS值，它突出显示并触发PMTUD的信号。

TCP会话详细信息（如R4上所示）— ACTIVE:

<#root>

```
! - Final stage as seen from R4 - Active

RP/0/0/CPU0:R4#show tcp detail pcb 0x153913b8
Wed May 12 10:09:43.246 UTC
==========================================================
Connection state is ESTAB, I/O status: 0, socket status: 0
Established at Wed May 12 09:02:07 2021

PCB 0x153913b8, SO 0x153917f0, TCPCB 0x1538fb58, vrfid 0x60000000,
Pak Prio: Medium, TOS: 192, TTL: 255, Hash index: 382
Local host: 192.168.0.4, Local port: 32449 (Local App PID: 1196319)
Foreign host: 192.168.0.1, Foreign port: 179
(Local App PID/instance/SPL_APP_ID: 1196319/1/0)

Current send queue size in bytes: 0 (max 24576)
Current receive queue size in bytes: 0 (max 32768)  mis-ordered: 0 bytes
Current receive queue size in packets: 0 (max 0)

Timer          Starts    Wakeups         Next(msec)
Retrans            72          1                0
SendWnd             0          0                0
TimeWait            0          0                0
AckHold            71         69                0
KeepAlive           1          0                0
PmtuAger            0          0                0
GiveUp              0          0                0
Throttle            0          0                0

   iss: 2130367566  snduna: 2130368957  sndnxt: 2130368957
sndmax: 2130368957  sndwnd: 31453       sndcwnd: 2920
   irs: 2847698412  rcvnxt: 2847700946  rcvwnd: 31799   rcvadv: 2847732745

SRTT: 220 ms,  RTTO: 300 ms,  RTV: 12 ms,  KRTT: 0 ms
minRTT: 9 ms,  maxRTT: 239 ms

ACK hold time: 200 ms, Keepalive time: 0 sec, SYN waittime: 30 sec
Giveup time: 0 ms, Retransmission retries: 0, Retransmit forever: FALSE
Connect retries remaining: 10, connect retry interval: 30 secs

State flags: none
Feature flags: Win Scale, Nagle,
```

Path MTU

Request flags: Win Scale

Datagrams (in bytes): MSS 1460, peer MSS 1460, min MSS 1460, max MSS 1460

Window scales: rcv 0, snd 0, request rcv 0, request snd 0
Timestamp option: recent 0, recent age 0, last ACK sent 0
Sack blocks {start, end}: none
Sack holes {start, end, dups, rxmit}: none

Socket options: SO_REUSEADDR, SO_REUSEPORT, SO_NBIO
Socket states: SS_ISCONNECTED, SS_PRIV
Socket receive buffer states: SB_DEL_WAKEUP
Socket send buffer states: SB_DEL_WAKEUP
Socket receive buffer: Low/High watermark 1/32768
Socket send buffer   : Low/High watermark 2048/24576, Notify threshold 0
Socket misc info   : Rcv data size (sb_cc) 0, so_qlen 0,
                       so_q0len 0, so_qlimit 0, so_error 0
                       so_auto_rearm 1

PDU information:
 #PDU's in buffer: 0
FIB Lookup Cache: IFH: 0x40 PD ctx: size: 0 data:
  Num Labels: 0 Label Stack:
Num of peers with authentication info: 0

RP/0/0/CPU0:R4#

R1上看到的TCP会话详细信息 — PASSIVE:

<#root>

```
! - Final stage as seen from R1 - Passive

RP/0/0/CPU0:R1#show tcp detail pcb 0x15393770
Wed May 12 10:12:41.432 UTC
================================================================
Connection state is ESTAB, I/O status: 240, socket status: 0
Established at Wed May 12 09:02:05 2021

PCB 0x15393770, SO 0x15394ea0, TCPCB 0x15391c0c, vrfid 0x60000000,
Pak Prio: Medium, TOS: 192, TTL: 255, Hash index: 382
Local host: 192.168.0.1, Local port: 179 (Local App PID: 1192224)
Foreign host: 192.168.0.4, Foreign port: 32449
(Local App PID/instance/SPL_APP_ID: 1192224/1/0)

Current send queue size in bytes: 0 (max 24576)
Current receive queue size in bytes: 0 (max 32768)  mis-ordered: 0 bytes
Current receive queue size in packets: 0 (max 0)

Timer          Starts    Wakeups         Next(msec)
Retrans            75          0                  0
SendWnd             0          0                  0
TimeWait            0          0                  0
```

```
AckHold            73         71               0
KeepAlive           1          0               0

PmtuAger           28         27           41595


GiveUp              0          0               0
Throttle            0          0               0

   iss: 2847698412  snduna: 2847701003  sndnxt: 2847701003
sndmax: 2847701003  sndwnd: 31742       sndcwnd: 4380
   irs: 2130367566  rcvnxt: 2130369014  rcvwnd: 31396   rcvadv: 2130400410

SRTT: 224 ms,  RTTO: 300 ms,  RTV: 23 ms,  KRTT: 0 ms
minRTT: 9 ms,  maxRTT: 259 ms

ACK hold time: 200 ms, Keepalive time: 0 sec, SYN waittime: 30 sec
Giveup time: 0 ms, Retransmission retries: 0, Retransmit forever: FALSE
Connect retries remaining: 0, connect retry interval: 0 secs

State flags: PMTU ager
Feature flags: Win Scale, Nagle,

Path MTU


Request flags: Win Scale


Datagrams (in bytes): MSS 1460, peer MSS 1460, min MSS 472, max MSS 1460



Window scales: rcv 0, snd 0, request rcv 0, request snd 0
Timestamp option: recent 0, recent age 0, last ACK sent 0
Sack blocks {start, end}: none
Sack holes {start, end, dups, rxmit}: none

Socket options: SO_REUSEADDR, SO_REUSEPORT, SO_NBIO
Socket states: SS_ISCONNECTED, SS_PRIV
Socket receive buffer states: SB_DEL_WAKEUP
Socket send buffer states: SB_DEL_WAKEUP
Socket receive buffer: Low/High watermark 1/32768
Socket send buffer   : Low/High watermark 2048/24576, Notify threshold 0
Socket misc info     : Rcv data size (sb_cc) 0, so_qlen 0,
                       so_q0len 0, so_qlimit 0, so_error 0
                       so_auto_rearm 1

PDU information:
 #PDU's in buffer: 0
FIB Lookup Cache:  IFH: 0x20  PD ctx: size: 0  data:
  Num Labels: 0  Label Stack:
Num of peers with authentication info: 0

RP/0/0/CPU0:R1#
```

最后，如果在任何给定时间存在ICMP(目的地不可达 — 类型3;需要分段 — 代码4)消息再次收到，然后PMTUD再次按之前所述操作。

从R1 - PASSIVE - PMTUD已再次触发：

<#root>

```
! - As seen from R1 - Passive
! - TCP PMTUD is again triggered upon new ICMP unreachable received
! - Behavior can be triggered via clearing redistributed, network and aggregate routes originated

RP/0/0/CPU0:R1#clear bgp ipv4 all self-originated
Wed May 12 10:19:06.836 UTC
RP/0/0/CPU0:R1#

! - New BGP update message is sourced from R1 after clear bgp command

1707   1.712657       192.168.0.1 192.168.0.4 BGP    1121   UPDATE Message

Frame 1707: 1121 bytes on wire (8968 bits), 1121 bytes captured (8968 bits) on interface 0
Ethernet II, Src: fa:16:3e:42:18:05 (fa:16:3e:42:18:05), Dst: fa:16:3e:5c:f1:80 (fa:16:3e:5c:f1:80)
Internet Protocol Version 4, Src: 192.168.0.1, Dst: 192.168.0.4
    0100 .... = Version: 4
    .... 0101 = Header Length: 20 bytes (5)
    Differentiated Services Field: 0xc0 (DSCP: CS6, ECN: Not-ECT)
    Total Length: 1107
    Identification: 0x1a38 (6712)
    Flags: 0x02 (Don't Fragment)
    Fragment offset: 0
    Time to live: 255
    Protocol: TCP (6)
    Header checksum: 0x52a3 [validation disabled]
    [Header checksum status: Unverified]
    Source: 192.168.0.1
    Destination: 192.168.0.4
    [Source GeoIP: Unknown]
    [Destination GeoIP: Unknown]
Transmission Control Protocol, Src Port: 179, Dst Port: 32449, Seq: 2705, Ack: 1562, Len: 1067
Border Gateway Protocol - UPDATE Message

! - ICMP Destination Unreachable / Fragmentation needed is received and triggers PMTUD

1708   0.001614       10.2.3.1       192.168.0.1 ICMP   110

Destination unreachable (Fragmentation needed)



Frame 1708: 110 bytes on wire (880 bits), 110 bytes captured (880 bits) on interface 0
Ethernet II, Src: fa:16:3e:5c:f1:80 (fa:16:3e:5c:f1:80), Dst: fa:16:3e:42:18:05 (fa:16:3e:42:18:05)
Internet Protocol Version 4, Src: 10.2.3.1, Dst: 192.168.0.1
    0100 .... = Version: 4
    .... 0101 = Header Length: 20 bytes (5)
    Differentiated Services Field: 0x00 (DSCP: CS0, ECN: Not-ECT)
    Total Length: 96
    Identification: 0x0002 (2)
    Flags: 0x00
    Fragment offset: 0
    Time to live: 255


Protocol: ICMP (1)
```

```
    Header checksum: 0xac96 [validation disabled]
    [Header checksum status: Unverified]
    Source: 10.2.3.1
    Destination: 192.168.0.1
    [Source GeoIP: Unknown]
    [Destination GeoIP: Unknown]
Internet Control Message Protocol
```

Type: 3 (Destination unreachable)

Code: 4 (Fragmentation needed)

```
    Checksum: 0x3b73 [correct]
    [Checksum Status: Good]
    Length: 17
    [Length of original datagram: 68]
    Unused: 0011
```

MTU of next hop: 512

```
    Internet Protocol Version 4, Src: 192.168.0.1, Dst: 192.168.0.4
        0100 .... = Version: 4
        .... 0101 = Header Length: 20 bytes (5)
        Differentiated Services Field: 0xc0 (DSCP: CS6, ECN: Not-ECT)
        Total Length: 1107
        Identification: 0x1a38 (6712)
        Flags: 0x02 (Don't Fragment)
        Fragment offset: 0
        Time to live: 254
        Protocol: TCP (6)
        Header checksum: 0x53a3 [validation disabled]
        [Header checksum status: Unverified]
        Source: 192.168.0.1
        Destination: 192.168.0.4
        [Source GeoIP: Unknown]
        [Destination GeoIP: Unknown]
    Transmission Control Protocol, Src Port: 179, Dst Port: 32449, Seq: 2847701117, Ack: 2130369128
    Border Gateway Protocol - UPDATE Message

! - Note new/updated MSS value and PmtuAger
! - MSS 472 ; Aligned with “MTU of next hop” value contained in ICMP message

RP/0/0/CPU0:R1#show tcp detail pcb 0x15393770
Wed May 12 10:19:31.494 UTC
================================================================
Connection state is ESTAB, I/O status: 240, socket status: 0
Established at Wed May 12 09:02:05 2021

PCB 0x15393770, SO 0x15394ea0, TCPCB 0x15391c0c, vrfid 0x60000000,
Pak Prio: Medium, TOS: 192, TTL: 255, Hash index: 382
Local host: 192.168.0.1, Local port: 179 (Local App PID: 1192224)
Foreign host: 192.168.0.4, Foreign port: 32449
(Local App PID/instance/SPL_APP_ID: 1192224/1/0)

Current send queue size in bytes: 0 (max 24576)
Current receive queue size in bytes: 0 (max 32768)  mis-ordered: 0 bytes
Current receive queue size in packets: 0 (max 0)

Timer          Starts    Wakeups         Next(msec)
```

```
Retrans            83          0               0
SendWnd             0          0               0
TimeWait            0          0               0
AckHold            80         77               0
KeepAlive           1          0               0

PmtuAger           32         30          575401


GiveUp              0          0               0
Throttle            0          0               0

   iss: 2847698412  snduna: 2847702184  sndnxt: 2847702184
sndmax: 2847702184  sndwnd: 32173       sndcwnd: 944
   irs: 2130367566  rcvnxt: 2130369147  rcvwnd: 32730   rcvadv: 2130401877

SRTT: 221 ms,  RTTO: 300 ms,  RTV: 16 ms,  KRTT: 0 ms
minRTT: 9 ms,  maxRTT: 259 ms

ACK hold time: 200 ms, Keepalive time: 0 sec, SYN waittime: 30 sec
Giveup time: 0 ms, Retransmission retries: 0, Retransmit forever: FALSE
Connect retries remaining: 0, connect retry interval: 0 secs

State flags: PMTU ager
Feature flags: Win Scale, Nagle,

Path MTU


Request flags: Win Scale


Datagrams (in bytes): MSS 472, peer MSS 1460, min MSS 472, max MSS 1460



Window scales: rcv 0, snd 0, request rcv 0, request snd 0
Timestamp option: recent 0, recent age 0, last ACK sent 0
Sack blocks {start, end}: none
Sack holes {start, end, dups, rxmit}: none

Socket options: SO_REUSEADDR, SO_REUSEPORT, SO_NBIO
Socket states: SS_ISCONNECTED, SS_PRIV
Socket receive buffer states: SB_DEL_WAKEUP
Socket send buffer states: SB_DEL_WAKEUP
Socket receive buffer: Low/High watermark 1/32768
Socket send buffer   : Low/High watermark 2048/24576, Notify threshold 0
Socket misc info     : Rcv data size (sb_cc) 0, so_qlen 0,
                       so_q0len 0, so_qlimit 0, so_error 0
                       so_auto_rearm 1

PDU information:
 #PDU's in buffer: 0
FIB Lookup Cache:  IFH: 0x20  PD ctx: size: 0  data:
  Num Labels: 0  Label Stack:
Num of peers with authentication info: 0

RP/0/0/CPU0:R1#
```

在Cisco Bug ID CSCvf10395影响的Cisco IOS XR版本上，将忽略ICMP错误消息中包含的下一跳，节点将尝试通过使用之前提到并由RFC1191 - Path MTU discovery记录的一组定义良好的平台(IP

MTU)级别中的下一个较低值来建立端到端最低IP MTU。这些尝试一直进行到成功传输，这意味着直到ICMP(目的地不可达 — 第3类；需要分段 — 代码4)消息不再接收。

从受思科漏洞ID CSCvf10395影响的思科IOS XR版本的节点可以看出：

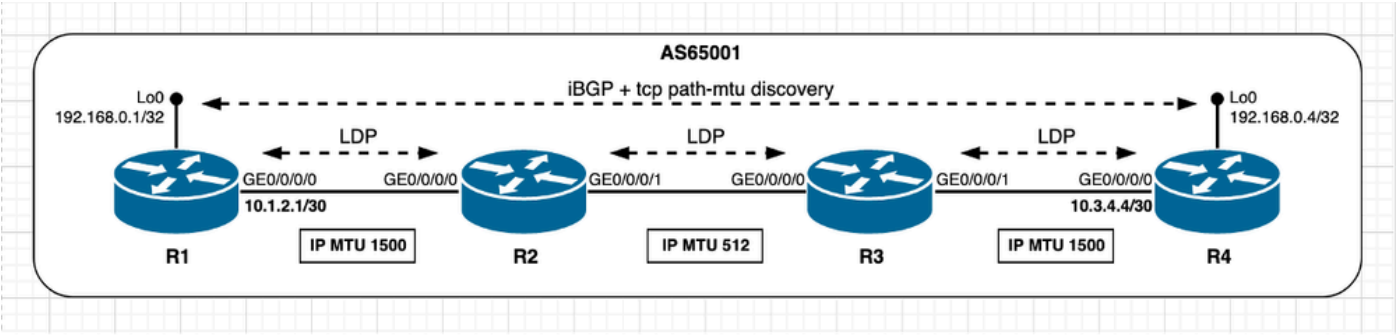
```
! - As seen from IOX XR node with a release impacted by Cisco bug ID CSCvf10395
! - Node ignores “MTU of next hop” and tries next lower plateau
! - This is observed till ICMP error messages are no longer received
! - Practical consequence is extra retransmissions occurrence

RP/0/0/CPU0:Feb 23 17:05:32.929 : tcp[399]: [t4] PCB 0x12152adc: Process ICMP Dest-unreach (next hop mtu
RP/0/0/CPU0:Feb 23 17:05:32.929 : tcp[399]: [t4] PCB 0x12152adc: Invalid next hop mtu (33554432), ignore
RP/0/0/CPU0:Feb 23 17:05:34.649 : tcp[399]: [t27] PCB 0x12152adc: Trying next lower MTU: 1452 <<<<<<<< H

RP/0/0/CPU0:Feb 23 17:05:35.519 : tcp[399]: [t4] PCB 0x12152adc: Process ICMP Dest-unreach (next hop mtu
RP/0/0/CPU0:Feb 23 17:05:35.519 : tcp[399]: [t4] PCB 0x12152adc: Invalid next hop mtu (33554432), ignore
RP/0/0/CPU0:Feb 23 17:05:37.239 : tcp[399]: [t27] PCB 0x12152adc: Trying next lower MTU: 966 <<<<<<<< HI

RP/0/0/CPU0:Feb 23 17:05:38.109 : tcp[399]: [t4] PCB 0x12152adc: Process ICMP Dest-unreach (next hop mtu
RP/0/0/CPU0:Feb 23 17:05:38.109 : tcp[399]: [t4] PCB 0x12152adc: Invalid next hop mtu (33554432), ignore
RP/0/0/CPU0:Feb 23 17:05:39.829 : tcp[399]: [t27] PCB 0x12152adc: Trying next lower MTU: 468 <<<<<<<< HI
```

作为下一步，请考虑所有接口上使用标签分发协议(LDP)的相同场景。此处的目标是了解在支持MPLS的环境中，从以前的场景中可观察到哪些差异。



映像3.3 — 启用PMTUD，路径段具有更低的IP MTU - MPLS方案。

首先，考虑在PMTUD触发之前建立的BGP会话的初始阶段，如下所示。

在R4上看到的TCP(BGP)初始状态 — 活动 — 启用MPLS的场景：

<#root>

```
! - as seen on R4 - Active
! - TCP path MTU discovery enabled
! - MPLS LDP enabled
! - TCP session initial state

RP/0/0/CPU0:R4#show tcp detail pcb 0x153bdaf0
Mon May 17 08:32:16.673 UTC
================================================================
Connection state is ESTAB, I/O status: 0, socket status: 0
```

```
Established at Mon May 17 08:31:57 2021

PCB 0x153bdaf0, SO 0x153acc80, TCPCB 0x153acea8, vrfid 0x60000000,
Pak Prio: Medium, TOS: 192, TTL: 255, Hash index: 757
Local host: 192.168.0.4, Local port: 57400 (Local App PID: 1196319)
Foreign host: 192.168.0.1, Foreign port: 179
(Local App PID/instance/SPL_APP_ID: 1196319/1/0)

Current send queue size in bytes: 0 (max 24576)
Current receive queue size in bytes: 0 (max 32768)  mis-ordered: 0 bytes
Current receive queue size in packets: 0 (max 0)

Timer          Starts    Wakeups         Next(msec)
Retrans             5          1                 0
SendWnd             0          0                 0
TimeWait            0          0                 0
AckHold             2          1                 0
KeepAlive           1          0                 0
PmtuAger            0          0                 0
GiveUp              0          0                 0
Throttle            0          0                 0

   iss: 1386459919  snduna: 1386460037  sndnxt: 1386460037
sndmax: 1386460037  sndwnd: 32726       sndcwnd: 4380
   irs: 3874414679  rcvnxt: 3874414864  rcvwnd: 32678   rcvadv: 3874447542

SRTT: 48 ms,  RTTO: 300 ms,  RTV: 228 ms,  KRTT: 0 ms
minRTT: 9 ms,  maxRTT: 229 ms

ACK hold time: 200 ms, Keepalive time: 0 sec, SYN waittime: 30 sec
Giveup time: 0 ms, Retransmission retries: 0, Retransmit forever: FALSE
Connect retries remaining: 10, connect retry interval: 30 secs

State flags: none
Feature flags: Win Scale, Nagle,
```

Path MTU

```
Request flags: Win Scale
```

Datagrams (in bytes): MSS 1460, peer MSS 1460, min MSS 1460, max MSS 1460

```
Window scales: rcv 0, snd 0, request rcv 0, request snd 0
Timestamp option: recent 0, recent age 0, last ACK sent 0
Sack blocks {start, end}: none
Sack holes {start, end, dups, rxmit}: none

Socket options: SO_REUSEADDR, SO_REUSEPORT, SO_NBIO
Socket states: SS_ISCONNECTED, SS_PRIV
Socket receive buffer states: SB_DEL_WAKEUP
Socket send buffer states: SB_DEL_WAKEUP
Socket receive buffer: Low/High watermark 1/32768
Socket send buffer   : Low/High watermark 2048/24576, Notify threshold 0
Socket misc info     : Rcv data size (sb_cc) 0, so_qlen 0,
                       so_q0len 0, so_qlimit 0, so_error 0
                       so_auto_rearm 1

PDU information:
 #PDU's in buffer: 0
```

```
FIB Lookup Cache:  IFH: 0x40  PD ctx: size: 0  data:
  Num Labels: 1  Label Stack:  0x5dc2
Num of peers with authentication info: 0

RP/0/0/CPU0:R4#
```

在R1上看到的TCP(BGP)初始状态 — 被动 — 启用MPLS的场景：

<#root>

```
! - as seen on R1 - Passive
! - TCP path MTU discovery enabled
! - MPLS LDP enabled
! - TCP session initial state

RP/0/0/CPU0:R1#show tcp detail pcb 0x153acc8c
Mon May 17 08:32:56.618 UTC
================================================================
Connection state is ESTAB, I/O status: 0, socket status: 0
Established at Mon May 17 08:31:55 2021

PCB 0x153acc8c, SO 0x153adad4, TCPCB 0x153adcfc, vrfid 0x60000000,
Pak Prio: Medium, TOS: 192, TTL: 255, Hash index: 757
Local host: 192.168.0.1, Local port: 179 (Local App PID: 1192224)
Foreign host: 192.168.0.4, Foreign port: 57400
(Local App PID/instance/SPL_APP_ID: 1192224/1/0)

Current send queue size in bytes: 0 (max 24576)
Current receive queue size in bytes: 0 (max 32768)  mis-ordered: 0 bytes
Current receive queue size in packets: 0 (max 0)

Timer          Starts    Wakeups         Next(msec)
Retrans             3          0                 0
SendWnd             0          0                 0
TimeWait            0          0                 0
AckHold             3          1                 0
KeepAlive           1          0                 0
PmtuAger            0          0                 0
GiveUp              0          0                 0
Throttle            0          0                 0

   iss: 3874414679  snduna: 3874414864  sndnxt: 3874414864
sndmax: 3874414864  sndwnd: 32678       sndcwnd: 4380
   irs: 1386459919  rcvnxt: 1386460037  rcvwnd: 32726   rcvadv: 1386492763

SRTT: 45 ms,  RTTO: 300 ms,  RTV: 239 ms,  KRTT: 0 ms
minRTT: 19 ms,  maxRTT: 229 ms

ACK hold time: 200 ms, Keepalive time: 0 sec, SYN waittime: 30 sec
Giveup time: 0 ms, Retransmission retries: 0, Retransmit forever: FALSE
Connect retries remaining: 0, connect retry interval: 0 secs

State flags: none
Feature flags: Win Scale, Nagle,

Path MTU


Request flags: Win Scale
```

```
Datagrams (in bytes): MSS 1460, peer MSS 1460, min MSS 1460, max MSS 1460


Window scales: rcv 0, snd 0, request rcv 0, request snd 0
Timestamp option: recent 0, recent age 0, last ACK sent 0
Sack blocks {start, end}: none
Sack holes {start, end, dups, rxmit}: none

Socket options: SO_REUSEADDR, SO_REUSEPORT, SO_NBIO
Socket states: SS_ISCONNECTED, SS_PRIV
Socket receive buffer states: SB_DEL_WAKEUP
Socket send buffer states: SB_DEL_WAKEUP
Socket receive buffer: Low/High watermark 1/32768
Socket send buffer   : Low/High watermark 2048/24576, Notify threshold 0
Socket misc info     : Rcv data size (sb_cc) 0, so_qlen 0,
                       so_q0len 0, so_qlimit 0, so_error 0
                       so_auto_rearm 1

PDU information:
 #PDU's in buffer: 0
FIB Lookup Cache:  IFH: 0x20  PD ctx: size: 0  data:
  Num Labels: 1  Label Stack:  0x5dc3
Num of peers with authentication info: 0

RP/0/0/CPU0:R1#
```

在此启用MPLS的场景中，观察到已建立TCP(LDP)会话的详细信息。请注意，前面介绍的有关TCP(BGP)会话的MSS计算的所有内容同样适用于TCP(LDP)会话。例如，节点R3和R2 TCP(LDP)会话MSS计算可以总结如下：

- R2和R3都使用512字节的非默认IP MTU。
- 已启用路径 MTU 发现.
- TCP对等体未直接连接（TCP会话在环回接口之间建立）。
    - R3管理LDP连接。
    - R3发送MSS为472字节的SYN。
        - 512（接口IP MTU） — 20(minTCP_H)- 20(minIP_H)。
    - R2发送SYN、ACK，其MSS为472字节。
        - 发送[Received MSS;本地初始MSS]。
        - 已接收472字节MSS;本地初始MSS 472字节。
    - 两个对等体上均使用最低MSS值。

TCP(LDP)会话详细信息，如在R3 - ACTIVE - MPLS启用方案中所示：

<#root>

```
! - as seen on R3 - Active
! - TCP path MTU discovery enabled
! - MPLS LDP enabled
! - TCP session initial state

RP/0/0/CPU0:R3#show tcp detail pcb 0x15393fbc
Mon May 17 08:33:30.627 UTC
```

```
================================================================
Connection state is ESTAB, I/O status: 0, socket status: 0
Established at Mon May 17 08:30:04 2021

PCB 0x15393fbc, SO 0x15393d94, TCPCB 0x153941b4, vrfid 0x60000000,
Pak Prio: Medium, TOS: 192, TTL: 255, Hash index: 970
Local host: 192.168.0.3, Local port: 57146 (Local App PID: 1151216)
Foreign host: 192.168.0.2, Foreign port: 646
(Local App PID/instance/SPL_APP_ID: 1151216/0/0)

Current send queue size in bytes: 0 (max 16384)
Current receive queue size in bytes: 0 (max 16384)  mis-ordered: 0 bytes
Current receive queue size in packets: 0 (max 60)

Timer          Starts    Wakeups         Next(msec)
Retrans             8          1                0
SendWnd             0          0                0
TimeWait            0          0                0
AckHold             6          4                0
KeepAlive           1          0                0
PmtuAger            0          0                0
GiveUp              0          0                0
Throttle            0          0                0

   iss: 2917752466  snduna: 2917752838  sndnxt: 2917752838
sndmax: 2917752838  sndwnd: 16013       sndcwnd: 944
   irs: 228184383   rcvnxt: 228184763   rcvwnd: 16005   rcvadv: 228200768

SRTT: 103 ms,  RTTO: 580 ms,  RTV: 477 ms,  KRTT: 0 ms
minRTT: 9 ms,  maxRTT: 279 ms

ACK hold time: 200 ms, Keepalive time: 0 sec, SYN waittime: 30 sec
Giveup time: 0 ms, Retransmission retries: 0, Retransmit forever: FALSE
Connect retries remaining: 1, connect retry interval: 3 secs

State flags: none
Feature flags: Win Scale, Nagle,

Path MTU


Request flags: Win Scale


Datagrams (in bytes): MSS 472, peer MSS 472, min MSS 472, max MSS 472



Window scales: rcv 0, snd 0, request rcv 0, request snd 0
Timestamp option: recent 0, recent age 0, last ACK sent 0
Sack blocks {start, end}: none
Sack holes {start, end, dups, rxmit}: none

Socket options: SO_REUSEADDR, SO_REUSEPORT, SO_NBIO
Socket states: SS_ISCONNECTED, SS_PRIV
Socket receive buffer states: SB_SEL, SB_DEL_WAKEUP
Socket send buffer states: SB_DEL_WAKEUP
Socket receive buffer: Low/High watermark 1/16384
Socket send buffer   : Low/High watermark 2048/16384, Notify threshold 0
Socket misc info     : Rcv data size (sb_cc) 0, so_qlen 0,
                       so_q0len 0, so_qlimit 0, so_error 0
                       so_auto_rearm 1
```

```
PDU information:
 #PDU's in buffer: 0
FIB Lookup Cache:  IFH: 0x40  PD ctx: size: 0  data:
  Num Labels: 1  Label Stack:  0x5dc2
Num of peers with authentication info: 0

RP/0/0/CPU0:R3#
```

TCP(LDP)会话详细信息，如R2上所示 — 被动 — 启用MPLS的场景：

<#root>

```
! - as seen on R2 - Passive
! - TCP path MTU discovery enabled
! - MPLS LDP enabled
! - TCP session initial state

RP/0/0/CPU0:R2#show tcp detail pcb 0x153a1f44
Mon May 17 08:34:28.843 UTC
================================================================
Connection state is ESTAB, I/O status: 0, socket status: 0
Established at Mon May 17 08:30:31 2021

PCB 0x153a1f44, SO 0x153a1d1c, TCPCB 0x153a213c, vrfid 0x60000000,
Pak Prio: Medium, TOS: 192, TTL: 255, Hash index: 970
Local host: 192.168.0.2, Local port: 646 (Local App PID: 1151216)
Foreign host: 192.168.0.3, Foreign port: 57146
(Local App PID/instance/SPL_APP_ID: 1151216/0/0)

Current send queue size in bytes: 0 (max 16384)
Current receive queue size in bytes: 0 (max 16384)  mis-ordered: 0 bytes
Current receive queue size in packets: 0 (max 60)

Timer          Starts    Wakeups         Next(msec)
Retrans             7          0                0
SendWnd             0          0                0
TimeWait            0          0                0
AckHold             7          5                0
KeepAlive           1          0                0
PmtuAger            0          0                0
GiveUp              0          0                0
Throttle            0          0                0

   iss: 228184383   snduna: 228184763   sndnxt: 228184763
sndmax: 228184763   sndwnd: 16005       sndcwnd: 944
   irs: 2917752466  rcvnxt: 2917752856  rcvwnd: 15995   rcvadv: 2917768851

SRTT: 95 ms,  RTTO: 561 ms,  RTV: 466 ms,  KRTT: 0 ms
minRTT: 0 ms,  maxRTT: 219 ms

ACK hold time: 200 ms, Keepalive time: 0 sec, SYN waittime: 30 sec
Giveup time: 0 ms, Retransmission retries: 0, Retransmit forever: FALSE
Connect retries remaining: 0, connect retry interval: 0 secs

State flags: none
Feature flags: Win Scale, Nagle,
```

Path MTU

```
Request flags: Win Scale
```

Datagrams (in bytes): MSS 472, peer MSS 472, min MSS 472, max MSS 472

```
Window scales: rcv 0, snd 0, request rcv 0, request snd 0
Timestamp option: recent 0, recent age 0, last ACK sent 0
Sack blocks {start, end}: none
Sack holes {start, end, dups, rxmit}: none

Socket options: SO_REUSEADDR, SO_REUSEPORT, SO_NBIO
Socket states: SS_ISCONNECTED, SS_PRIV
Socket receive buffer states: SB_SEL, SB_DEL_WAKEUP
Socket send buffer states: SB_DEL_WAKEUP
Socket receive buffer: Low/High watermark 1/16384
Socket send buffer   : Low/High watermark 2048/16384, Notify threshold 0
Socket misc info     : Rcv data size (sb_cc) 0, so_qlen 0,
                       so_q0len 0, so_qlimit 0, so_error 0
                       so_auto_rearm 1

PDU information:
 #PDU's in buffer: 0
FIB Lookup Cache:  IFH: 0x60  PD ctx: size: 0  data:
  Num Labels: 1  Label Stack:  0x5dc1
Num of peers with authentication info: 0

RP/0/0/CPU0:R2#
```

建立BGP会话后，R1发送BGP更新消息并接收ICMP消息(目的地不可达 — 类型3;需要分段 — 代码4)返回源自R2节点，该节点在节点R1上触发TCP PMTUD。出现这种情况是因为传送BGP更新消息的IP数据包设置了DF位，并且R2/R3网段使用的512字节的IP MTU小于IP数据包大小1116字节。与之前一样，接收此ICMP消息会触发PMTUD。与之前非MPLS方案相比，启用MPLS方案的差异在于节点R2 ICMP消息中包含的下一跳值(目的地不可达 — 类型3;需要分段 — 代码4)。 在此启用MPLS的场景中，下一跳MTU值占了4字节的额外MPLS开销，这意味着它占据R2上的出口MPLS标签堆栈，如以下输出所示。

在R1上看到的TCP路径MTU发现 — 被动 — 启用MPLS的场景：

<#root>

```
! - as seen from R1 - Passive
! - R1 sends BGP Update message with IP length of 1116 Bytes
! - Note MPLS Header as packet is to be label-switched (single label ; IGP label)
! - note IP Header Flags shows DF bit set

455    0.044859       192.168.0.1 192.168.0.4 BGP    1134   UPDATE Message, KEEPALIVE Message

Frame 455: 1134 bytes on wire (9072 bits), 1134 bytes captured (9072 bits) on interface 0
Ethernet II, Src: fa:16:3e:42:18:05 (fa:16:3e:42:18:05), Dst: fa:16:3e:5c:f1:80 (fa:16:3e:5c:f1:80)
```

MultiProtocol Label Switching Header

```
, Label: 24002, Exp: 6, S: 1, TTL: 255
Internet Protocol Version 4, Src: 192.168.0.1, Dst: 192.168.0.4
```

```
    0100 .... = Version: 4
    .... 0101 = Header Length: 20 bytes (5)
    Differentiated Services Field: 0xc0 (DSCP: CS6, ECN: Not-ECT)


Total Length: 1116


    Identification: 0xc6dd (50909)
    Flags: 0x02 (Don't Fragment)
        0... .... = Reserved bit: Not set


.1.. .... = Don't fragment: Set


        ..0. .... = More fragments: Not set
    Fragment offset: 0
    Time to live: 255
    Protocol: TCP (6)
    Header checksum: 0xa5f4 [validation disabled]
    [Header checksum status: Unverified]
    Source: 192.168.0.1
    Destination: 192.168.0.4
    [Source GeoIP: Unknown]
    [Destination GeoIP: Unknown]
Transmission Control Protocol, Src Port: 179, Dst Port: 57400, Seq: 242, Ack: 175, Len: 1076
Border Gateway Protocol - UPDATE Message
Border Gateway Protocol - KEEPALIVE Message
<snip>

! - as seen from R1 - Passive
! - IP MTU on R2/R3 of 512 bytes is lower than IP packet length and DF bit is set
! - R1 receives ICMP error message from R2
! - note R2 ICMP error message carries Next-Hop MTU
! - “The size in octets of the largest datagram that could be forwarded, along the path of
!   the original datagram, without being fragmented at this router. The size includes the
!   IP header and IP data, and does not include any lower-level headers.”
! - In present MPLS-enabled scenario Next-Hop MTU value is 508 bytes
! - In previous non-MPLS scenario Next-Hop MTU value was 512 bytes

456    0.014117       10.2.3.1       192.168.0.1 ICMP   182

Destination unreachable (Fragmentation needed)



Frame 456: 182 bytes on wire (1456 bits), 182 bytes captured (1456 bits) on interface 0
Ethernet II, Src: fa:16:3e:5c:f1:80 (fa:16:3e:5c:f1:80), Dst: fa:16:3e:42:18:05 (fa:16:3e:42:18:05)
Internet Protocol Version 4, Src: 10.2.3.1, Dst: 192.168.0.1
    0100 .... = Version: 4
    .... 0101 = Header Length: 20 bytes (5)
    Differentiated Services Field: 0x00 (DSCP: CS0, ECN: Not-ECT)
    Total Length: 168
    Identification: 0x001f (31)
    Flags: 0x00
        0... .... = Reserved bit: Not set
        .0.. .... = Don't fragment: Not set
        ..0. .... = More fragments: Not se
    Fragment offset: 0
    Time to live: 251

   Protocol: ICMP (1)
```

```
    Header checksum: 0xb031 [validation disabled]
    [Header checksum status: Unverified]
    Source: 10.2.3.1
    Destination: 192.168.0.1
    [Source GeoIP: Unknown]
    [Destination GeoIP: Unknown]
Internet Control Message Protocol
```

Type: 3 (Destination unreachable)

Code: 4 (Fragmentation needed)

```
    Checksum: 0x5199 [correct]
    [Checksum Status: Good]
    Length: 17
    [Length of original datagram: 68]
    Unused: 0011
```

MTU of next hop: 508

```
    Internet Protocol Version 4, Src: 192.168.0.1, Dst: 192.168.0.4
    Transmission Control Protocol, Src Port: 179, Dst Port: 57400, Seq: 3874414921, Ack: 1386460094
    Border Gateway Protocol - UPDATE Message

! - As seen from R1 - Passive
! - Hint is provided by ICMP unreachable message MTU of next-hop field: 508 bytes
! - R1 then considers this value and retransmits BGP Update split in three distinct packets
! - Sum of TCP length = 468 + 468 + 140 = 1076 bytes

457    0.006689       192.168.0.1 192.168.0.4 TCP    526    [TCP Retransmission] 179 → 57400 [ACK] Seq=24
```

Len=468

```
460    0.004001       192.168.0.1 192.168.0.4 TCP    526    [TCP Retransmission] 179 → 57400 [ACK] Seq=71
```

Len=468

```
461    0.001788       192.168.0.1 192.168.0.4 TCP    198    [TCP Retransmission] 179 → 57400 [PSH, ACK] S
```

Len=140

```
463    0.056695       192.168.0.4 192.168.0.1 TCP    54     57400 → 179 [ACK] Seq=175 Ack=1318 Win=31545 Le

! - As seen from R1 - Passive - 'debug tcp pmtud' and 'debug icmp' active
! - TCP PMTUD is triggered once ICMP unreachable received

RP/0/0/CPU0:May 17 08:29:56.131 UTC: tcp[399]: [t1] Try to enable path MTU discovery(neww age timer: 10
RP/0/0/CPU0:May 17 08:29:56.131 UTC: tcp[399]: [t1] Path mtu is ON (age-timer: 10)
RP/0/0/CPU0:May 17 08:35:51.726 UTC: ipv4_io[266]: ip_icmp_lib_ipv4_receive: Receiving pak(0xb0c07d8f)
RP/0/0/CPU0:May 17 08:35:51.726 UTC: ipv4_io[266]: Entering ipv4_mtu_update_cb
RP/0/0/CPU0:May 17 08:35:51.726 UTC: ipv4_io[266]: IPv4 ICMP: Received ICMP too big from 192.168.0.1 ab
RP/0/0/CPU0:May 17 08:35:51.726 UTC: ipv4_io[266]: ipv4_icmp_unreachable_rcvd ICMP unreach recvd: sendi
RP/0/0/CPU0:May 17 08:35:51.726 UTC: ipv4_io[266]: ip_icmp_lib_ipv4_receive: sending pak(0xb0c07d8f) to
RP/0/0/CPU0:May 17 08:35:51.726 UTC: tcp[399]: [t4] PCB 0x153acc8c: Process ICMP Dest-unreach (next hop

! - attempt new MSS 468 = MTU of next-hop(508) - TCP_H(20) - IP_H(20)
```

```
RP/0/0/CPU0:May 17 08:35:51.726 UTC: tcp[399]: [t4] PCB 0x153acc8c: Try to use new MSS: 468
RP/0/0/CPU0:May 17 08:35:51.726 UTC: tcp[399]: [t4] PCB 0x153acc8c, New path MTU decided to use: 468  c

! - over time PMTUD attempts to raise MSS as per egress interface configured MTU

RP/0/0/CPU0:May 17 08:45:51.745 UTC: tcp[399]: [t29] PCB 0x153acc8c: Trying next higher MTU: 966
RP/0/0/CPU0:May 17 08:47:51.757 UTC: tcp[399]: [t29] PCB 0x153acc8c: Trying next higher MTU: 1452
RP/0/0/CPU0:May 17 08:49:51.769 UTC: tcp[399]: [t29] PCB 0x153acc8c: Trying next higher MTU: 1460
```

从R1 — 被动 — TCP PMTUD触发 — 启用MPLS的场景可以看出：

<#root>

```
! - as seen on R1 - Passive
! - R1 session details after TCP PMTUD trigger

RP/0/0/CPU0:R1#show tcp detail pcb 0x153acc8c
Mon May 17 08:43:07.077 UTC
==============================================================
Connection state is ESTAB, I/O status: 240, socket status: 0
Established at Mon May 17 08:31:55 2021

PCB 0x153acc8c, SO 0x153adad4, TCPCB 0x153adcfc, vrfid 0x60000000,
Pak Prio: Medium, TOS: 192, TTL: 255, Hash index: 757
Local host: 192.168.0.1, Local port: 179 (Local App PID: 1192224)
Foreign host: 192.168.0.4, Foreign port: 57400
(Local App PID/instance/SPL_APP_ID: 1192224/1/0)

Current send queue size in bytes: 0 (max 24576)
Current receive queue size in bytes: 0 (max 32768)  mis-ordered: 0 bytes
Current receive queue size in packets: 0 (max 0)

Timer          Starts    Wakeups         Next(msec)
Retrans            15          0                0
SendWnd             0          0                0
TimeWait            0          0                0
AckHold            14          9                0
KeepAlive           1          0                0

PmtuAger            1          0           164599


GiveUp              0          0                0
Throttle            0          0                0

   iss: 3874414679  snduna: 3874416130  sndnxt: 3874416130
sndmax: 3874416130  sndwnd: 31412       sndcwnd: 936
   irs: 1386459919  rcvnxt: 1386460246  rcvwnd: 32517   rcvadv: 1386492763

SRTT: 180 ms,  RTTO: 509 ms,  RTV: 329 ms,  KRTT: 0 ms
minRTT: 19 ms,  maxRTT: 239 ms

ACK hold time: 200 ms, Keepalive time: 0 sec, SYN waittime: 30 sec
Giveup time: 0 ms, Retransmission retries: 0, Retransmit forever: FALSE
Connect retries remaining: 0, connect retry interval: 0 secs

State flags: PMTU ager
Feature flags: Win Scale, Nagle,
```

Path MTU

Request flags: Win Scale

Datagrams (in bytes): MSS 468, peer MSS 1460, min MSS 468, max MSS 1460

Window scales: rcv 0, snd 0, request rcv 0, request snd 0
Timestamp option: recent 0, recent age 0, last ACK sent 0
Sack blocks {start, end}: none
Sack holes {start, end, dups, rxmit}: none

Socket options: SO_REUSEADDR, SO_REUSEPORT, SO_NBIO
Socket states: SS_ISCONNECTED, SS_PRIV
Socket receive buffer states: SB_DEL_WAKEUP
Socket send buffer states: SB_DEL_WAKEUP
Socket receive buffer: Low/High watermark 1/32768
Socket send buffer : Low/High watermark 2048/24576, Notify threshold 0
Socket misc info : Rcv data size (sb_cc) 0, so_qlen 0,
so_q0len 0, so_qlimit 0, so_error 0
so_auto_rearm 1

PDU information:
#PDU's in buffer: 0
FIB Lookup Cache: IFH: 0x20 PD ctx: size: 0 data:
Num Labels: 1 Label Stack: 0x5dc3
Num of peers with authentication info: 0

RP/0/0/CPU0:R1#

请注意，在启用MPLS的场景中，节点R2 ICMP消息中包含的下一跳MTU的值用于出口MPLS标签堆栈。要进一步强化这一方面，请考虑下一个示例。如果在R2过滤的IP数据包与L3VPN服务相关联，则意味着以太网帧现在携带两个标签（IGP标签和VPN标签）。 然后，下一跳的MTU将反映所需的标签堆栈大小。请参阅这些输出。

如R1 - PASSIVE - L3 VPN服务数据包所示：

<#root>

```
! - as seen from R1 - Passive
! - L3 VPN service packet is sourced by node R1 and destined to node R4
! - Note presence of MPLS label stack - both IGP and VPN label are present
! - Note IP Total Length of 610 bytes higher than the IP MTU on R2/R3 segment
! - note IP Header Flags shows DF bit set

2024   0.302370       10.1.14.1      10.1.14.14     TELNET 632    Telnet Data ...

Frame 2024: 632 bytes on wire (5056 bits), 632 bytes captured (5056 bits) on interface 0
Ethernet II, Src: fa:16:3e:42:18:05 (fa:16:3e:42:18:05), Dst: fa:16:3e:5c:f1:80 (fa:16:3e:5c:f1:80)

MultiProtocol Label Switching Header, Label: 24002, Exp: 0, S: 0, TTL: 255

    0000 0101 1101 1100 0010 .... .... .... = MPLS Label: 24002
    .... .... .... .... .... 000. .... .... = MPLS Experimental Bits: 0
```

```
    .... .... .... .... .... ...0 .... .... = MPLS Bottom Of Label Stack: 0
    .... .... .... .... .... .... 1111 1111 = MPLS TTL: 255
```

MultiProtocol Label Switching Header, Label: 24005, Exp: 0, S: 1, TTL: 255

```
    0000 0101 1101 1100 0101 .... .... .... = MPLS Label: 24005
    .... .... .... .... .... 000. .... .... = MPLS Experimental Bits: 0
    .... .... .... .... .... ...1 .... .... = MPLS Bottom Of Label Stack: 1
    .... .... .... .... .... .... 1111 1111 = MPLS TTL: 255
Internet Protocol Version 4, Src: 10.1.14.1, Dst: 10.1.14.14
    0100 .... = Version: 4
    .... 0101 = Header Length: 20 bytes (5)
    Differentiated Services Field: 0x00 (DSCP: CS0, ECN: Not-ECT)
    Total Length: 610
    Identification: 0x7c9f (31903)
    Flags: 0x02 (Don't Fragment)
        0... .... = Reserved bit: Not set
```

.1.. .... = Don't fragment: Set

```
        ..0. .... = More fragments: Not set
    Fragment offset: 0
    Time to live: 255
    Protocol: TCP (6)
    Header checksum: 0xcce5 [validation disabled]
    [Header checksum status: Unverified]
    Source: 10.1.14.1
    Destination: 10.1.14.14
    [Source GeoIP: Unknown]
    [Destination GeoIP: Unknown]
Transmission Control Protocol, Src Port: 22008, Dst Port: 23, Seq: 34755, Ack: 93250, Len: 570
```

如在R1上看到的 — PASSIVE - L3 VPN服务 — ICMP类型3/代码4:

<#root>

```
! - as seen from R1 - Passive
! - IP MTU on R2/R3 of 512 bytes is lower than IP packet length and DF bit is set
! - R1 receives ICMP error message from R2
! - note R2 ICMP error message carries Next-Hop MTU
! - “The size in octets of the largest datagram that could be forwarded, along the path of
!   the original datagram, without being fragmented at this router. The size includes the
!   IP header and IP data, and does not include any lower-level headers.”
! - In present L3VPN MPLS-enabled scenario (dual-label) Next-Hop MTU value is 504 bytes
! - In previous MPLS scenario (single-label) Next-Hop MTU value was 508 bytes

2030   0.020299       10.2.3.1       10.1.14.1      ICMP   190
```

Destination unreachable (Fragmentation needed)

```
Frame 2030: 190 bytes on wire (1520 bits), 190 bytes captured (1520 bits) on interface 0
Ethernet II, Src: fa:16:3e:5c:f1:80 (fa:16:3e:5c:f1:80), Dst: fa:16:3e:42:18:05 (fa:16:3e:42:18:05)
MultiProtocol Label Switching Header, Label: 24005, Exp: 0, S: 1, TTL: 251
    0000 0101 1101 1100 0101 .... .... .... = MPLS Label: 24005
    .... .... .... .... .... 000. .... .... = MPLS Experimental Bits: 0
```

```
    .... .... .... .... .... ...1 .... .... = MPLS Bottom Of Label Stack: 1
    .... .... .... .... .... .... 1111 1011 = MPLS TTL: 251
Internet Protocol Version 4, Src: 10.2.3.1, Dst: 10.1.14.1
    0100 .... = Version: 4
    .... 0101 = Header Length: 20 bytes (5)
    Differentiated Services Field: 0x00 (DSCP: CS0, ECN: Not-ECT)
    Total Length: 172
    Identification: 0x002b (43)
    Flags: 0x00
        0... .... = Reserved bit: Not set
        .0.. .... = Don't fragment: Not set
        ..0. .... = More fragments: Not set
    Fragment offset: 0
    Time to live: 253


Protocol: ICMP (1)


    Header checksum: 0x9821 [validation disabled]
    [Header checksum status: Unverified]
    Source: 10.2.3.1
    Destination: 10.1.14.1
    [Source GeoIP: Unknown]
    [Destination GeoIP: Unknown]
Internet Control Message Protocol
    Type: 3 (Destination unreachable)



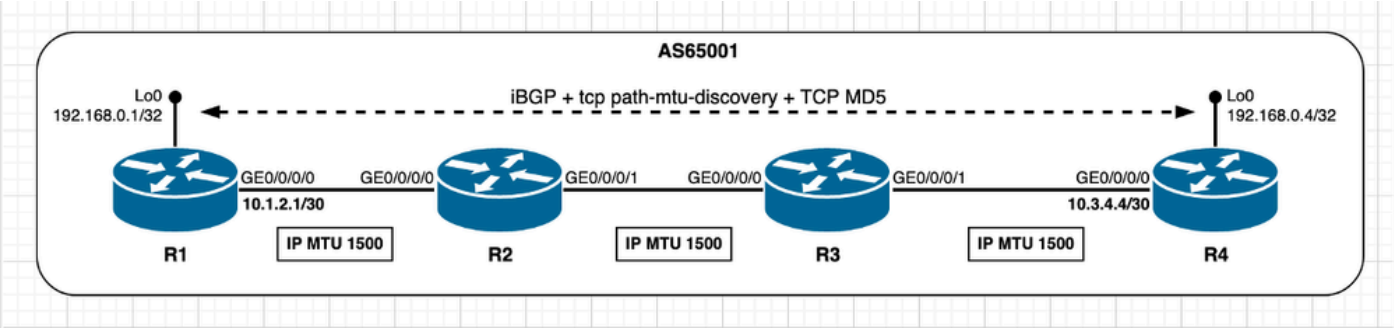
    Code: 4 (Fragmentation needed)


    Checksum: 0xbbac [correct]
    [Checksum Status: Good]
    Length: 17
    [Length of original datagram: 68]
    Unused: 0011


MTU of next hop: 504


    Internet Protocol Version 4, Src: 10.1.14.1, Dst: 10.1.14.14
        0100 .... = Version: 4
        .... 0101 = Header Length: 20 bytes (5)
        Differentiated Services Field: 0x00 (DSCP: CS0, ECN: Not-ECT)
        Total Length: 610
        Identification: 0x7c9f (31903)
        Flags: 0x02 (Don't Fragment)
            0... .... = Reserved bit: Not set
            .1.. .... = Don't fragment: Set
            ..0. .... = More fragments: Not set
        Fragment offset: 0
        Time to live: 255
        Protocol: TCP (6)
        Header checksum: 0xcce5 [validation disabled]
        [Header checksum status: Unverified]
        Source: 10.1.14.1
        Destination: 10.1.14.14
        [Source GeoIP: Unknown]
        [Destination GeoIP: Unknown]
    Transmission Control Protocol, Src Port: 22008, Dst Port: 23, Seq: 586828435, Ack: 754580617
```

PMTUD - TCP选项(MD5)



映像3.4 — 启用PMTUD和TCP MD5身份验证。

在启用TCP MD5身份验证的先前场景中，对PMTUD行为没有引入任何区别。与之前使用的TCP MD5身份验证共享一样，Cisco IOS XR考虑额外开销，并且活动TCP对等体初始MSS值反映相同。有关使用TCP选项的影响的其他详细信息，请参阅前面的使用TCP选项 — XR主动和使用TCP选项 — XR被动部分。此场景中的TCP MSS计算可概述如下：

- 所有节点使用默认IP MTU 1500字节。
- TCP路径MTU发现已启用。
- TCP对等体未直接连接。
- R1和R4上均启用了TCP MD5身份验证。
  - R4管理BGP连接。
  - R4发送MSS为1436字节的SYN。
    - 1500（接口IP MTU） — 20(minTCP_H)- 20(minIP_H)- 24字节（IOS XR TCP选项开销）。
  - R1发送SYN，MSS为1436字节的ACK。
    - 发送[Received MSS;本地初始MSS]。
    - 接收的MSS 1436字节；本地初始MSS 1460字节。
  - 两个对等体上均使用最低MSS值。

源自R4的TCP SYN:

<#root>

```
! - TCP SYN sourced from R4

2408   5.695076       192.168.0.4 192.168.0.1 TCP     82      59050 → 179 [SYN] Seq=0 Win=16384 Len=0

MSS=1436

 WS=1

Frame 2408: 82 bytes on wire (656 bits), 82 bytes captured (656 bits) on interface 0
Ethernet II, Src: fa:16:3e:d7:7e:f6 (fa:16:3e:d7:7e:f6), Dst: fa:16:3e:8f:8f:54 (fa:16:3e:8f:8f:54)
Internet Protocol Version 4, Src: 192.168.0.4, Dst: 192.168.0.1
Transmission Control Protocol, Src Port: 59050, Dst Port: 179, Seq: 0, Len: 0
    Source Port: 59050
    Destination Port: 179
    [Stream index: 8]
```

```
    [TCP Segment Len: 0]
    Sequence number: 0    (relative sequence number)
    Acknowledgment number: 0
    Header Length: 48 bytes
    Flags: 0x002 (SYN)
    Window size value: 16384
    [Calculated window size: 16384]
    Checksum: 0x20d7 [unverified]
    [Checksum Status: Unverified]
    Urgent pointer: 0
    Options: (28 bytes), Maximum segment size, Window scale, No-Operation (NOP),
```

TCP MD5 signature

```
, End of Option List (EOL)
        Maximum segment size: 1436 bytes
            Kind: Maximum Segment Size (2)
            Length: 4
```

MSS Value: 1436

```
        Window scale: 0 (multiply by 1)
        No-Operation (NOP)
        TCP MD5 signature
        End of Option List (EOL)
```

来自R1的TCP SYN、ACK:

<#root>

```
! - TCP SYN,ACK sourced from R1

2409   0.004352       192.168.0.1 192.168.0.4 TCP    82     179 → 59050 [SYN, ACK] Seq=0 Ack=1 Win=16384 Le
```

MSS=1436

```
 WS=1

Frame 2409: 82 bytes on wire (656 bits), 82 bytes captured (656 bits) on interface 0
Ethernet II, Src: fa:16:3e:8f:8f:54 (fa:16:3e:8f:8f:54), Dst: fa:16:3e:d7:7e:f6 (fa:16:3e:d7:7e:f6)
Internet Protocol Version 4, Src: 192.168.0.1, Dst: 192.168.0.4
Transmission Control Protocol, Src Port: 179, Dst Port: 59050, Seq: 0, Ack: 1, Len: 0
    Source Port: 179
    Destination Port: 59050
    [Stream index: 8]
    [TCP Segment Len: 0]
    Sequence number: 0    (relative sequence number)
    Acknowledgment number: 1    (relative ack number)
    Header Length: 48 bytes
    Flags: 0x012 (SYN, ACK)
    Window size value: 16384
    [Calculated window size: 16384]
    Checksum: 0xcbf8 [unverified]
    [Checksum Status: Unverified]
    Urgent pointer: 0
    Options: (28 bytes), Maximum segment size, Window scale, No-Operation (NOP),
```

TCP MD5 signature

```
, End of Option List (EOL)
        Maximum segment size: 1436 bytes
            Kind: Maximum Segment Size (2)
            Length: 4
```

MSS Value: 1436

```
        Window scale: 0 (multiply by 1)
        No-Operation (NOP)
        TCP MD5 signature
        End of Option List (EOL)
```

TCP会话详细信息（如R4上所示）— ACTIVE:

<#root>

```
! - as seen from R4 - Active

RP/0/0/CPU0:R4#show tcp detail pcb 0x121542c0
Tue Jan 12 13:27:23.526 UTC
==============================================================
Connection state is ESTAB, I/O status: 0, socket status: 0
Established at Tue Jan 12 13:25:41 2021

PCB 0x121542c0, SO 0x1213c0e4, TCPCB 0x12156010, vrfid 0x60000000,
Pak Prio: Medium, TOS: 192, TTL: 255, Hash index: 359
Local host: 192.168.0.4, Local port: 59050 (Local App PID: 1052958)
Foreign host: 192.168.0.1, Foreign port: 179

Current send queue size in bytes: 0 (max 24576)
Current receive queue size in bytes: 0 (max 32768)  mis-ordered: 0 bytes
Current receive queue size in packets: 0 (max 0)

Timer          Starts    Wakeups         Next(msec)
Retrans             6          1                 0
SendWnd             0          0                 0
TimeWait            0          0                 0
AckHold             3          2                 0
KeepAlive           1          0                 0
PmtuAger            0          0                 0
GiveUp              0          0                 0
Throttle            0          0                 0

   iss: 3299472269  snduna: 3299473445  sndnxt: 3299473445
sndmax: 3299473445  sndwnd: 31646       sndcwnd: 4308
   irs: 3225544359  rcvnxt: 3225545535  rcvwnd: 31665   rcvadv: 3225577200

SRTT: 89 ms,  RTTO: 530 ms,  RTV: 441 ms,  KRTT: 0 ms
minRTT: 19 ms,  maxRTT: 239 ms

ACK hold time: 200 ms, Keepalive time: 0 sec, SYN waittime: 30 sec
Giveup time: 0 ms, Retransmission retries: 0, Retransmit forever: FALSE
Connect retries remaining: 30, connect retry interval: 30 secs

State flags: none
Feature flags:
```

MD5

, Win Scale, Nagle,

Path MTU

Request flags: Win Scale

Datagrams (in bytes): MSS 1436, peer MSS 1436, min MSS 1436, max MSS 1436

Window scales: rcv 0, snd 0, request rcv 0, request snd 0
Timestamp option: recent 0, recent age 0, last ACK sent 0
Sack blocks {start, end}: none
Sack holes {start, end, dups, rxmit}: none

Socket options: SO_REUSEADDR, SO_REUSEPORT, SO_NBIO
Socket states: SS_ISCONNECTED, SS_PRIV
Socket receive buffer states: SB_DEL_WAKEUP
Socket send buffer states: SB_DEL_WAKEUP
Socket receive buffer: Low/High watermark 1/32768
Socket send buffer   : Low/High watermark 2048/24576, Notify threshold 0

PDU information:
 #PDU's in buffer: 0
FIB Lookup Cache: IFH: 0x40 PD ctx: size: 0 data:
  Num Labels: 0 Label Stack:

RP/0/0/CPU0:R4#

R1上看到的TCP会话详细信息 — PASSIVE:

<#root>

```
! - as seen from R1 - Passive

RP/0/0/CPU0:R1#show tcp detail pcb 0x121560ec
Tue Jan 12 13:25:59.310 UTC
==============================================================
Connection state is ESTAB, I/O status: 0, socket status: 0
Established at Tue Jan 12 13:25:31 2021

PCB 0x121560ec, SO 0x121556d4, TCPCB 0x121575bc, vrfid 0x60000000,
Pak Prio: Medium, TOS: 192, TTL: 255, Hash index: 359
Local host: 192.168.0.1, Local port: 179 (Local App PID: 983326)
Foreign host: 192.168.0.4, Foreign port: 59050

Current send queue size in bytes: 0 (max 24576)
Current receive queue size in bytes: 0 (max 32768)  mis-ordered: 0 bytes
Current receive queue size in packets: 0 (max 0)

Timer          Starts    Wakeups         Next(msec)
Retrans             3          0                  0
SendWnd             0          0                  0
TimeWait            0          0                  0
AckHold             3          2                  0
KeepAlive           1          0                  0
```

```
PmtuAger            0          0               0
GiveUp              0          0               0
Throttle            0          0               0

   iss: 3225544359  snduna: 3225545516  sndnxt: 3225545516
sndmax: 3225545516  sndwnd: 31684       sndcwnd: 4308
   irs: 3299472269  rcvnxt: 3299473426  rcvwnd: 31665   rcvadv: 3299505091

SRTT: 37 ms,  RTTO: 300 ms,  RTV: 244 ms,  KRTT: 0 ms
minRTT: 9 ms,  maxRTT: 239 ms

ACK hold time: 200 ms, Keepalive time: 0 sec, SYN waittime: 30 sec
Giveup time: 0 ms, Retransmission retries: 0, Retransmit forever: FALSE
Connect retries remaining: 0, connect retry interval: 0 secs

State flags: none
Feature flags:

MD5

, Win Scale, Nagle,

Path MTU


Request flags: Win Scale


Datagrams (in bytes): MSS 1436, peer MSS 1436, min MSS 1460, max MSS 1460



Window scales: rcv 0, snd 0, request rcv 0, request snd 0
Timestamp option: recent 0, recent age 0, last ACK sent 0
Sack blocks {start, end}: none
Sack holes {start, end, dups, rxmit}: none

Socket options: SO_REUSEADDR, SO_REUSEPORT, SO_NBIO
Socket states: SS_ISCONNECTED, SS_PRIV
Socket receive buffer states: SB_DEL_WAKEUP
Socket send buffer states: SB_DEL_WAKEUP
Socket receive buffer: Low/High watermark 1/32768
Socket send buffer   : Low/High watermark 2048/24576, Notify threshold 0

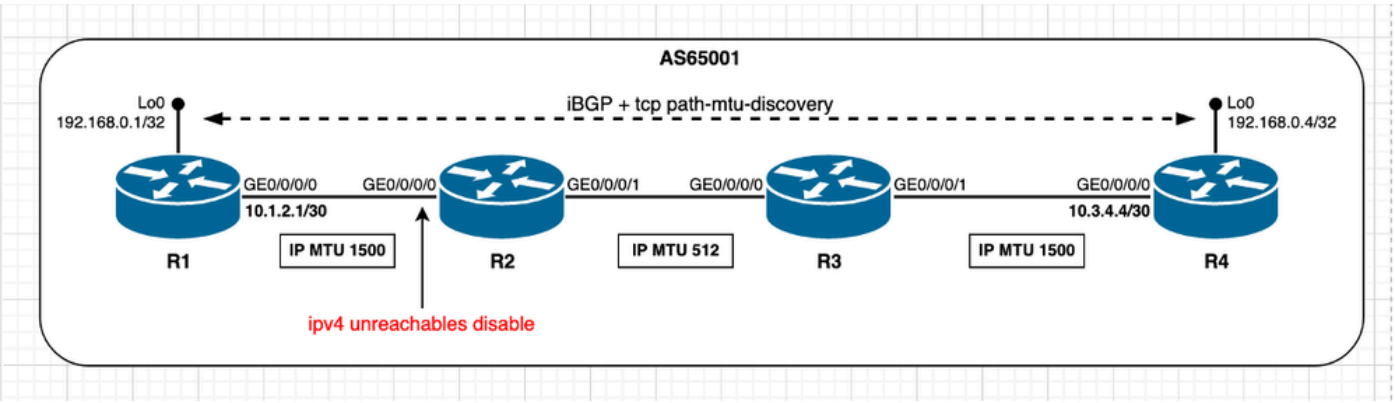
PDU information:
 #PDU's in buffer: 0
FIB Lookup Cache:  IFH: 0x40  PD ctx: size: 0  data:
  Num Labels: 0  Label Stack:

RP/0/0/CPU0:R1#
```

PMTUD — 黑洞检测

如PMTUD — 路径分段具有更低的IP MTU部分中所述，启用时的TCP PMTUD由接收ICMP(目的地不可达 — 类型3;需要分段 — 代码4)消息。在某些情况下，由于某些原因未收到这些消息，因此不会触发PMTUD。在这种情况下，不会获取TCP对等体之间路径的最低IP MTU。如果IP数据包设置了DF位，并且其大小高于最低IP MTU路径段，则此类场景将引入潜在的黑洞。这些数据包将以静默方式丢弃。

本节旨在重点介绍Cisco IOS XR如何检测此类潜在黑洞场景并对其执行操作。为此，在R2接口GE0/0/0/0上禁用IPv4不可达功能，如下一幅图像和CLI输出所示。



映像3.5 — 禁用R1/R4和R2 IPv4不可达的PMTUD。

在R2上禁用IPv4不可达：

<#root>

```
!- R2 - IP unreachables is disabled

RP/0/0/CPU0:R2#show run interface gigabitEthernet 0/0/0/0
Thu May 13 12:09:45.483 UTC
interface GigabitEthernet0/0/0/0
 ipv4 address 10.1.2.2 255.255.255.252
ipv4 unreachables disable


!

RP/0/0/CPU0:R2#show ipv4 interface gigabitEthernet 0/0/0/0
Thu May 13 12:10:04.112 UTC
GigabitEthernet0/0/0/0 is Up, ipv4 protocol is Up
  Vrf is default (vrfid 0x60000000)
  Internet address is 10.1.2.2/30
  MTU is 1514 (1500 is available to IP)
  Helper address is not set
  Multicast reserved groups joined: 224.0.0.2 224.0.0.1 224.0.0.5
      224.0.0.6
  Directed broadcast forwarding is disabled
  Outgoing access list is not set
  Inbound  common access list is not set, access list is not set
  Proxy ARP is disabled
  ICMP redirects are never sent


ICMP unreachables are never sent


  ICMP mask replies are never sent
  Table Id is 0xe0000000
```

Cisco IOS XR处理此黑洞场景的方法是重传相同数据包两次，如果仍不成功，即未收到预期的TCP

ACK，然后重试，但使用RFC1191 — 路径MTU发现中记录的下一个较低定义的高原值(请参阅PMTUD — 路径段具有较低的IP MTU部分，了解平台列表)。 总之，Cisco IOS XR假定数据包可能由于大小而丢弃到其目的地的路径中的某个位置，并尝试通过数据包重新传输来解决这个问题。在节点R1接口处捕获的数据包以及debug tcp pmtud的输出中，可以观察到此行为。

R1上的IOS-XR黑洞检测：

<#root>

```
! - at R1
! - Original BGP Update message is sent
! - Note IP Total Length of 1116 bytes and TCP Segment Length of 1076 bytes
! - R2 filters such packet and send and ICMP error message towards R1 which triggers PMTUD
! - But because IPv4 unreachables are disabled at R2 GE0/0/0/0 ICMP message is not sent
! - Hence BGP message is silently filtered at R2

562    7.638774       192.168.0.1 192.168.0.4 BGP    1130   UPDATE Message, KEEPALIVE Message

Frame 562: 1130 bytes on wire (9040 bits), 1130 bytes captured (9040 bits) on interface 0
Ethernet II, Src: fa:16:3e:42:18:05 (fa:16:3e:42:18:05), Dst: fa:16:3e:5c:f1:80 (fa:16:3e:5c:f1:80)
Internet Protocol Version 4, Src: 192.168.0.1, Dst: 192.168.0.4
    0100 .... = Version: 4
    .... 0101 = Header Length: 20 bytes (5)
    Differentiated Services Field: 0xc0 (DSCP: CS6, ECN: Not-ECT)


Total Length: 1116


    Identification: 0x4a37 (18999)
    Flags: 0x02 (Don't Fragment)
        0... .... = Reserved bit: Not set


.1.. .... = Don't fragment: Set


        ..0. .... = More fragments: Not set
    Fragment offset: 0
    Time to live: 255
    Protocol: TCP (6)
    Header checksum: 0x229b [validation disabled]
    [Header checksum status: Unverified]
    Source: 192.168.0.1
    Destination: 192.168.0.4
    [Source GeoIP: Unknown]
    [Destination GeoIP: Unknown]
Transmission Control Protocol, Src Port: 179, Dst Port: 57082, Seq: 318, Ack: 251, Len: 1076
Border Gateway Protocol - UPDATE Message
Border Gateway Protocol - KEEPALIVE Message
<snip>

! - at R1
! - No TCP ACK is received
! - Packet retransmission is attempted (2 attempts)
! - Note initial MSS value is of 1460 bytes

563    0.560058       192.168.0.1 192.168.0.4 TCP    1130   [TCP Retransmission] 179 → 57082 [PSH, ACK] S
564    1.101367       192.168.0.1 192.168.0.4 TCP    1130   [TCP Retransmission] 179 → 57082 [PSH, ACK] S
```

```
! - at R1
! - Still no TCP ACK received; previous retransmissions failed
! - Next lower plateau value is attempted - 1492 bytes
! - Packet retransmission is attempted (2 attempts)

RP/0/0/CPU0:May 13 10:20:44.251 UTC: tcp[399]: [t1] PCB 0x15392224: Trying next lower MTU: 1452

567    1.850294       192.168.0.1 192.168.0.4 TCP    1130   [TCP Retransmission] 179 → 57082 [PSH, ACK] S
568    1.111361       192.168.0.1 192.168.0.4 TCP    1130   [TCP Retransmission] 179 → 57082 [PSH, ACK] S

! - at R1
! - Still no TCP ACK received; previous retransmissions failed
! - Next lower plateau value is attempted - 1006 bytes
! - Packet retransmission is attempted (2 attempts)

RP/0/0/CPU0:May 13 10:20:47.560 UTC: tcp[399]: [t1] PCB 0x15392224: Trying next lower MTU: 966

569    2.198327       192.168.0.1 192.168.0.4 TCP    1020   [TCP Retransmission] 179 → 57082 [ACK] Seq=31
570    1.109602       192.168.0.1 192.168.0.4 TCP    1020   [TCP Retransmission] 179 → 57082 [ACK] Seq=31

! - at R1
! - Still no TCP ACK received; previous retransmissions failed
! - Next lower plateau value is attempted - 508 bytes
! - Original information (TCP Length of 1076 bytes) is split in three distinct packets
! - TCP Segment Lengths 468 + 468 + 140 = 1076
! - TCP ACK is received from peer R4

RP/0/0/CPU0:May 13 10:20:50.870 UTC: tcp[399]: [t1] PCB 0x15392224: Trying next lower MTU: 468

571    2.205552       192.168.0.1 192.168.0.4 TCP    522    [TCP Retransmission] 179 → 57082 [ACK] Seq=31
```

Len=468

```
573    0.004254       192.168.0.1 192.168.0.4 TCP    522    [TCP Retransmission] 179 → 57082 [ACK] Seq=78
```

Len=468

```
574    0.002724       192.168.0.1 192.168.0.4 TCP    194    [TCP Retransmission] 179 → 57082 [PSH, ACK] S
```

Len=140

```
! - Peer R4 TCP ACK is received

575    0.223172       192.168.0.4 192.168.0.1 TCP    54     57082 → 179 [ACK] Seq=251 Ack=1394 Win=31469 Le
```

关于此翻译
思科采用人工翻译与机器翻译相结合的方式将此文档翻译成不同语言，希望全球的用户都能通过各自的语言得到支持性的内容。

请注意：即使是最好的机器翻译，其准确度也不及专业翻译人员的水平。

Cisco Systems, Inc. 对于翻译的准确性不承担任何责任，并建议您总是参考英文原始文档（已提供链接）。