Verificar la Detección de MTU de Trayectoria en Cisco IOS XR y BGP

Contenido

Introducción

Este documento describe la detección de la unidad máxima de transmisión (MTU) de la ruta del protocolo de control de transmisión (TCP) (PMTUD) en los dispositivos Cisco IOS® XR.

Antecedentes

El mecanismo PMTUD intenta determinar el tamaño de paquete de protocolo de Internet (IP) más grande que no requiere fragmentación en ningún lugar a lo largo de la ruta entre dos hosts. El valor establecido es designado Path MTU y es igual a un mínimo de los valores MTU en cada salto. Si considera la MTU de Trayectoria cuando transmite información, le permite sacar el máximo partido de la capacidad de la red y evitar la fragmentación y la eficiencia de transmisión. La mecánica y la implementación de PMTUD se introducen en un conjunto diverso de escenarios con el uso del protocolo de gateway fronterizo (BGP) como protocolo de cliente que revela gradualmente el comportamiento de PMTUD.

TCP PMTUD y TCP MSS

TCP aprovecha el resultado de la PMTUD para influir en el tamaño máximo de segmento (MSS) local, lo que significa que se adapta dinámicamente a la MTU de ruta descubierta. Por lo tanto, antes de pasar a la PMTUD, puede revisar rápidamente el tamaño máximo de segmento (MSS)

de TCP y comprender lo que significa y su propósito.

Según la definición original de MSS de RFC879: Se puede establecer la definición de la opción MSS: El número máximo de octetos de datos que puede recibir el remitente de esta opción TCP en segmentos TCP sin opciones de encabezado TCP transmitidas en datagramas IP sin opciones de encabezado IP.

Para aclarar algunos aspectos y proporcionar asesoramiento a los encargados de la ejecución, RFC6691 destaca cómo se debe calcular el valor de MSS:

Cuando calcule el valor que se debe poner en la opción TCP MSS, el valor de MTU debe reducirse solamente en el tamaño de los encabezados IP y TCP fijos y no debe reducirse para tener en cuenta ninguna opción IP o TCP posible; a la inversa, el remitente DEBE reducir la longitud de datos TCP para tener en cuenta las opciones de IP o TCP que incluye en los paquetes que envía.

Se puede extraer una definición más detallada de MSS de la Guía de Configuración de Ruteo para Cisco ASR 9000 Series Routers, IOS XR versión 6.7.x:

MSS es la mayor cantidad de datos que un equipo o un dispositivo de comunicación puede recibir en un solo segmento TCP no fragmentado. Todas las sesiones TCP están limitadas por un límite en el número de bytes que se pueden transportar en un solo paquete; este límite es MSS. TCP divide los paquetes en fragmentos de una cola de transmisión antes de pasar los paquetes a la capa IP.

El valor TCP MSS depende de la MTU de una interfaz, que es la longitud máxima de datos que puede transmitir un protocolo en una instancia. La longitud máxima del paquete TCP la determina tanto la MTU de la interfaz saliente en el dispositivo de origen como el MSS anunciado por el dispositivo de destino durante el proceso de configuración TCP. Cuanto más cerca esté el MSS de la MTU, más eficiente será la transferencia de mensajes BGP. Cada dirección del flujo de datos puede utilizar un valor MSS diferente.

¿Cuál sería entonces el valor que TCP debería considerar para MSS en una sesión TCP dada? ¿Y cómo se calcula?

Para los valores predeterminados según RFC879, tiene: Los hosts no deben enviar datagramas mayores de 576 octetos a menos que tengan conocimiento específico de que el host de destino está preparado para aceptar datagramas más grandes. EL TAMAÑO MÁXIMO DEL SEGMENTO TCP ES EL TAMAÑO MÁXIMO DE DATAGRAMA IP DE 40.

El tamaño máximo de datagrama IP predeterminado es 576.

El tamaño máximo de segmento predeterminado de TCP es 536.

Esto toma en consideración un valor de MTU IP de 576 bytes. Pero si ignora el valor de MTU IP real, el cálculo de MSS de TCP se puede resumir de la siguiente manera:

- Peer activo: calcula y envía el MSS inicial con el paquete SYN.

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

- Pasivo Peer - calcula el MSS inicial, compara con el MSS recibido del Peer Activo, y envía SYN, ACK con el menor de estos valores MSS.

```
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.
```

No hay negociación con respecto al valor de la opción MSS. Cada nodo determina su propio valor y anuncia lo mismo en el establecimiento de la sesión TCP. Queda claro que si el valor de MTU de IP considerado para el cálculo de MSS se puede derivar de PMTUD, el valor de MSS se puede adaptar al valor más efectivo para una MTU de Trayectoria dada. El comportamiento de Cisco IOS XR tiene algunos detalles con respecto al cálculo de MSS y la función PMTUD resumidos aquí.

PMTUD está inhabilitado de forma predeterminada en Cisco IOS XR:

- El cálculo de MSS inicial local considera la MTU de IP de la siguiente manera: Si los peers conectados directamente - considere la MTU IP de la interfaz de egreso.Si los peers no conectados directamente - considere la MTU IP de 1280 bytes.Las opciones TCP configuradas influyen en el valor de MSS.

Cuando PMTUD está habilitado en Cisco IOS XR:

- El cálculo de MSS inicial local considera la MTU de IP de la siguiente manera: Independientemente de los peers conectados directa/no directamente: considere la MTU IP de la interfaz de salida.Las opciones TCP configuradas influyen en el valor de MSS.

Hay detalles adicionales sobre la mecánica y aplicación de la PMTUD que deben tenerse en cuenta y que este documento presenta mediante ejemplos prácticos resumidos en la siguiente tabla. Esta tabla también presenta MTU IP de peers TCP activos y pasivos, así como valores MSS seleccionados para cada escenario considerado.

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

Escenarios: TCP PMTUD desactivado

Usar valores predeterminados de MTU

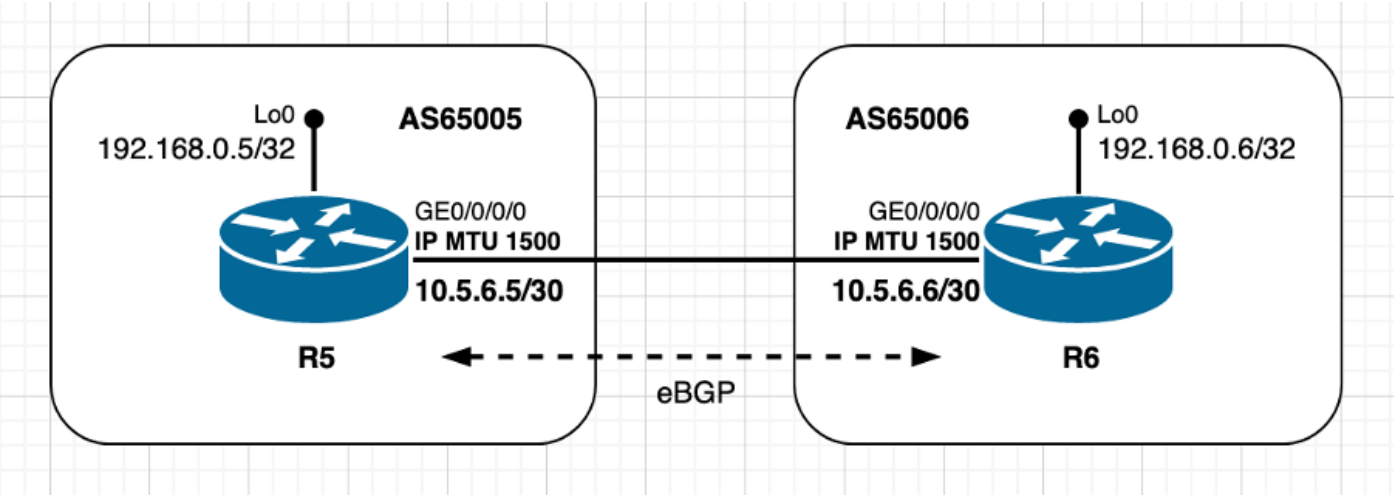

Imagen 2.1. Con valores de MTU predeterminados

En el caso de los peers eBGP que se muestran en la Imagen 2.1 R6 administra la conexión TCP, esto significa que desempeña el rol activo e inicia la sesión TCP con R5 en el puerto de destino 179. Los pares están conectados directamente y ambos utilizan los valores de MTU de IP predeterminados en las interfaces respectivas. Basándose en la información compartida al inicio de este documento, el cálculo de MSS en este escenario se puede resumir de la siguiente manera:

- Ambos nodos utilizan una MTU IP predeterminada de 1500 bytes
- La detección de MTU de trayecto TCP está inhabilitada de forma predeterminada
- Los peers TCP están conectados directamente R6 administra la conexión BGPR6 envía SYN con MSS de 1460 bytes 1500 (MTU de IP de interfaz) - 20 (minTCP_H) - 20 (minIP_H)R5 envía SYN, ACK con MSS de 1460 bytes Envía el valor inferior de [MSS recibido; MSS inicial local]MSS recibido 1460 bytes; MSS inicial local 1460 bytesEl valor MSS más bajo se utiliza en ambos pares

Detalles de la sesión TCP tal como se ven en R6 - ACTIVE:

```
! – 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
```

Detalles de la sesión TCP tal como se ven en R5 - PASSIVE:

```
! – 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#
```

Usar valor de MTU no predeterminado - Peer TCP activo

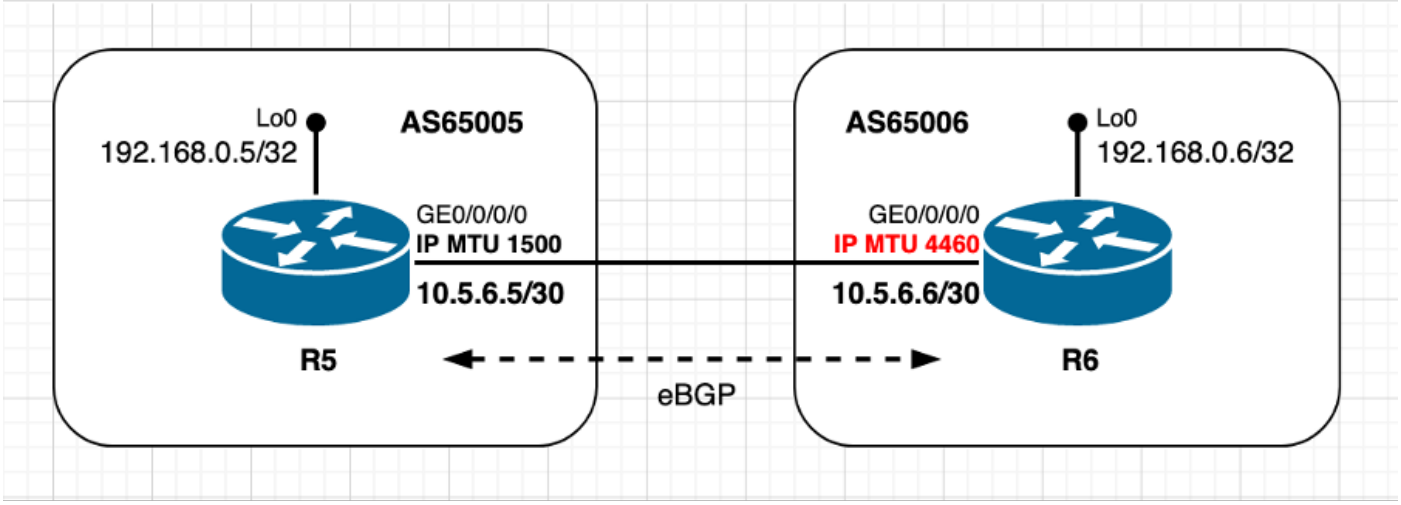

Imagen 2.2: el par ACTIVE utiliza un valor de MTU no predeterminado

Este escenario es esencialmente el mismo que el anterior, con la única diferencia que el peer TCP R6 activo ahora utiliza un valor de MTU IP no predeterminado. Observe cómo el cálculo inicial y la decisión sobre el valor de MSS es realizada por el par TCP pasivo R5. El cálculo de TCP MSS en este escenario se puede resumir de la siguiente manera:

- R6 utiliza una MTU IP no predeterminada de 4460 bytes
- La detección de MTU de trayecto TCP está inhabilitada de forma predeterminada
- Los peers TCP están conectados directamente R6 gestiona la conexión BGPR6 envía SYN con MSS de 4420 bytes 4460 (MTU de IP de interfaz) - 20 (minTCP_H) - 20 (minIP_H)R5 send SYN, ACK con MSS de 1460 bytes envía el valor inferior de [MSS recibido; MSS inicial local]MSS recibido 4420 bytes; MSS inicial local 1460 bytesEl valor MSS más bajo se utiliza en ambos pares

TCP SYN originado en R6:

```
! – 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)
```

TCP SYN, ACK originado en R5:

```
! – 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=16384 Len=0 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)
```

Detalles de la sesión TCP tal como se ven en R6 - ACTIVE:

```
! - 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#
```

Detalles de la sesión TCP tal como se ven en R5 - PASSIVE:

```
! - 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#
```

Utilizar valor de MTU no predeterminado - Peer TCP pasivo

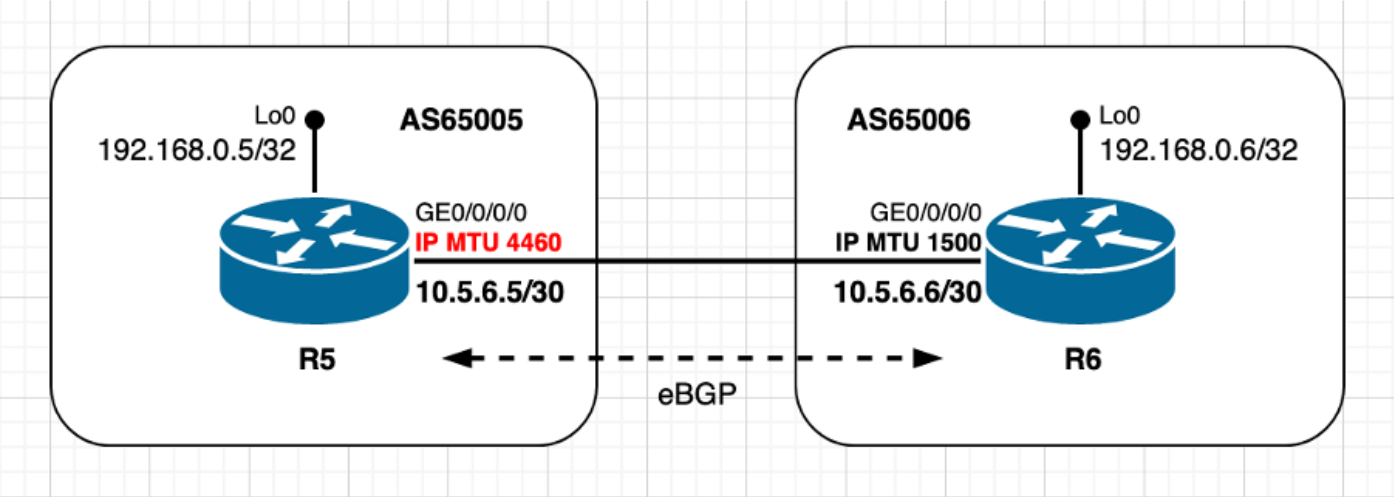

Imagen 2.3: el par PASSIVE utiliza un valor de MTU no predeterminado.

Con el mismo escenario eBGP, pero ahora con el peer TCP R5 pasivo configurado con la MTU IP no predeterminada y el peer TCP R6 activo con el valor de MTU IP predeterminado. Como en el escenario anterior, tenga en cuenta cómo el valor de MSS es seleccionado por el peer pasivo R5. El cálculo de TCP MSS en este escenario se puede resumir de la siguiente manera:

- R5 utiliza una MTU IP no predeterminada de 4460 bytes
- La detección de MTU de trayecto TCP está inhabilitada de forma predeterminada
- Los peers TCP están conectados directamente R6 gestiona la conexión BGPR6 envía SYN con MSS de 1460 bytes 1500 (MTU de IP de interfaz) - 20 (minTCP_H) - 20 (minIP_H)R5 send SYN, ACK con MSS de 1460 bytes envía el valor inferior de [MSS recibido; MSS inicial local]MSS recibido 1460 bytes; MSS inicial local de 4420 bytesEl valor MSS más bajo se utiliza en ambos pares

TCP SYN originado en R6:

```
! – 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)
```

TCP SYN, ACK originado en R5:

```
! – 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=16384 Len=0 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)
```

Detalles de la sesión TCP tal como se ven en R6 - ACTIVE:

```
! - 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#
```

Detalles de la sesión TCP tal como se ven en R5 - PASSIVE:

```
! - 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#
```

Usar opciones TCP - XR activo

Como se mencionó anteriormente en este documento, el uso de opciones TCP (como TCP MD5, TCP selective-ack o TCP timestamps) influye en el cálculo de MSS, ya que estas opciones llevan a bytes adicionales que se contabilizarán en el cálculo de MSS.

Esta sección, así como la siguiente, tiene como objetivo ilustrar el cálculo de MSS realizado por los peers cuando hay opciones de TCP. La opción de autenticación TCP MD5 se utiliza como ejemplo. Consulte el escenario de referencia en Imágenes 2.4 como se muestra en la imagen.

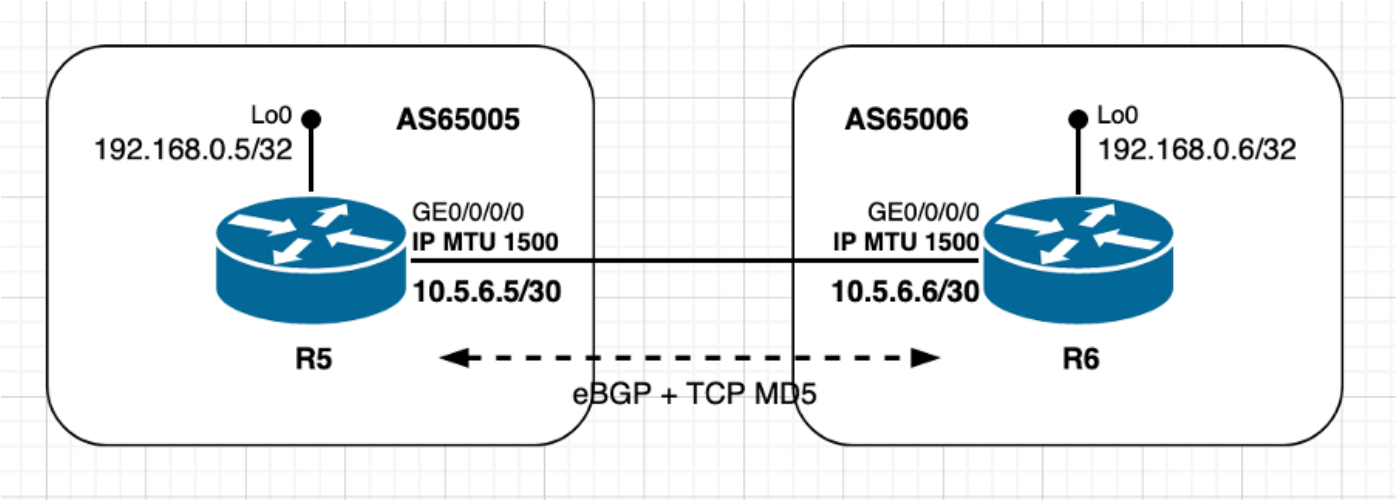

Imagen 2.4 - Use TCP Options (MD5) - XR Active.

En este escenario, ambos peers utilizan valores de MTU de IP predeterminados, están conectados directamente y el par R6 reproduce el rol activo de TCP. Como ya se ha compartido, la configuración y el uso de la cuenta de autenticación MD5 de TCP para obtener una sobrecarga adicional. El cálculo de TCP MSS en este escenario particular se puede resumir de la siguiente manera:

- Ambos nodos utilizan una MTU IP predeterminada de 1500 bytes
- La detección de MTU de trayecto TCP está inhabilitada de forma predeterminada
- Los peers TCP están conectados directamente
- Autenticación TCP MD5 habilitada en ambos nodos R6 gestiona la conexión BGPR6 envía SYN con MSS de 1436 bytes 1500 (MTU de IP de interfaz) - 20 (minTCP_H) - 20 (minIP_H) - 24 bytes (sobrecarga de opciones TCP IOS XR)R5 send SYN, ACK con MSS de 1436 bytes envía el valor inferior de [MSS recibido; MSS inicial local]MSS recibido 1436 bytes; MSS inicial local 1460 bytesEl valor MSS más bajo se utiliza en ambos pares

Como se ve en el resumen, el comportamiento de Cisco IOS XR no es estrictamente según RFC 879, y RFC 6691, que indican que las opciones TCP no se deben contabilizar en el cálculo de MSS.

La cuenta Cisco IOS XR de un factor adicional en la **longitud del encabezado tcp** se documenta con mayor detalle en el Id. de bug Cisco CSCvf20166:

"(...)Cuando XR está iniciando la conexión BGP, BGP primero crea el socket y luego establece las opciones de socket incluyendo **MD5**. Esto hace que la **longitud del encabezado de la opción tcp sea 24**. Y por lo tanto el MSS inicial se convierte en 1500 - 40 - 24 = 1436. Esto se envía al par y el par utiliza min(1436, 1460) = 1436.(...)

TCP SYN originado en R6:

```
! – 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)
```

TCP SYN, ACK originado en R5:

```
! – 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=16384 Len=0 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)
```

Detalles de la sesión TCP tal como se ven en R6 - ACTIVE:

```
! - 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#
```

Detalles de la sesión TCP tal como se ven en R5 - PASSIVE:

```
! - 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#
```

Se puede observar un comportamiento similar con otras opciones de TCP que, cuando se configuran, representan una sobrecarga adicional e influyen en el cálculo de MSS en Cisco IOS XR. Considere el mismo escenario y estos ejemplos que documentan el cálculo de MSS cuando se configuran las marcas de tiempo TCP y las opciones de paquete selectivo de TCP.

Detalles de la sesión TCP tal como se ven en R6 - ACTIVE - con opciones TCP indicación de fecha y hora y opciones de paquete selectivo configuradas:

```
! - 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>
```

Usar opciones TCP - XR pasivo

En el escenario anterior probablemente haya notado el comportamiento distintivo del nodo XR del IOS de Cisco cuando se encuentra en una función pasiva con respecto al cálculo MSS inicial. El nodo no tiene en cuenta la **longitud del encabezado de la opción tcp**. Este escenario apunta a resaltar este comportamiento distintivo que también se describe en la ID de bug de Cisco :

"(...): Cuando el par inicia la conexión, envía el MSS inicial como 1460. XR TCP crea socket, pcb, etc. y luego realiza dos acciones en un orden determinado:

- Primero, computa el MSS inicial después de restar **la longitud del encabezado de la opción tcp**. Esta es '0' ya que la opción MD5 todavía no se ha heredado a este socket del socket de escucha.

- Luego, hereda el 'MD5' y otras opciones y esto hace que la 'longitud de bytes del encabezado de opción' sea de 24.

Por lo tanto, en este caso, el TCP XR envía el MSS inicial como 1460 y, por lo tanto, ambos lo utilizan. (…)“

En este escenario, aunque el peer R8 de TCP activo es un nodo de Cisco IOS, este hecho no introduce ninguna diferencia o detalles específicos sobre lo que el escenario pretende resaltar. Sin embargo, y de manera interesante, observe que de manera diferente a Cisco IOS XR como se mostró con el escenario de sección anterior, aquí el peer TCP R8 activo no considera las opciones TCP en el cálculo MSS inicial.

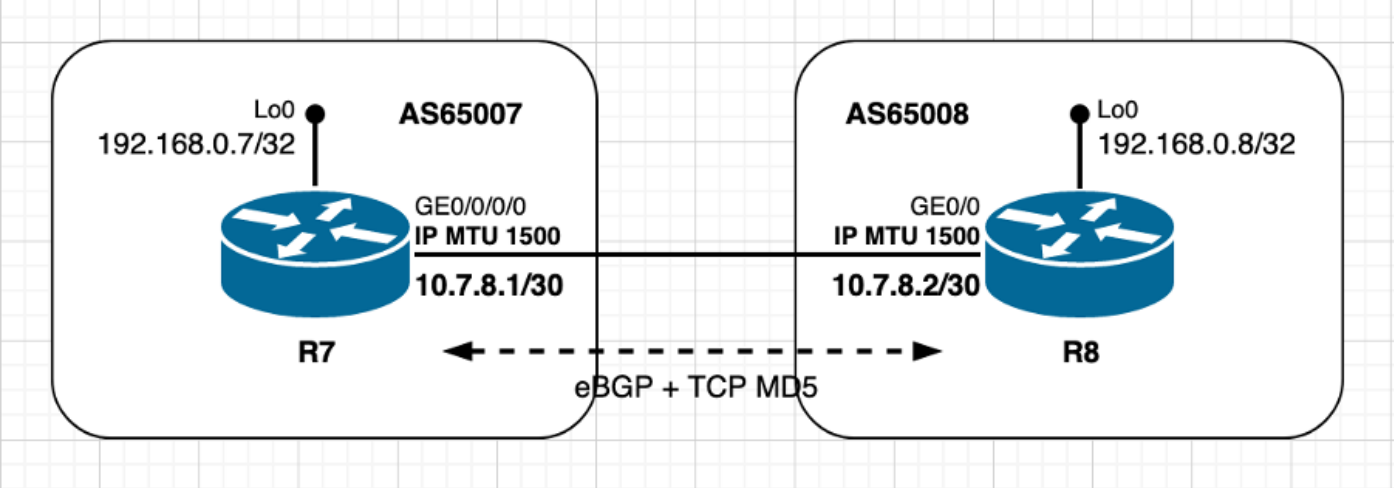

Imagen 2.5 - Use TCP Options (MD5) - XR Passive.

Ambos peers utilizan valores de MTU IP predeterminados y están conectados directamente. El par R8 del IOS de Cisco desempeña un papel activo. El cálculo de TCP MSS en este escenario se puede resumir de la siguiente manera:

- Ambos nodos utilizan una MTU IP predeterminada de 1500 bytes
- La detección de MTU de trayectoria TCP está inhabilitada de forma predeterminada en Cisco IOS XR R7
- La detección de MTU de trayectoria TCP está habilitada de forma predeterminada en el Cisco IOS R8
- Los pares TCP están conectados directamente
- Autenticación TCP MD5 habilitada en ambos nodos IOS R8 administra la conexión BGPIOS R8 envía SYN con MSS de 1460 bytes 1500 (MTU de IP de interfaz) - 20 (minTCP_H) - 20 (minIP_H)IOS XR R7 envía SYN, ACK con MSS de 1460 bytes envía el valor inferior de [MSS recibido; MSS inicial local]MSS recibido 1460 bytes; MSS inicial local 1460 bytesEl valor MSS más bajo se utiliza en ambos pares

TCP SYN originado en R8 - Cisco IOS:

```
! – 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)
```

TCP SYN, ACK originado en R7 - Cisco IOS XR:

```
! – 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=16384 Len=0 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)
```

Detalles de la sesión TCP tal como se ven en R8 - Cisco IOS - ACTIVE:

```
! – 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 data bytes: 213

 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#
```

Detalles de la sesión TCP tal como se ven en R7 - Cisco IOS XR - PASIVO:

```
! – 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#
```

Peers TCP no conectados directamente

Cuando los peers no están conectados directamente, la forma en que se realiza el cálculo inicial de TCP MSS cambia como se describió anteriormente en la sección introductoria de este documento. El escenario de una sesión iBGP con todos los peers configurados con valores de MTU de IP predeterminados se utiliza para recorrer el cálculo de MSS.

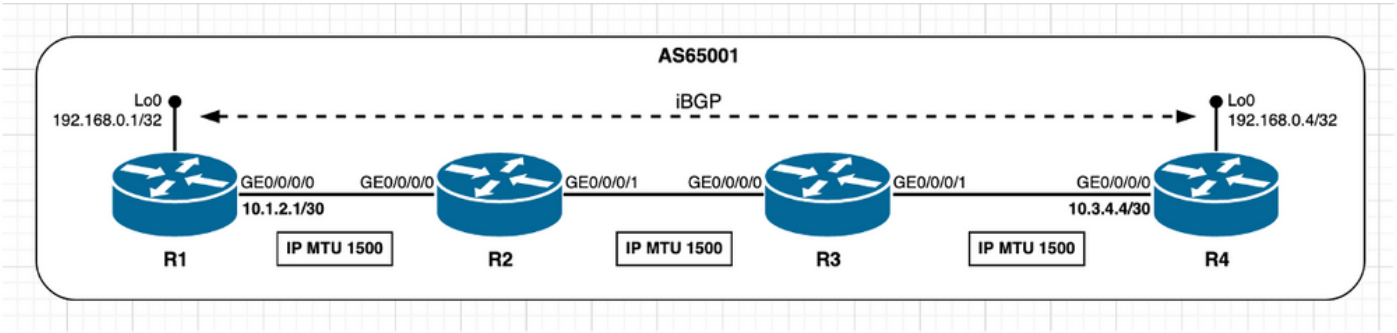

Imagen 2.6 - Peers TCP no conectados directamente - iBGP.

El aspecto importante a tener en cuenta es que cuando se inhabilita TCP Path MTU Discovery y los peers no están conectados directamente, por diseño, Cisco IOS XR utiliza un valor de MTU IP fija de 1280 bytes.

En la imagen anterior R4 desempeña un papel activo y administra la conexión TCP, R4 abre la sesión TCP con R1 en el puerto de destino 179. Ambos nodos utilizan el valor de MTU IP predeterminado en sus interfaces. El cálculo de MSS en este escenario se puede resumir de la siguiente manera:

- Todos los nodos utilizan una MTU IP predeterminada de 1500 bytes
- La detección de MTU de trayecto TCP está inhabilitada de forma predeterminada
- Los peers TCP no están conectados directamente R4 administra la conexión BGPR4 envía SYN con MSS de 1240 bytes La MTU de interfaz no se considera cuando los peers no están conectados directamente y la Detección de MTU de trayectoria TCP está inhabilitadaSegún el diseño XR de Cisco IOS, 1280 bytes se considera TCP_DEFAULT_MTU1280 (TCP_DEFAULT_MTU) - 20 (minTCP_H) - 20 (minIP_H)R1 envía SYN, ACK con MSS de 1240 bytes Envía el valor inferior de [MSS recibido; MSS inicial local]MSS recibido 1240 bytes; MSS inicial local 1240 bytesEl valor MSS más bajo se utiliza en ambos pares

TCP SYN originado en R4:

```
! – 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)
```

TCP SYN, ACK originado en R1:

```
! – 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 Len=0 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)
```

Detalles de la sesión TCP tal como se ven en R4 - ACTIVE:

```
! – 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#
```

Detalles de la sesión TCP tal como se ven en R1 - PASIVO:

```
! – 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#
```

Peers TCP no conectados directamente: utilice opciones TCP (MD5)

Para el escenario de peer no conectado directamente y en uso de la autenticación TCP MD5, no hay diferencia fundamental con los casos de prueba o escenarios anteriores ya descritos. Como se ha visto anteriormente con la autenticación TCP MD5, Cisco IOS XR considera sobrecarga adicional y el valor MSS inicial refleja lo mismo. Consulte las secciones anteriores Use TCP Options - XR Active y Use TCP Options - XR Passive para obtener más información sobre la influencia de las opciones TCP en el cálculo de TCP MSS.

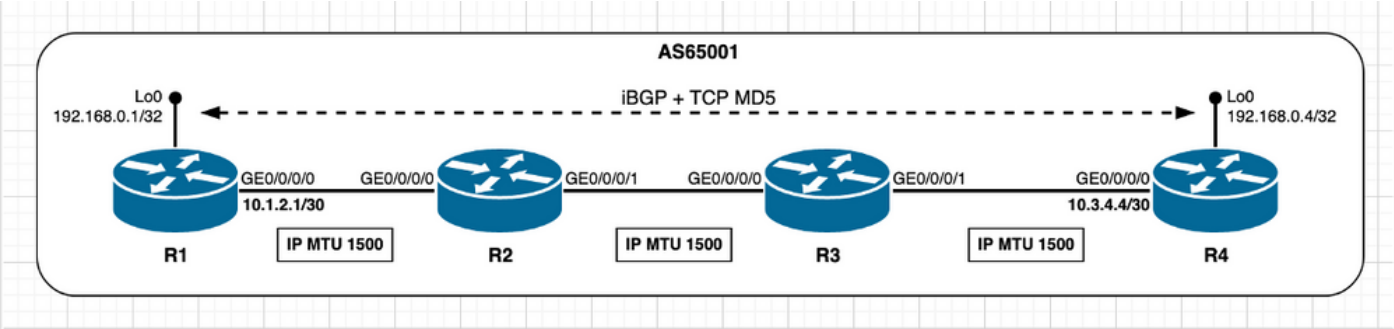

Imagen 2.7 - Peers TCP no conectados directamente - iBGP + TCP MD5.

El cálculo de TCP MSS en este escenario se puede resumir de la siguiente manera:

- Todos los nodos utilizan una MTU IP predeterminada de 1500 bytes
- La detección de MTU de trayecto TCP está inhabilitada de forma predeterminada
- Los pares TCP no están conectados directamente R4 administra la conexión BGPEl destino R1 no está conectado directamenteR4 envía SYN con MSS de 1216 bytes La MTU de interfaz no se considera cuando los peers no están conectados directamente y la Detección de MTU de trayectoria TCP está inhabilitadaSegún el diseño, 1280 bytes se considera TCP_DEFAULT_MTU1280 (TCP_DEFAULT_MTU) - 20 (minTCP_H) - 20 (minIP_H) - 24

bytes (sobrecarga de opciones TCP IOS XR)R1 envía SYN, ACK con MSS de 1216 bytes Envía el valor inferior de [MSS recibido; MSS inicial local]MSS recibido 1216 bytes; MSS inicial local 1240 bytesEl valor MSS más bajo se utiliza en ambos pares

TCP SYN originado en R4:

```
! – 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)
```

TCP SYN, ACK originado en R1:

```
! – 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 Len=0 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)
```

Detalles de la sesión TCP tal como se ven en R4 - ACTIVE:

```
! – 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#
```

Detalles de la sesión TCP tal como se ven en R1 - PASIVO:

```
! – 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#
```

Peers TCP no conectados directamente: el segmento de trayectoria tiene una MTU IP inferior

Con el siguiente escenario, el objetivo es observar y concluir lo que sucede si hay un segmento de trayectoria intermedia con una MTU de IP inferior mientras está en condición predeterminada, esto significa que la PMTUD de TCP está inhabilitada. Consulte esta imagen.

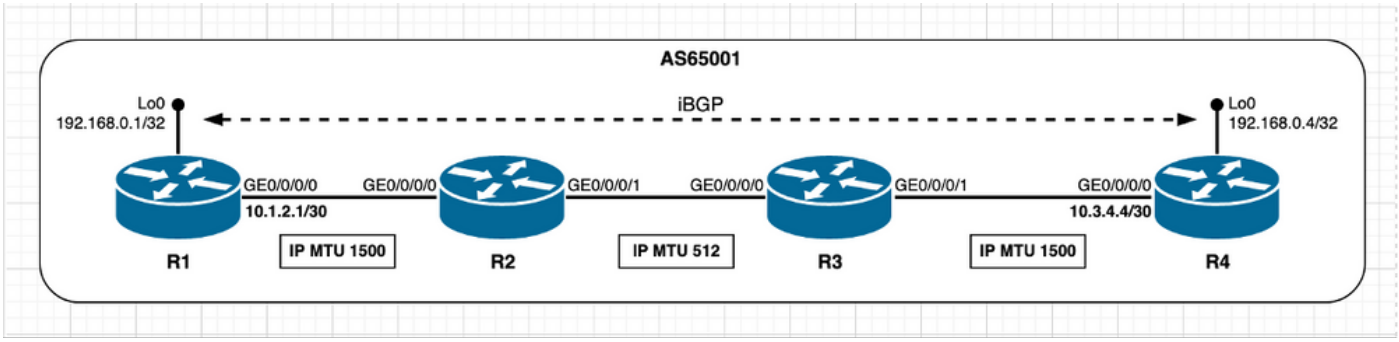

Imagen 2.8 - El segmento de trayecto R2/R3 tiene una MTU de IP inferior.

Como escenario inicial, considere que la información de BGP es mínima, es decir, que lo que sea necesario intercambiar entre peers BGP se puede lograr con paquetes IP que encajan en la MTU de trayectoria mínima de 512 bytes. Con esta suposición, el cálculo de MSS ocurre como se describe en la sección **Peers TCP no conectados directamente**. Tanto R1 como R4 seleccionan un valor MSS de 1240 bytes.

Detalles de la sesión TCP tal como se ven en R4 - ACTIVE:

```
! – 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>
```

Detalles de la sesión TCP tal como se ven en R1 - PASIVO:

```
! – 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>
```

Con la sesión BGP establecida ahora, considere que se activa un mensaje de actualización BGP con un tamaño superior al MTU de trayectoria mínima de 512 bytes. Como se puede observar de las salidas, Cisco IOS XR no configura el bit df con el mensaje de actualización BGP, lo que significa que la información BGP se transmite a expensas de la fragmentación de paquetes en los nodos intermedios.

Actualización de BGP originada por R1 - PASIVO:

```
! – 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
```

La fragmentación del mensaje de actualización de BGP originado por el nodo R1 se produce en el nodo R2, como puede observarse mediante la captura de tráfico realizada en la interfaz R2 GE0/0/0/1.

Fragmentación IP en el nodo R2:

```
! – 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=488, ID=7b41)
6      0.000122       192.168.0.1 192.168.0.4 IPv4   135    Fragmented IP protocol (proto=TCP 6,
off=976, ID=7b41)

! – 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>
```

Escenarios: TCP PMTUD habilitado

Habilitar PMTUD

Una vez que se habilita PMTUD, entonces independientemente de si los peers están conectados directa o no directamente, el cálculo inicial de MSS siempre toma en consideración la MTU IP de la interfaz de salida.

Esta situación proporciona información sobre el comportamiento esperado cuando se habilita PMTUD. Aquí, el nodo R4 de Cisco IOS XR desempeña la función activa, administra la conexión TCP y abre la sesión TCP con el nodo R1 de Cisco IOS XR en el puerto de destino 179. Ambos nodos utilizan los valores de MTU IP predeterminados en sus interfaces.

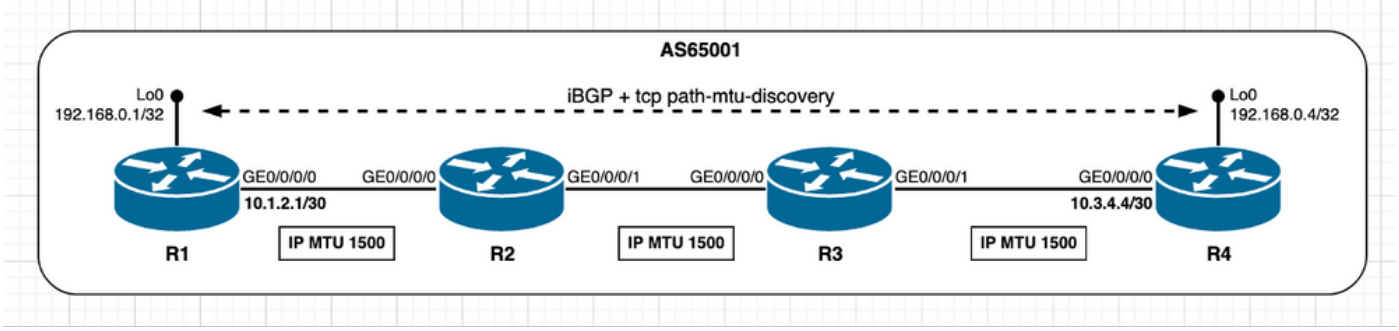

Imagen 3.1 - TCP PMTUD Enabled (PMTUD de TCP activado).

El cálculo de MSS en este escenario se puede resumir de la siguiente manera:

- Todos los nodos utilizan una MTU IP predeterminada de 1500 bytes
- Detección de MTU de Trayectoria TCP habilitada
- Los peers TCP no están conectados directamente R4 administra la conexión BGPR4 envía SYN con MSS de 1460 bytes 1500 (MTU de IP de interfaz) - 20 (minTCP_H) - 20 (minIP_H)R1 envía SYN, ACK con MSS de 1460 bytes Envía el valor inferior de [MSS recibido; MSS inicial local]MSS recibido 1460 bytes; MSS inicial local 1460 bytesEl valor MSS más bajo se utiliza en ambos pares

Para resaltar el cambio de comportamiento introducido desde habilitar PMTUD, los siguientes resultados ilustran la secuencia de eventos:

1. El estado inicial de la sesión TCP establecida en el escenario predeterminado de PMTUD inhabilitado;
2. PMTUD se configura y se habilita en ambos peers TCP R4 y R1;
3. Se reinicia la sesión TCP, se realiza el cálculo de MSS y la PMTUD de TCP influye.

Como se ve en R4 - ACTIVE - TCP PMTUD desactivado (valor predeterminado):

```
! – 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#
```

Como se ve en R1 - PASIVO - PMTUD TCP inhabilitado (valor predeterminado):

```
! – 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#
```

Como se ve en R4 - ACTIVE - TCP PMTUD habilitado:

```
! – '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 min)
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#
```

Como se ve en R1 - PASIVO - PMTUD TCP habilitado:

```
! – '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 min)
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#
```

Observe el comportamiento del temporizador del ager PMTU:

```
! – 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 min)
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
```

Como se ve en R4 - ACTIVE - BGP Session restart - TCP SYN:

```
! – 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)
```

Como se ve en R1 - PASSIVE - BGP Session restart - TCP SYN, ACK.

```
! – 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 Len=0 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)
```

Detalles de la sesión TCP tal como se ven en R4 - ACTIVE - después de que se habilita TCP PMTUD y se borra la sesión BGP:

```
! – 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#
```

Los detalles de la sesión TCP tal como se ven en R1 - PASSIVE - después de que se habilita TCP PMTUD y se borra la sesión BGP.

```
! – 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 - El segmento de la ruta tiene una MTU IP inferior

El escenario anterior ayudó a entender lo que sucede en el establecimiento inicial de la sesión TCP con PMTUD habilitado. Esta situación se basa en la parte superior y ayuda a entender cómo funciona TCP PMTUD y la influencia que tiene en las sesiones TCP establecidas.

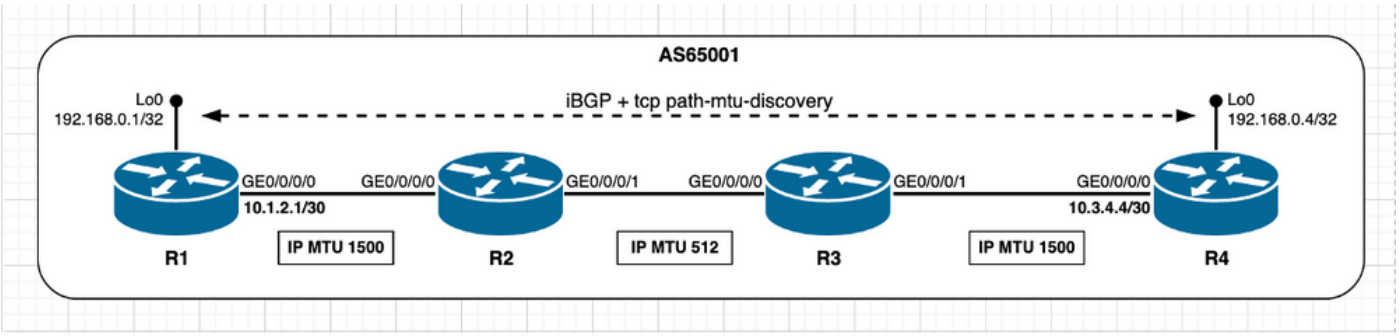

Imagen 3.2: PMTUD habilitado y el segmento de trayectoria tiene una MTU IP inferior.

Considere la imagen anterior como referencia, asuma que la sesión BGP se establece y R1 envía el mensaje de actualización BGP transportado por un paquete IP con un tamaño mayor que 512 bytes. Con la PMTUD activada, el bit DF (no fragmentar) está ahora configurado. Por lo tanto, el nodo R2 descarta el paquete IP y envía un mensaje ICMP (Protocolo de mensajes de control de Internet) (Destino inalcanzable - tipo 3; Fragmentación necesaria - código 4) de vuelta a R1. En el nodo R1 después de recibir el mensaje ICMP, se activa la PMTUD e intenta establecer la MTU IP de trayecto más bajo. Lo hace mediante el uso del siguiente valor inferior de un conjunto de niveles de meseta bien definidos, que considera un nuevo valor de MSS de sesión TCP. Luego, TCP retransmite la actualización BGP original con el nuevo valor MSS y este proceso se repite tantas veces como sea necesario hasta el mensaje ICMP (Destino inalcanzable - tipo 3; Fragmentación necesaria - Código 4) ya no se recibe. Esto significa que hasta que el valor de MSS en uso es tal que cada paquete enviado cae bajo la MTU IP del segmento de trayectoria más baja. A medida que pasa el tiempo, la PMTUD gobernada por el temporizador PmtuAger recorre los niveles de meseta en la dirección inversa y eleva el MSS a su valor máximo. En un momento dado, si un mensaje ICMP (Destino inalcanzable - tipo 3 ; Fragmentación necesaria: el código 4) se recibe de nuevo y la PMTUD actúa como se ha descrito anteriormente.

Los siguientes resultados recorren el comportamiento de PMTUD que se acaba de describir y comienzan desde el escenario de una sesión TCP establecida. Aquí, el nodo R4 del IOS de Cisco

desempeña una función activa, por lo que administra la conexión TCP y abre la sesión TCP con R1 en el puerto de destino 179. Ambos nodos utilizan los valores de MTU IP predeterminados en sus interfaces. El cálculo de MSS inicial en este escenario se puede resumir de la siguiente manera:

- El segmento intermedio entre los nodos R2 y R3 utiliza una MTU IP no predeterminada de 512 bytes.
- R1 y R4 utilizan valores de MTU predeterminados en sus interfaces.
- La detección de MTU de trayecto TCP está habilitada.
- Los pares TCP no están conectados directamente. R4 administra la conexión BGP.R4 envía SYN con MSS de 1460 bytes. 1500 (MTU de IP de interfaz) - 20 (minTCP_H) - 20 (minIP_H).R1 envía SYN, ACK con MSS de 1460 bytes. Envía la parte inferior de [MSS recibido ; MSS inicial local].MSS recibido 1460 bytes; MSS inicial local 1460 bytes.El valor MSS más bajo se utiliza en ambos peers.

TCP SYN originado en R4:

```
! – 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)
```

TCP SYN, ACK originado en R1:

```
! – 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=0 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)
```

Con la sesión BGP establecida, el nodo R1 envía el mensaje de actualización de BGP y recibe el mensaje ICMP (Destino inalcanzable - tipo 3 ; Fragmentación necesaria - Código 4) a cambio originada en el nodo R2.

Esto ocurre porque el paquete IP que transporta el mensaje de actualización de BGP tiene configurado el bit DF y la MTU IP de 512 bytes utilizada en el segmento R2/R3 es menor que el tamaño del paquete IP de 1116 bytes. Como se explicó anteriormente, la recepción del mensaje ICMP activa la PMTUD.

En el ICMP R1, se recibe el mensaje de tipo 3/código 4:

```
! – 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]
```

En el nodo R1, desencadenado por un mensaje ICMP, la PMTUD de TCP intenta establecer la MTU IP de extremo a extremo más baja mediante el uso del siguiente valor inferior de un conjunto de niveles de meseta (MTU de IP) bien definidos. Estos niveles de meseta se documentan en RFC1191 - Detección de MTU de Trayectoria.

```
MTU plateaus from RFC 1191
- values include both TCP and IP headers
65535
32000
17914
8166
4352
2002
1492
1006
508
296
68
```

Pero desde ICMP (Destino inalcanzable - tipo 3; Fragmentation needed - Code 4) message received by node R1 transmite la **MTU del siguiente salto** y luego como se muestra a continuación, el nodo R1 utiliza este valor, que en nuestro ejemplo es 512 bytes, y ajusta el valor de MSS de la sesión TCP. Tenga en cuenta que la longitud del segmento TCP original era de 1076 bytes, por lo que se necesitan tres paquetes para retransmitir el segmento TCP original.

Como se ve en R1 - PASSIVE - Operación PMTUD:

```
! – 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 Ack=251 Win=32593 Len=472
532    0.015374       192.168.0.1 192.168.0.4 TCP    526    [TCP Retransmission] 179  32449
[ACK] Seq=790 Ack=251 Win=32593 Len=472
533    0.004129       192.168.0.1 192.168.0.4 TCP    186    [TCP Retransmission] 179  32449
[PSH, ACK] Seq=1262 Ack=251 Win=32593 Len=132
```

Como se indicó anteriormente, una vez que todos los paquetes se han transmitido con el tiempo, PMTUD recorre los niveles de meseta en la dirección inversa regida por el temporizador PmtuAger e intenta elevar el MSS a su valor máximo según el escenario en vigor.

Como se ve en R1 - PMTUD a través de las plataformas definidas:

```
! – 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 about 192.168.0.4, MTU=512
RP/0/0/CPU0:May 12 09:09:22.763 UTC: ipv4_io[266]: ipv4_icmp_unreachable_rcvd ICMP unreach
recvd: sending pak(0xb0c07d8f) to transport: 6, tid: 5
RP/0/0/CPU0:May 12 09:09:22.763 UTC: ipv4_io[266]: ip_icmp_lib_ipv4_receive: sending
pak(0xb0c07d8f) to transport: 1, tid: 5
RP/0/0/CPU0:May 12 09:09:22.763 UTC: tcp[399]: [t4] PCB 0x15393770: Process ICMP Dest-unreach
(next hop mtu: 512)

! – 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 mtu: 512)
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  configured tp_user_mss 0

! – 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
```

Se puede observar el estado final en estos resultados. Observe en particular los valores MSS mínimos y máximos mostrados por el nodo R1, que resalta y señala que se activó PMTUD.

Detalles de la sesión TCP tal como se ven en R4 - ACTIVE:

```
! – 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#
```

Detalles de la sesión TCP tal como se ven en R1 - PASIVO:

```
! – 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#
```

Por último, si en un momento dado un ICMP (Destino inalcanzable - tipo 3 ; Fragmentación necesaria: el mensaje de código 4) se vuelve a recibir y, a continuación, PMTUD vuelve a actuar como se ha descrito anteriormente.

Como se ve desde R1 - PASSIVE - PMTUD se ha activado de nuevo:

```
! – 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#
```

En las versiones XR de Cisco IOS impactadas por el ID de bug de Cisco CSCvf10395, se ignora el siguiente salto contenido en el mensaje de error ICMP y el nodo intenta establecer la MTU IP de extremo a extremo más baja mediante el uso del siguiente valor inferior del conjunto de niveles de meseta (MTU IP) bien definidos mencionados anteriormente y documentados por RFC1119 1 - Detección de MTU de Trayectoria. Estos intentos ocurren hasta que la transmisión sea exitosa, esto significa hasta el ICMP (Destino inalcanzable - tipo 3 ; Fragmentación necesaria - Código 4) los mensajes ya no se reciben.

Como se ve desde un nodo con la versión Cisco IOS XR impactada por el ID de bug Cisco CSCvf10395:

```
! – 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: 33554432)
RP/0/0/CPU0:Feb 23 17:05:32.929 : tcp[399]: [t4] PCB 0x12152adc: Invalid next hop mtu
(33554432), ignore it
RP/0/0/CPU0:Feb 23 17:05:34.649 : tcp[399]: [t27] PCB 0x12152adc: Trying next lower MTU: 1452
<<<<<<<< HERE: Plateau 1492

RP/0/0/CPU0:Feb 23 17:05:35.519 : tcp[399]: [t4] PCB 0x12152adc: Process ICMP Dest-unreach (next
hop mtu: 33554432)
RP/0/0/CPU0:Feb 23 17:05:35.519 : tcp[399]: [t4] PCB 0x12152adc: Invalid next hop mtu
(33554432), ignore it
RP/0/0/CPU0:Feb 23 17:05:37.239 : tcp[399]: [t27] PCB 0x12152adc: Trying next lower MTU: 966
<<<<<<<< HERE: Plateau 1006

RP/0/0/CPU0:Feb 23 17:05:38.109 : tcp[399]: [t4] PCB 0x12152adc: Process ICMP Dest-unreach (next
hop mtu: 33554432)
RP/0/0/CPU0:Feb 23 17:05:38.109 : tcp[399]: [t4] PCB 0x12152adc: Invalid next hop mtu
(33554432), ignore it
RP/0/0/CPU0:Feb 23 17:05:39.829 : tcp[399]: [t27] PCB 0x12152adc: Trying next lower MTU: 468
<<<<<<<< HERE: Plateau 508
```

Como paso siguiente, considere la misma situación, pero con el protocolo de distribución de etiquetas (LDP) en todas las interfaces. El objetivo aquí es entender qué diferencias se pueden observar de escenarios anteriores en un entorno habilitado para MPLS.

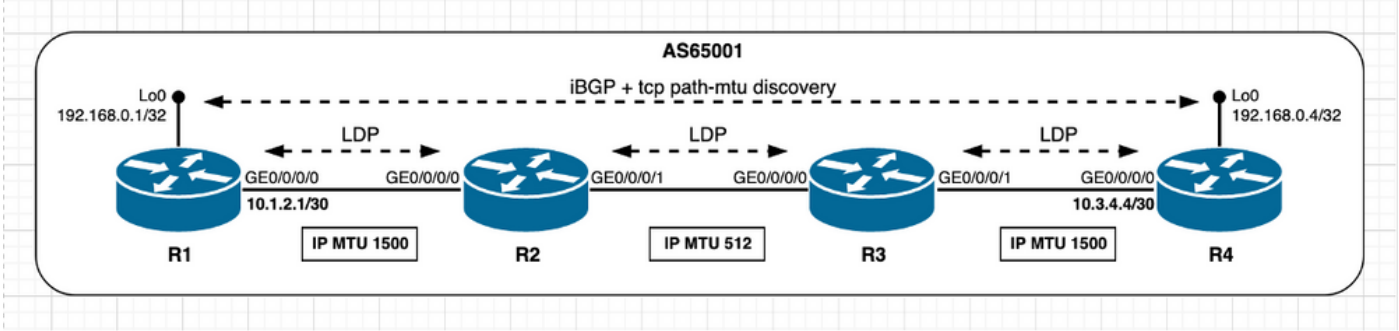

Imagen 3.3: PMTUD habilitado y el segmento de trayectoria tiene un escenario de MTU de IP -

MPLS más bajo.

Primero, considere la etapa inicial de la sesión BGP establecida antes del disparador PMTUD como se muestra aquí.

Estado inicial de TCP (BGP) visto en el escenario R4 - ACTIVE - habilitado para MPLS:

```
! – 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#
```

Estado inicial de TCP (BGP) tal como se ve en el escenario R1 - PASSIVE - habilitado para MPLS:

```
! – 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#
```

En este escenario habilitado para MPLS, se observa que se establecieron los detalles para las sesiones TCP (LDP). Tenga en cuenta que todo lo descrito anteriormente con respecto al cálculo de MSS para las sesiones TCP (BGP), se aplica también a las sesiones TCP (LDP). A modo de ejemplo, el cálculo de MSS de la sesión TCP (LDP) de los nodos R3 y R2 se puede resumir de la siguiente manera:

- Tanto R2 como R3 utilizan MTU IP no predeterminada de 512 bytes.
- La detección de MTU de trayecto está habilitada.
- Los peers TCP no están conectados directamente (la sesión TCP se establece entre las interfaces Loopback). R3 administra la conexión LDP.R3 envía SYN con MSS de 472 bytes. 512 (MTU de IP de interfaz) - 20 (minTCP_H) - 20 (minIP_H).R2 envía SYN, ACK con un MSS de 472 bytes. Envía el valor inferior de [MSS recibido; MSS inicial local].MSS recibido 472 bytes; MSS inicial local de 472 bytes.El valor MSS más bajo se utiliza en ambos peers.

Detalles de la sesión TCP (LDP) tal como se ven en el escenario R3 - ACTIVE - MPLS habilitado:

```
! – 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#
```

Detalles de la sesión TCP (LDP) tal como se ven en el escenario R2 - PASSIVE - habilitado para MPLS:

```
! – 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#
```

Después de establecer la sesión BGP, R1 envía el mensaje de actualización BGP y recibe el mensaje ICMP (Destino inalcanzable - tipo 3; Fragmentación necesaria - Código 4) a cambio originada en el nodo R2 que activa la PMTUD TCP en el nodo R1. Esto ocurre porque el paquete IP que transporta el mensaje de actualización de BGP tiene configurado el bit DF y la MTU IP de 512 bytes utilizada en el segmento R2/R3 es menor que el tamaño del paquete IP de 1116 bytes. Como antes, la recepción de este mensaje ICMP activa la PMTUD. La diferencia en el escenario habilitado para MPLS cuando se compara con los escenarios anteriores no MPLS es con respecto a la **MTU del** valor **del salto siguiente** incluido en el mensaje ICMP del nodo R2 (Destino inalcanzable - tipo 3; Fragmentación necesaria - Código 4). En este escenario habilitado para MPLS, la **MTU del** valor **del salto siguiente** representa la sobrecarga MPLS adicional de 4 bytes, lo que significa que representa la pila de etiquetas MPLS de salida en R2, como se ve en estas salidas.

Detección de MTU de trayectoria TCP en acción tal como se ve en el escenario R1 - PASSIVE - habilitado para MPLS:

```
! – 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=242 Ack=175 Win=32669 Len=468
460    0.004001       192.168.0.1 192.168.0.4 TCP    526    [TCP Retransmission] 179  57400
[ACK] Seq=710 Ack=175 Win=32669 Len=468
461    0.001788       192.168.0.1 192.168.0.4 TCP    198    [TCP Retransmission] 179  57400
[PSH, ACK] Seq=1178 Ack=175 Win=32669 Len=140
463    0.056695       192.168.0.4 192.168.0.1 TCP    54     57400  179 [ACK] Seq=175 Ack=1318
Win=31545 Len=0

! – 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 min)
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) tid: 5
```

```
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 about 192.168.0.4, MTU=508
RP/0/0/CPU0:May 17 08:35:51.726 UTC: ipv4_io[266]: ipv4_icmp_unreachable_rcvd ICMP unreach
recvd: sending pak(0xb0c07d8f) to transport: 6, tid: 5
RP/0/0/CPU0:May 17 08:35:51.726 UTC: ipv4_io[266]: ip_icmp_lib_ipv4_receive: sending
pak(0xb0c07d8f) to transport: 1, tid: 5
RP/0/0/CPU0:May 17 08:35:51.726 UTC: tcp[399]: [t4] PCB 0x153acc8c: Process ICMP Dest-unreach
(next hop mtu: 508)

! – 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  configured tp_user_mss 0

! – 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
```

Como se ve desde R1 - PASIVO - PMTUD TCP activado - escenario habilitado para MPLS:

```
! – 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#
```

Tenga en cuenta que en el escenario habilitado para MPLS, el valor de la **MTU del salto siguiente** incluido en el mensaje ICMP del nodo R2 representa la pila de etiquetas MPLS de salida. Para reforzar aún más este aspecto, consideremos el siguiente ejemplo. Si el paquete IP filtrado en R2 se asocia con un servicio L3VPN, significa que la trama Ethernet ahora lleva dos etiquetas (etiqueta IGP y etiqueta VPN). A continuación, la **MTU del salto siguiente** refleja el tamaño de pila de etiquetas requerido. Consulte estos resultados.

Como se ve en el paquete de servicio R1 - PASSIVE - L3 VPN:

```
! – 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
```

Como se ve en R1 - PASSIVE - servicio VPN L3 - ICMP tipo 3/código 4:

```
! – 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 - Opciones TCP (MD5)

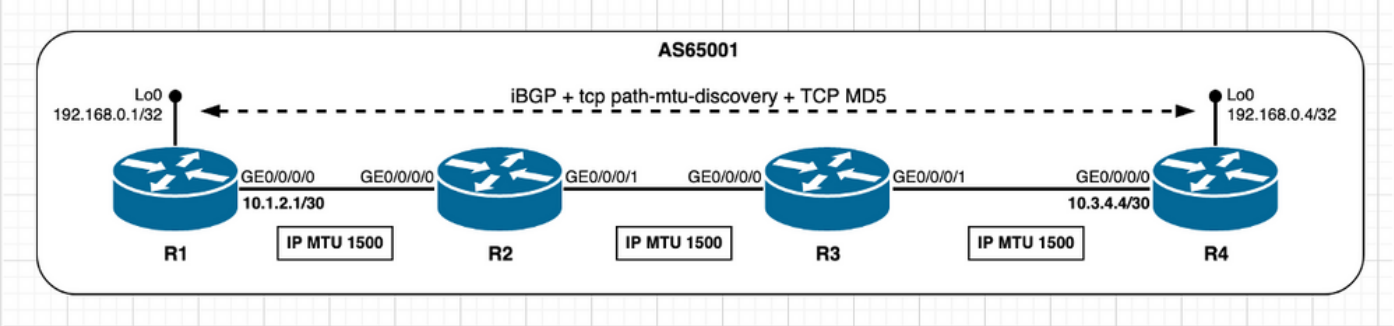

Imagen 3.4: PMTUD habilitado y autenticación TCP MD5.

No se introduce ninguna distinción con respecto al comportamiento de PMTUD a partir de lo que ya se ha descrito en los escenarios anteriores con la autenticación TCP MD5 habilitada. Como se compartió previamente con la autenticación TCP MD5 en uso, Cisco IOS XR considera sobrecarga adicional y el valor MSS inicial del peer TCP activo refleja lo mismo. Consulte las secciones anteriores **Use TCP Options - XR Active** y **Use TCP Options - XR Passive** para obtener más detalles sobre el impacto del uso de las opciones TCP. El cálculo de TCP MSS en este escenario se puede resumir de la siguiente manera:

- Todos los nodos utilizan una MTU IP predeterminada de 1500 bytes.
- La detección de MTU de trayecto TCP está habilitada.
- Los pares TCP no están conectados directamente.
- Autenticación TCP MD5 habilitada en R1 y R4. R4 administra la conexión BGP.R4 envía SYN con MSS de 1436 bytes. 1500 (MTU de IP de interfaz) - 20 (minTCP_H) - 20 (minIP_H) - 24 bytes (sobrecarga de opciones TCP IOS XR).R1 envía SYN, ACK con MSS de 1436 bytes.

envía el valor inferior de [MSS recibido ; MSS inicial local].MSS recibido 1436 bytes; MSS inicial local 1460 bytes.El valor MSS más bajo se utiliza en ambos peers.

TCP SYN originado en R4:

```
! – 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)
```

TCP SYN, ACK originado en R1:

```
! – 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 Len=0 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)
```

Detalles de la sesión TCP tal como se ven en R4 - ACTIVE:

```
! – 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#
```

Detalles de la sesión TCP tal como se ven en R1 - PASIVO:

```
! – 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: detección de agujeros negros

Como se explicó anteriormente en la sección **PMTUD - El Segmento de Trayectoria tiene MTU IP Menor**, la PMTUD TCP cuando se activa se activa mediante la recepción de un ICMP (Destino inalcanzable - tipo 3; Fragmentación necesaria - Código 4). Puede ser que estos mensajes no se reciban por alguna razón, lo que da como resultado que no se active PMTUD. En ese caso, no se aprende la MTU IP más baja de la trayectoria entre los peers TCP. Semejante escenario introduciría un agujero negro potencial si los paquetes IP tienen el bit DF configurado y si tienen un tamaño mayor que el segmento de trayectoria de MTU IP más bajo. Esos paquetes se descartarían silenciosamente.

Esta sección tiene como objetivo destacar cómo Cisco IOS XR detecta y actúa en ese escenario de agujero negro potencial. Para este fin, la función IPv4 unreachables se inhabilita en la interfaz R2 GE0/0/0/0, como se describe en la siguiente imagen y salida CLI.

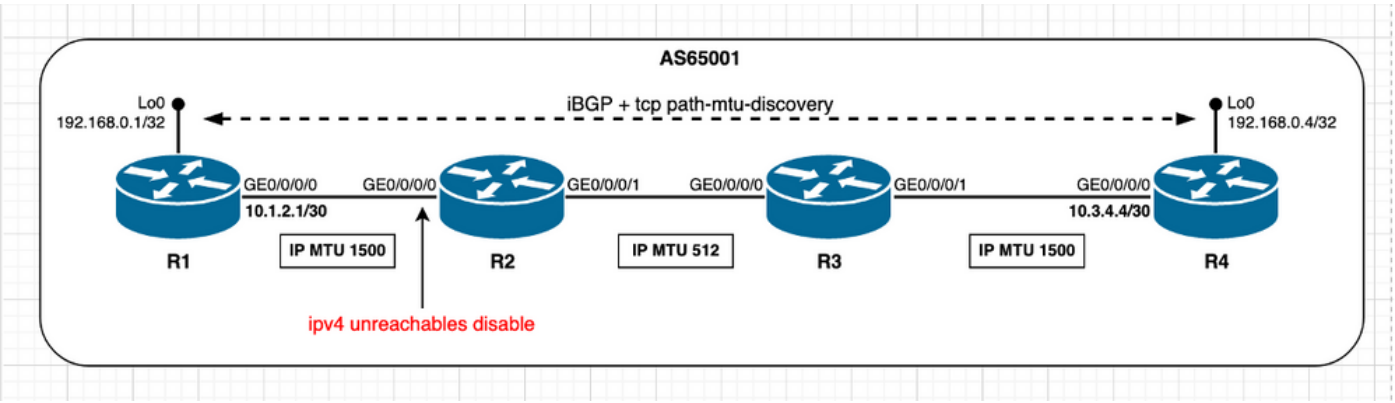

Imagen 3.5: PMTUD habilitado en IPv4 R1/R4 y R2 inalcanzables.

Inalcanzables IPv4 deshabilitados en R2:

```
!- 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
```

La manera en que Cisco IOS XR trata este escenario de agujero negro es retransmitir el mismo paquete dos veces y, si aún no se recibe el resultado, es decir, no se recibe el ACK de TCP esperado, vuelva a intentarlo pero utilice el siguiente valor de meseta bien definido inferior como se documenta en RFC1191 - Detección de MTU de Trayectoria (consulte la sección **PMTUD - El Segmento de Trayectoria tiene MTU de IP Inferior MTU** de MTU de IP eaus). En resumen, Cisco IOS XR asume que puede ser el caso de que los paquetes se descarten en algún lugar dentro de la trayectoria hacia su destino debido a su tamaño e intenta resolverlo a través de la retransmisión de paquetes. Este comportamiento se puede observar con el siguiente ejemplo de una captura de paquetes tomada en la interfaz del nodo R1 y el resultado de **debug tcp pmtud**.

Detección de agujero negro de IOS-XR en R1:

```
! – 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] Seq=318 Ack=251 Win=32593 Len=1076
564    1.101367       192.168.0.1 192.168.0.4 TCP    1130   [TCP Retransmission] 179  57082
[PSH, ACK] Seq=318 Ack=251 Win=32593 Len=1076

! – 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] Seq=318 Ack=251 Win=32593 Len=1076
568    1.111361       192.168.0.1 192.168.0.4 TCP    1130   [TCP Retransmission] 179  57082
[PSH, ACK] Seq=318 Ack=251 Win=32593 Len=1076

! – 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=318 Ack=251 Win=32593 Len=966
570    1.109602       192.168.0.1 192.168.0.4 TCP    1020   [TCP Retransmission] 179  57082
[ACK] Seq=318 Ack=251 Win=32593 Len=966

! – 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=318 Ack=251 Win=32593 Len=468
573    0.004254       192.168.0.1 192.168.0.4 TCP    522    [TCP Retransmission] 179  57082
[ACK] Seq=786 Ack=251 Win=32593 Len=468
574    0.002724       192.168.0.1 192.168.0.4 TCP    194    [TCP Retransmission] 179  57082
[PSH, ACK] Seq=1254 Ack=251 Win=32593 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 Len=0
```