Vérification de la détection MTU de chemin sur Cisco IOS XR et BGP

Table des matières

Introduction

Informations générales

TCP PMTUD et TCP MSS

Scénarios - TCP PMTUD désactivé

- Utiliser les valeurs MTU par défaut
- Utiliser une valeur MTU non par défaut - homologue TCP actif
- Utiliser une valeur MTU non par défaut - Pair TCP passif
- Utiliser les options TCP - XR actif
- Utiliser les options TCP - XR passif
- Homologues TCP non connectés directement
- Homologues TCP non connectés directement - Utiliser les options TCP (MD5)
- Les homologues TCP non connectés directement - Le segment de chemin a une MTU IP inférieure

Scénarios - TCP PMTUD activé

- Activer PMTUD
- PMTUD - Le segment de chemin a une MTU IP inférieure
- PMTUD - Options TCP (MD5)
- PMTUD - Détection des trous noirs

Introduction

Ce document décrit la détection PMTUD (Path Maximum Transmission Unit) du protocole de contrôle de transmission (TCP) sur les périphériques Cisco IOS® XR.

Informations générales

Le mécanisme PMTUD tente de déterminer la plus grande taille de paquet IP (Internet Protocol) qui ne nécessite aucune fragmentation sur le chemin entre deux hôtes. La valeur établie est appelée MTU du chemin et est égale à un minimum des valeurs MTU sur chaque saut. Si vous considérez le MTU de chemin lorsque vous transmettez des informations, il vous permet de tirer le meilleur parti de la capacité du réseau et d'éviter la fragmentation et l'efficacité de la transmission. La mécanique et la mise en oeuvre de PMTUD sont introduites dans un ensemble diversifié de scénarios avec l'utilisation du protocole BGP (Border Gateway Protocol) comme protocole client qui révèle progressivement le comportement de PMTUD.

TCP PMTUD et TCP MSS

Le protocole TCP tire parti des résultats de la PMTUD afin d'influencer la taille maximale de segment (MSS) locale, ce qui signifie qu'il s'adapte dynamiquement à la MTU de chemin découverte. Par conséquent, avant de passer à PMTUD, vous pouvez rapidement revoir la taille maximale de segment (MSS) TCP et comprendre ce qu'elle signifie et son objectif.

Selon la définition d'origine de MSS de RFC879 : La définition de l'option MSS peut être énoncée comme suit : Nombre maximal d'octets de données pouvant être reçus par l'expéditeur de cette option TCP dans des segments TCP sans options d'en-tête TCP transmis dans des datagrammes IP sans options d'en-tête IP.

Clarifier certains aspects et conseiller les responsables de la mise en oeuvre, RFC6691 met en évidence la façon dont la valeur MSS doit être calculée :

Lorsque vous calculez la valeur à mettre dans l'option MSS TCP, la valeur MTU doit être diminuée uniquement de la taille des en-têtes IP et TCP fixes et ne doit pas être diminuée pour tenir compte d'options IP ou TCP possibles ; inversement, l'expéditeur DOIT réduire la longueur des données TCP pour prendre en compte les options IP ou TCP qu'il inclut dans les paquets qu'il envoie.

Une définition plus détaillée de MSS peut être extraite du Guide de configuration du routage pour les routeurs de la gamme Cisco ASR 9000, IOS XR version 7.9.x :

MSS est la plus grande quantité de données qu'un ordinateur ou un périphérique de communication peut recevoir dans un seul segment TCP non fragmenté. Toutes les sessions TCP sont limitées par une limite sur le nombre d'octets qui peuvent être transportés dans un seul paquet ; cette limite est MSS. Le protocole TCP divise les paquets en segments dans une file d'attente de transmission avant de les transmettre à la couche IP.

La valeur MSS TCP dépend du MTU d'une interface, qui est la longueur maximale de données pouvant être transmises par un protocole à une instance. La longueur maximale des paquets TCP est déterminée à la fois par la MTU de l'interface de sortie sur le périphérique source et par la MSS annoncée par le périphérique de destination au cours du processus de configuration TCP. Plus le MSS est proche du MTU, plus le transfert de messages BGP est efficace. Chaque direction du flux de données peut utiliser une valeur MSS différente.

Quelle est alors la valeur que TCP doit prendre en compte pour MSS sur une session TCP donnée ? Et comment est-il calculé ?

Pour les valeurs par défaut selon RFC879, vous avez : Les hôtes ne doivent pas envoyer de datagrammes supérieurs à 576 octets, sauf s'ils savent que l'hôte de destination est prêt à accepter des datagrammes plus importants. LA TAILLE MAXIMALE DU SEGMENT TCP EST LA TAILLE MAXIMALE DU DATAGRAMME IP MOINS QUARANTE.

La taille maximale du datagramme IP par défaut est 576.

La taille maximale de segment TCP par défaut est 536.

Cela prend en compte une valeur de MTU IP de 576 octets. Cependant, si vous ignorez la valeur

MTU IP réelle, le calcul MSS TCP peut être résumé comme suit :

- Homologue actif : calcule et envoie le MSS initial avec le paquet SYN.

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

- Pair passif : calcule le MSS initial, le compare au MSS reçu de l'homologue actif et envoie SYN, ACK avec la valeur MSS la plus faible.

```
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.
```

Il n'y a pas de négociation concernant la valeur de l'option MSS. Chaque noeud détermine sa propre valeur et l'annonce lors de l'établissement de la session TCP. Il devient clair que si la valeur MTU IP considérée pour le calcul MSS peut être dérivée de PMTUD, alors la valeur MSS peut être adaptée à la valeur la plus efficace pour un MTU de chemin donné. Le comportement de Cisco IOS XR a quelques spécificités en ce qui concerne le calcul MSS et le rôle PMTUD résumés ici.

PMTUD est désactivé par défaut sur Cisco IOS XR :

- Le calcul de la MSS initiale locale prend en compte la MTU IP comme suit :
  - Si des homologues connectés directement - considérez l'interface de sortie IP MTU.
  - Si des homologues non connectés directement - considérez IP MTU de 1280 octets.
  - La valeur MSS dépend des options TCP configurées.

Lorsque PMTUD est activé sur Cisco IOS XR :

- Le calcul de la MSS initiale locale prend en compte la MTU IP comme suit :
  - Quels que soient les homologues connectés directement/non directement, considérez le MTU IP de l'interface de sortie.
  - La valeur MSS dépend des options TCP configurées.

Il y a d'autres détails sur la mécanique et la mise en oeuvre de la PMTUD qui doivent être pris en compte et que ce document présente à travers des exemples pratiques résumés dans le tableau

suivant. Ce tableau présente également le MTU IP des homologues TCP actif et passif ainsi que les valeurs MSS sélectionnées pour chaque scénario considéré.

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

Scénarios - TCP PMTUD désactivé

Utiliser les valeurs MTU par défaut

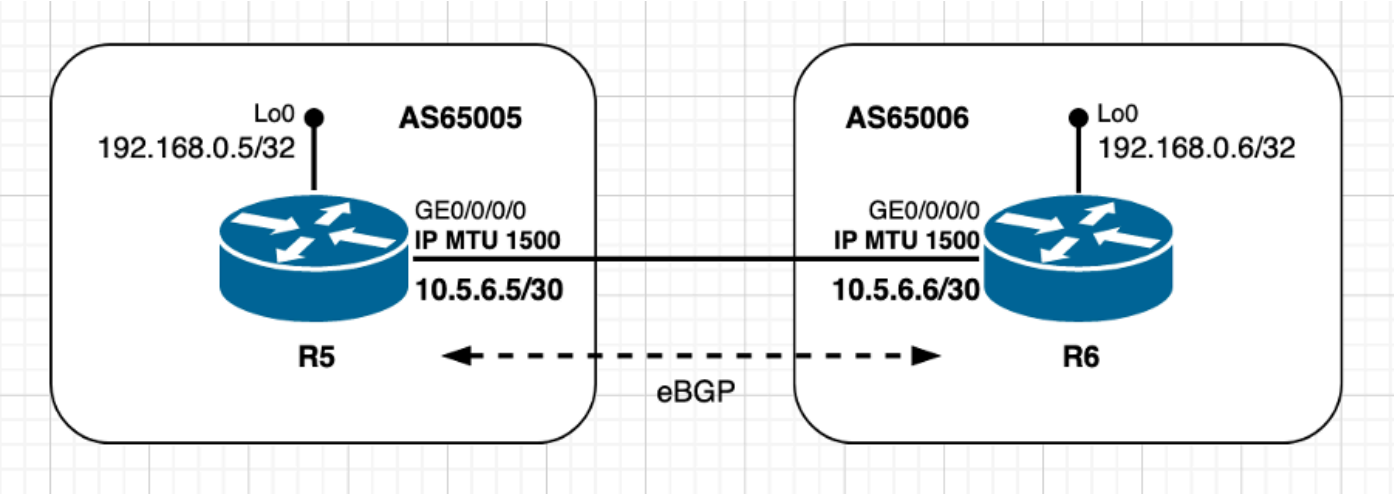


Image 2.1. Avec les valeurs MTU par défaut

Dans le cas des homologues eBGP illustrés dans l'image 2.1, R6 gère la connexion TCP, ce qui signifie qu'il joue le rôle actif et démarre la session TCP avec R5 sur le port de destination 179. Les homologues sont directement connectés et les deux utilisent les valeurs MTU IP par défaut sur leurs interfaces respectives. En fonction des informations partagées au début de ce document, le calcul MSS dans ce scénario peut être résumé comme suit :

- Les deux noeuds utilisent une MTU IP par défaut de 1 500 octets
- La détection MTU du chemin TCP est désactivée par défaut
- Les homologues TCP sont connectés directement
  - R6 gère la connexion BGP
  - R6 envoie SYN avec MSS de 1 460 octets
    - 1 500 (MTU IP d'interface) - 20 (minTCP_H) - 20 (minIP_H)
  - R5 envoie SYN, ACK avec MSS de 1 460 octets

- Envoie le plus faible de [Received MSS ; MSS initiale locale]
- MSS reçue 1460 octets ; MSS initiale locale 1 460 octets

- La valeur MSS la plus faible est utilisée sur les deux homologues

Détails de la session TCP tels qu'ils apparaissent sur R6 - ACTIVE :

<#root>

```
! - As seen on R6 - ACTIVE

RP/0/0/CPU0:R6#show interfaces gigabitEthernet 0/0/0/0
Fri Jan  8 09:35:48.553 UTC
GigabitEthernet0/0/0/0 is up, line protocol is up
Interface state transitions: 1
Hardware is GigabitEthernet, address is fa16.3e85.3dc2 (bia fa16.3e85.3dc2)
Internet address is 10.5.6.6/30
MTU 1514 bytes
, BW 1000000 Kbit (Max: 1000000 Kbit)
<snip>

RP/0/0/CPU0:R6#show tcp brief
Fri Jan  8 09:36:22.491 UTC
PCB      VRF-ID     Recv-Q Send-Q Local Address          Foreign Address        State
<snip>
0x121649fc 0x60000000      0      0  10.5.6.6:24454         10.5.6.5:179           ESTAB
<snip>

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

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

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

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

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

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

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

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


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



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

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

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

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

Détails de la session TCP tels qu'ils apparaissent sur R5 - PASSIVE :

<#root>

```
! - As seen on R5 - PASSIVE

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

MTU 1514 bytes

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

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

RP/0/0/CPU0:R5#show tcp detail pcb 0x12155884
```

```
Fri Jan  8 09:34:47.317 UTC
=============================================================
Connection state is ESTAB, I/O status: 0, socket status: 0
Established at Fri Jan  8 09:28:29 2021

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

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

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

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

SRTT: 47 ms,  RTTO: 300 ms,  RTV: 170 ms,  KRTT: 0 ms
minRTT: 19 ms,  maxRTT: 219 ms

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

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


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



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

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

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

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

Utiliser une valeur MTU non par défaut - homologue TCP actif

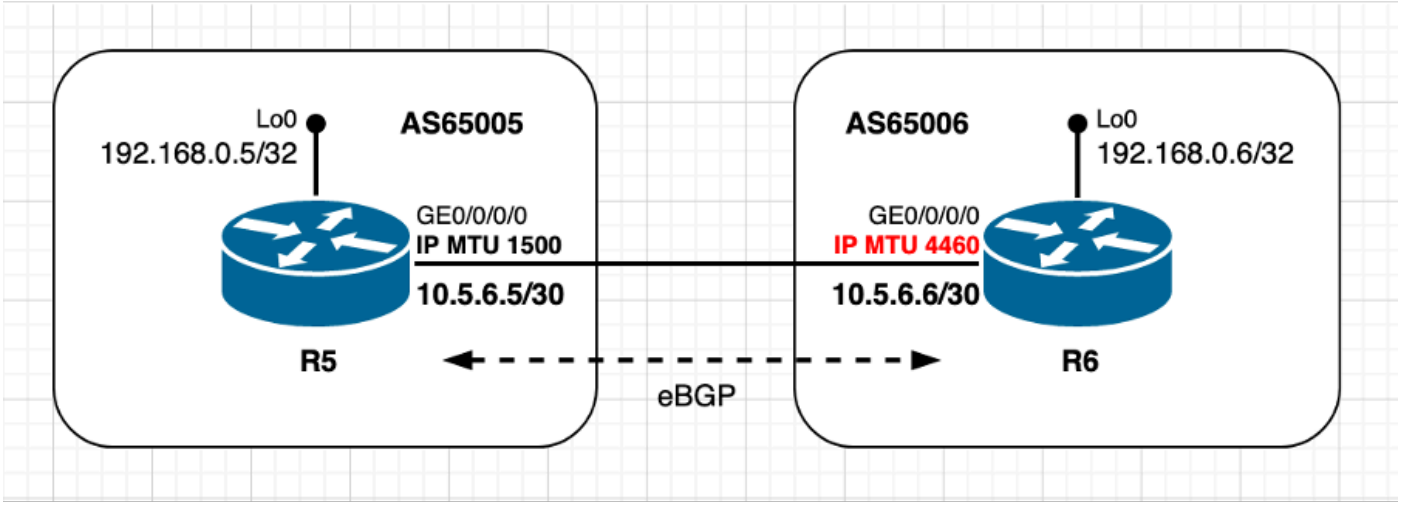


Image 2.2 - L'homologue ACTIF utilise une valeur MTU autre que la valeur par défaut

Ce scénario est essentiellement le même que le précédent, à la seule différence que l'homologue TCP actif R6 utilise désormais une valeur MTU IP autre que la valeur par défaut. Notez comment le calcul initial et la décision sur la valeur MSS sont effectués par l'homologue TCP passif R5. Le calcul MSS TCP dans ce scénario peut être résumé comme suit :

- R6 utilise un MTU IP autre que par défaut de 4 460 octets
- La détection MTU du chemin TCP est désactivée par défaut
- Les homologues TCP sont connectés directement
  - R6 gère la connexion BGP
  - R6 envoie SYN avec un MSS de 4 420 octets
    - 4460 (MTU IP d'interface) - 20 (minTCP_H) - 20 (minIP_H)
  - R5 envoie SYN, ACK avec MSS de 1 460 octets
    - envoie le plus faible de [Received MSS ; MSS initiale locale]
    - MSS reçue 4 420 octets ; MSS initiale locale 1 460 octets
  - La valeur MSS la plus faible est utilisée sur les deux homologues

SYN TCP provenant de R6 :

<#root>

```
! - TCP SYN sourced from R6

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

MSS=4420

 WS=1

Frame 140: 62 bytes on wire (496 bits), 62 bytes captured (496 bits) on interface 0
Ethernet II, Src: fa:16:3e:85:3d:c2 (fa:16:3e:85:3d:c2), Dst: fa:16:3e:ad:51:8f (fa:16:3e:ad:51:8f)
Internet Protocol Version 4, Src: 10.5.6.6, Dst: 10.5.6.5
Transmission Control Protocol, Src Port: 35502, Dst Port: 179, Seq: 0, Len: 0
    Source Port: 35502
```

```
    Destination Port: 179
    [Stream index: 6]
    [TCP Segment Len: 0]
    Sequence number: 0    (relative sequence number)
    Acknowledgment number: 0
    Header Length: 28 bytes
    Flags: 0x002 (SYN)
    Window size value: 16384
    [Calculated window size: 16384]
    Checksum: 0x219d [unverified]
    [Checksum Status: Unverified]
    Urgent pointer: 0
    Options: (8 bytes), Maximum segment size, Window scale, End of Option List (EOL)
        Maximum segment size: 4420 bytes
            Kind: Maximum Segment Size (2)
            Length: 4
```

MSS Value: 4420

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

TCP SYN, ACK provenant de R5 :

<#root>

```
! – TCP SYN, ACK sourced from R5

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

MSS=1460

```
 WS=1

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

MSS Value: 1460

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

Détails de la session TCP tels qu'ils apparaissent sur R6 - ACTIVE :

<#root>

```
! - as seen on R6 - Active

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

MTU 4474 bytes

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

RP/0/0/CPU0:R6#show tcp detail pcb 0x1215761c
Fri Jan  8 09:56:25.819 UTC
================================================================
Connection state is ESTAB, I/O status: 0, socket status: 0
Established at Fri Jan  8 09:51:46 2021

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

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

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

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

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

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

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


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



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

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

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

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

Détails de la session TCP tels qu'ils apparaissent sur R5 - PASSIVE :

<#root>

```
! - as seen on R5 - Passive

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

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

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

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

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

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

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

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

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


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

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

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

Utiliser une valeur MTU non par défaut - Pair TCP passif

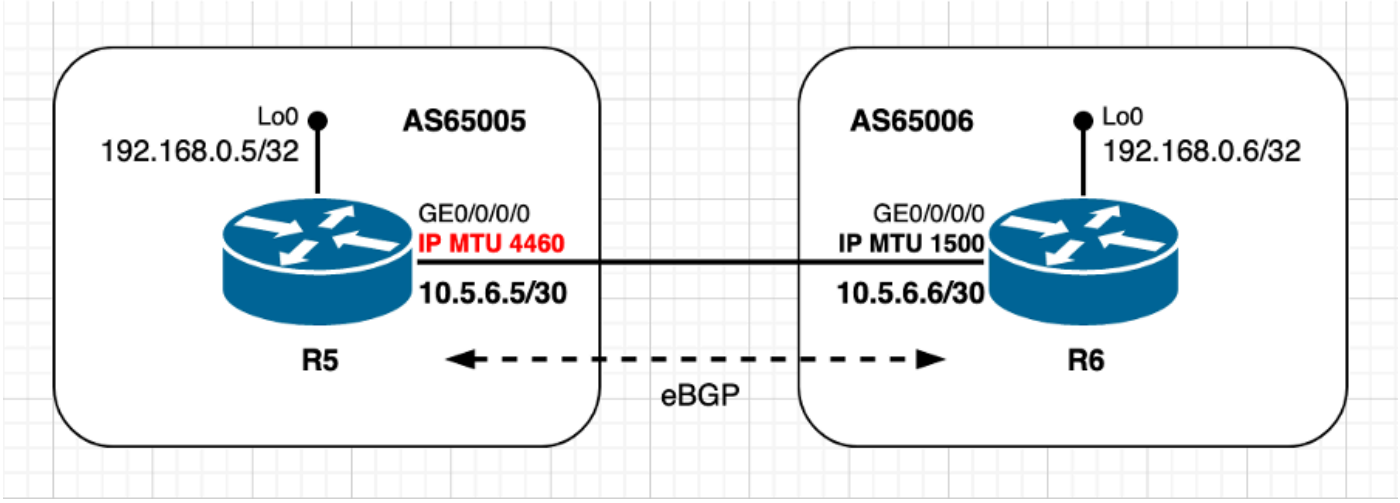


Image 2.3 - L'homologue PASSIVE utilise une valeur MTU autre que la valeur par défaut.

Avec toujours le même scénario eBGP, mais maintenant avec l'homologue TCP passif R5 configuré avec le MTU IP non par défaut et l'homologue TCP actif R6 avec la valeur MTU IP par défaut. Comme dans le scénario précédent, prenez note de la façon dont la valeur MSS est sélectionnée par l'homologue passif R5. Le calcul MSS TCP dans ce scénario peut être résumé comme suit :

- R5 utilise un MTU IP autre que par défaut de 4 460 octets
- La détection MTU du chemin TCP est désactivée par défaut
- Les homologues TCP sont connectés directement
    - R6 gère la connexion BGP
    - R6 envoie SYN avec MSS de 1 460 octets
        - 1 500 (MTU IP d'interface) - 20 (minTCP_H) - 20 (minIP_H)
    - R5 envoie SYN, ACK avec MSS de 1 460 octets
        - envoie le plus faible de [Received MSS ; MSS initiale locale]
        - MSS reçue 1460 octets ; MSS initiale locale 4 420 octets
    - La valeur MSS la plus faible est utilisée sur les deux homologues

SYN TCP provenant de R6 :

<#root>

```
! – TCP SYN sourced from R6

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

MSS=1460

 WS=1

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

MSS Value: 1460

        Window scale: 0 (multiply by 1)
```

```
        End of Option List (EOL)
```

TCP SYN, ACK provenant de R5 :

<#root>

```
! – TCP SYN, ACK sourced from R5

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

MSS=1460

 WS=1

Frame 238: 62 bytes on wire (496 bits), 62 bytes captured (496 bits) on interface 0
Ethernet II, Src: fa:16:3e:ad:51:8f (fa:16:3e:ad:51:8f), Dst: fa:16:3e:85:3d:c2 (fa:16:3e:85:3d:c2)
Internet Protocol Version 4, Src: 10.5.6.5, Dst: 10.5.6.6
Transmission Control Protocol, Src Port: 179, Dst Port: 47007, Seq: 0, Ack: 1, Len: 0
    Source Port: 179
    Destination Port: 47007
    [Stream index: 10]
    [TCP Segment Len: 0]
    Sequence number: 0    (relative sequence number)
    Acknowledgment number: 1    (relative ack number)
    Header Length: 28 bytes
    Flags: 0x012 (SYN, ACK)
    Window size value: 16384
    [Calculated window size: 16384]
    Checksum: 0x7078 [unverified]
    [Checksum Status: Unverified]
    Urgent pointer: 0
    Options: (8 bytes), Maximum segment size, Window scale, End of Option List (EOL)
        Maximum segment size: 1460 bytes
            Kind: Maximum Segment Size (2)
            Length: 4

MSS Value: 1460


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

Détails de la session TCP tels qu'ils apparaissent sur R6 - ACTIVE :

<#root>

```
! - as seen on R6 - Active

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

PCB 0x1215761c, SO 0x12162aac, TCPCB 0x12156f64, vrfid 0x60000000,
```

```
Pak Prio: Medium, TOS: 192, TTL: 1, Hash index: 103
Local host: 10.5.6.6, Local port: 47007 (Local App PID: 1011918)
Foreign host: 10.5.6.5, Foreign port: 179

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

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

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

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

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

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


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



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

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

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

Détails de la session TCP tels qu'ils apparaissent sur R5 - PASSIVE :

<#root>

```
! - as seen on R5 - Passive

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

MTU 4474 bytes

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

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

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

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

Timer          Starts    Wakeups         Next(msec)
Retrans             7          0                0
SendWnd             0          0                0
TimeWait            0          0                0
AckHold             7          2                0
KeepAlive           1          0                0
PmtuAger            0          0                0
GiveUp              0          0                0
Throttle            0          0                0

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

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

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

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

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

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

```
Sack holes {start, end, dups, rxmit}: none

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

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

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

Utiliser les options TCP - XR actif

Comme mentionné précédemment dans ce document, l'utilisation d'options TCP (telles que TCP MD5, TCP selective-ack, ou TCP timestamps) influence le calcul MSS car ces options conduisent à des octets supplémentaires à prendre en compte dans le calcul MSS.

Cette section ainsi que la suivante vise à illustrer le calcul MSS effectué par les homologues en présence d'options TCP. L'option d'authentification MD5 TCP est utilisée à titre d'exemple. Reportez-vous au scénario de référence des images 2.4, comme illustré dans l'image.

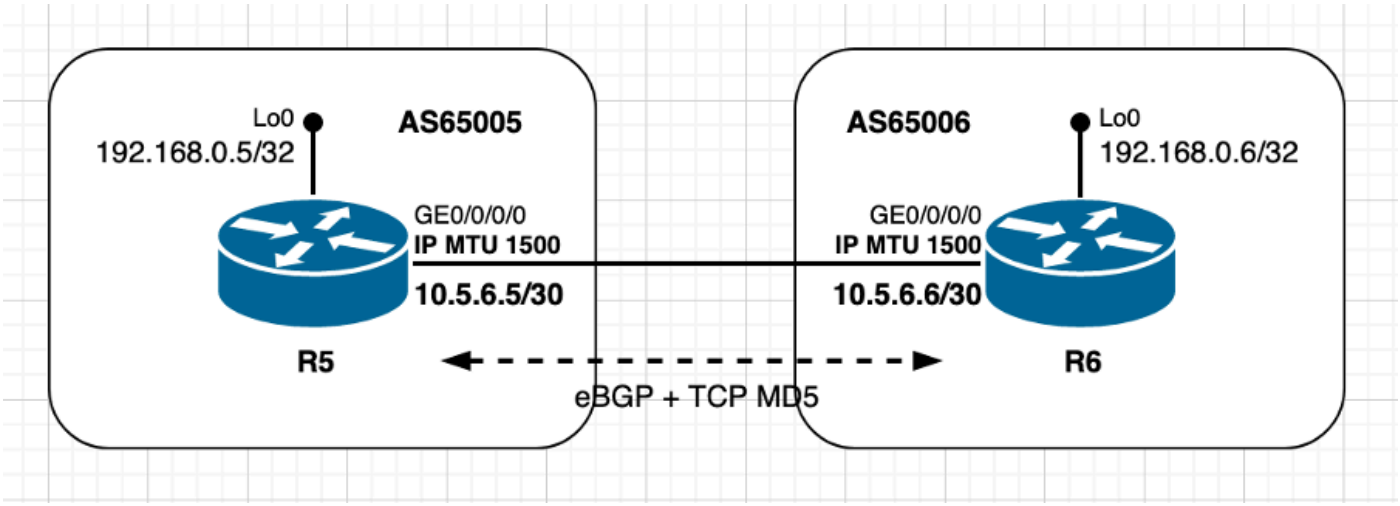


Image 2.4 - Utiliser les options TCP (MD5) - XR actif.

Dans ce scénario, les deux homologues utilisent les valeurs MTU IP par défaut, sont connectés directement et l'homologue R6 joue un rôle actif TCP. Comme déjà partagé la configuration et l'utilisation du compte d'authentification MD5 TCP pour une surcharge supplémentaire. Le calcul MSS TCP dans ce scénario particulier peut être résumé comme suit :

- Les deux noeuds utilisent une MTU IP par défaut de 1 500 octets
- La détection MTU du chemin TCP est désactivée par défaut
- Les homologues TCP sont connectés directement
- Authentification MD5 TCP activée sur les deux noeuds

- R6 gère la connexion BGP
- R6 envoie SYN avec MSS de 1 436 octets
  - 1 500 (MTU IP d'interface) - 20 (minTCP_H) - 20 (minIP_H) - 24 octets (surcharge des options TCP IOS XR)
- R5 envoie SYN, ACK avec MSS de 1 436 octets
  - envoie le plus faible de [Received MSS ; MSS initiale locale]
  - MSS reçue 1436 octets ; MSS initiale locale 1 460 octets
- La valeur MSS la plus faible est utilisée sur les deux homologues

Comme le montre le résumé, la façon dont Cisco IOS XR se comporte n'est pas strictement selon RFC 879, et RFC 6691, qui indiquent que les options TCP ne doivent pas être prises en compte dans le calcul MSS.

Le compte XR de Cisco IOS d'un facteur supplémentaire sur la longueur d'en-tête tcp est documenté plus loin sur l'ID de bogue Cisco CSCvf2016 :

«(...)Lorsque XR démarre la connexion BGP, BGP crée d'abord le socket, puis définit les options de socket, y compris MD5. Cela fait que la longueur d'en-tête de l'option tcp = 24. Et donc la valeur MSS initiale devient 1500 - 40 - 24 = 1436. Ceci est envoyé à l'homologue et l'homologue utilise min(1436, 1460) = 1436.(...)

SYN TCP provenant de R6 :

<#root>

```
! - TCP SYN sourced from R6

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

MSS=1436

 WS=1

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

TCP MD5 signature

, End of Option List (EOL)
        Maximum segment size: 1436 bytes
```

```
            Kind: Maximum Segment Size (2)
            Length: 4
```

MSS Value: 1436

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

TCP SYN, ACK provenant de R5 :

<#root>

```
! – TCP SYN, ACK sourced from R5

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

MSS=1436

```
 WS=1

Frame 431: 82 bytes on wire (656 bits), 82 bytes captured (656 bits) on interface 0
Ethernet II, Src: fa:16:3e:ad:51:8f (fa:16:3e:ad:51:8f), Dst: fa:16:3e:85:3d:c2 (fa:16:3e:85:3d:c2)
Internet Protocol Version 4, Src: 10.5.6.5, Dst: 10.5.6.6
Transmission Control Protocol, Src Port: 179, Dst Port: 24785, Seq: 0, Ack: 1, Len: 0
    Source Port: 179
    Destination Port: 24785
    [Stream index: 14]
    [TCP Segment Len: 0]
    Sequence number: 0    (relative sequence number)
    Acknowledgment number: 1    (relative ack number)
    Header Length: 48 bytes
    Flags: 0x012 (SYN, ACK)
    Window size value: 16384
    [Calculated window size: 16384]
    Checksum: 0xe83d [unverified]
    [Checksum Status: Unverified]
    Urgent pointer: 0
    Options: (28 bytes), Maximum segment size, Window scale, No-Operation (NOP),
```

TCP MD5 signature

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

MSS Value: 1436

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

Détails de la session TCP tels qu'ils apparaissent sur R6 - ACTIVE :

<#root>

```
! - as seen on R6 - Active

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

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

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

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

   iss: 1379482495  snduna: 1379482819  sndnxt: 1379482819
sndmax: 1379482819  sndwnd: 32498       sndcwnd: 4308
   irs: 3750694052  rcvnxt: 3750694376  rcvwnd: 32517   rcvadv: 3750726893

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

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

State flags: none
Feature flags:
```

MD5

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

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

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

Socket options: SO_REUSEADDR, SO_REUSEPORT, SO_NBIO
```

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

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

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

Détails de la session TCP tels qu'ils apparaissent sur R5 - PASSIVE :

<#root>

```
! - as seen on R5 - Passive

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

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

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

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

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

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

State flags: none
Feature flags:
```

MD5

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

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

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

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

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

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

Un comportement similaire peut être observé avec d'autres options TCP qui, lorsqu'elles sont configurées, prennent en compte une surcharge supplémentaire et influencent le calcul MSS dans Cisco IOS XR. Considérez le même scénario et ces exemples qui documentent le calcul MSS lorsque les horodatages TCP et les options de retour sélectif TCP sont configurés.

Détails de la session TCP tels qu'ils apparaissent sur R6 - ACTIVE - avec les options TCP timestamp et selective-ack configurées :

<#root>

```
! - as seen on R6 – Active
```

! -- tcp timestamp configured

```
! -- 12 bytes of additional overhead

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

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

```
<snip>

! - as seen on R6 – Active
```

! -- tcp selective-ack configured

```
! -- 36 bytes of additional overhead

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


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


<snip>

! - as seen on R6 – Active

! -- tcp selective-ack and tcp timestamp configured


! -- 40 bytes of additional overhead

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


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


<snip>

! - as seen on R6 – Active

! -- MD5 and tcp selective-ack configured


! -- 36 bytes of additional overhead

RP/0/0/CPU0:R6#show tcp detail pcb 0x1539b3cc
<snip>
Feature flags: Sack, MD5, Win Scale, Nagle
Request flags: Sack, Win Scale


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


<snip>

! - as seen on R6 – Active

! -- MD5 and tcp timestamp configured


! -- 36 bytes of additional overhead

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

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


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

```
<snip>

! - as seen on R6 – Active

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


! -- 40 bytes of additional overhead

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


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


<snip>
```

Utiliser les options TCP - XR passif

D'après le scénario précédent, vous avez probablement remarqué le comportement distinct du noeud Cisco IOS XR lorsqu'il joue un rôle passif en ce qui concerne le calcul MSS initial. Le noeud ne tient pas compte de la longueur d'en-tête de l'option tcp. Ce scénario vise à mettre en évidence ce comportement distinct qui est également décrit par l'ID de bogue Cisco :

« (...) - Lorsque l'homologue initie la connexion, il envoie le MSS initial comme 1460. XR TCP crée socket, pcb etc., puis il prend ces deux actions dans l'ordre donné :

- D'abord, il calcule le MSS initial après soustraction de la longueur d'en-tête de l'option tcp. C'est '0' car l'option MD5 n'est pas encore héritée de cette socket à partir de la socket d'écoute.

- Ensuite, il hérite de 'MD5' et d'autres options, ce qui fait 'option header bytes length' à 24.

Dans ce cas, le protocole TCP XR envoie le MSS initial sous la forme 1460 et est donc utilisé par les deux. (…)“

Dans ce scénario, bien que l'homologue TCP actif R8 soit un noeud Cisco IOS, ce fait n'introduit aucune différence ou précision sur ce que le scénario vise à mettre en évidence. Néanmoins, et de manière intéressante, notez que différemment de Cisco IOS XR comme indiqué dans le scénario de la section précédente, ici l'homologue TCP actif R8 ne prend pas en compte les options TCP sur le calcul MSS initial.

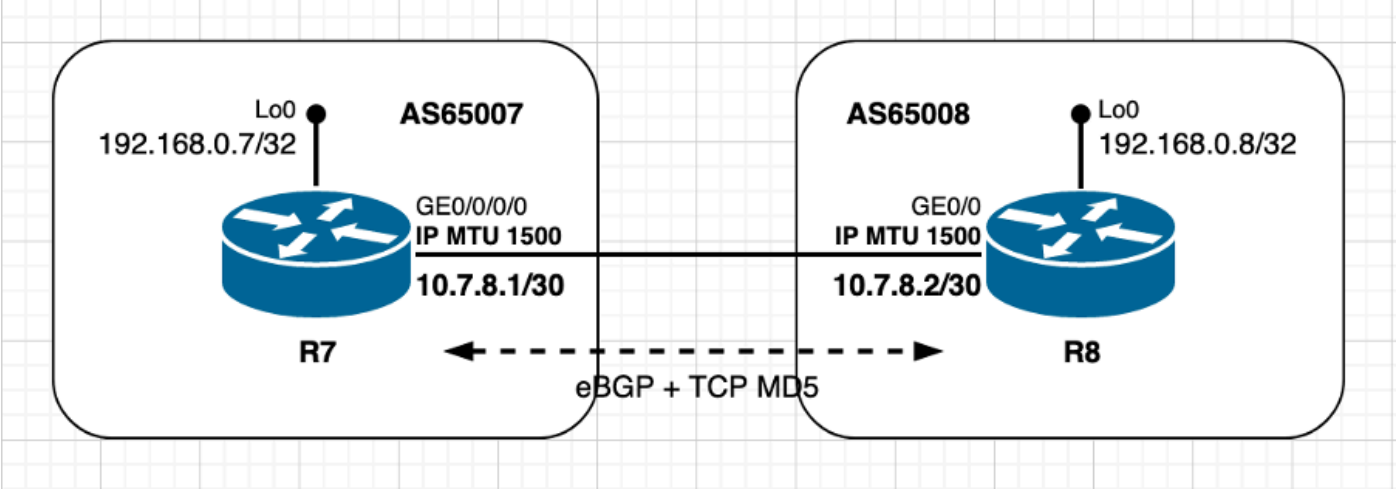


Image 2.5 - Utiliser les options TCP (MD5) - XR passif.

Les deux homologues utilisent des valeurs de MTU IP par défaut et sont directement connectés. L'homologue R8 de Cisco IOS joue un rôle actif. Dans ce scénario, le calcul MSS TCP peut être résumé comme suit :

- Les deux noeuds utilisent une MTU IP par défaut de 1 500 octets
- La détection MTU du chemin TCP est désactivée par défaut sur Cisco IOS XR R7
- La détection MTU du chemin TCP est activée par défaut sur Cisco IOS R8
- Les homologues TCP sont connectés directement
- Authentification MD5 TCP activée sur les deux noeuds
    - IOS R8 gère la connexion BGP
    - IOS R8 envoie SYN avec MSS de 1 460 octets
        - 1 500 (MTU IP d'interface) - 20 (minTCP_H) - 20 (minIP_H)
    - IOS XR R7 envoie SYN, ACK avec MSS de 1 460 octets
        - envoie le plus faible de [Received MSS ; MSS initiale locale]
        - MSS reçue 1460 octets ; MSS initiale locale 1 460 octets
    - La valeur MSS la plus faible est utilisée sur les deux homologues

SYN TCP provenant de R8 - Cisco IOS :

<#root>

```
! – TCP SYN sourced from R8

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

MSS=1460

```
Frame 96: 78 bytes on wire (624 bits), 78 bytes captured (624 bits) on interface 0
Ethernet II, Src: fa:16:3e:58:21:ba (fa:16:3e:58:21:ba), Dst: fa:16:3e:68:d9:e5 (fa:16:3e:68:d9:e5)
Internet Protocol Version 4, Src: 10.7.8.2, Dst: 10.7.8.1
Transmission Control Protocol, Src Port: 52975, Dst Port: 179, Seq: 0, Len: 0
    Source Port: 52975
    Destination Port: 179
    [Stream index: 3]
    [TCP Segment Len: 0]
```

```
    Sequence number: 0    (relative sequence number)
    Acknowledgment number: 0
    Header Length: 44 bytes
    Flags: 0x002 (SYN)
    Window size value: 16384
    [Calculated window size: 16384]
    Checksum: 0xb612 [unverified]
    [Checksum Status: Unverified]
    Urgent pointer: 0
    Options: (24 bytes), Maximum segment size,
```

TCP MD5 signature

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

MSS Value: 1460

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

TCP SYN, ACK provenant de R7 - Cisco IOS XR :

<#root>

```
! – TCP SYN,ACK sourced from R7

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

MSS=1460

```
Frame 97: 78 bytes on wire (624 bits), 78 bytes captured (624 bits) on interface 0
Ethernet II, Src: fa:16:3e:68:d9:e5 (fa:16:3e:68:d9:e5), Dst: fa:16:3e:58:21:ba (fa:16:3e:58:21:ba)
Internet Protocol Version 4, Src: 10.7.8.1, Dst: 10.7.8.2
Transmission Control Protocol, Src Port: 179, Dst Port: 52975, Seq: 0, Ack: 1, Len: 0
    Source Port: 179
    Destination Port: 52975
    [Stream index: 3]
    [TCP Segment Len: 0]
    Sequence number: 0    (relative sequence number)
    Acknowledgment number: 1    (relative ack number)
    Header Length: 44 bytes
    Flags: 0x012 (SYN, ACK)
    Window size value: 16384
    [Calculated window size: 16384]
    Checksum: 0xfb47 [unverified]
    [Checksum Status: Unverified]
    Urgent pointer: 0
    Options: (24 bytes), Maximum segment size,
```

TCP MD5 signature

```
, End of Option List (EOL)
        Maximum segment size: 1460 bytes
```

```
            Kind: Maximum Segment Size (2)
            Length: 4
```

MSS Value: 1460

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

Détails de la session TCP tels qu'ils apparaissent sur R8 - Cisco IOS - ACTIVE :

<#root>

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

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

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

 For address family: IPv4 Unicast
  Session: 10.7.8.1
  BGP table version 1, neighbor version 1/0
  Output queue size : 0
  Index 6, Advertise bit 0
  6 update-group member
  Slow-peer detection is disabled
  Slow-peer split-update-group dynamic is disabled
                               Sent       Rcvd
  Prefix activity:             ----       ----
    Prefixes Current:             0          0
    Prefixes Total:               0          0
    Implicit Withdraw:            0          0
    Explicit Withdraw:            0          0
    Used as bestpath:           n/a          0
```

```
    Used as multipath:            n/a          0
    Used as secondary:            n/a          0

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

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

Transport(tcp) path-mtu-discovery is enabled

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

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

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

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

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

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

md5

IP Precedence value : 6

Datagrams (max data segment is 1460 bytes)

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

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

R8#

Détails de la session TCP tels qu'ils apparaissent sur R7 - Cisco IOS XR - PASSIVE :

<#root>

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

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

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

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

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

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

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

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

```
State flags: none
Feature flags:
```

MD5

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

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

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

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

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

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

Homologues TCP non connectés directement

Lorsque les homologues ne sont pas directement connectés, la façon dont le calcul initial MSS TCP est effectué change comme décrit précédemment dans la section d'introduction de ce document. Le scénario d'une session iBGP avec tous les homologues configurés avec des valeurs de MTU IP par défaut est utilisé pour parcourir le calcul MSS.

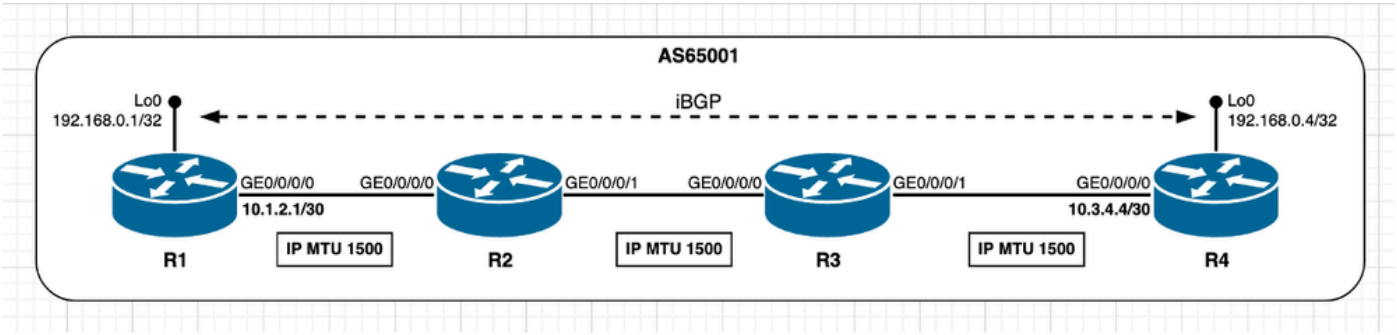


Image 2.6 - Homologues TCP non connectés directement - iBGP.

Il est important de noter que lorsque la détection MTU du chemin TCP est désactivée et que les homologues ne sont pas directement connectés, par conception, Cisco IOS XR utilise une valeur MTU IP fixe de 1 280 octets.

Dans l'image précédente, R4 joue un rôle actif et gère la connexion TCP. R4 ouvre la session TCP avec R1 sur le port de destination 179. Les deux noeuds utilisent la valeur MTU IP par défaut sur leurs interfaces. Le calcul de la MSS dans ce scénario peut être résumé comme suit :

- Tous les noeuds utilisent une MTU IP par défaut de 1 500 octets
- La détection MTU du chemin TCP est désactivée par défaut
- Les homologues TCP ne sont pas connectés directement
    - R4 gère la connexion BGP
    - R4 envoie SYN avec MSS de 1 240 octets
        - Le MTU d'interface n'est pas pris en compte lorsque les homologues ne sont pas directement connectés et que la détection du MTU du chemin TCP est désactivée
        - Selon la conception de Cisco IOS XR, 1 280 octets sont considérés comme TCP_DEFAULT_MTU
        - 1280 (TCP_DEFAULT_MTU) - 20 (minTCP_H) - 20 (minIP_H)
    - R1 envoie SYN, ACK avec MSS de 1 240 octets
        - Envoie le plus faible de [Received MSS ; MSS initiale locale]
        - MSS reçue 1240 octets ; MSS initiale locale 1 240 octets
    - La valeur MSS la plus faible est utilisée sur les deux homologues

SYN TCP provenant de R4 :

<#root>

```
! – TCP SYN sourced from R4

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

MSS=1240

 WS=1

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

MSS Value: 1240

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

TCP SYN, ACK provenant de R1 :

<#root>

```
! – TCP SYN,ACK sourced from R1

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

MSS=1240

```
 WS=1

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

MSS Value: 1240

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

Détails de la session TCP tels qu'ils apparaissent sur R4 - ACTIVE :

<#root>

```
! – as seen on R4 - Active

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

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

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

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

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

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

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

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

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


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



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

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

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

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

Détails de la session TCP tels qu'ils apparaissent sur R1 - PASSIVE :

<#root>

```
! - as seen on R1 - Passive

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

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

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

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

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

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

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

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


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



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

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

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

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

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

Homologues TCP non connectés directement - Utiliser les options TCP (MD5)

Dans le cas d'un scénario homologue connecté non directement et dans l'utilisation de l'authentification TCP MD5, il n'existe aucune différence fondamentale par rapport aux scénarios ou scénarios de test précédents déjà décrits. Comme nous l'avons vu précédemment avec l'authentification MD5 TCP, Cisco IOS XR considère une surcharge supplémentaire et la valeur MSS initiale reflète la même chose. Reportez-vous aux sections précédentes Utiliser les options TCP - XR actif et Utiliser les options TCP - XR passif pour plus de détails sur l'influence des options TCP sur le calcul TCP MSS.

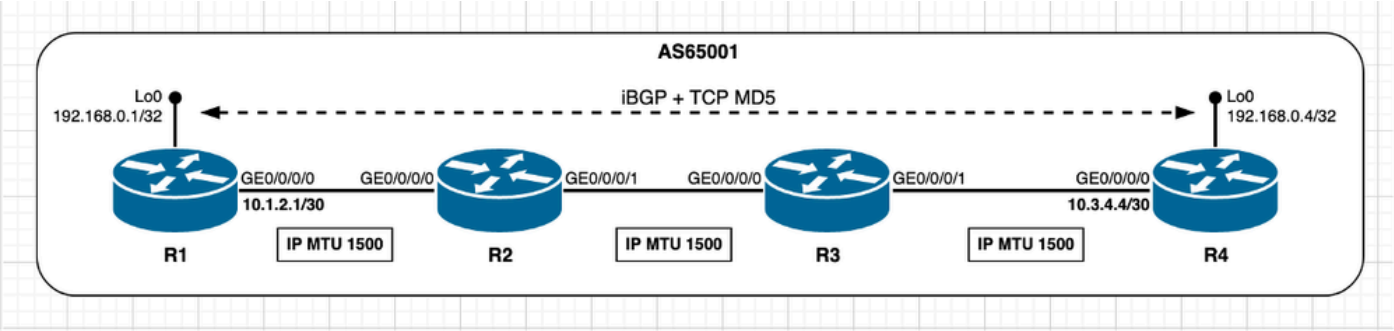


Image 2.7 - Homologues TCP non connectés directement - iBGP + TCP MD5.

Dans ce scénario, le calcul MSS TCP peut être résumé comme suit :

- Tous les noeuds utilisent une MTU IP par défaut de 1 500 octets
- La détection MTU du chemin TCP est désactivée par défaut
- Les homologues TCP ne sont pas connectés directement
    - R4 gère la connexion BGP
    - La destination R1 n'est pas connectée directement
    - R4 envoie SYN avec MSS de 1 216 octets
        - Le MTU d'interface n'est pas pris en compte lorsque les homologues ne sont pas directement connectés et que la détection du MTU du chemin TCP est désactivée
        - Selon la conception, 1 280 octets sont considérés comme TCP_DEFAULT_MTU
        - 1280 (TCP_DEFAULT_MTU) - 20 (minTCP_H) - 20 (minIP_H) - 24 octets (surcharge des options TCP IOS XR)
    - R1 envoie SYN, ACK avec MSS de 1 216 octets
        - Envoie le plus faible de [Received MSS ; MSS initiale locale]
        - MSS reçue 1216 octets ; MSS initiale locale 1 240 octets
    - La valeur MSS la plus faible est utilisée sur les deux homologues

SYN TCP provenant de R4 :

<#root>

```
! - 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 provenant de R1 :

<#root>

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

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

MSS=1216

 WS=1
```

```
Frame 3426: 82 bytes on wire (656 bits), 82 bytes captured (656 bits) on interface 0
Ethernet II, Src: fa:16:3e:8f:8f:54 (fa:16:3e:8f:8f:54), Dst: fa:16:3e:d7:7e:f6 (fa:16:3e:d7:7e:f6)
Internet Protocol Version 4, Src: 192.168.0.1, Dst: 192.168.0.4
Transmission Control Protocol, Src Port: 179, Dst Port: 42135, Seq: 0, Ack: 1, Len: 0
    Source Port: 179
    Destination Port: 42135
    [Stream index: 10]
    [TCP Segment Len: 0]
    Sequence number: 0    (relative sequence number)
    Acknowledgment number: 1    (relative ack number)
    Header Length: 48 bytes
    Flags: 0x012 (SYN, ACK)
    Window size value: 16384
    [Calculated window size: 16384]
    Checksum: 0xbb05 [unverified]
    [Checksum Status: Unverified]
    Urgent pointer: 0
    Options: (28 bytes), Maximum segment size, Window scale, No-Operation (NOP),
```

TCP MD5 signature

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

MSS Value: 1216

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

Détails de la session TCP tels qu'ils apparaissent sur R4 - ACTIVE :

<#root>

```
! - as seen from R4 - Active

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

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

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

Timer          Starts    Wakeups         Next(msec)
Retrans            14          1                  0
```

```
SendWnd             0          0               0
TimeWait            0          0               0
AckHold            11          9               0
KeepAlive           1          0               0
PmtuAger            0          0               0
GiveUp              0          0               0
Throttle            0          0               0


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

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

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

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


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



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

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

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

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

Détails de la session TCP tels qu'ils apparaissent sur R1 - PASSIVE :

<#root>

```
! - as seen from R1 - Passive

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

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

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

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

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

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

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

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


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



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

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

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

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

Les homologues TCP non connectés directement - Le segment de chemin a une MTU IP inférieure

Dans le scénario suivant, l'objectif est d'observer et de conclure ce qui se passe s'il y a un segment de chemin intermédiaire avec un MTU IP inférieur alors qu'il est dans la condition par défaut, ce qui signifie que TCP PMTUD est désactivé. Reportez-vous à cette image.

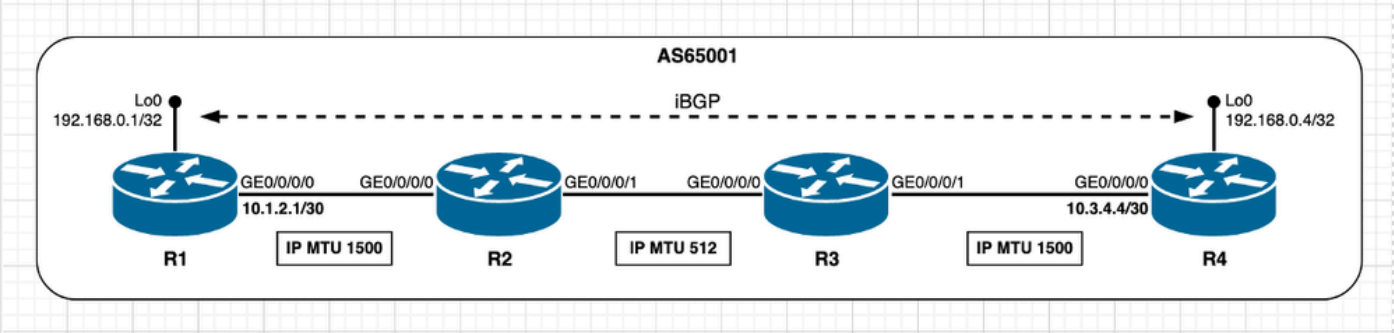


Image 2.8 - Le segment de chemin R2/R3 a une MTU IP inférieure.

Comme scénario initial, considérez que les informations BGP sont minimales, c'est-à-dire que tout ce qui est requis pour être échangé entre des homologues BGP peut être accompli avec des paquets IP qui tiennent sous le MTU de chemin minimum de 512 octets. Avec cette hypothèse, le calcul MSS se produit comme décrit dans la section Paires TCP non connectés directement. R1 et R4 sélectionnent une valeur MSS de 1 240 octets.

Détails de la session TCP tels qu'ils apparaissent sur R4 - ACTIVE :

<#root>

```
! - as seen from R4 - Active

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

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

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

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

```
Throttle           0          0               0

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

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

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

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

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

<snip>

Détails de la session TCP tels qu'ils apparaissent sur R1 - PASSIVE :

<#root>

```
! - as seen from R1 - Passive

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

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

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

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

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

SRTT: 202 ms,  RTTO: 355 ms,  RTV: 153 ms,  KRTT: 0 ms
```

```
minRTT: 9 ms,  maxRTT: 309 ms

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

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


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


<snip>
```

Maintenant que la session BGP est établie, considérez qu'un message de mise à jour BGP avec une taille supérieure à la MTU de chemin minimale de 512 octets est déclenché. Comme on peut l'observer à partir des résultats, Cisco IOS XR ne définit pas le df-bit avec le message de mise à jour BGP, ce qui signifie que les informations BGP sont transmises au détriment de la fragmentation des paquets sur les noeuds intermédiaires.

Mise à jour BGP provenant de R1 - PASSIVE :

<#root>

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

23     3.450878       192.168.0.1 192.168.0.4 BGP    1111   UPDATE Message

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


Total Length: 1097


    Identification: 0x5841 (22593)
    Flags: 0x00
        0... .... = Reserved bit: Not set
        .0.. .... = Don't fragment: Not set
        ..0. .... = More fragments: Not set
    Fragment offset: 0
    Time to live: 255
    Protocol: TCP (6)
    Header checksum: 0x54a4 [validation disabled]
    [Header checksum status: Unverified]
    Source: 192.168.0.1
    Destination: 192.168.0.4
    [Source GeoIP: Unknown]
    [Destination GeoIP: Unknown]
Transmission Control Protocol, Src Port: 179, Dst Port: 39046, Seq: 20, Ack: 20, Len: 1057
Border Gateway Protocol - UPDATE Message
```

```
    Marker: ffffffffffffffffffffffffffffffff
    Length: 1057
    Type: UPDATE Message (2)
    Withdrawn Routes Length: 0
    Total Path Attribute Length: 1034
    Path attributes
        Path Attribute - MP_REACH_NLRI
        Path Attribute - ORIGIN: INCOMPLETE
        Path Attribute - AS_PATH: empty
        Path Attribute - MULTI_EXIT_DISC: 0
        Path Attribute - LOCAL_PREF: 100
```

La fragmentation du message de mise à jour BGP provenant du noeud R1 a lieu au niveau du noeud R2, comme cela peut être observé par la capture du trafic effectuée au niveau de l'interface GE0/0/0/1 de R2.

Fragmentation IP au noeud R2 :

<#root>

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

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

! - Captured frame details

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

Total Length: 508



Identification: 0x7b41 (31553)


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

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



Fragment offset: 0


    Time to live: 254
```

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

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


Total Length: 508



Identification: 0x7b41 (31553)


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


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



Fragment offset: 488


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

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


Total Length: 121



Identification: 0x7b41 (31553)
```

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

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

Fragment offset: 976

    Time to live: 254
    Protocol: TCP (6)
    Header checksum: 0x35fa [validation disabled]
    [Header checksum status: Unverified]
    Source: 192.168.0.1
    Destination: 192.168.0.4
    [Source GeoIP: Unknown]
    [Destination GeoIP: Unknown]
Data (101 bytes)
<snip>
```

Scénarios - TCP PMTUD activé

Activer PMTUD

Une fois que PMTUD est activé, indépendamment du fait que les homologues sont directement ou non directement connectés, le calcul initial MSS prend toujours en compte le MTU IP de l'interface de sortie.

Ce scénario fournit des informations sur le comportement attendu lorsque PMTUD est activé. Ici, le noeud Cisco IOS XR R4 joue le rôle actif, gère la connexion TCP et ouvre la session TCP avec le noeud Cisco IOS XR R1 sur le port de destination 179. Les deux noeuds utilisent les valeurs MTU IP par défaut sur leurs interfaces.

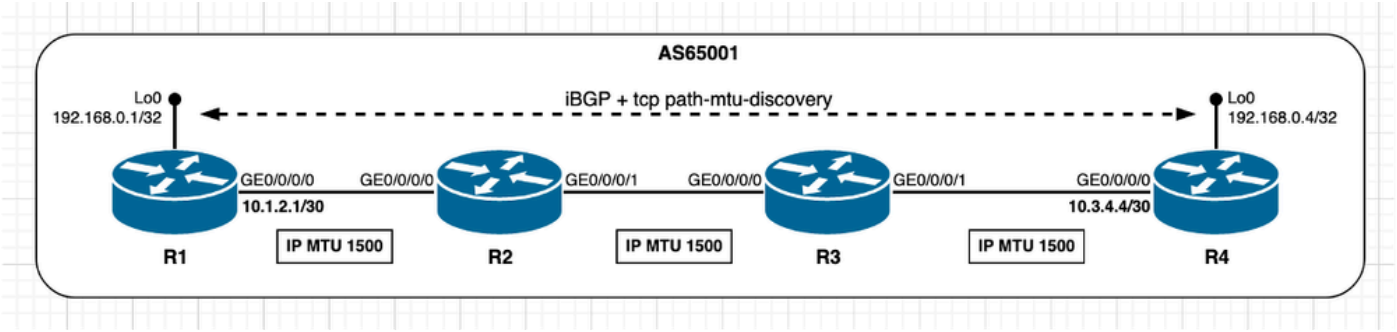


Image 3.1 : TCP PMTUD activé.

Le calcul de la MSS dans ce scénario peut être résumé comme suit :

- Tous les noeuds utilisent une MTU IP par défaut de 1 500 octets

- La détection MTU du chemin TCP est activée
- Les homologues TCP ne sont pas connectés directement
    - R4 gère la connexion BGP
    - R4 envoie SYN avec MSS de 1 460 octets
        - 1 500 (MTU IP d'interface) - 20 (minTCP_H) - 20 (minIP_H)
    - R1 envoie SYN, ACK avec MSS de 1 460 octets
        - Envoie le plus faible de [Received MSS ; MSS initiale locale]
        - MSS reçue 1460 octets ; MSS initiale locale 1 460 octets
    - La valeur MSS la plus faible est utilisée sur les deux homologues

Afin de mettre en évidence le changement de comportement introduit à partir de enable PMTUD, les sorties suivantes illustrent la séquence d'événements :

1. L'état initial de la session TCP établie dans le scénario par défaut de PMTUD désactivé ;
2. PMTUD est configuré et activé sur les homologues TCP R4 et R1 ;
3. La session TCP est redémarrée, le calcul MSS a lieu et est influencé par la PMTUD TCP.

Comme indiqué sur R4 - ACTIVE - TCP PMTUD désactivé (par défaut) :

<#root>

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Comme indiqué sur R1 - PASSIVE - TCP PMTUD désactivé (par défaut) :

<#root>

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

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

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

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

Timer          Starts    Wakeups         Next(msec)
Retrans             3          0                 0
SendWnd             0          0                 0
```

```
TimeWait            0          0               0
AckHold             3          1               0
KeepAlive           1          0               0
PmtuAger            0          0               0
GiveUp              0          0               0
Throttle            0          0               0

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

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

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

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


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



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

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

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

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

Comme indiqué sur R4 - ACTIVE - TCP PMTUD activé :

<#root>

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

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

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

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

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

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

Timer          Starts    Wakeups         Next(msec)
Retrans            10          1                  0
SendWnd             0          0                  0
TimeWait            0          0                  0
AckHold             7          4                  0
KeepAlive           1          0                  0

PmtuAger            1          0             508096


GiveUp              0          0                  0
Throttle            0          0                  0

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

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

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

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

Path MTU


Request flags: Win Scale


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



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

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

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

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

Comme indiqué sur R1 - PASSIVE - TCP PMTUD activé :

<#root>

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

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

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

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

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

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

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

PmtuAger            1          0            562042


GiveUp              0          0                 0
Throttle            0          0                 0

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

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

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

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

Path MTU

```
Request flags: Win Scale
```

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

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

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

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

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

Notez le comportement du minuteur de l'unité 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 m
RP/0/0/CPU0:Jan  8 16:09:25.214 : tcp[399]: [t21] Path mtu is ON (age-timer: 10)
RP/0/0/CPU0:Jan  8 16:19:25.233 : tcp[399]: [t21] PCB 0x12157020: Trying next higher MTU: 1240
RP/0/0/CPU0:Jan  8 16:21:25.245 : tcp[399]: [t21] PCB 0x12157020: Trying next higher MTU: 1240
RP/0/0/CPU0:Jan  8 16:23:25.256 : tcp[399]: [t21] PCB 0x12157020: Trying next higher MTU: 1240
```

Comme indiqué sur R4 - ACTIF - Redémarrage de session BGP - SYN TCP :

<#root>

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

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

MSS=1460

WS=1

```
Frame 2734: 62 bytes on wire (496 bits), 62 bytes captured (496 bits) on interface 0
Ethernet II, Src: fa:16:3e:d7:7e:f6 (fa:16:3e:d7:7e:f6), Dst: fa:16:3e:8f:8f:54 (fa:16:3e:8f:8f:54)
Internet Protocol Version 4, Src: 192.168.0.4, Dst: 192.168.0.1
Transmission Control Protocol, Src Port: 32077, Dst Port: 179, Seq: 0, Len: 0
    Source Port: 32077
    Destination Port: 179
    [Stream index: 25]
    [TCP Segment Len: 0]
    Sequence number: 0    (relative sequence number)
    Acknowledgment number: 0
    Header Length: 28 bytes
    Flags: 0x002 (SYN)
    Window size value: 16384
    [Calculated window size: 16384]
    Checksum: 0x6398 [unverified]
    [Checksum Status: Unverified]
    Urgent pointer: 0
    Options: (8 bytes), Maximum segment size, Window scale, End of Option List (EOL)
        Maximum segment size: 1460 bytes
            Kind: Maximum Segment Size (2)
            Length: 4
```

MSS Value: 1460

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

Comme indiqué sur R1 - PASSIVE - BGP Session restart - TCP SYN, ACK.

<#root>

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

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

MSS=1460

WS=1

```
Frame 2735: 62 bytes on wire (496 bits), 62 bytes captured (496 bits) on interface 0
Ethernet II, Src: fa:16:3e:8f:8f:54 (fa:16:3e:8f:8f:54), Dst: fa:16:3e:d7:7e:f6 (fa:16:3e:d7:7e:f6)
Internet Protocol Version 4, Src: 192.168.0.1, Dst: 192.168.0.4
Transmission Control Protocol, Src Port: 179, Dst Port: 32077, Seq: 0, Ack: 1, Len: 0
    Source Port: 179
    Destination Port: 32077
    [Stream index: 25]
    [TCP Segment Len: 0]
    Sequence number: 0    (relative sequence number)
    Acknowledgment number: 1    (relative ack number)
    Header Length: 28 bytes
    Flags: 0x012 (SYN, ACK)
    Window size value: 16384
```

```
    [Calculated window size: 16384]
    Checksum: 0xbf77 [unverified]
    [Checksum Status: Unverified]
    Urgent pointer: 0
    Options: (8 bytes), Maximum segment size, Window scale, End of Option List (EOL)
        Maximum segment size: 1460 bytes
            Kind: Maximum Segment Size (2)
            Length: 4
```

MSS Value: 1460

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

Détails de la session TCP tels qu'ils apparaissent sur R4 - ACTIVE - une fois que la PMTUD TCP est activée et que la session BGP est effacée :

<#root>

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

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

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

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

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

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

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

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

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

Path MTU

```
Request flags: Win Scale
```

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

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

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

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

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

Les détails de la session TCP, tels qu'ils apparaissent sur R1 - PASSIVE - une fois que la PMTUD TCP est activée et que la session BGP est effacée.

<#root>

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

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

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

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

Timer          Starts    Wakeups         Next(msec)
Retrans             6          0                 0
SendWnd             0          0                 0
```

```
TimeWait            0          0               0
AckHold             6          3               0
KeepAlive           1          0               0
PmtuAger            0          0               0
GiveUp              0          0               0
Throttle            0          0               0

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

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

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

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

Path MTU

```
Request flags: Win Scale
```

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

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

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

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

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

PMTUD - Le segment de chemin a une MTU IP inférieure

Le scénario précédent a aidé à comprendre ce qui se passe lors de l'établissement initial d'une session TCP avec PMTUD activé. Ce scénario s'appuie sur le protocole TCP PMTUD et aide à comprendre son fonctionnement et son influence sur les sessions TCP établies.

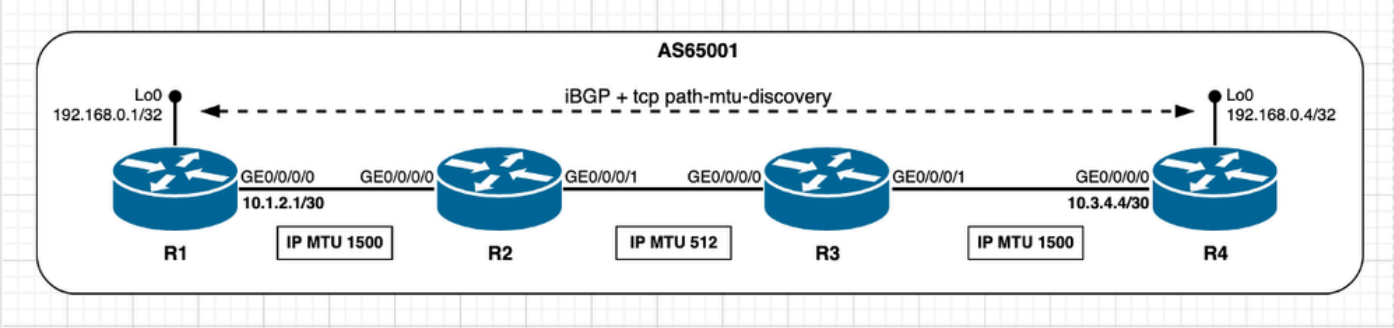


Image 3.2 : PMTUD activé et segment de chemin ayant une MTU IP inférieure.

Considérez l'image précédente comme référence, supposez que la session BGP est établie et que R1 envoie le message de mise à jour BGP transporté par un paquet IP d'une taille supérieure à 512 octets. Lorsque PMTUD est activé, le bit DF (Don’t Fragment) est maintenant défini. Par conséquent, le noeud R2 abandonne le paquet IP et envoie un message ICMP (Internet Control Message Protocol) (Destination Unreachable - type 3 ; Fragmentation requise - code 4) vers R1. Au niveau du noeud R1, après la réception du message ICMP, la PMTUD est déclenchée et tente d’établir le MTU IP le plus bas du chemin. Il le fait en utilisant la valeur inférieure suivante d'un ensemble de niveaux de plateau bien définis, c'est-à-dire une nouvelle valeur MSS de session TCP. TCP retransmet ensuite la mise à jour BGP d'origine avec la nouvelle valeur MSS et ce processus est répété autant de fois que nécessaire jusqu'au message ICMP (Destination Unreachable - type 3 ; Fragmentation requise - Le code 4) n'est plus reçu. Cela signifie que jusqu'à ce que la valeur MSS utilisée soit telle que chaque paquet envoyé tombe sous la valeur MTU IP du segment de chemin le plus bas. Au fil du temps, la PMTUD réglée par le temporisateur PmtuAger passe par les niveaux de plateau dans la direction inverse et remonte la MSS à sa valeur maximale. À tout moment, si un message ICMP (Destination Unreachable - type 3 ; Fragmentation nécessaire - Le code 4) est de nouveau reçu, puis PMTUD agit comme décrit précédemment.

Les sorties suivantes parcourent le comportement PMTUD qui vient d'être décrit et commencent à partir du scénario d'une session TCP établie. Ici, le noeud Cisco IOS XR R4 joue un rôle actif et gère donc la connexion TCP et ouvre la session TCP avec R1 sur le port de destination 179. Les deux noeuds utilisent les valeurs MTU IP par défaut sur leurs interfaces. Le calcul initial de la MSS dans ce scénario peut être résumé comme suit :

- Le segment intermédiaire entre les noeuds R2 et R3 utilise une MTU IP autre que par défaut de 512 octets.
- R1 et R4 utilisent les valeurs MTU par défaut sur leurs interfaces.
- La détection MTU du chemin TCP est activée.
- Les homologues TCP ne sont pas connectés directement.
    - R4 gère la connexion BGP.
    - R4 envoie SYN avec un MSS de 1 460 octets.
        - 1500 (MTU IP d'interface) - 20 (minTCP_H) - 20 (minIP_H).
    - R1 envoie SYN, ACK avec MSS de 1 460 octets.
        - Envoie le plus faible de [MSS reçu ; MSS initiale locale].
        - MSS reçue 1460 octets ; MSS initiale locale de 1 460 octets.
    - La valeur MSS la plus faible est utilisée sur les deux homologues.

SYN TCP provenant de R4 :

<#root>

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

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

MSS=1460

WS=1

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

MSS Value: 1460

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

TCP SYN, ACK provenant de R1 :

<#root>

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

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

MSS=1460

WS=1

```
Frame 393: 62 bytes on wire (496 bits), 62 bytes captured (496 bits) on interface 0
Ethernet II, Src: fa:16:3e:42:18:05 (fa:16:3e:42:18:05), Dst: fa:16:3e:5c:f1:80 (fa:16:3e:5c:f1:80)
Internet Protocol Version 4, Src: 192.168.0.1, Dst: 192.168.0.4
```

```
Transmission Control Protocol, Src Port: 179, Dst Port: 32449, Seq: 0, Ack: 1, Len: 0
    Source Port: 179
    Destination Port: 32449
    [Stream index: 10]
    [TCP Segment Len: 0]
    Sequence number: 0    (relative sequence number)
    Acknowledgment number: 1    (relative ack number)
    Header Length: 28 bytes
    Flags: 0x012 (SYN, ACK)
    Window size value: 16384
    [Calculated window size: 16384]
    Checksum: 0x509e [unverified]
    [Checksum Status: Unverified]
    Urgent pointer: 0
    Options: (8 bytes), Maximum segment size, Window scale, End of Option List (EOL)
        Maximum segment size: 1460 bytes
            Kind: Maximum Segment Size (2)
            Length: 4
```

MSS Value: 1460

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

Une fois la session BGP établie, le noeud R1 envoie le message de mise à jour BGP et reçoit le message ICMP (Destination Unreachable - type 3 ; Fragmentation requise - Code 4) dans le retour provenant du noeud R2.

Cela se produit parce que le paquet IP qui transporte le message de mise à jour BGP a le bit DF défini et le MTU IP de 512 octets utilisé au niveau du segment R2/R3 est inférieur à la taille de paquet IP de 116 octets. Comme expliqué précédemment, la réception du message ICMP déclenche la PMTUD.

Au niveau du protocole ICMP de R1, un message de type 3/code 4 est reçu :

<#root>

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

528    5.893055       192.168.0.1 192.168.0.4 BGP
```

1130

```
   UPDATE Message, KEEPALIVE Message

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

Total Length: 1116

    Identification: 0x8c37 (35895)

Flags: 0x02 (Don't Fragment)

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

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

529   0.002423   10.2.3.1   192.168.0.1 ICMP   110

Destination unreachable (Fragmentation needed)

Frame 529: 110 bytes on wire (880 bits), 110 bytes captured (880 bits) on interface 0
Ethernet II, Src: fa:16:3e:5c:f1:80 (fa:16:3e:5c:f1:80), Dst: fa:16:3e:42:18:05 (fa:16:3e:42:18:05)
Internet Protocol Version 4, Src: 10.2.3.1, Dst: 192.168.0.1
    0100 .... = Version: 4
    .... 0101 = Header Length: 20 bytes (5)
    Differentiated Services Field: 0x00 (DSCP: CS0, ECN: Not-ECT)
    Total Length: 96
    Identification: 0x0001 (1)
    Flags: 0x00
    Fragment offset: 0
    Time to live: 255

    Protocol: ICMP (1)

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

    Type: 3 (Destination unreachable)

    Code: 4 (Fragmentation needed)

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

  MTU of next hop: 512


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

Au niveau du noeud R1, déclenché par un message ICMP, TCP PMTUD tente d'établir le MTU IP le plus bas de bout en bout en utilisant la valeur inférieure suivante à partir d'un ensemble de niveaux de plateau bien définis (IP MTU). Ces niveaux de plateau sont documentés sur RFC1191 - Path MTU discovery.

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

Mais puisque ICMP (Destination inaccessible - type 3 ; Fragmentation requise - Code 4) Le message reçu par le noeud R1 transporte le MTU du saut suivant puis, comme illustré ci-après, le noeud R1 utilise cette valeur, qui dans notre exemple est de 512 octets, et ajuste la valeur MSS

de la session TCP. Notez que la longueur du segment TCP d'origine était de 1 076 octets. Trois paquets sont donc nécessaires pour retransmettre le segment TCP d'origine.

Comme indiqué sur le fonctionnement de R1 - PASSIVE - PMTUD :

<#root>

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

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

Len=472


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

Len=472


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

Len=132
```

Comme indiqué précédemment, une fois que tous les paquets ont été transmis dans le temps, PMTUD parcourt les niveaux de plateau dans le sens inverse régi par le temporisateur PmtuAger et tente d'élever le MSS à sa valeur maximale selon le scénario en place.

Comme on le voit sur R1 - PMTUD sur des plateaux définis :

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

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

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

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

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

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

L'état final peut être observé sur ces sorties. Notez en particulier les valeurs MSS min et max présentées par le noeud R1, qui met en évidence et signale que la PMTUD a été déclenchée.

Détails de la session TCP tels qu'ils apparaissent sur R4 - ACTIVE :

<#root>

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

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

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

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

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

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

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

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

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

Path MTU

```
Request flags: Win Scale
```

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

```
Window scales: rcv 0, snd 0, request rcv 0, request snd 0
Timestamp option: recent 0, recent age 0, last ACK sent 0
```

```
Sack blocks {start, end}: none
Sack holes {start, end, dups, rxmit}: none

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

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

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

Détails de la session TCP tels qu'ils apparaissent sur R1 - PASSIVE :

<#root>

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

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

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

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

Timer          Starts    Wakeups         Next(msec)
Retrans            75          0                 0
SendWnd             0          0                 0
TimeWait            0          0                 0
AckHold            73         71                 0
KeepAlive           1          0                 0

PmtuAger           28         27             41595


GiveUp              0          0                 0
Throttle            0          0                 0

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

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

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

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

Path MTU

```
Request flags: Win Scale
```

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

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

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

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

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

Enfin, si à un moment donné un ICMP (Destination Unreachable - type 3 ; Fragmentation requise - Le message Code 4) est de nouveau reçu, puis PMTUD agit de nouveau comme décrit précédemment.

Comme indiqué par R1 - PASSIVE - PMTUD s'est déclenché à nouveau :

<#root>

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

RP/0/0/CPU0:R1#clear bgp ipv4 all self-originated
```

```
Wed May 12 10:19:06.836 UTC
RP/0/0/CPU0:R1#

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

1707   1.712657       192.168.0.1 192.168.0.4 BGP    1121   UPDATE Message

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

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

1708   0.001614       10.2.3.1       192.168.0.1 ICMP   110

Destination unreachable (Fragmentation needed)



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


Protocol: ICMP (1)


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

    Type: 3 (Destination unreachable)
```

Code: 4 (Fragmentation needed)

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

MTU of next hop: 512

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

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

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

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

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

Timer          Starts    Wakeups         Next(msec)
Retrans            83          0                 0
SendWnd             0          0                 0
TimeWait            0          0                 0
AckHold            80         77                 0
KeepAlive           1          0                 0

PmtuAger           32         30            575401

GiveUp              0          0                 0
Throttle            0          0                 0
```

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

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

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

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

Path MTU


Request flags: Win Scale


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



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

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

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

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

Sur les versions de Cisco IOS XR affectées par l'ID de bogue Cisco CSCvf10395, le saut suivant contenu dans le message d'erreur ICMP est ignoré et le noeud tente d'établir la MTU IP la plus basse de bout en bout en utilisant la valeur inférieure suivante de l'ensemble des niveaux de plateau bien définis (IP MTU) mentionnés précédemment et documentés par RFC1191 - Path MTU discovery. Ces tentatives se produisent jusqu'à la transmission réussie, ce qui signifie jusqu'à ICMP (Destination Unreachable - type 3 ; Fragmentation requise : les messages de code 4 ne sont plus reçus.

Comme on peut le voir à partir d'un noeud avec la version de Cisco IOS XR affectée par l'ID de bogue 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
RP/0/0/CPU0:Feb 23 17:05:32.929 : tcp[399]: [t4] PCB 0x12152adc: Invalid next hop mtu (33554432), ignor
RP/0/0/CPU0:Feb 23 17:05:34.649 : tcp[399]: [t27] PCB 0x12152adc: Trying next lower MTU: 1452 <<<<<<<< H

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

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

À l’étape suivante, considérez le même scénario, mais avec le protocole LDP (Label Distribution Protocol) sur toutes les interfaces. L'objectif ici est de comprendre quelles différences peuvent être observées par rapport aux scénarios précédents dans un environnement compatible MPLS.

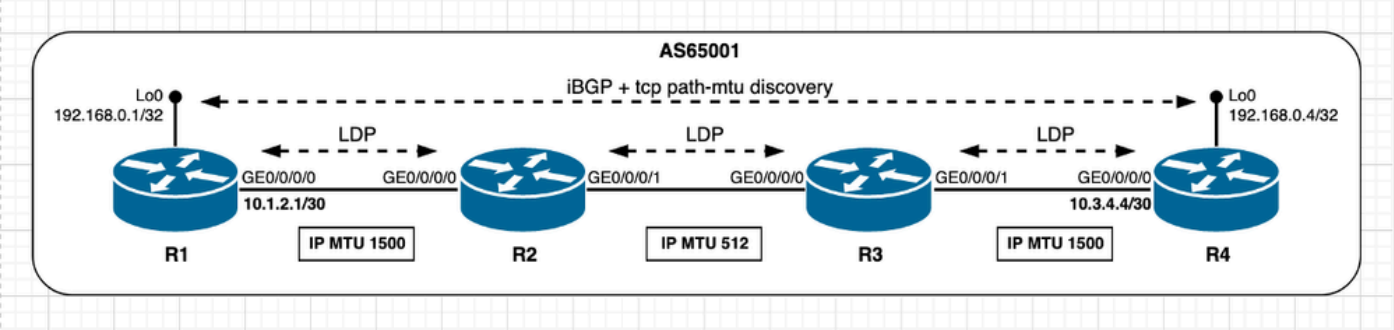


Image 3.3 : PMTUD activé et segment de chemin avec MTU IP inférieur - scénario MPLS.

Tout d'abord, considérez l'étape initiale de la session BGP établie avant le déclenchement PMTUD comme illustré ici.

État initial TCP (BGP) tel qu'il apparaît sur R4 - ACTIF - scénario MPLS activé :

<#root>

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

RP/0/0/CPU0:R4#show tcp detail pcb 0x153bdaf0
Mon May 17 08:32:16.673 UTC
==========================================================
Connection state is ESTAB, I/O status: 0, socket status: 0
Established at Mon May 17 08:31:57 2021

PCB 0x153bdaf0, SO 0x153acc80, TCPCB 0x153acea8, vrfid 0x60000000,
Pak Prio: Medium, TOS: 192, TTL: 255, Hash index: 757
```

```
Local host: 192.168.0.4, Local port: 57400 (Local App PID: 1196319)
Foreign host: 192.168.0.1, Foreign port: 179
(Local App PID/instance/SPL_APP_ID: 1196319/1/0)

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

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

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

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

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

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

Path MTU

```
Request flags: Win Scale
```

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

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

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

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

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

État initial TCP (BGP) tel qu'il apparaît sur R1 - PASSIVE - scénario MPLS :

<#root>

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

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

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

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

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

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

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

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

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

Path MTU


Request flags: Win Scale


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

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

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

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

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

Dans ce scénario compatible MPLS, il est observé que les détails des sessions TCP (LDP) ont été établis. Notez que tous les éléments décrits précédemment en ce qui concerne le calcul MSS pour les sessions TCP (BGP) s'appliquent également aux sessions TCP (LDP). À titre d'exemple, le calcul MSS des sessions TCP (LDP) des noeuds R3 et R2 peut être résumé comme suit :

- R2 et R3 utilisent tous deux une MTU IP non par défaut de 512 octets.
- La détection MTU du chemin est activée.
- Les homologues TCP ne sont pas directement connectés (la session TCP est établie entre les interfaces de bouclage).
    - R3 gère la connexion LDP.
    - R3 envoie SYN avec un MSS de 472 octets.
        - 512 (MTU IP d'interface) - 20 (minTCP_H) - 20 (minIP_H).
    - R2 envoie SYN, ACK avec une MSS de 472 octets.
        - Envoie le plus faible de [Received MSS ; MSS initiale locale].
        - MSS reçue 472 octets ; MSS initiale locale de 472 octets.
    - La valeur MSS la plus faible est utilisée sur les deux homologues.

Détails de session TCP (LDP) tels qu'ils apparaissent sur R3 - ACTIVE - scénario MPLS activé :

<#root>

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

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

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

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

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

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

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

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

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

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

Path MTU


Request flags: Win Scale


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



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

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

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

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

Détails de session TCP (LDP) tels qu'ils apparaissent sur R2 - PASSIVE - scénario MPLS :

<#root>

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

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

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

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

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

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

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

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

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

Path MTU

```
Request flags: Win Scale
```

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

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

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

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

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

Une fois la session BGP établie, R1 envoie le message de mise à jour BGP et reçoit le message ICMP (Destination Unreachable - type 3 ; Fragmentation requise - Code 4) en retour provenant du noeud R2 qui déclenche la PMTUD TCP au niveau du noeud R1. Cela se produit parce que le paquet IP qui transporte le message de mise à jour BGP a le bit DF défini et que la MTU IP de 512 octets utilisée au niveau du segment R2/R3 est inférieure à la taille de paquet IP de 116 octets. Comme précédemment, la réception de ce message ICMP déclenche la PMTUD. La différence dans le scénario MPLS-enabled par rapport aux scénarios non-MPLS précédents est en ce qui concerne la MTU de la valeur de saut suivant incluse sur le message ICMP du noeud R2 (Destination Unreachable - type 3 ; Fragmentation nécessaire - Code 4). Dans ce scénario compatible MPLS, le MTU de la valeur de tronçon suivant tient compte de la surcharge MPLS supplémentaire de 4 octets, ce qui signifie qu'il tient compte de la pile d'étiquettes MPLS de sortie au niveau de R2, comme indiqué dans ces sorties.

Détection de MTU de chemin TCP en action, comme indiqué sur R1 - PASSIVE - scénario MPLS activé :

<#root>

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

455    0.044859       192.168.0.1 192.168.0.4 BGP    1134   UPDATE Message, KEEPALIVE Message
```

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

MultiProtocol Label Switching Header

```
, Label: 24002, Exp: 6, S: 1, TTL: 255
Internet Protocol Version 4, Src: 192.168.0.1, Dst: 192.168.0.4
    0100 .... = Version: 4
    .... 0101 = Header Length: 20 bytes (5)
    Differentiated Services Field: 0xc0 (DSCP: CS6, ECN: Not-ECT)
```

Total Length: 1116

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

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

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

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

456    0.014117       10.2.3.1       192.168.0.1 ICMP   182
```

Destination unreachable (Fragmentation needed)

```
Frame 456: 182 bytes on wire (1456 bits), 182 bytes captured (1456 bits) on interface 0
Ethernet II, Src: fa:16:3e:5c:f1:80 (fa:16:3e:5c:f1:80), Dst: fa:16:3e:42:18:05 (fa:16:3e:42:18:05)
Internet Protocol Version 4, Src: 10.2.3.1, Dst: 192.168.0.1
    0100 .... = Version: 4
    .... 0101 = Header Length: 20 bytes (5)
    Differentiated Services Field: 0x00 (DSCP: CS0, ECN: Not-ECT)
    Total Length: 168
    Identification: 0x001f (31)
    Flags: 0x00
```

```
        0... .... = Reserved bit: Not set
        .0.. .... = Don't fragment: Not set
        ..0. .... = More fragments: Not se
    Fragment offset: 0
    Time to live: 251
```

Protocol: ICMP (1)

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

Type: 3 (Destination unreachable)

Code: 4 (Fragmentation needed)

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

MTU of next hop: 508

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

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

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

Len=468

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

Len=468

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

Len=140

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

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

RP/0/0/CPU0:May 17 08:29:56.131 UTC: tcp[399]: [t1] Try to enable path MTU discovery(neww age timer: 10
RP/0/0/CPU0:May 17 08:29:56.131 UTC: tcp[399]: [t1] Path mtu is ON (age-timer: 10)
RP/0/0/CPU0:May 17 08:35:51.726 UTC: ipv4_io[266]: ip_icmp_lib_ipv4_receive: Receiving pak(0xb0c07d8f)
```

```
RP/0/0/CPU0:May 17 08:35:51.726 UTC: ipv4_io[266]: Entering ipv4_mtu_update_cb
RP/0/0/CPU0:May 17 08:35:51.726 UTC: ipv4_io[266]: IPv4 ICMP: Received ICMP too big from 192.168.0.1 ab
RP/0/0/CPU0:May 17 08:35:51.726 UTC: ipv4_io[266]: ipv4_icmp_unreachable_rcvd ICMP unreach recvd: sendi
RP/0/0/CPU0:May 17 08:35:51.726 UTC: ipv4_io[266]: ip_icmp_lib_ipv4_receive: sending pak(0xb0c07d8f) to
RP/0/0/CPU0:May 17 08:35:51.726 UTC: tcp[399]: [t4] PCB 0x153acc8c: Process ICMP Dest-unreach (next hop

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

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

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

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

Comme le montre R1 - PASSIVE - TCP PMTUD déclenché - scénario MPLS activé :

<#root>

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

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

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

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

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

PmtuAger            1          0           164599


GiveUp              0          0                0
Throttle            0          0                0

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

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

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

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

Path MTU

```
Request flags: Win Scale
```

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

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

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

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

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

Notez que dans le scénario MPLS-enabled, la valeur du MTU du saut suivant inclus sur le noeud R2 ICMP message compte pour la pile d'étiquettes MPLS de sortie. Pour renforcer cet aspect, examinons l'exemple suivant. Si le paquet IP filtré sur R2 est associé à un service L3VPN, cela signifie que la trame Ethernet porte désormais deux étiquettes (étiquette IGP et étiquette VPN). Ensuite, le MTU du saut suivant reflète la taille de pile d'étiquette requise. Reportez-vous à ces sorties.

Comme indiqué sur le paquet de service VPN de couche 3 de R1 - PASSIVE :

<#root>

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

```
2024   0.302370       10.1.14.1      10.1.14.14     TELNET 632    Telnet Data ...

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

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


    0000 0101 1101 1100 0010 .... .... .... = MPLS Label: 24002
    .... .... .... .... .... 000. .... .... = MPLS Experimental Bits: 0
    .... .... .... .... .... ...0 .... .... = MPLS Bottom Of Label Stack: 0
    .... .... .... .... .... .... 1111 1111 = MPLS TTL: 255

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


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


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


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

Comme indiqué sur R1 - PASSIVE - L3 VPN service - ICMP Type 3/Code 4 :

<#root>

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

2030 0.020299 10.2.3.1 10.1.14.1 ICMP 190

Destination unreachable (Fragmentation needed)

Frame 2030: 190 bytes on wire (1520 bits), 190 bytes captured (1520 bits) on interface 0
Ethernet II, Src: fa:16:3e:5c:f1:80 (fa:16:3e:5c:f1:80), Dst: fa:16:3e:42:18:05 (fa:16:3e:42:18:05)
MultiProtocol Label Switching Header, Label: 24005, Exp: 0, S: 1, TTL: 251
    0000 0101 1101 1100 0101 .... = MPLS Label: 24005
    .... 000. .... = MPLS Experimental Bits: 0
    ....1 .... = MPLS Bottom Of Label Stack: 1
    .... 1111 1011 = MPLS TTL: 251
Internet Protocol Version 4, Src: 10.2.3.1, Dst: 10.1.14.1
    0100 .... = Version: 4
    .... 0101 = Header Length: 20 bytes (5)
    Differentiated Services Field: 0x00 (DSCP: CS0, ECN: Not-ECT)
    Total Length: 172
    Identification: 0x002b (43)
    Flags: 0x00
        0... .... = Reserved bit: Not set
        .0.. .... = Don't fragment: Not set
        ..0. .... = More fragments: Not set
    Fragment offset: 0
    Time to live: 253

Protocol: ICMP (1)

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

    Type: 3 (Destination unreachable)

    Code: 4 (Fragmentation needed)

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

MTU of next hop: 504

    Internet Protocol Version 4, Src: 10.1.14.1, Dst: 10.1.14.14
        0100 .... = Version: 4
        .... 0101 = Header Length: 20 bytes (5)
        Differentiated Services Field: 0x00 (DSCP: CS0, ECN: Not-ECT)
        Total Length: 610
        Identification: 0x7c9f (31903)
        Flags: 0x02 (Don't Fragment)
            0... .... = Reserved bit: Not set
            .1.. .... = Don't fragment: Set
            ..0. .... = More fragments: Not set

```
    Fragment offset: 0
    Time to live: 255
    Protocol: TCP (6)
    Header checksum: 0xcce5 [validation disabled]
    [Header checksum status: Unverified]
    Source: 10.1.14.1
    Destination: 10.1.14.14
    [Source GeoIP: Unknown]
    [Destination GeoIP: Unknown]
Transmission Control Protocol, Src Port: 22008, Dst Port: 23, Seq: 586828435, Ack: 754580617
```

PMTUD - Options TCP (MD5)

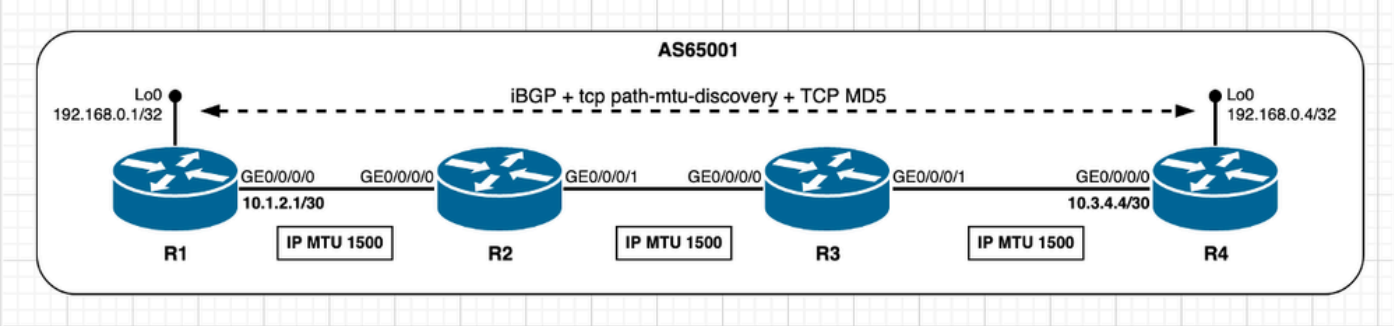


Image 3.4 : PMTUD activé et authentification TCP MD5.

Aucune distinction n'est faite en ce qui concerne le comportement PMTUD par rapport à ce qui a déjà été décrit dans les scénarios précédents avec l'authentification MD5 TCP activée. Comme précédemment partagé avec l'authentification TCP MD5 en cours d'utilisation, Cisco IOS XR considère une surcharge supplémentaire et la valeur MSS initiale de l'homologue TCP actif reflète la même. Reportez-vous aux sections précédentes Utiliser les options TCP - XR actif et Utiliser les options TCP - XR passif pour plus de détails sur l'impact des options TCP utilisées. Dans ce scénario, le calcul MSS TCP peut être résumé comme suit :

- Tous les noeuds utilisent une MTU IP par défaut de 1 500 octets.
- La détection MTU du chemin TCP est activée.
- Les homologues TCP ne sont pas connectés directement.
- Authentification MD5 TCP activée sur R1 et R4.
    - R4 gère la connexion BGP.
    - R4 envoie SYN avec un MSS de 1 436 octets.
        - 1500 (MTU IP d'interface) - 20 (minTCP_H) - 20 (minIP_H) - 24 octets (surcharge des options TCP IOS XR).
    - R1 envoie SYN, ACK avec MSS de 1 436 octets.
        - envoie le plus faible de [MSS reçu ; MSS initiale locale].
        - MSS reçue 1436 octets ; MSS initiale locale de 1 460 octets.
    - La valeur MSS la plus basse est utilisée sur les deux homologues.

SYN TCP provenant de 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 provenant de R1 :

<#root>

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

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

MSS=1436

```
 WS=1

Frame 2409: 82 bytes on wire (656 bits), 82 bytes captured (656 bits) on interface 0
Ethernet II, Src: fa:16:3e:8f:8f:54 (fa:16:3e:8f:8f:54), Dst: fa:16:3e:d7:7e:f6 (fa:16:3e:d7:7e:f6)
Internet Protocol Version 4, Src: 192.168.0.1, Dst: 192.168.0.4
Transmission Control Protocol, Src Port: 179, Dst Port: 59050, Seq: 0, Ack: 1, Len: 0
    Source Port: 179
```

```
    Destination Port: 59050
    [Stream index: 8]
    [TCP Segment Len: 0]
    Sequence number: 0    (relative sequence number)
    Acknowledgment number: 1    (relative ack number)
    Header Length: 48 bytes
    Flags: 0x012 (SYN, ACK)
    Window size value: 16384
    [Calculated window size: 16384]
    Checksum: 0xcbf8 [unverified]
    [Checksum Status: Unverified]
    Urgent pointer: 0
    Options: (28 bytes), Maximum segment size, Window scale, No-Operation (NOP),
```

TCP MD5 signature

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

MSS Value: 1436

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

Détails de la session TCP tels qu'ils apparaissent sur R4 - ACTIVE :

<#root>

```
! - as seen from R4 - Active

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

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

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

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

```
Throttle            0          0                0

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

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

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

State flags: none
Feature flags:
```

MD5

```
, Win Scale, Nagle,
```

Path MTU

```
Request flags: Win Scale
```

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

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

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

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

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

Détails de la session TCP tels qu'ils apparaissent sur R1 - PASSIVE :

<#root>

```
! - as seen from R1 - Passive

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

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

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

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

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

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

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

State flags: none
Feature flags:
```

MD5

, Win Scale, Nagle,

Path MTU

```
Request flags: Win Scale
```

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

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

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

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

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

PMTUD - Détection des trous noirs

Comme expliqué précédemment dans la section PMTUD - Path Segment has Lower IP MTU, la PMTUD TCP lorsqu'elle est activée est déclenchée par la réception d'un ICMP (Destination Unreachable - type 3 ; Fragmentation requise - message de code 4). Il peut arriver que ces messages ne soient pas reçus pour une raison quelconque, ce qui signifie que la PMTUD n'est pas déclenchée. Dans ce cas, le MTU IP le plus bas du chemin entre les homologues TCP n'est pas appris. Un tel scénario introduirait un trou noir potentiel si les paquets IP ont le bit DF défini et s'ils ont une taille supérieure au segment de chemin MTU IP le plus bas. Ces paquets seraient supprimés en silence.

Cette section vise à souligner comment Cisco IOS XR détecte et agit sur un tel scénario de trou noir potentiel. À cette fin, la fonctionnalité IPv4 unreachables est désactivée sur l'interface GE0/0/0/0 de R2, comme indiqué dans l'image suivante et le résultat de l'interface de ligne de commande.

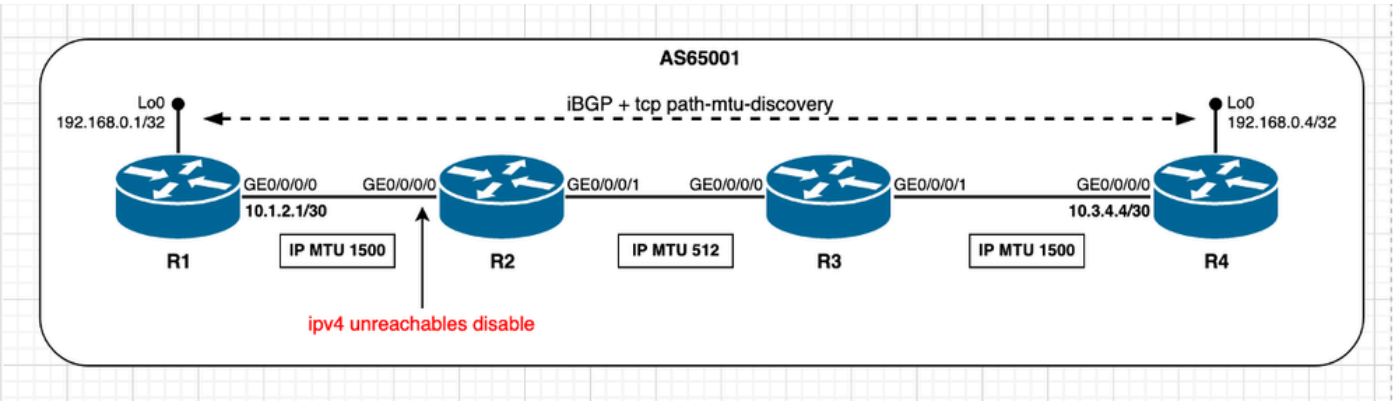


Image 3.5 : PMTUD activé sur les inaccessibles IPv4 de R1/R4 et R2 désactivés.

Inaccessibles IPv4 désactivés sur R2 :

<#root>

```
!- R2 - IP unreachables is disabled

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

ipv4 unreachables disable

```
!

RP/0/0/CPU0:R2#show ipv4 interface gigabitEthernet 0/0/0/0
Thu May 13 12:10:04.112 UTC
GigabitEthernet0/0/0/0 is Up, ipv4 protocol is Up
```

```
  Vrf is default (vrfid 0x60000000)
  Internet address is 10.1.2.2/30
  MTU is 1514 (1500 is available to IP)
  Helper address is not set
  Multicast reserved groups joined: 224.0.0.2 224.0.0.1 224.0.0.5
      224.0.0.6
  Directed broadcast forwarding is disabled
  Outgoing access list is not set
  Inbound  common access list is not set, access list is not set
  Proxy ARP is disabled
  ICMP redirects are never sent
```

ICMP unreachables are never sent

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

La façon dont Cisco IOS XR traite ce scénario de trou noir est de retransmettre le même paquet deux fois et si toujours échoue, c'est-à-dire que l'ACK TCP attendu n'est pas reçu, puis réessayer mais utiliser la valeur de plateau suivante bien définie comme documenté dans RFC1191 - Découverte de MTU de chemin (référez-vous à la section PMTUD - Le segment de chemin a MTU IP inférieure pour la liste des paliers). En résumé, Cisco IOS XR suppose que les paquets peuvent être abandonnés quelque part sur le chemin vers leur destination en raison de leur taille et tente de contourner ce problème par le biais de la retransmission de paquets. Ce comportement peut être observé avec l'exemple suivant d'une capture de paquet prise au niveau de l'interface du noeud R1 et la sortie de debug tcp pmtud.

Détection des trous noirs IOS-XR sur R1 :

<#root>

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

562    7.638774       192.168.0.1 192.168.0.4 BGP    1130   UPDATE Message, KEEPALIVE Message

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

Total Length: 1116

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

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


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

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

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

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

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

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

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

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

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

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

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

571    2.205552       192.168.0.1 192.168.0.4 TCP    522    [TCP Retransmission] 179 → 57082 [ACK] Seq=31
Len=468


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

Len=468

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

Len=140

! – Peer R4 TCP ACK is received

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

À propos de cette traduction
Cisco a traduit ce document en traduction automatisée vérifiée par une personne dans le cadre d'un service mondial permettant à nos utilisateurs d'obtenir le contenu d'assistance dans leur propre langue.

Il convient cependant de noter que même la meilleure traduction automatisée ne sera pas aussi précise que celle fournie par un traducteur professionnel.