Überprüfen der MTU-Pfaderkennung in Cisco IOS XR und BGP

Inhalt

Einleitung

In diesem Dokument wird die PMTUD (Transmission Control Protocol (TCP) Path Maximum Transmission Unit (MTU) Discovery) auf Cisco IOS® XR-Geräten beschrieben.

Hintergrundinformationen

Der PMTUD-Mechanismus versucht, die größte IP-Paketgröße (Internet Protocol) zu ermitteln, die an keiner Stelle im Pfad zwischen zwei Hosts fragmentiert werden muss. Der ermittelte Wert ist als Pfad-MTU festgelegt und entspricht mindestens den MTU-Werten für jeden Hop. Wenn Sie bei der Datenübertragung die MTU-Größe für den Pfad berücksichtigen, können Sie die Netzwerkkapazität optimal nutzen und Fragmentierung und Übertragungseffizienz vermeiden. Die PMTUD-Mechanik und -Implementierung wird in einer Vielzahl von Szenarien mit Border Gateway Protocol (BGP) als Clientprotokoll eingeführt, das das PMTUD-Verhalten schrittweise aufzeigt.

TCP-PMTUD und TCP-MSS

TCP nutzt das PMTUD-Ergebnis, um die lokale Maximum Segment Size (MSS) zu beeinflussen. Das bedeutet, es passt sich dynamisch an die erkannte Path-MTU an. Bevor Sie zur PMTUD übergehen, können Sie daher schnell die maximale TCP-Segmentgröße (MSS) überprüfen und verstehen, was sie bedeutet und welchen Zweck sie hat.

Gemäß MSS-Originaldefinition aus RFC 879: Die Definition der MSS-Option kann wie folgt festgelegt werden: Die maximale Anzahl von Datenoctets, die vom Absender dieser TCP-Option in TCP-Segmenten empfangen werden können, ohne dass in IP-Datagrammen ohne IP-Headeroptionen übertragen wurden.

Um einige Aspekte zu klären und Implementierungsexperten zu beraten, RFC 6691 hebt die Berechnung des MSS-Werts hervor:

Wenn Sie den Wert berechnen, der in die TCP-MSS-Option eingegeben werden soll, sollte der MTU-Wert nur um die Größe der festen IP- und TCP-Header verringert und nicht verringert werden, um mögliche IP- oder TCP-Optionen zu berücksichtigen. Umgekehrt MUSS der Sender die TCP-Datenlänge reduzieren, um alle IP- oder TCP-Optionen zu berücksichtigen, die er in den von ihm gesendeten Paketen berücksichtigt.

Eine ausführlichere Definition von MSS kann aus dem Routing-Konfigurationshandbuch für Cisco Router der Serie ASR 9000, IOS XR, Version 6.7.x, extrahiert werden:

MSS ist die größte Datenmenge, die ein Computer oder ein Kommunikationsgerät in einem einzigen, nicht fragmentierten TCP-Segment empfangen kann. Alle TCP-Sitzungen werden durch eine Beschränkung der Byteanzahl begrenzt, die in einem einzigen Paket übertragen werden kann. dieser Grenzwert ist MSS. TCP unterteilt Pakete in Chunks in einer Übertragungswarteschlange, bevor Pakete an die IP-Schicht weitergeleitet werden.

Der TCP-MSS-Wert ist von der MTU einer Schnittstelle abhängig. Dies ist die maximale Länge von Daten, die von einem Protokoll in einer Instanz übertragen werden können. Die maximale TCP-Paketlänge wird sowohl durch die MTU der ausgehenden Schnittstelle auf dem Quellgerät als auch durch die MSS bestimmt, die das Zielgerät während des TCP-Setup-Prozesses angekündigt hat. Je näher die MSS an der MTU liegt, desto effizienter ist die Übertragung von BGP-Nachrichten. Jede Datenflussrichtung kann einen anderen MSS-Wert verwenden.

Welchen Wert sollte TCP dann für MSS einer TCP-Sitzung berücksichtigen? Und wie wird es berechnet?

Für die Standardwerte gemäß RFC879 verfügen Sie über: Hosts dürfen keine Datagramme mit mehr als 576 Oktetten senden, es sei denn, sie verfügen über spezifische Kenntnisse, dass der Zielhost zum Akzeptieren größerer Datagramme bereit ist. DIE MAXIMALE TCP-SEGMENTGRÖSSE IST DIE MAXIMALE IP-DATAGRAMMGRÖSSE VON MINUS VIER.

Die maximale IP-Datagrammgröße ist standardmäßig 576.

Die standardmäßige maximale TCP-Segmentgröße beträgt 536.

Dabei wird ein IP-MTU-Wert von 576 Byte berücksichtigt. Wenn Sie jedoch den tatsächlichen IP-MTU-Wert ignorieren, kann die TCP-MSS-Berechnung wie folgt zusammengefasst werden:

- Active Peer - berechnet und sendet die anfängliche MSS mit einem SYN-Paket.

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

```
sizeof(minimum TCPHDR) = 20 bytes.
sizeof(minimum IPHDR) = 20 bytes.
```

- Passive Peer - berechnet die anfängliche MSS, vergleicht die empfangene MSS von Active Peer und sendet SYN, ACK mit der niedrigeren dieser MSS-Werte.

```
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.
```

Es gibt keine Verhandlung über den Wert der MSS-Option. Jeder Knoten legt seinen eigenen Wert fest und gibt diesen bei der TCP-Sitzungserstellung an. Wenn der für die MSS-Berechnung berücksichtigte IP-MTU-Wert von der PMTUD abgeleitet werden kann, kann der MSS-Wert auf den effektivsten Wert für eine gegebene Path-MTU angepasst werden. Das Cisco IOS XR-Verhalten weist einige Besonderheiten hinsichtlich der MSS-Berechnung und der PMTUD-Rolle auf, die hier zusammengefasst werden.

Die PMTUD ist in Cisco IOS XR standardmäßig deaktiviert:

- Bei der Berechnung der lokalen anfänglichen MSS wird die IP-MTU wie folgt berücksichtigt: Bei direkt verbundenen Peers sollten Sie die IP-MTU der Ausgangsschnittstelle berücksichtigen.Bei Peers ohne direkte Verbindung: IP-MTU von 1280 Byte.Der MSS-Wert wird durch konfigurierte TCP-Optionen beeinflusst.

Wenn die PMTUD auf Cisco IOS XR aktiviert ist:

- Bei der Berechnung der lokalen anfänglichen MSS wird die IP-MTU wie folgt berücksichtigt: Unabhängig von Peers mit direkter/nicht direkter Verbindung - IP-MTU für die Ausgangsschnittstelle in Betracht ziehen.Der MSS-Wert wird durch konfigurierte TCP-Optionen beeinflusst.

Es gibt weitere Einzelheiten zu den PMTUD-Mechanismen und deren Umsetzung, die berücksichtigt werden müssen und die in diesem Dokument anhand praktischer Beispiele vorgestellt werden, die in der nächsten Tabelle zusammengefasst sind. In dieser Tabelle werden außerdem die IP-MTU der aktiven und passiven TCP-Peers sowie die ausgewählten MSS-Werte für jedes betrachtete Szenario aufgeführt.

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

Szenarien - TCP-PMTUD deaktiviert

Standard-MTU-Werte verwenden

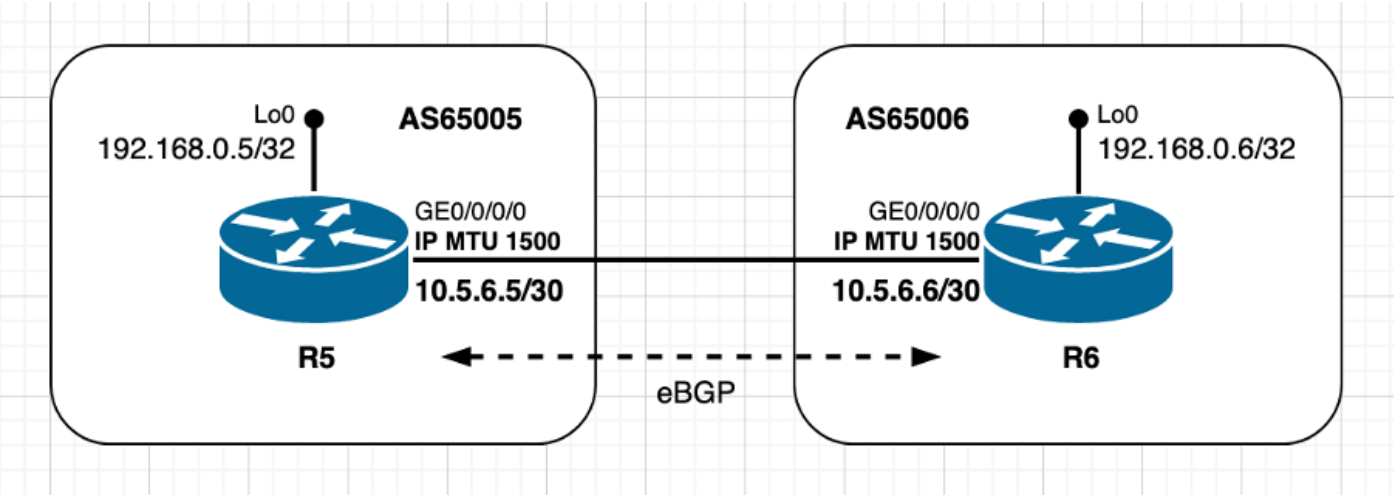

Bild 2.1. Mit Standard-MTU-Werten

Bei den eBGP-Peers, die in Bild 2.1 R6 dargestellt sind, wird die TCP-Verbindung verwaltet, d. h. es übernimmt die aktive Rolle und startet die TCP-Sitzung mit R5 am Ziel-Port 179. Peers sind direkt verbunden, und beide verwenden die standardmäßigen IP-MTU-Werte für die jeweiligen Schnittstellen. Auf der Grundlage der zu Beginn dieses Dokuments geteilten Informationen kann die MSS-Berechnung in diesem Szenario wie folgt zusammengefasst werden:

- Beide Knoten verwenden eine Standard-IP-MTU von 1.500 Byte.
- Die TCP-Pfad-MTU-Erkennung ist standardmäßig deaktiviert.
- TCP-Peers sind direkt verbunden R6 verwaltet die BGP-VerbindungR6 sendet SYN mit einer MSS von 1460 Byte 1500 (Interface IP MTU) - 20 (minTCP_H) - 20 (minIP_H)R5 sendet SYN, ACK mit MSS von 1460 Byte Sendet die untere von [empfangenes MSS; Lokale anfängliche MSS]Empfangene MSS 1460 Byte; Lokale anfängliche MSS 1460 ByteDer niedrigste MSS-Wert wird auf beiden Peers verwendet.

Einzelheiten zu TCP-Sitzungen finden Sie unter R6 - ACTIVE (AKTIV):

```
! – 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
```

Einzelheiten zur TCP-Sitzung finden Sie unter R5 - PASSIVE:

```
! – As seen on R5 – PASSIVE

RP/0/0/CPU0:R5#show interfaces gigabitEthernet 0/0/0/0
Fri Jan  8 09:33:04.564 UTC
GigabitEthernet0/0/0/0 is up, line protocol is up
Interface state transitions: 1
Hardware is GigabitEthernet, address is fa16.3ead.518f (bia fa16.3ead.518f)
Internet address is 10.5.6.5/30
MTU 1514 bytes, BW 1000000 Kbit (Max: 1000000 Kbit)
<snip>

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Nicht standardmäßigen MTU-Wert verwenden - Aktiver TCP-Peer

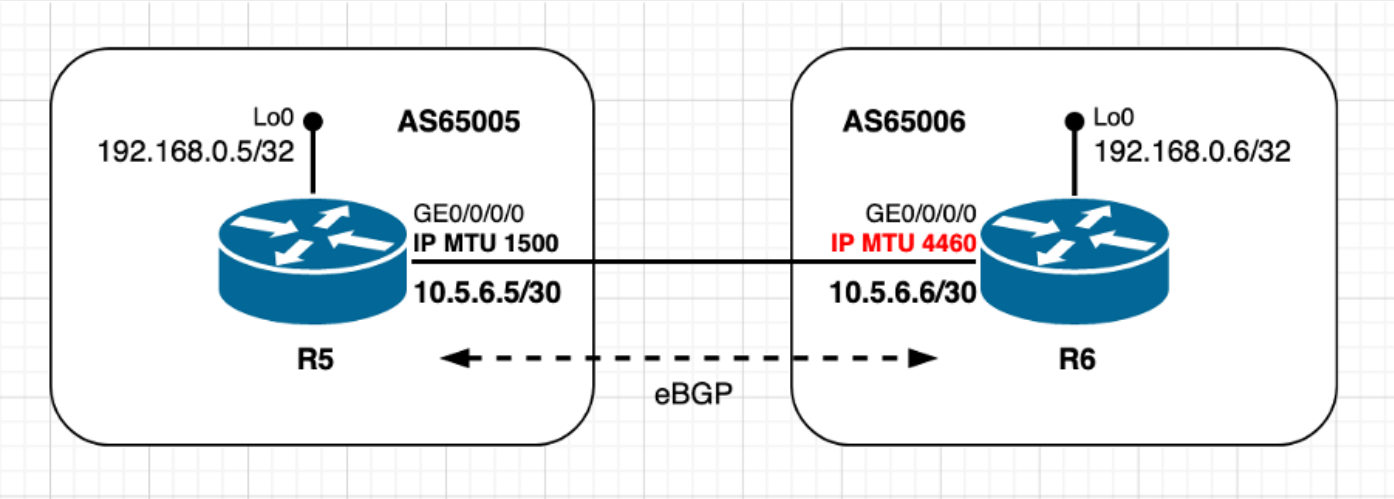

Image 2.2 - ACTIVE Peer verwendet einen nicht standardmäßigen MTU-Wert.

Dieses Szenario ist im Wesentlichen dasselbe wie das vorherige, mit dem einzigen Unterschied, dass aktive TCP-Peer-R6 jetzt einen nicht standardmäßigen IP-MTU-Wert verwendet. Beachten Sie, dass die anfängliche Berechnung und die Entscheidung über den MSS-Wert von passivem TCP Peer R5 erfolgt. Die TCP-MSS-Berechnung in diesem Szenario kann wie folgt zusammengefasst werden:

- R6 verwendet eine nicht standardmäßige IP-MTU-Größe von 4460 Byte.
- Die TCP-Pfad-MTU-Erkennung ist standardmäßig deaktiviert.
- TCP-Peers sind direkt verbunden R6 verwaltet BGP-VerbindungR6 sendet SYN mit einer MSS von 4420 Byte 4460 (Interface IP MTU) - 20 (minTCP_H) - 20 (minIP_H)R5 Send SYN, ACK mit MSS von 1460 Byte sendet die untere von [empfangenes MSS; Lokale anfängliche MSS]MSS 4420 Byte empfangen; Lokale anfängliche MSS 1460 ByteDer niedrigste MSS-Wert wird auf beiden Peers verwendet.

TCP SYN ausgehend von R6:

```
! – TCP SYN sourced from R6

140    1598.150521    10.5.6.6       10.5.6.5       TCP    62     35502  179 [SYN] Seq=0
Win=16384 Len=0 MSS=4420 WS=1

Frame 140: 62 bytes on wire (496 bits), 62 bytes captured (496 bits) on interface 0
Ethernet II, Src: fa:16:3e:85:3d:c2 (fa:16:3e:85:3d:c2), Dst: fa:16:3e:ad:51:8f
```

```
(fa:16:3e:ad:51:8f)
Internet Protocol Version 4, Src: 10.5.6.6, Dst: 10.5.6.5
Transmission Control Protocol, Src Port: 35502, Dst Port: 179, Seq: 0, Len: 0
   Source Port: 35502
   Destination Port: 179
   [Stream index: 6]
   [TCP Segment Len: 0]
   Sequence number: 0    (relative sequence number)
   Acknowledgment number: 0
   Header Length: 28 bytes
   Flags: 0x002 (SYN)
   Window size value: 16384
   [Calculated window size: 16384]
   Checksum: 0x219d [unverified]
   [Checksum Status: Unverified]
   Urgent pointer: 0
   Options: (8 bytes), Maximum segment size, Window scale, End of Option List (EOL)
       Maximum segment size: 4420 bytes
           Kind: Maximum Segment Size (2)
           Length: 4
           MSS Value: 4420
       Window scale: 0 (multiply by 1)
       End of Option List (EOL)
```

TCP-SYN, ACK von R5 bezogen:

```
! – TCP SYN, ACK sourced from R5

141    1598.154866    10.5.6.5       10.5.6.6       TCP    62     179  35502 [SYN, ACK] Seq=0
Ack=1 Win=16384 Len=0 MSS=1460 WS=1

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

Einzelheiten zu TCP-Sitzungen finden Sie unter R6 - ACTIVE (AKTIV):

```
! - 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#
```

Einzelheiten zur TCP-Sitzung finden Sie unter R5 - PASSIVE:

```
! - as seen on R5 – Passive

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

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

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

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

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

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

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

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

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

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

PDU information:
```

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

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

Nicht standardmäßiger MTU-Wert verwenden - Passiver TCP-Peer

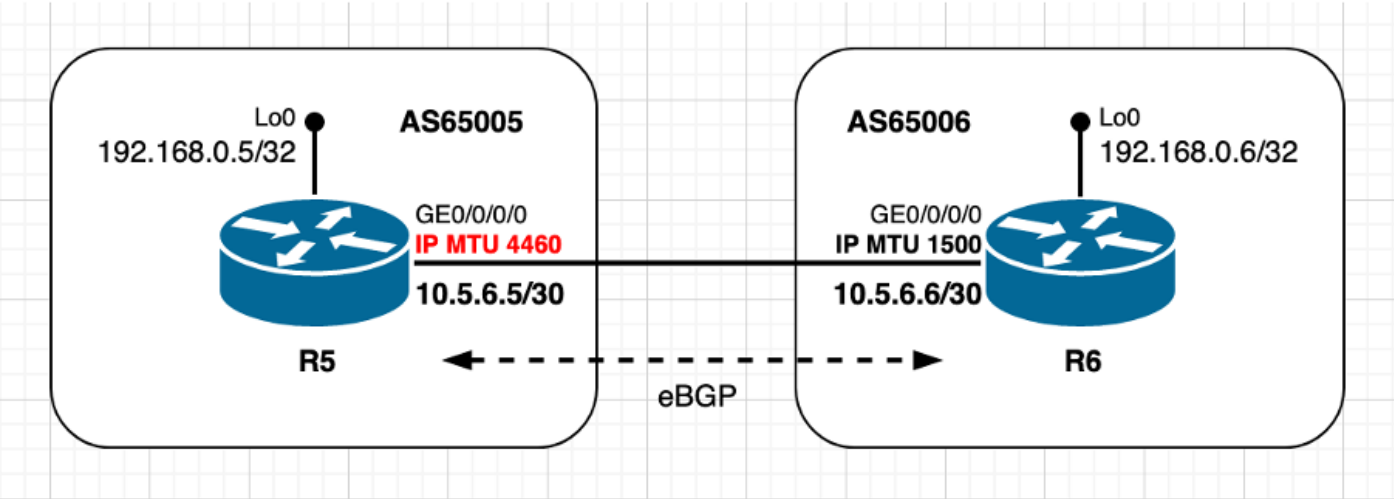

Bild 2.3 - PASSIVE Peer verwendet einen nicht standardmäßigen MTU-Wert.

Mit immer noch demselben eBGP-Szenario, aber jetzt mit passivem TCP-Peer R5 konfiguriert mit einer nicht standardmäßigen IP-MTU und aktivem TCP-Peer R6 mit Standard-IP-MTU-Wert. Beachten Sie wie im vorherigen Szenario, wie der MSS-Wert vom passiven Peer R5 ausgewählt wird. Die TCP-MSS-Berechnung in diesem Szenario kann wie folgt zusammengefasst werden:

- R5 verwendet eine nicht standardmäßige IP-MTU-Größe von 4460 Byte.
- Die TCP-Pfad-MTU-Erkennung ist standardmäßig deaktiviert.
- TCP-Peers sind direkt verbunden R6 verwaltet BGP-VerbindungR6 sendet SYN mit einer MSS von 1460 Byte 1500 (Interface IP MTU) - 20 (minTCP_H) - 20 (minIP_H)R5 Send SYN, ACK mit MSS von 1460 Byte sendet die untere von [empfangenes MSS; Lokale anfängliche MSS]Empfangene MSS 1460 Byte; Lokale anfängliche MSS 4420 ByteDer niedrigste MSS-Wert wird auf beiden Peers verwendet.

TCP SYN ausgehend von R6:

```
! – TCP SYN sourced from R6

237    2696.666481    10.5.6.6       10.5.6.5       TCP    62     47007  179 [SYN] Seq=0
Win=16384 Len=0 MSS=1460 WS=1

Frame 237: 62 bytes on wire (496 bits), 62 bytes captured (496 bits) on interface 0
Ethernet II, Src: fa:16:3e:85:3d:c2 (fa:16:3e:85:3d:c2), Dst: fa:16:3e:ad:51:8f
(fa:16:3e:ad:51:8f)
Internet Protocol Version 4, Src: 10.5.6.6, Dst: 10.5.6.5
Transmission Control Protocol, Src Port: 47007, Dst Port: 179, Seq: 0, Len: 0
   Source Port: 47007
   Destination Port: 179
   [Stream index: 10]
   [TCP Segment Len: 0]
   Sequence number: 0    (relative sequence number)
   Acknowledgment number: 0
   Header Length: 28 bytes
```

```
   Flags: 0x002 (SYN)
   Window size value: 16384
   [Calculated window size: 16384]
   Checksum: 0x2025 [unverified]
   [Checksum Status: Unverified]
   Urgent pointer: 0
   Options: (8 bytes), Maximum segment size, Window scale, End of Option List (EOL)
       Maximum segment size: 1460 bytes
           Kind: Maximum Segment Size (2)
           Length: 4
           MSS Value: 1460
       Window scale: 0 (multiply by 1)
       End of Option List (EOL)
```

TCP-SYN, ACK von R5 bezogen:

```
! – TCP SYN, ACK sourced from R5

238    2696.702792    10.5.6.5       10.5.6.6       TCP    62     179  47007 [SYN, ACK] Seq=0
Ack=1 Win=16384 Len=0 MSS=1460 WS=1

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

Einzelheiten zu TCP-Sitzungen finden Sie unter R6 - ACTIVE (AKTIV):

```
! - 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#
```

Einzelheiten zur TCP-Sitzung finden Sie unter R5 - PASSIVE:

```
! - as seen on R5 – Passive

RP/0/0/CPU0:R5#show interfaces gigabitEthernet 0/0/0/0
Fri Jan  8 10:10:39.110 UTC
GigabitEthernet0/0/0/0 is up, line protocol is up
 Interface state transitions: 1
 Hardware is GigabitEthernet, address is fa16.3ead.518f (bia fa16.3ead.518f)
 Internet address is 10.5.6.5/30
 MTU 4474 bytes, BW 1000000 Kbit (Max: 1000000 Kbit)
<snip>

RP/0/0/CPU0:R5#show tcp detail pcb 0x121550fc
```

```
Fri Jan  8 10:14:20.105 UTC
==========================================================
Connection state is ESTAB, I/O status: 0, socket status: 0
Established at Fri Jan  8 10:10:05 2021

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

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

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

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

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

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

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

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

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

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

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

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

TCP-Optionen verwenden - XR Aktiv

Wie bereits zuvor in diesem Dokument erwähnt, beeinflusst die Verwendung von TCP-Optionen

(z. B. TCP MD5, TCP Selektiv-Rückruf oder TCP-Zeitstempel) die MSS-Berechnung, da diese Optionen zu zusätzlichen Byte führen, die in der MSS-Berechnung berücksichtigt werden.

Dieser Abschnitt und der nächste Abschnitt sollen die MSS-Berechnung veranschaulichen, die von Peers in Gegenwart von TCP-Optionen vorgenommen wird. Als Beispiel wird die TCP-MD5-Authentifizierungsoption verwendet. Weitere Informationen finden Sie im Referenzszenario in Images 2.4, wie im Bild gezeigt.

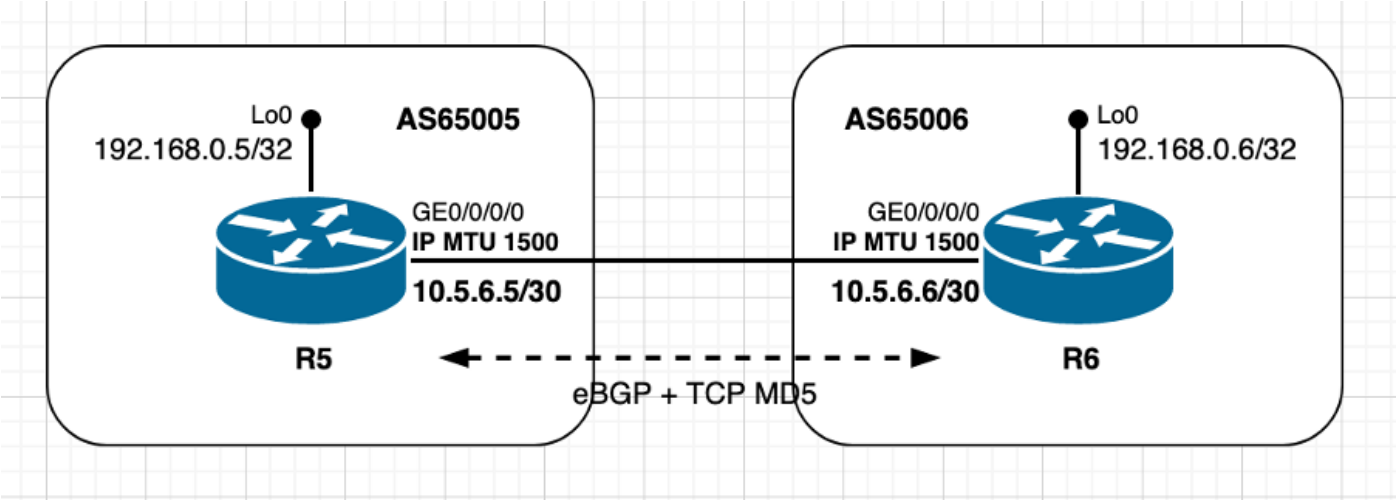

Bild 2.4 - TCP-Optionen (MD5) verwenden - XR Active (Aktiv).

In diesem Szenario verwenden beide Peers standardmäßige IP-MTU-Werte, sind direkt verbunden, und Peer R6 übernimmt die aktive Rolle des TCP. Wie bereits freigegeben, wird die Konfiguration und Verwendung des TCP-MD5-Authentifizierungskontos für zusätzlichen Overhead. Die TCP-MSS-Berechnung in diesem speziellen Szenario kann wie folgt zusammengefasst werden:

- Beide Knoten verwenden eine Standard-IP-MTU von 1.500 Byte.
- Die TCP-Pfad-MTU-Erkennung ist standardmäßig deaktiviert.
- TCP-Peers sind direkt verbunden
- TCP-MD5-Authentifizierung auf beiden Knoten aktiviert R6 verwaltet BGP-VerbindungR6 sendet SYN mit einer MSS von 1436 Byte 1500 (Interface IP MTU) - 20 (minTCP_H) - 20 (minIP_H) - 24 Byte (IOS XR TCP Options Overhead)R5 Send SYN, ACK mit MSS von 1436 Byte sendet die untere von [empfangenes MSS; Lokale anfängliche MSS]Empfangene MSS 1436 Byte Lokale anfängliche MSS 1460 ByteDer niedrigste MSS-Wert wird auf beiden Peers verwendet.

Wie aus der Zusammenfassung ersichtlich, entspricht das Verhalten von Cisco IOS XR nicht ausschließlich dem RFC 879 und RFC 6691, in denen angegeben ist, dass die TCP-Optionen nicht in der MSS-Berechnung berücksichtigt werden sollten.

Der zusätzliche Faktor für die **TCP-Headerlänge** von Cisco IOS XR ist in der Cisco Bug-ID CSCvf20166 weiter dokumentiert:

"(..) Wenn XR die BGP-Verbindung startet, erstellt BGP zunächst den Socket und legt dann die Socket-Optionen einschließlich **MD5 fest**. Dadurch wird die **Länge des Headers "tcp option" 24**. Daher wird die anfängliche MSS zu 1500 - 40 - 24 = 1436. Diese wird an Peer- und Peer-Benutzer mit min(1436, 1460) = 1436.(...) gesendet.

TCP SYN ausgehend von R6:

```
! – TCP SYN sourced from R6

430    5775.839420    10.5.6.6       10.5.6.5       TCP    82     24785  179 [SYN] Seq=0
Win=16384 Len=0 MSS=1436 WS=1

Frame 430: 82 bytes on wire (656 bits), 82 bytes captured (656 bits) on interface 0
Ethernet II, Src: fa:16:3e:85:3d:c2 (fa:16:3e:85:3d:c2), Dst: fa:16:3e:ad:51:8f
(fa:16:3e:ad:51:8f)
Internet Protocol Version 4, Src: 10.5.6.6, Dst: 10.5.6.5
Transmission Control Protocol, Src Port: 24785, Dst Port: 179, Seq: 0, Len: 0
   Source Port: 24785
   Destination Port: 179
   [Stream index: 14]
   [TCP Segment Len: 0]
   Sequence number: 0    (relative sequence number)
   Acknowledgment number: 0
   Header Length: 48 bytes
   Flags: 0x002 (SYN)
   Window size value: 16384
   [Calculated window size: 16384]
   Checksum: 0xd62b [unverified]
   [Checksum Status: Unverified]
   Urgent pointer: 0
   Options: (28 bytes), Maximum segment size, Window scale, No-Operation (NOP), TCP MD5
signature, End of Option List (EOL)
       Maximum segment size: 1436 bytes
           Kind: Maximum Segment Size (2)
           Length: 4
           MSS Value: 1436
       Window scale: 0 (multiply by 1)
       No-Operation (NOP)
       TCP MD5 signature
       End of Option List (EOL)
```

TCP-SYN, ACK von R5 bezogen:

```
! – TCP SYN, ACK sourced from R5

431    5775.845744    10.5.6.5       10.5.6.6       TCP    82     179  24785 [SYN, ACK] Seq=0
Ack=1 Win=16384 Len=0 MSS=1436 WS=1

Frame 431: 82 bytes on wire (656 bits), 82 bytes captured (656 bits) on interface 0
Ethernet II, Src: fa:16:3e:ad:51:8f (fa:16:3e:ad:51:8f), Dst: fa:16:3e:85:3d:c2
(fa:16:3e:85:3d:c2)
Internet Protocol Version 4, Src: 10.5.6.5, Dst: 10.5.6.6
Transmission Control Protocol, Src Port: 179, Dst Port: 24785, Seq: 0, Ack: 1, Len: 0
   Source Port: 179
   Destination Port: 24785
   [Stream index: 14]
   [TCP Segment Len: 0]
   Sequence number: 0    (relative sequence number)
   Acknowledgment number: 1    (relative ack number)
   Header Length: 48 bytes
   Flags: 0x012 (SYN, ACK)
   Window size value: 16384
   [Calculated window size: 16384]
   Checksum: 0xe83d [unverified]
   [Checksum Status: Unverified]
   Urgent pointer: 0
   Options: (28 bytes), Maximum segment size, Window scale, No-Operation (NOP), TCP MD5
signature, End of Option List (EOL)
       Maximum segment size: 1436 bytes
```

```
           Kind: Maximum Segment Size (2)
           Length: 4
           MSS Value: 1436
       Window scale: 0 (multiply by 1)
       No-Operation (NOP)
       TCP MD5 signature
       End of Option List (EOL)
```

Einzelheiten zu TCP-Sitzungen finden Sie unter R6 - ACTIVE (AKTIV):

```
! - 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#
```

Einzelheiten zur TCP-Sitzung finden Sie unter R5 - PASSIVE:

```
! - as seen on R5 – Passive

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

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

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

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

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

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

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

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

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

Socket options: SO_REUSEADDR, SO_REUSEPORT, SO_NBIO
Socket states: SS_ISCONNECTED, SS_PRIV
Socket receive buffer states: SB_DEL_WAKEUP
```

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

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

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

Ähnliches Verhalten kann auch bei anderen TCP-Optionen beobachtet werden, die bei der Konfiguration zusätzliche Gemeinkosten verursachen und die MSS-Berechnung in Cisco IOS XR beeinflussen. Betrachten Sie das gleiche Szenario und diese Beispiele, in denen die MSS-Berechnung dokumentiert wird, wenn TCP-Zeitstempel und TCP-Optionen für selektive Rückkopplung konfiguriert werden.

TCP-Sitzungsdetails wie unter R6 - ACTIVE - mit konfiguriertem Zeitstempel und selektiven Rückgabeoptionen:

```
! - 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>
```

TCP-Optionen verwenden - XR passiv

Im vorherigen Szenario ist Ihnen vermutlich das unterschiedliche Verhalten des Cisco IOS XR-Knotens aufgefallen, wenn er in passiver Funktion in Bezug auf die anfängliche MSS-Berechnung auftritt. Der Knoten berücksichtigt nicht die **Headerlänge der tcp-Option**. Dieses Szenario soll dieses unterschiedliche Verhalten hervorheben, das auch durch die Cisco Bug-ID beschrieben wird:

"(...) - Wenn der Peer die Verbindung initiiert, sendet er die erste MSS als 1460. XR TCP erstellt Socket, pcb usw. Anschließend werden in der angegebenen Reihenfolge zwei Aktionen ausgeführt:

- Zunächst wird die anfängliche MSS nach Subtrahieren der **Headerlänge der tcp-Option** berechnet. Dies ist '0', da die MD5-Option noch nicht von Listen-Socket auf diesen Socket geerbt wurde.

- Dann erbt es die 'MD5' und andere Optionen und das macht 'option header bytes länge' zu 24.

In diesem Fall sendet XR TCP also die erste MSS als 1460 und wird daher von beiden verwendet. (..)"

In diesem Szenario führt der aktive TCP-Peer R8 zwar zu einem Cisco IOS-Knoten, aber diese Tatsache führt nicht zu einem Unterschied oder zu Einzelheiten hinsichtlich der Zielsetzung des Szenarios. Beachten Sie jedoch, und interessanterweise, dass anders als Cisco IOS XR, wie im vorherigen Abschnitt Szenario gezeigt, hier der aktive TCP-Peer R8 bei der anfänglichen MSS-Berechnung keine TCP-Optionen berücksichtigt.

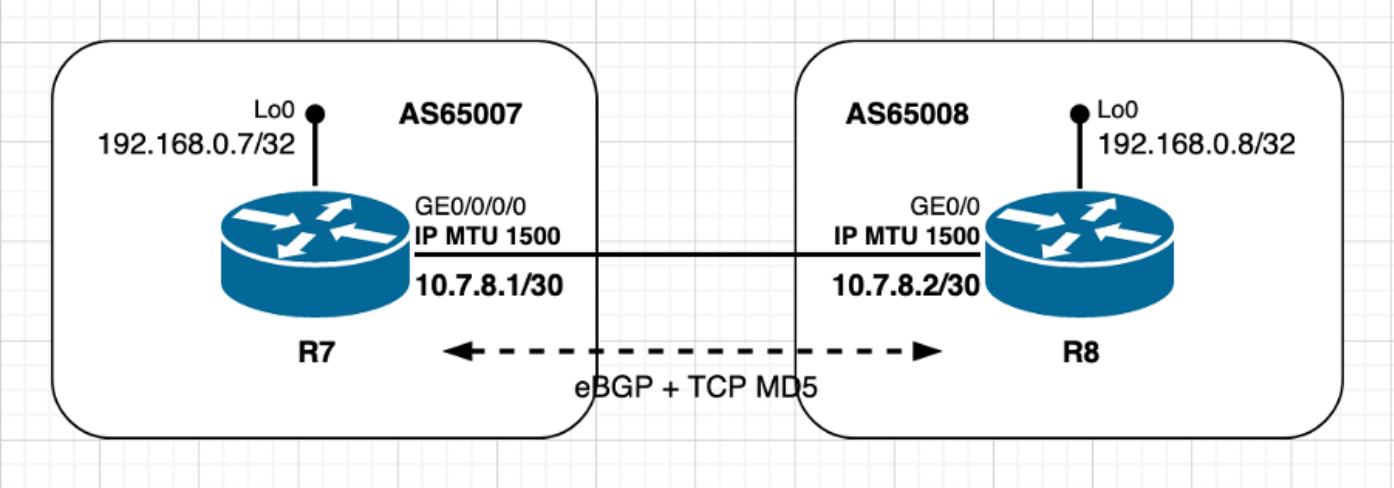

Bild 2.5 - TCP-Optionen (MD5) - XR Passiv.

Beide Peers verwenden Standard-IP-MTU-Werte und sind direkt verbunden. Cisco IOS Peer R8 spielt eine aktive Rolle. Die TCP-MSS-Berechnung in diesem Szenario kann wie folgt zusammengefasst werden:

- Beide Knoten verwenden eine Standard-IP-MTU von 1.500 Byte.
- Die TCP-Pfad-MTU-Erkennung ist in Cisco IOS XR R7 standardmäßig deaktiviert.
- Die TCP-Pfad-MTU-Erkennung ist in Cisco IOS R8 standardmäßig aktiviert.
- TCP-Peers sind direkt verbunden
- TCP-MD5-Authentifizierung auf beiden Knoten aktiviert IOS R8 verwaltet die BGP-VerbindungIOS R8 sendet SYN mit einer MSS von 1460 Byte 1500 (Interface IP MTU) - 20 (minTCP_H) - 20 (minIP_H)IOS XR7 sendet SYN, ACK mit MSS von 1460 Byte sendet die untere von [empfangenes MSS; Lokale anfängliche MSS]Empfangene MSS 1460 Byte; Lokale anfängliche MSS 1460 ByteDer niedrigste MSS-Wert wird auf beiden Peers verwendet.

TCP SYN ausgehend von R8 - Cisco IOS:

```
! – TCP SYN sourced from R8

96     5.907127       10.7.8.2       10.7.8.1       TCP    78     52975  179 [SYN] Seq=0
Win=16384 Len=0 MSS=1460

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

```
   Options: (24 bytes), Maximum segment size, TCP MD5 signature, End of Option List (EOL)
       Maximum segment size: 1460 bytes
           Kind: Maximum Segment Size (2)
           Length: 4
           MSS Value: 1460
       TCP MD5 signature
       End of Option List (EOL)
```

TCP SYN, ACK von R7 - Cisco IOS XR:

```
! – TCP SYN,ACK sourced from R7

97     0.003446       10.7.8.1       10.7.8.2       TCP    78     179  52975 [SYN, ACK] Seq=0
Ack=1 Win=16384 Len=0 MSS=1460

Frame 97: 78 bytes on wire (624 bits), 78 bytes captured (624 bits) on interface 0
Ethernet II, Src: fa:16:3e:68:d9:e5 (fa:16:3e:68:d9:e5), Dst: fa:16:3e:58:21:ba
(fa:16:3e:58:21:ba)
Internet Protocol Version 4, Src: 10.7.8.1, Dst: 10.7.8.2
Transmission Control Protocol, Src Port: 179, Dst Port: 52975, Seq: 0, Ack: 1, Len: 0
   Source Port: 179
   Destination Port: 52975
   [Stream index: 3]
   [TCP Segment Len: 0]
   Sequence number: 0    (relative sequence number)
   Acknowledgment number: 1    (relative ack number)
   Header Length: 44 bytes
   Flags: 0x012 (SYN, ACK)
   Window size value: 16384
   [Calculated window size: 16384]
   Checksum: 0xfb47 [unverified]
   [Checksum Status: Unverified]
   Urgent pointer: 0
   Options: (24 bytes), Maximum segment size, TCP MD5 signature, End of Option List (EOL)
       Maximum segment size: 1460 bytes
           Kind: Maximum Segment Size (2)
           Length: 4
           MSS Value: 1460
       TCP MD5 signature
       End of Option List (EOL)
```

Einzelheiten zu TCP-Sitzungen finden Sie unter R8 - Cisco IOS - ACTIVE:

```
! – as seen from R8 - Cisco IOS

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

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

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

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

 Address tracking is enabled, the RIB does have a route to 10.7.8.1
 Connections established 6; dropped 5
 Last reset 00:06:18, due to BGP Notification received of session 1, Administrative Reset
 External BGP neighbor configured for connected checks (single-hop no-disable-connected-check)
 Interface associated: GigabitEthernet0/1 (peering address in same link)
 Transport(tcp) path-mtu-discovery is enabled
 Graceful-Restart is disabled
 SSO is disabled
Connection state is ESTAB, I/O status: 1, unread input bytes: 0
Connection is ECN Disabled, Mininum incoming TTL 0, Outgoing TTL 1
Local host: 10.7.8.2, Local port: 52975
Foreign host: 10.7.8.1, Foreign port: 179
Connection tableid (VRF): 0
Maximum output segment queue size: 50

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

Event Timers (current time is 0x15DD97):
Timer          Starts    Wakeups            Next
```

```
Retrans            10          0             0x0
TimeWait            0          0             0x0
AckHold             9          5             0x0
SendWnd             0          0             0x0
KeepAlive           0          0             0x0
GiveUp              0          0             0x0
PmtuAger            1          0         0x195465
DeadWait            0          0             0x0
Linger              0          0             0x0
ProcessQ            0          0             0x0

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

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

SRTT: 737 ms, RTTO: 2506 ms, RTV: 1769 ms, KRTT: 0 ms
minRTT: 7 ms, maxRTT: 1000 ms, ACK hold: 200 ms
uptime: 372981 ms, Sent idletime: 16648 ms, Receive idletime: 16431 ms
Status Flags: active open
Option Flags: nagle, path mtu capable, md5
IP Precedence value : 6

Datagrams (max data segment is 1460 bytes):
Rcvd: 18 (out of order: 0), with data: 8, total data bytes: 209
Sent: 16 (retransmit: 0, fastretransmit: 0, partialack: 0, Second Congestion: 0), with data: 9,
total data bytes: 213

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

R8#
```

Einzelheiten zu TCP-Sitzungen finden Sie unter R7 - Cisco IOS XR - PASSIVE:

```
! – 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#
```

TCP-Peers nicht direkt verbunden

Wenn Peers nicht direkt miteinander verbunden sind, ändert sich die Vorgehensweise für die anfängliche TCP-MSS-Berechnung, wie bereits im einleitenden Abschnitt dieses Dokuments beschrieben. Das Szenario einer iBGP-Sitzung mit allen Peers, die mit standardmäßigen IP-MTU-Werten konfiguriert sind, wird für die MSS-Berechnung verwendet.

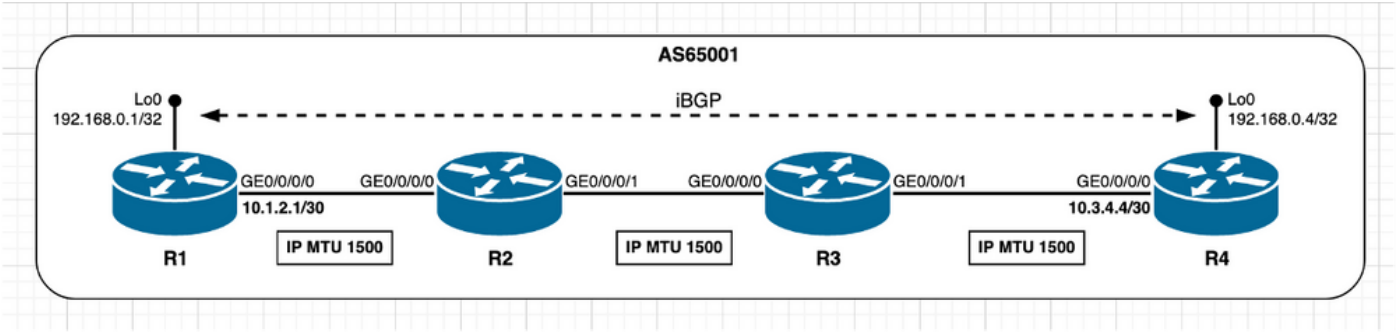

Bild 2.6 - Nicht direkt verbundene TCP-Peers - iBGP.

Der wichtige Aspekt ist, dass Cisco IOS XR einen festen IP-MTU-Wert von 1280 Byte verwendet, wenn die TCP-Path-MTU-Erkennung deaktiviert ist und Peers nicht direkt verbunden sind.

Im vorherigen Bild übernimmt R4 die aktive Rolle und verwaltet die TCP-Verbindung. R4 öffnet die

TCP-Sitzung mit R1 am Zielport 179. Beide Knoten verwenden den IP-MTU-Standardwert ihrer Schnittstellen. Die MSS-Berechnung in diesem Szenario kann wie folgt zusammengefasst werden:

- Alle Knoten verwenden eine Standard-IP-MTU von 1.500 Byte.
- Die TCP-Pfad-MTU-Erkennung ist standardmäßig deaktiviert.
- TCP-Peers sind nicht direkt verbunden R4 verwaltet die BGP-VerbindungR4 sendet SYN mit MSS von 1240 Byte Die Schnittstellen-MTU wird nicht berücksichtigt, wenn Peers nicht direkt verbunden sind und die TCP-Pfad-MTU-Erkennung deaktiviert ist.Laut Cisco IOS XR-Design gelten 1280 Byte als TCP_DEFAULT_MTU.1280 (TCP_DEFAULT_MTU) - 20 (minTCP_H) - 20 (minIP_H)R1 sendet SYN, ACK mit MSS von 1240 Byte Sendet die untere von [empfangenes MSS; Lokale anfängliche MSS]MSS 1240 Byte empfangen; Lokale anfängliche MSS 1240 ByteDer niedrigste MSS-Wert wird auf beiden Peers verwendet.

TCP SYN ausgehend von R4:

```
! – TCP SYN sourced from R4

194    434.274181     192.168.0.4 192.168.0.1 TCP    62     37740  179 [SYN] Seq=0 Win=16384
Len=0 MSS=1240 WS=1

Frame 194: 62 bytes on wire (496 bits), 62 bytes captured (496 bits) on interface 0
Ethernet II, Src: fa:16:3e:d7:7e:f6 (fa:16:3e:d7:7e:f6), Dst: fa:16:3e:8f:8f:54
(fa:16:3e:8f:8f:54)
Internet Protocol Version 4, Src: 192.168.0.4, Dst: 192.168.0.1
Transmission Control Protocol, Src Port: 37740, Dst Port: 179, Seq: 0, Len: 0
   Source Port: 37740
   Destination Port: 179
   [Stream index: 7]
   [TCP Segment Len: 0]
   Sequence number: 0    (relative sequence number)
   Acknowledgment number: 0
   Header Length: 28 bytes
   Flags: 0x002 (SYN)
   Window size value: 16384
   [Calculated window size: 16384]
   Checksum: 0x8643 [unverified]
   [Checksum Status: Unverified]
   Urgent pointer: 0
   Options: (8 bytes), Maximum segment size, Window scale, End of Option List (EOL)
       Maximum segment size: 1240 bytes
           Kind: Maximum Segment Size (2)
           Length: 4
           MSS Value: 1240
       Window scale: 0 (multiply by 1)
       End of Option List (EOL)
```

TCP-SYN, ACK von R1 stammt:

```
! – TCP SYN,ACK sourced from R1

195    434.277985     192.168.0.1 192.168.0.4 TCP    62     179  37740 [SYN, ACK] Seq=0 Ack=1
Win=16384 Len=0 MSS=1240 WS=1

Frame 195: 62 bytes on wire (496 bits), 62 bytes captured (496 bits) on interface 0
Ethernet II, Src: fa:16:3e:8f:8f:54 (fa:16:3e:8f:8f:54), Dst: fa:16:3e:d7:7e:f6
(fa:16:3e:d7:7e:f6)
Internet Protocol Version 4, Src: 192.168.0.1, Dst: 192.168.0.4
```

```
Transmission Control Protocol, Src Port: 179, Dst Port: 37740, Seq: 0, Ack: 1, Len: 0
   Source Port: 179
   Destination Port: 37740
   [Stream index: 7]
   [TCP Segment Len: 0]
   Sequence number: 0    (relative sequence number)
   Acknowledgment number: 1    (relative ack number)
   Header Length: 28 bytes
   Flags: 0x012 (SYN, ACK)
   Window size value: 16384
   [Calculated window size: 16384]
   Checksum: 0xd8f7 [unverified]
   [Checksum Status: Unverified]
   Urgent pointer: 0
   Options: (8 bytes), Maximum segment size, Window scale, End of Option List (EOL)
       Maximum segment size: 1240 bytes
           Kind: Maximum Segment Size (2)
           Length: 4
           MSS Value: 1240
       Window scale: 0 (multiply by 1)
       End of Option List (EOL)
```

Einzelheiten zur TCP-Sitzung finden Sie unter R4 - ACTIVE (AKTIV):

```
! – 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#
```

Einzelheiten zur TCP-Sitzung finden Sie unter R1 - PASSIVE:

```
! – 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#
```

Nicht direkt verbundene TCP-Peers - TCP-Optionen (MD5) verwenden

Bei einem Peer-Szenario ohne direkte Verbindung und bei Verwendung der TCP-MD5-Authentifizierung gibt es keinen grundsätzlichen Unterschied zu den zuvor beschriebenen Testfällen oder Szenarien. Wie bereits bei der TCP-MD5-Authentifizierung festgestellt, berücksichtigt Cisco IOS XR den zusätzlichen Overhead und den anfänglichen MSS-Wert spiegelt diesen wider. Weitere Informationen zu TCP-Optionen, die sich auf die TCP-MSS-Berechnung auswirken, finden Sie in den vorherigen Abschnitten TCP-Optionen verwenden - XR Aktiv und TCP-Optionen verwenden - XR Passiv.

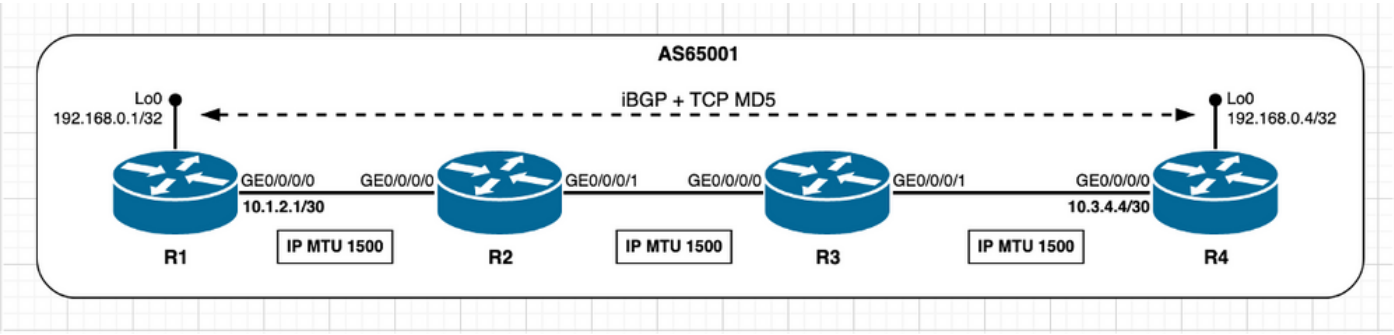

Bild 2.7 - Nicht direkt verbundene TCP-Peers - iBGP + TCP MD5.

Die TCP-MSS-Berechnung in diesem Szenario kann wie folgt zusammengefasst werden:

- Alle Knoten verwenden eine Standard-IP-MTU von 1.500 Byte.
- Die TCP-Pfad-MTU-Erkennung ist standardmäßig deaktiviert.
- TCP-Peers sind nicht direkt verbunden R4 verwaltet die BGP-VerbindungZiel R1 ist nicht direkt verbundenR4 sendet SYN mit MSS von 1216 Byte Die Schnittstellen-MTU wird nicht berücksichtigt, wenn Peers nicht direkt verbunden sind und die TCP-Pfad-MTU-Erkennung deaktiviert ist.Laut Design gelten 1280 Byte als TCP_DEFAULT_MTU.1280 (TCP_DEFAULT_MTU) - 20 (minTCP_H) - 20 (minIP_H) - 24 Byte (IOS XR TCP Options

Overhead)R1 sendet SYN, ACK mit MSS von 1216 Byte Sendet die untere von [empfangenes MSS; Lokale anfängliche MSS]MSS 1216 Byte empfangen; Lokale anfängliche MSS 1240 ByteDer niedrigste MSS-Wert wird auf beiden Peers verwendet.

TCP SYN ausgehend von R4:

```
! – TCP SYN sourced from R4

3425   3.691042       192.168.0.4 192.168.0.1 TCP    82     42135  179 [SYN] Seq=0 Win=16384
Len=0 MSS=1216 WS=1

Frame 3425: 82 bytes on wire (656 bits), 82 bytes captured (656 bits) on interface 0
Ethernet II, Src: fa:16:3e:d7:7e:f6 (fa:16:3e:d7:7e:f6), Dst: fa:16:3e:8f:8f:54
(fa:16:3e:8f:8f:54)
Internet Protocol Version 4, Src: 192.168.0.4, Dst: 192.168.0.1
Transmission Control Protocol, Src Port: 42135, Dst Port: 179, Seq: 0, Len: 0
   Source Port: 42135
   Destination Port: 179
   [Stream index: 10]
   [TCP Segment Len: 0]
   Sequence number: 0    (relative sequence number)
   Acknowledgment number: 0
   Header Length: 48 bytes
   Flags: 0x002 (SYN)
   Window size value: 16384
   [Calculated window size: 16384]
   Checksum: 0xc503 [unverified]
   [Checksum Status: Unverified]
   Urgent pointer: 0
   Options: (28 bytes), Maximum segment size, Window scale, No-Operation (NOP), TCP MD5
signature, End of Option List (EOL)
       Maximum segment size: 1216 bytes
           Kind: Maximum Segment Size (2)
           Length: 4
           MSS Value: 1216
       Window scale: 0 (multiply by 1)
       No-Operation (NOP)
       TCP MD5 signature
       End of Option List (EOL)
```

TCP-SYN, ACK von R1 stammt:

```
! – TCP SYN,ACK sourced from R1

3426   0.004186       192.168.0.1 192.168.0.4 TCP    82     179  42135 [SYN, ACK] Seq=0 Ack=1
Win=16384 Len=0 MSS=1216 WS=1

Frame 3426: 82 bytes on wire (656 bits), 82 bytes captured (656 bits) on interface 0
Ethernet II, Src: fa:16:3e:8f:8f:54 (fa:16:3e:8f:8f:54), Dst: fa:16:3e:d7:7e:f6
(fa:16:3e:d7:7e:f6)
Internet Protocol Version 4, Src: 192.168.0.1, Dst: 192.168.0.4
Transmission Control Protocol, Src Port: 179, Dst Port: 42135, Seq: 0, Ack: 1, Len: 0
   Source Port: 179
   Destination Port: 42135
   [Stream index: 10]
   [TCP Segment Len: 0]
   Sequence number: 0    (relative sequence number)
   Acknowledgment number: 1    (relative ack number)
   Header Length: 48 bytes
   Flags: 0x012 (SYN, ACK)
   Window size value: 16384
```

```
   [Calculated window size: 16384]
   Checksum: 0xbb05 [unverified]
   [Checksum Status: Unverified]
   Urgent pointer: 0
   Options: (28 bytes), Maximum segment size, Window scale, No-Operation (NOP), TCP MD5
signature, End of Option List (EOL)
       Maximum segment size: 1216 bytes
           Kind: Maximum Segment Size (2)
           Length: 4
           MSS Value: 1216
       Window scale: 0 (multiply by 1)
       No-Operation (NOP)
       TCP MD5 signature
       End of Option List (EOL)
```

Einzelheiten zur TCP-Sitzung finden Sie unter R4 - ACTIVE (AKTIV):

```
! – 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#
```

Einzelheiten zur TCP-Sitzung finden Sie unter R1 - PASSIVE:

```
! – 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#
```

Nicht direkt verbundene TCP-Peers - Pfadsegment hat niedrigere IP-MTU

Im nächsten Szenario soll beobachtet und entschieden werden, was passiert, wenn ein Zwischenpfad-Segment mit einer niedrigeren IP-MTU vorhanden ist, während sich die TCP-PMTUD in der Standardeinstellung befindet. Weitere Informationen finden Sie in diesem Bild.

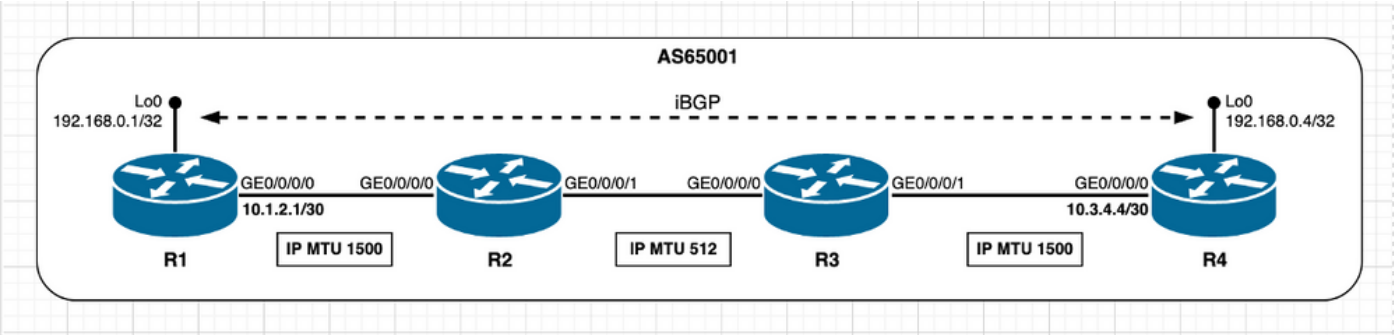

Image 2.8 - R2/R3-Pfadsegment hat eine niedrigere IP-MTU.

Betrachten wir zunächst, dass die BGP-Informationen minimal sind, d. h., dass alles, was zwischen BGP-Peers ausgetauscht werden muss, mit IP-Paketen erreicht werden kann, die unter die MTU des minimalen Pfads von 512 Byte passen. Bei dieser Annahme erfolgt die MSS-Berechnung wie im Abschnitt **Nicht direkt verbundene TCP-Peers** beschrieben. Sowohl R1 als auch R4 wählen einen MSS-Wert von 1240 Byte aus.

Einzelheiten zur TCP-Sitzung finden Sie unter R4 - ACTIVE (AKTIV):

```
! – 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>
```

Einzelheiten zur TCP-Sitzung finden Sie unter R1 - PASSIVE:

```
! – 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>
```

Wenn die BGP-Sitzung jetzt eingerichtet ist, stellen Sie fest, dass eine BGP-Update-Nachricht mit einer Größe über der MTU für den minimalen Pfad von 512 Byte ausgelöst wird. Wie aus den Ausgängen ersichtlich ist, legt Cisco IOS XR das df-Bit nicht mit der BGP-Aktualisierungsnachricht fest, d. h., dass BGP-Informationen auf Kosten der Paketfragmentierung auf Zwischenknoten übertragen werden.

BGP-Update von R1 - PASSIVE:

```
! – 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
```

Die Fragmentierung der BGP Update-Nachricht, die von Knoten R1 bezogen wird, findet an Knoten R2 statt, wie durch die Erfassung des Datenverkehrs an der R2-Schnittstelle GE0/0/0/1 festgestellt werden kann.

IP-Fragmentierung am Knoten R2:

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

4      1.334852       192.168.0.1 192.168.0.4 BGP    522    UPDATE Message
5      0.000289       192.168.0.1 192.168.0.4 IPv4   522    Fragmented IP protocol (proto=TCP 6,
off=488, ID=7b41)
6      0.000122       192.168.0.1 192.168.0.4 IPv4   135    Fragmented IP protocol (proto=TCP 6,
off=976, ID=7b41)

! – Captured frame details

Frame 4: 522 bytes on wire (4176 bits), 522 bytes captured (4176 bits) on interface 0
Ethernet II, Src: fa:16:3e:61:25:f0 (fa:16:3e:61:25:f0), Dst: fa:16:3e:23:ab:27
(fa:16:3e:23:ab:27)
Internet Protocol Version 4, Src: 192.168.0.1, Dst: 192.168.0.4
   0100 .... = Version: 4
   .... 0101 = Header Length: 20 bytes (5)
   Differentiated Services Field: 0xc0 (DSCP: CS6, ECN: Not-ECT)
   Total Length: 508
   Identification: 0x7b41 (31553)
   Flags: 0x01 (More Fragments)
       0... .... = Reserved bit: Not set
       .0.. .... = Don't fragment: Not set
       ..1. .... = More fragments: Set
   Fragment offset: 0
   Time to live: 254
   Protocol: TCP (6)
   Header checksum: 0x14f1 [validation disabled]
   [Header checksum status: Unverified]
   Source: 192.168.0.1
   Destination: 192.168.0.4
   [Source GeoIP: Unknown]
   [Destination GeoIP: Unknown]
Transmission Control Protocol, Src Port: 179, Dst Port: 39046, Seq: 4276759681, Ack: 1991250830
Border Gateway Protocol - UPDATE Message
<snip>

Frame 5: 522 bytes on wire (4176 bits), 522 bytes captured (4176 bits) on interface 0
Ethernet II, Src: fa:16:3e:61:25:f0 (fa:16:3e:61:25:f0), Dst: fa:16:3e:23:ab:27
(fa:16:3e:23:ab:27)
Internet Protocol Version 4, Src: 192.168.0.1, Dst: 192.168.0.4
   0100 .... = Version: 4
   .... 0101 = Header Length: 20 bytes (5)
   Differentiated Services Field: 0xc0 (DSCP: CS6, ECN: Not-ECT)
   Total Length: 508
   Identification: 0x7b41 (31553)
   Flags: 0x01 (More Fragments)
       0... .... = Reserved bit: Not set
       .0.. .... = Don't fragment: Not set
       ..1. .... = More fragments: Set
   Fragment offset: 488
```

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

Frame 6: 135 bytes on wire (1080 bits), 135 bytes captured (1080 bits) on interface 0
Ethernet II, Src: fa:16:3e:61:25:f0 (fa:16:3e:61:25:f0), Dst: fa:16:3e:23:ab:27
(fa:16:3e:23:ab:27)
Internet Protocol Version 4, Src: 192.168.0.1, Dst: 192.168.0.4
   0100 .... = Version: 4
   .... 0101 = Header Length: 20 bytes (5)
   Differentiated Services Field: 0xc0 (DSCP: CS6, ECN: Not-ECT)
   Total Length: 121
   Identification: 0x7b41 (31553)
   Flags: 0x00
       0... .... = Reserved bit: Not set
       .0.. .... = Don't fragment: Not set
       ..0. .... = More fragments: Not set
   Fragment offset: 976
   Time to live: 254
   Protocol: TCP (6)
   Header checksum: 0x35fa [validation disabled]
   [Header checksum status: Unverified]
   Source: 192.168.0.1
   Destination: 192.168.0.4
   [Source GeoIP: Unknown]
   [Destination GeoIP: Unknown]
Data (101 bytes)
<snip>
```

Szenarien - TCP-PMTUD aktiviert

PMTUD aktivieren

Sobald die PMTUD aktiviert ist, wird bei der MSS-Erstberechnung immer die IP-MTU der Ausgangsschnittstelle berücksichtigt, unabhängig davon, ob Peers direkt oder nicht direkt verbunden sind.

Dieses Szenario bietet Einblicke in das erwartete Verhalten bei Aktivierung der PMTUD. Hier übernimmt der Cisco IOS XR-Knoten R4 die aktive Rolle, verwaltet die TCP-Verbindung und öffnet die TCP-Sitzung mit dem Cisco IOS XR-Knoten R1 am Zielport 179. Beide Knoten verwenden die standardmäßigen IP-MTU-Werte für ihre Schnittstellen.

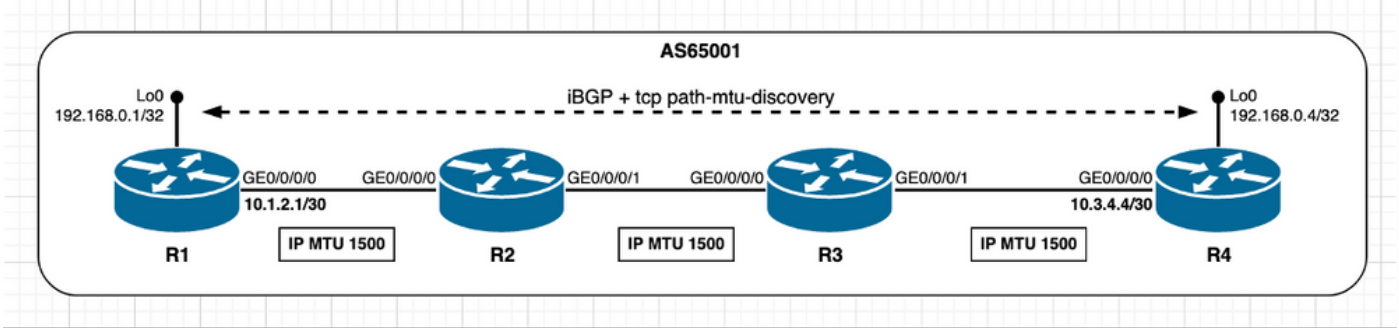

Image 3.1 - TCP-PMTUD aktiviert.

Die MSS-Berechnung in diesem Szenario kann wie folgt zusammengefasst werden:

- Alle Knoten verwenden eine Standard-IP-MTU von 1.500 Byte.
- Die TCP-Path-MTU-Erkennung ist aktiviert.
- TCP-Peers sind nicht direkt verbunden R4 verwaltet die BGP-VerbindungR4 sendet SYN mit MSS von 1460 Byte 1500 (Interface IP MTU) - 20 (minTCP_H) - 20 (minIP_H)R1 sendet SYN, ACK mit MSS von 1460 Byte Sendet die untere von [empfangenes MSS; Lokale anfängliche MSS]Empfangene MSS 1460 Byte; Lokale anfängliche MSS 1460 ByteDer niedrigste MSS-Wert wird auf beiden Peers verwendet.

Um die Verhaltensänderungen zu unterstreichen, die durch enable PMTUD eingeführt wurden, veranschaulichen die nächsten Ausgaben die Abfolge von Ereignissen:

1. Der Anfangsstatus der eingerichteten TCP-Sitzung im Standardszenario der PMTUD ist deaktiviert.
2. Die PMTUD ist konfiguriert und auf den TCP-Peers R4 und R1 aktiviert.
3. Die TCP-Sitzung wird neu gestartet, die MSS-Berechnung erfolgt und wird durch die TCP-PMTUD beeinflusst.

Wie bei R4 - ACTIVE - TCP PMTUD deaktiviert (Standard) gezeigt:

```
! – 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#
```

Wie unter R1 - PASSIVE - TCP PMTUD disabled (Standardeinstellung) zu sehen:

```
! – 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#
```

Wie bei R4 - ACTIVE - TCP PMTUD aktiviert:

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

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

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

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

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

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

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

```
SendWnd             0          0                0
TimeWait            0          0                0
AckHold             7          4                0
KeepAlive           1          0                0
PmtuAger            1          0           508096
GiveUp              0          0                0
Throttle            0          0                0

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

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

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

State flags: PMTU ager
Feature flags: Win Scale, Nagle, Path MTU
Request flags: Win Scale

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

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

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

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

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

Wie bei R1 - PASSIVE - TCP PMTUD aktiviert:

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

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

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

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

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

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

Timer          Starts    Wakeups         Next(msec)
Retrans             7          0                0
SendWnd             0          0                0
TimeWait            0          0                0
AckHold             7          4                0
KeepAlive           1          0                0
PmtuAger            1          0           562042
GiveUp              0          0                0
Throttle            0          0                0

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

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

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

State flags: PMTU ager
Feature flags: Win Scale, Nagle, Path MTU
Request flags: Win Scale

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

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

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

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

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

Beachten Sie das Verhalten des PMTU-Zeitgebers:

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

RP/0/0/CPU0:Jan  8 16:09:25.214 : tcp[399]: [t21] Try to enable path MTU discovery(neww age
```

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

Wie unter R4 - ACTIVE - BGP Session Restart - TCP SYN:

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

2734   4.810311       192.168.0.4 192.168.0.1 TCP    62     32077  179 [SYN] Seq=0 Win=16384
Len=0 MSS=1460 WS=1

Frame 2734: 62 bytes on wire (496 bits), 62 bytes captured (496 bits) on interface 0
Ethernet II, Src: fa:16:3e:d7:7e:f6 (fa:16:3e:d7:7e:f6), Dst: fa:16:3e:8f:8f:54
(fa:16:3e:8f:8f:54)
Internet Protocol Version 4, Src: 192.168.0.4, Dst: 192.168.0.1
Transmission Control Protocol, Src Port: 32077, Dst Port: 179, Seq: 0, Len: 0
   Source Port: 32077
   Destination Port: 179
   [Stream index: 25]
   [TCP Segment Len: 0]
   Sequence number: 0    (relative sequence number)
   Acknowledgment number: 0
   Header Length: 28 bytes
   Flags: 0x002 (SYN)
   Window size value: 16384
   [Calculated window size: 16384]
   Checksum: 0x6398 [unverified]
   [Checksum Status: Unverified]
   Urgent pointer: 0
   Options: (8 bytes), Maximum segment size, Window scale, End of Option List (EOL)
       Maximum segment size: 1460 bytes
           Kind: Maximum Segment Size (2)
           Length: 4
           MSS Value: 1460
       Window scale: 0 (multiply by 1)
       End of Option List (EOL)
```

Wie unter R1 - PASSIVE - BGP Session Restart - TCP SYN, ACK.

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

2735   0.003879       192.168.0.1 192.168.0.4 TCP    62     179  32077 [SYN, ACK] Seq=0 Ack=1
Win=16384 Len=0 MSS=1460 WS=1

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

```
   Flags: 0x012 (SYN, ACK)
   Window size value: 16384
   [Calculated window size: 16384]
   Checksum: 0xbf77 [unverified]
   [Checksum Status: Unverified]
   Urgent pointer: 0
   Options: (8 bytes), Maximum segment size, Window scale, End of Option List (EOL)
       Maximum segment size: 1460 bytes
           Kind: Maximum Segment Size (2)
           Length: 4
           MSS Value: 1460
       Window scale: 0 (multiply by 1)
       End of Option List (EOL)
```

TCP-Sitzungsdetails wie unter R4 - ACTIVE (aktiv) angezeigt - nachdem die TCP-PMTUD aktiviert und die BGP-Sitzung gelöscht wurde:

```
! – 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#
```

TCP-Sitzungsdetails wie unter R1 - PASSIVE - nach Aktivierung der TCP-PMTUD und Löschen der BGP-Sitzung.

```
! – 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 - Pfadsegment hat niedrigere IP-MTU

Das vorherige Szenario half dabei zu verstehen, was beim erstmaligen Einrichten einer TCP-Sitzung mit aktivierter PMTUD geschieht. Dieses Szenario baut auf und hilft dabei zu verstehen, wie TCP PMTUD funktioniert und welchen Einfluss es auf etablierte TCP-Sitzungen hat.

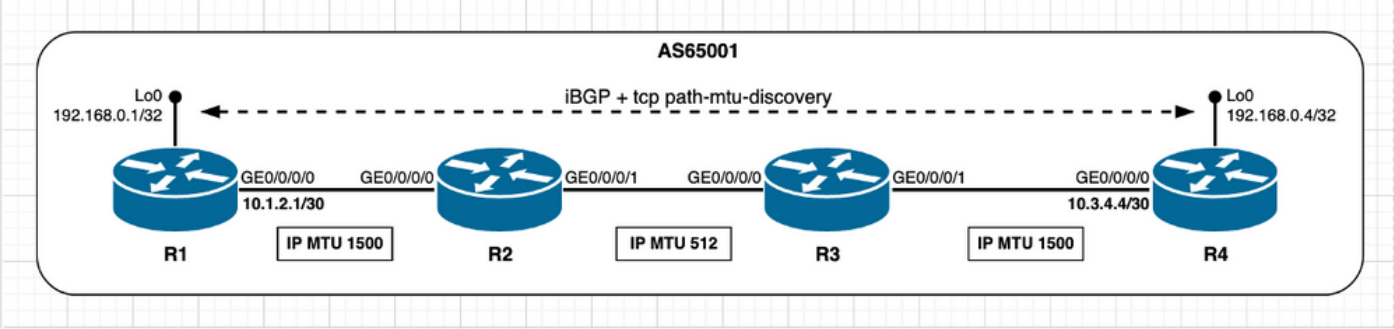

Image 3.2 - PMTUD aktiviert, und das Pfadsegment hat eine niedrigere IP-MTU.

Betrachten Sie das vorherige Bild als Referenz, nehmen Sie an, dass die BGP-Sitzung eingerichtet ist, und R1 sendet die BGP Update-Nachricht, die von einem IP-Paket mit einer Größe von mehr als 512 Byte übertragen wird. Bei aktivierter PMTUD ist jetzt das DF-Bit (Don't Fragment) festgelegt. Daher verwirft Knoten R2 das IP-Paket und sendet eine ICMP-Meldung (Internet Control Message Protocol) (Ziel nicht erreichbar - Typ 3). Fragmentierung erforderlich - Code 4) zurück zu R1. Bei Knoten R1 nach dem Empfang der ICMP-Meldung wird die PMTUD ausgelöst und versucht, die IP-MTU für den niedrigsten Pfad festzulegen. Dabei wird der nächste niedrigere Wert aus einer Reihe klar definierter Plateau-Ebenen verwendet, d. h. ein neuer TCP-Session-MSS-Wert. TCP überträgt dann das ursprüngliche BGP-Update erneut mit dem neuen MSS-Wert. Dieser Prozess wird so oft wiederholt, wie es erforderlich ist, bis die ICMP-Nachricht (Destination Unreachable - Typ 3) gesendet wird. Fragmentierung erforderlich - Code 4) wird nicht mehr empfangen. Dies bedeutet, dass der verwendete MSS-Wert so hoch ist, dass jedes gesendete Paket unter die IP-MTU des niedrigsten Pfadoptsegments fällt. Mit der Zeit durchläuft die PMTUD, die vom PmtuAger-Timer gesteuert wird, die Plateauebenen in umgekehrter Richtung und erhöht die MSS wieder auf ihren maximalen Wert. Wenn eine ICMP-Nachricht (Ziel nicht erreichbar - Geben Sie zu jedem Zeitpunkt 3 ein; Fragmentierung erforderlich - Code 4) wird erneut empfangen, dann handelt die PMTUD wie oben beschrieben.

Die nächsten Ausgaben führen das eben beschriebene PMTUD-Verhalten durch und beginnen mit dem Szenario einer etablierten TCP-Sitzung. In diesem Fall übernimmt der Cisco IOS XR-Knoten R4 eine aktive Rolle und verwaltet so die TCP-Verbindung und öffnet die TCP-Sitzung mit R1 am Zielport 179. Beide Knoten verwenden die standardmäßigen IP-MTU-Werte für ihre Schnittstellen. Die anfängliche MSS-Berechnung in diesem Szenario kann wie folgt zusammengefasst werden:

- Das Zwischensegment zwischen R2- und R3-Knoten verwendet eine nicht standardmäßige IP-MTU von 512 Byte.
- R1 und R4 verwenden auf ihren Schnittstellen die MTU-Standardwerte.
- Die TCP-Pfad-MTU-Erkennung ist aktiviert.
- TCP-Peers sind nicht direkt verbunden. R4 verwaltet die BGP-Verbindung.R4 sendet SYN mit einer MSS von 1460 Byte. 1500 (Interface IP MTU) - 20 (minTCP_H) - 20 (minIP_H).R1 sendet SYN, ACK mit MSS von 1460 Byte. Sendet die untere von [Received MSS ; Lokale anfängliche MSS].Empfangene MSS 1460 Byte; Lokale anfängliche MSS 1460 Byte.Der niedrigste MSS-Wert wird für beide Peers verwendet.

TCP SYN ausgehend von R4:

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

392    6.752774       192.168.0.4 192.168.0.1 TCP    62     32449  179 [SYN] Seq=0 Win=16384
Len=0 MSS=1460 WS=1

Frame 392: 62 bytes on wire (496 bits), 62 bytes captured (496 bits) on interface 0
Ethernet II, Src: fa:16:3e:5c:f1:80 (fa:16:3e:5c:f1:80), Dst: fa:16:3e:42:18:05
(fa:16:3e:42:18:05)
Internet Protocol Version 4, Src: 192.168.0.4, Dst: 192.168.0.1
Transmission Control Protocol, Src Port: 32449, Dst Port: 179, Seq: 0, Len: 0
   Source Port: 32449
   Destination Port: 179
   [Stream index: 10]
   [TCP Segment Len: 0]
   Sequence number: 0    (relative sequence number)
   Acknowledgment number: 0
   Header Length: 28 bytes
   Flags: 0x002 (SYN)
   Window size value: 16384
   [Calculated window size: 16384]
   Checksum: 0x6858 [unverified]
   [Checksum Status: Unverified]
   Urgent pointer: 0
   Options: (8 bytes), Maximum segment size, Window scale, End of Option List (EOL)
       Maximum segment size: 1460 bytes
           Kind: Maximum Segment Size (2)
           Length: 4
           MSS Value: 1460
       Window scale: 0 (multiply by 1)
       End of Option List (EOL)
```

TCP-SYN, ACK von R1 stammt:

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

393    0.003628      192.168.0.1 192.168.0.4 TCP    62     179  32449 [SYN, ACK] Seq=0 Ack=1
Win=16384 Len=0 MSS=1460 WS=1
```

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

Wenn die BGP-Sitzung eingerichtet ist, sendet Knoten R1 die BGP-Update-Nachricht und empfängt die ICMP-Nachricht (Ziel nicht erreichbar - Typ 3; Fragmentierung erforderlich - Code 4) im Gegenzug von Knoten R2.

Dies liegt daran, dass das IP-Paket, das die BGP-Update-Nachricht enthält, über einen DF-Bit-Satz verfügt und die IP-MTU von 512 Byte, die im R2/R3-Segment verwendet wird, unter der IP-Paketgröße von 116 Byte liegt. Wie bereits erläutert, löst der Empfang der ICMP-Nachricht die PMTUD aus.

Bei R1 ICMP wird die Meldung Typ 3/Code 4 empfangen:

```
! – 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]
```

Bei Knoten R1, ausgelöst durch ICMP-Nachricht, versucht die TCP-PMTUD, die End-to-End-niedrigste IP-MTU durch Verwendung des nächsten niedrigeren Werts aus einer Reihe klar definierter Plateau-Ebenen (IP MTU) zu ermitteln. Diese Plateau-Ebenen sind in der RFC1191 - MTU-Pfaderkennung dokumentiert.

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

Seit ICMP (Destination Unreachable - Typ 3; Fragmentierung erforderlich - Code 4)-Nachricht, die von Knoten R1 empfangen wird, zeigt die **MTU des nächsten Hop an**, dann, wie dargestellt, Knoten R1 verwendet diesen Wert, der in unserem Beispiel 512 Byte beträgt, und passen Sie den MSS-Wert der TCP-Sitzung an. Beachten Sie, dass die ursprüngliche Länge des TCP-Segments 1076 Byte betrug. Daher sind drei Pakete erforderlich, um das ursprüngliche TCP-Segment neu zu übertragen.

Wie bei R1 - PASSIVE - PMTUD-Betrieb gezeigt:

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

530    0.007497       192.168.0.1 192.168.0.4 TCP    526    [TCP Out-Of-Order] 179  32449 [ACK]
Seq=318 Ack=251 Win=32593 Len=472
532    0.015374       192.168.0.1 192.168.0.4 TCP    526    [TCP Retransmission] 179  32449
[ACK] Seq=790 Ack=251 Win=32593 Len=472
533    0.004129       192.168.0.1 192.168.0.4 TCP    186    [TCP Retransmission] 179  32449
[PSH, ACK] Seq=1262 Ack=251 Win=32593 Len=132
```

Wie bereits erwähnt, durchläuft die PMTUD nach der Übertragung aller Pakete im Laufe der Zeit die Plateauebenen in der vom PmtuAger-Timer vorgegebenen umgekehrten Richtung und versucht, die MSS auf den maximalen Wert gemäß dem vorhandenen Szenario zu erhöhen.

Siehe R1 - PMTUD für definierte Plattformen:

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

RP/0/0/CPU0:May 12 09:09:22.763 UTC: ipv4_io[266]: IPv4 ICMP: Received ICMP too big from
192.168.0.1 about 192.168.0.4, MTU=512
RP/0/0/CPU0:May 12 09:09:22.763 UTC: ipv4_io[266]: ipv4_icmp_unreachable_rcvd ICMP unreach
recvd: sending pak(0xb0c07d8f) to transport: 6, tid: 5
RP/0/0/CPU0:May 12 09:09:22.763 UTC: ipv4_io[266]: ip_icmp_lib_ipv4_receive: sending
pak(0xb0c07d8f) to transport: 1, tid: 5
RP/0/0/CPU0:May 12 09:09:22.763 UTC: tcp[399]: [t4] PCB 0x15393770: Process ICMP Dest-unreach
(next hop mtu: 512)
```

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

RP/0/0/CPU0:May 12 09:09:22.763 UTC: tcp[399]: [t4] PCB 0x15393770: Process ICMP Dest-unreach
(next hop mtu: 512)
RP/0/0/CPU0:May 12 09:09:22.763 UTC: tcp[399]: [t4] PCB 0x15393770: Try to use new MSS: 472
RP/0/0/CPU0:May 12 09:09:22.763 UTC: tcp[399]: [t4] PCB 0x15393770, New path MTU decided to use:
472  configured tp_user_mss 0

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

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

Der Endzustand kann bei diesen Ausgaben beobachtet werden. Beachten Sie insbesondere die vom Knoten R1 angezeigten MSS-Werte min und max., was die Auslösung der PMTUD unterstreicht und anzeigt.

Einzelheiten zur TCP-Sitzung finden Sie unter R4 - ACTIVE (AKTIV):

```
! – 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#
```

Einzelheiten zur TCP-Sitzung finden Sie unter R1 - PASSIVE:

```
! – 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#
```

Letztlich, wenn zu einem bestimmten Zeitpunkt ein ICMP (Destination Unreachable - Typ 3); Fragmentierung erforderlich - Code 4) Nachricht wird erneut empfangen, und die PMTUD verhält sich wieder wie oben beschrieben.

Wie aus R1 - PASSIVE - Die PMTUD hat erneut ausgelöst:

```
! – 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#
```

Bei Cisco IOS XR-Versionen, die von der Cisco Bug-ID CSCvf10395 betroffen sind, wird der Next-Hop in der ICMP-Fehlermeldung ignoriert und der Knoten versucht, die End-to-End-niedrigste IP-MTU mithilfe des nächsten niedrigeren Werts aus den zuvor erwähnten und von der RFC1 dokumentierten gut definierten Plateau-Ebenen zu ermitteln. 191 - MTU-Pfaderkennung. Diese Versuche erfolgen bis zur erfolgreichen Übertragung, das heißt bis ICMP (Destination Unreachable - Typ 3; Fragmentierung erforderlich - Code 4) Nachrichten werden nicht mehr empfangen.

Ein Knoten mit Cisco IOS XR-Version, der von der Cisco Bug-ID CSCvf10395 betroffen ist, hat Folgendes gesehen:

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

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

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

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

Als nächster Schritt sollten Sie dasselbe Szenario betrachten, jedoch mit Label Distribution Protocol (LDP) über alle Schnittstellen hinweg. Ziel hierbei ist es, die Unterschiede zu verstehen, die in einer MPLS-fähigen Umgebung im Vergleich zu vorherigen Szenarien beobachtet werden können.

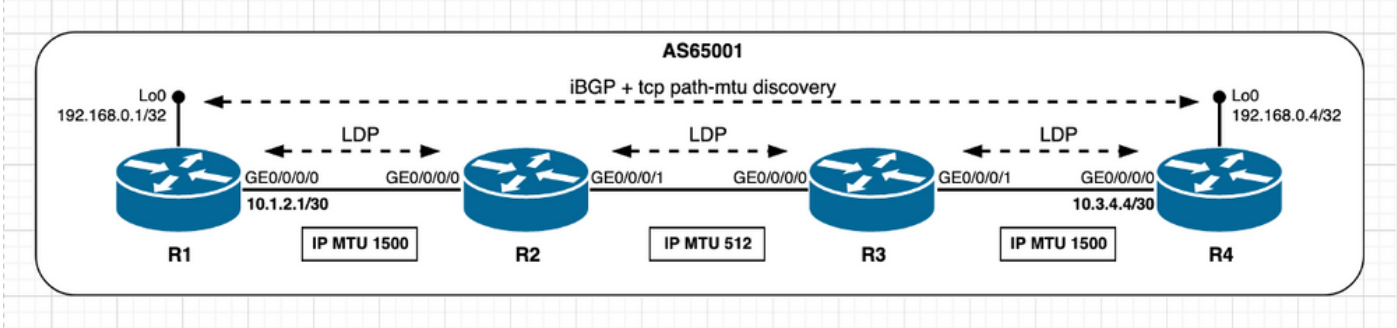

Image 3.3 - PMTUD aktiviert, und das Pfadsegment weist eine niedrigere IP-MTU - MPLS-Szenario auf.

Betrachten Sie zunächst die Anfangsphase der BGP-Sitzung, die vor dem PMTUD-Trigger eingerichtet wurde, wie hier gezeigt.

Der Anfangsstatus von TCP (BGP) wird in einem Szenario mit aktiviertem MPLS-Protokoll von R4 (AKTIV) angezeigt:

```
! – 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#
```

Der Anfangsstatus von TCP (BGP) wird für R1 - PASSIVE - MPLS-aktiviertes Szenario angezeigt:

```
! – 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#
```

In diesem MPLS-aktivierten Szenario wurden die Details für die TCP-Sitzungen (LDP) festgelegt. Bitte beachten Sie, dass alle zuvor beschriebenen Schritte bezüglich der MSS-Berechnung für TCP-Sitzungen (BGP) auch für TCP-Sitzungen (LDP) gelten. Die MSS-Berechnung für die Sitzungen der Knoten R3 und R2 TCP (LDP) kann beispielsweise wie folgt zusammengefasst werden:

- Sowohl R2 als auch R3 verwenden eine nicht standardmäßige IP-MTU von 512 Byte.
- Die MTU-Pfaderkennung ist aktiviert.
- TCP-Peers sind nicht direkt verbunden (TCP-Sitzung wird zwischen Loopback-Schnittstellen eingerichtet). R3 verwaltet die LDP-Verbindung.R3 sendet SYN mit einer MSS von 472 Byte. 512 (Interface IP MTU) - 20 (minTCP_H) - 20 (minIP_H).R2 sendet SYN, ACK mit einer MSS von 472 Byte. Sendet die untere von [empfangenes MSS; Lokale anfängliche MSS].Empfangene MSS 472 Byte Lokale anfängliche MSS 472 Byte.Der niedrigste MSS-Wert wird für beide Peers verwendet.

Einzelheiten zu TCP-Sitzungen (LDP), wie im Szenario "R3 - ACTIVE - MPLS" dargestellt:

```
! – 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#
```

Einzelheiten zu TCP-Sitzungen (LDP), wie unter R2 - PASSIVE - MPLS-aktiviertes Szenario dargestellt:

```
! – 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#
```

Nachdem die BGP-Sitzung eingerichtet wurde, sendet R1 die BGP-Update-Nachricht und empfängt die ICMP-Meldung (Destination Unreachable - Typ 3). Fragmentierung erforderlich - Code 4) im Gegenzug von Knoten R2, der TCP-PMTUD an Knoten R1 auslöst. Dies liegt daran, dass das IP-Paket, das die BGP-Update-Nachricht enthält, über einen DF-Bit-Satz verfügt und die IP-MTU von 512 Byte, die im R2/R3-Segment verwendet wird, unter der IP-Paketgröße von 116 Byte liegt. Wie zuvor wird die PMTUD durch den Empfang dieser ICMP-Nachricht ausgelöst. Der Unterschied im MPLS-aktivierten Szenario im Vergleich zu den vorherigen Nicht-MPLS-Szenarien besteht im Hinblick auf die **MTU des nächsten Hop-**Werts in der ICMP-Meldung des Knotens R2 (Ziel nicht erreichbar - Typ 3). Fragmentierung erforderlich - Code 4). In diesem MPLS-aktivierten Szenario berücksichtigt die **MTU des nächsten Hop-**Werts den zusätzlichen MPLS-Overhead von 4 Byte, d. h., er berücksichtigt den Ausgangs-MPLS-Label-Stack bei R2, wie in diesen Ausgaben zu sehen ist.

TCP-Pfad-MTU-Erkennung in Aktion, wie im Szenario R1 - PASSIVE - MPLS-aktiviertes Szenario gezeigt:

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

455    0.044859       192.168.0.1 192.168.0.4 BGP    1134   UPDATE Message, KEEPALIVE Message

Frame 455: 1134 bytes on wire (9072 bits), 1134 bytes captured (9072 bits) on interface 0
Ethernet II, Src: fa:16:3e:42:18:05 (fa:16:3e:42:18:05), Dst: fa:16:3e:5c:f1:80
(fa:16:3e:5c:f1:80)
MultiProtocol Label Switching Header, Label: 24002, Exp: 6, S: 1, TTL: 255
Internet Protocol Version 4, Src: 192.168.0.1, Dst: 192.168.0.4
   0100 .... = Version: 4
   .... 0101 = Header Length: 20 bytes (5)
   Differentiated Services Field: 0xc0 (DSCP: CS6, ECN: Not-ECT)
   Total Length: 1116
   Identification: 0xc6dd (50909)
   Flags: 0x02 (Don't Fragment)
       0... .... = Reserved bit: Not set
       .1.. .... = Don't fragment: Set
       ..0. .... = More fragments: Not set
   Fragment offset: 0
   Time to live: 255
   Protocol: TCP (6)
   Header checksum: 0xa5f4 [validation disabled]
   [Header checksum status: Unverified]
   Source: 192.168.0.1
   Destination: 192.168.0.4
```

```
   [Source GeoIP: Unknown]
   [Destination GeoIP: Unknown]
Transmission Control Protocol, Src Port: 179, Dst Port: 57400, Seq: 242, Ack: 175, Len: 1076
Border Gateway Protocol - UPDATE Message
Border Gateway Protocol - KEEPALIVE Message
<snip>

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

456    0.014117       10.2.3.1       192.168.0.1 ICMP   182    Destination unreachable
(Fragmentation needed)

Frame 456: 182 bytes on wire (1456 bits), 182 bytes captured (1456 bits) on interface 0
Ethernet II, Src: fa:16:3e:5c:f1:80 (fa:16:3e:5c:f1:80), Dst: fa:16:3e:42:18:05
(fa:16:3e:42:18:05)
Internet Protocol Version 4, Src: 10.2.3.1, Dst: 192.168.0.1
   0100 .... = Version: 4
   .... 0101 = Header Length: 20 bytes (5)
   Differentiated Services Field: 0x00 (DSCP: CS0, ECN: Not-ECT)
   Total Length: 168
   Identification: 0x001f (31)
   Flags: 0x00
       0... .... = Reserved bit: Not set
       .0.. .... = Don't fragment: Not set
       ..0. .... = More fragments: Not se
   Fragment offset: 0
   Time to live: 251
   Protocol: ICMP (1)
   Header checksum: 0xb031 [validation disabled]
   [Header checksum status: Unverified]
   Source: 10.2.3.1
   Destination: 192.168.0.1
   [Source GeoIP: Unknown]
   [Destination GeoIP: Unknown]
Internet Control Message Protocol
    Type: 3 (Destination unreachable)
    Code: 4 (Fragmentation needed)
   Checksum: 0x5199 [correct]
   [Checksum Status: Good]
   Length: 17
   [Length of original datagram: 68]
   Unused: 0011
   MTU of next hop: 508
   Internet Protocol Version 4, Src: 192.168.0.1, Dst: 192.168.0.4
   Transmission Control Protocol, Src Port: 179, Dst Port: 57400, Seq: 3874414921, Ack:
1386460094
   Border Gateway Protocol - UPDATE Message

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

457    0.006689       192.168.0.1 192.168.0.4 TCP    526    [TCP Retransmission] 179  57400
[ACK] Seq=242 Ack=175 Win=32669 Len=468
460    0.004001       192.168.0.1 192.168.0.4 TCP    526    [TCP Retransmission] 179  57400
```

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

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

RP/0/0/CPU0:May 17 08:29:56.131 UTC: tcp[399]: [t1] Try to enable path MTU discovery(neww age
timer: 10 min)
RP/0/0/CPU0:May 17 08:29:56.131 UTC: tcp[399]: [t1] Path mtu is ON (age-timer: 10)
RP/0/0/CPU0:May 17 08:35:51.726 UTC: ipv4_io[266]: ip_icmp_lib_ipv4_receive: Receiving
pak(0xb0c07d8f) tid: 5
RP/0/0/CPU0:May 17 08:35:51.726 UTC: ipv4_io[266]: Entering ipv4_mtu_update_cb
RP/0/0/CPU0:May 17 08:35:51.726 UTC: ipv4_io[266]: IPv4 ICMP: Received ICMP too big from
192.168.0.1 about 192.168.0.4, MTU=508
RP/0/0/CPU0:May 17 08:35:51.726 UTC: ipv4_io[266]: ipv4_icmp_unreachable_rcvd ICMP unreach
recvd: sending pak(0xb0c07d8f) to transport: 6, tid: 5
RP/0/0/CPU0:May 17 08:35:51.726 UTC: ipv4_io[266]: ip_icmp_lib_ipv4_receive: sending
pak(0xb0c07d8f) to transport: 1, tid: 5
RP/0/0/CPU0:May 17 08:35:51.726 UTC: tcp[399]: [t4] PCB 0x153acc8c: Process ICMP Dest-unreach
(next hop mtu: 508)

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

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

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

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

Wie aus dem Szenario R1 - PASSIVE - TCP PMTUD Triggered - MPLS-enabled (Ausgelöstes Szenario für TCP-PMTUD) ersichtlich:

```
! – 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#
```

Beachten Sie, dass im MPLS-aktivierten Szenario der Wert der **MTU des nächsten Hop, der** in der ICMP-Nachricht des Knotens R2 enthalten ist, für den MPLS-Labelstack für den Ausgang berücksichtigt wird. Um diesen Aspekt weiter zu stärken, sehen Sie sich das nächste Beispiel an. Wenn das bei R2 gefilterte IP-Paket einem L3VPN-Service zugeordnet ist, bedeutet dies, dass der Ethernet-Frame jetzt zwei Labels enthält (IGP-Label und VPN-Label). Anschließend spiegelt die **MTU des nächsten Hop** die erforderliche Label-Stack-Größe wider. Siehe diese Ausgaben.

Wie bei R1 - PASSIVE - L3 VPN-Servicepaket gezeigt:

```
! – 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
```

Wie unter R1 - PASSIVE - L3 VPN-Service - ICMP Typ 3/Code 4 zu sehen:

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

2030   0.020299       10.2.3.1       10.1.14.1      ICMP   190    Destination unreachable
(Fragmentation needed)

Frame 2030: 190 bytes on wire (1520 bits), 190 bytes captured (1520 bits) on interface 0
Ethernet II, Src: fa:16:3e:5c:f1:80 (fa:16:3e:5c:f1:80), Dst: fa:16:3e:42:18:05
(fa:16:3e:42:18:05)
MultiProtocol Label Switching Header, Label: 24005, Exp: 0, S: 1, TTL: 251
   0000 0101 1101 1100 0101 .... .... .... = MPLS Label: 24005
   .... .... .... .... .... 000. .... .... = MPLS Experimental Bits: 0
   .... .... .... .... .... ...1 .... .... = MPLS Bottom Of Label Stack: 1
   .... .... .... .... .... .... 1111 1011 = MPLS TTL: 251
Internet Protocol Version 4, Src: 10.2.3.1, Dst: 10.1.14.1
   0100 .... = Version: 4
   .... 0101 = Header Length: 20 bytes (5)
   Differentiated Services Field: 0x00 (DSCP: CS0, ECN: Not-ECT)
```

```
   Total Length: 172
   Identification: 0x002b (43)
   Flags: 0x00
       0... .... = Reserved bit: Not set
       .0.. .... = Don't fragment: Not set
       ..0. .... = More fragments: Not set
   Fragment offset: 0
   Time to live: 253
   Protocol: ICMP (1)
   Header checksum: 0x9821 [validation disabled]
   [Header checksum status: Unverified]
   Source: 10.2.3.1
   Destination: 10.1.14.1
   [Source GeoIP: Unknown]
   [Destination GeoIP: Unknown]
Internet Control Message Protocol
    Type: 3 (Destination unreachable)
    Code: 4 (Fragmentation needed)
   Checksum: 0xbbac [correct]
   [Checksum Status: Good]
   Length: 17
   [Length of original datagram: 68]
   Unused: 0011
   MTU of next hop: 504
   Internet Protocol Version 4, Src: 10.1.14.1, Dst: 10.1.14.14
       0100 .... = Version: 4
       .... 0101 = Header Length: 20 bytes (5)
       Differentiated Services Field: 0x00 (DSCP: CS0, ECN: Not-ECT)
       Total Length: 610
       Identification: 0x7c9f (31903)
       Flags: 0x02 (Don't Fragment)
           0... .... = Reserved bit: Not set
           .1.. .... = Don't fragment: Set
           ..0. .... = More fragments: Not set
       Fragment offset: 0
       Time to live: 255
       Protocol: TCP (6)
       Header checksum: 0xcce5 [validation disabled]
       [Header checksum status: Unverified]
       Source: 10.1.14.1
       Destination: 10.1.14.14
       [Source GeoIP: Unknown]
       [Destination GeoIP: Unknown]
   Transmission Control Protocol, Src Port: 22008, Dst Port: 23, Seq: 586828435, Ack: 754580617
```

PMTUD - TCP-Optionen (MD5)

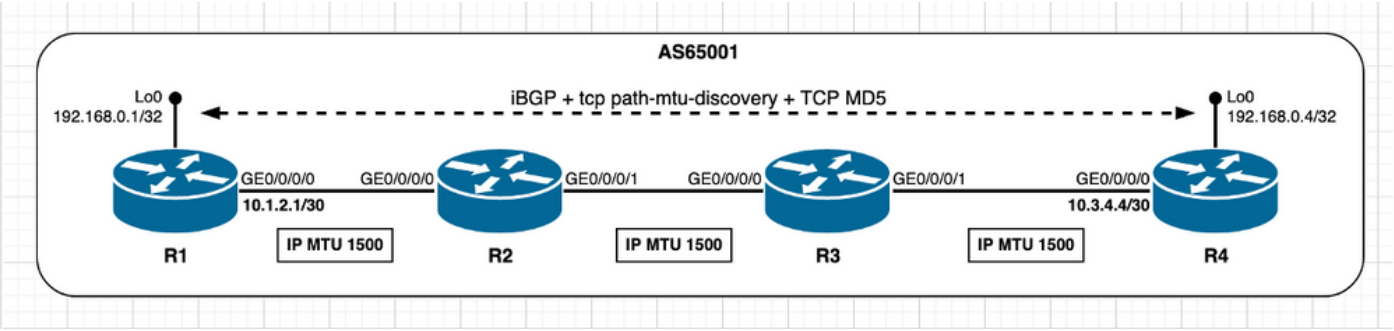

Image 3.4 - PMTUD-fähige und TCP-MD5-Authentifizierung.

Bei aktivierter TCP-MD5-Authentifizierung gibt es keine Unterscheidung hinsichtlich des PMTUD-Verhaltens gegenüber dem, was bereits in den vorherigen Szenarien beschrieben wurde. Wie

zuvor bereits bei der verwendeten TCP-MD5-Authentifizierung verwendet, berücksichtigt Cisco IOS XR zusätzliche Overhead und den ursprünglichen MSS-Wert des aktiven TCP-Peers. In den vorherigen Abschnitten **TCP-Optionen verwenden - XR Aktiv** und **TCP-Optionen verwenden - XR Passiv** finden Sie weitere Informationen zu den Auswirkungen der TCP-Optionen. Die TCP-MSS-Berechnung in diesem Szenario kann wie folgt zusammengefasst werden:

- Alle Knoten verwenden eine IP-StandardMTU von 1.500 Byte.
- Die TCP-Pfad-MTU-Erkennung ist aktiviert.
- TCP-Peers sind nicht direkt verbunden.
- Die TCP-MD5-Authentifizierung ist sowohl auf R1 als auch auf R4 aktiviert. R4 verwaltet die BGP-Verbindung.R4 sendet SYN mit einer MSS von 1436 Byte. 1500 (Interface IP MTU) - 20 (minTCP_H) - 20 (minIP_H) - 24 Byte (IOS XR TCP Options Overhead).R1 sendet SYN, ACK mit MSS von 1436 Byte. sendet die untere von [Empfangs-MSS; Lokale anfängliche MSS].Empfangene MSS 1436 Byte Lokale anfängliche MSS 1460 Byte.Der niedrigste MSS-Wert wird auf beiden Peers verwendet.

TCP SYN ausgehend von 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 von R1 stammt:

```
! – TCP SYN,ACK sourced from R1

2409   0.004352       192.168.0.1 192.168.0.4 TCP    82     179  59050 [SYN, ACK] Seq=0 Ack=1
Win=16384 Len=0 MSS=1436 WS=1
```

```
Frame 2409: 82 bytes on wire (656 bits), 82 bytes captured (656 bits) on interface 0
Ethernet II, Src: fa:16:3e:8f:8f:54 (fa:16:3e:8f:8f:54), Dst: fa:16:3e:d7:7e:f6
(fa:16:3e:d7:7e:f6)
Internet Protocol Version 4, Src: 192.168.0.1, Dst: 192.168.0.4
Transmission Control Protocol, Src Port: 179, Dst Port: 59050, Seq: 0, Ack: 1, Len: 0
   Source Port: 179
   Destination Port: 59050
   [Stream index: 8]
   [TCP Segment Len: 0]
   Sequence number: 0    (relative sequence number)
   Acknowledgment number: 1    (relative ack number)
   Header Length: 48 bytes
   Flags: 0x012 (SYN, ACK)
   Window size value: 16384
   [Calculated window size: 16384]
   Checksum: 0xcbf8 [unverified]
   [Checksum Status: Unverified]
   Urgent pointer: 0
   Options: (28 bytes), Maximum segment size, Window scale, No-Operation (NOP), TCP MD5
signature, End of Option List (EOL)
       Maximum segment size: 1436 bytes
           Kind: Maximum Segment Size (2)
           Length: 4
           MSS Value: 1436
       Window scale: 0 (multiply by 1)
       No-Operation (NOP)
       TCP MD5 signature
       End of Option List (EOL)
```

Einzelheiten zur TCP-Sitzung finden Sie unter R4 - ACTIVE (AKTIV):

```
! – 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#
```

Einzelheiten zur TCP-Sitzung finden Sie unter R1 - PASSIVE:

```
! – 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 - Blackhole-Erkennung

Wie bereits im Abschnitt **PMTUD - Pfadsegment hat niedrigere IP-MTU** erläutert, wird die TCP-PMTUD bei Aktivierung durch den Empfang eines ICMP (Destination Unreachable - Typ 3) ausgelöst. Fragmentierung erforderlich - Code 4)-Nachricht. Es kann sein, dass diese Meldungen aus irgendeinem Grund nicht empfangen werden, wodurch die PMTUD nicht ausgelöst wird. In diesem Fall wird die niedrigste IP-MTU des Pfads zwischen den TCP-Peers nicht ermittelt. Ein solches Szenario würde ein potenzielles Blackhole einführen, wenn IP-Pakete über ein festgelegtes DF-Bit verfügen und eine größere Größe als das Segment des niedrigsten IP-MTU-Pfads haben. Diese Pakete würden unbemerkt verworfen.

In diesem Abschnitt wird erläutert, wie Cisco IOS XR ein solches potenzielles Blackhole-Szenario erkennt und entsprechend reagiert. Zu diesem Zweck ist die Funktion für nicht erreichbare IPv4 an der R2-Schnittstelle GE0/0/0/0 deaktiviert, wie im nächsten Image und der CLI-Ausgabe dargestellt.

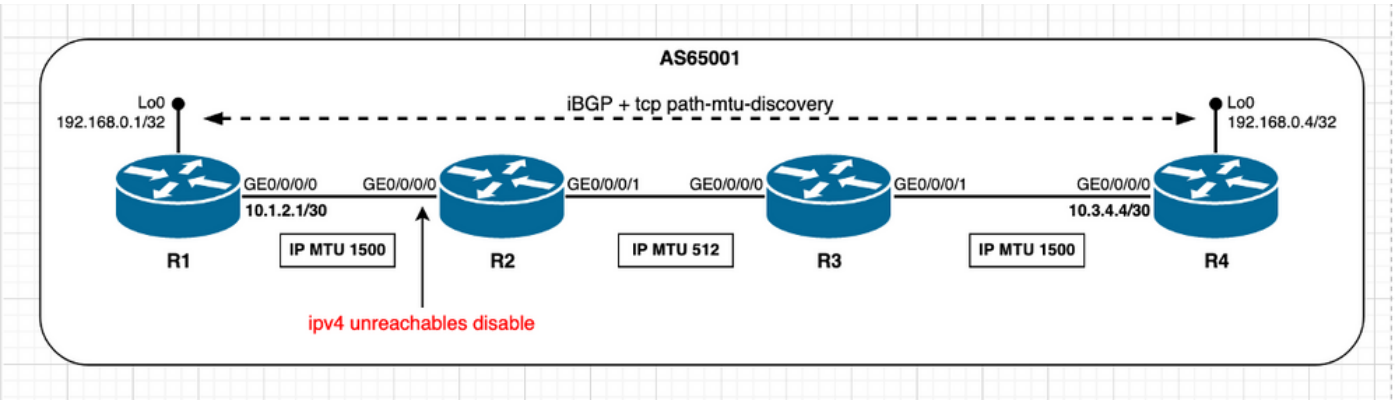

Image 3.5 - Die PMTUD ist auf R1/R4- und R2-IPv4-Unerreichbaren aktiviert.

IPv4-Unerreichbare bei R2 deaktiviert:

```
!- R2 - IP unreachables is disabled

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

RP/0/0/CPU0:R2#show ipv4 interface gigabitEthernet 0/0/0/0
Thu May 13 12:10:04.112 UTC
GigabitEthernet0/0/0/0 is Up, ipv4 protocol is Up
 Vrf is default (vrfid 0x60000000)
 Internet address is 10.1.2.2/30
 MTU is 1514 (1500 is available to IP)
 Helper address is not set
 Multicast reserved groups joined: 224.0.0.2 224.0.0.1 224.0.0.5
     224.0.0.6
 Directed broadcast forwarding is disabled
 Outgoing access list is not set
 Inbound  common access list is not set, access list is not set
 Proxy ARP is disabled
 ICMP redirects are never sent
 ICMP unreachables are never sent
 ICMP mask replies are never sent
 Table Id is 0xe0000000
```

Cisco IOS XR behandelt dieses Blackhole-Szenario folgendermaßen: Das gleiche Paket wird zweimal übertragen, und wenn es immer noch nicht erfolgreich ist, wird die erwartete TCP-ACK nicht empfangen. Versuchen Sie es dann erneut, aber verwenden Sie den nächstniedrigeren, klar definierten Plateauwert, wie in RFC1191 beschrieben: MTU-Pfaderkennung (siehe Abschnitt **PMTUD - Pfadsegment hat niedrigere IP-MTU-Liste** für**) Plateaus).** Zusammenfassend geht Cisco IOS XR davon aus, dass Pakete aufgrund ihrer Größe irgendwo im Pfad zum Ziel verworfen werden und über die Paketreübertragung umgangen werden. Dieses Verhalten kann im nächsten Beispiel aus einer Paketerfassung an der Knoten-R1-Schnittstelle und der Ausgabe von **debug tcp pmtud** beobachtet werden.

IOS-XR Blackhole-Erkennung bei R1:

```
! – at R1
! – Original BGP Update message is sent
! – Note IP Total Length of 1116 bytes and TCP Segment Length of 1076 bytes
```

```
! – R2 filters such packet and send and ICMP error message towards R1 which triggers PMTUD
! – But because IPv4 unreachables are disabled at R2 GE0/0/0/0 ICMP message is not sent
! – Hence BGP message is silently filtered at R2

562    7.638774       192.168.0.1 192.168.0.4 BGP    1130   UPDATE Message, KEEPALIVE Message

Frame 562: 1130 bytes on wire (9040 bits), 1130 bytes captured (9040 bits) on interface 0
Ethernet II, Src: fa:16:3e:42:18:05 (fa:16:3e:42:18:05), Dst: fa:16:3e:5c:f1:80
(fa:16:3e:5c:f1:80)
Internet Protocol Version 4, Src: 192.168.0.1, Dst: 192.168.0.4
   0100 .... = Version: 4
   .... 0101 = Header Length: 20 bytes (5)
   Differentiated Services Field: 0xc0 (DSCP: CS6, ECN: Not-ECT)
   Total Length: 1116
   Identification: 0x4a37 (18999)
   Flags: 0x02 (Don't Fragment)
       0... .... = Reserved bit: Not set
       .1.. .... = Don't fragment: Set
       ..0. .... = More fragments: Not set
   Fragment offset: 0
   Time to live: 255
   Protocol: TCP (6)
   Header checksum: 0x229b [validation disabled]
   [Header checksum status: Unverified]
   Source: 192.168.0.1
   Destination: 192.168.0.4
   [Source GeoIP: Unknown]
   [Destination GeoIP: Unknown]
Transmission Control Protocol, Src Port: 179, Dst Port: 57082, Seq: 318, Ack: 251, Len: 1076
Border Gateway Protocol - UPDATE Message
Border Gateway Protocol - KEEPALIVE Message
<snip>

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

563    0.560058       192.168.0.1 192.168.0.4 TCP    1130   [TCP Retransmission] 179  57082
[PSH, ACK] Seq=318 Ack=251 Win=32593 Len=1076
564    1.101367       192.168.0.1 192.168.0.4 TCP    1130   [TCP Retransmission] 179  57082
[PSH, ACK] Seq=318 Ack=251 Win=32593 Len=1076

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

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

567    1.850294       192.168.0.1 192.168.0.4 TCP    1130   [TCP Retransmission] 179  57082
[PSH, ACK] Seq=318 Ack=251 Win=32593 Len=1076
568    1.111361       192.168.0.1 192.168.0.4 TCP    1130   [TCP Retransmission] 179  57082
[PSH, ACK] Seq=318 Ack=251 Win=32593 Len=1076

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

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

569    2.198327       192.168.0.1 192.168.0.4 TCP    1020   [TCP Retransmission] 179  57082
[ACK] Seq=318 Ack=251 Win=32593 Len=966
```

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

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

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

571    2.205552       192.168.0.1 192.168.0.4 TCP    522    [TCP Retransmission] 179  57082
[ACK] Seq=318 Ack=251 Win=32593 Len=468
573    0.004254       192.168.0.1 192.168.0.4 TCP    522    [TCP Retransmission] 179  57082
[ACK] Seq=786 Ack=251 Win=32593 Len=468
574    0.002724       192.168.0.1 192.168.0.4 TCP    194    [TCP Retransmission] 179  57082
[PSH, ACK] Seq=1254 Ack=251 Win=32593 Len=140

! – Peer R4 TCP ACK is received

575    0.223172       192.168.0.4 192.168.0.1 TCP    54     57082  179 [ACK] Seq=251 Ack=1394
Win=31469 Len=0
```