

# Cisco IOS XR 및 BGP에서 경로 MTU 검색 확인

## 목차

[소개](#)

[배경 정보](#)

[TCP PMTUD 및 TCP MSS](#)

[시나리오 - TCP PMTUD 사용 안 함](#)

[기본 MTU 값 사용](#)

[기본값이 아닌 MTU 값 사용 - 활성 TCP 피어](#)

[기본값이 아닌 MTU 값 사용 - 수동 TCP 피어](#)

[TCP 옵션 사용 - XR Active](#)

[TCP 옵션 사용 - XR 수동](#)

[직접 연결되지 않은 TCP 피어](#)

[직접 연결되지 않은 TCP 피어 - TCP 옵션 사용\(MD5\)](#)

[직접 연결되지 않은 TCP 피어 - 경로 세그먼트의 IP MTU 감소](#)

[시나리오 - TCP PMTUD 사용](#)

[PMTUD 사용](#)

[PMTUD - 경로 세그먼트에 더 낮은 IP MTU가 있음](#)

[PMTUD - TCP 옵션\(MD5\)](#)

[PMTUD - 블랙홀 탐지](#)

## 소개

이 문서에서는 Cisco IOS® XR 장치의 TCP(Transmission Control Protocol) MTU(Path Maximum Transmission Unit) Discovery(PMTUD)에 대해 설명합니다.

## 배경 정보

PMTUD 메커니즘은 두 호스트 사이의 경로에 따라 프래그먼트화가 필요하지 않은 가장 큰 IP(Internet Protocol) 패킷 크기를 결정하려고 시도합니다. 설정된 값은 경로 MTU로 지정되며 각 홉의 최소 MTU 값과 같습니다. 정보를 전송할 때 Path MTU를 고려하는 경우 네트워크 용량을 최대한 활용하고 조각화 및 전송 효율성을 방지할 수 있습니다. PMTUD 메커니즘과 구현은 PMTUD 동작을 점차적으로 보여주는 클라이언트 프로토콜로 BGP(Border Gateway Protocol)를 사용하여 다양한 시나리오 세트에 도입됩니다.

## TCP PMTUD 및 TCP MSS

TCP는 PMTUD 결과를 활용하여 로컬 MSS(Maximum Segment Size)에 영향을 줍니다. 즉, 검색된 경로 MTU에 동적으로 적용됩니다. 따라서 PMTUD로 이동하기 전에 TCP MSS(Maximum Segment Size)를 신속하게 검토하고 MSS가 의미하는 내용과 그 목적을 파악할 수 있습니다.

[RFC879](#)의 MSS 원래 정의에 따라: MSS 옵션의 정의는 다음과 같습니다. IP 헤더 옵션이 없는 IP 데이터그램에서 전송되는 TCP 헤더 옵션이 없는 TCP 세그먼트에서 이 TCP 옵션의 발신자가 수신할 수 있는 최대 데이터 패킷 수입니다.

몇 가지 측면을 명확히 하고 구현자에게 조언을 제공하기 위해 [RFC6691](#) MSS 값을 계산하는 방법을 강조 표시합니다.

TCP MSS 옵션에 넣을 값을 계산할 때 MTU 값은 고정 IP 및 TCP 헤더의 크기만 줄이고 가능한 IP 또는 TCP 옵션을 고려하도록 줄이지 않아야 합니다. 반대로 발신자는 전송한 패킷에 포함된 IP 또는 TCP 옵션을 고려하려면 TCP 데이터 길이를 줄여야 합니다.

MSS에 대한 보다 자세한 정의는 [Cisco ASR 9000 Series Routers, IOS XR Release 6.7.x용 라우팅 컨피그레이션 가이드에서](#) 추출할 수 있습니다.

MSS는 컴퓨터 또는 통신 디바이스가 단편화되지 않은 단일 TCP 세그먼트에서 수신할 수 있는 최대 데이터 양입니다. 모든 TCP 세션은 단일 패킷으로 전송할 수 있는 바이트 수의 제한에 의해 제한됩니다. 이 제한은 MSS입니다. TCP는 패킷을 IP 레이어로 전달하기 전에 전송 큐의 청크로 분할합니다.

TCP MSS 값은 인터페이스의 MTU에 따라 달라집니다. 이는 한 인스턴스에서 프로토콜에 의해 전송할 수 있는 데이터의 최대 길이입니다. 최대 TCP 패킷 길이는 소스 디바이스에서 아웃바운드 인터페이스의 MTU와 TCP 설정 프로세스 중에 대상 디바이스에서 발표한 MSS에 의해 결정됩니다. MSS가 MTU에 가까울수록 BGP 메시지를 전송하는 것이 더 효율적입니다. 데이터 흐름의 각 방향은 서로 다른 MSS 값을 사용할 수 있습니다.

그러면 TCP가 지정된 TCP 세션의 MSS에 대해 고려해야 하는 값은 무엇입니까? 어떻게 계산하죠?

RFC879에 따라 기본값에 대해 다음을 수행합니다. 대상 호스트가 더 큰 데이터그램을 수락할 준비가 되어 있다는 특별한 지식이 없는 경우, 호스트는 576보다 큰 데이터그램을 전송해서는 안 됩니다. TCP 최대 세그먼트 크기는 IP MAXIMUM DATAGRAM SIZE MINUS FORTY입니다.

기본 IP 최대 데이터그램 크기는 576입니다.

기본 TCP 최대 세그먼트 크기는 536입니다.

이는 576바이트의 IP MTU 값을 고려합니다. 그러나 실제 IP MTU 값을 무시하면 TCP MSS 계산을 다음과 같이 요약할 수 있습니다.

- Active Peer - SYN 패킷으로 초기 MSS를 계산하고 전송합니다.

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

Where,

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

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

- Passive Peer - 초기 MSS를 계산하고, Active Peer에서 받은 MSS와 비교하고, SYN, ACK를 이 러한 MSS 값 중 낮은 값으로 전송합니다.

```
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.

MSS 옵션 값과 관련된 협상은 없습니다. 각 노드는 고유한 값을 결정하고 TCP 세션 설정 시 이를 알립니다. MSS 계산에 고려되는 IP MTU 값을 PMTUD에서 파생시킬 수 있는 경우 MSS 값을 지정된 경로 MTU에 대해 가장 효과적인 값으로 조정할 수 있습니다. Cisco IOS XR 동작에는 MSS 계산 및 PMTUD 역할과 관련된 몇 가지 세부 사항이 요약되어 있습니다.

PMTUD는 Cisco IOS XR에서 기본적으로 비활성화되어 있습니다.

- 로컬 초기 MSS 계산에서는 IP MTU를 다음과 같이 간주합니다. 직접 연결된 피어가 있는 경우 이그레스 인터페이스 IP MTU를 고려하십시오. 직접 연결되지 않은 피어의 경우 IP MTU가 1280바이트인 것을 고려하십시오. MSS 값은 구성된 TCP 옵션의 영향을 받습니다.

Cisco IOS XR에서 PMTUD가 활성화된 경우:

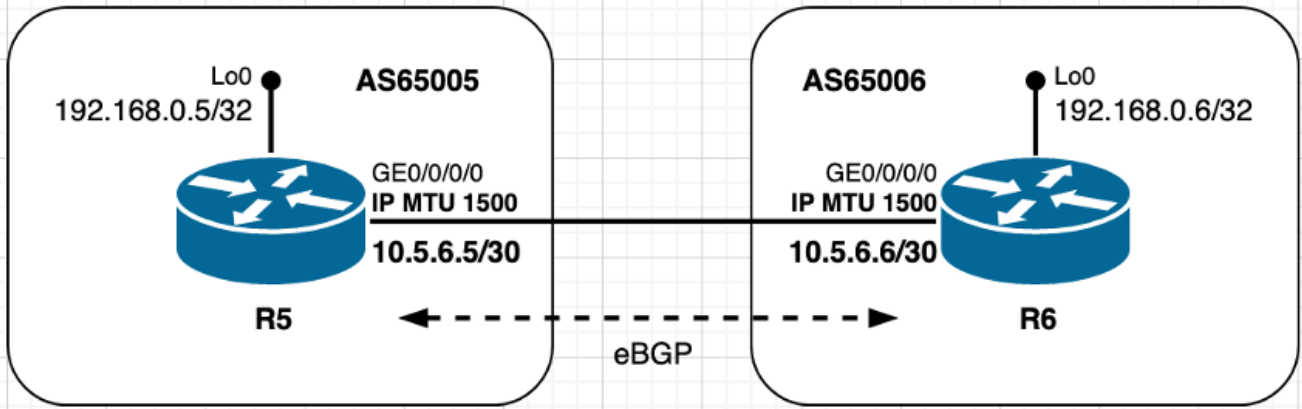
- 로컬 초기 MSS 계산에서는 IP MTU를 다음과 같이 간주합니다. 직접/직접 연결되지 않은 피어에 관계없이 이그레스 인터페이스 IP MTU를 고려하십시오. MSS 값은 구성된 TCP 옵션의 영향을 받습니다.

PMTUD 역학 및 구현에 대한 추가 세부 사항이 있으며, 이 문서에서는 다음 표에 요약된 실제 예를 소개합니다. 이 표는 또한 Active 및 Passive TCP 피어 IP MTU 및 고려된 각 시나리오에 대해 선택한 MSS 값을 나타냅니다.

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

## 시나리오 - TCP PMTUD 사용 안 함

### 기본 MTU 값 사용



이미지 2.1. 기본 MTU 값 포함

이미지 2.1 R6에 표시된 eBGP 피어가 TCP 연결을 관리하는 경우, 이는 활성 역할을 수행하고 대상 포트 179에서 R5로 TCP 세션을 시작합니다. 피어가 직접 연결되며, 둘 다 각 인터페이스에서 기본 IP MTU 값을 사용합니다. 이 문서의 시작 부분에서 공유된 정보를 기준으로 이 시나리오의 MSS 계산을 다음과 같이 요약할 수 있습니다.

- 두 노드 모두 기본 IP MTU 1500바이트를 사용합니다.
- TCP 경로 MTU 검색은 기본적으로 비활성화되어 있습니다.
- TCP 피어가 직접 연결된 R6은 BGP 연결을 관리합니다. R6은 MSS가 1460바이트인 SYN을 전송합니다.  $1500(\text{인터페이스 IP MTU}) - 20(\text{minTCP\_H}) - 20(\text{minIP\_H})$  R5는 SYN, ACK를 1460바이트의 MSS로 전송합니다. [Received MSS; 로컬 초기 MSS]수신된 MSS 1460바이트; 로컬 초기 MSS 1460바이트가 가장 낮은 MSS 값은 두 피어 모두에서 사용됩니다.

R6 - ACTIVE에 표시된 TCP 세션 세부 정보:

! - 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

**R5 - PASSIVE에 표시된 TCP 세션 세부 정보:**

! - 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

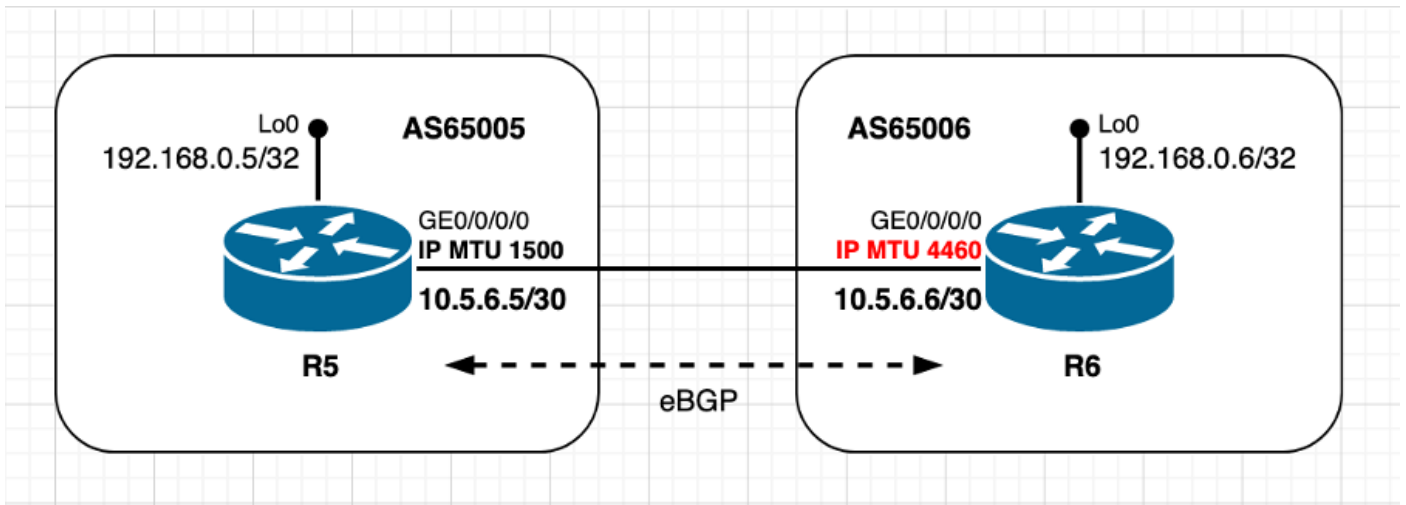
Socket send buffer : Low/High watermark 2048/24576, Notify threshold 0

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#
```

## 기본값이 아닌 MTU 값 사용 - 활성 TCP 피어



이미지 2.2 - 활성 피어가 기본이 아닌 MTU 값을 사용합니다.

이 시나리오는 기본적으로 이전 시나리오와 동일하며, 현재 활성 TCP 피어 R6에서 기본값이 아닌 IP MTU 값을 사용하는 유일한 차이점입니다. MSS 값에 대한 초기 계산 및 결정은 패시브 TCP 피어 R5에 의해 수행되는 방법에 유의하십시오. 이 시나리오의 TCP MSS 계산은 다음과 같이 요약할 수 있습니다.

- R6은 기본이 아닌 IP MTU 4460바이트 사용
- TCP 경로 MTU 검색은 기본적으로 비활성화되어 있습니다.
- TCP 피어가 직접 연결된 R6은 BGP 연결을 관리합니다. R6은 MSS가 4420바이트인 SYN을 전송합니다.  $4460(\text{인터페이스 IP MTU}) - 20(\text{minTCP\_H}) - 20(\text{minIP\_H})$  R5 전송 SYN, ACK(MSS 1460바이트) [Received MSS; 로컬 초기 MSS] 수신된 MSS 4420바이트; 로컬 초기 MSS 1460바이트가 가장 낮은 MSS 값은 두 피어 모두에서 사용됩니다.

R6에서 제공된 TCP SYN:

```
! - 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)
```

## R5에서 제공된 TCP SYN, ACK:

! - 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)
```

## R6 - ACTIVE에 표시된 TCP 세션 세부 정보:

! - 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#

**R5 - PASSIVE에 표시된 TCP 세션 세부 정보:**

! - 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, KRRTT: 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:

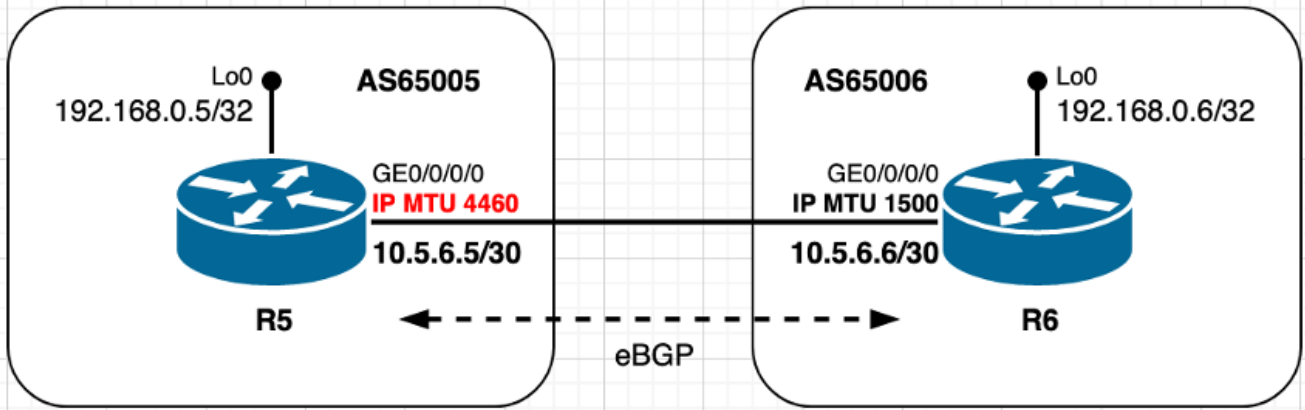
#PDU's in buffer: 0

FIB Lookup Cache: IFH: 0x40 PD ctx: size: 0 data:

Num Labels: 0 Label Stack:

RP/0/0/CPU0:R5#

**기본값이 아닌 MTU 값 사용 - 수동 TCP 피어**



이미지 2.3 - PASSIVE 피어가 기본이 아닌 MTU 값을 사용합니다.

여전히 동일한 eBGP 시나리오로, 이제 기본 IP MTU가 아닌 액티브 TCP 피어 R5가 구성된 패시브 TCP 피어 R6과 기본 IP MTU 값이 있는 액티브 TCP 피어 R6을 사용합니다. 이전 시나리오와 마찬가지로, 패시브 피어 R5에서 MSS 값을 선택하는 방법을 기록해 두십시오. 이 시나리오의 TCP MSS 계산은 다음과 같이 요약할 수 있습니다.

- R5는 기본이 아닌 IP MTU 4460바이트 사용
- TCP 경로 MTU 검색은 기본적으로 비활성화되어 있습니다.
- TCP 피어가 직접 연결된 R6은 BGP 연결을 관리합니다. R6은 MSS가 1460바이트인 SYN을 전송합니다.  $1500(\text{인터페이스 IP MTU}) - 20(\text{minTCP\_H}) - 20(\text{minIP\_H})$  R5 전송 SYN, ACK(MSS 1460바이트) [Received MSS; 로컬 초기 MSS] 수신된 MSS 1460바이트; 로컬 초기 MSS 4420바이트가 가장 낮은 MSS 값은 두 피어 모두에서 사용됩니다.

R6에서 제공된 TCP SYN:

! - 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)

## R5에서 제공된 TCP SYN, ACK:

! - 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)

## R6 - ACTIVE에 표시된 TCP 세션 세부 정보:

! - 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   rcvadp: 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#

## R5 - PASSIVE에 표시된 TCP 세션 세부 정보:

! - 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

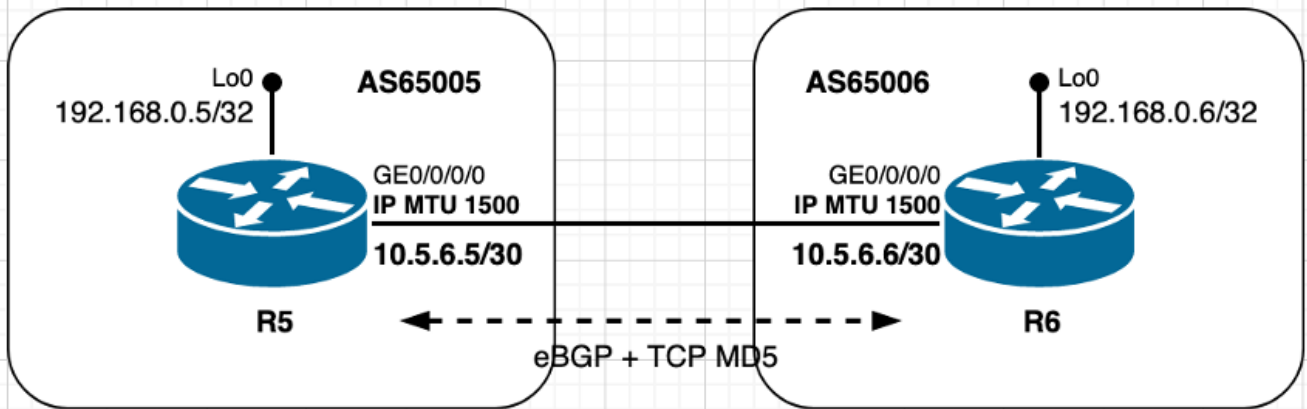
PDU information:  
#PDU's in buffer: 0  
FIB Lookup Cache: IFH: 0x40 PD ctx: size: 0 data:  
Num Labels: 0 Label Stack:

RP/0/0/CPU0:R5#

## TCP 옵션 사용 - XR Active

이 문서에서 앞서 언급한 대로, TCP 옵션(예: [TCP MD5](#), [TCP selective-ack](#) 또는 [TCP 타임스탬프](#))을 사용하면 MSS 계산에서 고려될 추가 바이트가 생성되므로 MSS 계산에 영향을 줍니다.

이 섹션과 다음 섹션에서는 TCP 옵션이 있을 때 피어가 수행한 MSS 계산을 설명합니다. TCP MD5 인증 옵션이 예로 사용됩니다. 이미지에 표시된 이미지 2.4의 참조 시나리오를 참조하십시오.



이미지 2.4 - TCP 옵션 사용(MD5) - XR 활성화

이 시나리오에서는 두 피어가 기본 IP MTU 값을 사용하고, 직접 연결되며, 피어 R6은 TCP 활성화 역할을 합니다. 이미 공유한 것처럼 추가 오버헤드를 위해 TCP MD5 인증 계정의 컨피그레이션 및 사용을 참조하십시오. 이 특정 시나리오의 TCP MSS 계산은 다음과 같이 요약할 수 있습니다.

- 두 노드 모두 기본 IP MTU 1500바이트를 사용합니다.
- TCP 경로 MTU 검색은 기본적으로 비활성화되어 있습니다.
- TCP 피어가 직접 연결됨
- 두 노드에서 모두 활성화된 TCP MD5 인증 R6은 BGP 연결을 관리합니다. R6은 MSS가 1436바이트인 SYN을 전송합니다.  $1500(\text{인터페이스 IP MTU}) - 20(\text{minTCP\_H}) - 20(\text{minIP\_H}) - 24\text{바이트}(\text{IOS XR TCP 옵션 오버헤드})$  R5 전송 SYN, ACK(MSS 1436바이트) [Received MSS; 로컬 초기 MSS]수신된 MSS 1436바이트; 로컬 초기 MSS 1460바이트가장 낮은 MSS 값은 두 피어 모두에서 사용됩니다.

요약에서 볼 수 있듯이 Cisco IOS XR의 동작 방식은 [RFC 879](#)와 [RFC 6691](#)에 따라 엄격하게 동작하지 않으며, MSS 계산에서 TCP 옵션을 계상해서는 안 된다고 명시되어 있습니다.

tcp 헤더 길이에 대한 추가 요소의 Cisco IOS XR 어카운트는 Cisco 버그 ID CSCvf20166에 [자세히 설명되어 있습니다](#).

"(...)XR에서 BGP 연결을 시작할 때 BGP는 먼저 소켓을 생성한 다음 MD5를 포함한 소켓 옵션을 설정합니다. 이렇게 하면 tcp 옵션 헤더 길이가 24가 됩니다. 따라서 초기 MSS는  $1500 - 40 - 24 = 1436$ 이 됩니다. 피어로 전송되며 피어가  $\min(1436, 1460) = 1436$ (...)을 사용합니다.

R6에서 제공된 TCP SYN:

! - 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)

```

### R5에서 제공된 TCP SYN, ACK:

! - 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)

```

### R6 - ACTIVE에 표시된 TCP 세션 세부 정보:

! - 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#

**R5 - PASSIVE에 표시된 TCP 세션 세부 정보:**

! - 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#

다른 TCP 옵션에서도 유사한 동작을 관찰할 수 있습니다. 이 옵션은 추가 오버헤드를 위해 계정을 구성하고 Cisco IOS XR에서 MSS 계산에 영향을 줍니다. TCP 타임스탬프 및 TCP selective-ack 옵션이 구성된 경우 MSS 계산을 문서화하는 동일한 시나리오와 이러한 예를 고려하십시오.

## R6 - ACTIVE에 표시된 TCP 세션 세부 정보 - TCP 옵션 타임스탬프 및 선택적 ACK 옵션이 구성된 경우:

```
! - 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 옵션 사용 - XR 수동

이전 시나리오에서 초기 MSS 계산과 관련하여 패시브 역할에서 Cisco IOS XR 노드의 고유한 동작을 감지했을 것입니다. 노드는 **tcp 옵션 헤더 길이**를 고려하지 않습니다. 이 시나리오는 Cisco 버그 ID에 의해 설명되는 이 고유한 동작을 강조하기 위한 것입니다.

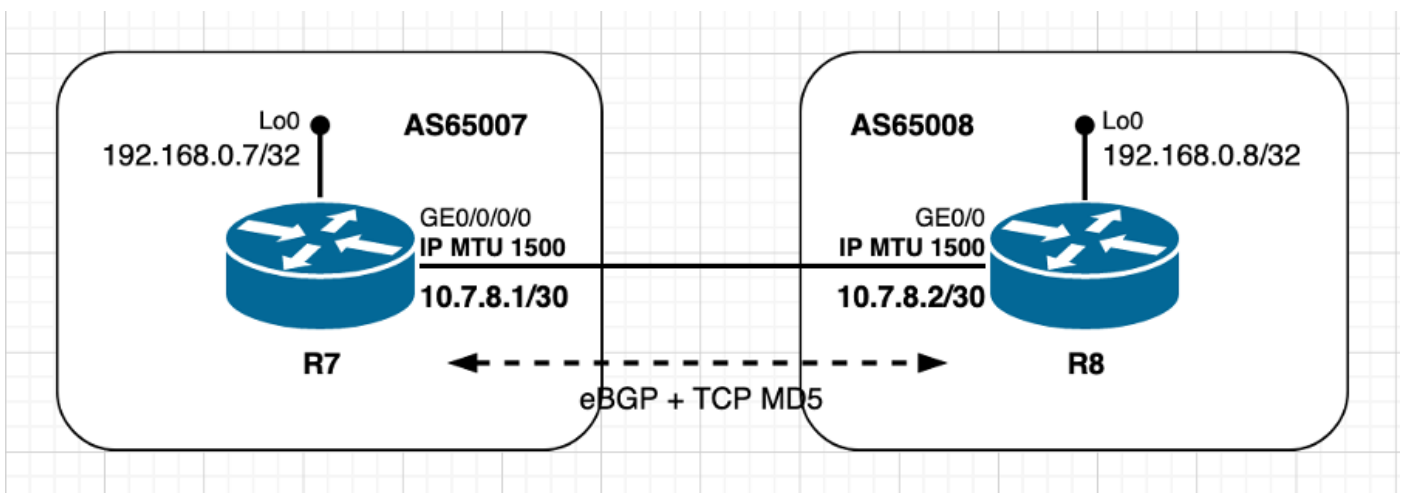
"(...) - 피어가 연결을 시작하면 초기 MSS를 1460으로 보냅니다. XR TCP는 소켓, pcb 등을 생성한 다음 지정된 순서대로 두 가지 작업을 수행합니다.

- 먼저 **tcp 옵션 헤더 길이**를 뺀 후 초기 MSS를 계산합니다. MD5 옵션은 수신 소켓에서 이 소켓에 아직 상속되지 않았으므로 '0'입니다.

- 그런 다음 'MD5' 및 기타 옵션을 상속하므로 '옵션 헤더 바이트 길이'가 24로 설정됩니다.

따라서 이 경우 XR TCP는 초기 MSS를 1460으로 전송하기 때문에 이 두 가지가 모두 사용됩니다. (...)"

이 시나리오에서는 활성 TCP 피어 R8이 Cisco IOS 노드이지만 이 사실은 시나리오가 무엇을 강조 표시하는지에 대한 차이점이나 세부 사항을 설명하지 않습니다. 그럼에도 불구하고, 이전 섹션 시나리오에서 보여주는 Cisco IOS XR과 다른 점은, 액티브 TCP 피어 R8은 초기 MSS 계산에서 TCP 옵션을 고려하지 않는다는 점입니다.



이미지 2.5 - TCP 옵션 사용(MD5) - XR 패시브.

두 피어 모두 기본 IP MTU 값을 사용하며 직접 연결됩니다. Cisco IOS 피어 R8은 적극적인 역할을 수행합니다. 이 시나리오의 TCP MSS 계산은 다음과 같이 요약할 수 있습니다.

- 두 노드 모두 기본 IP MTU 1500바이트를 사용합니다.
- TCP 경로 MTU 검색은 Cisco IOS XR R7에서 기본적으로 비활성화되어 있습니다.
- TCP 경로 MTU 검색은 Cisco IOS R8에서 기본적으로 활성화되어 있습니다.
- TCP 피어가 직접 연결됨
- 두 노드에서 모두 활성화된 TCP MD5 인증 IOS R8은 BGP 연결을 관리합니다. IOS R8은 1460바이트의 MSS로 SYN을 전송합니다. 1500(인터페이스 IP MTU) - 20(minTCP\_H) - 20(minIP\_H) IOS XR R7은 SYN, ACK를 1460바이트의 MSS로 전송합니다. [Received MSS; 로컬 초기 MSS]수신된 MSS 1460바이트; 로컬 초기 MSS 1460바이트가장 낮은 MSS 값은 두 피어 모두에서 사용됩니다.

#### R8에서 제공된 TCP SYN - 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, R7 - Cisco IOS XR에서 제공된 ACK:

! - 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)

```

## R8 - Cisco IOS - ACTIVE에 표시된 TCP 세션 세부 정보:

! - 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 NLRI 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, Minimum 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#

## R7 - Cisco IOS XR - PASSIVE에 표시된 TCP 세션 세부 정보:

! - 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

```

```

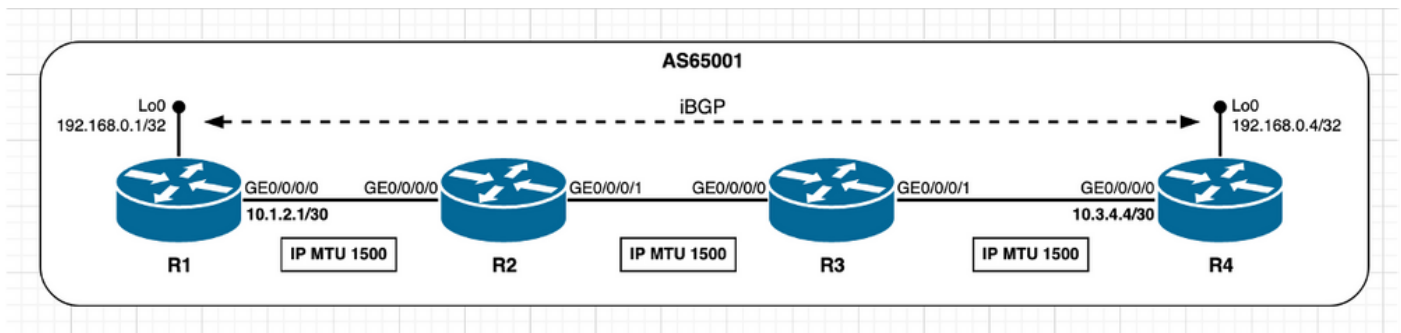
PDU information:
#PDU's in buffer: 0
FIB Lookup Cache: IFH: 0x40 PD ctx: size: 0 data:
Num Labels: 0 Label Stack:

```

RP/0/0/CPU0:R7#

## 직접 연결되지 않은 TCP 피어

피어가 직접 연결되지 않은 경우 TCP MSS 초기 계산을 수행하는 방법은 이 문서의 소개 섹션에서 앞서 설명한 대로 변경됩니다. 기본 IP MTU 값으로 구성된 모든 피어가 포함된 iBGP 세션의 시나리오는 MSS 계산을 거치는 데 사용됩니다.



이미지 2.6 - 직접 연결되지 않은 TCP 피어 - iBGP.

주목해야 할 중요한 점은 TCP 경로 MTU 검색을 비활성화하고 피어가 직접 연결되지 않은 경우 설계에 따라 Cisco IOS XR에서 고정 IP MTU 값 1280바이트를 사용한다는 것입니다.

이전 이미지에서 R4는 활성 역할을 수행하고 TCP 연결을 관리하며, R4는 대상 포트 179에서 R1을 사용하는 TCP 세션을 엽니다. 두 노드는 인터페이스에서 기본 IP MTU 값을 사용합니다. 이 시나리오의 MSS 계산은 다음과 같이 요약할 수 있습니다.

- 모든 노드는 1500바이트의 기본 IP MTU를 사용합니다
- TCP 경로 MTU 검색은 기본적으로 비활성화되어 있습니다.
- TCP 피어가 직접 연결되지 않음 R4는 BGP 연결을 관리합니다.R4는 MSS가 1240바이트인 SYN을 전송합니다. 피어가 직접 연결되지 않고 TCP 경로 MTU 검색을 비활성화한 경우 인터페이스 MTU는 고려되지 않습니다.Cisco IOS XR 설계에 따라 1280바이트는 TCP\_DEFAULT\_MTU로 간주됩니다.1280(TCP\_DEFAULT\_MTU) - 20(minTCP\_H) - 20(minIP\_H)R1은 1240바이트의 MSS로 SYN, ACK 전송 [Received MSS; 로컬 초기 MSS]수신된 MSS 1240바이트; 로컬 초기 MSS 1240바이트가장 낮은 MSS 값은 두 피어 모두에서 사용됩니다.

R4에서 제공된 TCP SYN:

! - 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, R1에서 제공된 ACK:

! - 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)

```

### R4에 표시된 TCP 세션 세부 정보 - 활성:

! - 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#

**R1에 표시된 TCP 세션 세부 정보 - 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

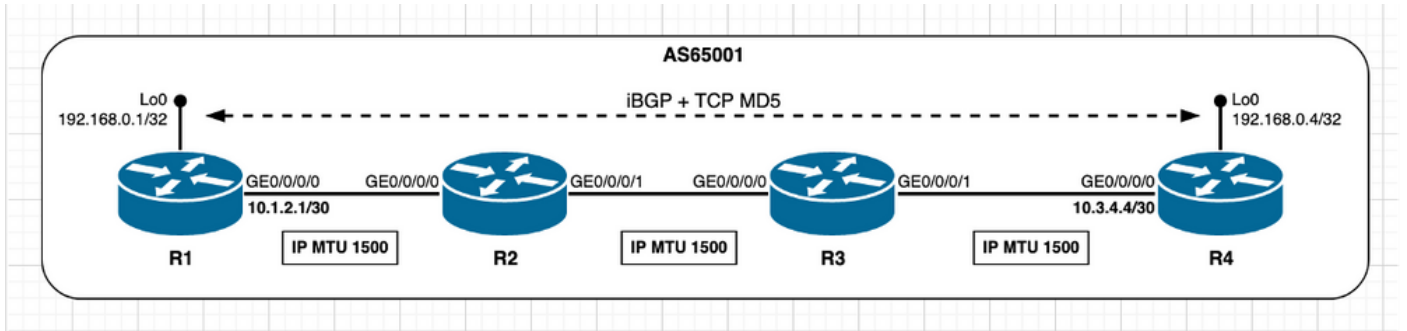
Socket options: SO\_REUSEADDR, SO\_REUSEPORT, SO\_NBIO  
Socket states: SS\_ISCONNECTED, SS\_PRIV  
Socket receive buffer states: SB\_DEL\_WAKEUP  
Socket send buffer states: SB\_DEL\_WAKEUP  
Socket receive buffer: Low/High watermark 1/32768  
Socket send buffer : Low/High watermark 2048/24576, Notify 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#

## 직접 연결되지 않은 TCP 피어 - TCP 옵션 사용(MD5)

직접 연결되지 않은 피어 시나리오 및 TCP MD5 인증을 사용하는 경우, 이전에 설명한 테스트 사례 또는 시나리오와 근본적인 차이가 없습니다. 이전에 TCP MD5 인증에서 살펴본 것처럼 Cisco IOS XR에서는 추가 오버헤드를 고려하며 초기 MSS 값은 동일하게 반영됩니다. TCP MSS 계산에 영향을 주는 TCP 옵션에 대한 자세한 내용은 이전 섹션 Use TCP Options - XR Active and Use TCP Options - XR Passive를 참조하십시오.



이미지 2.7 - 직접 연결되지 않은 TCP 피어 - iBGP + TCP MD5.

이 시나리오의 TCP MSS 계산은 다음과 같이 요약할 수 있습니다.

- 모든 노드는 1500바이트의 기본 IP MTU를 사용합니다
- TCP 경로 MTU 검색은 기본적으로 비활성화되어 있습니다.
- TCP 피어가 직접 연결되지 않음 R4는 BGP 연결을 관리합니다. 대상 R1이 직접 연결되지 않음 R4는 MSS가 1216바이트인 SYN을 전송합니다. 피어가 직접 연결되지 않고 TCP 경로 MTU 검색을 비활성화한 경우 인터페이스 MTU는 고려되지 않습니다. 설계에 따라 1280바이트는 TCP\_DEFAULT\_MTU로 간주됩니다.  $1280(\text{TCP\_DEFAULT\_MTU}) - 20(\text{minTCP\_H}) - 20(\text{minIP\_H}) - 24\text{바이트}(\text{IOS XR TCP 옵션 오버헤드})$  R1은 1216바이트의 MSS로 SYN, ACK 전송 [Received MSS; 로컬 초기 MSS] 수신된 MSS 1216바이트; 로컬 초기 MSS 1240바이트가 가장 낮은 MSS 값은 두 피어 모두에서 사용됩니다.

R4에서 제공된 TCP SYN:

! - TCP SYN sourced from R4

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

```
Frame 3425: 82 bytes on wire (656 bits), 82 bytes captured (656 bits) on interface 0
Ethernet II, Src: fa:16:3e:d7:7e:f6 (fa:16:3e:d7:7e:f6), Dst: fa:16:3e:8f:8f:54
(fa:16:3e:8f:8f:54)
```

```
Internet Protocol Version 4, Src: 192.168.0.4, Dst: 192.168.0.1
```

```
Transmission Control Protocol, Src Port: 42135, Dst Port: 179, Seq: 0, Len: 0
```

```
Source Port: 42135
```

```
Destination Port: 179
```

```
[Stream index: 10]
```

```
[TCP Segment Len: 0]
```

```
Sequence number: 0 (relative sequence number)
```

```
Acknowledgment number: 0
```

```
Header Length: 48 bytes
```

```
Flags: 0x002 (SYN)
```

```
Window size value: 16384
```

```
[Calculated window size: 16384]
```

```
Checksum: 0xc503 [unverified]
```

```
[Checksum Status: Unverified]
```

```
Urgent pointer: 0
```

```
Options: (28 bytes), Maximum segment size, Window scale, No-Operation (NOP), TCP MD5
signature, End of Option List (EOL)
```

Maximum segment size: 1216 bytes  
Kind: Maximum Segment Size (2)  
Length: 4  
**MSS Value: 1216**  
Window scale: 0 (multiply by 1)  
No-Operation (NOP)  
TCP MD5 signature  
End of Option List (EOL)

## TCP SYN, R1에서 제공된 ACK:

! - 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)

## R4에 표시된 TCP 세션 세부 정보 - 활성:

! - 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 rcvadp: 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#

## R1에 표시된 TCP 세션 세부 정보 - 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

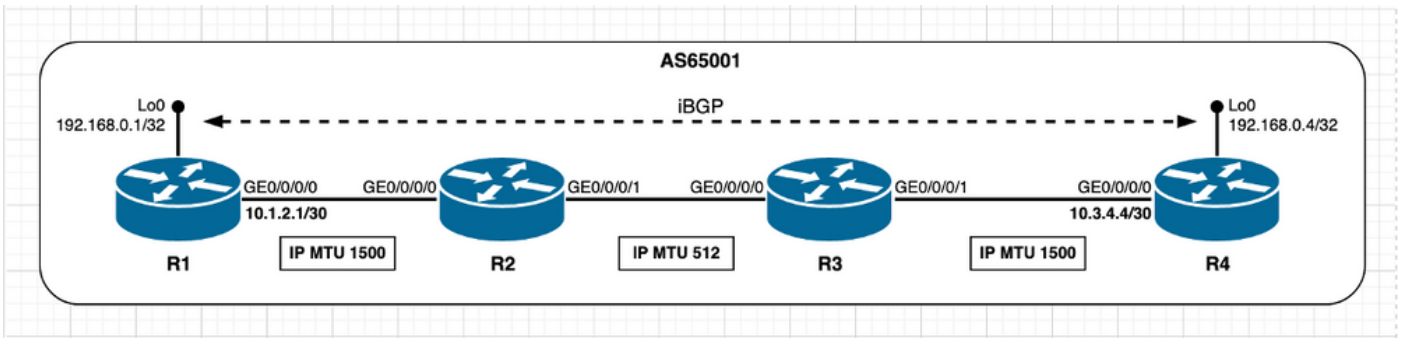
PDU information:  
#PDU's in buffer: 0  
FIB Lookup Cache: IFH: 0x40 PD ctx: size: 0 data:  
Num Labels: 0 Label Stack:

RP/0/0/CPU0:R1#

## 직접 연결되지 않은 TCP 피어 - 경로 세그먼트의 IP MTU 감소

다음 시나리오에서 목표는 IP MTU가 더 낮은 중간 경로 세그먼트가 있는 경우, 기본 상태에서는 TCP PMTUD가 비활성화됨을 의미합니다. 이 이미지를 참조하십시오.





이미지 2.8 - R2/R3 경로 세그먼트의 IP MTU가 낮습니다.

초기 시나리오에서는 BGP 정보가 최소로 간주되므로, BGP 피어 간에 교환해야 하는 모든 작업은 최소 경로 MTU 512바이트에 맞는 IP 패킷으로 수행할 수 있습니다. 이 가정을 통해 MSS 계산은 직접 연결되지 않은 TCP 피어 섹션에 설명된 대로 수행됩니다. R1과 R4 모두 1240바이트의 MSS 값을 선택합니다.

R4에 표시된 TCP 세션 세부 정보 - 활성:

! - 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 rcvadp: 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>

## R1에 표시된 TCP 세션 세부 정보 - 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 rcvadp: 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>

이제 BGP 세션이 설정되면 최소 경로 MTU가 512바이트보다 큰 BGP 업데이트 메시지가 트리거되는 것을 고려하십시오. 출력에서 볼 수 있듯이 Cisco IOS XR는 BGP 업데이트 메시지와 함께 df-bit를 설정하지 않습니다. 즉, BGP 정보는 중간 노드의 패킷 조각화의 비용으로 전송됩니다.

## R1에서 제공하는 BGP 업데이트 - 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: ff

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

노드 R1에서 제공하는 BGP 업데이트 메시지의 조각화는 R2 인터페이스 GE0/0/0/1에서 트래픽 캡처가 수행하는 것처럼 노드 R2에서 발생합니다.

노드 R2의 IP 단편화:

! - 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

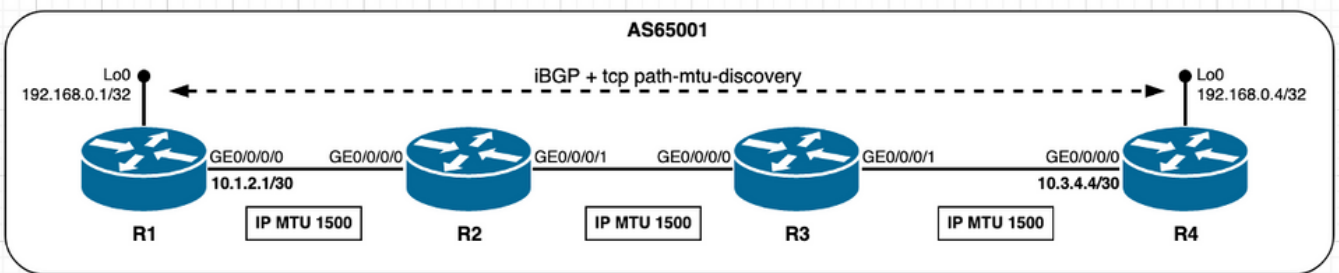
```
[Source GeoIP: Unknown]
[Destination GeoIP: Unknown]
Data (101 bytes)
<snip>
```

# 시나리오 - TCP PMTUD 사용

## PMTUD 사용

PMTUD가 활성화되면 피어가 직접 연결되거나 직접 연결되지 않은 경우에도 MSS 초기 계산에서는 항상 이그레스 인터페이스 IP MTU를 고려합니다.

이 시나리오는 PMTUD가 활성화된 경우 예상되는 동작에 대한 통찰력을 제공합니다. 여기서 Cisco IOS XR 노드 R4는 활성 역할을 수행하고, TCP 연결을 관리하고, 대상 포트 179에서 Cisco IOS XR 노드 R1을 사용하여 TCP 세션을 엽니다. 두 노드는 인터페이스에서 기본 IP MTU 값을 사용합니다.



이미지 3.1 - TCP PMTUD 활성화됨

이 시나리오의 MSS 계산은 다음과 같이 요약할 수 있습니다.

- 모든 노드는 1500바이트의 기본 IP MTU를 사용합니다
- TCP 경로 MTU 검색이 활성화됨
- TCP 피어가 직접 연결되지 않음 R4는 BGP 연결을 관리합니다.R4는 MSS가 1460바이트인 SYN을 전송합니다. 1500(인터페이스 IP MTU) - 20(minTCP\_H) - 20(minIP\_H)R1은 1460바이트의 MSS로 SYN, ACK 전송 [Received MSS; 로컬 초기 MSS]수신된 MSS 1460바이트; 로컬 초기 MSS 1460바이트가장 낮은 MSS 값은 두 피어 모두에서 사용됩니다.

enable PMTUD에서 도입된 동작 변경 사항을 강조하기 위해 다음 출력은 이벤트의 순서를 설명합니다.

1. PMTUD의 기본 시나리오에서 설정된 TCP 세션의 초기 상태입니다.
2. PMTUD는 TCP 피어 R4 및 R1 모두에서 구성 및 활성화됩니다.
3. TCP 세션이 재시작되고 MSS 계산이 수행되며 TCP PMTUD의 영향을 받습니다.

R4 - ACTIVE - TCP PMTUD 비활성화됨(기본값):

```
! - 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#

**R1 - PASSIVE - TCP PMTUD 비활성화됨(기본값):**

! - 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#  
**R4 - ACTIVE - TCP PMTUD가 활성화되었습니다.**

! - '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
<b>PmtuAger</b>	<b>1</b>	<b>0</b>	<b>508096</b>
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 rcvadp: 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#

## R1 - PASSIVE - TCP PMTUD 활성화됨:

! - 'debug tcp pmtud' output on R1  
! - tcp path mtu discovery is enabled and uses default Path MTU aging timer (10 min / 60000 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
<b>PmtuAger</b>	<b>1</b>	<b>0</b>	<b>562042</b>
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 rcvadp: 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#

**PMTU 관리자 타이머 동작을 확인합니다.**

! - 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

**R4 - ACTIVE - BGP 세션 재시작 - 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)

## R1 - PASSIVE - BGP 세션 재시작 - TCP SYN, ACK에서 볼 수 있습니다.

! - 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 PMTUD가 활성화되고 BGP 세션이 지워진 후 R4 - ACTIVE에 표시된 TCP 세션 세부 정보:

! - 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 PMTUD가 활성화되고 BGP 세션이 지워진 후 R1 - PASSIVE에 표시된 TCP 세션 세부 정보

```
! - 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

```

```

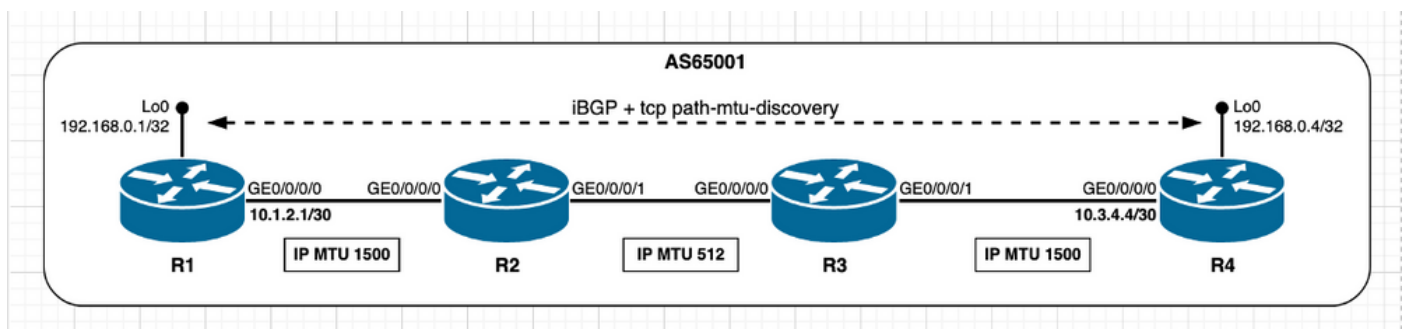
PDU information:
#PDU's in buffer: 0
FIB Lookup Cache: IFH: 0x40 PD ctx: size: 0 data:
Num Labels: 0 Label Stack:

```

RP/0/0/CPU0:R1#

## PMTUD - 경로 세그먼트에 더 낮은 IP MTU가 있음

이전 시나리오는 PMTUD가 활성화된 초기 TCP 세션 설정 시 발생하는 상황을 파악하는 데 도움이 되었습니다. 이 시나리오는 위에 구축되며 TCP PMTUD가 작동하는 방식 및 설정된 TCP 세션에 미치는 영향을 이해하는 데 도움이 됩니다.



### 이미지 3.2 - PMTUD가 활성화되고 경로 세그먼트의 IP MTU가 낮습니다.

이전 이미지를 참조로 가정하고, BGP 세션이 설정되었다고 가정하고, R1은 512바이트보다 큰 IP 패킷에서 전달되는 BGP 업데이트 메시지를 전송합니다. PMTUD가 활성화되면 DF 비트(Don't Fragment)가 설정됩니다. 따라서 노드 R2는 IP 패킷을 삭제하고 ICMP(Internet Control Message Protocol) 메시지(Destination Unreachable - type 3; 프래그먼트화 필요 - 코드 4) 다시 R1로 돌아갑니다. ICMP 메시지가 수신되면 노드 R1에서 PMTUD가 트리거되고 경로 최하위 IP MTU를 설정하려고 시도합니다. 잘 정의된 고원 수준 집합에서 다음 낮은 값을 사용하여 새 TCP 세션 MSS 값을 고려합니다. 그런 다음 TCP는 원래 BGP 업데이트를 새 MSS 값으로 재전송하고 이 프로세스는 ICMP 메시지(Destination Unreachable - type 3; 조각화 필요 - 코드 4)가 더 이상 수신되지 않습니다. 이는 전송 중인 모든 패킷이 가장 낮은 경로 세그먼트 IP MTU에 속하도록 MSS 값이 사용 중일 때까지입니다. 시간이 경과하면 PmtuAger 타이머에 의해 제어되는 PMTUD가 평원 수준을 반대 방향으로 통과하고 MSS를 최대 값으로 다시 높입니다. 지정된 시간에 ICMP 메시지(Destination Unreachable - type 3, 프래그먼트화 필요 - 코드 4)가 다시 수신되고 PMTUD가 이전에 설명한 대로 작동합니다.

다음 출력은 방금 설명한 PMTUD 동작을 거치고 설정된 TCP 세션의 시나리오에서 시작합니다. 여기서 Cisco IOS XR 노드 R4는 활성 역할을 수행하므로 TCP 연결을 관리하고 대상 포트 179에서 R1을 사용하는 TCP 세션을 엽니다. 두 노드는 인터페이스에서 기본 IP MTU 값을 사용합니다. 이 시나리오의 초기 MSS 계산은 다음과 같이 요약할 수 있습니다.

- R2와 R3 노드 간의 중간 세그먼트는 기본이 아닌 IP MTU 512바이트를 사용합니다.
- R1 및 R4는 인터페이스에서 기본 MTU 값을 사용합니다.
- TCP 경로 MTU 검색이 활성화되어 있습니다.
- TCP 피어가 직접 연결되지 않습니다. R4는 BGP 연결을 관리합니다. R4는 MSS가 1460바이트인 SYN을 전송합니다. 1500(인터페이스 IP MTU) - 20(minTCP\_H) - 20(minIP\_H) R1은 1460바이트의 MSS로 SYN, ACK를 전송합니다. [Received MSS; 로컬 초기 MSS]. 수신된 MSS 1460바이트; 로컬 초기 MSS 1460바이트가 가장 낮은 MSS 값은 두 피어 모두에서 사용됩니다.

R4에서 제공된 TCP SYN:

```
! - 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, R1에서 제공된 ACK:

```
! - 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)
```

BGP 세션이 설정된 상태에서 노드 R1은 BGP Update 메시지를 보내고 ICMP 메시지를 받습니다 (Destination Unreachable - type 3; 프래그먼트화 필요 - 코드 4) 노드 R2에서 소싱된 반환.

이 문제는 BGP 업데이트 메시지를 전달하는 IP 패킷에 DF 비트 집합이 있고 R2/R3 세그먼트에서 사용된 IP MTU가 1116바이트의 IP 패킷 크기보다 작기 때문입니다. 앞에서 설명한 대로 ICMP 메시지 수신은 PMTUD를 트리거합니다.

R1 ICMP에서 Type 3/Code 4 메시지가 수신됩니다.

```
! - 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]
```

ICMP 메시지에 의해 트리거된 노드 R1에서 TCP PMTUD는 IP MTU(well-defined platection) 수준 집합에서 다음 낮은 값을 사용하여 종단간 최저 IP MTU를 설정하려고 시도합니다. 이러한 고원 레벨은 [RFC1191 - 경로 MTU 검색에](#) 설명되어 있습니다.

```
MTU plateaus from RFC 1191
```

```
- values include both TCP and IP headers
```

```
65535
32000
17914
8166
4352
2002
1492
1006
508
296
68
```

그러나 ICMP(Destination Unreachable - type 3; 프래그먼트화 필요 - 코드 4) 노드 R1이 수신한 메시지는 다음 값의 MTU를 전달한 다음 표시 시 노드 R1은 이 값을 사용합니다. 이 값은 512바이트입니다. 이 예에서는 TCP 세션 MSS 값을 조정합니다. 원래 TCP 세그먼트 길이는 1076바이트이므로 원래 TCP 세그먼트를 재전송하려면 3개의 패킷이 필요합니다.

R1 - PASSIVE - PMTUD 작업에서 볼 수 있듯이:

```
! - 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
```

앞에서 설명한 것처럼, 모든 패킷이 시간 경과에 따라 전송되면 PMTUD는 PmtuAger 타이머가 제어하는 반대 방향으로 고원 레벨을 통과하고, 상황에 따라 MSS를 최대 값으로 높이려고 시도합니다.

R1 - PMTUD에서 볼 수 있듯이 정의된 플랫폼 간:

```
! - 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

이러한 출력에서 최종 상태를 확인할 수 있습니다. 특히 노드 R1에서 표시한 최소 및 최대 MSS 값은 PMTUD가 트리거되었음을 강조 표시하고 알립니다.

## R4에 표시된 TCP 세션 세부 정보 - 활성:

! - 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 rcvnx: 2847700946 rcvwnd: 31799 rcvad: 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#

### R1에 표시된 TCP 세션 세부 정보 - 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  rcvadp: 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#

마지막으로, 지정된 시간에 ICMP(Destination Unreachable - type 3, 조각화 필요 - 코드 4) 메시지가 다시 수신되고 PMTUD가 이전에 설명한 대로 다시 작동합니다.

R1 - PASSIVE - PMTUD가 다시 트리거되었습니다.

```
! - 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
<b>PmtuAger</b>	<b>32</b>	<b>30</b>	<b>575401</b>
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#

Cisco 버그 ID CSCvf10395의 영향을 받는 Cisco IOS XR 릴리스에서는 ICMP 오류 메시지에 포함된 next-hop이 무시되고 노드는 이전에 언급되고 문서화된 IP MTU(well-defined plateau) 레벨에서 다음 낮은 값을 사용하여 엔드 투 엔드 IP MTU를 설정하려고 시도합니다 - [RFC111911911191 TU 검색](#). 이러한 시도는 전송이 성공할 때까지 발생하며, 이는 ICMP(Destination Unreachable - type 3, 프래그먼트화 필요 - 코드 4) 메시지가 더 이상 수신되지 않습니다.

Cisco IOS XR 릴리스가 Cisco 버그 ID CSCvf10395의 영향을 받는 노드에서 볼 수 있습니다.

```
! - 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
```

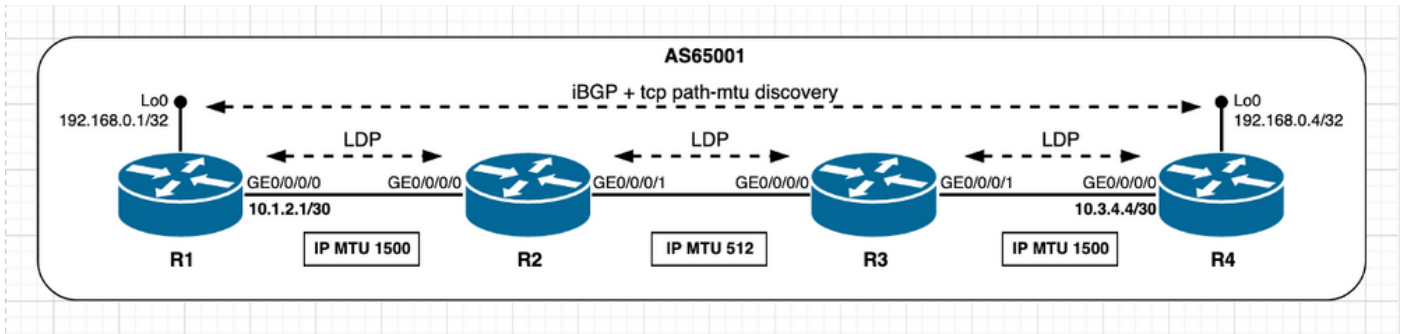
```
RP/0/0/CPU0:Feb 23 17:05:37.239 : tcp[399]: [t27] PCB 0x12152adc: Trying next lower MTU: 966
<<<<<<<< HERE: Plateau 1006
```

```
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
```

다음 단계에서는 동일한 시나리오를 고려하지만 모든 인터페이스에서 LDP(Label Distribution Protocol)를 사용합니다. 여기서 목표는 MPLS 지원 환경의 이전 시나리오와 어떤 차이가 관찰되는지 파악하는 것입니다.



이미지 3.3 - PMTUD가 활성화되고 경로 세그먼트는 IP MTU - MPLS 시나리오가 낮습니다.

먼저 PMTUD 트리거 전에 설정된 BGP 세션의 초기 단계를 여기에 표시된 대로 고려하십시오.

R4에 표시된 TCP(BGP) 초기 상태 - ACTIVE - MPLS 지원 시나리오:

- ! - 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 rcvad: 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#

## R1에 표시된 TCP(BGP) 초기 상태 - PASSIVE - MPLS 지원 시나리오:

! - 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 rcvadp: 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#

이 MPLS 지원 시나리오에서는 TCP(LDP) 세션에 대한 세부 정보가 설정된 것으로 관찰됩니다. 앞서 TCP(BGP) 세션의 MSS 계산에 대해 설명한 모든 내용은 TCP(LDP) 세션에도 적용됩니다. 예를 들어 노드 R3 및 R2 TCP(LDP) 세션 MSS 계산은 다음과 같이 요약할 수 있습니다.

- R2와 R3 모두 기본이 아닌 IP MTU를 512바이트로 사용합니다.
- 경로 MTU 검색이 사용하도록 설정되었습니다.
- TCP 피어가 직접 연결되지 않음(루프백 인터페이스 간에 TCP 세션이 설정됨). R3는 LDP 연결을 관리합니다. R3는 MSS가 472바이트인 SYN을 전송합니다. 512(인터페이스 IP MTU) - 20(minTCP\_H) - 20(minIP\_H) R2는 MSS가 472바이트인 SYN, ACK를 전송합니다. [Received MSS; 로컬 초기 MSS]. 수신된 MSS 472바이트; 로컬 초기 MSS 472바이트가 가장 낮은 MSS 값은 두 피어 모두에서 사용됩니다.

R3 - ACTIVE - MPLS 지원 시나리오에서 볼 수 있는 TCP(LDP) 세션 세부 정보:

```
! - 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 rcvadp: 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#

R2에 표시된 TCP(LDP) 세션 세부 정보 - PASSIVE - MPLS 지원 시나리오:

! - 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 rcvadp: 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#
```

BGP 세션이 설정되면 R1은 BGP Update 메시지를 보내고 ICMP 메시지(Destination Unreachable - type 3, 프래그먼트화 필요 - 노드 R1에서 TCP PMTUD를 트리거하는 노드 R2에서 소싱된 반환 코드 4). BGP 업데이트 메시지를 전달하는 IP 패킷에 DF 비트 세트가 있고 R2/R3 세그먼트에서 사용되는 IP MTU가 116바이트의 IP 패킷 크기보다 작기 때문입니다. 이전과 마찬가지로, 이 ICMP 메시지의 수신이 PMTUD를 트리거합니다. 이전 비 MPLS 시나리오와 비교할 때 MPLS 지원 시나리오의 차이점은 노드 R2 ICMP 메시지(Destination Unreachable - type 3; 조각화 필요 - 코드 4). 이 MPLS 지원 시나리오에서 **다음 hop 값의 MTU**는 추가 MPLS 오버헤드가 4바이트인 것을 고려합니다. 즉, 이 출력에 표시된 대로 R2의 이그레스 MPLS 레이블 스택을 고려합니다.

TCP 경로 MTU 검색 실행 중(R1 - PASSIVE - MPLS 사용 시나리오):

```
! - 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 set

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
```

## R1 - PASSIVE - TCP PMTUD 트리거됨 - MPLS 활성화 시나리오:

! - 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
<b>PmtuAger</b>	<b>1</b>	<b>0</b>	<b>164599</b>
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 rcvadiv: 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#

MPLS 활성화 시나리오에서 이그레스 MPLS 레이블 스택에 대한 노드 R2 ICMP 메시지 계정에 포함된 다음 **출의 MTU** 값에 유의하십시오. 이러한 측면을 더욱 강화하려면 다음 예를 고려하십시오. R2에서 필터링된 IP 패킷이 L3VPN 서비스와 연결된 경우 이더넷 프레임이 이제 두 개의 레이블 (IGP 레이블 및 VPN 레이블)을 전달함을 의미합니다. 그런 다음 **다음 출의 MTU**는 필요한 레이블 스택 크기를 반영합니다. 이러한 출력을 참조하십시오.

R1 - PASSIVE - L3 VPN 서비스 패킷에 표시된 대로:

! - 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

### R1 - PASSIVE - L3 VPN 서비스 - ICMP Type 3/Code 4:

```

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

```

```

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

```

```

Frame 2030: 190 bytes on wire (1520 bits), 190 bytes captured (1520 bits) on interface 0
Ethernet II, Src: fa:16:3e:5c:f1:80 (fa:16:3e:5c:f1:80), Dst: fa:16:3e:42:18:05
(fa:16:3e:42:18:05)

```

```

MultiProtocol Label Switching Header, Label: 24005, Exp: 0, S: 1, TTL: 251
  0000 0101 1101 1100 0101 .... .... .... = MPLS Label: 24005
  .... .... .... .... .... 000. .... .... = MPLS Experimental Bits: 0
  .... .... .... .... .... ...1 .... .... = MPLS Bottom Of Label Stack: 1
  .... .... .... .... .... .... 1111 1011 = MPLS TTL: 251

```

```

Internet Protocol Version 4, Src: 10.2.3.1, Dst: 10.1.14.1
0100 .... = Version: 4
.... 0101 = Header Length: 20 bytes (5)
Differentiated Services Field: 0x00 (DSCP: CS0, ECN: Not-ECT)
Total Length: 172
Identification: 0x002b (43)
Flags: 0x00
  0... .... = Reserved bit: Not set
  .0.. .... = Don't fragment: Not set
  ..0. .... = More fragments: Not set
Fragment offset: 0
Time to live: 253
Protocol: ICMP (1)
Header checksum: 0x9821 [validation disabled]
[Header checksum status: Unverified]
Source: 10.2.3.1
Destination: 10.1.14.1
[Source GeoIP: Unknown]
[Destination GeoIP: Unknown]

```

```

Internet Control Message Protocol
Type: 3 (Destination unreachable)

```

**Code: 4 (Fragmentation needed)**

Checksum: 0xbbac [correct]

[Checksum Status: Good]

Length: 17

[Length of original datagram: 68]

Unused: 0011

**MTU of next hop: 504**

Internet Protocol Version 4, Src: 10.1.14.1, Dst: 10.1.14.14

0100 .... = Version: 4

.... 0101 = Header Length: 20 bytes (5)

Differentiated Services Field: 0x00 (DSCP: CS0, ECN: Not-ECT)

Total Length: 610

Identification: 0x7c9f (31903)

Flags: 0x02 (Don't Fragment)

0... .... = Reserved bit: Not set

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

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

Fragment offset: 0

Time to live: 255

Protocol: TCP (6)

Header checksum: 0xcce5 [validation disabled]

[Header checksum status: Unverified]

Source: 10.1.14.1

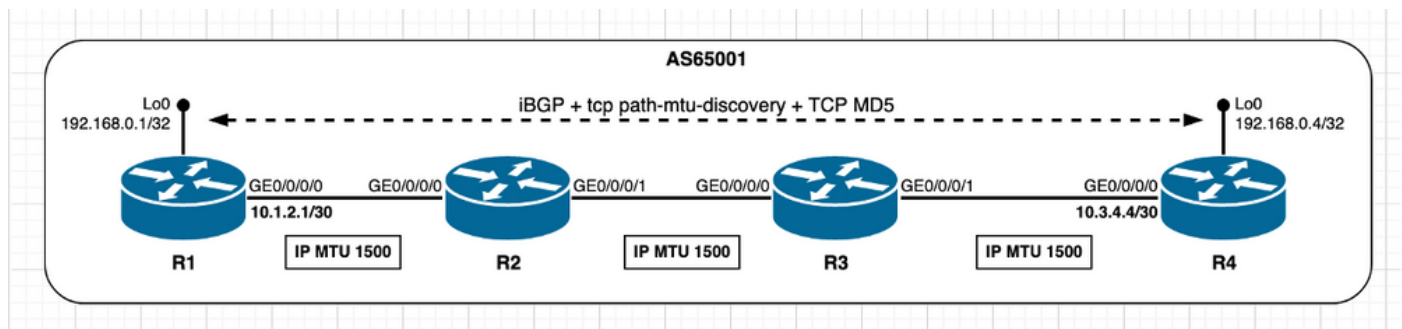
Destination: 10.1.14.14

[Source GeoIP: Unknown]

[Destination GeoIP: Unknown]

Transmission Control Protocol, Src Port: 22008, Dst Port: 23, Seq: 586828435, Ack: 754580617

## PMTUD - TCP 옵션(MD5)



이미지 3.4 - PMTUD 활성화 및 TCP MD5 인증

PMTUD 동작과 관련된 구분은 TCP MD5 인증이 활성화된 이전 시나리오에서 이미 설명한 것에서도 도입되지 않습니다. 이전에 사용 중인 TCP MD5 인증과 공유된 것처럼, Cisco IOS XR에서는 추가 오버헤드를 고려하며 활성 TCP 피어 초기 MSS 값은 동일하게 반영됩니다. TCP 옵션 사용의 영향에 대한 자세한 내용은 이전 섹션 **Use TCP Options - XR Active** and **Use TCP Options - XR Passive**를 참조하십시오. 이 시나리오의 TCP MSS 계산은 다음과 같이 요약할 수 있습니다.

- 모든 노드는 1500바이트의 기본 IP MTU를 사용합니다.
- TCP 경로 MTU 검색이 활성화되어 있습니다.
- TCP 피어가 직접 연결되지 않습니다.
- R1 및 R4 모두에서 TCP MD5 인증이 활성화되었습니다. R4는 BGP 연결을 관리합니다. R4는 MSS가 1436바이트인 SYN을 전송합니다.  $1500(\text{인터페이스 IP MTU}) - 20(\text{minTCP\_H}) - 20(\text{minIP\_H}) - 24\text{바이트}(\text{IOS XR TCP 옵션 오버헤드})$  R1은 SYN, ACK를 1436바이트의 MSS로 전송합니다. 예서는 [Received MSS; 로컬 초기 MSS]. 수신된 MSS 1436바이트; 로컬 초기 MSS 1460바이트가 가장 낮은 MSS 값은 두 피어 모두에서 사용됩니다.

R4에서 제공된 TCP SYN:

! - 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, R1에서 제공된 ACK:

! - 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)

## R4에 표시된 TCP 세션 세부 정보 - 활성:

! - 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#

## R1에 표시된 TCP 세션 세부 정보 - 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:

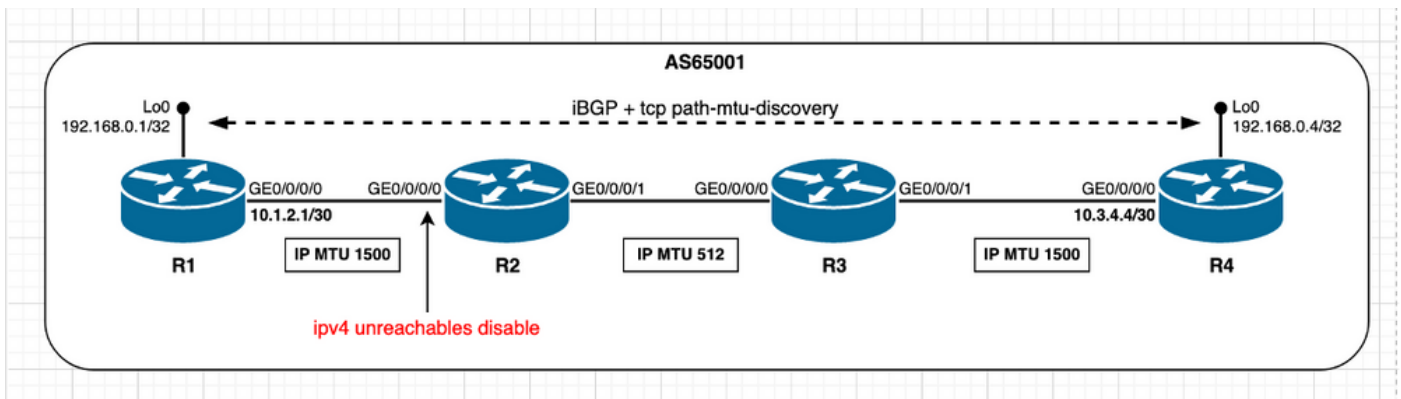
```

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

## PMTUD - 블랙홀 탐지

앞서 PMTUD 섹션에서 설명한 대로 **Path Segment has Lower IP MTU**, ICMP(Destination Unreachable - type 3; 조각화 필요 - 코드 4) 메시지입니다. PMTUD가 트리거되지 않는 어떤 이유로 이러한 메시지가 수신되지 않을 수 있습니다. 이 경우 TCP 피어 간 경로의 가장 낮은 IP MTU는 학습되지 않습니다. 이러한 시나리오에서는 IP 패킷에 DF 비트가 설정되어 있고 가장 낮은 IP MTU 경로 세그먼트보다 크기가 큰 경우 잠재적인 블랙홀이 발생할 수 있습니다. 이러한 패킷은 자동으로 삭제됩니다.

이 섹션에서는 Cisco IOS XR에서 이러한 잠재적 블랙홀 시나리오를 탐지하고 어떻게 행동하는지 중점적으로 살펴봅니다. 이 용도로 IPv4 연결 불가능 기능은 다음 이미지 및 CLI 출력에 표시된 대로 R2 인터페이스 GE0/0/0/0에서 비활성화됩니다.



이미지 3.5 - R1/R4 및 R2 IPv4에서 활성화된 PMTUD를 사용할 수 없습니다.

R2에서 비활성화된 IPv4 연결 불가능:

```

!- R2 - IP unreachable 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 unreachable 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 unreachable are never sent
ICMP mask replies are never sent
Table Id is 0xe0000000
```

이 블랙홀 시나리오에서 Cisco IOS XR이 처리하는 방법은 동일한 패킷을 두 번 재전송하는 것이며, 그래도 실패하면 예상 TCP ACK가 수신되지 않는 것입니다. 다시 시도한 후 [RFC1191 - 경로 MTU 검색에](#) 설명된 대로 다음으로 낮은 잘 정의된 고원 값을 사용합니다([경로 세그먼트가 낮은 IP MTU](#) 섹션 참조). 요약하자면, Cisco IOS XR에서는 패킷의 크기가 패킷 재전송을 통해 패킷을 복구하려고 시도하여 패킷이 목적지로 향하는 경로 내에 있는 어느 곳에 삭제될 수 있다고 가정합니다. 이 동작은 노드 R1 인터페이스에서 가져온 패킷 캡처와 디버그 tcp pmtud의 출력에서 다음 예에서 볼 수 있습니다.

R1에서 IOS-XR 블랙홀 탐지:

```
! - 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