مدیریت MTU در GRE Tunnel

یکی از مشکلاتی که بعد از ایجاد تونل GRE و یا انواع دیگر تونل ها مانند تونل IPSec بین دو سایت (در سناریوهای Site-to-Site VPN) و یا بین کاربر با سایت (در سناریوهای Remote Access VPN) ایجاد می شود، مشکل MTU است. این مشکل بدین صورت دیده می شود که ارسال ترافیک های کوچک مانند ping و telnet بین دو طرف ارتباط تونل امکان پذیر است اما نمی توان ترافیک های بزرگ مانند http و یا ftp جابجا نمود. در این بخش قصد داریم راه حل های موجود برای رفع مشکل MTU در تونل GRE را مورد بررسی قرار می دهیم. البته مشابه این راه حل ها در بقیه روش های Tunneling نیز قابل پیاده سازی است.

قبل از هر چیز لازم است تا تفاوت بین واژه های TCP MSS، IP MTU و Interface MTU را بشناسیم. در ذیل تعریفی کوتاه نسبت به هر یک ارائه می شود.

Interface MTU: هر اینترفیسی پارامتری به نام Interface MTU دارد که حداکثر سایز مجاز بسته ارسالی روی آن اینترفیس را تعیین می کند. این سایز شامل سربار لایه ۲ نمی شود. در ذیل Interface MTU بعضی از اینترفیس های معروف آمده است.

Ethernet: ۱۵۰۰

Serial: ۱۵۰۰

Token Ring: ۴۴۶۴

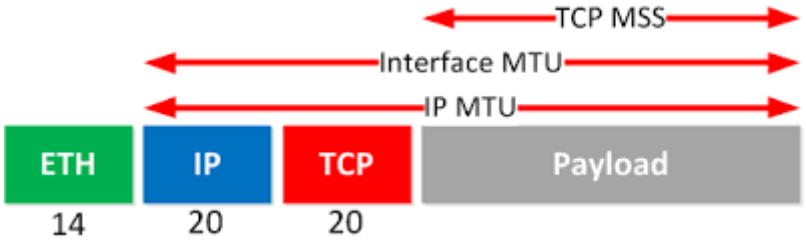
ATM: ۴۴۷۰

FDDI: ۴۴۷۰

- IP MTU: این پارامتر حداکثر سایز بسته های IP را تعیین می کند. در ترافیک های IP بین Inteface MTU و IP MTU تفاوتی دیده نمی شود. در صورتی که مقدار Interface MTU و IP MTU یکسان نباشند، مقدار کوچکتر در ارسال ترافیک IP مد نظر قرار می گیرد. همانطور که قبلا مشاهده نمودیم برای تغییر IP MTU از دستور ip mtu در محیط اینترفیس استفاده می کنیم.
- TCP MSS : TCP MSS حداکثر سایز داده را نشان می دهد که توسط پروتکل TCP قابل بسته بندی است. این سایز بدون احتساب سربار لایه۲، لایه ۳ و لایه ۴ است. در پروتکل TCP و در پروسه three-way handshaking هر یک از طرفین حداکثر سایز MSS مجاز را به طرف مقابل اعلام می کند. بنابراین هر یک با توجه به MSS اعلام شده توسط طرف مقابل، داده های دریافتی از لایه application را به بخش های کوچکتر تقسیم می کند. نکته دیگر اینکه اگر

فرض کنیم حداکثر سایز مجاز بسته IP، 1500 بایت است (که در اکثر اینترفیس ها این چنین است) و با توجه به اینکه سربار IP و TCP هر کدام ۲۰ بایت و در مجموع ۴۰ بایت است، بنابراین حداکثر مقدار مجاز MSS، 1460 بایت خواهد بود. دستور تغییر TCP MSS را در ادامه شرح خواهیم داد.

شکل زیر تفاوت واژه های Interface MTU، IP MTU و TCP MSS را دقیق تر بیان می کند.



با فرض اینکه MTU اینترفیس فیزیکی که تونل روی آن سوار شده است ۱۵۰۰ بایت است، بعد از ایجاد تونل GRE مقدار MTU در اینترفیس Tunnel را به ۱۴۶۴ کاهش می دهیم (به اندازه حداکثر سایز سربار ناشی از تونل GRE از ۱۵۰۰ کسر می نماییم) تا بعد از اضافه شدن سربار GRE، سایز بسته ارسالی به ۱۵۰۰ برسد که قابل ارسال توسط اینترفیس فیزیکی باشد.

پیکربندی تغییر IP MTU در اینترفیس تونل:

```
interface tunnel0
 ip mtu 1464
```

بعد از این تعریف اولیه، مشکل ارسال ترافیک های ftp و http را در تونل های GRE در چه می بینید؟ اگر فرض کنیم، اندازه بسته های ترافیک ارسالی http و ftp حداکثر سایز مجاز باشد، آنگاه با اضافه کردن سربار GRE و سربار IP بیرونی حداقل به مقدار ۲۴ بایت (شامل ۴ بایت GRE و ۲۰ بایت IP) و حداکثر به مقدار ۳۶ بایت (شامل ۱۶ بایت GRE و ۲۰ بایت IP) به سایز بسته IP اضافه می شود و از آنجایی که اینترفیس های شبکه قابلیت انتقال ترافیک های بیش از ۱۵۰۰ بایت را ندارند، لذا ترافیک در ارسال دچار مشکل می شود.

حتما این واقعیت در ذهن شما شکل می گیرد که پروتکل IP در صورتی که سایز بسته ارسالی بیش از MTU اینترفیس باشد، ترافیک را Fragment (بخش بندی) و سپس ارسال می کند. با توجه به اینکه پیش تر مقدار MTU اینترفیس Tunnel به ۱۴۶۴ کاهش داده شده است، ترافیک های با سایز بیش از ۱۴۶۴ وقتی وارد اینترفیس Tunnel می شوند،

قائدتا باید به بسته های کوچکتر بخش بندی شده و سپس توسط اینترفیس Tunnel بسته بندی شده و تحویل اینترفیس فیزیکی شوند که در این صورت نباید مشکلی در ارسال وجود داشته باشد.

در پاسخ باید گفت در هدر TCP، Flag ای به نام DF یا DONOT FRAGMENT وجود دارد که خیلی از application ها با ست کردن این Flag اجازه fragmentation را به پروتکل IP نمی دهند و در چنین شرایطی است که ارسال ترافیک های بزرگتر از سایز MTU با مشکل مواجه می گردد. در این صورت روتر ابتدای تونل که امکان ارسال ترافیک بیش از ۱۴۶۴ را ندارد با ارسال پیغام ICMP با Type شماره ۳ و code شماره ۴ هم مبدا را از این رخداد آگاه می کند و هم آنکه حداکثر سایز MTU مجاز را به مبدا اعلام می کند و مبدا نیز سایز بسته را متناسب با آن کاهش می دهد. نمونه این پیغام در ذیل آمده است.

ICMP: dst (10.10.10.10) frag. needed and DF set unreachable sent to 10.1.3.4

اما اگر پیغام ICMP در طول مسیر block شود و به مبدا نرسد، مبدا سایز بسته ارسالی را تغییر نمی دهد و مشکل برطرف نخواهد شد.

برای درک بهتر توضیحات فوق در ذیل با توجه به سناریوی قبل از روتر Client ترافیک ICMP که در آن فیلد DF ست شده است با سایز از ۱۴۶۰ تا ۱۴۶۶ به مقصد سرور ارسال شده است. همانطور که مشاهده می کنید، ترافیک های ارسالی با سایز از ۱۴۶۰ تا ۱۴۶۴ بدون مشکل به مقصد رسیده اند اما به محض آنکه سایز ترافیک ارسالی ۱۴۶۵ و یا بالاتر می شود، بسته ارسالی با مشکل مواجه شده و به مقصد نمی رسد. روتر ابتدای تونل با ارسال پیغام ICMP حداکثر سایز مجاز بسته ارسالی را نیز اعلام می کند.

پیکربندی حداکثر سایز مجاز بسته ارسالی در تونل GRE:

```
CLIENT#ping
 Protocol [ip]:
Target IP address: 192.168.2.2
Repeat count [5]: 1
Datagram size [100]:
Timeout in seconds [2]:
Extended commands [n]: y
Source address or interface:
Type of service [0]:
```

Set DF bit in IP header? [no]: yes

Validate reply data? [no]:

Data pattern [0xABCD]:

Loose, Strict, Record, Timestamp, Verbose[none]: Verbose

Loose, Strict, Record, Timestamp, Verbose[V]:

Sweep range of sizes [n]: y

Sweep min size [36]: 1460

Sweep max size [18024]: 1466

Sweep interval [1]:

Type escape sequence to abort.

Sending 7, [1460..1466]-byte ICMP Echos to 192.168.2.2, timeout is 2 seconds:

Packet sent with the DF bit set

Reply to request 0 (48 ms) (size 1460)

Reply to request 1 (44 ms) (size 1461)

Reply to request 2 (48 ms) (size 1462)

Reply to request 3 (40 ms) (size 1463)

Reply to request 4 (40 ms) (size 1464)

Unreachable from 192.168.1.1, maximum MTU 1464 (size 1465)

Request 6 timed out (size 1466)

Success rate is 71 percent (5/7), round-trip min/avg/max = 40/44/48 ms

چه راه حل هایی برای رفع مشکل وجود دارد. در ذیل لیست بعضی از راه حل ها و روش پیاده سازی آن نشان داده شده است.

۱ – سایز MTU را روی کارت شبکه همه client ها و همچنین سرور به ۱۴۶۴ کاهش دهید (به اندازه ۳۶ بایت حداکثر سربار ناشی از تونل GRE از سایز مجاز MTU کسر نمایید). این راه حل عملا امکان پذیر نیست.

۲– در همه مسیر سایز MTU را از ۱۵۰۰ بایت به ۱۵۳۶ افزایش دهیم. از آنجایی که بستر ارتباطی در اختیار ما نیست، این راه حل هم نیز عملا امکان پذیر نخواهد بود.

در ادامه بعضی از **راه حل های شدنی** را مورد بررسی قرار می دهیم.

۱- فیلد DF ترافیک های دریافتی در مبدا تونل را قبل از ارسال روی تونل به ۰ تغییر دهید و یا به عبارت دیگر clear نمایید تا ترافیک های بزرگتر از سایز MTU تعریف شده در اینترفیس تونل، قابل بخش بندی باشند. برای پیاده سازی این روش از ابزار PBR استفاده می شود.

پیکربندی۸-۴حل مشکل MTU در GRE با Clear کردن فیلد DF بسته های دریافتی:

```
SITE1(config)#route-map CLEAR-DF permit 10
SITE1(config-route-map)#match length ?
  <0-2147483647>  Minimum packet length
!
SITE1(config-route-map)#match length 1465 ?
  <0-2147483647>  Maximum packet length
SITE1(config-route-map)#match length 1465 1500
!
route-map CLEAR-DF permit 10
 match length 1465 1500
 set ip df 0
!
interface Ethernet0/0
 ip policy route-map CLEAR-DF
```

پیکربندی بررسی حداکثر سایز مجاز بسته ارسالی در تونل GRE بعد از clear کردن فیلد DF:

```
CLIENT#ping
Protocol [ip]:
Target IP address: 192.168.2.2
```

Repeat count [5]: 1

Datagram size [100]:

Timeout in seconds [2]:

Extended commands [n]: y

Source address or interface:

Type of service [0]:

Set DF bit in IP header? [no]: yes

Validate reply data? [no]:

Data pattern [0xABCD]:

Loose, Strict, Record, Timestamp, Verbose[none]: V

Loose, Strict, Record, Timestamp, Verbose[V]:

Sweep range of sizes [n]: y

Sweep min size [36]: 1460

Sweep max size [18024]: 1500

Sweep interval [1]:

Type escape sequence to abort.

Sending 41, [1460..1500]-byte ICMP Echos to 192.168.2.2, timeout is 2 seconds:

Packet sent with the DF bit set

Reply to request 0 (44 ms) (size 1460)

Reply to request 1 (48 ms) (size 1461)

Reply to request 2 (44 ms) (size 1462)

Reply to request 3 (44 ms) (size 1463)

Reply to request 4 (44 ms) (size 1464)

Reply to request 5 (60 ms) (size 1465)

Reply to request 6 (52 ms) (size 1466)

Reply to request 7 (52 ms) (size 1467)

Reply to request 8 (48 ms) (size 1468)

Reply to request 9 (52 ms) (size 1469)

Reply to request 10 (52 ms) (size 1470)

Reply to request 11 (56 ms) (size 1471)

Reply to request 12 (52 ms) (size 1472)

Reply to request 13 (60 ms) (size 1473)

Reply to request 14 (56 ms) (size 1474)

Reply to request 15 (52 ms) (size 1475)

Reply to request 16 (52 ms) (size 1476)

Reply to request 17 (52 ms) (size 1477)

Reply to request 18 (56 ms) (size 1478)

Reply to request 19 (56 ms) (size 1479)

Reply to request 20 (48 ms) (size 1480)

Reply to request 21 (52 ms) (size 1481)

Reply to request 22 (52 ms) (size 1482)

Reply to request 23 (60 ms) (size 1483)

Reply to request 24 (60 ms) (size 1484)

Reply to request 25 (52 ms) (size 1485)

Reply to request 26 (60 ms) (size 1486)

Reply to request 27 (56 ms) (size 1487)

Reply to request 28 (56 ms) (size 1488)

Reply to request 29 (56 ms) (size 1489)

Reply to request 30 (52 ms) (size 1490)

Reply to request 31 (60 ms) (size 1491)

Reply to request 32 (56 ms) (size 1492)

Reply to request 33 (52 ms) (size 1493)

Reply to request 34 (52 ms) (size 1494)

Reply to request 35 (60 ms) (size 1495)

Reply to request 36 (44 ms) (size 1496)

Reply to request 37 (52 ms) (size 1497)

Reply to request 38 (56 ms) (size 1498)

Reply to request 39 (56 ms) (size 1499)

Reply to request 40 (52 ms) (size 1500)

Success rate is 100 percent (41/41), round-trip min/avg/max = 44/53/60 ms

– با ایجاد دستور ip tcp adjust-mss 1424 در اینترفیس Tunnel مقدار فیلد MSS بسته SYN در ترافیک TCP از مقدار ۱۴۶۰ به اندازه ۳۶ بایت، که حداکثر سایز ناشی از تونل GRE است، کسر می شود. به عبارت دیگر مقدار MSS به ۱۴۲۴ کاهش می یابد تا پروتکل TCP در هر یک از طرفین داده را به بخش های کوچکتری تقسیم نماید تا نهایتا با اضافه شدن سربار GRE سایز بسته ارسالی به ۱۵۰۰ بایت برسد که در ارسال روی اینترفیس فیزیکی با مشکلی مواجه نشود. بدیهی است که این روش فقط روی ترافیک های TCP تاثیر می گذارد و در سناریوی قبل که در آن ترافیک ارسالی از نوع ICMP است، در صورت ارسال ترافیک های با سایز بزرگتر از ۱۴۶۴ با مشکل مواجه می شود.

پیکربندی تغییر مقدار MSS از ۱۴۶۰ به ۱۴۲۴ برای حل مشل MTU در تونل GRE:

interface Tunnel0

ip tcp adjust-mss 1424

!

!

!

CLIENT#ping 192.168.2.2 size 1490 df-bit

Type escape sequence to abort.

Sending 5, 1490-byte ICMP Echos to 192.168.2.2, timeout is 2 seconds:

Packet sent with the DF bit set

M.M.M

Success rate is 0 percent (0/5)

!

*Jan 23 21:29:09.107: ICMP: dst (192.168.1.2) frag. needed and DF set unreachable rcv from 192.168.1.1.M

– در شیوه سوم مقدار IP MTU اینترفیس Tunnel را از ۱۴۶۴ به ۱۵۰۰ افزایش دهید. بنابراین ترافیک های دریافتی روی اینترفیس Tunnel که حداکثر سایز آنها ۱۵۰۰ بایت است نیاز به بخش بندی نخواهند داشت. بعد از اضافه کردن سربار GRE سایز بسته به ۱۵۳۶ بایت افزایش می یابد اما از آنجایی که فیلد DF از هدر بسته IP اصلی به هدر IP بیرونی، که مربوط به تونل GRE است، کپی نمی شود، لذا اینترفیس فیزیکی به راحتی می تواند ترافیک های با سایز بیش از ۱۵۰۰ بایت را بخش بندی نموده و سپس ارسال نماید. مشکل این روش در این است که روتر طرف مقابل قبل از حذف سربار GRE باید ترافیک های GRE دریافتی را سرهم[۱] نماید. از آنجایی که سرهم کردن ترافیک های بخش بندی شده GRE توسط CPU انجام می شود، کارایی ترافیک های GRE را بسیار کاهش می دهد. اضافه کردن این نکته نیز حائز اهیمت است که به صورت معمول ترافیک های fragment شده در مقصد نهایی سرهم بندی می شود اما در صورت استفاده از این روش ترافیک های fragment شده در روتر که انتهای تونل است، سرهم بندی انجام می پذیرد. این بدان دلیل است که مقصد ترافیک های GRE آدرس روتر سایت مقصد است و نه کاربر نهایی. بدین ترتیب روتر درگیر سرهم کردن ترافیک های دریافتی خواهد شد و سربار آن به مراتب کاهش می یابد.

پیکربندی تغییر IP MTU از ۱۴۶۴ به ۱۵۰۰ در اینترفیس تونل به عنوان یکی از راه حل های رفع مشکل MTU در GRE:

interface tunnel 0

ip mtu 1500

!

!

CLIENT#ping 192.168.2.2 size 1500 df-bit

Type escape sequence to abort.

Sending 5, 1500-byte ICMP Echos to 192.168.2.2, timeout is 2 seconds:

Packet sent with the DF bit set

!!!!!

Success rate is 100 percent (5/5), round-trip min/avg/max = 44/54/60 ms

!

CLIENT#ping 192.168.2.2 size 1501 df-bit

Type escape sequence to abort.

Sending 5, 1501-byte ICMP Echos to 192.168.2.2, timeout is 2 seconds:

Packet sent with the DF bit set

.....

Success rate is 0 percent (0/5)