VPN技术中的MTU问题

万欣 2008-11-25 发表

VPN技术中的MTU问题

【问题简述】

在VPN技术中经常遇到隧道已经建立但是某些应用程序无法正常通讯的问题，一般所来是由于路径MTU不匹配的缘故，下面就其中的数据传输流程做一些分析，以说明遇到出现此类问题该如何解决。

【问题分析】

先从一个实验说起

实验拓扑如下：

PC1-------VPN网关1————————VPN网关2--------PC2

在上述组网中，VPN网关1和2之间建立了一条GRE隧道，PC1与PC2通过该GRE隧道互访。

这时，我们会发现，PC1在ping 1448的报文时，在PC2上抓包发现没有被分片，而ping 1449时，PC2上会发现PC1过来的报文已经被分片了。为什么呢?

PC1出来的IP报文长度为1448＋8（ICMP报头长）＋20（IP报头长），到达VPN网关1，VPN网关1发到GRE隧道口，封装GRE头（4字节），再加上外层IP头，到达VPN网关外层以太口，这时IP报文的长度已经变为：1448＋8＋20＋4＋20＝1500字节，刚好等于以太口的MTU，于是被顺利传送。而ping 1449时，到达外层以太口为1501字节，超出了1500的MTU，又因为报文DF位未置1，即可以分片，VPN网关于是将该报文分片发送出去。这就是我们所看到的现象。

在上述实验中，由于应用程序是ping，所以报文可以被分片，因此互通没有问题。但是如果是WEB访问等应用，则有些报文是不允许分片的，这样在外层以太口就会将超过1500的报文丢掉，导致无法通讯。

从上述实验可以看到，由于VPN会额外加入一些报文头，如果通讯双方的MTU不能随之改变的话，就容易产生不通的问题。

下面以HTTP为例，说明为何产生此问题并如何解决。

先看看HTTP为何无法像ICMP那样自动分片通讯。

假设PC1/2建立了HTTP连接后，PC2希望从PC1下载一个大的网页。PC2开始发送，其IP的DF位置1，不允许分片，IP报文长度为1500字节。到达VPN网关1的外网口后，VPN网关1发现其长度超过了1500个字节，于是将其丢弃，并给PC1发回一个目的地址不可达的ICMP信息，出错代码为”Fragmentation needed”，表示需要分片，但不允许分片，同时给出”MTU of next hop: 1500”。PC1接收到该消息后，又按照1500字节对外发送，又被丢弃，于是就形成了循环，无法通讯。

根据上述的分析，很容易得到如下解决方式，在VPN网关1的出接口设置MTU为1500 - 4 - 20＝1476，这样VPN网关1返回ICMP不可达消息时将给出”MTU of next hop: 1476”。PC2将以1476作为自己的最大MTU对外发送，到达VPN网关1，封装GRE和外层IP头后就不会超过1500而顺利发到对端。

这时仅解决了下载的问题，如果PC2需要将大文件上传到PC1，同样需要设置VPN网关2的出接口MTU值小于1476。

当然，还可以更改VPN网关1的出接口的TCP MSS数值，将其更改为1500-4-20-20（TCP头）=1456字节，也可保证HTTP等TCP应用顺利通过。但该情况仅适用于TCP应用。

上述解决方式同样适用于其他隧道技术，在L2TP、IPSEC等应用时可以相应的根据其包头数值设置MTU或MSS。