

IP分片原理及分析

在应用程序中，必须关心IP数据的长度。如果它超过网络的MTU，那么就要对IP数据报进行分片。如果需要，源目的端之间的每个网络都要进行分片，并不只是发送端主机连接第一个网络才这样做。在此文中，将讨论IP分片的一些原理及其在分片中所需要的几个元素（此文只讨论MTU最大传输单位和MSS最大数据分段），本文对以太网的例子作了初略的分解。

关键字：IP分片，MTU，MSS

引言

分片是分组交换的思想体现，也是IP协议解决的两个主要问题之一。在IP协议中的分片算法主要解决异种网最大传输单元(MTU)的不同.但是分组在传输过程中不断地分片和重组会带来很大的工作量还会增加一些不安全的因素。

正文：

一、什么是IP分片

IP分片是网络上传输IP报文的一种技术手段。IP协议在传输数据包时，将数据报文分为若干分片进行传输，并在目标系统中进行重组。这一过程称为分片（fragmentation）。

二、为什么要进行IP分片

通常要传输的IP报文的大小超过最大传输单位MTU(Maximum Transmission Unit)时就会产生IP分片情况。IP分片通常发生在网络环境中。比如说，在以太网（Ethernet）环境中可传输最大IP报文大小

（MTU）为1500字节。而传输的报文大小要比1500字节(不包括以太协议的首部和尾部18个字节)大，这个时候就需要利用到分片技术，经分片后才能传输此报文。另外，使用UDP很容易导致IP分片，而很难强迫TCP发送一个需要进行分片的报文。

三、IP分片原理及分析

分片和重新组装的过程对传输层是透明的，其原因是当IP数据报进行分片之后，只有当它到达下一站时，才可进行重新组装，且它是由目的端的IP层来完成的。分片之后的数据报根据需要也可以再次进行分片。

IP分片和完整IP报文差不多拥有相同的IP头，ID域对于每个分片都是一致的，这样才能在重新组装的时候识别出来自同一个IP报文的分片。在IP头里面，16位识别号唯一记录了一个IP包的ID（ipid），具有同一个ID的IP分片将会重新组装；而13位片偏移则记录了某IP片相对整个包的位置；而这两个表中间的3位标志则标志着该分片后面是否还有新的分片。这三个域就组成了IP分片的所有信息，接受方就可以利用这些信息对IP数据进行重新组织。

1、标志字段的作用

标志字段在分片数据报中起了很大作用，在数据报分片时把它的值复制到每片中。标志字段的其中一个比特称作“不分片”位，用其中一个比特来表示“更多的片”。除了最后一片外，其他每个组成数据报的片都要把该比特置1。片偏移字段指的是该片偏移原始数据报开始处的位置。另外，当数据报被分片后，每个片的总长度值要改为该片的长度值。如果将标志字段的“不分片”比特置1，则IP将不对数据报进行分片。相反把数据报丢弃并发送一个ICMP差错报文并通知源主机废弃的原因。如果不是特殊需要，则不应该置1；最右比特置1表示该报文不是最后一个IP分片。

故意发送部分IP分片而不是全部，则会导致目标主机总是等待分片消耗并占用系统资源。某些分片风暴攻击就是这种原理。

这里以以太网为例，由于以太网传输电气方面的限制，每个以太网帧都有最小的大小64bytes最大不能超过1518bytes，抛去以太网帧的帧头(DMAC目的MAC地址48bit=6Bytes+SMAC源MAC地址48bit=6Bytes+Type域2bytes)14Bytes和帧尾CRC校验部分4Bytes，那么剩下承载上层协议的地方也就是Data域最大就只能是1500Bytes，这就是前面所说的MTU的值。这个也是网络层协议非常关心的地方，因为网络层的IP协议会根据这个值来决定是否把上层传达下来的数据进行分片。就好比一个盒子没法装下一大块面包，我们需要把面包切成片，装在多个盒子里面一样的道理。

2、MTU原理

当两台远程PC互联的时候，它们的数据需要穿过很多的路由器和各种各样的网络媒介才能到达对端，网络中不同媒介的MTU各不相同，就好比一长段的水管，由不同粗细的水管组成(MTU不同)通过这段水管最大水量就要由中间最细的水管决定。

对于网络层的上层协议而言(这里以TCP/IP协议族为例)它们对水管粗细不在意它们认为这个是网络层的事情。网络层IP协议会检查每个从上层

协议下来的数据包的大小，并根据本机MTU的大小决定是否作“分片”处理。分片最大的坏处就是降低了传输性能，本来一次可以搞定的事情，分成多次搞定，所以在网络层更高一层(就是传输层)的实现中往往会对此加以注意!有些高层因为某些原因就会要求我这个面包不能切片，我要完整地面包，所以会在IP数据包包头里面加上一个标签:DF(Donot Fragment)。这样当这个IP数据包在一大段网络(水管里面)传输的时候，如果遇到MTU小于IP数据包的情况，转发设备就会根据要求丢弃这个数据包。然后返回一个错误信息给发送者。这样往往会造成某些通讯上的问题，不过幸运的是大部分网络链路MTU都是1500或者大于1500。

对于UDP协议而言，这个协议本身是无连接的协议，对数据包的到达顺序以及是否正确到达不甚关心，所以一般UDP应用对分片没有特殊要求。

对于TCP协议而言就不一样了，这个协议是面向连接的协议，对于TCP协议而言它非常在意数据包的到达顺序以及是否传输中有错误发生。所以有些TCP应用对分片有要求---不能分片(DF)。

3、MSS的原理

MSS就是TCP数据包每次能够传输的最大数据分段。为了达到最佳的传输效能TCP协议在建立连接的时候通常要协商双方的MSS值，这个值TCP协议在实现的时候往往用MTU值代替(需要减去IP数据包包头的大小20Bytes和TCP数据段的包头20Bytes)所以往往MSS为1460。通讯双方会根据双方提供的MSS值得最小值确定为这次连接的最大MSS值。

当IP数据报被分片后，每一片都成为一个分组，具有自己的IP首部，并在选择路由时与其他分组独立。这样，当数据报的这些片到达目的端时有可能会失序，但是在IP首部中有足够的信息让接收端能正确组装这些数据报片。

尽管IP分片过程看起来是透明的，但有一点让人不想使用它：即使只丢失一片数据也要重传整个数据报。因为IP层本身没有超时重传的机制——由更高层来负责超时和重传（TCP有超时和重传机制，但UDP没有。一些UDP应用程序本身也执行超时和重传）。当来自TCP报文段的某一片丢失后，TCP在超时后会重发整个TCP报文段，该报文段对应于一份IP数据报。没有办法只重传数据报中的一个数据报片。事实上，如果对数据报分片的是中间路由器，而不是起始端系统，那么起始端系统就无法知道数据报是如何被分片的。就这个原因，经常要避免分片。

五、IP分片算法的原理

分片重组是IP层一个最重要的工作，其处理的主要思想：当数据包从一个网络A进入另一个网络B时，若原网络的数据包大于另一个网络的最大数据包的长度，必须进行分片。因而在IP数据包的报头有若干标识域注明分片包的共同标识号、分片的偏移量、是否最后一片及是否允许分片。传输途中的网关利用这些标识域进行分片，目有主机把收到的分片进行重组以恢重数据。因此，分片包在经过网络监测设备、安全设备、系统管理设备时，为了获取信息、处理数据，都必须完成数据包的分片或重组。

Identification R DF MF Fragment Offset

R：保留未用；DF：Don't Fragment,“不分片”位，如果将这一比特置1，IP层将不对数据报进行分片；MF：More Fragment,“更多的片”，除了最后一片外，其它每个组成数据报的片都要把比特置1；Fragment Offset：该片偏移原始数据包开始处的位置。偏移的字节数是该值乘以8。

六、总结

此论文阐述了什么是IP分片，并举例说明了IP的分片现像及其特点。简单论述了IP分片的原理。并加以重点分析了MTU，阐述了MTU在IP分片技术中所起的作用，简略讨论了ICMP不可达差错，利用UDP观察路径MTU发现过程。简述了MSS的作用以及其值是怎样得来的。本文只是对IP分片作了一个小小的论述。