IP分片原理及分析 - 李忠阳 - 开源中国社区

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

每 一种物理网络都会规定链路层数据帧的最大长度，称为链路层MTU(Maximum Transmission Unit).IP协议在传输数据包时，若IP数据报加上数据帧头部后长度大于MTU，则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。比如 说，在以太网环境中可传输最大IP报文大小（MTU）为1500字节。如果要传输的数据帧大小超过1500字节，即IP数据报长度大于 1472(1500-20-8=1472，普通数据报)字节，则需要分片之后进行传输。

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，只有当它到达目的站时，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，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。这三个标志就组成了IP分片的所有信息(将在后面介绍)，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 片中的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。如果将标志字段的比特置1，则IP将不对数据报 进行分片，若在某个中间路由器上需要对其分片，则仅仅把数据报丢弃并发送一个ICMP不可达差错报文给源主机。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域最大就只能有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。

下面是标志位在IP首部中的格式以及各个标志的意义：

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。

2、MTU原理

当两台远程PC需要通信的时候，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(Don‘t 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(仅X.25网络的 576和点对点网络的296小于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并不关心，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(Maxmum Sigmentation Size)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20字节和TCP数据段的包头20字节)所以往往MSS为1460。通讯双方会根据双方提 供的MSS值的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，经常需要避免分片。

四、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或者接口的MTU长度，则 需要进行分片(若设置DF为1，则丢弃，并回送ICMP不可达差错报文)。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可能的再行分片，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。

五、IP分片的安全问题

IP 分片是在网络上传输IP报文时常采用的一种技术，但是其中存在一些安全隐患。Ping of Death, teardrop等攻击可能导致某些系统在重组IP分片的过程中宕机或者重新启动。一些IP分片攻击除了用于进行拒绝服务攻击之外，还常用于躲避防火墙或 者网络入侵检测系统的一种手段。部分路由器或者基于网络的入侵检测系统（NIDS），由于IP分片重组能力的欠缺，导致无法进行正常的过滤或者检测。

介绍一下Tiny fragment 攻击：

所 谓Tiny fragment攻击是指通过恶意操作，发送极小的分片来绕过包过滤系统或者入侵检测系统的一种攻击手段。攻击者通过恶意操作，可将TCP报头(通常为 20字节)分布在2个分片中，这样一来，目的端口号可以包含

在第二个分片中。对于包过滤设备或者入侵检测系统来说，首先通过判断目的端口号来采取允许/禁 止措施。但是由于通过恶意分片使目的端口号位于第二个分片中，因此包过滤设备通过判断第一个分片,决定后续的分片是否允许通过。但是这些分片在目标主机上 进行重组之后将形成各种攻击。通过这种方法可以迂回一些入侵检测系统及一些安全过滤系统。目前一些智能的包过滤设备直接丢掉报头中未包含端口信息的分片。

六、总结

本 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，简述了MSS的作用以及其值是怎样得来的，简略讨论了IP分片所引起的安全问题。本文只是对IP分片作了一个小小的论述。