我们已经学习了ARP协议，它可以让我们通过IP地址获得对应的MAC地址。

除了ARP协议，OSI第3层还有其他的协议，本课我们要学习的就是很有用的 ICMP协议 。

其实，我们到目前为止已经不止一次使用过这个协议了，在我们之前用ping命令或traceroute命令（在Windows中是tracert命令）的时候，其实都使用了ICMP协议。

借助ICMP我们往往可以大概清楚网络发生了什么错误，因此我称之为网络信鸽，把对方的错误信息携带回来，下面由我来好好介绍这个新协议！

一、ICMP协议概述

我们在之前的路由实践中已经看到，配置网络中的路由并不总是那么容易。而且当配置不成功时，要找出错误的来源也不容易。

ICMP 协议的目的之一就是使网络调试变得更加容易。大致说来，ICMP 协议使我们能够快速了解网络问题的根源，并为我们提供排查网络问题的工具。

ICMP 是 Internet Control Message Protocol 的缩写，表示 “互联网控制报文协议”。

ICMP报文被封装在IP数据报中发送，ICMP协议也不会与IP协议竞争，因为它的目的不是传输信息。ICMP协议的作用是控制传输错误，并帮助我们进行网络调试。

因此注意， ICMP是帮助我们快速排查网络问题的工具。因此，ICMP协议是IP协议的 “补充”，或更准确地说，是TCP/IP协议栈的 “补充”，ICMP协议使得我们在出现问题时更容易理解网络上到底发生了什么

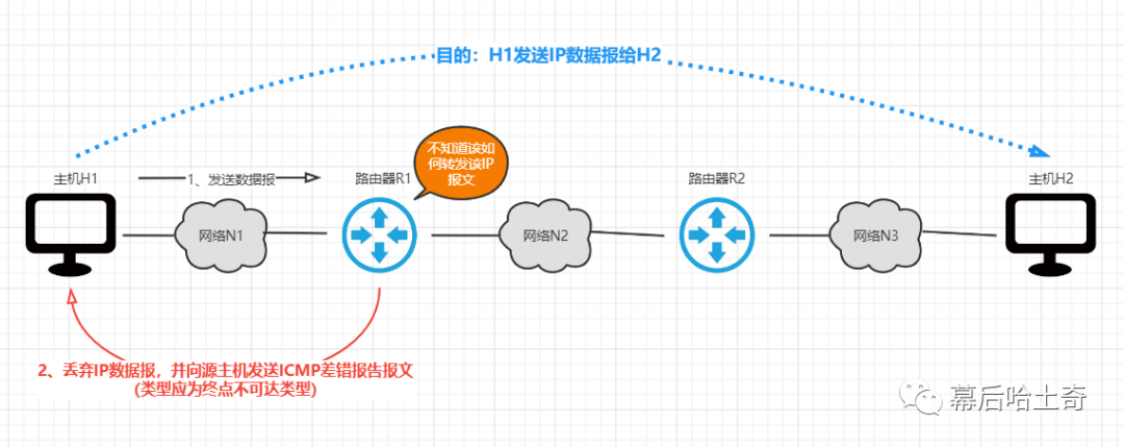
二、ICMP五大差错报告报文

ICMP为传递网络中发生的错误信息而生，作为它最神圣的使命，我们好好来看看ICMP五种差错报告报文。

①终点不可达

当路由器或主机不能交付数据报时，就向源点发送终点不可达报文，具体可再根据ICMP的代码字段细分为目的网络不可达、目的主机不可达、目的协议不可达、目的端口不可达、目的网络未知、目的主机未知等13种错误。

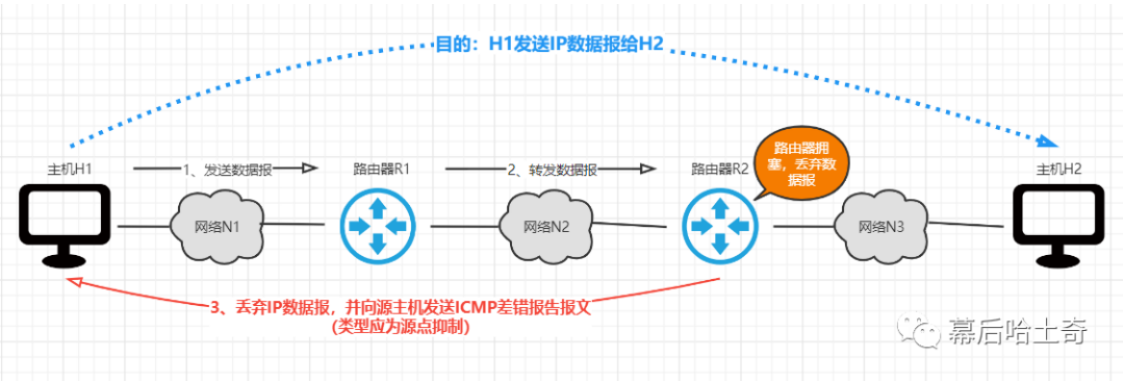
下面举个小例子来辅助理解。



假设主机H1要给主机H2发送IP数据报，H1首先将IP数据报发送给路由器R1，由R1帮其转发，若R1的路由表中没有网络N3的路由记录、默认路由以及主机H2的特定主机路由，则R1就不知道如何转发该数据报，只能将其丢弃，并向发送该数据报的源主机H1发送ICMP差错报告报文，其类型为终点不可达。

②源点抑制

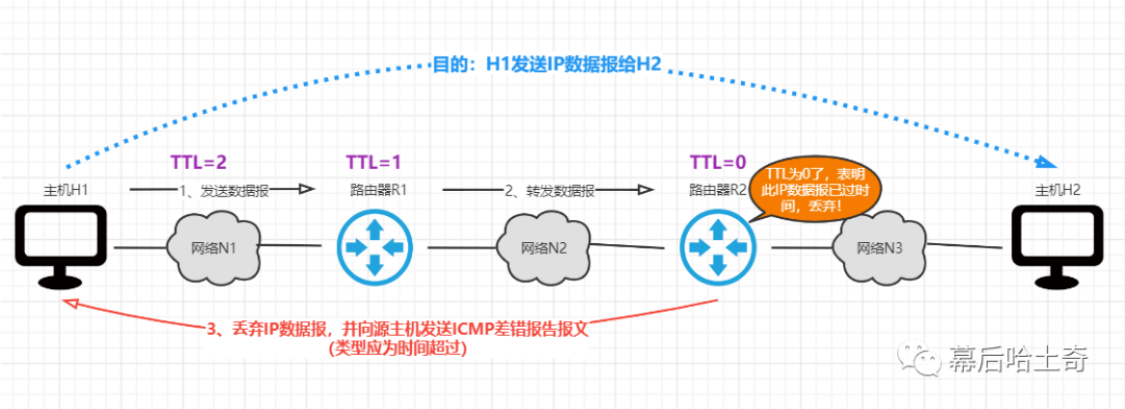
当路由器或主机由于拥塞而丢弃数据报时，就向源点发送源点抑制报文，使源点知道应当把数据报的发送速率放慢。



如图所示，主机H1要给主机H2发送IP数据报，H1首先将IP数据报发送给路由器R1，由R1帮其转发给路由器R2，由于R2拥塞，也就是R2比较繁忙，R2根据自己的丢包策略丢弃了该数据报，并向发送该数据报的源主机H1发送ICMP差错报告报文，其类型为源点抑制。

③时间超过

当路由器收到一个目的IP地址不是自己的IP数据报时，会将其生存时间TTL字段的值减一。若结果不为0，则将该IP数据报继续转发出去；若结果为0，除丢弃该IP数据报外，还要向源点发送时间超过报文。

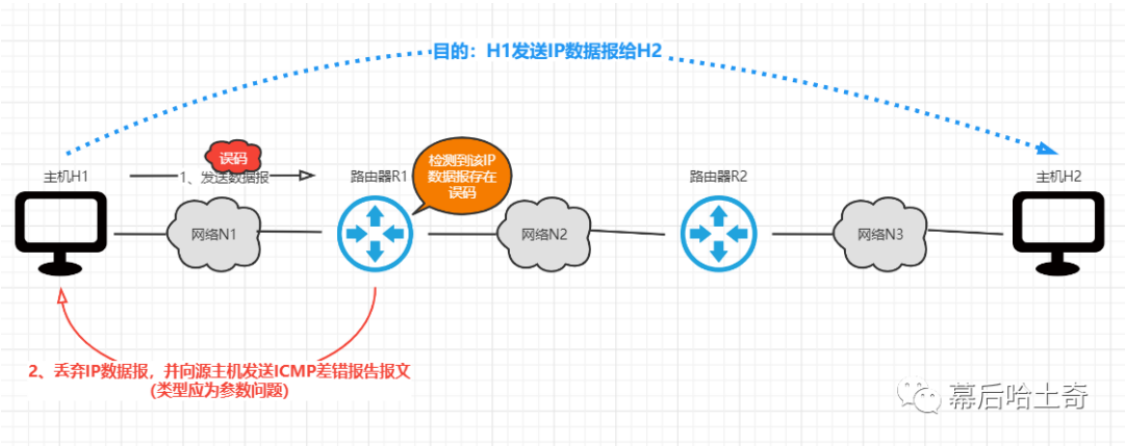


例如某生存时间为2的数据报传送到了R1，R1将其生存时间减1后发现结果为1，则继续转发给路由器R2，路由器R2将其生存时间再减1，此时变成了0，那么表示该数据报生存时间结束了，则丢失该数据报，并向发送该数据报的源主机H1发送ICMP差错报告报文，其类型为时间超过。

另外，当终点在预先规定的时间内不能收到一个数据报的全部数据切片时，就把已收到的数据切片都丢弃，也会向源点发送时间超过报文。

④参数问题

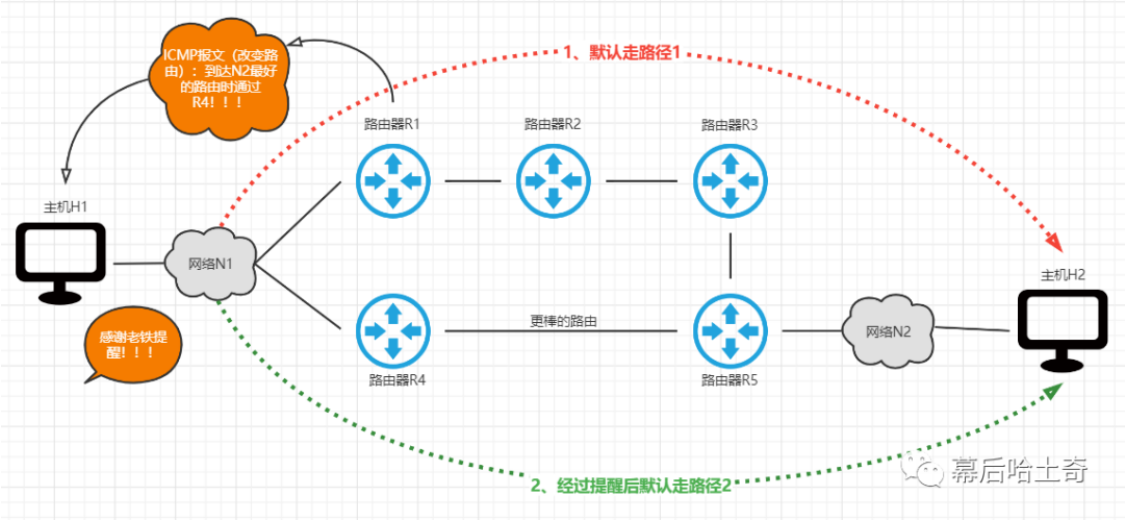
当路由器或目的主机收到IP数据报后，根据其首部中的检验和字段发现首部在传输过程中出现了误码，就丢弃该数据报，并向源点发送参数问题报文。



如图所示，主机H1要给主机H2发送IP数据报，假设该数据报在传输过程中受到了干扰，其首部出现了误码，当这个出错的数据报到达路由器R1，R1检测出该数据报出错，则丢弃该数据报，并向发送该数据报的源主机H1发送ICMP差错报告报文，其类型为参数问题。

⑤改变路由（重定向）

路由器把改变路由报文发送给主机，让主机知道下次应将数据报发送给另外的路由器（可通过更好的路由）。



假设我们给主机H1指定的默认网关是路由器R1，则H1要发往主机H2的IP数据报就会先传输给R1，路径是R1-》R2-》R3-》R5-》H2。

当R1发现发给主机H2的最佳路由不应该是经过R1，而应该是经过R4时，就用改变路由报文把这个情况告诉主机H1，于是H1就在自己的路由表中添加一条记录：到达主机H2所在网络N2应经过路由器R4，而不是R1。

以上路由器之间能互相知晓，可得益于我们之前《27 | 网络层篇：路由选择协议三剑客：OSPF协议》学习的OSPF协议，这里就不再赘述了。

三、ICMP两大询问报文

除了本职的差错报告报文外，还有兼职，比如询问报文。常用的ICMP询问报文分为以下两种：

回送请求和回答

也可以叫做回显请求和回显应答，ICMP回送请求报文是由主机或路由器向一个特定的目的主机发出的询问，收到此报文的主机必须给源主机或路由器发送ICMP回送应答报文。

这种询问报文用来测试目的站是否可达及了解其有关状态，下面说到的ping就用到了这种询问报文。

时间戳请求和回答

ICMP时间戳请求报文是请某个主机或路由器回答当前的日期和时间。

在ICMP时间戳回答报文中有一个32位的字段，其中写入的整数代表从1900年1月1日起到当前时刻一共有多少秒。

四、ICMP应用举例及原理剖析

①第一个应用是分组网间探测PING，英文全称是Packet InterNet Groper：

- 用来测试主机或路由器之间的连通性；
- 应用层直接使用网际层ICMP，没有通过传输层的TCP或UDP协议；
- 原理是使用了ICMP的回送请求报文“echo request”和回送应答报文“echo reply”。

很显然，一台机器发送“echo request”消息，目标机器以“echo reply”消息响应。这就是为什么当我们从一台机器成功ping通另一台机器时，我们知道两个方向的路由都是正确的。

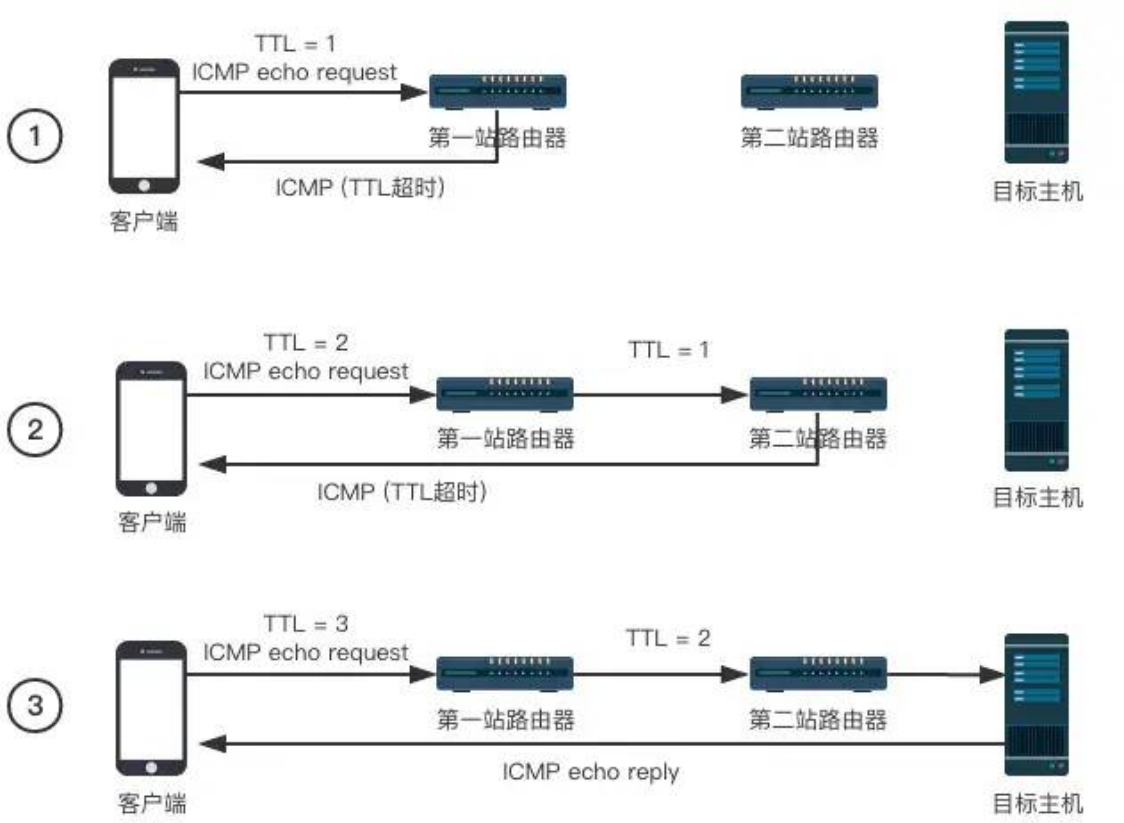
②第二个应用是跟踪路由traceroute，用来测试从源主机到达目的主机要经过哪些路由器：

- 应用层直接使用网际层ICMP，没有通过传输层的TCP或UDP协议；
- 原理是使用了ICMP的回送请求报文“echo request”、回送应答报文“echo reply”、差错报告报文（类型为时间超过类型）。

时间超过类型的ICMP差错报告报文：当路由器收到一个目的IP地址不是自己的IP数据报时，会将其生存时间TTL字段的值减一。若结果不为0，则将该IP数据报继续转发出去；若结果为0，除丢弃该IP数据报外，还要向源点发送时间超过报文。

第一次将IP数据报中的TTL的值设置为1，第一个路由器接收到数据报文，将数据报文中的 TTL值减1，因此TTL的值就变为0，第一个路由器就必须将该消息丢弃，并向我这台机器发送ICMP的“TTL exceeded”（时间超过）错误消息。这样，我就能知道第一个路由器的IP地址了！

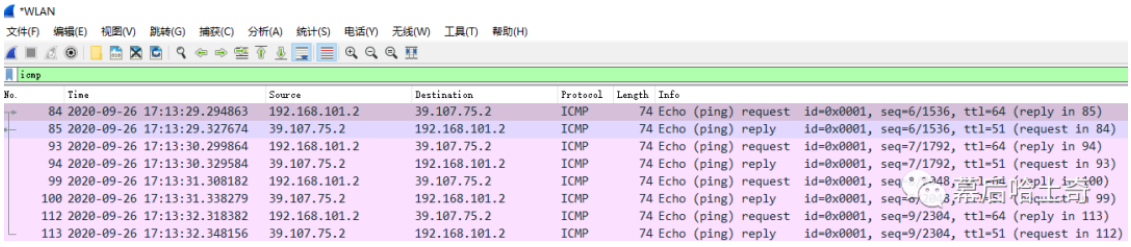
那么要知道第二台路由器的地址，只需将数据报中的TTL值设置为2即可。以此类推，我们就可以知道经过的所有路由器的IP地址了！



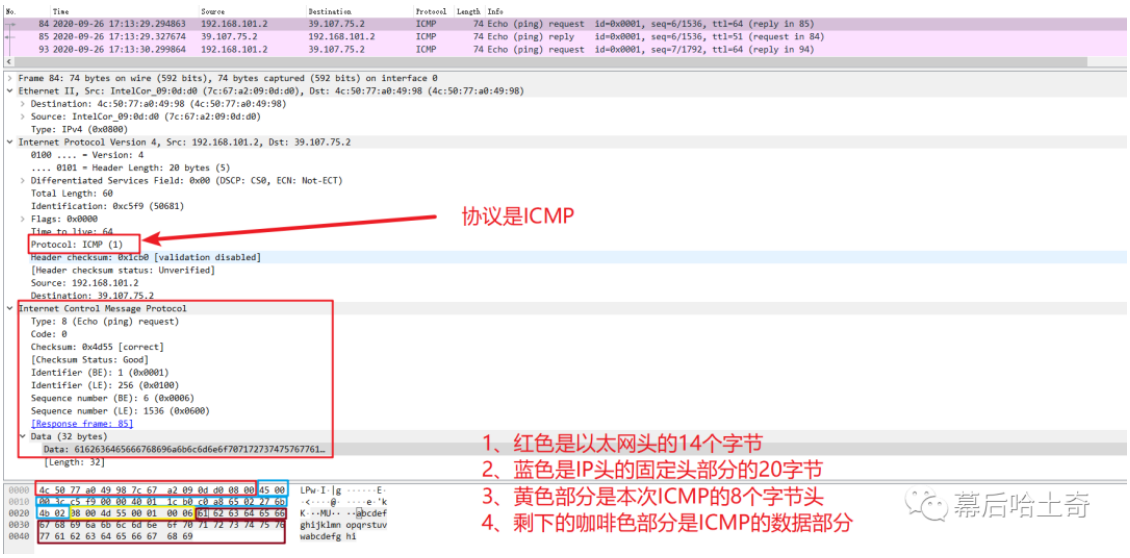
- 1、客户端发送一个TTL为1的ICMP回送请求报文，在第一跳的时候超时并返回一个ICMP超时数据报文，得到第一跳的地址。
- 2、客户端发送一个TTL为2的ICMP回送请求报文，返回ICMP超时数据报文得到第二跳的地址。
- 3、客户端发送一个TTL为3的ICMP回送请求报文，到达目标主机，目标主机返回一个ICMP回送应答报文，traceroute结束。

五、实践验证理论-对ping抓包分析

找一个目标IP做下ping的操作，我们看到在windows下默认会尝试4次的交互，因此我们抓包中也可以看到四对请求和回执：



我们拿84号的请求帧来看看里面的情况：



我们结合实际抓包，实际上整个ICMP报文的格式如下图所示：

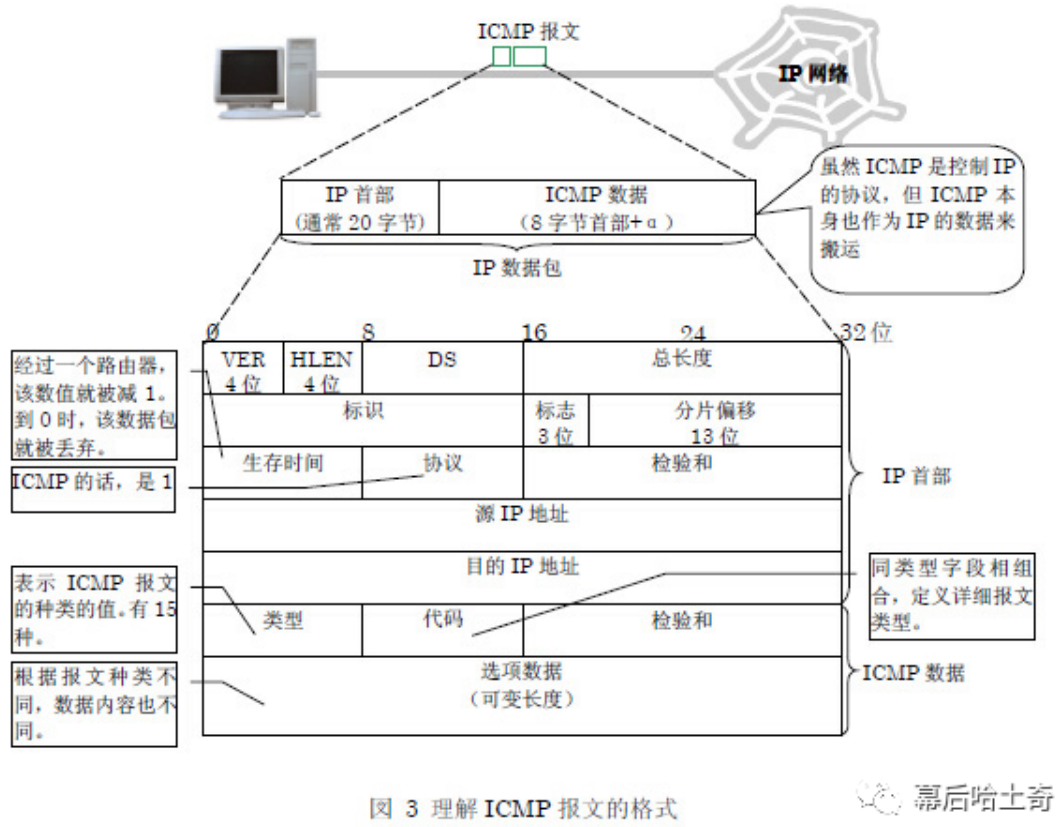


图 3 理解 ICMP 报文的格式

学习要学会抓重点，ICMP 报文的头部（header）中有两类有用的信息：type 和 code。

- type 表示“类型”，因此 type 用于说明 ICMP 消息的用途；
- code 表示“代码”，因此 code 用于说明 ICMP 消息的角色。

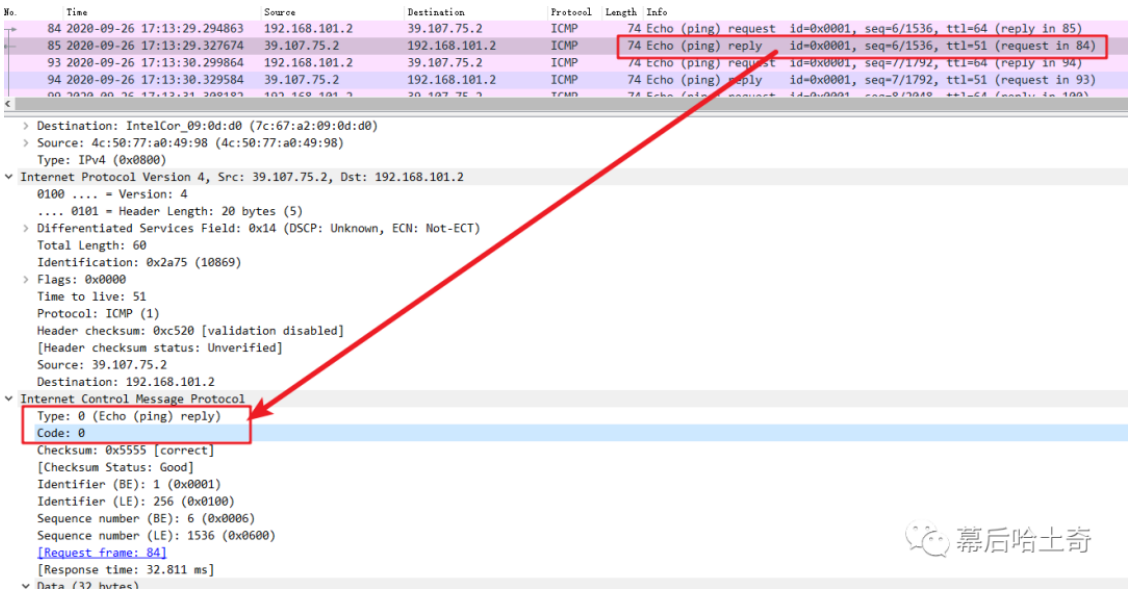
比如这里的type为8，code为0，则表示该ICMP报文是ICMP回送请求报文，我们看下RFC关于type和code的组合表示情况：

ICMP类型

TYPE	CODE	Description	Query	Error
0	0	Echo Reply——回显应答（Ping应答）	x	
3	0	Network Unreachable——网络不可达		x
3	1	Host Unreachable——主机不可达		x
3	2	Protocol Unreachable——协议不可达		x
3	3	Port Unreachable——端口不可达		x
3	4	Fragmentation needed but no frag. bit set——需要进行分片但设置不分片比特		x
3	5	Source routing failed——源站选路失败		x
3	6	Destination network unknown——目的网络未知		x
3	7	Destination host unknown——目的主机未知		x
3	8	Source host isolated (obsolete)——源主机被隔离（作废不用）		x
3	9	Destination network administratively prohibited——目的网络被强制禁止		x
3	10	Destination host administratively prohibited——目的主机被强制禁止		x
3	11	Network unreachable for TOS——由于服务类型TOS，网络不可达		x
3	12	Host unreachable for TOS——由于服务类型TOS，主机不可达		x
3	13	Communication administratively prohibited by filtering——由于过滤，通信被强制禁止		x
3	14	Host precedence violation——主机越权		x
3	15	Precedence cutoff in effect——优先中止生效		x
4	0	Source quench——源端被关闭（基本流控制）		
5	0	Redirect for network——对网络重定向		
5	1	Redirect for host——对主机重定向		
5	2	Redirect for TOS and network——对服务类型和网络重定向		
5	3	Redirect for TOS and host——对服务类型和主机重定向		
8	0	Echo request——回显请求（Ping请求）	x	
9	0	Router advertisement——路由器通告		
10	0	Route solicitation——路由器请求		
11	0	TTL equals 0 during transit——传输期间生存时间为0		x
11	1	TTL equals 0 during reassembly——在数据报组装期间生存时间为0		x
12	0	IP header bad (catchall error)——坏的IP首部（包括各种差错）		x
12	1	Required options missing——缺少必需的选项		x
13	0	Timestamp request (obsolete)——时间戳请求（作废不用）	x	
14		Timestamp reply (obsolete)——时间戳应答（作废不用）	x	
15	0	Information request (obsolete)——信息请求（作废不用）	x	
16	0	Information reply (obsolete)——信息应答（作废不用）	x	
17	0	Address mask request——地址掩码请求		
18	0	Address mask reply——地址掩码应答		

type表示类型，比如【终点不可达】对应的type一般是3，改变路由（重定向）对应的type是5。消息的code则会告诉我到底是什么导致了此网络问题。

从表上可以看到ICMP回送应答报文应该是type等于0，code也等于0，下面来看下抓包是不是真的如此！我们简单看下reply包里面字段验证下：



显然符合预期！

当然了，上面演示的都是正常无错误的情况，实际上只有发生错误的时候，这里的code和type才更加有存在意义。

正如上面说过，type 为 3 的 ICMP 数据报文说明“终点不可达”（Destination Unreachable）。

如果我将数据报文发送到一台机器，却收到type为 3 的ICMP消息，那么我就知道网络上存在问题，不能到达目的地。

当type为 3 时，接下来根据消息的code看到底是什么导致了此网络问题：

- 如果code为 0，则说明该【网络不可达】，Network Unreachable。一般来说，是因为途中的一个路由器的路由表中没有到达此目标网络的路由记录；
- 如果code为 1，则说明该【机器不可达】，Host Unreachable。应该是最后一个路由器发送了ARP请求，但没有得到响应。

其他的类型，我们只要掌握了上面的ICMP五大差错报告报文和ICMP两大询问报文，基本就扫平ICMP协议了，本文完。