第5章 ICMP协议分析

IP协议负责把数据从一个网络传输到另一个网络，而IP协议自身没有差错控制机制，如果在传输中出现因某种原因不能发送的IP数据，ICMP将被用来传输差错报文及其他需要注意的信息。作为一个重要的错误处理和信息处理协议，ICMP是TCP/IP协议族中不可或缺的一部分。本章讨论ICMP报文的类型、结构及各种ICMP报文的应用，分析ICMP回显请求与应答、ICMP重定向差错报文的格式和工作过程，介绍ping程序和traceroute程序的用法和工作过程、IP记录路由选项、时间戳选项和源站选路选项等有关内容。

5.1 ICMP的作用

作为网络层重要的协议，ICMP可以提供有关网络可连接性的信息，获得基于数据报或无连接协议不能传输的路由行为的信息。

如果要诊断和修复TCP/IP连接性问题，就必须知道从什么地方可以得到IP互联网上数据包如何从源位置传输到目的位置的信息。通常网络的可达性表述为：对于任何与另一个网络节点进行通信和交互数据的网络节点来说，一定存在从发送方到接收方转发数据的某种方法。正常情况下，可用转发路径可在位于发送方和接收方之间各种中间设备的本地IP路由表的内容中发现。

由于IP自身是不可靠传输，不能提供可达性、交互错误、路由错误报告以及控制信息，因此，由ICMP提供一种将信息返回给发送方的方法。通过采用特殊的ICMP消息格式，消息包含数据包在转发过程中经历的路由器信息(包括可达性信息)，并提供了一种当路由或可达性问题阻止交付IP数据包时返回出错信息的方法，这种能力很好地补充了IP的数据包交付服务。

需要指出的是，ICMP虽然也是网络层的协议，但却封装在IP报文中。从这个意义上看，ICMP消息不过是特殊格式的IP数据报，与一般网络流量中其他IP数据报受到相同的限制。另外，ICMP报告错误、阻塞以及其他网络状况的能力对于增强IP的尽最大努力交付方法本身并没有任何直接的好处。因此，即使ICMP能够报告错误或网络阻塞，如果要借助它来改变网络的通信状况，则依赖于接收消息的主机操作这些消息内容的方式。

典型的例子是ICMP重定向消息的处理。当网关和路由器转发数据报时发现有更好的路径去往目标主机，则把一条ICMP消息提供给发送方，把主机引导到一条更好的网络路由上，即发送一条重定向消息。主机对这条ICMP消息的处理则各有不同，一般默认使用网络

上的最佳路径传输数据，但也可以丢弃这条消息而不使用新的路由。

RFC 792 提供了有关 ICMP 协议的基础规范，并定义了各种 ICMP 信息和服务的类型。在这个标准中，明确了 ICMP 是 IP 基础支持的一部分，为网关或目的主机提供了一种与源主机通信的机制；规定采用特殊格式的 IP 数据报，使用特殊的关联消息类型和代码，同时为了防止出现错误消息的循环，ICMP 不传输有关自身的任何消息，并且仅提供任何分片数据包序列中的第一个分片的消息。

5.2 ICMP 报文及类型

5.2.1 ICMP 报文格式

ICMP 报文是封装在 IP 数据报内作为 IP 报文的数据被传输的，如图 5-1 所示。但 ICMP 并非更高层的协议，仍被认为是网络层的一个组成部分。

ICMP 报文的种类很多，而且各自又有自己的代码和处理信息内容，因此，ICMP 并没有一个统一的报文格式以供全部 ICMP 报文来使用。ICMP 报文结构如图 5-2 所示。

图 5-1 ICMP 报文的 IP 封装

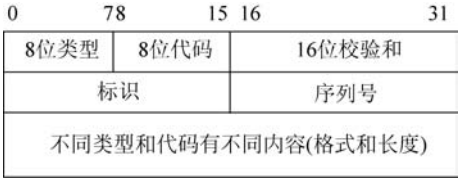

图 5-2 ICMP 报文格式

尽管不同的 ICMP 报文类别有不同的报文字段，但 ICMP 报文在首部内容上还是一致的，即前 4 字节有统一的格式，共有类型、代码和校验和 3 个字段。接着的 4 字节内容与 ICMP 报文的类型有关(图 5-2 中的标识和序列号是最常见的内容)。

不同类别的 ICMP 报文由类型和代码字段共同来区分，代码是为了进一步区分某种类型中的不同情况。

校验和字段覆盖整个 ICMP 报文。使用的算法与 IP 首部的校验和算法相同，采用二进制反码求和的方式来得到校验和。ICMP 的校验和是必需的。

在报文首部后面的是数据字段，其长度取决于 ICMP 报文的类型。

5.2.2 ICMP 报文类型

ICMP 将在网络里传输数据过程中需要报告给数据发送者的错误用一个预先定义好的消息集合来表示，每种消息都提供专用的功能，并用 ICMP 类型和代码来表示。对网络控制或探测的其他功能，ICMP 也提供了相应的类型和代码来处理。

ICMP 报文虽然细分为很多类别，但总的来看，可以分为如表 5-1 所示的两大类：差错报告和查询报文，也常把差错报告进一步分为差错报文和控制报文。报文的不同功能由报文中的类型字段和代码字段共同决定。

ICMP 所有报文类型字段可以有 15 个不同的值,用以描述特定类型的 ICMP 报文。其中类型 15 和 16 已废弃。大多数 ICMP 报文还使用代码字段来进一步描述不同的条件。

表 5-1 ICMP 报文的分类

ICMP 报文种类	类 型	描 述
差错报告	3	目的不可达
	4	源站抑制
	5	路由重定向
	11	超时
	12	参数问题
查询报文	8 或 0	回显请求或应答
	10 或 9	路由器询问或通告
	13 或 14	时间戳请求或应答
	17 或 18	地址掩码请求或应答

根据不同的 ICMP 报文类型和代码,ICMP 报文的其他部分也有不同的内容,本章将分别对各类报文及其用途做详细介绍。

5.2.3 ICMP 差错报告

1. ICMP 差错报告的特点

ICMP 差错报告具有以下特点。

- 只报告差错,但不负责纠正错误,纠错工作留给高层协议去完成。
- 发现出错的设备只向信源报告差错。
- 差错报告作为一般数据传输,不享受特别优先权和可靠性。
- 产生 ICMP 差错报告的同时,会丢弃出错的 IP 数据报。

需要注意的是,下列情况下将不会产生 ICMP 差错报文。

- 在对 ICMP 差错报文进行响应时,永远不会生成另一份 ICMP 差错报文。但 ICMP 查询报文可以产生 ICMP 差错报文。
- 广播或多播的 IP 数据报。
- 作为链路层广播的数据报。
- 不是 IP 分片的第一片。
- 源地址是零、环回地址、广播或多播地址。

当发送一份 ICMP 差错报文时,报文始终包含产生 ICMP 差错的 IP 报文的首部和 IP 报文中数据的前 8 字节。这样,接收 ICMP 差错报文的模块就会把它与某个特定的协议(根据 IP 数据报首部中的协议字段来判断)和用户进程(根据包含在 IP 数据报前 8 字节中的 TCP 或 UDP 报文首部中的 TCP 或 UDP 端口号来判断)联系起来,进而可以进行针对性的处理。

ICMP 差错报文主要有目的不可达、超时和参数问题三类,源站抑制和路由重定向是具有控制作用的差错报文。

2. ICMP 目的不可达

目的不可达差错报文主要用于当路由和传输错误阻止 IP 数据报抵达其目的地时记入文档。网关在寻找路由和转发数据报的时候，可能因为各种原因不能够转发成功，例如目标机不在运行中(关机或故障)、目标地址不存在或者网络没有去往目的地的路由，这时网关都会产生目的不可达的 ICMP 差错，返回给源主机。

如图 5-3 所示是类型 3(目的不可达)的差错报文格式。

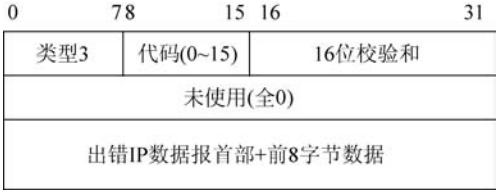

图 5-3 ICMP 目的不可达报文格式

目的不可达需要区分的情况比较多，这时代码值就十分重要。代码及对应的含义如表 5-2 所示，处理方法请参阅参考文献或有关教材。

表 5-2 ICMP 类型 3 代码描述

代码	描 述	代码	描 述
0	网络不可达	8	源主机被隔离
1	主机不可达	9	目的网络被强制禁止
2	协议不可达	10	目的主机被强制禁止
3	端口不可达	11	因服务类型 TOS 网络不可达
4	需要分片但设置了不分片	12	因服务类型 TOS 主机不可达
5	源站选路失败	13	因过滤通信被强制禁止
6	目标网络不认识	14	主机越权
7	目标主机不认识	15	优先权终止生效

从表 5-2 可以看出，所谓目标有 4 个层次的概念，从大到小依次为网络、主机、协议和端口，各层次存在相应的依赖关系。例如，全局性的协议地址包括网络地址、主机地址和协议地址，全局性的主机地址则包含网络地址和主机地址，否则不能在 Internet 上使用，发往某个协议地址则可能引发网络不可达、主机不可达或协议不可达错误。

由于寻找路由是基于网络地址的，所以网络不可达说明存在寻找路由故障。

如果出现主机不可达错误，则必然不会发生网络不可到达的故障，而且说明寻找路由是正常的，因此主机不可达的问题是传输过程中的问题。例如，目的主机不在运行中或者目的主机不存在，这些问题是路由中的最后一个网关，即目的主机所在网络上的网关，通过网络硬件提供的应答机制发现的。

协议不可达和端口不可达，这两种报文涉及更高级的协议，由目的主机本身所产生。实际上是 IP 报文虽然到达目的主机，但是没有办法被高层应用软件接收。由于高层软件往往采用多重协议，而同一协议则可能通过不同协议端口同时处理多个访问，因此 IP 数据报的信宿可能深入到协议和端口的深层结构。协议号和端口号如同网络地址和主机地址一样，

也作为数据报目的地址的一部分,因此协议和端口不可达在概念上也如同网络和主机不可达一样。

源主机可以从返回的 ICMP 差错报文中出错数据报的报头发现通信双方的有关信息,如目的主机的确切地址和协议类型等信息,从 IP 数据的前 8 字节里获得上层协议的端口地址等内容。

3. 超时

在 Internet 中,为了防止出现路由循环,TCP/IP 采取了两个措施:一是每个 IP 数据报的报头设置 TTL(Time To Live,生存时间)域,二是对分片数据报采用定时器计数。其核心思想就是通过定时来限制数据包在网络中的逗留时间,以防出现不可忍受的传输延迟,从而提高网络的传输率。在上述措施中,一旦报文的定时时间到,网关或信宿机都要立即抛弃本数据报并向信源机发送 ICMP 超时报告,网络上所有的路由器都不会转发超时的报文。

由此可见,超时差错报文用于指示 IP 数据报的 TTL 或分片 IP 数据报的重组定时器超时,其类型为 11,代码取值为 0 或 1,其中 0 代表 TTL 超时,1 代表分片重组超时。

在路径探测中,超时报文也具有特别的重要作用,将在后面详细介绍。

4. 参数问题

参数问题主要是用于指示在入站数据报的 IP 首部的数据或者 IP 选项的数据发生了某些问题,网关或主机不得不抛弃数据报时,将会向源主机发送参数错误的 ICMP 报文。

不过 ICMP 参数问题差错报文的功能很弱,仅能够指出报文首部中引起故障的字节,对一些模糊的问题进行表述,一般都还需要进一步的处理。

5.2.4 ICMP 控制报文

ICMP 差错报告中具有控制功能的报文包括源站抑制报文和重定向报文,其中源站抑制报文用于拥塞控制,路由重定向报文则用于路径控制。

1. 源站抑制

源站抑制也叫源抑制,是指通过向相应的信源发送 ICMP 源抑制报文,信源根据收到的源抑制报文中所带的先前发出的 IP 数据报的首部信息,决定对去往某一特定信宿的信息流进行抑制,通常是减缓信源发出数据报的速率,以实现拥塞控制。

在网络通信过程中,当大量的数据报进入路由器或信宿时,会造成缓冲区溢出,即出现拥塞。引起网络拥塞的原因,可能是网关的处理速度太慢,不能完成对大量用户数据包的处理,或者是网关输入数据的速率大于输出线路的容量,许多数据同时通过同一网关转发就可能导致拥塞。从本质上讲,拥塞的原因都在于没有足够的数据缓冲区,只要有足够的数据缓冲区,网关或主机总能将数据存入队列等待处理。

拥塞控制的方法很多,TCP/IP 采用源抑制技术,即抑制信源发出数据包的速率。通常,网关周期性地测试每条输出线路,一旦发现某条输出线路发生拥塞,立即向相应的信源机发送 ICMP 源抑制报文,根据网关输出队列的情况会有不同的发送方式。信源机收到源

抑制报文后，按一定的速率降低发往某信宿的数据报速率。拥塞解除后，信源机要恢复数据报传输速率。恢复的过程与 ICMP 无关，完全由主机自行解决。

源抑制报文类型为 4，码值只有一个，即 0。

2. 路由重定向

路由重定向差错是指通过路由器发送重定向报文，网络上主机中的路由表也可以得到更新。

网络上的路由器和主机中都存有一个路由表，路由表决定了去往目的地的下一跳路由器的地址。路由器上的路由表通过不同的路由协议在路由器之间定期交换路由信息（参见第 6 章），以保证其能及时地反映网络结构的变化。

主机路由表所给出的下一跳路由器可能并非是去往信宿的最佳的下一跳路由器。当主机的下一跳路由器收到数据报后，该路由器根据它的路由表判断本路由器是否是去往信宿的最佳选择，如果不是，该路由器仍然会向信宿网络转发该数据报，但在转发的同时会产生一个 ICMP 重定向报文，通知信源主机修改它的路由表，重定向报文中将给出信源最佳下一跳路由器的 IP 地址。重定向报文格式如图 5-4 所示。

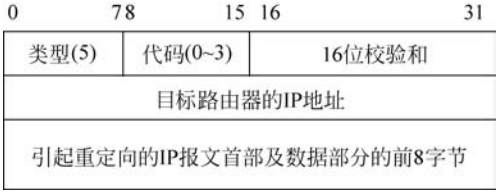

图 5-4 ICMP 重定向报文格式

重定向是路由器向主机发送的、请求主机改变路由的 ICMP 差错报文，主机操作系统决定了对重定向报文的处理。Windows 系统和许多 UNIX 系统都支持 ICMP 重定向，即对于网关返回的 ICMP 重定向报文，系统会在主机的路由表中修改或添加一项主机路由。

对路由器而言，收到 ICMP 重定向报文的一般处理是丢弃。但在关闭 IP 路由的情况下，某些类型的路由器，如 Cisco 路由器，会接收 ICMP 重定向报文并修改自己的路由表，即在 IP 路由关闭的情况下，路由器会作为主机执行操作。若这个漏洞被攻击者利用来发送伪造的 ICMP 报文，可能导致 IOS 路由表被修改，从而破坏或截获通信。

如图 5-4 所示，ICMP 重定向报文类型为 5，代码有 4 个可选值，即 0～3，其中 0 和 2 与网络重定向有关，1 表示主机重定向，3 表示对服务类型和主机重定向。报文中目的路由器的 IP 地址即去往信宿的最佳下一跳路由器的 IP 地址。

特别要注意的是，原则上重定向报文是由路由器产生而供主机使用的。

5.2.5 ICMP 查询报文

ICMP 查询报文的出现使得互联网上的任何主机或路由器可以向其他主机或路由器发送请求并获得应答。

通过 ICMP 查询报文，网络管理员、用户或应用程序可以对网络进行检测，了解设备的

可达性、地址掩码的设置、时钟的同步等情况,利用这些有用的信息对网络进行故障诊断和控制。

1. 回显请求与回显应答

回显请求报文用于向特定的信宿机发送一个回显请求,其中包含一个任选的数据区。信宿机收到回显请求报文就发回相应的回显应答,其中包含一个请求中数据区的副本。

假如回显请求发出后,成功地收到一个回显应答,同时应答中的数据副本与请求中的数据完全一致,则不但说明信宿机可以到达,而且说明数据报传输系统的工作整体上是正常的,至少信源机和信宿机的 IP 协议软件、ICMP 协议软件是工作正常的,请求与应答经过的中间网关也能正常寻找路由。

由此可见,ICMP 回显请求和应答不仅可以被用来测试主机或路由器的可达性,还可以测试 IP 协议的工作情况。这对于在网络工程实践中判定网络状况有直接的帮助。

ICMP 回显请求与应答报文的格式如图 5-5 所示。

类型"8"表明是回显请求报文,代码只有一个,为 0。类型"0"表明是回显应答报文,代码也只有一个,为 0。数据部分则由于协议的不同实现,其内容和长度会有所不同。

图 5-5 ICMP 回显请求与应答报文

协议未对标识符和序列号字段进行正式定义,通常将标识符和序列号用于匹配请求与应答,标识符一般为发起请求进程的进程 ID,回显请求与应答报文的标识符和序列号要一致。

最典型和常用的 ping 命令的功能就是利用 ICMP 回显请求与应答报文来实现的。

2. 路由器询问或通告

路由器询问或通告是利用 ICMP 来实现路由器初始化路由表的一种方法。

一般认为,主机在引导以后,要广播或多播传送一份路由器询问报告,也叫路由器请求报文。路由器请求报文的格式如图 5-6 所示,报文中没有更多的内容,通过类型 10 和代码 0 表明这是一个 ICMP 的路由器请求报文。

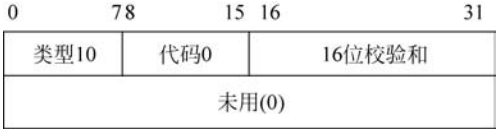

图 5-6 ICMP 路由器请求报文

网络上的一台或多台路由器响应一份路由器通告报文，报文的格式如图 5-7 所示。

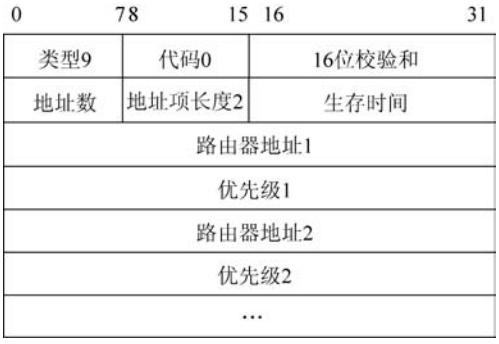

图 5-7 ICMP 路由器通告报文

路由器通告报文类型为 9，代码为 0，报文中的地址数指在数据包中公告的路由器地址个数；地址项长度则用于定义所公告的每个路由器地址按 4 字节计算的个数，这个值始终为 2；生存时间指这个路由信息可以被认为有效的最大秒数；路由器地址和优先级可以有一对或多对，表示发送的可用路由器的 IP 地址和优先级，优先级值越大，代表该地址的路由器越可能成为用于本地主机的默认网关。从报文结构可见，一条路由器通告中可以通告多个地址。

除了当路由器启动的时候会定期地在所有广播和多播传输接口上发送路由器通告报文以外，路由器也会定期地广播或多播传输其路由器通告报文，以允许每个正在监听的主机相应地更新它们的路由表。另外，路由器还要监听来自主机的请求报文并发送作为应答的路由器通告报文。

在较复杂的网络里，往往采用动态路由协议来实现路由通告，比如 RIP（Route Information Protocol，路由信息协议）等，这部分内容将在第 6 章介绍。

3. 时间戳请求与应答

ICMP 时间戳请求允许系统向另一个系统查询当前的时间，返回的建议值采用 UTC（Universal Time Coordinated，协调的统一时间）计时方式自 24：00 开始计算的毫秒数。

ICMP 时间戳请求和应答的报文格式如图 5-8 所示。

图 5-8 ICMP 时间戳请求和应答报文

类型 13 为时间戳请求报文，类型 14 为时间戳应答报文，代码值固定为 0。

请求端填写发起时间戳,然后发送报文。应答系统收到请求报文时填写接收时间戳,在发送应答时填写传输时间戳。实际上,大多数的实现把后面两个字段都设成相同的值(提供3个字段的原因是可以让发送方分别计算发送请求的时间和发送应答的时间)。这个报文格式是固定的,没有可选数据,所以其长度是固定的。

时间戳请求和应答报文可以用于估算请求主机和信宿机两地的时间差。首先计算出时间戳请求和应答报文的往返时间,并把这个时间作为一般数据包的往返时间:根据初始时间戳与信源机收到应答时的当前时间,两者相减便是往返时间。再通过接收时间戳和发起时间戳,计算出报文到达信宿机的时间,用接收时间戳减去发起时间戳即可。最后用这个时间减去往返时间的一半,计算出两地时差。当然,由于数据报在网络上传输的随机性,事实上上述的往返时间也不太准确,甚至采用多次测量求平均值的方法也不一定准确。

由于UTC是基于原子时的,因此这种ICMP报文的好处是提供了毫秒级的分辨率。但其不足之处是由于返回的时间是从24:00开始计算的,因此调用者必须通过其他方法获知当时的日期。

更严格的计时器使用NTP(Network Time Protocol,网络时间协议),该协议在RFC 1305中给出了描述。最新的NTP版本是2010年6月发布的第4版(NTP Version 4),其标准化文档为RFC 5905。2016年3月发布的RFC 7822和2019年6月发布的RFC 8573对其进行了补充。NTP是用来使网络中的各个计算机时间同步的一种协议,可以提供高精准度的时间校正,在局域网内可达0.1ms,在互联网上绝大多数的地方其精度可以达到1～50ms。NTP还提供一定的安全机制来防止网络攻击。

目前网络时间同步技术还在向更高精度、更强的兼容性和多平台的适应性方向发展。

4. 地址掩码请求与应答

ICMP地址掩码请求用于无盘系统在引导过程中获取自己的子网掩码。与利用RARP来获取IP地址类似,系统在引导过程中广播地址掩码请求报文,希望从网络上获取子网掩码。RFC规定,除非系统是地址掩码的授权代理,否则它不能发送地址掩码应答,大多数主机在收到请求时都发送一个应答,甚至有一些主机还发送错误的应答。

ICMP地址掩码请求与应答报文格式如图5-9所示。

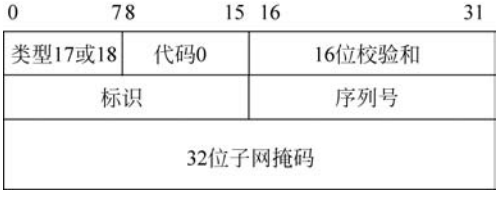

图5-9 ICMP地址掩码请求与应答报文

类型17为地址掩码请求,类型18为地址掩码应答。代码固定为0。ICMP报文中的标识符和序列号字段由发送端任意选择设定,这些值将在应答中被返回,这样,发送端就可以把应答与请求进行匹配。获得地址掩码的另一个方法是通过BOOTP实现。

5.3 ICMP 测试和故障诊断程序

目前,网络中用于 ICMP 测试和故障诊断的主要应用程序就是 ping 和 traceroute。

5.3.1 ping 程序

ping 是调试网络的基本工具,利用的就是最常用的 ICMP 回显请求与应答机制,最基本的用途就是测试网络的连通性。ping 检查是否有数据报被丢弃、复制或重传,这一般是通过在程序中连续地发送多个有不同序列号的 ICMP 请求,比较收到的 ICMP 应答的序列号来实现。ping 程序还校验每个收到的数据报,确定数据是否损坏。

ping 程序还能够通过在其所发送的数据报中存放发送请求的时间值,根据应答返回时的时间信息计算数据报的往返时间(Round Trip Time,RTT),据此推断网络通信状况。

不同操作系统下 ping 程序的功能都类似,但命令格式特别是参数有所不同。实验 5-1 中给出了 Windows 系统下 ping 命令的基本使用格式和参数。其他操作系统中的 ping 命令及参数的用法请参考相关资料。

利用 IP 选项,ping 程序还可以支持记录路由和时间戳信息。不同版本的 ping 程序都具有-r 选项,以提供记录路由(Record Route,RR)的功能,让 ping 程序在发送出去的 IP 数据报中设置 IP RR 选项(该 IP 数据报包含 ICMP 回显请求报文)。每个处理该数据报的路由器都把它的 IP 地址(通常是路由器数据出口的地址)放入选项字段中。当数据报到达目的端时,IP 地址清单应该复制到 ICMP 回显应答中。这样返回途中所经过的路由器地址也被加入清单中。当 ping 程序收到回显应答时,它就打印出这份 IP 地址清单。在具体实现上,UNIX 类系统就是如此,并且记录路由选项的路由器总是把出口的 IP 地址加入清单。

IP 首部中选项的最大字节数是 40,这样记录路由的最大问题是 IP 首部中用来存放路由器 IP 地址的空间很有限。IP 首部记录路由的一般格式如图 5-10 所示。选项说明字段用去前 3 字节,这样只剩下 37 字节来存放 IP 地址清单,也就是说只能存放 9 个 IP 地址。1 字节长的 code 指明 IP 选项的类型。对于 RR 选项来说,它的值为 7。length 是 RR 选项总字节数,ping 程序总是提供 39 字节的选项字段,对目前的网络来说,这已经不够用了。

ptr 称为指针字段。它是一个基于 1 的指针,指向存放着下一个路由器 IP 地址的位置。它的最小值为 4,指向存放第一个 IP 地址的位置。随着每个 IP 地址存入清单,ptr 的值相应增加,如图 5-10 所示中 ptr 指示的那样。

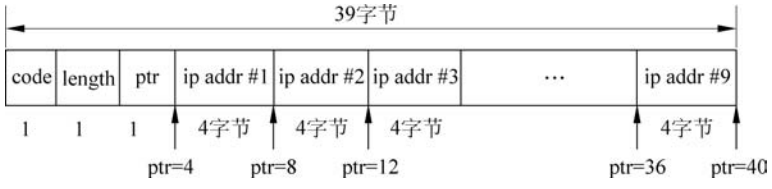

图 5-10 IP 首部记录路由选项的格式

IP 时间戳选项的处理上与记录路由选项类似,但选项说明字段占用 4 字节,报文格式如图 5-11 所示。其中,code 的值为 0x44,表示 IP 选项是时间戳选项,length 和 ptr 字段与记录路由选项相同。另外,增加了两个长度都是 4 位的字段 OF 和 FL。OF 为溢出字段,取 1 表示数据溢出,即选项空间不够完全记录数据。FL 为标志字段,取 0 时选项只记录时间戳;取 1 时选项要记录每台路由器的 IP 地址和时间戳。

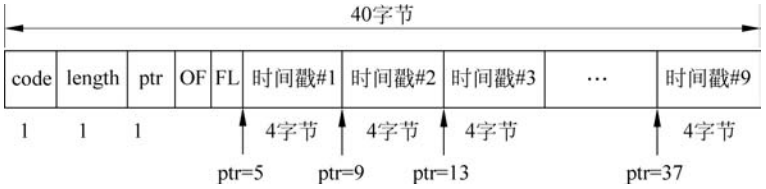

图 5-11 IP 首部记录时间戳选项的格式

若要在 IP 选项中同时记录时间戳处的路由器地址,就必须用 8 字节才能够同时记录位置和时间,这样 IP 选项最多记录 4 个路由器的时间戳。

ping 程序的使用也在随着技术的发展而变化。曾经还可以作出这样没有限定的断言:如果不能 ping 到某台主机,那么就不能 Telnet 或 FTP 到那台主机。随着 Internet 安全意识的增强,出现了提供访问控制清单的路由器和防火墙,那么像这样没有限定的断言就不再成立。一台主机的可达性可能不只取决于 IP 层是否可达,还取决于使用何种协议以及端口号。ping 程序的运行结果可能显示某台主机不可达,但仍然可以用 Telnet 远程登录到该台主机的某些端口,例如 25 号端口(邮件服务器)。

5.3.2 traceroute 程序

traceroute 程序(Windows 下程序名称为 tracert.exe)可以使用户获得 IP 数据报从一台主机传输到另一台主机所经过的路由。traceroute 还可以利用 IP 选项来支持源站选路。

如果要查看 IP 数据报经过的路径,使用 ping 程序的 IP 记录路由选项就可以实现。但由于网络上不是所有的路由器都支持记录路由选项,并且利用 ping 程序记录下来的路由器地址要记录往返的节点地址,这样来回记录使得数据量增加了一倍。更重要的是 IP 首部中记录路由选项的空间有限,而目前网络的规模又越来越大,这些都大大地限制了利用 ping 程序的 IP 记录路由选项来获得路径的效能,而 traceroute 则是代替其功能的有效实现。

traceroute 的工作原理主要是利用 ICMP 差错报文中的超时机制和 IP 首部中的 TTL 字段设置来实现的。不同的操作系统或环境中,traceroute 程序的具体实现上有所不同,主要有基于 ICMP 或基于 UDP 两种方式。

Windows 使用基于 ICMP 回显请求和响应的方法,其他包括 UNIX、Linux 和 Cisco 路由器中都使用基于 UDP 端口不可达机制。

1. 基于 ICMP 的 traceroute

traceroute 程序基于 ICMP 回显请求(Echo Request)、回显应答(Echo Reply)和超时(TTL-exceeded)来实现,完全基于 ICMP,因而也可简称为 ICMP traceroute。

这时,程序的工作机制描述如下。

(1) 首先源主机发出 ICMP Echo Request,第一次 Echo request 的 TTL 设置为 1,第二次 Echo request 的 TTL 设置为 2,依次递增,直至第 30 次,实际程序中每次一般会发出多个(常常是 3 个)Echo request 报文来避免网络传输带来的偶然错误。

(2) 中间的路由器对收到的 ICMP 报文中 IP 首部的 TTL 执行递减操作,如果 TTL 值为 0 或 1,就对源主机送回 ICMP TTL 超时报文(TTL-exceeded,ICMP 类型为 11),ICMP 请求报文同时因 TTL 超时而被丢弃。

(3) 源主机依次收到中间路由器发回的 ICMP TTL-exceeded 差错报文,由此知晓去往目的地所经过的每个路由器。

(4) 最后的 ICMP Echo Request 报文到达目标节点时,送回 ICMP Echo Reply,源主机收到这个 Echo Reply 报文便知道已经完成了路径探索,就不再发送 TTL 增加的 Echo Request 报文而是结束程序。

2. 基于 UDP 的 traceroute

这种 traceroute 程序的源主机发出的是 UDP 数据报,使用特别的 UDP 端口号,利用 ICMP 超时(TTL-exceeded: 类型 11)和 ICMP 端口不可达(port unreachable: 类型 3, code 3)差错报文来实现。由于程序基于 UDP 报文发送,因此可称为 UDP traceroute。

UDP traceroute 程序的工作机制描述如下。

(1) 源主机发出 UDP 数据报(可把这样的报文称为 UDP 探针)。探针报文的源端口使用随机的任何大于 32 768 的高段端口,报文的目的端口则从 33 434 开始,在后续的每个数据报中依次递增,直至 33 434+29 即 33 463。同时承载这些 UDP 探针的 IP 报文的 TTL 从 1 开始依次递增,直至 30(最多发送 30 个 UDP 探针)。

(2) 和 ICMP traceroute 程序的工作过程一样,中间的路由器会送回 ICMP TTL-exceeded 差错报文,使得源主机得知中间的每个路由器。

(3) UDP 探针报文到达最后的目的节点时,因为任何主机上都没有应用在使用 UDP 端口大于 32 768 这样的高段端口,所以目的节点发回 ICMP 端口不可达(port unreachable)差错报文。

traceroute 工作时会因为中间路由器的设置使得路由器不回送 TTL-exceeded 包,这样源主机上将看不到中间路由器地址,但却看得到报文最后到达目的主机时回送的响应。

某些网络设备,如 Cisco 路由器,可以使用 extended-traceroute 命令修改 UDP 探针使用的起始端口号 33 434。

3. IP 源站选路选项

traceroute 和 ping 命令都提供了源路由选项。下面对此做简要说明。

源路由即源站选路(source routing),其思想是由发送者指定路由。通常源路由分为两种形式:严格源路由和宽松源路由,其差别是严格源路由所指定的下一个路由器不在其直接连接的网络上,那么就返回一个“源站路由失败”的 ICMP 差错报文(类型为 3,代码为 5),而宽松源路由则允许数据报在清单上指明的任意两个地址之间通过其他路由器。

源路由的实现是通过采用 IP 选项来记录路由信息,其报文格式与图 5-10 所示的 IP 记录路由选项相同。宽松源路由的 code 取值为 0x83,严格源路由的 code 取值为 0x89。

源路由数据包在发送的过程中,会对选项中的IP地址清单进行更新。发送主机从应用程序接到源路由清单后,先将第一个表项取出,将所有剩余的地址向左移一格位置,并将最终目的地址作为清单的最后一项,再把ptr指针指向清单的第一项,然后将取出的第一个表项地址作为下一跳地址来发送报文。数据包到达目的主机时,如果指针大小比选项长度小,路由器会将指针指向的IP地址填入数据包的目的IP地址字段,将数据包外出接口(outgoing interface)的IP地址填入到指针指向的位置,然后将指针加4,指向下一个IP地址,再重新发送给这个新的IP地址。如果指针大小比选项长度大,说明已经到达了列表末尾,这台主机就是最终的目的主机。

如果数据包含有宽松源路由选项,那么数据包转发过程中,如果转发路由器不是目的主机,则会继续转发,不对IP列表进行操作。

当一个应用程序接收到有源路由指定路由的数据时,在发送应答时,应该读出接收到的路由值,并提供反向路由。

IP源站选路曾经是网络攻击者借用的手段,因此目前许多路由器对带有源站选路的报文都会设置为不予处理。相关内容请参考网络安全方面的资料。

5.4 小结

(1) ICMP可提供有关网络可连接性的信息,提供可达性、交互错误、路由错误报告及控制信息等IP不能够提供的信息,并把信息返回给发送方,是TCP/IP网络层的重要协议。

(2) ICMP报文是封装在IP数据报内作为IP报文的数据被传输的。ICMP报文除了首部4字节一致外,并没有一个统一的报文格式,而是采用不同的类型和代码值来区分各种类别的ICMP报文。

(3) ICMP差错报文对IP通信中产生的各种差错向源端进行报告,差错报文始终包含产生ICMP差错的IP报文的首部和IP报文中数据的前8字节。

(4) ICMP差错报告中具有控制功能的报文包括源站抑制报文和重定向报文,其中源站抑制报文用于拥塞控制,路由重定向报文则用于路径控制。重定向报文可以对主机路由表进行更新。

(5) ICMP查询报文中最常用的是ICMP回显请求与应答报文,其他还有路由器询问与通告、地址掩码请求和应答及时间戳请求和应答报文。这些都属于典型的请求—应答报文,通过ICMP报文中的标识符和序列号,应用程序可以在应答和请求之间进行匹配。

(6) 利用ICMP的重要应用程序有ping程序和traceroute程序,在网络工程实践中用作测试网络状况、获取路径信息的工具。

(7) 记录路由、时间戳和源站选路等应用都可以利用IP选项、ping程序和traceroute程序相结合来实现。

5.5 习题

1. 分析ping程序实现IP记录路由选项和时间戳选项的原理。

2. 分析traceroute工作的原理,尝试验证具体系统中traceroute程序的不同实现方法。

3. 什么叫宽松的 IP 源站选路和严格的源站选路?

4. 如何利用 netstat 命令查看主机收发的 ICMP 报文类型?

实验

实验 5-1 ICMP 回显查询报文

1. 实验说明

通过运行 ping 程序,在真实网络环境中观察分析 ICMP 回显请求和应答报文,理解 ICMP 查询报文的格式和工作特点。

Windows 系统下 ping 命令的基本格式和参数如下:

```
ping [ - t] [ - a] [ - n count] [ - l length] [ - f] [ - i TTL] [ - v TOS] [ - r count] [ - s count]
  [[ - j host - list] : [ - k host - list]] [ - w timeout] target_name
```

主要参数的用法及含义如下。

(1) -t: 校验与指定计算机的连接,直到用户中断。

(2) -a: 将地址解析为计算机名。

(3) -n count: 发送由 count 指定数量的回显请求报文,即 ECHO 报文,默认值为 4。

(4) -l length: 发送包含由 length 指定长度的数据的回显请求报文,默认值为 32 字节,最大值为 8192 字节。

(5) -f: 在 IP 首部中设置"不分片"标志,使包不被网络上的路由器分片。

(6) -i TTL: 将 IP 首部的"生存时间"字段设置为 TTL 指定的数值。

(7) -v TOS: 将 IP 首部的"服务类型"字段设置为 TOS 指定的数值。

(8) -r count: 在 IP 首部选项中的"记录路由"字段中记录发出报文和返回报文的路由。指定的 count 值最小为 1,最大为 9。

(9) -s count: 在 IP 首部选项中的"时间戳选项"字段中记录由 count 指定的转发次数的时间戳,也可以同时记录转发节点的 IP 地址。

(10) -j host-list: 经过由 host-list 指定的计算机列表的路由报文。中间网关可能分隔连续的计算机(松散的源路由)。允许的最大 IP 地址数目是 9。

(11) -k host-list: 经过由 host-list 指定的计算机列表的路由报文。中间网关不能分隔连续的计算机(严格的源路由)。允许的最大 IP 地址数目是 9。

(12) -w timeout: 以毫秒为单位指定超时间隔。

(13) target_name: 指定要校验连接的远程计算机。

ping 命令的参数较多,部分参数的用法和 Linux 系统中有所不同,实验时要注意区分。

2. 实验环境

Windows 或 Linux 操作系统及网络环境(主机有以太网卡并已连接局域网或 Internet),安装有 Wireshark。

3. 实验步骤

步骤 1 在实验计算机上先启动 Wireshark,过滤器可以设置为只查看 ICMP 报文。

步骤 2 在实验计算机上打开命令行窗口,对网络中存在的主机节点,如本局域网中某主机或互联网上的主机,运行如下 ping 命令:

```
C:\> ping 192.168.0.100
```

或

```
ping www.baidu.com
```

Windows 系统中的 ping 程序一般都是发送 4 个 32 字节数据的 ICMP 回显请求报文,查看捕获的数据包中回显请求与应答包的对应关系。

步骤 3 对网络中不存在的 IP 地址,如本局域网中或互联网上不存在的主机,运行 ping 命令,观察能捕获到什么样的报文。

4. 实验报告

记录实验过程和实验结果,分析 ICMP 回显请求和应答报文的组成,理解 ICMP 查询报文的实现过程和工作特点。

5. 思考

(1) 比较 Windows 和 Linux 系统中 ping 命令的用法的异同。

(2) UNIX 环境下的 ping 程序实现中,采用在发出的 ICMP 回显请求报文里存放发送请求的时间值,再根据应答返回时的时间信息计算 RTT 的方法。Windows 也是这样得到 RTT 的吗? 读者打算如何搞清楚这一点?

实验 5-2 ping 程序和 IP 选项

1. 实验说明

通过运行 ping 程序,指定记录路由或时间戳,在真实网络环境中观察、分析带有 IP 记录路由选项和 IP 时间戳选项的 ICMP 回显请求和响应报文,理解 IP 记录路由选项和时间戳选项的报文格式和工作特点。

2. 实验环境

Windows 或 Linux 操作系统及网络环境(主机有以太网卡并已连接局域网或 Internet),安装有 Wireshark。

3. 实验步骤

步骤 1 在实验计算机上先启动 Wireshark,过滤器可以设置为只查看 ICMP 报文。

步骤 2 在实验计算机上打开命令行窗口,查看 ping 程序的用法,Windows 下执行命令“ping /?”,Linux 下执行命令“man ping”。

步骤 3 选择网络中与实验主机之间有多个路由器的网络主机节点，如互联网上的主机，运行带有记录路由选项的 ping 命令（注意 Windows 和 Linux 下的命令格式不同）。

```
C:\> ping -r n www.sohu.com
```

命令中的 n 取值范围为 1～9，表示选项中要记录的路由数。

查看捕获的回显请求与应答包中的 IP 报文，特别注意观察 IP 首部和选项部分的内容，比较发出的请求报文和对应的应答报文中 IP 选项部分的内容。

步骤 4 选择网络中与实验主机之间有多个路由器的网络主机节点，如互联网上的主机，运行带有时间戳选项的 ping 命令（注意 Windows 和 Linux 下的命令格式不同）：

```
C:\> ping -s n www.sohu.com
```

命令中的 n 取值范围为 1～4，表示选项中要记录的时间戳数。

查看捕获的回显请求与应答包中的 IP 报文，特别注意观察 IP 首部和选项部分的内容，比较发出的请求报文和对应的应答报文中 IP 选项部分的内容。

4. 实验报告

记录实验过程和实验结果，分析 IP 记录路由选项和时间戳选项的报文组成，理解 IP 记录路由和时间戳选项的实现过程和工作特点。

5. 思考

32 位的时间戳数值的表示方法是怎样的？计算实验中获得的时间戳数值所表示的时间。

实验 5-3 ICMP 重定向差错报文

1. 实验说明

利用 GNS3 构建实验虚拟网络，通过捕获路由器转发过程中数据包的分析，观察 ICMP 重定向现象，掌握 ICMP 重定向报文格式和重定向工作原理，理解重定向更新路由表的方式及重定向对网络安全的影响。

2. 实验环境

Windows 操作系统及网络环境（主机有以太网卡并已连接局域网），安装有 GNS3（安装时选择安装 VPCS）和 Wireshark。

3. 实验步骤

步骤 1 在 GNS3 中建立如图 5-12 所示的实验拓扑，图中除标识 IP 地址外，还把几个主要的 MAC 地址作为示例也标注出来，实验时以实际 MAC 为准。图中 C1 为 Virtual PC，启动 VPCS，运行以下命令：

```
VPCS[1]ip 10.1.1.1 /24 10.1.1.2
```

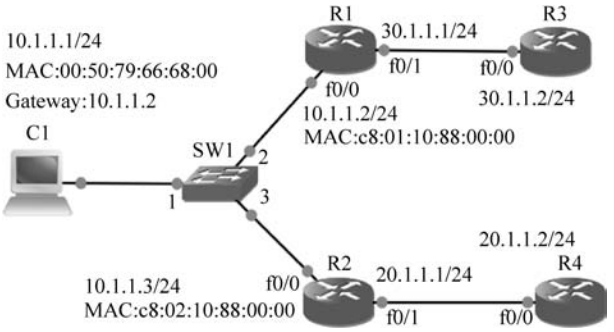

图 5-12 ICMP 重定向实验拓扑

将 C1 的 IP 地址设为 10.1.1.1/24,默认网关指向路由器 R1(10.1.1.2)。

各路由器均选择 Cisco 2621,均配置运行 RIPv2。各路由器的配置过程都相似,下面仅以 R1 为例,说明路由器接口配置。R1 配置命令如下:

```
R1(config)#inte f0/0
R1(config-if)#no shut
R1(config-if)#ip addr 10.1.1.2 255.255.255.0
R1(config-if)#exit
R1(config)#inte f0/1
R1(config-if)#no shut
R1(config-if)#ip addr 30.1.1.1 255.255.255.0
```

配置 RIPv2 命令如下:

```
R1(config)#router rip
R1(config-router)#network 10.1.1.0
R1(config-router)#network 30.1.1.0
R1(config-router)#version 2
```

将 4 个路由器都配置好后,很快各个路由器上均有了路由表信息,可以用 show ip route 命令在各路由器上查看验证。

步骤 2 观察 ICMP 重定向对数据包路由的影响。

实验观察 C1 发往 R4 的数据转发情况。在 R1 的 f0/0、R2 的 f0/0 接口处启动抓包,然后在 C1 的命令提示符下,输入以下命令:

```
VPCS[1]ping 20.1.1.2
```

重定向功能在路由器上是默认启用的,记录在 C1 上看到的输出信息,观察重定向的发生情况。

注意,R1 的 f0/0 处收发的分组链路层帧的 MAC 地址,记录重定向发生前后 C1 上 ping 程序产生的 ICMP 分组的 MAC 地址。

观察由 R1 的 f0/0 发往 R2 的 f0/0 接口的分组的变化情况,注意在 Wireshark 中捕获到的 ICMP 重定向报文的内容。

步骤 3 观察关闭重定向后的数据转发。

运行下列命令,在路由器 R1 的 f0/0 接口关闭 ICMP 重定向。

```
R1(config)#int f0/0
R1(config-if)#no ip redirect
```

再次在C1上运行命令“ping 20.1.1.2”，由于R1已经关闭重定向功能，C1发出的数据将首先经过默认网关再转发，即沿着C1→R1→R2→R4这样的走向。在R1上进行抓包并观察验证。

步骤4 修改图5-12的C1为直接利用路由器在GNS3中模拟主机。

观察用路由器模拟主机、重复上述实验的过程。

路由器在关闭路由功能后将接收ICMP重定向，对路由器的恶意攻击有可能利用这一点来实施。

4. 实验报告

记录实验过程和实验结果，分析ICMP重定向差错报文的组成，理解ICMP重定向的原理和工作特点。认识关闭路由器路由功能时重定向引起的安全问题。

5. 思考

(1) 用VMware虚拟机充当主机进行ICMP重定向，实验将如何实现？

(2) 要设计怎样的网络工作状况才可以观察到其他类型的ICMP差错报文？如主机不可达或端口不可达。

实验5-4 traceroute程序

1. 实验说明

通过捕获traceroute程序工作过程中收发数据包的分析，掌握traceroute程序的工作原理，掌握ICMP超时差错报文格式，理解traceroute程序基于ICMP和UDP的不同实现方式，理解IP源站选路的工作特点和报文格式。

traceroute程序的命令格式与其在不同操作系统中的具体实现有关。

Windows下的tracert.exe命令格式如下：

```
tracert [-d] [-h maximum_hops] [-j host-list] [-w timeout] target_name
```

命令的参数说明如下。

(1) -d：指定不对计算机名解析地址。

(2) -h maximum_hops：指定查找目的的跳转的最大数目。

(3) -j host-list：指定在host-list中宽松源路由。

(4) -w timeout：等待由timeout对每个应答指定的毫秒数。

(5) target_name：目的计算机的名称。

Linux系统中的traceroute命令格式如下：

```
traceroute [options] <IP-address or domain-name> [data size]
```

命令参数说明如下。

(1) [options]的常用选项有：

① -d：使用 Socket 层级的排错功能。

② -f first_ttl：设置第一个检测数据包的存活数值 TTL 的大小。

③ -F：设置不分片位。

④ -g gate：设置源路由网关，最多可设置 8 个。

⑤ -i device：使用指定的网络界面送出数据包。

⑥ -I：使用 ICMP 回应取代 UDP 资料信息。

⑦ -m max_ttl：设置检测数据包的最大存活数值 TTL 的大小。

⑧ -n：直接使用 IP 地址而非主机名称。

⑨ -p port：设置 UDP 传输协议的通信端口(默认为 33 434)。

⑩ -q nqueries：设置测试报文数(默认为 3)。

⑪ -r sendwait：忽略普通的路由表，直接将数据包送到远端主机上。

⑫ -s src_addr：设置本地主机送出数据包的 IP 地址。

⑬ -t tos：设置检测数据包的 TOS 数值。

⑭ -v：详细显示指令的执行过程。

⑮ -w waittime：设置等待远端主机响应的时间。

⑯ -x：开启或关闭数据包的正确性检验。

(2) [data size]：每次测试包的数据字节数(默认为 38)。

还有一些不经常使用的选项，需要了解时可以参考程序提供的帮助信息。

2. 实验环境

Windows 操作系统及网络环境(主机有以太网卡并已连接 Internet)，Wireshark。

3. 实验步骤

(1) 基于 ICMP 的 traceroute 工作过程。

步骤 1 在一台能够连接 Internet 的 Windows 主机上，启动 Wireshark，设置过滤器为 ICMP。

步骤 2 在 Windows 命令行窗口运行下列命令：

```
C:\> tracert www.sohu.com  (也可以选择其他可跟踪跃点的目的节点)
```

观察命令执行过程中输出的跟踪跃点的内容。

步骤 3 对照分析捕获的 tracert 程序发出的 ICMP 回显请求报文、ICMP 超时差错报文、到达目的节点的 ICMP 回显应答报文，注意各 IP 报文的 TTL 时间。

(2) IP 源站选路选项观察。

步骤 1 执行命令“tracert /?”，查看程序的参数选项，了解宽松的源路由的命令格式。

步骤 2 继续上面(1)中的实验，在其命令输出的、去往 www.sohu.com 的路径上，选择两个路由器地址，如 61.139.45.197 和 171.208.202.97，作为指定的源路由(去往目的地的路径可能不唯一，以实际执行命令时得到的信息为准)，执行下列命令：

```
C:\> tracert -j 61.139.45.197 171.208.202.97 www.sohu.com
```

观察命令输出结果。

步骤 3 查看捕获的 ICMP 回显请求报文中 IP 选项的内容，记录选项里 code、ptr 的值，记录选项中 IP 地址的内容。

由于不少路由器对带 IP 源站选路选项的报文都做了限制，因此不一定能够捕获 ICMP 回显应答报文。如果能够捕获，记录并分析其中的上述各项内容的值。

4. 实验报告

记录实验过程和实验结果，分析 ICMP 超时差错报文的组成，分析、理解利用 ICMP 回显请求应答报文和超时差错报文实现 traceroute 的工作原理；分析 IP 源路由选项中 ptr、IP 地址的内容及变化与 traceroute 工作特点的关系。

5. 思考

(1) 如何能够观察到基于 UDP 的 traceroute 实现？

(2) 如何根据带有源站选路的 traceroute 命令的输出，画出到目的节点的拓扑路径？

6. 延伸学习

IP 源路由可以被攻击者利用，以欺骗目的节点将数据报发往本不应该经过的网络，进而被窃听者盗用。

防范 IP 源路由欺骗的方法主要有：配置好路由器，使它抛弃那些由外部网进来的、声称是内部主机的报文；关闭主机和路由器上的源路由功能。

Cisco 路由器关闭源路由功能的命令如下：

```
R(config)#no ip source-route
```

请在课后阅读相关资料，学习防范 IP 源路由欺骗的方法。