基于ICMP协议的指纹探测技术研究

On the Fingerprints Detection Technology Based upon ICMP Protocol

孙乐昌　刘京菊　王永杰　陆余良

（电子工程学院网络工程教研室　合肥230037）

Abstract　To obtain the OS's type and version used by remote host on the Internet is very important on studing the network security. The fingerprints detection technology is an effective solution method to this problem. In this article, the principium, method and actualization of the fingerprints detection technology based upon ICMP protocol are studied especially.

Keywords　ICMP protocol, Fingerprint detection, Request packet, Echo packet, Error packet

准确地获取计算机网络中远程主机的操作系统类型及其版本号是研究计算机网络安全漏洞的一个重要内容，因为大多数的安全漏洞都是和一定的操作系统相联系的。指纹是指能够通过网络连接获取的唯一标识某一具体操作系统的类型及其版本号的一组特征信息的集合。作为指纹的特征信息集必须满足完备性和唯一性的要求。所谓指纹探测技术，就是根据不同类型、不同版本的操作系统在协议栈实现上的细微区别，通过向目标系统发送特定格式的数据包，并从反馈数据包中提取操作系统指纹信息的远程操作系统探测技术。

指纹探测技术基本上可以分为两大类，一类是在传输层实现的基于TCP协议的指纹探测技术；另一类是在IP层实现的基于ICMP协议的指纹探测技术。

基于TCP协议的指纹探测技术主要通过探测TCP包的窗口大小、ACK值、ISN值、BOGUS(伪造)标记等字段的取值特征或者通过向目标机分别发送置FIN位、ACK位等标志位的TCP包，根据反馈的不同指纹特征来鉴别不同的操作系统。

基于ICMP协议的指纹探测技术主要是构造并向目标系统发送各种可能的ICMP报文或者UDP报文，然后对接收到的ICMP响应报文或错误报文的指纹特征进行提取分析，从而识别目标系统的操作系统。本文将重点研究讨论基于ICMP协议的指纹探测技术的原理、方法和技术实现。

1　ICMP协议简介

ICMP经常被认为是IP层的一个组成部分，它通常被IP层或更高协议(TCP或UDP)使用。ICMP报文是在IP数据报内部被传输的，其报文一般格式如图1所示，前面20个字节是IP数据报头，紧跟着是ICMP报文，报文的前4个字节都是一样的，但是剩下的数据则对应不同类型和不同代码的报文。

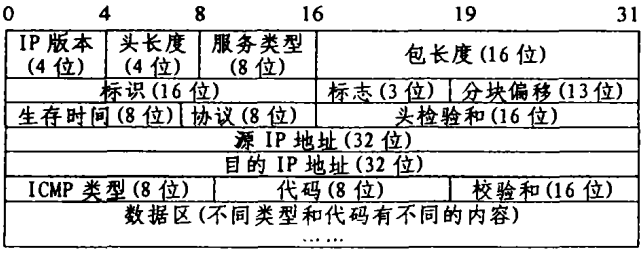

图1　封装在IP数据报内部的ICMP报文格式

ICMP报文的类型域是一字节整数，指出ICMP报文的类型，代码域也是一字节整数，提供关于报文类型的进一步信息；校验域共两字节长，提供整个ICMP报文的校验和。

ICMP报文可以大致分为查询/回应报文和差错报文两大部分。查询报文包括回应请求/应答报文、时戳请求/应答报文、信息请求/应答报文和地址掩码请求/应答报文几种，差错报文包括信宿不可达报文、数据报超时报文和参数错报文等类型。通过发送常规的和刻意构造的ICMP查询报文，然后对返回的ICMP回应报文或ICMP差错报文中的特征值进行分析和归纳，就能够获知目标主机的操作系统类型及其版本号。下面将对这一技术进行详细探讨。另外，在本文讨论的指纹探测技术中，还用到了不少IP数据报头中的字段信息，读者可参考相关的资料。

2　常规的ICMP协议指纹探测技术

2.1　用DF位来识别SUN Solaris，HP-UX10.20和AIX4.3.x

在RFC791中定义了IP数据包头中的三位控制标志。第0位是保留标志，必须为零。第1位是不分片标志，只可取两个值。0代表可以分片，1表示不能分片。若此标志被置位，则IP层的包分片将不被允许，反之亦然。第2位是片未完标志。它也可取两值，0表示此为最后一个分片，1表示后面还有分片将到达。

IP数据包头中的下一字段是分片偏移字段，指出本片数据在初始数据报数据区中的偏移量，偏移量以8个字节为单位。

经测试发现一些操作系统在他们的回应报文里将置位分片标志。

下面是用Tcpdump截获的Sun Solaris2.7对ICMP回应请求报文的响应信息。

可见，Sun Solaris在应答报文中默认把DF标志置位。但HP-UX10.30&11.0x和AIX4.3x对连续发送的询问报文，其应答报文中的DF标志将发生变化，有可能在第一个应答报文中，其DF位就被置为1，也有可能在数个应答报文发送后才置DF位。分析这些细节将有助于区分Sun Solaris，HP-UX10.30&11.0x和AIX4.3x等操作系统。

HP-UX10.30&11.0x和AIX4.3x等操作系统之所以有DF位的设置区别，是因为它们在利用ICMP回应请求报文来进行PMTU发现。

```
17:10:19.538020 if 4>y.y.y.y>x.x.x.x : icmp: echo request (ttl 255, id 13170)
                   4500 0024 3372 0000 f f 01 9602 yyyy yyyy
                   xxxx xxxx 0800 54a4 8d04 0000 cbe7 bc39
                   8635 0800
17:10:19.905254 if 4<x.x.x.x<y.y.y.y :icmp: echo reply (ttl 233, id 24941)
                   4500 0024 616d 4000 e901 3e07 xxxx xxxx
                   yyyy yyyy 0000 54a4 8d04 0000 cbe7 bc39
                   8635 0800
```

若发送一个 ICMP 回应请求报文给一台 HP-UX 11.0计算机，其第一个应答报文很正常，DF 未被置位，但是接着 HP-UX 11.0计算机又主动发送了一个 ICMP 回应请求报文，并置了 DF 位，包的大小为1500字节，这是以太网络的最大传输单元 MTU。在途经的路由中，MTU 小于1500的路由器将向 HP-UX 11.0计算机回应 ICMP 出错报文，从而使它根据 PMTU 的发现算法发送长度短一些的数据包，直到接收到 ICMP 回应响应报文，说明 PMTU 值已经被确定。

这是一种非常简单的指纹探测技术，它不需要附加任何特别的模式设置。这种指纹探测技术的使用也有一定的局限性，因为对于 Sun Solaris 和 HP-UX 类操作系统可以通过改变配置选项来禁止使用 ICMP 请求响应报文中的 DF 位。这样 HP-UX10.30&11.0x 将不允许通过 ICMP 请求响应来实现路径 MTU 发现。在此情况下，利用 DF 位的指纹探测技术将失效。

2.2 用 TTL 字段来进行指纹探测

IP 数据报中的 TTL 值表示数据包在网络中的最大生存时间，它设置了数据报可以经过的最多路由器数。TTL 的初始值由源主机设置(通常为32或64)，一旦经过一个处理其路由器，其值就减去1。当该子段的值为0时，则此数据报将被丢弃，并发送 ICMP 报文通知源主机。

在 ICMP 请求报文和 ICMP 应答报文中都存在 TTL 字段。使用 TTL 字段值有助于识别或分类某些操作系统，而且提供了一种最简单的主机操作系统识别准则。

(1)ICMP 回应应答报文中的 TTL 字段　各类操作系统 ICMP 回应应答报文中的 TTL 值都有不同的设置：

1)UNIX 类操作系统在 ICMP 回应应答报文中使用255作为 TTL 字段值；

2)Compaq Tru64 5.0和 LINUX2.0.x 比较例外，他们在 ICMP 回应应答报文中把 TTL 值设为64；

3)Window 类操作系统的 TTL 值设为128；

4)Windows 95是唯一在 ICMP 回应应答报文中把 TTL 字段置为32的操作系统。

(2)ICMP 回应请求报文中的 TTL 字段　和 ICMP 回应应答报文中的 TTL 字段类似，ICMP 回应请求报文中的 TTL 字段也可用来进行指纹探测。

1)内核为2.0.x，2.2.x 和2.4.x 的 LINIX 在 ICMP 回应请求报文中使用64作为 TTL 字段值；

2) FreeBSD4.1，4.0，3.4；Sun Solaris2.5.1，2.6，2.7，2.8；OpenBSD 2.6，2.7，NetBSD 和 HP UX 10.20在 ICMP 回应请求报文中把 TTL 值设为255；

3) Window95/98/98SE/ME/NT4 WRKS SP3，SP4，SP6a/NT4 Server SP4等操作系统都把 TTL 值设为128；

4)Windows2000在 ICMP 回应请求报文中把 TTL 字段置为128。

(3)综合探测　综合上述两种指纹探测技术，可以明确地区分如表1所示的几种常用的操作系统。

表1　根据 ICMP 回应请求报文和回应应答报文中的 TTL 字段可区分的操作系统列表

操作系统	ICMP 回应请求报文	ICMP 回应应答报文
Microsoft Windows 类	32	128
*BSD 和 Solaris	255	255
Linux2.2.x&2.4.x	64	255
Linux 2.0.x	64	64
Microsoft Windows2000	128	128
Microsoft Windows95	33	32

2.3 用分片的 ICMP 地址掩码请求来识别操作系统

如上述表1所示，某些操作系统将对 ICMP 地址掩码请求有响应，那些操作系统包括 ULTRIX OpenVMS，Window95/98/98SE/ME，HP-UX11.0和 Sun Solaris。为了进一步区分它们，可以采用分片的 ICMP 地址掩码请求指纹探测技术，其实现过程如图2所示：首先发送分片数据包，其中 IP 数据部分只

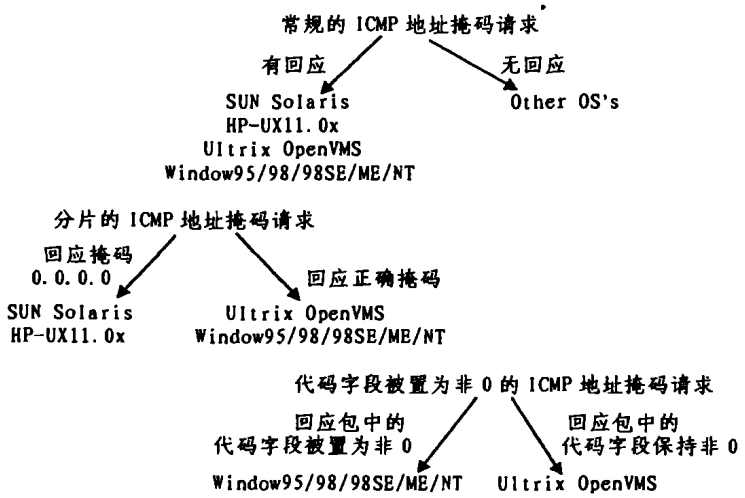

图2　使用 ICMP 地址掩码请求进行指纹探测

有8字节，则 SUN Solaris 和 HP-UX 将以0.0.0.0作为地址掩码返回，而其他几种操作系统则以真实的地址掩码返回；然后再向以真实的地址掩码应答的操作系统发送 ICMP 报头中的代码字段被置为非0的地址掩码请求包，则 Ultrix OpenVMS 仍以非0的代码字段返回，而 Window95/98/98SE/ME/NT 则将代码字段置0。

3 非规则的 ICMP 协议指纹探测技术

ICMP 报文中 TOS 字段格式如下：

Precedence	TOS	MBZ

每个 IP 数据包有一个8字节的服务类型(TOS)字段，它包括一个3 bit 的优先权子字段，4 bit 的 TOS 子字段和1 bit 未使用但必须置0的"MBZ"位。4 bit 的 TOS 分别代表：最小时延、最大吞吐量、最高可靠性和最小费用。4 bit 中只能置其中1 bit，若所有4 bit 均为0，那么就意味着是一般服务。

3.1 用优先权子字段进行指纹探测

AIX4.3等大多数操作系统在 ICMP 回应应答报文中仍然使用 ICMP 回应请求报文中的优先权子字段(Precedence)值，但是另外一些操作系统如 Windows 2000和 ULTRIX 在 ICMP 回应应答报文中将优先权子字段置为0。而 HPUX11.0 的处理方式更加特别，在第一个应答报文中保持优先权子字段值不变，随后发送一个1500字节大小的 ICMP 回应请求包，用于 PMTU 发现，以后所有的响应包的优先权子字段都被置为0，这与上述用 DF 位来识别 SUN Solaris，HP-UX10.20和 AIX4.3.x 时所发生的现象相似。图3描述了上述的探测过程。

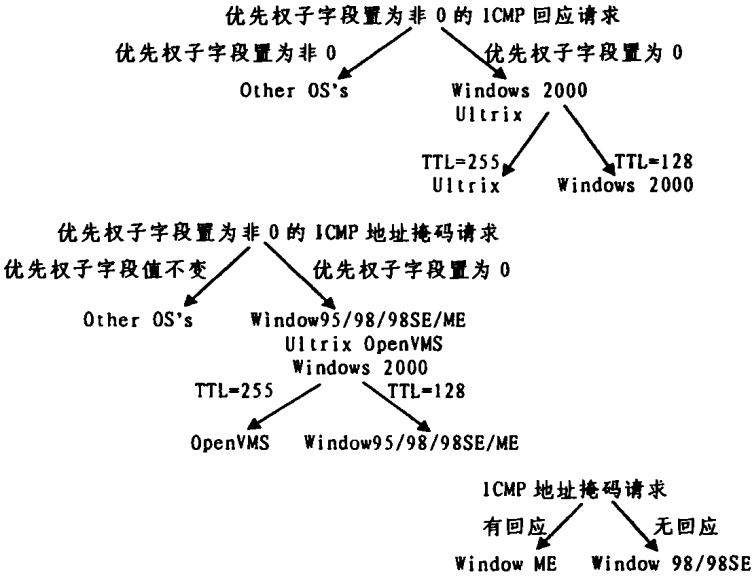

图3 使用优先权子字段为非0的 ICMP 回应请求进行指纹探测

3.2 用 TOS 子字段进行指纹探测

服务类型字段中的4 bit 的 TOS 子字段取值及含义如表2所示。

表2 TOS 子字段取值列表

TOS 值	服务	TOS 值	服务
0000	一般	0100	最大吞吐量
1000	最小延时	0010	最高可靠性
0001	最小费用	1111	最高安全

RFC 1349中定义了 ICMP 报文中 TOS 域的用法。在 ICMP 差错报文、ICMP 请求报文和 ICMP 回应报文中 TOS 的用法都有区别。简单的规则如下：

·ICMP 差错报文的 TOS 默认值为0x00

·ICMP 请求报文的 TOS 可取任意值

·ICMP 回应报文的 TOS 值与 ICMP 请求报文的 TOS 值一致

通过实验，我们发现 Windows 2000、Ultrix 和 Novell Netware 并未遵守此规则，而是在 ICMP 回应报文中把 TOS 置为0。据此就可以区分几种操作系统，实现过程如图4所示。

3.3 用 MBZ 子字段进行指纹探测

RFC1349规定"MBZ"位不被使用，且必须为0，路由器和主机将忽略这一位。若收到"MBZ"位被置为1的 ICMP 回应请求报文，大多数操作系统的回应应答报文中"MBZ"位仍然为1。而 Windows 2000和 ULTRIX 则将"MBZ"位置为0。

4 ICMP 差错报文指纹探测技术

当发送一份 ICMP 差错报文时，报文数据区包含出错数据报 IP 首部及产生 ICMP 差错报文的 IP 数据报的前8个字节，这样，接收 ICMP 差错报文的模块就会把它与某个特定的协议和用户进程联系起来。几乎所有实现只送回 IP 请求头外加8 字节。然而，Solaris 送回的稍多，而 Linux 更多。

另外，某些操作系统在实现 ICMP 出错报文时，可能会改变所引用的出错数据包的原 IP 报头。

若向运行 AIX 4.2.1的机器未被打开的 UDP 端口发送 UDP 数据包，将监听到如下的数据包：

```
12:33:17.319275 ppp0>x.x.x.x.2160>y.y.y.y.0: udp 0 [tos 0x10] (ttl 64, id 47349)
                4510 001c b8f5 0000 4011 9bea xxxx xxxx
                yyyy yyyy 0870 0000 0008 d18c
12:33:17.614823 ppp0 < y.y.y.y > x.x.x.x: icmp: y.y.y.y udp port 0 unreachable offending pkt: x.x.x.x.2160 >
```

```
y.y.y.y.0: udp 0 [ttl ox10] (ttl 49 , id 47349, bad cksum area!)[tos 0x10] (ttl 241, id 17965)
                4510 0038 462d 0000 f101 5da6 yyyy yyyy
                xxxx xxxx 0303 f470 0000 0000 4510 0030
                b8f5 0000 3111 aaea xxxx xxxx yyyy yyyy
                0870 0000 0008 0000
```

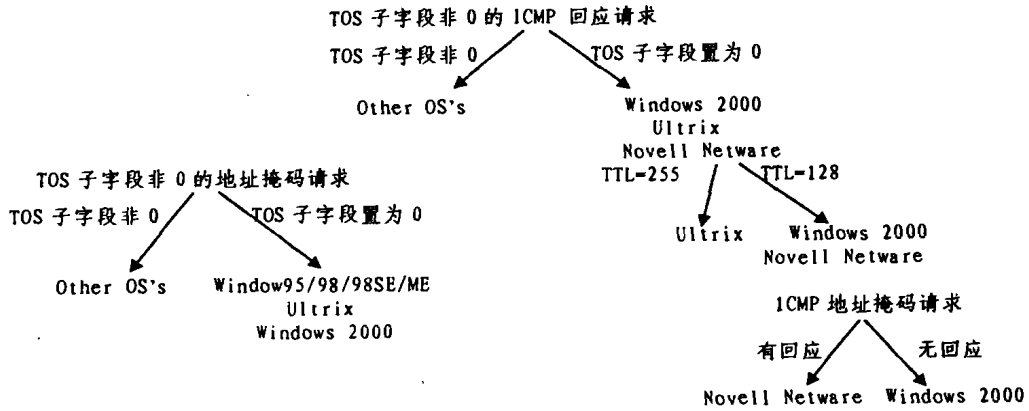

图4 使用 TOS 子字段非0的 ICMP 回应请求等进行指纹探测

从上面的数据包中可以看出原始 IP 报头和 ICMP 端口不可达出错报文数据部分中 IP 报头不同的地方：

1)IP 报头16位的总长度字段 原 UDP 数据报大小为28字节，但回应包引用的出错数据部分 IP 报头中的总长度字段值却为48，多了20个字节。

2)TTL 字段值 回应包引用的出错数据部分 IP 报头中的 TTL 字段值为49，比原 UDP 数据报中的 TTL 值小15，说明中间经过了15跳点。

3)部校验和字段 因为总长度字段和 TTL 字段都变化了，回应包引用的出错数据部分 IP 报头中的首部校验和字段值自然也发生了改变。

4)UDP 首部校验和字段 回应包引用的出错数据部分前64比特数据中的 UDP 首部校验和字段等于0。

在实验中还发现 AIX 4.1在返回的数据包中除了上述第四个不同点外，其余3个不同点都存在。根据这些特征就可以区分不同版本的 AIX 操作系统。

向运行 BSDI 4.x 的机器关闭的 UDP 端口发送 UDP 数据包，监听数据包后发现不同点：

1)IP 报头 16 位的总长度字段 原 UDP 数据报大小为28字节，但回应包引用的出错数据部分 IP 报头中的总长度字段值却为48，多了20个字节。

2)IP TTL 字段值 回应包引用的出错数据部分 IP 报头中的 TTL 字段值为53，中间经过了64-53＝11跳点。

3)IP 首部校验和字段 回应包引用的出错数据部分 IP 报头中的 IP 首部校验和字段等于0。其他操作系统也有各自的特征。

结束语 指纹探测技术是一项非常复杂的技术，它需要了解和掌握各种操作系统之间的细微特征差异，并能够远程捕获这些差异。遵循 ICMP 协议可以构造出种类众多的报文格式，恰好满足了指纹特征多样性的要求，而且各种操作系统的 ICMP 协议栈在实现上确实存在一定的差异。因此，基于 ICMP 协议的指纹探测技术一定会有广阔的发展空间。

参 考 文 献

1 Stevens W R. TCP/IP 详解 卷1：协议. 机械工业出版社，2000

2 Arkin O. Understanding some of the ICMP Protocol's Hazards. 2000

3 Comer D E. 用 TCP/IP 进行网际互联(第一卷). 电子工业出版社，1998

(上接第48页)

表1 算法在不同规模下的平均精度 p×100

训练集大小	平均精度 p×100				
(每类篇数)	KNFL	KNN	NFL	NN	NB
10	72.45	67.34	60.87	45.89	70.43
20	74.64	69.50	70.34	50.56	75.68
30	76.37	75.42	66.45	63.43	78.53
50	80.45	77.56	69.59	68.57	79.23
90	82.57	77.95	67.53	69.54	80.34

结论及今后工作 kNFL 结合了 kNN 与 NFL 的优点，扩大了训练集规模，改进了平均精度，实验表明对于高维样本和训练规模不大时，kNFL 有明显的优势。kNFL 是一种正确率较高的、简单的分类算法。kNFL 和 NFL 可以应用于更多的模式识别的领域，特别是向量空间具有明显连续性的模型中。算法的通用性和应用价值需要进一步在实验中检验。

参 考 文 献

1 边肇祺. 模式识别. 北京：清华大学出版社，第二版，1999. 136～161

2 Dasarathy B V. Nearest Neighbor(NN) Norms: NN Pattern Classification Techniques. McGraw-Hill Computer Science Series. IEEE Computer Society Press, Las Alamitos, California, 1991

3 Lam W, Ho C Y. Using a generalized instance set for automatic text categorization. In: Proc. of the 21th Ann Int ACM SIGIR Conf. on Research and Development in Information Retrieval, 1998. 81～89

4 Vapnic V. The Nature of Statistical Learning Theory. Springer, New York, 1995

5 Rennie J, McCallum A. Efficient Web Spidering with Reinforcement Learning. ICML-99, 1999

6 Li S Z, Lu J. Face Recognition Using the Nearest Feature Line Method. IEEE Trans. Neural Networks, 1999, 10(2): 439～443

7 Wong S G A, Yang C S. A vector space model for automatic indexing. Communications of the ACM, 1975, 18(11): 613～620