第 3 章　网络安全问题的产生

本章探讨互联网安全为什么会有这么多问题，其中互联网体系架构和软件系统脆弱性是网络安全一内一外的两大重要原因。

3.1　互联网的巨大成功——必要条件

基于 TCP/IP 技术的互联网获得了巨大成功，战胜了通信工业提出的 ATM\X.25 等其他网络技术方案成为事实的计算机网络标准，成为人类社会的基础信息网络框架。同时，WWW 的兴起为互联网的内容提供了一种标准的媒体描述方式，使得互联网不仅作为计算机网络而存在，同时具有丰富的媒体内容资源，促进了互联网爆炸式的增长。在此基础上，随着信息化的浪潮，人类社会的社会行为纷纷迁移到了互联网上，形成了电子政务、电子商务、电子金融、电子教学等，使得互联网的使用成为人类的生活习惯，互联网安全问题随之受到大家的关注，以致成为热点问题。

归结互联网的成功(参考互联网基本原理章节)，互联网技术作为传统电信网络的颠覆者，是一个开放的平台，从技术发展角度去考虑设计问题，提供了丰富的媒体内容，顺应用户的需求，顺应新事物由弱到强、由小到大的发展过程。主要因素如下。

(1) 从不可靠的网络出发，构建可靠的网络。

(2) 从互信、未考虑安全因素出发，构建安全网络。

(3) 从对等角度，构建大规模系统。

表 3-1 给出了互联网与电信网络的对比说明。

表 3-1　电信网络与 Internet 的比较

	电 信 网 络	Internet
设计用途	话音传送为主	数据传输网络
服务内容	电话、传真、短信	WWW Web、多媒体等
设计目的	商业化、可运营化	健壮、抗毁、自愈
商业模式	收取通信费用	初期无商业化、多元化
设计要求	保证通信质量	尽力传送，保证互连互通
结构模式	同质，从上到下，集中管理	异质、对等、分布式、自治
发展模式	集中式架构，集中规范，集中升级	代议制/工程师联盟/IETF 管理架构和方式
网络终端	固定通话，终端地址	计算机
设计理念	网络设计复杂，终端功能简单	网络简单，尽量把复杂功能留给终端

是否需要重新设计 Internet？答案是肯定的，也就是设计下一代互联网或者未来网络。

目前 IPv6 应用越来越广阔，事实上 IPv6 只是新一代互联网的一种而已。

如何设计呢？这主要是架构与工程的问题，理论问题不多。事实上，从 2010 年开始，各国都兴起了新的网络设计热潮，美国国家自然科学基金(NSF)在 2010 年一次资助了 4 个未来的互联网研究项目，以解决当前互联网存在的网络安全、移动性、应用僵化和可管理性等问题。我国在未来互联网研究中投资巨大，目前，清华大学已经和 NDN 项目组进行合作，以期望在该领域有所斩获。

其中最新潮的解决思路称为信息中心网络(Information Centric Networking，ICN)，强调信息内容的互连。这其中最引人注目的是命名数据网络 NDN/CCN (Named Data Networking/ Content Centric Networking)，由加州大学洛杉矶分校(UCLA)计算机科学系张丽霞教授领导 12 个学校联合开发。

3.2 计算机系统的安全漏洞

当前随着计算机系统的功能越来越丰富，系统的规模越来越庞大，软硬件系统自身的健壮性由于系统的复杂性而降低，造成系统存在很多安全“漏洞”。在软件开发过程中，开发的复杂性通过编程语言、编程类库、编译工具和操作系统调用而大大增加了安全漏洞的产生。同时在商业化的竞争压力下，在系统的性能功能与安全性之间的平衡中，系统开发对安全功能的重视不够，因为安全的代价是以大大牺牲性能为代价。软件系统漏洞增长趋势如图 3-1所示。

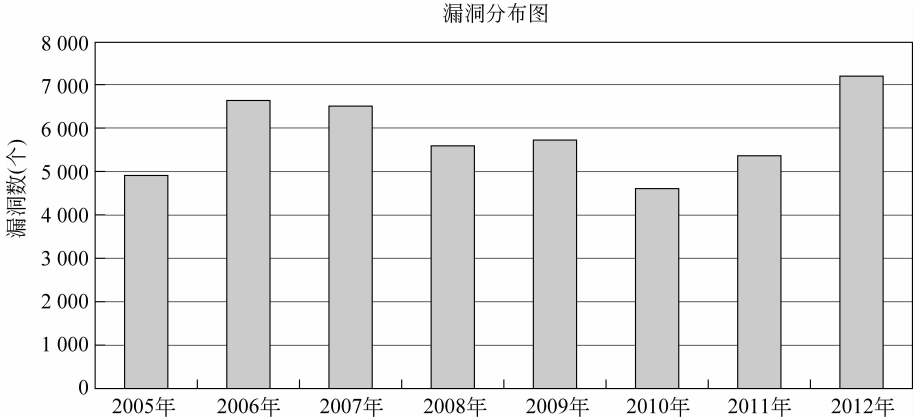

图 3-1 软件系统漏洞增长趋势

安全与用户友好性是一个比较大的矛盾：一个典型的案例是微软公司的 Windows XP，因其简单易用成为大众网吧的首选操作系统，虽然其安全性比较低，但易用性极好；相反，Windows Vista 系统吸收了 Windows XP 的安全问题，却因为重视安全性而导致易用性和性能的下降，大大减少了用户的接受程度。因此，微软公司不得不在 2009 年推出了 Windows 7，通过降低 Vista 系统的安全级别，以获得用户对其易用性的认可。从这个例子可以看出，互联网安全有其根本性的原因(人的因素)。

3.3 互联网的安全性

随着计算机网络技术的不断普及，尤其是互联网成为信息基础框架后，全世界的计算机采用 TCP/IP，通过路由器、交换机而连接起来，形成了全球互连互通的网络。这样，原本是单个计算机系统的安全漏洞通过网络的互连效应被大大地放大了。

更大的问题是，Internet 是一个分布式自组织的网络，其体系架构中缺乏统一的安全和管理框架。网络环境造成了攻击成本低，造成危害大的不利局面。比如动态 IP 地址分配、IP 路径重配置、NAT(网络地址转换)、P2P 覆盖网络等技术使得互联网安全事件的审计和追踪非常困难，安全攻击追踪受制于管理域问题而不能得到应有的“惩戒”。

3.4 互联网安全的“黑金”现象

目前很多病毒，如蠕虫(Worm)、木马(Torjan)和机器人网络(Botnet)的背后都有地下经济的黑影。互联网安全事件从单纯的“找乐”(For Fun)、显示“个人成就”等动机，迅速转化为受经济利益驱动的具有黑金性质的地下产业。制作病毒，传播病毒，控制机器人网络，散发兜售广告邮件(SPAM)成为一条完整的地下经济产业链。

加州大学伯克利分校的 Vern Paxson 教授研究表明，通过控制机器人网络来散发兜售广告邮件，是成本收益比非常低的销售手段，从而为机器人网络扩散提供了很强的经济动机。

此外，通过控制机器人网络，对政府、商业等网站进行分布式拒绝服务(Distributed Denial of Service，DDoS)攻击，表达政治诉求或者谋求经济利益，也是目前互联网安全事件频发的因素之一。

第 4 章　互联网的基本原理

前面提到了互联网的安全性问题，为了进一步说明其深层的原因，下面对其基本原理进行简要介绍。

4.1　互联网的定义

解决人人之间的通信问题，最基本的手段是给每对人之间都拉上专线光纤，形成点对点网络，如图 4-1 所示。但是因为经济成本的问题，这样做是不可行的，所以需要设计一个接入网络，人人先接入局部网络（或叫局域网，对于世界范围，局部也是相对的），然后由路由器（Router）将这些小网络连接起来，形成更大的世界范围的网络，这就是互联网的端到端的基本思想（End-to-End），如图 4-2 所示。

图 4-1　点对点网络

图 4-2　端到端的跨网络相连

实际的互联网拓扑图如图 4-3 所示。

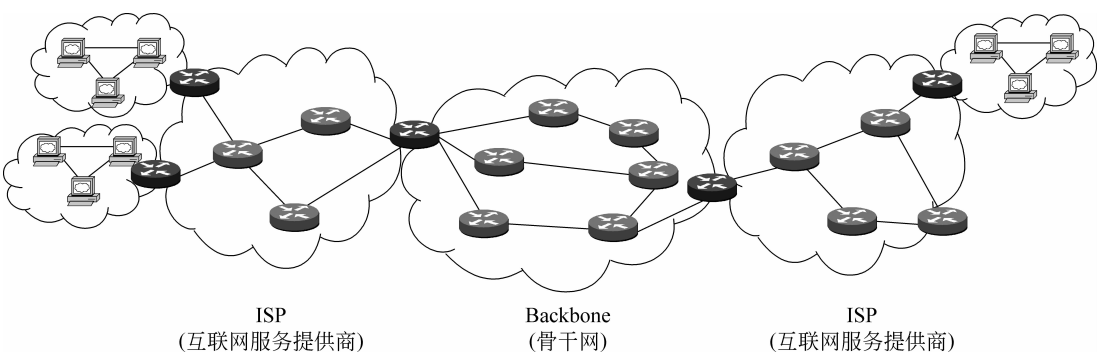

图 4-3　实际的互联网拓扑图

互联网采用 TCP/IP，由 Vinton G. Cerf 和 Robert E. Kahn 于 20 世纪 70 年代设计，并于 1980 年 9 月公布为 IETF RFC 791 标准。TCP/IP 为不同的数据网络（局部网络）提供了一套标准的协议，使得不同的数据网络按照这一套标准语义和语法，能够建立路由和通信。为此，TCP/IP 设计者获得了 2004 年的 ACM 图灵奖。

互联网的设计与人类社会运行几千年的邮政系统是一致的。正如书面信函一样，只要将信函写好，装上信封，写上收信人地址和寄信人地址，投入邮局的邮箱就可以了。互联网采用 TCP/IP 一样的运作模式。

首先，互联网采用 IP 协议通过编址为全世界的计算机都分配一个 IP 地址（固定地址，类比于收信人和寄信人地址），IP 地址中包含了该计算机所在区域号（即 IP 地址的网络地址，Network Address，类比于邮政编码 Postal Code），以及在该区域内的详细地址，即主机地址（Host Address）。

其次，用户要发送的消息分解成许多数据包（又叫做网包，Packet，类比于一封封信函），送给的第一个路由器叫网关（也叫默认网关，Default Gateway，类比于邮局邮箱），路由器先本地保存这份数据，然后等待机会发送到下一跳的路由器上，这样一跳一跳地路由（这种模式叫做存储转发，类比于邮政职工将信件从一个邮局送往下一个邮局）。

如何确定信件的下一个路由器需要 IP 网络的路由系统（Routing System）来确定。路由系统确定了从 A 点到 B 点路由过程中的传输路径，它是通过在每个路由器上维护一张路由表来实现。每个路由器独立维护自己的路由表。每个路由器都会主动向邻居路由器通报自己的路由表信息，根据相邻的路由器的路由表来更新自己的路由表。

当一个网包到达之后，通过查找路由表信息，就可以将这份数据传送到下一个路由器上，最终到达目的计算机上，在目的计算机上进行数据重组，呈现给用户。

这样就完成了信息从 A 点计算机转移到 B 点计算机的过程。

4.2 互联网的人机接口——域名系统

互联网通过所有计算机遵守 IP 协议完成了信息在 A 点计算机和 B 点计算机之间的数据传送。

这时专门用于提供信息内容的计算机出现了。但是信息的内容如何显示与创建、如何发布需要规范，这就是 W3C 的 WWW 的标准。浏览器是显示这些标准格式内容的软件。

随着浏览器的普及，以 W3C 标准的 HTML 的 Web 内容越来越多，这时提供这些 Web 内容的网站也越来越多，如 Google、Sina、Sohu 等门户网站。怎么搜索定位这些内容的计算机就成了问题，如何将 sina.com 映射成 IP 地址，这就有了域名系统（Domain System），而存储这些域名与 IP 地址映射关系的机器就是域名服务器。

域名系统是互联网的人机接口，为每个提供内容的计算机确定一个可读的名字。互联网的内容通过域名系统就能够精确定位，从而形成了一个提供内容和服务的互联网，这才是互联网的最重要的用处，也就是“上网”的意义。

4.3 互联网的交通运输——路由系统

IP 网络的每个网包因为具有独立的发信人地址和寄信人地址，因此路由器对这样的网包进行独立路由选择，互联网的信息选路和信息传送是同步进行的。

首先，IP 网络中的每个遵循 IP 协议的机器都是一个路由器，根据路由表来转发数据，如果路由表项的下一条是自己，则说明自己是这个信息的接收者，否则的话，就尽力把信息传送给其他路由器，如默认网关路由器或者其他相邻路由器。

互联网的路由系统也是分布式的，按照网络管理域分为域间路由和域内路由。分别采用不同的路由协议来更新每个路由器的路由表。

对于用户而言，互联网的路由系统是不可见的，用户只要配置好自己的默认路由器就可以实现网络服务的接入，因此是没有网络接入访问控制的。

4.4 IP 协 议

IP 协议尽力而为传送数据到达指定地址，不确认数据是否正确到达，是一种无连接(Connectionless)的协议。

1. IP 地址

当前 IP 协议版本 4(Internet Protocol Version 4，IPv4)采用 IP 地址辨别互联网上设备或主机，是唯一的识别码，具有全局效应，适用于不同网络间寻址。

IP 地址由 32b 组成，理论上讲可有 $2^{32}=4\ 294\ 967\ 296$ 个地址。

IP 地址采用分层结构，由网络地址(Network Address)和主机地址(Host Address)两部分组成。依照网络地址和主机地址的分配，分成 5 类地址，即 A 类、B 类、C 类、D 类和 E 类地址，如图 4-4 所示。分成 5 类地址主要原因是网络的规模(内部的计算机数目)多种多样。

<table>
<tr><th></th><th colspan="2">1st字节</th><th>2nd字节</th><th>3rd字节</th><th>4th字节</th><th>主机数</th></tr>
<tr><td>类A</td><td>0</td><td>网络地址</td><td colspan="3">主机地址</td><td>16 777 216</td></tr>
<tr><td>类B</td><td colspan="2">10</td><td colspan="2">网络地址</td><td>主机地址</td><td>65 536</td></tr>
<tr><td>类C</td><td colspan="2">110</td><td colspan="3">网络地址</td><td>主机地址</td><td>256</td></tr>
<tr><td>类D</td><td colspan="2">111</td><td colspan="4">多播</td><td>多播</td></tr>
<tr><td>类E</td><td colspan="2">1111</td><td colspan="4">保留</td><td>保留</td></tr>
</table>

图 4-4　IP 地址的分类

清华大学的 IP 地址为 59.66.0.1/16(A 类)和 166.111.0.1/16(B 类)。采用 ipconfig 命令可以有效确定 IP 地址，信息如下。

```
C:\Users\zhenchen>ipconfig
Windows IP 配置
以太网适配器 Local Area Connection:
  连接特定的 DNS 后缀 . . . . . : tsinghua.edu.cn
  IPv4 地址 . . . . . : 166.111.137.197
  子网掩码 . . . . . : 255.255.255.0
  默认网关. . . . . : 166.111.137.1
```

2. 子网掩码

如何获取 IP 地址的网络地址呢？这个问题由子网掩码(Subnet Mask)来解决，如图 4-5 所示。

3. 分割子网

将一较大的网络区段切割成几组较小的网络，使用于内部路由选择协议(Interior Routing Protocol)进行路由交换。

例如，B 级网络 59.66.0.0/16 可拆分成 256 个较小网络。

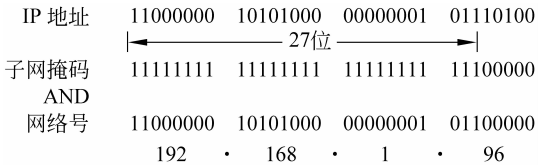

图 4-5 子网掩码

59.66.0.0 / 24
59.66.1.0 / 24
59.66.2.0 / 24
⋮
59.66.255.0 / 24

4. IP 协议数据格式

IP 协议头就是普通书面信函的信封，其格式如图 4-6 所示。信封内装的信纸，就是这里的数据。

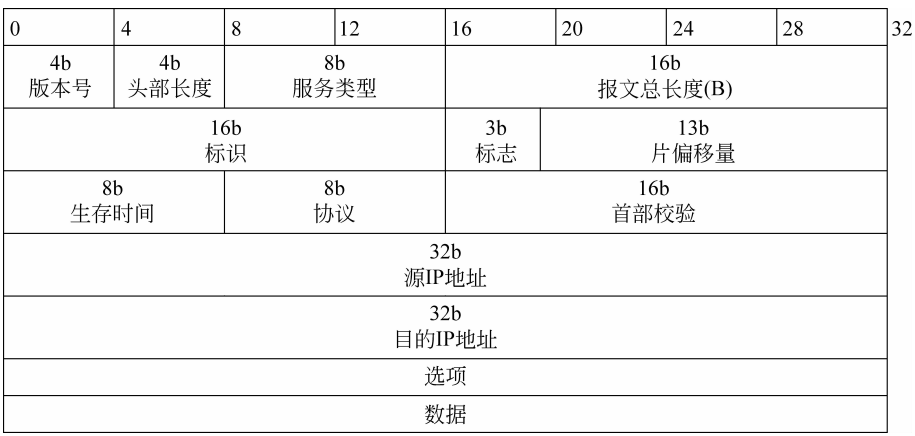

图 4-6 IPv4 协议头

为什么寄信人地址要写在收信人之前呢？因为 IP 协议是美国人设计的，大家可以注意一下西方的普通书信格式是寄信人在前，收信人在后，体现对收信人的礼貌，因此这种设计就是文化传统的问题，如果让中国人来设计的话，可能就会把这两项给颠倒过来。

4.5 ICMP 协议

互联网的运作离不开故障的管理，ICMP（Internet Control Message Protocol，互联网控制消息协议）设计用来处理网络设备之间的错误，报告网络环境中错误状态的发生，但是无法确保能把消息确实送达或是转回发送地。

ICMP 的一个应用是 Ping，用来测试机器之间的连通性，如下例所示。

```
C:\Users\zhenchen> ping www.tsinghua.edu.cn
正在 Ping www.d.tsinghua.edu.cn [2001:da8:200:200::4:100] 从 2001:da8:200:900e:2
```

```
00:5efe:166.111.137.197 具有 32 字节的数据:
来自 2001:da8:200:200::4:100 的回复: 时间<1ms
来自 2001:da8:200:200::4:100 的回复: 时间<1ms
来自 2001:da8:200:200::4:100 的回复: 时间<1ms
来自 2001:da8:200:200::4:100 的回复: 时间<1ms
2001:da8:200:200::4:100 的 Ping 统计信息:
  数据包: 已发送=4,已接收=4,丢失=0 (0% 丢失),
往返行程的估计时间(以毫秒为单位):
  最短=0ms,最长=0ms,平均=0ms
```

ICMP 的另外应用是 traceroute(UNIX 下)和 tracert(Windows 下),如下面的 tracert 一例。

```
C:\Users\zhenchen> tracert www.berkeley.edu
通过最多 30 个跃点跟踪
到 www.w3.berkeley.edu [169.229.131.81] 的路由:
1     <1 ms    <1 ms    <1 ms  SECURITY [192.168.128.1]
2      1 ms    <1 ms    <1 ms  166.111.137.1
3     <1 ms    <1 ms    <1 ms  tu128098.ip.tsinghua.edu.cn [166.111.128.98]
4      1 ms    <1 ms    <1 ms  tu128101.ip.tsinghua.edu.cn [166.111.128.101]
5      1 ms     1 ms    <1 ms  th004133.ip.tsinghua.edu.cn [59.66.4.133]
6      4 ms     1 ms     1 ms  118.229.2.10
7      1 ms     1 ms     1 ms  118.229.2.14
8      1 ms     1 ms     1 ms  118.229.2.2
9     10 ms    11 ms    11 ms  th002237.ip.tsinghua.edu.cn [59.66.2.237]
10    11 ms    11 ms    10 ms  pku0.cernet.net [202.112.38.73]
11    11 ms    11 ms    11 ms  202.112.53.169
12     3 ms     3 ms     3 ms  202.112.61.158
13    12 ms    11 ms    11 ms  202.112.53.18
14   100 ms   100 ms   100 ms  tpr5-ae0-25.jp.apan.net [203.181.194.125]
15   215 ms   214 ms   215 ms  losa-tokyo-tp2.transpac2.net [192.203.116.145]
16   206 ms   205 ms   206 ms  cenichpr-1-lo-jmb-702.lsanca.pacificwave.net
                               [207.231.240.129]
17   214 ms   213 ms   214 ms  svl-hpr--lax-hpr-10ge.cenic.net [137.164.25.13]
18   215 ms   215 ms   215 ms  oak-hpr--svl-hpr-10ge.cenic.net [137.164.25.9]
19   217 ms   215 ms   216 ms  hpr-ucb-ge--oak-hpr.cenic.net [137.164.27.130]
20   225 ms   225 ms   225 ms  t2-3.inr-202-reccev.Berkeley.EDU [128.32.0.39]
21   222 ms   217 ms   216 ms  t1-1.inr-211-srb.Berkeley.EDU [128.32.255.43]
22   216 ms   217 ms   216 ms  webfarm.Berkeley.EDU [169.229.131.81]
跟踪完成。
```

4.6 应用协议端口 Port

IP 使得互联网上的任何两台计算机之间能够建立尽力而为的通信,但是当计算机上的多个通信程序同时使用网络时,还需要在 IP 上设计数据传输协议,并引入端口的概念来区

分不同通信应用程序。

IP 网络类似于邮政系统，如在信件传递的基础上，邮政系统又推出了快件、加急快件服务（如 EMS）、平信和挂号信等服务。类比于互联网，互联网的 TCP 和 UDP 就是在普通信件服务上增加的这些专用的服务，以保证不同业务的服务要求。

IP 之上的数据传输协议主要包括 UDP 和 TCP。应用程序在互相通信时需要选择数据传输协议 UDP 或者 TCP，以及相应的传输端口，以区别其他应用程序，如图 4-7 所示。

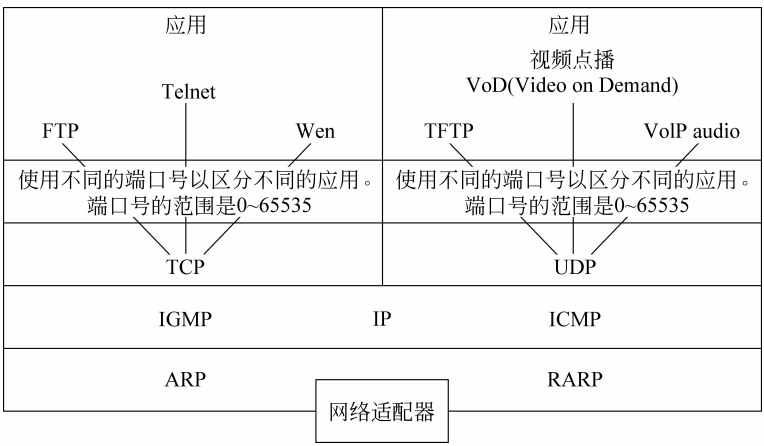

图 4-7　TCP/IP 协议栈

4.7　UDP

讲究传输速度且允许一点数据丢失（Data Loss）或是数据包丢失（Packet Loss）的常用传输协议是 UDP（User Datagram Protocol）。例如 GTalk、MSN、Skype、QQ 等使用语音通信的应用程序均使用了 UDP。

UDP 定义于 RFC 768，提供应用程序能够在最低的协议机制下发送消息给其他应用程序，但不保证是否可靠或有没有依照传送顺序到达。

UDP 数据格式如图 4-8 所示。

图 4-8　UDP 数据格式

4.8　TCP

IP 网络偶尔存在丢包的情况，为了在这种情况下还能有效保障传输的质量，需使用 TCP，如文件传输服务。例如，从 FTP 服务器下载电影需要将一个 1GB 大的电影文件分成很多网包在网络上传输，如果其中一份数据丢失将导致整个文件无法打开（因为校验不通过）。

为了解决可靠传输的问题，TCP 的基本思想是为每次通信会话保留状态，为传输的每段数据编号，通过发送和接收端的互相通报以及保存的状态信息来确定可能的网包丢失情况，通过重新传输丢弃的网包来保证可靠性。

TCP 提供以下三个基本功能。

(1) 可靠性(Reliability)：克服包丢失现象，传送的数据依照顺序交给程序。

(2) 复用性(Multiplexing)：通过不同的端口可使同一个 IP 地址(一台计算机)可以同时提供上层不同的多种服务，如 FTP、Telnet、HTTP Web 服务等。

(3) 流控(Flow Control 或 Congestion Avoidance and Control)：避免网络或接收端拥塞，以便提供高效率的传输。

TCP 的会话叫做连接(Connection)，TCP 的会话建立需要三次握手，如图 4-9 所示，会话拆除也需要握手。这种握手协议也是 TCP SYN 攻击的原因，一种异常的 TCP SYN 泛洪攻击如图 4-10 所示。

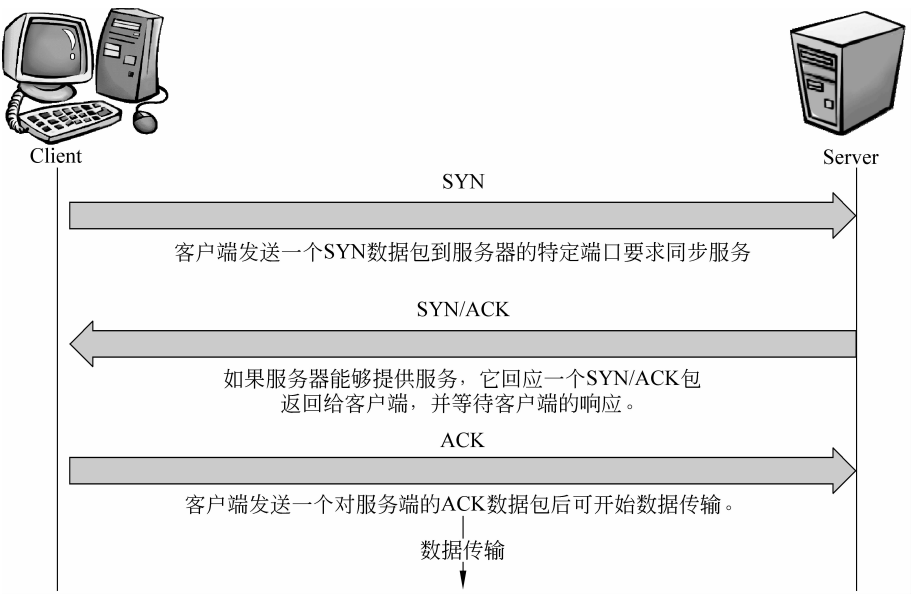

图 4-9　TCP 三次握手

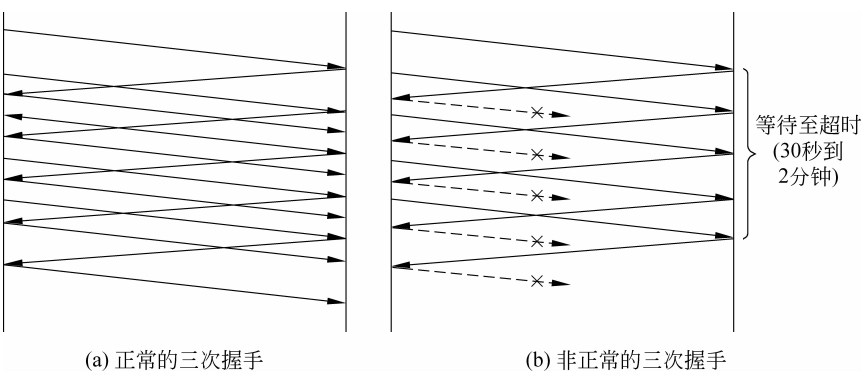

图 4-10　一种异常的 TCP SYN 泛洪攻击

TCP 的格式如图 4-11 所示。

图 4-11　TCP 的格式

4.9　以　太　网

1. 以太网(Ethernet)简介

以太网是使用最广泛的局域网类型。10M/100M 以太网采用共享网络介质,属于共享带宽;吉比特以太网 (Gigabit Ethernet) 采用交换网络方式,属于独享型带宽。

2. 以太网协议

以太网协议见 RFC 894,以太网协议帧格式如图 4-12 所示。

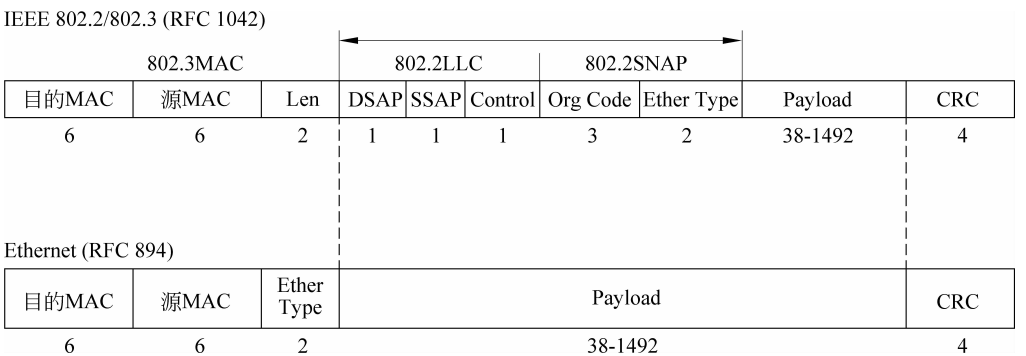

图 4-12　以太网协议帧格式

3. 5 类以太网络线

以太网网线为 10Base-T 及 100Base-T,使用 8 芯(4 对线)的无遮蔽双绞线(Unshielded Twisted Pair,UTP)网络线,接头为 RJ-45,接法有平行接法和串接接法,具体连接关系如图 4-13 所示。

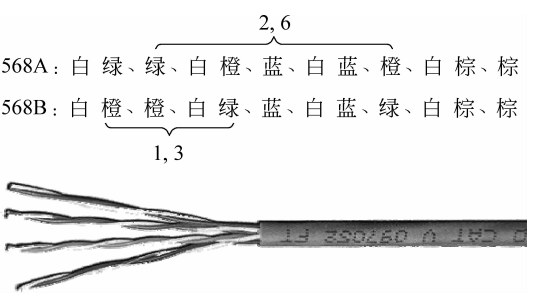

图 4-13　以太网 RJ-45 接口走线