

Internet 顾问(连载三)

第三章 Internet 如何运作?

本章论述关于 Internet 是如何工作的问题。Internet 专家们喜欢制造许多各式各样的名词,比如:域名、系统、主机、TCP/IP、URL、WWW 等等。本章我们就来谈谈这些专用词语。同时也将涉及使用 Internet 的一些基本知识。

1. 常听人谈起 Internet 主机(host),什么是主机?

在 Internet 上,主机是指任何一个物理地连接到网络上的计算机系统。更具体地讲,它是指任何一个具有唯一标识——一个名字和一个网络地址的计算机。

注意,在以后的论述中,我们经常对主机(host)、节点(site)、计算机(Computer)不加区别,互换使用。

每一个 Internet 主机都拥有一个“系统·域”(System. domain)形成的名字。系统(system)是这台计算机自己的标识符,域(domain)则包含了这台计算机所归属的组织的信息。

2. Internet 上的计算机如何相互通信?

什么是 TCP/IP?

TCP/IP 是一百多个数据通讯协议所组成的集合的统称,这些数据通讯协议将各式各样的计算机有机地结合起来,构成网络。组成 Internet 的计算机通过 TCP/IP 协议语言来相互通信。

TCP/IP 为 Internet 上的计算机制定了一套寻址方案。TCP/IP 规定了网络上的计算机和程序之间如何移动数据。这些协议是计算机为了将不同类型的数据从一个地方转移到另一个地方所必须遵循的规定。

通过 TCP/IP 传送的信息称为组(packet)。我们可以把在 Internet 上发送的每一组信息看作是一封信件。TCP/IP 将每一封信放在一个信封中,然后在信封上写明发信人和收信人的信息,最后发出这封信。这些组被设计得很小——通常只有 1500 B 左右。我们在 Internet 上发送和接收的东西(e-mail 消息、Usenet 邮件、文件等等)通常都大于一组的容量,因此,TCP/IP 将这些信息拆解开来,分成组发送。一旦这些组抵达目的地,TCP/IP 又将它们重新组装成一篇连贯的信息。

事实上,TCP 和 IP 是两个能够协调运作但又相互独立的协议。IP 负责将各组信息送往目的地,而 TCP 负责检查各个分组的完整性并将它们按原先的顺序排列。

事实上,将你的信息从源地址发送到目的地址也很好理解。Internet 是一个存储转发网络(store and forward network),这意味着这些分组信息在通往其目的地的道路上可以经过任何一台计算机。如果在两个节点之间有一个直接的网络连接——也就是说在两台计算机之间有一条物理电缆相连通——那么那些分组信息可以一步到位。但大多数情况下并没有直接连接的通路。因此,发送方的计算机将分组信息,发送到离接收方计算机较近的一台计算机上,这台计算机再传送下去,以此类推,直到分组信息抵达终点。有时一组横跨全美的分组信息从发送端到接收端要经过 20 至 30 多次中转接力。但大多数情况下,这一切发生的非常快。从加利福尼亚使用 Telnet 连接到纽约或者芬兰(如果顺利的话),你几乎觉察不到时间上的延迟。

3. 什么是域名(domain name)?

计算机和人有着不同的工作方式和习惯。计算机喜欢同数字打交道,而人类更倾向于使用单词和名字。

Internet 的每一台主机都拥有一个地址,它由四个数字组成,每个数字都小于 256,由句号分隔开。尽管计算机很喜欢这种编址方式,但人类并不太满意。因此,为了人们使用方便,Internet 的计算机也纷纷起了名字。

每一台计算机的地址——正式称为 IP 地址(Internet Protocol address)——是由用句号分隔开的四个数字组成,例如:

137.150.188.22

192.160.13.1

139.130.4.6

140.174.1.1

一般地,既可以使用它的名字也可以使用它的地址。比如,你可以键入:ftp archie.all 或键入:教室 139.130.4.6。用两种方法都可以使你连接到同一台机器上。但要注意,e-mail 是一个例外,你在使用 e-mail 时只能使用系统的名字,而不可使用其地址。E-mail 地址例如:savetz@rahul.net 千万不可写成 savetz@192.160.13.1。

举例说明之:一台汉勃特州立大学(Humboldt State University)的计算机名为

turing.cnrs.humboldt.edu

在这个例子中,有四个被句号分隔开的单词。计算机的名字或称宿主主机名字(hostname)是 turing。而 cnrs.humboldt.edu 是这台计算机的域名。(域中的每一个单词称为一个子域:subdomain。)

域提供了该计算机的相关信息,最左边的是最具体的信息,最右边的是最笼统的信息。turing.cnrs.humboldt.edu 是这台宿主机的完整的域名,由计算机名加它的 IP 地址组成。这台计算机——连同其名字——由汉勃特州立大学(Humboldt State University)的自然资源科学学院(College of Natural Resources and Sciences, 缩写为 cnrs)所拥有。Humboldt 是一个包括所有教育机构的全国性组织 edu 的一部分。因此,仔细研究一台计算机的名字(破译一些缩略语)往往可以增加我们对一个 Internet 节点的了解。

4. 什么是域名系统(domain name system)?

Internet 上的计算机需要一种将节点名翻译成相应的数字地址的方法。Internet 有一个类似电话号码簿的东西:一台计算机可以用名字在其上进行查寻,得到另一系统的地址。但这并不像听起来那么简单。数以百万计的宿主主机使得 Internet 上的宿主主机号码本非常非常庞大,即使是电子化的也数目惊人。并且当 Internet 上两台不同的计算机使用了同一个名字时,将会发生什么情形呢?哪一个地址是正确的呢?计算机可不喜欢模棱两可的回答。

当 Internet 的规模远远比今天小的时候,管理 Internet 地址查寻号码本的工作还比较容易。网络信息中心(NIC: Network Information Center)保留了 Internet 节点的注册登记记录。这一文档称为宿主主机文件(hosts file),被定期分发给 Internet 上的各个节点。但这些日子已经一去不复返了。随着 Internet 的不断成长壮大,管理和发送这样一个庞大的宿主主机文件渐渐变得非常棘手。

域名服务机构(the Domain Name Service, 或称为域名系统:the Domain Name System, 简记作 DNS)取代了过时的宿主主机文件。这一方法通过授权给每一节点组织,让该节点管理本地名字的方法来管理 Internet 系统名字。这一方法使得由宿主主机名字到其地址的转换不再依赖于一个集中统一管理的文件。

从此再没有一个统一的节点表了。取而代之的是每一个组织分别管理本组织内部的计算机。各家自扫门前雪。如果一个汉勃特大学的用户想要了解弗雷德的计算机发出查询信息,然后等待弗雷德的计算回答。简括地

说,这就是域名系统。

如果汉勃特州立大学自然资源科学学院的计算机实验室要想新接一台计算机上 Internet 网的话,他并不需要经过网络信息中心(NIC)同意,也不必让人把新机器加入宿主主机文件中。因为有了域名系统的支持,这一切都可以由用户自己完成。

尽管 NIC 至今仍在少量分发一些 IP 地址块,但 IP 地址的分配不能乱来。在将一台计算机连接到 Internet 上之前,必须从 NIC 获得地址块。你所得到的地址块的大小取决于你的需求量的大小。最小的一种地址块称为“C类”地址(例如 137.150.188.*),可供 254 台计算机上网。“B类”地址(如 137.150.*.*)适合于更大的组织使用,它可容纳 64,516 台宿主主机。最后,“A类”地址(如 137.*.*.*)拥有 16M 的地址容量。

相似地,如果有人想在汉勃特州立大学(HSU)创立一个新的小组(比如 journalism),并且将有三台计算机(分别为:murrow, rather. 和 hearst)加入这个小组的话,他可以不经任何人的批准。这样,这些机器的全名为:

```
turing.chrs.humboldt.edu
murrow.fournalism.humboldt.edu
rather.journalism.humboldt.edu
hearst.journalism.humboldt.edu
```

只要两台机器的域名不重复就没关系。如果每个节点的系统管理员肯定他所分配的名字都是互不相同的,那么就一定不会发生用名冲突。例如以下所给出的宿主主机名字都是对上述例子的有效的补充,它们在 Internet 上互不冲突:

```
murrow.chrs.humboldt.edu
turing.journalism.humboldt.edu
rather.sonoma.edu
```

5. 域名当中的.com、.net 或 .edu 是什么意思?

你经常在 Internet 的域名中看到象.com、.net、.deu 或 .mil 之类的后缀。这些“顶级”(top-level)域是在域系统建立时所建立的。下面是传统的域名后缀表:

```
.arpa 老式的 ARPAnet 地址(现已不再使用)
.com 商业节点
.edu 教育机构
.gov 政府节点
.mil 美国军方
.nato 北大西洋公约组织 NATO (现已不再使用)
.net 网络
```

.org 其他组织(通常为非赢利性组织)

这一命名方案试图将网络地址分门别类,以便用户对所连接的机构或所要发送邮件的接收方有些了解,但这一方案并不十分理想。当 Internet 还仅仅限于美国国内使用的时候,这一方案相当有效,但是随着数以百计的国家纷纷加入 Internet 以及各式各样成千上万的组织的涌现,这一方案的缺点逐渐暴露出来。比方说 .gov 这个扩展名代表政府节点的意思,但如果你不知道究竟哪一个国家拥有这台计算机的话,这一扩展名也说明不了多大问题。

现在有一种新的域名寻址方式,在这种方式中最后的字母指示计算机的地理位置,而不是它的组织所在的领域。比如节点 Well.sf.ca.us 在圣弗朗西斯科市(San Francisco),加利福尼亚州(California)、美国(United States)。但是令人遗憾的是这一方式在美国大受冷落。(这也许是因为美国人生性不愿意被一些死板的东西所束缚的原因,比如说:国家代码、公制系统等等。)总之,到目前为止还没有一个最后正式定形的读解域名的方案。

6. 我既有主机名又有 IP 地址,该用哪个呢?

两个都可以使用,但你应该养成使用主机名而不是其 IP 地址的习惯。如果一台主机物理上发生了位移,那么它的 IP 地址往往也要更动,但主机名却总是不变的。使用这个不变的 hostname 可以使你不受一定地域的限制。事情是千变万化的,但是使用 hostname 而非 IP 地址可以使这些变化对你产生更小的影响。

7. 什么是 URL?

统一资源指示器(URL:Uniform Resource Locator)是用来给出 Internet 上的对象(object)——文件、Usenet 新闻组、Telnet 节点以及其他的工具和资源——的地址的标记。URL 提供简单易读的联机提示,帮助你使用 Internet 上的各种服务功能。不仅人类读 URL 很容易,它也很适合于计算机理解。如果你使用过 World Wide Web(WWW)或其他联机超文本(hypertext)工具的话,你会经常发现当按钮被按下去时往往伴以一些事件的发生——比如运行一个程序、下载一个文件、Telnet 到某一节点等。这些按钮是与 URL 相关连的,按下该按钮,与之相关连的 URL 便被激活了。URL 的使用主要是随着 WWW 的发展而普及起来的。

8. 什么是 Telnet?

Telnet 是一个程序,它可以使你与其他的计算机连接上,并且运行那台计算机上的软件。通常有两种登录

方式,或者进入某一外壳(shell)环境(例如远程计算机上的 UNIX 操作系统),或者进入某一实用程序,例如天气预报服务或游戏。

要想通过 Telnet 连接到另一台计算机上,你必须知道该机的名字或者是它的 IP 地址:

```
telnet bolero.rahul.net
```

```
telnet 139.130.4.6
```

有些服务功能要求你在使用 Telnet 时明确指出所要连接到的远程计算机上的具体端口号。此时,这一端口往往是某一服务专用的,所以你一旦连通便可立即使用该程序或工具。在 Internet 地址后面,如果有端口号的话,一定要键入该端口号,例如:

```
telnet nri.reston.va.us 185
```

```
telnet lambda.parc.xerox.com 8888
```

9. 什么是 Gopher?

Gopher 是一个面向菜单的巡游和浏览 Internet 的工具。Gopher 以易于认读的指点式菜单的形式向你提供了 Internet 上的文件、程序、资源、服务机构等等菜单。你只需移动光标或输入适当的项目序号便可作出选择;从而只靠这些简便的键盘操作就巡游了大部分 Internet。

全世界有 4,500 多家企业、政府机构、个人通过 Gopher 向 Internet 上的用户提供其信息服务,这使得 Gopher 成为当今 Internet 上最受欢迎的工具之一。

Gopher 使那些管理人员将这几千个联机文件划分成分级的菜单结构,用户通过使用一些箭头键、回车键、鼠标键的组合动作或者输入数字来巡游这些文档。

每一个菜单都由若干选项组成,每一选项通常只包含一行文本描述信息。一个 Gopher 菜单选项可能是一个文本文件;也可能是其他类型的文件(比如一张天气预报图、一封邮件、或一帧图象);也可能是其他服务权限,比如 Telnet、FTP、WAIS、finger、一个查寻工具等等。当一个文件被选中后,它将被检索到并显示在用户的显示屏上。Gopher 也允许你将所选中的文件存在你的本地帐户上,或者把它作为 e-mail 发送给别人,或者将其打印出来。

一个 Gopher 菜单可以包括许多种上述项目,最多可达数百种。使用 Gopher,还可以定义书签(bookmark),用它可以快速存储和重新定位某一信息,而不用一步一步去查找。

Gopher 中的 server 程序负责管理文件服务功能。每一个用户都运行一个 Gopher client 程序,它用来处理诸如显示接收到的菜单和文件的事务。

有一些 Internet 用户把 Gopher 当作与其他用户分享信息的便捷手段。Gopherspace——Internet 上用户可以使用的 Gopher 服务器——很快地从 1,000 升到 2,000, 最后升到 3,000 个, 总共包含了 2,000,000 个菜单项!

Gopher client 适用于各类流行的计算机环境——DOS、Windows、Mac、Amiga 和 UNIX(使用 ASCII 和 X Windows)。Gopher 几乎适用于任何类型的计算机。

10. 什么是 Veronica?

Veronica(Very Easy Rodent - OrientedNet - wide Index to Computerized Archives; 简明全网计算机文档索引)是 Gopher 菜单项的一套索引系统, 用户使用它可以快速地查找到所需的信息。

不幸的是 Veronica 被证明有许多问题, 有些问题是 Gopherspace 所固有的。Gopherspace 的本质特点——入口可以轻易地指向远程服务器以及某一给定的服务器, 本身具有复杂的甚至循环嵌套的结构——使得正常的检索变得相当困难。

另外, Gopher 的菜单项也过于精简, 在脱离了其菜单上下文之后显得含意模糊、晦涩难解。因此, 你很难断定一次 Veronica 查寻的结果到底是一些不同的选项呢? 还是一些多级列表。而且, 目前总共只有三四家分开的服务机构提供 Veronica 服务。

11. 什么是 WWW(World Wide Web)?

全球广域网(WWW: World Wide Web)有点儿像 Gopher, 它是一个在相关连的 Internet 文件、资源以及服务之间进行组织、连接并提供触点式进入权限的系统。

WWW 使用超文本(hypertext), 或称“超媒体”(hypermedia)的手段。在这种方式中, 交叉引用(cross-reference)嵌在文档或其他入口当中。每一个交叉引用都是一个指向另一文件、动作、表格或菜单的指针。你可以这样想象它的工作方式: 当你点触一个注释之后, 你被立即带到另一本相关的书本中并且听到语音的解释, 或者你被自动连接到其他的服务项去, 并且登录到该服务项, 比方说登录入国会图书馆(Library of Congress)。

在 1993 到 1994 年期间, 作为一种享用 Internet 网络信息的快捷手段, 并且加上 Mosaic 和 Lynx 这两个 WWW“浏览器”的推动与促进, WWW 的使用迅速普及并且逐渐赶上了 Gopher。

WWW 浏览器或称为前端(front-end), 是你为了获取网上存储的信息所运行的一个程序。现在有两个流行的 WWW 浏览器: 一个是 Lynx, 它是基于 ASCII 码的; 另一个是 Mosaic, 它是一个多媒体浏览器, 可以在 Macin-

tosh 和 Microsoft Windows 上运行。你们在 Internet 上及本书中都会多次遇到 Mosaic 这个词, 它是一个改变了人们使用 Internet 的方式的软件, 它使用户可以轻而易举地获得文本、声音以及图像信息。

12. 为什么有如此多的不同的 Internet 工具?

由于种种原因 Internet 上有许许多多不同的工具。第一, 由于 Internet 的用户有许多的事情要处理, 因此他们需要大量不同的服务工具。最初, 每一种服务都针对某一特定的功能: FTP 用来传输文件; Archie 是 FTP 文件的数据库; Gopher 是一个分布式的基于菜单的文档服务器。比如, 用户需要 Telnet 来进行远程登录, 而传输文件时则需要 FTP 等。每一种工具有它自己的用途。由于这一原因, 产生了大量的工具。第二, 由于 Internet 的用户数量多, 范围广, 分布较分散, 很有可能产生这样的现象: 许多不同的个人和组织(有时对于对方互不了解)为了解决某个相近的问题而分头去研究, 最终产生了对该问题的各种不同的解决方法。比如, Gopher、Hytelnet、TechInfo 和 WWW 在某种意义上讲代表了对若干种服务功能的不同实现途径。同样在“怎样发现某个的 e-mail 地址”这一领域内分别有 Michael Schwartz 的 NetFind, Daniel Kegel 的 uwho 和 Whois++ 等等。许多问题可以有不止一种解决方法。

每一种工具都是为了满足该组织的需要而创造出来的, 因此它们工作方式各异, 往往带有各自的特点。由于各自的工作方法不相同, 你可能偏爱某一种工具而不喜欢其他的几种; 也许你要根据其价格、性能以及所需的计算机环境来选择某一合适的工具。

13. 什么是索引器(indexer)和导航器(havigator)?

索引器包含几方面的内容: 收集数据、建立索引、处理用户查询等等。例如 Archie 便是这样一种索引器, 它为成千上万的匿名 FTP 节点上的信息编目分类。索引是用来对 Internet 上数以千计的服务器作虚拟查找用的, 例如对匿名 FTP 节点或对 Gopher 服务器的菜单项的查找就需要用到索引。它的原理是: 从每一个参与该次查询的服务器上收集一张列表, 然后将这些列表放入一个可以查找的数据库内, 从而一次查询便看上去象是对整个 Internet 的一次搜索。

导航器给用户提供了这些信息的视图, 并且用户可以通过导航器进行查找、浏览以及选择。它提供给用户的是一个全 Internet 范围的视图而非只限于某一节点。比如, 一屏 Gopher 菜单可能会包括了全球数十个不同地点的信息!

(吴小钧 编译)