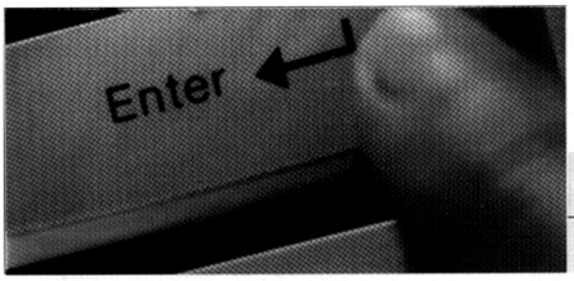
校园网反向代理服务器的应用与开发



曲 波 吴兆芝 (大连海事大学计算机学院 116026)

摘 要	本文阐述了在校园网中反向代理服务器的应用及其开发方法。
关键词	校园网 反向代理服务器 应用与开发

1 反向代理服务器在校园网中的应用

随着Internet的不断发展，越来越多的机构把自己的内部网络连接到Internet上。如何保护内部网络免受外部的恶意攻击，是网络系统管理人员面临的一个重要任务和严峻挑战。防火墙技术是实现这一目标的一个重要手段，绝大多数内部网络都通过各种形式的防火墙与Internet互连。防火墙技术的实现目前主要有两类：IP包过滤防火墙与代理型防火墙。

包过滤防火墙基于网络结构的底层，截取流经防火墙的所有IP包，识别其是否符合管理员预先设定的过滤规则。只有符合一定要求的IP包才被正确转发，其余一律废弃。通常包过滤的防火墙根据IP的类型屏蔽所有由外部发起的连接请求，从而保护内部网络。如果将内部网络的对外服务器(如Web服务器、FTP服务器、BBS服务器等)放在防火墙之内，就必须设置相应的过滤规则，允许外部网络对这些服务器以及它们使用的端口的访问。包过滤防火墙对保护内部网站有一定的局限性，一旦外部网络的主机可以向内部网络服务器发起连接请求，攻击者就可以在网络外部尝试进行连接；一旦攻击者进入内部网络服务器，整个内部网络就暴露在攻击者面前。因此包过滤防火墙只是一种简单的保护内部服务器的方法，安全程度不高。

代理型防火墙又称代理服务器，有基于应用型和基于Socks两种。基于应用型代理服务器工作在网络结构的应用层，针对不同的应用协议提供不同的代理服务；基于Socks的代理服务器工作在网络结构的传输层，针对TCP或UDP协议提供代理服务，与具体的应用协议无关。代理型防火墙采用网桥方式，跨接在外网与内网之间，将外网与内网有效隔离。内部用户对外的访问通过代理服务器提供代理服务，其内部IP地址及端口等对外是不可见的，外部网络所见到的只是代理服务器的IP地址和端口号。所以代理型防火墙可更有效地保护内部用户。

普通的代理服务器对用户提供多对多的代理访问服务，即对多个内部用户提供对多个外部网站的访问。为确保内部用户能够通过代理服务器实现对外访问，内部用户必须显式地配置代理服务器的地址及端口号。显然，对访问内部网络的外部用户来说，无法限制其必须配置相应的代理服务器，因此普通的代理服务器只适于内部用户对外访问，而不适于外部用户对内部网络的访问。

反向代理型防火墙可有效解决这个问题。反向代理服务器对外部网络表现为一个应用协议的服务器，但其上并无提供相应服务的任何内容。外部网络把反向代理服务器当作标准的应用协议服务器对其访问，反向代理服务器接收到外部网络的访问请求之后，代替外部网络对内部网络相应的服务器提出访问请求，然后将内部服务器所返回的服务结果返回给外部网络。

由此可见，反向代理服务器作为内部服务器的防火墙，使内部服务器对外部网络来说完全不可见，从而可靠地保证了内部网络服务器的安全。而反向代理服务器本身又没有任何相关的信息资源，因此具有较高的抗攻击能力。

©《计算机系统应用》编辑部 http://www.c-s-a.org.cn

目前，许多校园网采用基于用户的流量计费系统，用户通过代理服务方式访问外网，系统通过代理服务器截取用户流量信息。校园网内部服务器上同样可安装用户的客户机软件，从而使其一身二任：对外提供某种服务时它是一个服务器，使用客户机程序时它又是一个客户机。对这样的服务器如果简单地配置网络防火墙使其能够穿透防火墙，则同时也使其能够作为客户机不需代理而直接访问外部网络，从而逃避系统对其流量计费。而反向代理服务器则可有效解决这个问题，外部网络可通过反向代理服务器访问内部服务器，而内部服务器上的客户程序却不能通过反向代理服务器访问外部网络。

不难看出，反向代理服务器在校园网中可起到其他软件不可替代的作用：作为虚拟服务器对外部网络用户提供透明的代理服务，同时又作为内部服务器的对外防火墙。反向代理防火墙具有单向性(外部网络可通过反向代理服务器访问内部网络服务器，内部网络无法通过反向代理服务器访问外部网络)、透明性(反向代理服务器对外部网络表现为对某种应用协议提供服务的应用协议服务器，外部网络不需作任何额外配置并把它当作校园网的内部服务器来访问)和屏蔽性(外部网络看到的是反向代理服务器的IP地址和端口号，内部服务器的地址和端口号对外部网络是完全不可见的)。

2 反向代理服务器的开发

反向代理服务器对外部用户表现为对某种应用协议提供服务的应用协议服务器，工作在网络结构的应用层，根据所服务的应用协议不同，其程序结构也各有不同。就一般校园网而言，可提供的反向代理有Http(Web访问)、FTP、BBS等。

由于Linux操作系统允许一台主机具有多个IP地址，所以可在一台主机上用多个IP地址实现多个反向代理服务器，也可以用一个IP地址、多个端口号实现多个反向代理服务器。

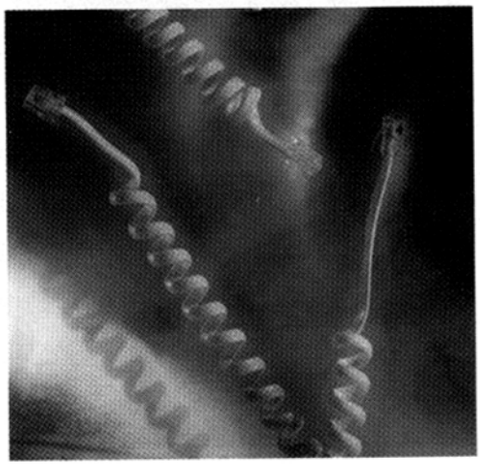
BBS反向代理服务器的原理并不复杂，本质上是一个接收转发服务器。FTP反向代理服务器由于要支持建立第二信息通道传送数据信息，因此不能简单地接收转发，但其程序结构与普通应用型FTP代理服务器基本相同，程序稍加修改即可实现。Http反向代理服务器本质上也是一个接收转发服务器，但由于涉及到主页信息，所以要略复杂一些。下面以笔者开发的Http反向代理服务器为例，说明开发反向代理服务器需要解决的几个问题。

2.1 反向代理服务器的程序结构

反向代理服务器既要面向内部用户，又要面向外部用户，对其响应速度要求较高，因此笔者采用并发多进程方式设计反向代理服务器。对于Http反向代理服务器而言，应能识别用户的GET请求，以便必要时返回镜像主页。

GET请求页面由Http请求页面的首行确定，GET请求页页的首行为：

GET请求首行格式为：GET URL HTTP协议的版本号



其中URL表示请求访问的主页地址。若URL的内容为“/”，则表示请求访问主页的默认首页。此时，可以把此请求页面直接转发给目标服务器，然后将目标服务器返回的主页页面返回给客户端；也可以从反向代理服务器上读取经过修改的镜像页面，返回给客户端。

2.2 从CERNET访问

CERNET是大连海事大学校园网的上级主干网，校园网的www服务器以通过校园网防火墙对外开放，无需反向代理。需要反向代理的是学校各部门的Web服务器，如图书馆、出版社、质管办、教务处等。

第一步，将各部门Web服务器分成两类：对外服务器和对内服务器。

第二步，在反向代理服务器主机上为每一个对外服务器设置一个对外IP地址。

第三步，在校园网DNS上为每一个对外主机建立一个域名，其地址为反向代理服务器主机上为其设置的IP地址。

第四步，在反向代理服务器主机上为每一个对象上主机域名(对应不同的IP地址)启动一个反向代理服务器进程，将其目标服务器设置为所对应的内部对外服务器的内部实际IP地址，例如：

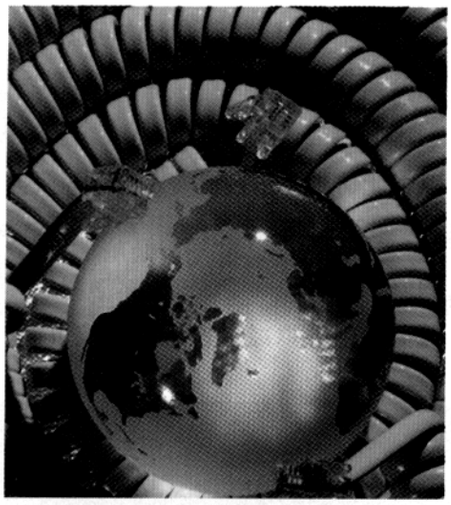
lib.dlmu.edu.cn反向代理服务器，目标服务器为202.118.84.1(图书馆Web服务器)；

press.dlmu.edu.cn反向代理服务器，目标地址为202.118.88.4(出版社Web服务器)。

第五步，将校园网www服务器主页上各对外服务器的链接地址修改为相应的域名地址。

当外部网络通过CERNET访问大连海事大学校园网主页时，得到对

©《计算机系统应用》编辑部 http://www.c-s-a.org.cn



外服务器的域名，如lib.dlmu.edu.cn。接下来访问该域名时，实际访问的是反向代理服务器主机上对应该域名所启动的与之对应的反向代理服务器进程，由该反向代理服务器进程完成代理服务工作。

2.3 从CHINANET访问

大连海事大学校园的CHINANET入口IP地址只有一个(202.107.54.198)，不能用不同的IP地址对应不同的内部对外服务器，因此使用同一个IP地址，用不同端口对应不同对外服务器，例如：

202.107.54.198.80 对应www.dlmu.edu.cn(校园网www服务器)；

202.107.54.198.8001 对应202.118.84.1(图书馆Web服务器)；

202.107.54.198.8002 对应202.118.88.4(出版社Web服务器)。

从CHINANET入口访问大连海事大学校园网的目地是加快对其访问的速度，但由于大连海事大学校园网主页上内部对外服务器的链接地址仍然是CERNET的域名和地址，因此当外部网络通过主页链接点访问这些内部对外服务器时，仍然需经由CERNET进入大连海事大学校园网，所以并未从根本上解决问题。

笔者采用镜像主页的方式解决了这个问题。在反向代理服务主机上存放一份校园网Web服务器的主页首页，将其上的链接点地址修改为反向代理服务器主机的IP地址及其相应的端口号，例如：

将lib.dlmu.edu.cn改为202.107.54.198:8001 (对应图书馆Web服务器)；

将press.dlmu.edu.cn改为202.107.54.198.8002 (对应出版社Web服务器)。

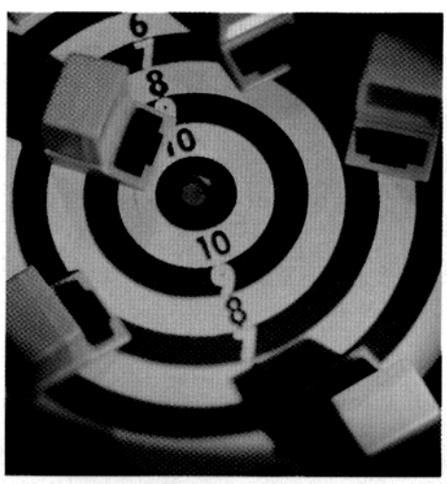
在反向代理服务器主机上对上述端口启动相应的反向代理服务器进程，其目标地址设置为相应的内部对外服务器IP地址。

当外部网络用户用浏览器通过CHINANET访问202.107.54.198时，所得到大连海事大学校园主页是放在202.107.54.198主机上的镜像主页，该镜像主页的链接点已被修改，当外部用户通过链接点访问大连海事大学图书馆时，其访问地址为202.107.54.198:8001，该地址所对应的反向代理服务器进程为外部用户实现对图书馆Web服务器的访问代理服务。

这种方法要求将对外服务器的链接点都体现在首页上。若内部对外服务器链接了本部门另外的对外服务器，也可以采用这种镜像主页的办法。

2.4 访问控制

反向代理服务器应对用户的源地址具有访问控制功能，以拒绝非法或恶意用户对其访问。用户的源地址可通过accept函数获得，也可通过getpeername函数获得。



在配置文件中将拒绝的源IP地址全部列出，在反向代理服务器启动时一次性读入，然后在反向代理之前将用户源IP与列表中的每一个IP地址比较，若相同则拒绝服务。为节省内存及提高比较速度，对IP地址列表采用子网/子网掩码结构，连续的源IP地址通过子网/子网掩码来描述。

2.5 流量统计

反向代理服务器应具有流量统计功能。由于内部用户应免费使用反向代理服务器，所以反向代理服务器应能区分是否内部用户，若是则不统计流量。若不是内部用户，则读取客户端请求时，累加其流量作为流入量；向客户端返回结果时，累加其流量作为流出量。当一次TCP连接结束后，将累加的流入量及流出量存盘。

由于CERNET对国内流入量和国际流入量使用不同的计费价格，因此反向代理服务器应能区别国内流量和国际流量。CERNET网络中心不定期公布CERNET收费政策定义的国内IP地址表，可将作为判断国内、国际流量的依据。

流量统计数据以二进制文件方式存盘，通过专门程序显示其数据。

3 结束语

实践证明，在校园网中使用反向代理服务器可较好地解决外部网络对内部网络的安全、快速访问，同时也较好地解决了内部网络的泄露问题，具有很高的实用价值。开展对反向代理服务器的研究与开发，很有实际意义。■

©《计算机系统应用》编辑部 http://www.c-s-a.org.cn