Talk:Group8

取自 Wiki for the Forum of Campus Networks

第八组第三次作业--穿越GFW技术及其控制方法

凹建勋 李理 贾斌 戴魏巍

（清华大学计算机系 网络所 北京 100084）

摘 要：本文分析了GFW所采用的主要技术，介绍了几种突破GFW封锁的软件及其技术原理，并针对它们使用的技术介绍了相应的控制方法，并进一步分析了可能的突破封锁的技术。

关键词：GFW；加密代理；穿越；破网控制；透明Web Cache

目录

一、引言

WWW空前广泛的应用，正在影响和改变人们的生活方式。但在WWW庞大的网络信息空间中，夹杂着大量的有害信息，主要包括：垃圾信息、虚假信息、政治渗透信息、种族歧视信息和恶意代码等，这些信息的泛滥对Internet造成了严重的信息污染。

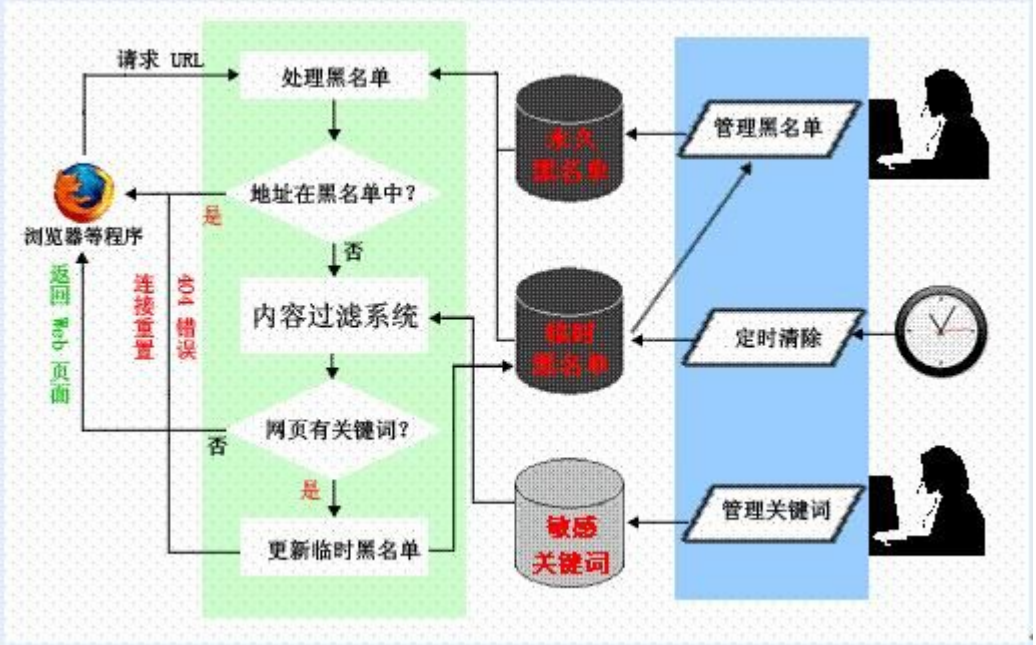
对网络空间的监控能有效地阻止有害信息的传播，控制计算机犯罪。放置在可信任网络和不可信任网络之间的防火墙，是运用非常广泛和效果最好的选择[1]。 Internet可以分为国内网络与国外网络两部分。由于各国的安全策略各不相同，因此不同的国家对有害信息的认定有不同的标准。对于我国来说，不良信息主要集中在国外网络。防止信息污染不仅要保证国内网络空同的洁净，同时要防止国外网络不良信息的侵蚀。目前我国的国际互联网出口的核心节点设在北京、上海和广州，国内的计算机信息网络进行国际联网，必须使用国家公用电信网提供的国际出入口信道。为了有效的控制信息流动，在出口处安装了防火墙[2]。

防火长城，也称中国防火墙或中国国家防火墙，这是对“国家公共网络监控系统”的俗称，是指中华人民共和国政府在其管辖互联网内部建立的多套网络审查系统的总称，包括相关行政审查系统。其英文名称Great Firewall of China，缩写为GFW[3]，国内简称“防火长城”，国外也叫“功夫网”。

二、GFW及其主要技术

1、概述

GFW主要指公共网络监控系统，尤其是指对境外涉及敏感内容的网站、IP地址、关键词、网址等的过滤。GFW的效果通常为，国内网络用户无法访问某些国外网站或者网页；或者国外网络用户无法访问国内的某些网站或者网页。这里的无法访问，有永久性的无法访问（比如某些色情网站），也有因为URL中含有敏感关键词或者网页上有敏感内容而暂时性的无法访问。国家防火墙并非中国的专利。其他国家也有类似的防火墙，对危害其国家安全的信息进行侦听，而中国的国家防火墙会直接切断敏感连接。伊朗、巴基斯坦、乌兹别克斯坦、北非共和国、叙利亚、缅甸、马尔代夫、古巴、北韩、南韩、沙特阿拉伯、阿拉伯联合酋长国、也门使用与GFW类似的国家防火墙。以下是猜想的GFW工作原理图[5]。



2、GFW所采用的关键技术

（1）、国家入口网关的IP封锁

从90年代初期开始，中国大陆只有教育网、高能所和公用数据网3个国家级网关出口，我国政府对认为违反国家法律法规的站点进行IP封锁，这是有效的封锁技术。对于IP封锁，用普通Proxy技术就可以绕过。只要找到一个普通的海外Proxy，然后通过Proxy就可以浏览自己平时看不到的信息了。所以，网络安全部门现在通常会将特别反动的网站的网址加入关键字过滤系统，以防止网民透过普通海外HTTP代理服务器访问。

一般情况下，GFW对于海外非法网站会采取独立IP封锁技术。然而，部分非法网站使用的是由虚拟主机服务提供商提供的多域名、单（同）IP的主机托管服务，这就会造成了封禁某个IP，就会造成所有使用该服务提供商服务的其他使用相同IP的网站用户一同遭殃，就算是内容健康、正当的网站，也不能幸免。例如如森美的个人网站，内容并无不当之处，但网站使用的是虚拟主机托管服务，而因为有一个香港BBS亦使用该托管服务，这就造成了GFW为了封锁该BBS，直接把这个固定IP：203.80.210.5封禁了。随之，有82个香港网站由于GFW封锁了这个IP地址，不论合法与否，都不能在中国大陆访问。

（2）、主干路由器关键词过滤拦截

主干路由器关键字过滤拦截在2002年左右开始，中国公安部门研发了一套系统，并规定各个因特网服务提供商必须使用。思科等公司的高级路由设备帮助中国大陆实现了关键字过滤，最主要的就是IDS（Intrusion Detection System）--- 入侵检测系统。它能够从计算机网络系统中的关键节点（如国家级网关）收集分析信息，过滤、嗅探出指定的关键字，并进行智能识别，检查网络中是否有违反安全策略的行为。

IDS主要进行IP数据包内容的过滤，如果符合既定的规则，则向该连接两端的计算机发送IP RST包，这可以从前后IP报头TTL值相差较大的特点可推测出来，用这种方法干扰两个通信终端间的正常TCP边接，使数据流中断，而在终端主机上会显示连接失败。这种关键字过滤-重置技术只对TCP连接有效。而广泛应用的HTTP协议正是使用TCP作为传输层协议，从目前来看，GFW对HTTP报文的过滤仅限于HTTP头，通常URL请求就位于HTTP的头部分，而GFW对HTTP数据部分很可能不作过滤，这正是某些用PHP编写的HTTP在线代理能避开关键词过滤的原因，例如PHProxy，它将明文的URL请求放

在HTTP数据部分，而不是放在HTTP的头部。对UDP（DNS通常使用UDP，GFW对捕获的DNS查询报文也进行关键词过滤并返回伪DNS响应，但因UDP没有复位标志而无法进行传输层的干扰）及其他第四层协议无效，对明文数据有效，对加密数据无效。不同的IDS有可能在一段预定或随机的时间内持续干扰刚刚被中断的两计算机间的所有TCP通信。所以在访问境外网站时，如果数据流里有敏感字词，即会立即被提示“该页无法显示”或网页开启一些后突然停止，随后在1-3分钟或更长时间内无法用同一IP浏览此域名或IP地址上的内容，屏蔽时间可能与敏感词等级以及所属网站有关。此种过滤是双向的，也就是说，国内含有关键词的网站在国外不可访问，国外含有关键词的网站在国内不可访问。以上所述的技术，也称为域名劫持，原理如下图所示。

某些特定的海外网站网址会被列入关键词过滤，即使IP地址未被封锁，也不能访问。 不过，GFW对于网页中含有的关键词字符并不是100%可以过滤成功，即使某些网页被成功过滤并导致“该页无法显示”，此时只要在浏览器进行多次刷新就有机会显示出来。而且，GFW还会偶尔出现故障而导致关键词过滤系统失效，此时部分只被网址关键词过滤的网站就能正常使用。

对于Google.com的查询返回结果可能是专门过滤的，即GFW针对Google.com返回结果中的网页地址进行过滤，对关键词的过滤并不严格。

从GFW的分布来看，审查过滤系统主要位于国际出口处，但最近通过对审查过滤系统返回的RST复位包IP头进行TTL值分析，发现存在两个欺骗源，其一位于国际出口处，另一个位于骨干网省级接入处。因此推测GFW对于境内的非法内容也具有一定审查能力。对于境内网络内容的审查可能主要是通过ICP备案来实现的。

从2007年2月前后，GFW开始对境外及境内的WAP网站含有的敏感字符进行过滤，原本在移动版Google可以打开的维基百科中文版现已不能通过Google网页转换功能进行访问，连带的就是在访问含有“zh.wikipedia.org”的Google连结后，5分钟内再次访问Google被拦截。

关键字过滤的弱点就是对已加密的信息无能为力，而网址的关键字和网页的关键字都可以用不同的手段来加密，从而使这样的信息过滤系统从根本上失去作用。不同的加密手段也是后来所有突破网络封锁软件的基础。

（3）、关键词过滤-复位包分析

有些网站含有大量的有用信息，同时也夹杂着大量的有害信息，如Google搜索引擎，如果使用域名重定向、lP地址过滤或者URL过滤都会禁止用户访问合法的信息。在这种情况下，可以使用基于内容的过滤，即只屏蔽掉含有有害信息的页面。

通常使用网址的关键字和网页的关键字过滤的方法屏蔽有害页面。防火墙建有一个敏感词词库，一旦网址或Web页面中的内容含有这个词库中的词时，防火墙将截获该网页，阻止对该页面的访问。

这种过滤是一种细粒度的过滤，实际上是对报文数据内容的过滤。在应用层可以实现对URL的过滤以及报文内容的过滤。应用层有害内容过滤不可避免地降低了互联网的通行效率，并且一般其有较大的误报率，但总的来说监控效果较好。

当前基于内容的过滤主要针对文本内容，对图像、音频、视频等多媒体内容的过滤仍未达到实用阶

段。

由文[7]的试验，可得GFW具体的过滤方式：采用嗅探软件记录HTTP客户端进出站数据包，且只考虑TCP连接。从进站RST复位包IP头TTL域值的分析，可认为逻辑上存在两个欺骗源（实际可能只是初始TTL不同），可分别称为“伪源1”和“伪源2”，伪源1离客户端路由跳计数较大，逻辑位置大致在因特网运营商国际出口处，伪源2离客户端路由跳计数较小，逻辑位置大致在因特网运营商骨干网省级节点处。

1）IP头部分：

Identification（标识）字段：在第一批RST包中，伪源1和伪源2将其设置为一个固定的值，而正常的处理方式是发送的每个IP报文都有不同的标识值，一般按生成次序递增。观察中发现伪源2的第二批RST包中该域值会改变。

Flags（分片标志）字段：伪源1和伪源2处理方式不同，例如伪源1将DF（不分片）标志置0，伪源2将DF标志置1。

Time to Live（生存时间）字段：如前所述，伪源1的RST包到达客户端PC时经过的跳计数较大，而伪源2较小，且可推测与真正的源物理位置有差距。

2）TCP头部分：

Sequence number（序列号）字段：关键词过滤系统很可能会偶而繁忙导致本地出口堵塞，以致RST包发送延迟并晚于真正的源发回的数据包到达客户端PC，造成RST包被客户端PC丢弃，从而整个过滤干预行为失败。考虑到这个因素，伪源还具有序列号预测功能，例如伪源2相邻的3个RST包中该值分别相差1460（以太网默认MSS值）和2920（即1460*2）。

Window size（窗口大小）字段：伪源1和伪源2处理方式不同，例如伪源1似乎为该字段设置了一个随机值，伪源2将其置0。正常的RST包是将该字段置0。 此外还包括HTTPS证书过滤、对破网软件的反制、对电子邮件的通讯的拦截等技术。

从以上的分析可知，GFW的主要技术手段大概有两种：

1）、IP封锁 这种方法主要针对国外知名的新闻网站，比如：http://news.bbc.co.uk/，http://wikipedia.org等，从技术上直接禁止了国内对这些IP地址的访问，或者利用的是国内的域名解析服务，可以将某些网站导向到广告网站或者警告网站。但是，这样的手段只能是重点防卫，而不能全面使用。为了规避IP封锁，只能通过借用国外代理服务器的方式，以国外的代理服务器为跳板，间接的访问这些被封锁的网站，具体的工具如无界浏览器、加拿大大学研究人员开发的Psiphon等。

2）、关键字过滤

针对多若繁星的个人网站，博客网站，社群网站，采用IP封锁的方法就不合适了，对这些网站的防卫主要依靠关键字过滤，比如说，一旦发现内容中包含了china，中国共产党这样的关键字，就切断连接。但是，这种技术手段很难在骨干网和骨干路由器上实现，否则骨干路由器的负担太重，难以保持合理的运行速度。所以，一般的做法是在接入网末端部署具备关键字过滤功能的防火墙，一旦检测到不和谐的关键字，这些防火墙就向两端都发送TCP RST包，让两端的机器以为连接中断了，实际上，原始的TCP包已经通过了防火墙，路本来是通的，只是亮了一下红灯，如果假装没看到红灯，闭着眼睛走过去，反而不会有任何障碍。

三、几种破网软件的原理

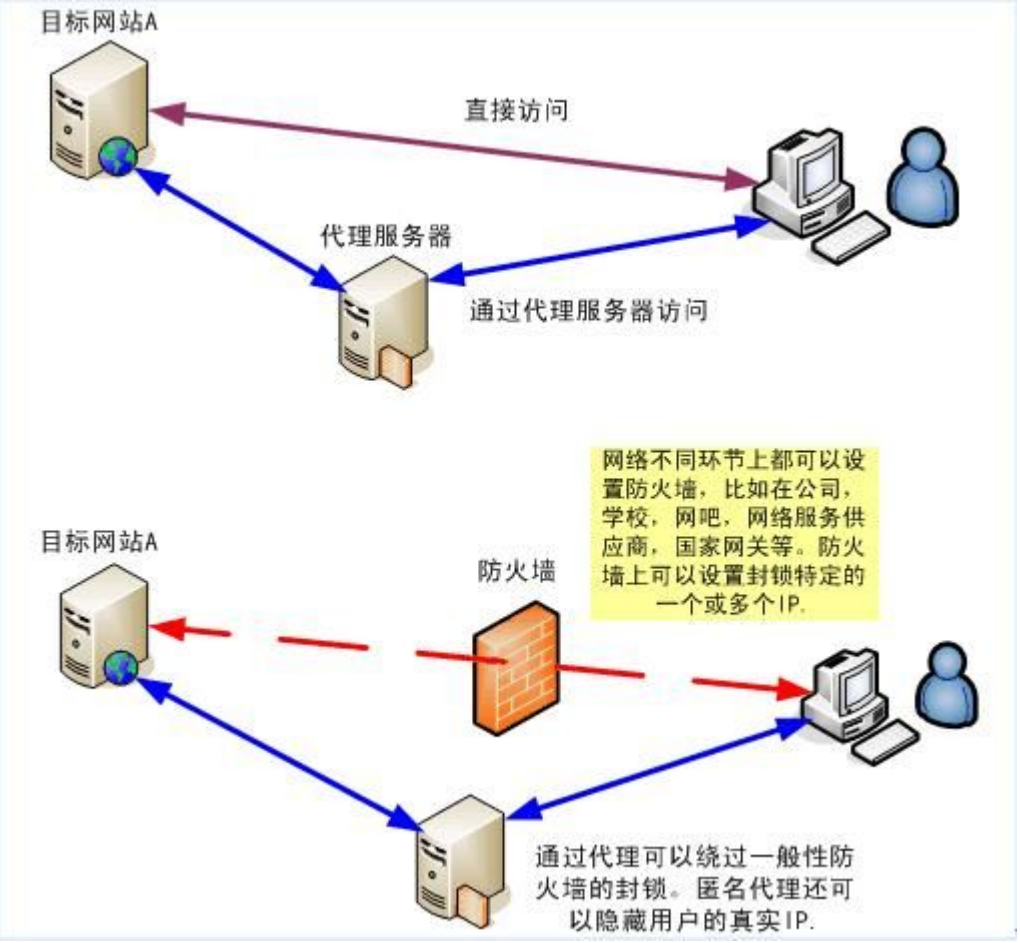
早期的破网软件都是普通代理工具的进一步完善。因为当时还没有采用内容和网址的过滤技术，所以只要找到合适的代理，在浏览器中设置代理服务器和端口，基本就可以畅通无阻。这些工具擅长于代

理的搜索、校验和动态切换，比较有代表性的是“代理猎手”和“MultiProxy”等[6]。

2002年采用关键字过滤技术后，各种加密的代理也就应运而生了。其中比较有名的是SSL加密页面代理，它能够根据用户的请求，把其他网站的内容抓过来，然后用SSL的加密传递给用户。用户使用加密代理，就能够浏览其他各种被审查封锁的网站，而所有的信息都是加密传输的，包括网址URL。但随着软件升级，金盾可以嗅出个别固定域名网站的证书，维基百科443端口的SSL加密浏览也被封住了。

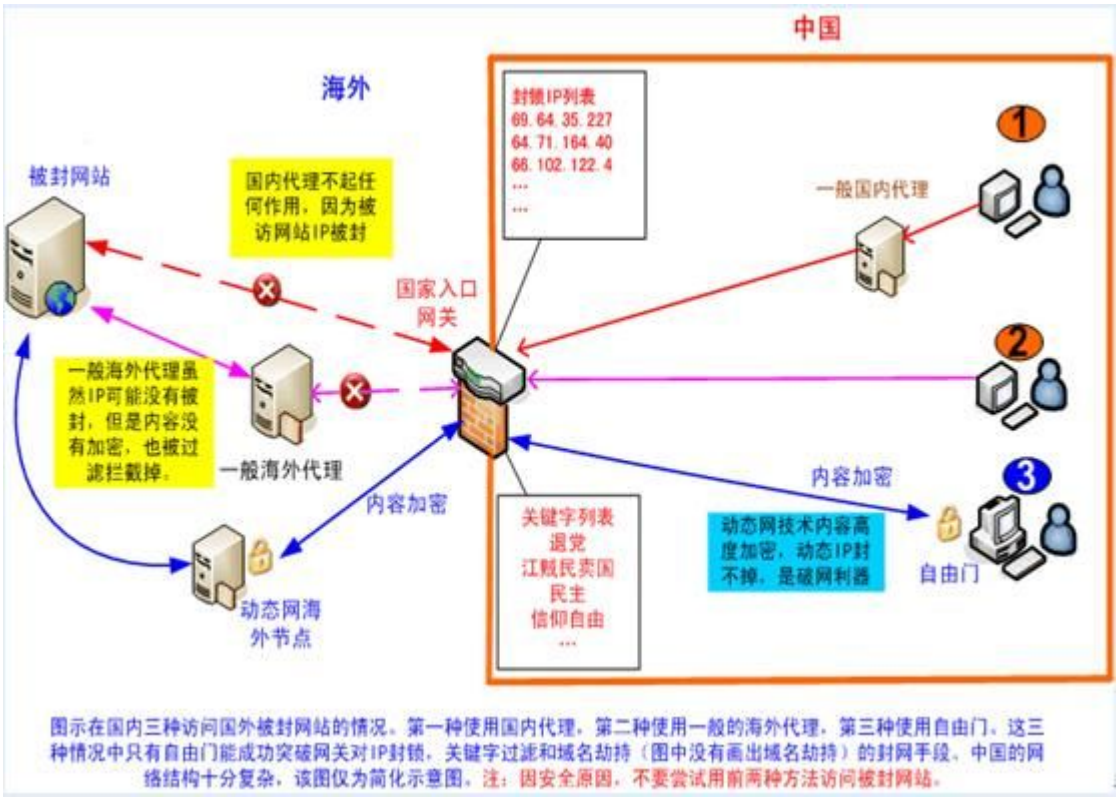
此外利用服务端和客户端的软件，自己定义加密手段，把服务端软件安装在海外的机器后，就可以用客户端软件加密浏览海外的信息了。随着各种加密代理的成熟，它们互相之间也开始吸取各自的长处，有些甚至在易用性上进展很大，做到了用户点击运行就可以自由浏览的程度。

使用这些技术的软件包括无界浏览，自由门，花园软件，世界通，火凤凰等等。其中前三个被称为“破网三剑客”。它们的基本工作原理基本上是一样的：运行软件后自动寻找预置软件服务器列表中的代理服务器，通常寻找最快的那几个，连接成功后自动设置IE，使IE成为代理访问模式。这样就可以直接用IE访问几乎任何网站了。通过代理返回的数据包经过加密，可以有效穿过关键词的过滤，达到可以访问任何信息的目的。这些软件里内置的代理服务器大多设置在国外。



代理服务器可以用于突破防火墙对IP的封锁。如上图所示。但是要突破GFW对海外网站的封锁，一般的代理就远远不够了。

下图中显示了三种用代理访问海外被封网站的情况。第一个用一般国内的代理，第二个用一般的海外代理，第三个用自由门。对于网页浏览器（如IE或火狐FireFox)而言，使用自由门时自由门就是网页浏览器的代理，所有的数据流都是经过自由门加密后传输的，所以也叫加密代理。从图中可以看到，三种情况中只有自由门可以有效的突破网络封锁。如下图所示。



1、Tor原理分析

（1）概述、功能

Tor（The Onion Router），中文叫“洋葱路由”，是一种点对点的代理软件，依靠网络上的众多电脑运行的Tor服务来提供代理，帮助用户抵御流量分析系统，对个人的自由与隐私等进行保护。

流量分析是一种对网络的监视行为，它能够从计算机网络系统中的关键点（如国家级网关）收集分析信息，过滤、嗅探指定的关键字，并进行智能识别，检查网络中是否有违反安全策略的行为，主要进行网址的过滤和网页内容的过滤，如果符合既定的规则，会干扰用户与服务器的连接，使数据流中断，达到禁止访问的目的。 Tor软件将用户的通信通过一个由遍及全球的志愿者运行的中继（relay）所组成的分布式网络转发，以此来保护用户的安全，它令监视用户的Internet连接的那些流量分析系统无法知道所访问的站点，它还令所访问的站点无法知道用户的物理位置。Tor能与现有的许多应用程序配合工作，包括 Web 浏览器、即时通讯客户端、远程登录和基于Internet的TCP协议的其他应用程序。

（2）工作原理

利用Tor软件可以构建一个分布式、匿名的网络来抵御流量分析系统。

Tor有助于降低简单的和高级的流量分析的风险，把用户的流量分散到互联网上的多个地点，所以不存在单一的一点可以把用户和目的地联系起来。在Tor网络上，来源和目的地是由一条通过数台中继的随机的路径连接的，数据包在这条路径上传输，因此，不存在在任何单一点上的观察者能够知道数据从哪里来、到哪里去。

Tor的一个特色是，只要用户运行了Tor server，用户的电脑就成为一个Tor节点，别人可以通过这个节点访问其它节点，用户也可以通过别人的节点进行访问。在Tor节点和节点之间的通信是完全加密的（SSL），所以不用担心你的通信会泄密。当要访问一个地址时，Tor利用一种路由算法在众多Tor节

点中找到一条可达路径。数据经过几“跳”以后，最终能够到达一个可以访问目标资源的Tor节点。Tor的选路过程并不是按照最优的原则，而是随机的。这和它的设计目的：防止数据追踪有关。使用随机的路由就使得数据追踪几乎不可能。以下图1--图3是Tor的工作原理图。

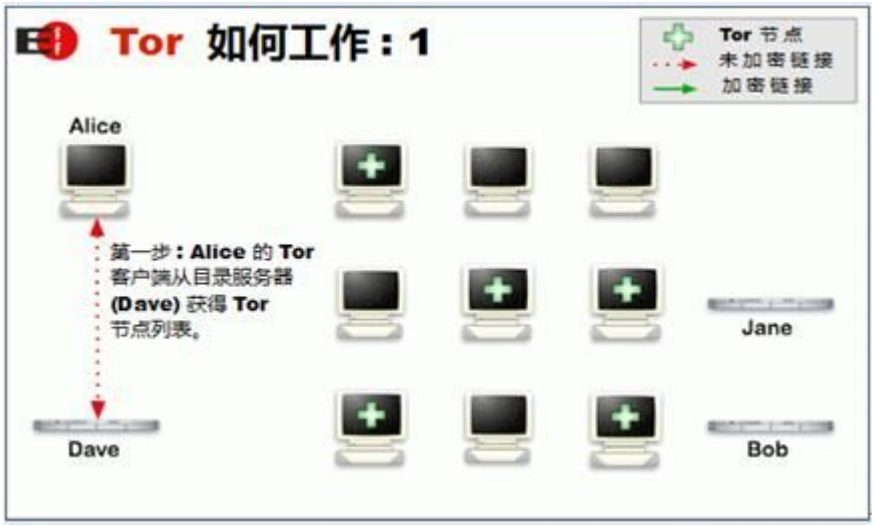


图1 工作原理一用Tor创建一条私有路径时，用户的软件或客户端通过网络上的中继递增地建立一条由若干加密连接组成的电路（circuit）。电路一次扩展一跳（hop），电路上的中继仅仅知道它从哪一个中继接收数据以及向哪一个中继发送数据。没有一台单独的中继会知道数据包的完整路径。客户端为电路上的每一跳分配独立的加密密钥以保证连接数据通过时不被跟踪。

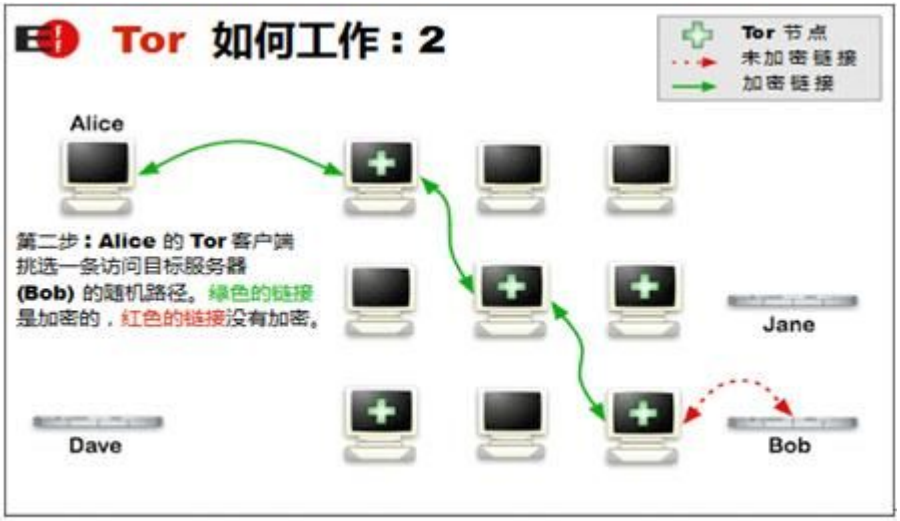


图 2 工作原理二

图2 工作原理二一旦一条电路建立完成，多种类型的数据可以进行交换，不同种类的软件应用程序也可以在Tor网络上部署。因为每一台中继最多只能知道电路中的一跳，窃听者（eavesdropper）或者被入侵的中继（compromised relay）都无法通过流量分析把连接的来源和目的地联系起来。Tor仅作用于TCP数据流，任何支持SOCKS的应用程序都可以使用它。

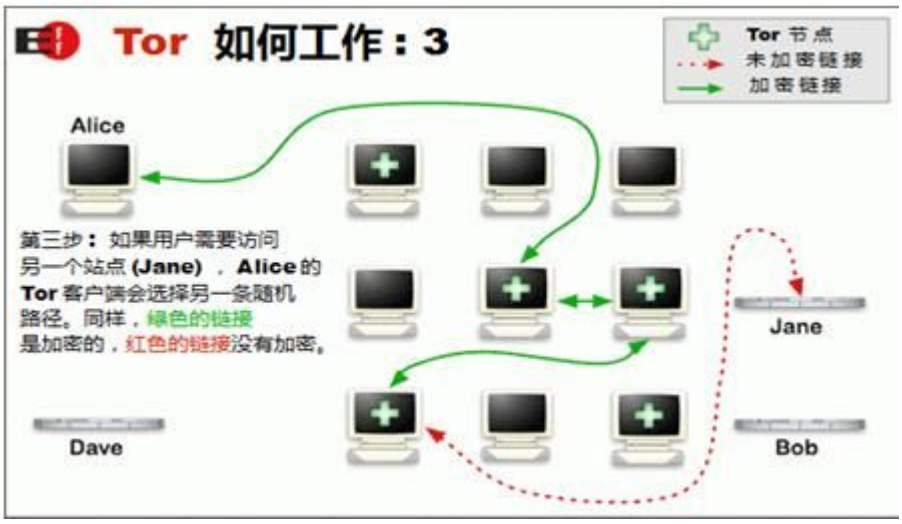


图3 工作原理三出于有效性，Tor 为大约在相同的十分钟内发生的连接分配同一电路。以后的请求被分配不同的电路，这样攻击者就不能把你早先的行为和新的行为联系起来。

（3）技术原理

Tor网络是一个overlay网络，每个节点（onion router，OR）都是运行在用户级，不需要内核级等特权；而且与其他的OR维持着一个TLS的连接；运行本地的代理软件（onion proxy，OP）去获取服务器目录，建立电路回路等。

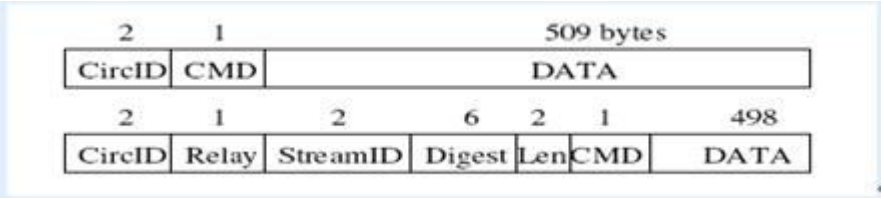
每个OR都维持着一个长期的identity key和一个短期的onion key。Identity key用来给TLS证书做签名，给OR的路由描述符（密钥，地址，带宽，出口规则，等）做签名，给目录做签名（有目录服务器做）。Onion key用来解密用户的请求，然后建立一个电路，协商一个临时密钥。

传输单元

OR同其他的OR，或者OP通讯时，使用了协商好的临时密钥，经由TLS连接进行通讯，将数据隐藏起来，安全的进行转发，阻止了攻击者对数据的修改。

流量按照固定的大小单元（cell）在链路上进行传输。每个单元是512字节，由一个头（header）和一个有效载荷（payload）组成。这个头中包括：一个电路的标识（circID），指明数据单元要经过的电路号；一个命令字段，表明将要对payload做何处理。基于命令类型，cell要么是控制单元（control cell），通常由接收者进行解释处理；要么是中继单元（relay cell），携带着端到端的数据流。控制单元的命令主要有：padding（用来保持存活等），create或created（建立一个新的电路），destroy（销毁一个电路）。

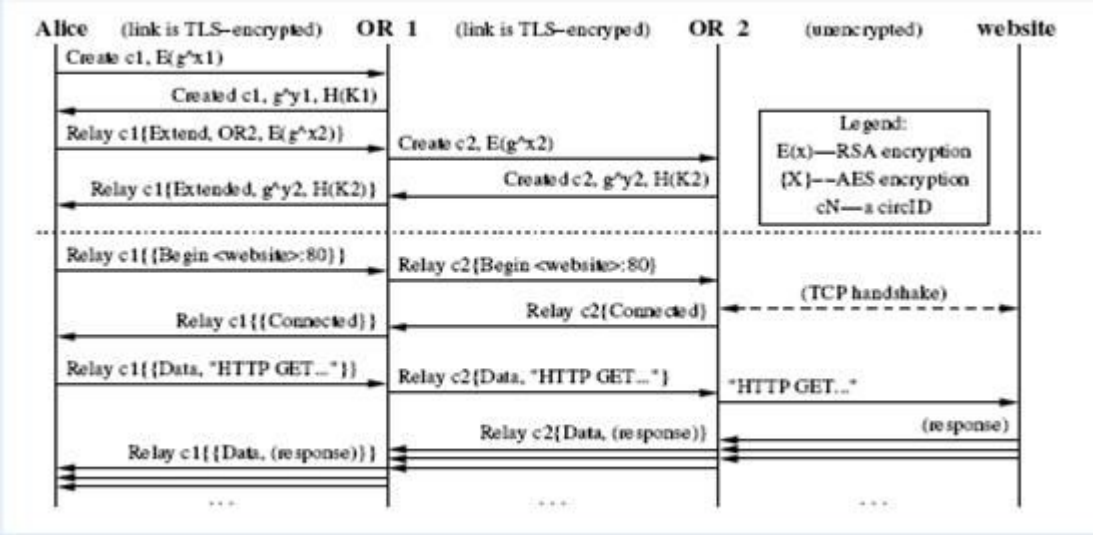
而中继单元在有效载荷的开始又有自己额外的中继头，包含：一个流标识（streamID），一个端到端的校验和（用来做完整性检查），一个中继载荷的长度，一个中继命令。整个的中继头和中继载荷采用了128-比特的AES加密后进行传输。下图为传输单元的格式。



cell结构

电路的建立

下图表明了电路如何建立，密钥如何商议，以及如何访问网页。



Alice建立一个2跳的电路，并访问一个网页

Alice与OR1和OR2的密钥协商：

1） Alice发送一个create命令的包，其中包含了电路标识C1，以及用OR1的公钥加密的密钥，这个密钥是Alice单方选择的。C1是Alice和OR1之间需要建立的电路的标识号。

2） OR1接收到后，回复一个created命令的包，其中包含了，电路标识C1，OR1选择的密钥，以及对双方协商好的密钥K1的Hash值。

3） Alice发送一个relay（中继包）命令给OR1，包的载荷中有一个用K1加密的信息，这个信息中包含了extend命令，OR2名字，和用OR2的公钥加密的密钥。

4） OR1接收到后，用K1进行解密，然后创建一个create包，选取一个OR1和OR2间没有用过的电路标识C2，连同用OR2的公钥加密的密钥一起发送给OR2。

5） OR2接收到后，回复一个created包给OR1，其中包含了自己选取的密钥，和对与Alice协商好的密钥K2的Hash值。

6） OR1接收到后，给Alice发送一个用K1加密的包，包中含有extended命令，OR2选取的密钥，以及K2的Hash值。

7） Alice接收到后，用K1解密，获得与OR2协商好的密钥K2。至此，Alice便与OR1和OR2协商好了密钥。 Alice经由上面的电路访问网页：

1） Alice将要访问的网页先用K2加密，然后用K1加密，将结果放进一个relay包中给OR1。

2） OR1用K1解开后，将结果信息也放进一个relay包中发送给OR2。

3） OR2接收到后，用K2解密，然后与目标网址进行TCP连接的建立。成功后，用K2加密connected命令包发送给OR1。

4） OR1接收到后，替换掉电路标识，然后对加密的信息再用K1进行加密，将最终结果放进relay包中发送给Alice。

5） Alice接收到connected信息之后，就发送Http请求进行访问，以后的通讯与前几步类似。

中继包

一旦Alice建立好了电路之后，就可以发送中继包。接收到中继包的OR，会查询相应的电路，对数据包进行解密，来查看是否有一个有效的校验。如果无效，OR查询电路上的下一个OR和电路标识，将原来的电路标识替换后，把解密后的relay包传送给下一个OR。若最后一个OR不能识别这个relay包，表明有错误发生，这条电路就会被销毁掉。

每个OP对待relay包基本上都是相同的操作，反复对relay包的头和有效载荷用相应的密钥进行解密。如果哪个阶段校验信息是有效的，则表明这个包是在这层的OR上创建的。要针对一个确定的OR创建relay包，Alice先确定一个摘要信息，然后对relay头和有效载荷进行重复加密，所用的密钥是与路径上由远及近的OR商议好的密钥。Relay包中这个摘要信息只有在经过最后一个OR的时候，才会发现是一个有效的值，这个OR得到了Alice真正的目的明文，然后进行目的地的访问。

当这个OR返回给Alice信息时，先用协商好的密钥对这个relay包进行加密，然后按原路返回，每经过一个OR都会进行相应的加密处理，最终传送到Alice时，Alice需要进行多次解密才能提取到真正的信息。

数据流的传输开始与关闭

当Alice的应用程序想要与指定的地址和端口进行一个TCP连接时，她会要求OP去完成这个连接。OP选择一条最新的或者建立一条电路，选择一个合适的OR作为出口节点与目标机连接。然后OP向出口节点发送一个relay begin包，随机选取一个流标识。一旦出口节点连接到目标机，便回复一个relay connected包。OP接收到后，会给应用程序发送一个连接成功的回应信息。于是OP便开始接收应用程序的数据，将数据打包成relay data包进行发送。

当数据传输完毕需要关闭时，也存在与TCP类似的方法：进行两次握手的正常操作，一个握手的非正常操作。非正常关闭，只需要发送一个relay teardown包。而正常关闭是发送一个relay end包。

目录服务器

Tor用一些共知的OR来记录网络拓扑的变化，节点的状态，包括密钥和出口策略。每一个这样的OR称为目录服务器（directory server），就像一个HTTP服务器，客户端可以从这里获取当前网络的状态和路由列表，其他的OR也可以上传自己的状态信息。OR会定期的向每一个目录服务器发布他们的状态信息的签名声明。所有客户端软件都事先加载了服务器的列表资料和他们的密钥。

当目录服务器接收到一个OR的签名声明后，会检查这个OR的identity key是否被识别，服务器不会对未识别的OR进行发布。

2、使用SSH穿越GFW

1）SSH简介

传统的网络服务程序，如 FTP、Pop和Telnet在传输机制和实现原理上是没有考虑安全机制的，其本质上都是不安全的；因为它们在网络上用明文传送数据、用户帐号和用户口令，攻击者可以轻易获得这些数据。而且，这些网络服务程序的简单安全验证方式很容易受到"中间人"（man-in-the-middle）攻击。

SSH是英文Secure Shell的简写形式。通过使用SSH，可以加密所有传输的数据，以抵御"中间人"攻击，而且还能够防止DNS欺骗和IP欺 骗。使用SSH，还有一个额外的好处就是传输的数据是经过压缩的，所以可以加快传输的速度。SSH有很多功能，它既可以代替Telnet，又可以为 FTP、Pop、甚至为PPP提供一个安全的"通道"。

最初的SSH是由芬兰的一家公司开发的。但是因为受版权和加密算法的限制，现在很多人都转而使用OpenSSH。OpenSSH是SSH的替代软件包，而且是免费的，可以预计将来会有越来越多的人使用它而不是SSH。

SSH在运行方式也不像其他的TCP/IP应用，SSH被设计为工作于自己的基础之上，而不是利用包装（wrappers）或通过Internet守护进程inetd来进行。

2）SSH协议的内容

SSH协议是建立在应用层和传输层基础上的安全协议，它主要由以下三部分组成，共同实现SSH的安全保密机制。

传输层协议，它提供诸如认证、信任和完整性检验等安全措施，此外它还可以任意地提供数据压缩功能。通常情况下，这些传输层协议都建立在面向连接的TCP数据流之上。

用户认证协议层，用来实现服务器的跟客户端用户之间的身份认证，它运行在传输层协议之上。

连接协议层，分配多个加密通道至一些逻辑通道上，它运行在用户认证层协议之上。

当安全的传输层连接建立之后，客户端将发送一个服务请求。当用户认证层连接建立之后将发送第二个服务请求。这就允许新定义的协议可以和以前的协议共存。连接协议提供可用作多种目的通道，为设置安全交互Shell会话和传输任意的TCP/IP端口和X11连接提供标准方法。

3）SSH的安全验证

从客户端来看，SSH提供两种级别的安全验证。 第一种级别（基于口令的安全验证），只要你知道自己的帐号和口令，就可以登录到远程 主机，并且所有传输的数据都会被加密。但是，这种验证方式不能保证你正在连接的服务器就是你想连接的服务器。可能会有别的服务器在冒充真正的服务器，也就 是受到"中间人"攻击。

第二种级别（基于密钥的安全验证），需要依靠密匙，也就是客户端必须为自己创建一对密钥，并把公钥放在远程服务器上。如果客户端要连接到SSH服务器上，客户端软件就会向服务器发出请求，请求用客户端的密钥进行安全验证。服务器收到请求之后，先在该服务器的用户根目录下寻找客户端的公钥，然后把它和客户端发送过来的公钥进行比较。如果两个密钥一致，服务器就用公钥加密"质询" （challenge）并把它发送给客户端软件。客户端软件收到"质询"之后就可以用客户端的私钥解密再把它发送给服务器。

与第一种级别相比，第二种级别不需要在网络上传送用户口令。另外，第二种级别不仅加密所有传送的数据，而"中间人"攻击也是不可能的（因为攻击者没有私钥）。但是整个登录的过程可能慢一些。

4）SSH的应用

首先，SSH最常见的应用就是，用它来取代传统的Telnet、FTP等网络应用程序，通过SSH登录到远程主机并执行工作或命令。在不安全的网络环境中，它提供了很强的验证机制与非常安全的通讯环境。实际上，SSH开发者的原意是设计它来取代原UNIX系统上的rcp、rlogin、rsh等指令程序的；但经过适当包装后，完全可以取代传统的Telnet、FTP等应用程序。

而用来替代r系列指令的SSH，则在安全方面做了极大的强化，不但对通讯内容可以进行极为安全的加密保护，同时也强化了对身份验证的安全机制，它应用了在密码学中已发展出来的数种安全加密机制，如 Symmetric Key Cryptography，Asymmetric Key Cryptography， One-way Hash Function，Random-number Generation等，来加强对于身份验证与通讯内容的安全保护。通讯时资料的加密有IDEA，three-key triple DES，DES，RC4-128，TSS，Blowfish 等多种安全加密算法可供选择，加密的key则是通过 RSA 进行交换的。资料的加密可以对抗IP spoofing，RSA这种非对称性的加密机制则可用来对抗DNS spoofing与IP routing spoofing，同时RSA也可以进行对主机身份的验证。

其次，通过使用用SSH可以在本地主机和远程服务器之间设置“加密通道”，并且这样设置的“加密

通道”可以跟常见的Pop应用程序、X应用程序、Linuxconf应用程序相结合，提供安全保障。 SSH的“加密通道”是通过“端口转发”来实现的。客户端可以在本地端口和在远程服务器上运行的某个服务的端口之间建立“加密通道”。然后只要连接到本地端口。所有对本地端口的请求都被SSH加密并且转发到远程服务器的端口。当然只有远程服务器上运行SSH服务器软件的时候“加密通道”才能工作。

5）利用SSH突破GFW

我们知道，GFW对国外敏感网站的封锁主要是通过ip限制与关键词过滤来实现，SSH显然可以突破关键词过滤拦截，即SSH通过创建一条加密隧道来防止特定关键词被GFW发现。以下介绍利用SSH突破GFW的限制的原理。

几个相关的概念。

SSH客户端,SSH服务端,应用程序客户端,应用程序服务端。

就突破GFW限制这个应用而言，SSH客户端与应用程序客户端都位于本地，而SSH服务端与应用程序服务端位于远端。（注：也可以在本地创建SSH服务端，让应用程序客户端与SSH服务端都位于本地，而SSH客户端与应用程序服务端都位于远端。然后通过一个SSH的反向连接来达到目的。然而，多数的应用还是前面讲述的情形）。那么，为了突破GFW的封锁，我们有什么要做的呢？

首先，应当在国外有一个运行SSH服务的主机，至于如何获得这样的主机不在本文的讨论范围内（无论你是通过黑客入侵的手段，还是租用此类服务等）。SSH的服务应当位于国外，否则将无法规避GFW的IP限制。

其次需要运行一个国外代理服务器，如果没有，本地的应用将受到限制。因为此时本地的应用将被局限在某一个被限制的站点，而不是一些。如果在运行SSH服务的主机上有足够的权限，当然也可以在其上安装相应的代理服务。

第三是让本地的SSH客户端监听某一端口，一旦该端口有数据要传送，SSH客户端将会在在本地与远端的SSH服务端建立一个加密连接。然后被监听端口的数据将加密传送给SSH服务端，而SSH服务端在收到数据后，将做相应的端口转发。这样做能规避GFW的原因在于：

1）运行SSH服务的主机并非被禁止的IP,因此你可以与其建立连接。

2）本地与远程SSH服务的主机所传送的数据都经过加密，因此GFW防火墙并不能轻易发现敏感信息。

下面具体举一例说明：

比如想通过web方式访问维基百科，而通常情形下这是被GFW所禁止的。所要做的是，先在本地SSH客户端执行如下命令： ssh －L9999:proxy:proxy_port sshd_server 这个命令的用途为：让ssh客户端监听9999端口，如果有数据，将其加密传输至ssh服务端，而服务端则相应向proxy的proxy_port转发。命令说明：9999是客户端的本地端口，也就是SSH客户端所监听的端口。Proxy,proxy_port分别为代理服务器的名字（或IP）以及它的服务监听端口。sshd_server是运行SSH服务的主机。这么做以后，只需将本地的浏览器代理设置为localhost:9999。接下来就可以顺利访问维基百科了。（本例中proxy可为http代理或socks代理）。

3、自由门

自由门可以非常安全的让你自由畅游网络世界。其加密强度程度可以与国际金融系统的相比，会自动隐藏你的IP，任何人绝对看不到你的IP。其最新版为自由门6.34版，修复了个别情况下代理密码的设置问题。

该软件的功能就是连接代理服务器，客户端和代理服务器端之间的数据经过高强度的SSL加密，数据传输速度很快，它自动搜索到代理服务器，连接并使用，访问美国骨干网络的速度快很多，能突破多种限制。

自由门软件工作模式分为：代理模式和经典模式。在经典模式下程序会自动设好与密道有关的设置文件并设好IE代理。

自由门软件其实质是用加密代理服务器的技术。为确保包含关键字信息的数据包不被GFW截获，加密代理服务系统器需要与客户端进行安全通信。一次安全通信分为密

钥建立和保密通信两部分。这里详细介绍密钥建立和广泛应用的网络安全套件SSL协议。自由门软件界面如下图所示。

(1)密钥建立

在客户端和加密代理服务器开始交互大量信息之前，两者需要建立一个安全的信道。于是，双方需要进行密钥建立以确定本次通讯所使用的公共密钥。

密钥建立分为无服务器的对称密钥建立和基于服务器的密钥建立，可根据代理服务器的类型和客户端的数量自行设计。无服务器的对称密钥建立又包括点对点密钥更新(有共享长期密钥)和无预先共享密钥的密钥建立两种。无预先共享密钥的密钥建立可以使用Diffie-Hellman密钥交换协议来实现分发公共密钥。其缺点是容易受到中间人攻击。

(2)SSL

SSL协议用来在客户端和服务器之间建立安全的TCP连接，并向基于TCP/IP协议的客户/服务器应用程序提供客户端和服务器的验证、数据完整性及信息保密性等安全措施。主要用于浏览器和Web服务器之间建立安全的数据传输通道，还适用于Telnet、FTP和NNTP等服务。

然而，SSL协议在具体使用中还面临如下3个问题：

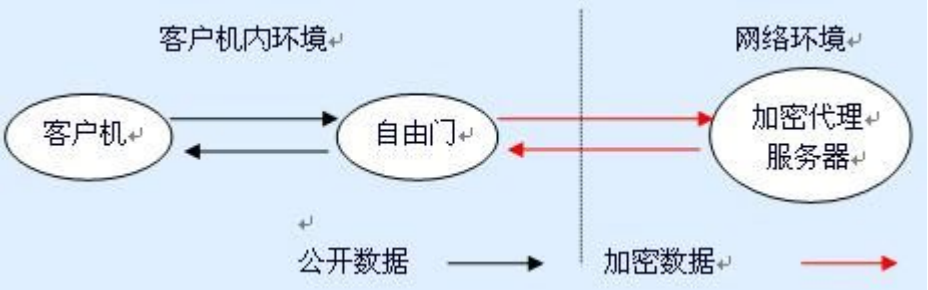
①客户端对服务器的身份确认；

②服务器对客户的身份确认；

③在服务器和客户之间建立安全的传输信道。

(3)加密代理服务器的实现

加密代理服务技术使用了加解密技术。只有加密通信双方使用相同的协议，客户端才能连接到加密服务器的端口进行访问。因此，如下图所示，客户端需要运行加密代理软件，来实现应用程序和加密服务器之间的信息转发和明密文转换。



(4)自由门技术的改进

目前GFW侦测类似自由门软件的方法主要根据它们数据包中的特征码方式进行侦测，并进行截获。并对数据包进行IP包进行跟踪，查获到境外的代理服务器后，将其IP地址列入黑名单中。

在自由门的新版本中改变以前的特征码，使数据包的特征码在不断变化中。以及改变数据包经过SSL加密后的特征码。定期对加密代理服务器进行IP更换措施，并将更换后的加密代理服务器的IP地址通过BBS,Email，聊天室内进行公布。

4、无界浏览器

无网界浏览 8.8是由美国极景网络科技公司推出的高品质软件产品。原理是根据其内置（或软件服务器上）的代理服务器列表不停查找选择速度最快的代理服务器。联机完成后会自动帮IE设定好HTTP代理服务器：127.0.0.1:9666，如果用其他浏览器也可以手动设定使用。在无界退出时，因为自动清除了主机上的所有使用信息，所以如要再使用无界浏览器，代理软件将重新搜索可用的加密代理服务器。使用动态SSL代理服务器，完全规避了防火墙的屏蔽。

（1）无界浏览的使用方法

1）启动该软件的可执行文件；

2）如果需要的话，添入代理（局域网需要在这里添代理，普通的拨号用户不需要）；

3）无界浏览会自动寻找海外代理，找到后，右下角会弹出一个金黄色的锁，这表示无界漫游启动成功了；

4）接着一般会弹出一个Internet Explorer窗口，此时的IE已经被设置了无界浏览的代理127.0.0.1:9666，默认访问的网站，就是无界网络；

5）用户再通过IE访问网络时，都通过无界浏览找到的加密代理，进行数据传输。

6）如果使用其它浏览器（如Mozilla Firefox）访问，无界网络启动成功后，可自己手动设置代理127.0.0.1:9666即可。

（2）版本沿革

1）6.9版

2）网络上曾出现无界浏览7.0版，极景网络公司声明于此版本无关，并非自己所出品，完全系被恶意冒用，并跳过7.0版直接推出了8.0版。

3）8.2版在2007年8月初遭到封锁。

4）8.3版于2007年8月6日推出的。

5）8.4版在2007年8月16日推出。

5）8.5版在2007年8月19日（美国时间）推出。

6）8.6版在2007年9月17日（美国时间）推出。

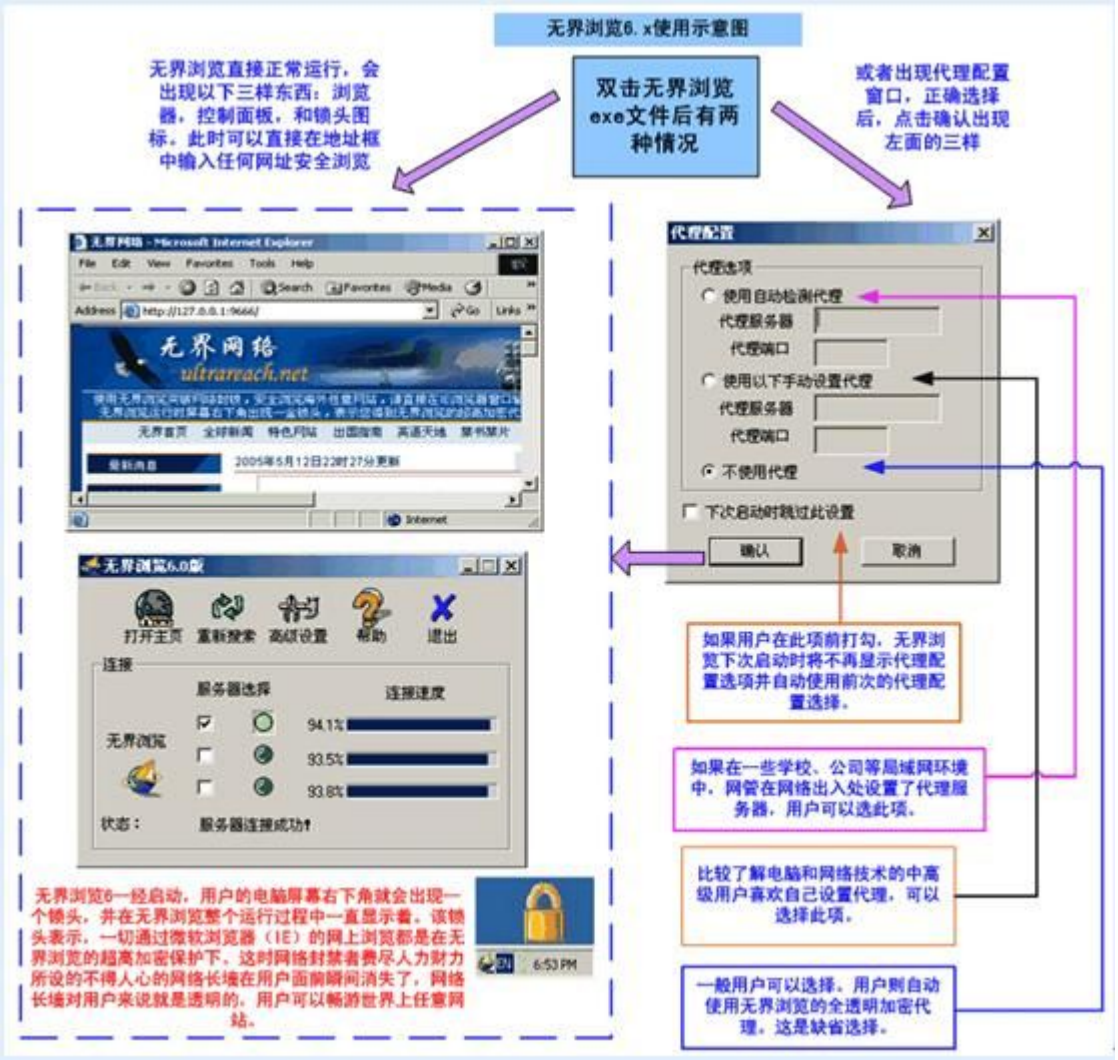
7）8.7a版(测试版)在2007年10月26日推出。

8） 8.7b版(测试版)在2007年10月27日推出。

9） 8.7正式版在2007年10月30日推出。

10）8.8版在2007年11月20日推出。

（3）无界浏览器使用示意图：



（4）特点：

1）、几乎可以访问国外所有的网站。
2）、速度快，比一般的代理服务器速度快很多，而且越多人访问的网站，速度越快。

3）、传送过程高度加密。在传输过程中将网页地址(URL)和内容都加密。

4）、当退出运行时，会清除所有访问记录。如果是非正常退出，当重新运行并正常退出后，所有访问记录也都将被清除。

5）、支持多媒体文件传送与下载，包括声音与图像文件。

6）、网站集锦栏目特别收集海外被禁网站连结，为有兴趣的用户浏览导航。

7）、特别支持Google, AltaVista 等搜索引擎，使搜索到的内容不被过滤。

使用无界软件无法访问国内网站，这可能与GFW屏蔽了国外破网软件的IP地址有关。解决方法是退出无界再重新打开，直到无界的出口IP换成没被屏蔽的IP地址为止。

5、其他穿越GFW的技术

除了以上介绍的四种破网技术外，还可以通过忽略TCP RST包的方式以通过具备关键字过滤功能的防

火墙。也有更专业的技术论文详细解释了其技术原理和实践方法，具体请请参考ignoring.pdf。所提出的穿越GFW的解决手段，对于Linux，用如下指令：

iptables -A INPUT -p tcp —tcp-flags RST RST -j DROP

对于FreeBSD，请用对应的指令：

ipfw add 1000 drop tcp from any to me tcpflags rst in 点击不能访问的链接，即可实现正常访问。

四、对破网软件的控制

针对突破防火长城的各类破网软件，防火长城也在技术上做了应对措施以减弱破网软件的穿透能力。比如每年的特定关键时间点，无界等软件就可能会无法正常连接或连接异常缓慢，这时境内外的正常网络互联也会受到干扰。

1、对Tor

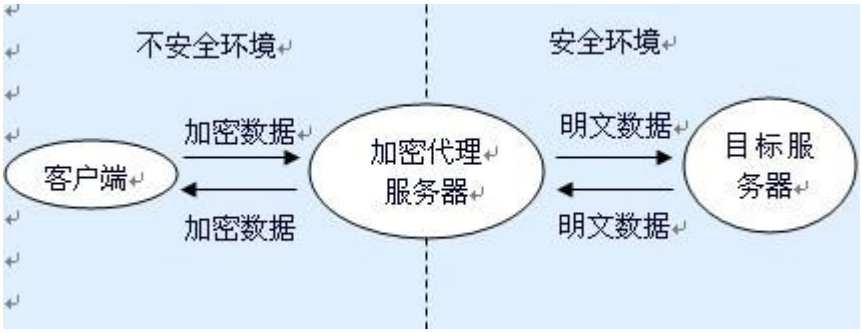
可采取建立虚假Tor节点的封锁措施。鉴于无法真正的完全封锁Tor，网络安全部门可在国内网络中安装了大量虚假 Tor节点服务器，所有经过这些“节点”的信息都将被最大程度的审查，与此同时，所有到达这些虚假节点的网络请求都将被屏蔽。有意见认为因为此举会暴露防火长城的位置，网络审查部门对虚假节点的设立有所节制。但另一方面，tor节点的大量增加很可能仅仅是因为国内用户增加的缘故，即使存在有虚假节点，对于使用图形界面Vidalia的用户也可以轻松将含有境内节点的路由删除，以确保安全。

2、对加密代理型浏览器

主要针对无界浏览器进行分析。

（1）无界加密代理的工作原理

加密代理服务器的工作原理见下图。



（2）网络活动概述

网络活动步骤如下：

1)、探测。访问国内外知名网站，用来探测是否接入互联阴。

2)、访问。访问“专职DNS服务器”。这些DNS服务器专门为无界提供信息更新支持。

3)、加密。访问加密代理服务器。

（3）各步骤的发包规律

1)、探测。随机选择4~6个知名网站进行测试。由于程序存储有限，因此可以找到所有被该程序记录的网站的IP。

2)、访问。这一部分是规律最多、最容易识别的部分。所有的DNS数据包解析地址均为[ns 1．4546355dc．net，即“ns”+数字1或2+“．”+9位十六进制编码+“ net”。由于这样的域名与常规域名存在较大的差异，因此容易发现并跟踪。但是，这个步骤仅出现于程序的第1次运行中。捕获这样的数据包的概率相对较少。

3)、加密。加密代理把秘密信息装扮成普通HTTP数据包的形式。实现和机密服务器的保密通信。由于网络上有海量的HTTP数据包，因此即使找到加密数据的规律也很难及时地进行数据包过滤查找。但是，当发现加密代理向加密服务器申请信息时，HTTP数据包中总有“GET”字段，所接收的URL地址有如下规律：

1)包含4~6个由“/'，分隔的部分；

2)每部分由3~6的英文小写字母组成；

3)字母随机组合且不是单词。

（4）侦察监控方法

无界v6．9的侦控方法主要有IP跟踪法和数据包筛选法两种。其中，IP跟踪法速度快效率高，但准确性欠佳；而数据包筛选法则有较好的准确性。

1）、IP跟踪法

通过运行加密代理软件，采用截取数据包的方法，可以获得相应的IP地址信息。对IP地址信息的跟踪分为3类：

a)、对专职DNS服务器的跟踪。这是侦察工作的重要环节，也是唯一可以全面控制加密信息来源的环节。专职DNS服务器的特点是：1)固定。每个版本的加密代理软件只能包含有限个专职DNS的IP地址。即使加密代理软件随机地使用全部IP中的几个，也可以通过多次跟踪找到绝大多数的IP地址。2)专职。这种DNS服务器只向相应的加密代理软件提供所谓的“域名解析”，而不提供通常意义的域名解析。将获取的专职DNS服务器的IP地址存放在数据库表DNSIP中。

b)、对加密代理服务器的跟踪。尽管加密代理服务器的IP地址是经常变换的，但是它在短时间内具有相对的稳定性，比如同一地区同一天内不会变化。可以建立一个IP地址采集机制，对保存时问超过一天的IP地址进行重新检测，以确认它是否已作废。将获取的加密代理服务器的IP地址存放在数据库表CPRIP中。

c)、对加密代理软件用户的跟踪。通过对专职DNS和加密代理服务器IP地址的收集，可通过数据包的简单分析查找到使用加密代理服务的用户的IP地址。将这些IP地址保存起来就可以对它们进行进一步的分析，从而找到可疑对象。将获取的加密代理软件用户的IP地址存放在数据库表USRIP中。

2)、数据包筛选法

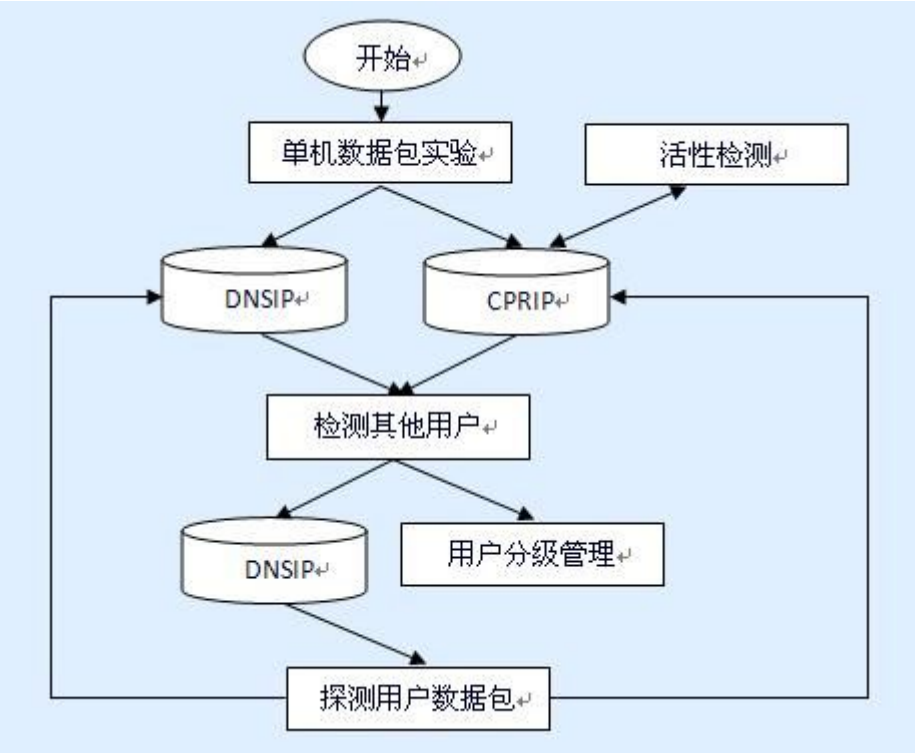
面对海量的DNS数据包和HTTP数据包，为了提高效率，将针对数据包的信息过滤作为IP跟踪法的辅助方法。即重点跟踪IP地址在表USRIP中的用户的DNS数据包和HTTP数据包，从而找到新的专职DNS和加密代理服务器的IP地址。利用上文中加密代理数据包的特征规律，对DNS数据包和HTTP数据包采用不同的过滤方式：

a)、DNS：过滤域名为“ns”+数字1或2+“．”+9位十六进制编码+“．net”的数据包。

b)、HTTP：过滤URL有上文提到的3个特点的数据包。确认获取的DNS和Web服务器是所需的专职服务器和加密代理服务器。将获得新的专职DNS和加密代理服务器的IP地址存放在表CPRIP和表CPRIP中。

（5）具体工作流程

通过以上的2种方法，就可以实现对该加密代理软件的应用范围的扩大寻找，对本地区网络的全面覆盖。具体操作流程见下图。



具体工作步骤如下：

1)、通过单机截包实验，找到为实验机提供信息的“专职DNS服务器”和“加密代理服务器”。并将IP地址分别存放在数据库表DNSIP和CPRIP中。

2)、通过在大规模的网络出口节点(例如各省的网络出口)截取具有数据库表DNSIP和CPRIP中IP地址的数据包，将使用该软件的用户的IP存放进数据库表USRIP中。

3)、对数据库表USRIP中的IP按照获取次数的频率进行分级。

4)、对数据库表USRIP中的频率级别较高IP进行跟踪，通过数据包筛选法，寻找新的“专职DNS服务器”和“加密代理服务器”，并将IP地址分别添加在数据库表DNSIP和CPRIP中。

5)、对数据库表CPRIP中的IP进行活性检测，定期去除已经失效的“加密代理服务器”。在整个监控系统初具规模后，可以做以下的工作：

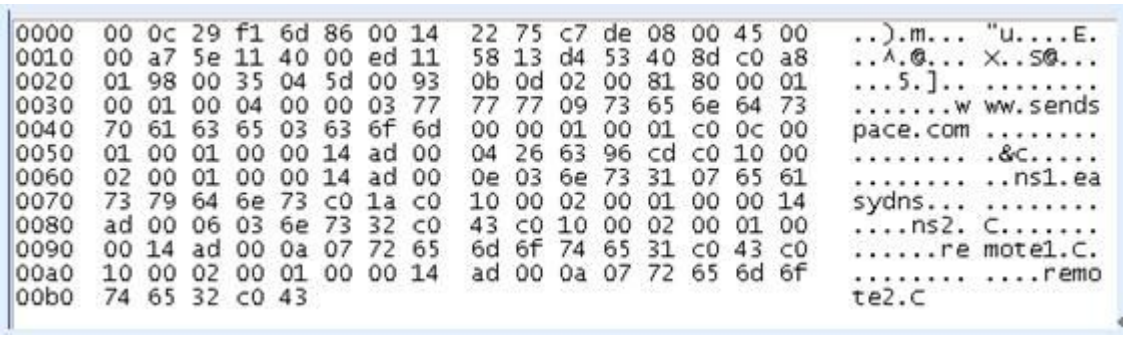
1)、全面封杀该破网软件。因为一个版本的无界浏览器只可能预设有限个“专职DNS服务器”，所以只需要在数据库表DNSIP相对稳定时，封锁所有来往于所有“专职DNS服务器”的数据包即可。

2)、掌握使用无界浏览器的用户情况。通过对数据库表USRIP的操作，可以发现使用该破网软件的用

户的分布情况，还可以对经常性的用户进行重点控制和深入的走访排查。

采用本方法，可以在较短的时间里收集到网络上的相关破网软件的详细信息，为掌控代理工作情况、用户状况等相关信息提供了必要的保障。破网与补网、渗透与反渗透对于国家安全来说是一对矛盾。破网软件设计者可以在以后的版本中通过技术改进，逃避现存的侦察手段。但是，只要它们还是面向普通网络用户的宣传工具，就可以找到数据包的规律，进而实现对它们的跟踪和监控。

对无界8.9的侦察仍然可以采用以上的方法来实现，下图为用WireShark在局域网中捕获无界第一次运行时的数据包，从中可见，仍然具有上文所述的特征。



对花园浏览器和自由门浏览器的侦测依然可以采用本文中的方法。就是先对一个普通的浏览器客户端进行小规模的局域网内的截包实验。在找到相应的数据包规律后，在大型网络的出口上进行动态的监测。

上述方法是基于有特定格式数据包的检测来实现的，如果在数据包中不使用有格式的字符串，或定期变换这些字符串，则可成功规避GFW的检测。

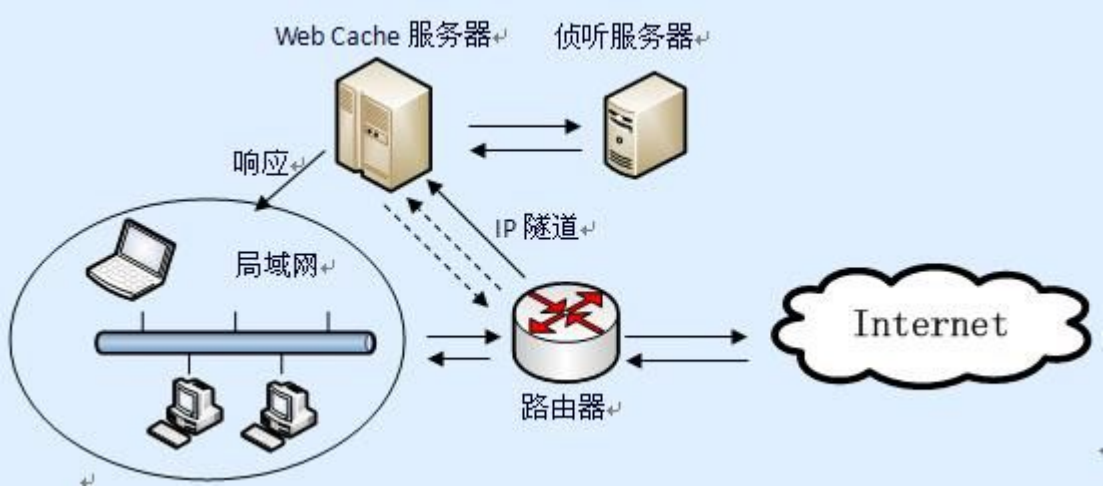
另外，如果引入新的密码学技术，例如使用信息隐藏技术，将“加密代理服务器”的IP地址隐藏在一些知名门户网站的合法信息中，这样也可以有效地避免这种“专职通告”机制受到监控和破坏。

3、一种基于透明Web Cache的内容过滤方法

通用的内容过滤技术都是基于实时的“事先判别”技术，即在用户浏览网页之前先期进行内容判别，对网页进行内容分析和过滤。其缺点是时延长，实时性差，准确率低，往往影响用户的浏览速度，而对设备性能的也要求较高。面对实时内容过滤中存在的这些问题，文[8]提出的基于高速缓存服务器(Cache Server)的过滤系统框架，对系统缓存的网页数据进行“事后审计”方式的内容过滤，据此生成过滤用的黑名单数据库，并配合侦听匹配阻断服务系统。该模式采用内容分析和网址过滤相互协同、分时工作的方法，可很好地提高内容过滤的准确性和实时性。

1）逻辑架构图

本系统架构是在位于出口路由器上的缓存服务器旁边加装了一台侦听服务器，该侦听服务器不占用路由器资源，它本身通过网络和缓存服务器相连，路由器支持WCCP(WebCache Communication Protoco1)协议，缓存服务器工作在透明模式下。



2）透明缓存服务器的工作原理

透明缓存的意思是客户端根本不需要知道有高速缓存服务器的存在，客户不需要在浏览器中设置任何代理，只需要设置缺省网关，客户访问外部网络的数据包都被发送到缺省网关，而这时缺省网关处运行有一个缓存服务器，数据实际上被重定向到缓存服务器的代理端口(如3128)，即由本地缓存代理服务器向外请求或直接提供所需数据，然后拷贝给客户端。

要完成透明缓存代理，目前所普遍采用的技术就是WCCP。WCCP协议是由Cisco公司提出并于1997年正式发布的，至今已有V1和V2两种版本。WCCP V1所要服务的资料类型仅仅是HTTP的资料类型。WCCP协议主要的功能是提供路由器和缓存引擎之间透明重定向的机制，将用户的请求在经过路由器时，利用GRE(Generic Routing Encapsulation，通用路由封装)技术封装起来，再送往缓存服务器，缓存服务器收到之后，解开GRE封包，并解读其中的HTTP请求，然后检索缓存内容，如果缓存服务器储存了符合用户所要求的资料，则缓存服务器便直接将该资料回应给用户；否则缓存服务器再向外界网站抓取资料，抓取完成之后，缓存服务器将资料同样用GRE封包，送回路由器，接着路由器改写封包还给发出请求的客户。WCCP这种运作机制对于用户来说是毫不知觉的。而如果WCCP沟通失败或缓存服务器发生问题时，用户的请求会完全不受影响地被路由器传送到目的地，用户需要的服务也不会遭受任何的中断。WCCP技术可以有效地降低Intemet网络流量，节省广域网链路费用。

3）黑名单—— 不良网址过滤数据库

数据库过滤技术是将用户请求的IP或URL与不良信息库进行比对，阻断数据库中存在的不良站点。不良信息库即黑名单，它采用一个驻留于缓存服务器中的类似于语义识别网络机器人技术或单一功能的进程，针对Cache中的缓存数据，根据过滤规则在缓存服务器相对空闲时进行内容安全方面的过滤，并记录符合过滤规则的网页对应的URL地址和IP地址，然后登记于文件中，再传给侦听服务器处理，进而形成过滤数据库。因为IP地址和URL地址都是分级的，如URL=协议名称+宿主名+路径与文件名，所以数据库采用目录型数据库，按照宿主机名、路径、文件名等组成分级树型数据结构，还可根据需要生成必要的索引。

过滤数据库中包括了：色情、恐怖、邪教、赌博、暴力、毒品、黑客等类型的不良站点，且每天都可更新，以确保内容过滤引擎和互联网的发展相一致。

4）网络数据包的侦听、匹配与阻断

利用网络中信息的传输是在用户端与服务器端之间进行这一特性，可以在这两端之间进行数据监听。在以太网上，任何一台主机发出的数据包都是在共享或交换以太网传输介质上传输的，每个数据包的包头部分都包含了源地址和目的地址。如果需要让一台主机能够接收所有的数据包，即进行网络数据包的“侦听”，只要设置该主机的网卡工作在“混杂模式”下(对交换网络需要设置流量镜像)，则不

论数据包的目的地址是否本机，都能够截获并传递给上层进行处理。

网络数据包的侦听可以使用一些现成的开发包来实现，如WinPcap和libpcap是比较著名的开发包，提供了较强的网络数据包截获功能，或者利用现有的监听工具，如Snifer、Netxray、Tcpdump等工具就可以轻而易举地截取数据包。

对于截获的数据包，拆包进程经过P—＞PH—＞GREH—＞IP—＞TCP—＞HTTP等层层拆包，提取出IP、URL等信息，通过将最常用的数据放入内存的预取策略，首先在内存中与部分黑名单数据进行快速匹配，如果匹配不成功，再进行全面匹配；而如果匹配成功，则启动阻断进程，进入会话阻断阶段。

阻断的方式有：与防火墙联动、中断TCP会话、阻塞HTTP请求、模拟SYN/ACK等。“阻断会话”机制是目前IDS最常用的方式，它既不需要外部设备的支持(如防火墙)，而且易于实现。可利用TCP/IP的工作原理来设计 。TCP使用端到端的连接，即TCP用源IP，源TCP端口号，目的lP，目的TCP端口号来唯一标识每一条已经建立连接的TCP链路。TCP对话通过三次握手来完成。三次握手的目的是使数据段的发送和接收同步；告诉其它主机一次可接收的数据量，并建立虚连接。其三次握手的简单过程如下：

a)发出请求的主机通过一个同步标志置位的数据段发出会话请求。

b)接收主机通过发回具有以下项目的数据段表示回复：同步标志置位、即将发送的数据段的起始字节的顺序号、应答并带有将收到的下一个数据段的字节顺序号。

c)发出请求的主机再回送一个数据段，并带有确认顺序号和确认号。

若能匹配出合适的信号则会向通信的两端各发送一个TCP RESET包，从而实现主动切断连接的目的，此时通信双方的堆栈将会把这个RESET包解释为另一端的回应，然后停止整个通信过程，释放缓冲区并撤销所有TCP状态信息。阻断的流程如下：

a)伪装成Server给Client发一个RAT包；

b)伪装成Client给Server发一个数据包；

c)Server回一个ACK包给Client；

d)因为Cleint的连接已经给RESET掉了。所以Client回一个RST包给Server。

五、可能的破网方法：

1、使用内容压缩、加密、变换和信息隐藏

对网页内容进行压缩、加密和变换后，可极大地增加基于内容过滤防火墙的过滤难度。文本内容压缩后，为保证防火墙的工作效率，一般不会对压缩包解压后过滤；通信内容加密后，在文件名或网页上公布解密密钥，防火墙对此是无能为力的；也可以对关键字做变换，由于目前基于多媒体内容的过滤还有较大的难度。文本可以以图片方式显示在页面上，这可以逃避基于文本的过滤，或者通过文本置乱(如在文本中夹杂一些特殊字符)、文本代换(如以谐音、拼音、外文代替)等简单的信息隐藏方式逃避过滤；利用复杂信息隐藏技术完全能够将网页的受限信息隐藏于可见的Web页面中，达到保护真实信息的目的，封防火墙来说发现网页中隐藏的信息将是一项不可胜任的工作。

2、利用主机漏洞搭建代理，自己创建动态代理服务器。

漏洞是指硬件、软件或策略上的缺陷，使系统受到未经授权的访问。攻击者利用扫描软件捕获国外有漏洞的一台普通的计算机主机，在该傀儡机上搭建加密代理(服务器)后，进行加密连接，再利用该主机去获得想要得到的受限资料，加密传回，达到突破络络封锁的目的。由于傀儡机本身是受害机器，

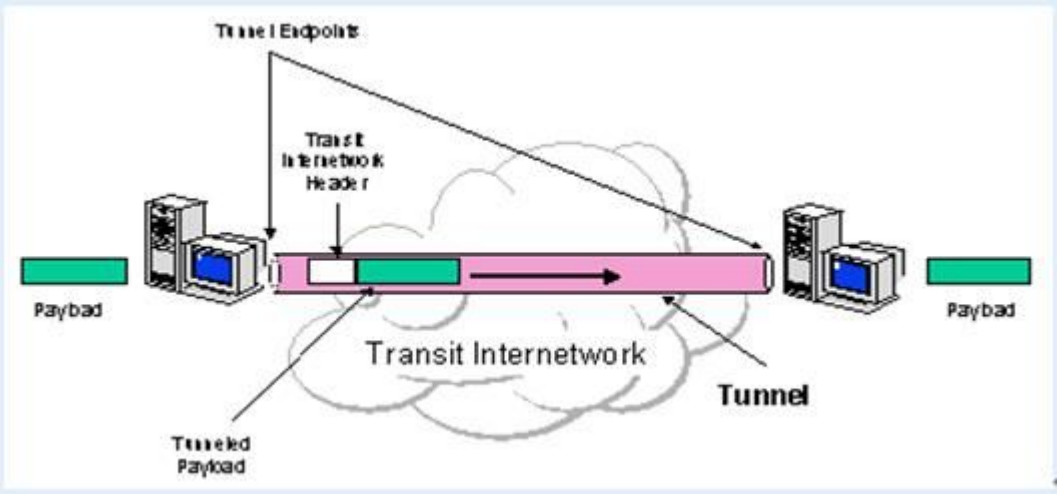
并其地域分布具有较大的随机性，这给防火墙监控提出了难题。

为了使利用漏洞搭建的代理长期有效，还可以在该傀儡机上装载代理型木马。代理型木马的服务端(被控制端)定时监测客户端(控制端)的存在，一旦发现控制端上线就立即主动连接控制端。为了隐蔽起见，控制端的被动端口一般开在80，这样，即使用户使用端口扫描软件，也会以为是在浏览Web。代理型木马的服务端可以通过代理获取控制端的lP地址。如事先约定好一个个人主页的空同，控制者上线后自动上传一个文本文件，内容是通过加密的lP地址。木马每隔一定时间取一次这种文件，如果文件内容为空，就什么都不做，如果有内容就按照文本文件中的数据计算出控制端的IP地址。代理型木马全部使用HTTP协议进行通信，能够有效地逃避包过滤。攻击者可以通过代理型木马(甚至多级代理型木马)访问受限站点，有效地逃避监控。

3、VPN技术

VPN，又叫虚拟专用网，相对于以上介绍的两种方法，VPN可能是一种更快、更受青睐、更正式的方法。本质上，VPN是用正常的信道建立一条专属的加密信道。VPN可以将国内客户机连上海外的某个服务器。客户机的下载及浏览请求就会传送到美国、芬兰或是日本的服务器，然后这个服务器去发现并将客户机要找的东西加密传输回来。GFW将无法阻止这种经过加密的通信。目前，在中国的外国公司几乎都在使用这样的网络。而且，VPN在国内的使用没有受到限制，因而个人也能使用。缺点可能是个人需要支付一定的费用。

VPN的具体实现可参见[15]。下图为使用VPN“隧道”技术示意图。



六、结语

从上面介绍的几种软件成功破网的事实可知，当前GFW的策略是：如果通过GFW的信息由于加密而不能识别，就挥手放行。在技术上讲，GFW可以随时切断所有代理服务器和VPN连接，但这种做法的后果是极其严重的。因为银行、外国制造商、零售商、软件厂商等与国外有业务往来的单位或部门都需要成功穿越GFW的应用技术才能存在。可以想象，如果商业明文信息通过公众互联网或GFW传送，会有什么样的后果，可能没有哪个公司能冒这样的风险。同样，如果GFW关闭免费、容易操作的代理服务器，也会遇到这样的问题，只不过是结果更温和一些。

除了下载及浏览请求外，加密的邮件也能不通过审查进行传送。Web界面邮件系统的用户能够通过将通常使用的“http”前缀更改为“https”来建立加密通道。例如，使用如下方式来实现邮件的安全传送：

https://mail.yahoo.com,

https://mail.google.com, e.com/。

为了有一个有利的国际环境，GFW必须在采取的措施中允许例外——即使知道许多网民会借机“透气”。

参考文献

[1]WWW 的信息监控研究　通讯和计算机 Journal of Communication and Compme~ISSN1 548-7709，USA 曹天杰 ，林柬岱 ，薛锐

[2] 防火墙技术与网络安全　徐向文

[3] The Great Firewall of China Charles R. Smith Friday, May 17, 2002 http://archive.newsmax.com/archives/articles/2002/5/17/25858.shtml

[4] Asia Pacific Root servers http://www.apnic.net/services/rootserver/index.html

[5] 维基百科 http://zh.wikipedia.org/w/index.php?title=%E9%98%B2%E7%81%AB%E9%95%BF%E5%9F%8E&variant=zh-cn

[6] 突破网络审查 http://zh.wikipedia.org/wiki/%E7%AA%81%E7%A0%B4%E7%BD%91%E7%BB%9C%E5%AE%A1%E6%9F%A5

[7] ConceptDoppler: A Weather Tracker for Internet Censorship Jedidiah R. Crandall，Daniel Zinn，Michael Byrd 14th ACM Conference on Computer and Communications Security, Oct. 29-Nov. 2, 2007

[8] 一种基于透明Web Cache的内容过滤实现框架计算机应用　第24卷第6期肖宗水，许艳美等

[9]王 宇，施燕姝．卢 昱．利用代理服务器实现透明的安全隧道 / 香港国际计算机会议论文集．1999

[10]施威铭研究室lnternet协议概念与实践 北京：清华大学出版社，2001．

[11]Bragg R Mark R O．网络安全完全手册 北京：电子工业出版社，2005．10．

[12] Ian Clarke，Oskar Sandberg Brandon Wiley， theodore WI-long．Freenet：A Distributed Anonymous Information Storage and Retrieval System． Designing Privacy Enhancing rechnologies： lntemational Workshop on Design Issues in Anonymity an d Unobservability,LNCS 2OO9．Spdnger-Verlag Berlin Heidelberg．PP．46-66．

[13] P2P网络中对等节点间安全通信研究叶润国 宋成等 2004年第21卷第6期 微电子学与计算机

[14]胡健伟，汤建龙，杨绍全．网络对抗原理 西安：西安电子科技术大学出版社，2004．06．

[15]VPN技术的研究与实现冯伟　冯登国等计算机工程与设计　2002.2

任务分工

凹建勋:引言、GFW技术、无界浏览器原理、对破网软件的控制、可能的破网方法、结语

戴魏巍:SSH穿越GFW

贾斌:Tor原理

李理:自由门原理、GFW技术

- 本页面最后修订：2008年6月22日,14:22.