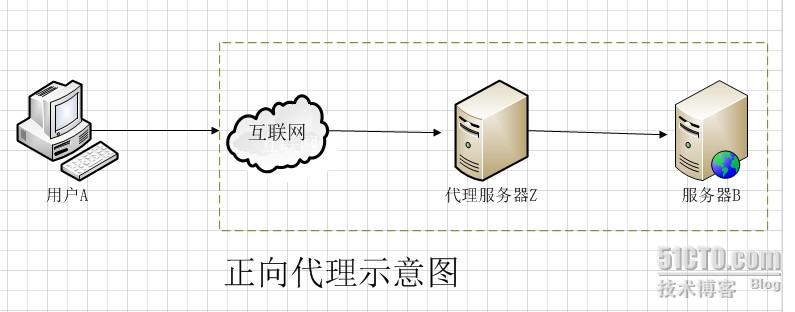
**Proxy**

一、正向代理(Forward Proxy)

一般情况下，如果没有特别说明，代理技术默认说的是正向代理技术。关于正向代理的概念如下：

正向代理(forward)是一个位于客户端【用户A】和原始服务器(origin server)【服务器B】之间的服务器【代理服务器Z】，为了从原始服务器取得内容，用户A向代理服务器Z发送一个请求并指定目标(服务器B)，然后代理服务器Z向服务器B转交请求并将获得的内容返回给客户端。客户端必须要进行一些特别的设置才能使用正向代理。如下图1.1



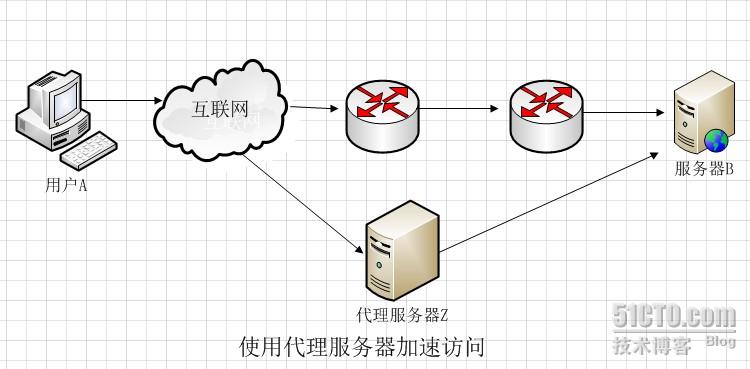
（图1.1）

从上面的概念中，我们看出，文中所谓的正向代理就是代理服务器替代访问方【用户A】去访问目标服务器【服务器B】

这就是正向代理的意义所在。而为什么要用代理服务器去代替访问方【用户A】去访问服务器B呢？这就要从代理服务器使用的意义说起。

使用正向代理服务器作用主要有以下几点：

1、访问本无法访问的服务器B，如下图1.2



（图1.2）

我们抛除复杂的网络路由情节来看图1.2，假设图中路由器从左到右命名为R1,R2

假设最初用户A要访问服务器B需要经过R1和R2路由器这样一个路由节点，如果路由器R1或者路由器R2发生故障，那么就无法访问服务器B了。但是如果用户A让代理服务器Z去代替自己访问服务器B，由于代理服务器Z没有在路由器R1或R2节点中，而是通过其它的路由节点访问服务器B，那么用户A就可以得到服务器B的数据了。

现实中的例子就是“翻墙”。不过自从VPN技术被广泛应用外，“翻墙”不但使用了传统的正向代理技术，有的还使用了VPN技术。

2、加速访问服务器B

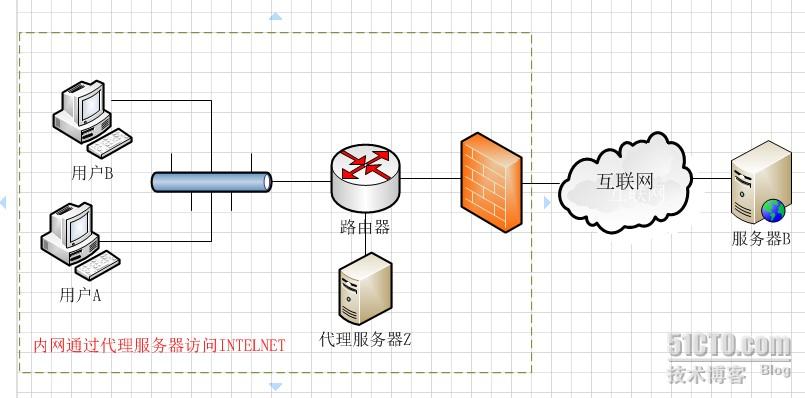
这种说法目前不像以前那么流行了，主要是带宽流量的飞速发展。早期的正向代理中，很多人使用正向代理就是提速。还是如图1.2，假设用户A到服务器B，经过R1路由器和R2路由器，而R1到R2路由器的链路是一个低带宽链路。而用户A到代理服务器Z，从代理服务器Z到服务器B都是高带宽链路。那么很显然就可以加速访问服务器B了。

3、Cache作用

Cache（缓存）技术和代理服务技术是紧密联系的（不光是正向代理，反向代理也使用了Cache（缓存）技术。还如上图所示，如果在用户A访问服务器B某数据J之前，已经有人通过代理服务器Z访问过服务器B上得数据J，那么代理服务器Z会把数据J保存一段时间，如果有人正好取该数据J，那么代理服务器Z不再访问服务器B，而把缓存的数据J直接发给用户A。这一技术在Cache中术语就叫Cache命中。如果有更多的像用户A的用户来访问代理服务器Z，那么这些用户都可以直接从代理服务器Z中取得数据J，而不用千里迢迢的去服务器B下载数据了。

4、客户端访问授权

这方面的内容现今使用的还是比较多的，例如一些公司采用ISA SERVER做为正向代理服务器来授权用户是否有权限访问互联网，挼下图1.3

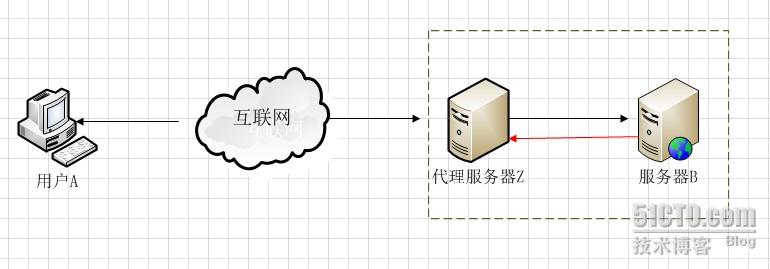


（图1.3）

图1.3防火墙作为网关，用来过滤外网对其的访问。假设用户A和用户B都设置了代理服务器，用户A允许访问互联网，而用户B不允许访问互联网（这个在代理服务器Z上做限制）这样用户A因为授权，可以通过代理服务器访问到服务器B，而用户B因为没有被代理服务器Z授权，所以访问服务器B时，数据包会被直接丢弃。

5、隐藏访问者的行踪

如下图1.4 我们可以看出服务器B并不知道访问自己的实际是用户A，因为代理服务器Z代替用户A去直接与服务器B进行交互。如果代理服务器Z被用户A完全控制（或不完全控制），会惯以“肉鸡”术语称呼。



（图1.4）

我们总结一下 正向代理是一个位于客户端和原始服务器(origin server)之间的服务器，为了从原始服务器取得内容，客户端向代理发送一个请求并指定目标(原始服务器)，然后代理向原始服务器转交请求并将获得的内容返回给客户端。客户端必须设置正向代理服务器，当然前提是要知道正向代理服务器的IP地址，还有代理程序的端口。

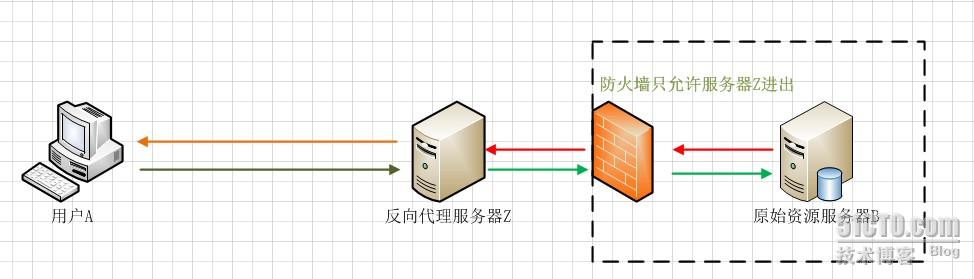
二、反向代理（reverse proxy）

反向代理正好与正向代理相反，对于客户端而言代理服务器就像是原始服务器，并且客户端不需要进行任何特别的设置。客户端向反向代理的命名空间(name-space)中的内容发送普通请求，接着反向代理将判断向何处(原始服务器)转交请求，并将获得的内容返回给客户端。

使用反向代理服务器的作用如下：

1、  保护和隐藏原始资源服务器

如下图2.1

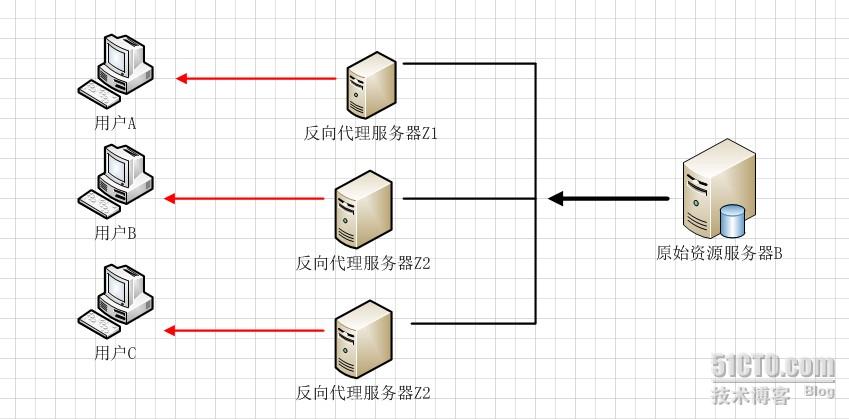


（图2.1）

用户A始终认为它访问的是原始服务器B而不是代理服务器Z，但实用际上反向代理服务器接受用户A的应答，从原始资源服务器B中取得用户A的需求资源，然后发送给用户A。由于防火墙的作用，只允许代理服务器Z访问原始资源服务器B。尽管在这个虚拟的环境下，防火墙和反向代理的共同作用保护了原始资源服务器B，但用户A并不知情。

2、  负载均衡

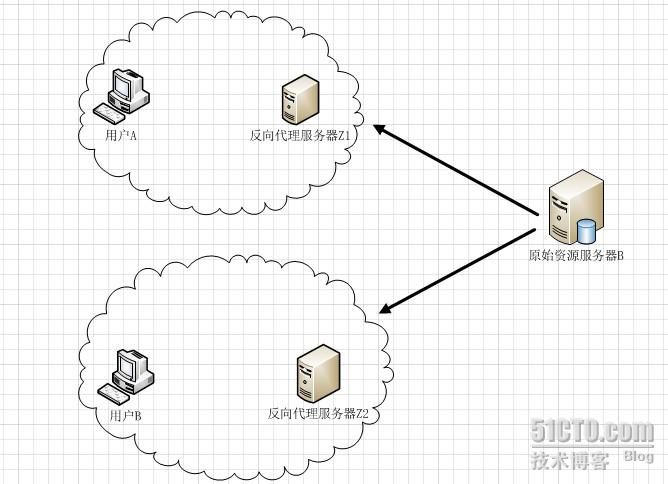
如下图2.2



（图2.2）

   当反向代理服务器不止一个的时候，我们甚至可以把它们做成集群，当更多的用户访问资源服务器B的时候，让不同的代理服务器Z（x）去应答不同的用户，然后发送不同用户需要的资源。

当然反向代理服务器像正向代理服务器一样拥有CACHE的作用，它可以缓存原始资源服务器B的资源，而不是每次都要向原始资源服务器B请求数据，特别是一些静态的数据，比如图片和文件，如果这些反向代理服务器能够做到和用户X来自同一个网络，那么用户X访问反向代理服务器X，就会得到很高质量的速度。这正是CDN技术的核心。如下图2.3



（图2.3）

我们并不是讲解CDN，所以去掉了CDN最关键的核心技术智能DNS。只是展示CDN技术实际上利用的正是反向代理原理这块。

反向代理结论与正向代理正好相反，对于客户端而言它就像是原始服务器，并且客户端不需要进行任何特别的设置。客户端向反向代理的命名空间(name-space)中的内容发送普通请求，接着反向代理将判断向何处(原始服务器)转交请求，并将获得的内容返回给客户端，就像这些内容原本就是它自己的一样。

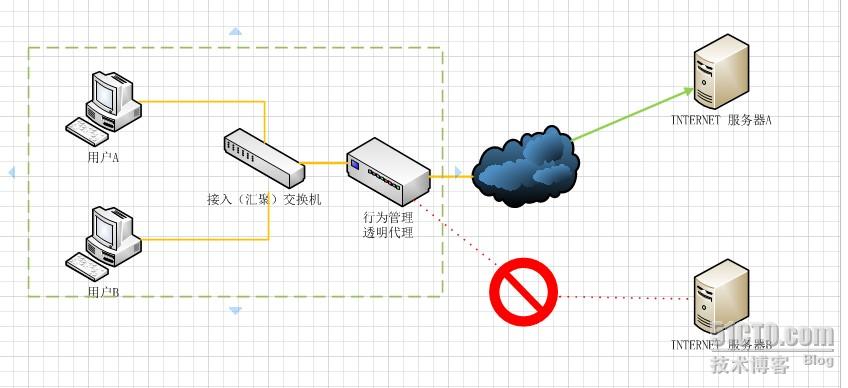
基本上，网上做正反向代理的程序很多，能做正向代理的软件大部分也可以做反向代理。开源软件中最流行的就是squid，既可以做正向代理，也有很多人用来做反向代理的前端服务器。另外MS ISA也可以用来在WINDOWS平台下做正向代理。反向代理中最主要的实践就是WEB服务，近些年来最火的就是Nginx了。网上有人说NGINX不能做正向代理，其实是不对的。NGINX也可以做正向代理，不过用的人比较少了。

三、透明代理

    如果把正向代理、反向代理和透明代理按照人类血缘关系来划分的话。那么正向代理和透明代理是很明显堂亲关系，而正向代理和反向代理就是表亲关系了 。

透明代理的意思是客户端根本不需要知道有代理服务器的存在，它改编你的request fields（报文），并会传送真实IP。注意，加密的透明代理则是属于匿名代理，意思是不用设置使用代理了。

透明代理实践的例子就是时下很多公司使用的行为管理软件。如下图3.1

（图3.1）

  用户A和用户B并不知道行为管理设备充当透明代理行为，当用户A或用户B向服务器A或服务器B提交请求的时候，透明代理设备根据自身策略拦截并修改用户A或B的报文，并作为实际的请求方，向服务器A或B发送请求，当接收信息回传，透明代理再根据自身的设置把允许的报文发回至用户A或B，如上图，如果透明代理设置不允许访问服务器B，那么用户A或者用户B就不会得到服务器B的数据。

1. 代理服务器

代理服务器英文全称是Proxy Server，其功能就是代理网络用户去取得网络信息。形象的说：它是网络信息的中转站。在一般情况下，我们使用网络浏览器直接去连接其他Internet站点取得网络信息时，须送出Request信号来得到回答，然后对方再把信息以bit方式传送回来。代理服务器是介于浏览器和Web服务器之间的一台服务器，有了它之后，浏览器不是直接到Web服务器去取回网页而是向代理服务器发出请求，Request信号会先送到代理服务器，由代理服务器来取回浏览器所需要的信息并传送给你的浏览器。而且，大部分代理服务器都具有缓冲的功能，就好象一个大的Cache，它有很大的存储空间，它不断将新取得数据储存到它本机的存储器上，如果浏览器所请求的数据在它本机的存储器上已经存在而且是最新的，那么它就不重新从Web服务器取数据，而直接将存储器上的数据传送给用户的浏览器，这样就能显着提高浏览速度和效率。更重要的是：Proxy Server(代理服务器)是Internet链路级网关所提供的一种重要的安全功能，它的工作主要在开放系统互联(OSI)模型的对话层。主要的功能有：

(1).突破自身IP访问限制，访问国外站点。教育网、169网等网络用户可以通过代理访问国外网站。

(2).访问一些单位或团体内部资源，如某大学FTP(前提是该代理地址在该资源 的允许访问范围之内)，使用教育网内地址段免费代理服务器，就可以用于对教育 网开放的各类FTP下载上传，以及各类资料查询共享等服务。

(3).突破中国电信的IP封锁：中国电信用户有很多网站是被限制访问的，这种 限制是人为的，不同Serve对地址的封锁是不同的。所以不能访问时可以换一个国 外的代理服务器试试。

(4).提高访问速度：通常代理服务器都设置一个较大的硬盘缓冲区，当有外界 的信息通过时，同时也将其保存到缓冲区中，当其他用户再访问相同的信息时， 则直接由缓冲区中取出信息，传给用户，以提高访问速度。

(5).隐藏真实IP：上网者也可以通过这种方法隐藏自己的IP，免受攻击。

使用代理服务器应

注意事项

代理服务器除了网络服务商为了各种目的而开设外，大部分是新建网络服务器设置的疏漏！虽然法律尚无具体规定，但没有经过允许而使用他人的服务器当然还是不太好！虽然目的主机一般只能得到您使用的代理服务器IP，似乎有效的遮掩了你的行程，但是值得一提的是：网络服务商开通的专业级代理服务器一般都有路由和流程记录，因此可以轻易的通过调用历史纪录来查清使用代理服务器地址的来路。当然，利用多层代理会增加被捕获的难度，但也不是不可能的。去年报上就有报道有人使用代理服务器进攻“天府热线”，进行非法活动而被抓的消息。因此，建议大家不要利用代理服务器来进行特别行动！只要你不使用代理进行非法活动，一般是没有关系的。

1. CDN(Content Delivery Network)

CDN的基本原理为反向代理，反向代理（Reverse Proxy）方式是指以代理服务器来接受internet上的连接请求，然后将请求转发给内部网络上的服务器，并将从服务器上得到的结果返回给internet上请求连接的客户端，此时代理服务器对外就表现为一个节点服务器。通过部署更多的反向代理服务器，来达到实现多节点CDN的效果。

传统的未加缓存服务的访问过程:

用户提交域名→浏览器对域名进行解析→得到目的主机的IP地址→根据IP地址访问发出请求→得到请求数据并回复

由上可见，用户访问未使用CDN缓存网站的过程为:

1)、用户向浏览器提供要访问的域名；

2)、浏览器调用域名解析函数库对域名进行解析，以得到此域名对应的IP地址；

3)、浏览器使用所得到的IP地址，向域名的服务主机发出数据访问请求；

4)、浏览器根据域名主机返回的数据显示网页的内容。

通过以上四个步骤，浏览器完成从用户处接收用户要访问的域名到从域名服务主机处获取数据的整个过程。CDN网络是在用户和服务器之间增加Cache层，如何将用户的请求引导到Cache上获得源服务器的数据，主要是通过接管DNS实现，下面让我们看看访问使用CDN缓存后的网站的过程：

使用了CDN缓存后的网站的访问过程变为：

1)、用户向浏览器提供要访问的域名；

2)、浏览器调用域名解析库对域名进行解析，由于CDN对域名解析过程进行了调整，所以解析函数库一般得到的是该域名对应的CNAME记录，为了得到实际IP地址，浏览器需要再次对获得的CNAME域名进行解析以得到实际的IP地址；在此过程中，使用的全局负载均衡DNS解析，如根据地理位置信息解析对应的IP地址，使得用户能就近访问。

3)、此次解析得到CDN缓存服务器的IP地址，浏览器在得到实际的IP地址以后，向缓存服务器发出访问请求；

4)、缓存服务器根据浏览器提供的要访问的域名，通过Cache内部专用DNS解析得到此域名的实际IP地址，再由缓存服务器向此实际IP地址提交访问请求；

5)、缓存服务器从实际IP地址得得到内容以后，一方面在本地进行保存，以备以后使用，另一方面把获取的数据返回给客户端，完成数据服务过程；

6)、客户端得到由缓存服务器返回的数据以后显示出来并完成整个浏览的数据请求过程。

通过以上的分析我们可以得到，为了实现既要对普通用户透明(即加入缓存以后用户客户端无需进行任何设置，直接使用被加速网站原有的域名即可访问，又要在为指定的网站提供加速服务的同时降低对ICP的影响，只要修改整个访问过程中的域名解析部分，以实现透明的加速服务，下面是CDN网络实现的具体操作过程。

1)、作为ICP，只需要把域名解释权交给CDN运营商，其他方面不需要进行任何的修改；操作时，ICP修改自己域名的解析记录，一般用cname方式指向CDN网络Cache服务器的地址。

2)、作为CDN运营商，首先需要为ICP的域名提供公开的解析，为了实现sortlist，一般是把ICP的域名解释结果指向一个CNAME记录；

3)、当需要进行sortlist时，CDN运营商可以利用DNS对CNAME指向的域名解析过程进行特殊处理，使DNS服务器在接收到客户端请求时可以根据客户端的IP地址，返回相同域名的不同IP地址；

4)、由于从cname获得的IP地址，并且带有hostname信息，请求到达Cache之后，Cache必须知道源服务器的IP地址，所以在CDN运营商内部维护一个内部DNS服务器，用于解释用户所访问的域名的真实IP地址；

5)、在维护内部DNS服务器时，还需要维护一台授权服务器，控制哪些域名可以进行缓存，而哪些又不进行缓存，以免发生开放代理的情况。

ICP是指各地通信管理部门核发的《中华人民共和国电信与信息服务业务经营许可证》

1. 智能DNS

普通的DNS服务器只负责为用户解析出IP记录，而不去判断用户从哪里来，这样会造成所有用户都只能解析到固定的IP地址上。智能DNS颠复了这个概念。智能DNS会判断用户的来路，而做出一些智能化的处理，然后把智能化判断后的IP返回给用户。比如，DNSPod的智能DNS就会自动判断用户的上网线路是网通还是电信，然后智能返回网通或者电信服务器的IP。