LVS-NAT模式 网络地址转换协议

LVS集群 实现三种IP负载均衡技术
NAT:网络地址转换协议。 virtual server via network address translation
TUN:隧道模式。 virtual server via ip tunneling
DR :直接路由模式。 virtual server via direct routing

1、基于应用层的负载均衡调度。 典型调度器 Zeus，pWeb SWEB Reverse-Poroxy
调度器 分析请求，根据每个服务器的负载情况，选出一台服务器，
重写请求并向选出的服务器访问，取得结果返回给用户。
存在的问题和解决方法：1.系统处理开销特别大，致使系统的伸缩性有限。
2.调度器对不同应用的兼容性不高。
2、F5使用的是 NAT 网络地址转换协议

NAT 只能在局域网内进行。 内部为局域网 基本工作在第三层网络层
通过NAT实现虚拟服务器。
NAT协议可将内部地址转化为Internets上可用的外部地址。

集群节点跟Direcator必须在同一个IP网络中。
RIP(真实IP)通常是私有地址，仅仅用于个集群之间通信。
director位于Client和real server之间，并负责处理进出的所有通信。
realservet必须将网关指向DIP。同时也支持端口影射
Realserver可以使用任意OS
在较大应用规模当中，单个Director会出现瓶颈
大概可以带10个左右的SERVER就会出现瓶颈
服务器和调度器之间通过 switch/hub链接

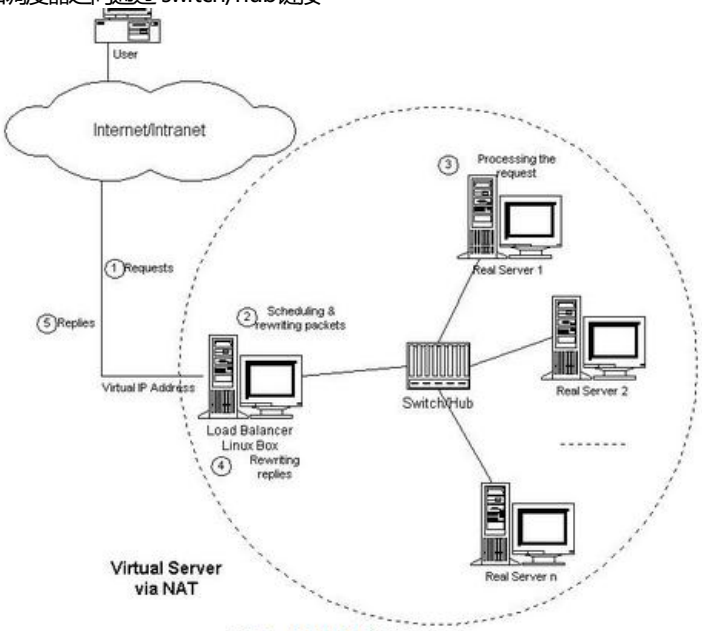
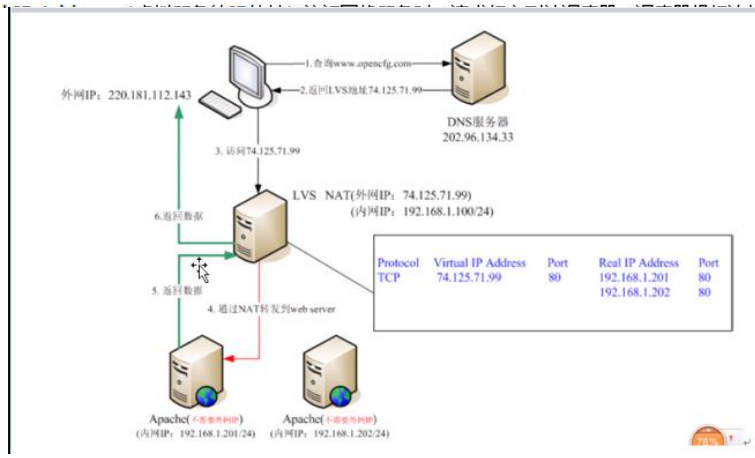


图2：VS/NAT的体系结构



。

VS/NAT 的配置如下表所示，所有到IP地址为202.103.106.5和端口为80的流量都被负载均衡地调度的真实服务器172.16.0.2:80和 172.16.0.3:8000上。目标地址为202.103.106.5:21的报文被转移到172.16.0.3:21上。而到其他端口的报文将被拒 绝。

Protocol	Virtual IP Address	Port	Real IP Address	Port	Weight
TCP	202.103.106.5	80	172.16.0.2	80	1
			172.16.0.3	8000	2
TCP	202.103.106.5	21	172.16.0.3	21	1

从以下的例子中，我们可以更详细地了解报文改写的流程。

访问Web服务的报文可能有以下的源地址和目标地址：

SOURCE	202.100.1.2:3456	DEST	202.103.106.5:80

调度器从调度列表中选出一台服务器，例如是172.16.0.3:8000。该报文会被改写为如下地址，并将它发送给选出的服务器。

SOURCE	202.100.1.2:3456	DEST	172.16.0.3:8000

从服务器返回到调度器的响应报文如下：

SOURCE	172.16.0.3:8000	DEST	202.100.1.2:3456

响应报文的源地址会被改写为虚拟服务的地址，再将报文发送给客户：

SOURCE	202.103.106.5:80	DEST	202.100.1.2:3456

这样，客户认为是从202.103.106.5:80服务得到正确的响应，而不会知道该请求是服务器172.16.0.2还是服务器172.16.0.3处理的。

图2：VS/NAT的体系结构

客 户通过Virtual IP Address（虚拟服务的IP地址）访问网络服务时，请求报文到达调度器，调度器根据连接调度算法从一组真实服务器中选出一台服务器，将报文的目标地址 Virtual IP Address改写成选定服务器的地址，报文的目标端口改写成选定服务器的相应端口，最后将修改后的报文发送给选出的服务器。同时，调度器在连接Hash 表中记录这个连接，当这个连接的下一个报文到达时，从连接Hash表中可以得到原选定服务器的地址和端口，进行同样的改写操作，并将报文传给原选定的服务 器。当来自真实服务器的响应报文经过调度器时，调度器将报文的源地址和源端口改为Virtual IP Address和相应的端口，再把报文发给用户。我们在连接上引入一个状态机，不同的报文会使得连接处于不同的状态，不同的状态有不同的超时值。在TCP 连接中，根据标准的TCP有限状态机进行状态迁移，这里我们不一一叙述，请参见W. Richard Stevens的《TCP/IP Illustrated Volume I》；在UDP中，我们只设置一个UDP状态。不同状态的超时值是可以设置的，在缺省情况下，SYN状态的超时为1分钟，ESTABLISHED状态的超 时为15分钟，FIN状态的超时为1分钟；UDP状态的超时为5分钟。当连接终止或超时，调度器将这个连接从连接Hash表中删除。

这样，客户所看到的只是在Virtual IP Address上提供的服务，而服务器集群的结构对用户是透明的。对改写后的报文，应用增里调整Checksum的算法调整TCP Checksum的值，避免了扫描整个报文来计算Checksum的开销。

在 一些网络服务中，它们将IP地址或者端口号在报文的数据中传送，若我们只对报文头的IP地址和端口号作转换，这样就会出现不一致性，服务会中断。所以，针 对这些服务，需要编写相应的应用模块来转换报文数据中的IP地址或者端口号。我们所知道有这个问题的网络服务有FTP、IRC、H.323、 CUSeeMe、Real Audio、Real Video、Vxtreme / Vosiac、VDOLive、VIVOActive、True Speech、RSTP、PPTP、StreamWorks、NTT AudioLink、NTT SoftwareVision、Yamaha MIDPlug、iChat Pager、Quake和Diablo。