基于 VRRP 的 LVS 高可用性研究

吴国才，施 婧，郑勇明
（厦门大学 计算机科学系 福建 厦门 361005）

【摘 要】：本文介绍了 VRRP(虚拟路由器冗余协议)，并描述一种基于 VRRP 的解决方案。该方案可用于 LVS 解决其单点故障问题，提高 LVS 的高可用性。

【关键字】：LVS；高可用性；VRRP

1.LVS 介绍

LVS[2](Linux Virtual Server)是一个开源项目，成立于 1998 年，它提供了一套用于构建 Linux 集群的解决方案。LVS 通过在 Linux 内核实现基于 IP 层和基于内容请求分发的负载平衡调度解决方法，将一组服务器构成一个实现可伸缩的、高可用网络服务的虚拟服务器。

LVS 的体系结构如图 1 所示，一组服务器通过高速的局域网或者地理分布的广域网相互连接，在它们的前端有一个负载均衡器(Load Balancer)。负载调度器能无缝地将网络请求调度到真实服务器上，从而使得服务器集群的结构对客户是透明的，客户访问集群系统提供的网络服务就像访问一台高性能、高可用的服务器一样。客户程序不受服务器集群的影响不需作任何修改。系统的伸缩性通过在服务机群中透明地加入和删除一个节点来达到，通过检测节点或服务进程故障和正确地重置系统达到高可用性[3]。

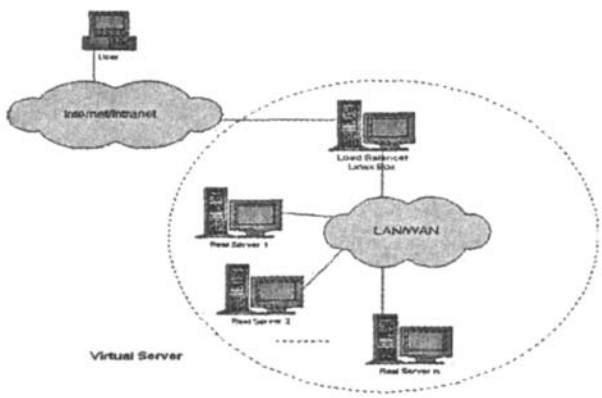

图 1 虚拟服务器结构

由于 LVS 所有的服务都要通过前端负载均衡器传递，因此当 LVS 负载过重时，可能出现由于负载均衡器的故障而导致整个集群系统瘫痪的单点失效。以下将描述一种基于 VRRP 的方案解决 LVS 的单点失效。

2. VRRP 简介

VRRP[3]是一种 LAN 接入设备容错协议，它将局域网的一组路由器组织成一个虚拟路由器，指定一种选举协议，动态的为局域网中每个 VRRP 路由器分配虚拟路由器的任务，其中与虚拟路由器联合控制 IP 地址的 VRRP 路由器被称做主控路由器(Master)，并负责转发目的地址为这些 IP 地址的包。这样局域网中任何一个虚拟路由的 IP 地址都可以被终端作为默认路由。如果主控路由器不可用，则选举过程提供了转发任务的动态恢复。

3.一种基于 VRRP 的 LVS 方案

利用 VRRP 的特性，设置两个负载均衡器，在其中一个负载均衡器出现故障时，由另一个负载均衡器接管其任务。基于 VRRP 的 LVS 逻辑结构如(如图 2 所示)：

3.1 VRRP 术语

该系统使用的一些术语如下：

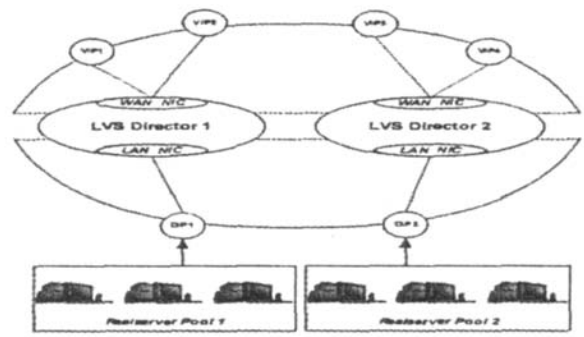

图 2 基于 VRRP 的 LVS 结构

· VRRP 实例：一个实例，控制操纵指定的一组 IP 地址。一个 VRRP 实例可以成为一个或者多个 VRRP 实例的备分。在图 2 中，有 4 个 VRRP 实例，实例 VI_1 拥有地址(VIP1,VIP2)，实例 VI_3 拥有地址(VIP3,VIP4)，实例 VI_2 拥有地址(DIP1)，实例 VI_4 拥有地址(DIP2)，它们可以参与一个或者多个虚拟路由。

· 主控状态：是 VRRP 实例的状态，处于该状态的实例负责转发目的地址是该 VRRP 实例拥有地址的包。

· 备份状态：VRRP 实例的状态，当 VRRP 主控状态失效时，接替主控实例负责转发数据包的任务。

· 实际负载均衡器：一个 LVS 负载均衡器，其上运行一个或多个 VRRP 实例。

· 虚拟负载均衡器：一组实际负载均衡器组成。

· 同步实例：需要同步的 VRRP 实例。

· 广告：一种简单的 VRRP 包，主控实例发送给一组 VRRP 实例声明状态。

3.2 系统工作流程

在图 2 的结构中，服务器池 2 的主机是属于虚拟 IP 地址 VIP3 和 VIP4 的，因此远程客户访问服务器池 2 的主机必须通过 VIP3 或 VIP4，但另一方面服务器池 2 的主机转发数据包是通过 DIP2 的，这样 DIP2 和 VIP3/VIP4 在同一个负载均衡器上必须保持同时激活的状态，才能维持完整的转发路径。这就意味着，如果负载均衡器 2 的局域网(LAN)接口失效，广域网(WAN)接口拥有的 IP 地址必须设置在冗余的负载均衡器 1 上的广域网接口上。这样每个 LVS 负载均衡器必须知道整个 LVS 的冗余结构，所以每个 LVS 负载均衡器上同时运行四个 VRRP 实例，两个 VRRP 实例处于主控状态，两个处于备份状态，所以在每个负载均衡器上 VRRP 实例配置时主控/备份状态是对称的，在负载均衡器 1 上，VRRP 实例的配置情况如下：

· VI_1：处于主控状态，拥有 IP 地址 VIP1 和 VIP2，并保持与 VI_2 同步。

· VI_2: 处于主控状态，拥有 IP 地址 DIP，并保持与 VI_1 同步。

· VI_3: 处于备份状态，备份 IP 地址 VIP3 和 VIP4，并保持与 VI_4 同步。

· VI_4: 处于备份状态，备份 IP 地址 DIP，并保持与 VI_3 同步。

在负载均衡器 2 上，VI_1 和 VI_2 处于备份状态，而 VI_3 和 VI_4 处于主控状态。

由于某种原因 DIP2 不可用，为了保持路由路径，则 VRRP 实例状态转换情况如图 3:

下图中，因为 DIP2 不可用，那么运行于负载均衡器 1 上处于备份状态的 VI_4，将在设定广告时间间隔内收不到主控 VI_4 的状态广告，因此备份 VI_4 推断 DIP2 不可用，所以这个备份状态的 VI_4 就切换到主控状态。

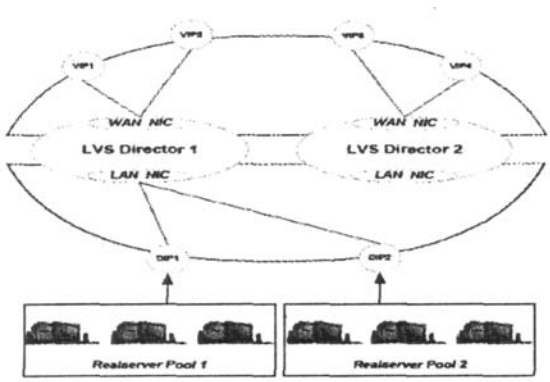

图 3　DIP2 任务接管情况

在同一负载均衡器上，实例 VI_3 和 VI_4 必须保持同步状态，否则通过 VIP3 和 VIP4 请求的客户无法得到 DIP2 的响应，这样在负载均衡器 1 的负责 VIP3 和 VIP4 的实例 VI_3 必须切换到主控状态，保持与运行在该负载均衡器上实例 VI_4 同步，实例 VI_3 的状态转换如下图 4:

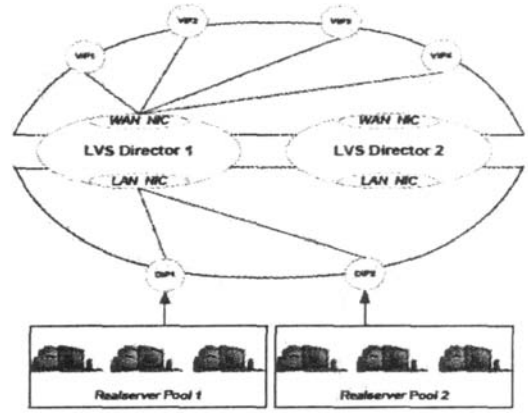

图 4 VRRP VIP3&VIP4 同步

最终 VIP3 和 VIP4 被负载均衡器 1 所接管，保持了与 DIP2 的同步，这就是在负载均衡器 2 局域网接口失效后应该达到的最终状态。此时所有的转发任务全部由负载均衡器 1 来完成。这种状态只是一种暂时的接管状态，当负载均衡器 2 的局域网接口恢复正常，所有的 VRRP 实例将切换到最初的状态。

4. 小结

本文描述了一种使用 VRRP 协议解决 LVS 的负载均衡器器单点故障的方案。该方案设立两个负载均衡器，在主负载均衡器出现故障时，由备份负载均衡器接管主负载均衡器的所有任务，提高 LVS 的高可用性。

参考文献:

1.章文嵩，吴婷婷，金士尧，吴泉源. 一个虚拟 Internet 服务器的设计与实现. 软件学报，2001，11(1): 122- 125

2.Alexandre Cassen. Linux Virtual High Ability using VRRPv2. http://www.LinuxVirtualSever.org,2001

3.白欣，宋博，左继章，向建军. 高可用性冗余实时集群系统的设计与构建. 计算机工程，2004，1

(上接第 142 页)

及一个 14 位的 EAN/UCC 代码。例如选择 SGTIN 类型，代码为 18801234567896。根据 SGTIN 的代码规则，通过一个代码转换模块，将代码转换成为一个 EPC URN: urn:epc:48.8801234.156789.400。其中，48 表示 EPC 的类型为 SGTIN，8001234 是公司代号，156789 是产品代号，400 是一个序列号。

③: 将得到的 URN 去除头部和 EPC 的序列号，再将剩下的 8801234.156789 的位置进行交换，得到 156789.8801234。最后加上一个 TLD，形成一个 FQDN (全域名): 156789.8801234.ods.or.kr。然后为这个 FQDN 向 ONS 服务器发送一个服务请求，要求返回一个或多个 NAPTR 格式的记录。

④: ONS 服务器接收到请求后，会根据接收到的 FQDN 找出与之相关的记录，并且以 NAPTR 的格式返回。在这里，order 和 pref 并未被使用，flag 部分为"u"，表示 regexp 部分是一个 URL，service 部分规定了使用何种类型文件描述产品信息 (XML 等)，regexp 部分指定了产品信息存放位置的 URI。

⑤: 通过解析接收到的 NAPTR 记录，可以从中得到产品相关信息存放的 URL 地址，并将这些 URL 地址发送到 ISP (Information Service Provider)模块中。

⑥,⑦,⑧: ISP 向 EPC- IS 服务器发出访问请求，根据得到的 URL，查看相关的信息。返回的信息以 Web 页的形式显示。

4.结束语

本文通过一个 EPC 网络原型的设计与实现，详细说明了 EPC 网络各组成部分的实现架构，并基于原型实现了一个 Web 应用程序，对了解 EPC 网络的工作原理，进一步的开发 EPC 网络的应用有着一定的参考意义。

参考文献:

1."ONS Registry Overview and Implementation Guide", Version 2.0, VeriSign Inc. August 2003

2.EPCGlobal, "EPCTM Tag Data Standards Version 1.1 Rev.1.23", Last Call Working Draft on 16 February,2004

3.Global Trade Item Numbers(GTIN) Application, Uniform Code Council, Inc

4.SCC (Serial Shipping Container Code) Implementation, Uniform Code Council, Inc

5.SGLN (Serialized Global Location Number) Implementation, Uniform Code Council, Inc

6.GRAI(Global Returnable Asset Identifier) Implementation, Uniform Code Council, Inc

7.GIAI (Global Individual Asset Identifier) Implementation, Uniform Code Council, Inc