Red Hat Enterprise Linux 5

Virtual Server Administration

Linux Virtual Server (LVS) for Red Hat Enterprise Linux

Red Hat Enterprise Linux 5 Virtual Server Administration
Linux Virtual Server (LVS) for Red Hat Enterprise Linux
版 5

Copyright © 2009 Red Hat, Inc.

The text of and illustrations in this document are licensed by Red Hat under a Creative Commons Attribution—Share Alike 3.0 Unported license ("CC-BY-SA"). An explanation of CC-BY-SA is available at http://creativecommons.org/licenses/by-sa/3.0/. In accordance with CC-BY-SA, if you distribute this document or an adaptation of it, you must provide the URL for the original version.

Red Hat, as the licensor of this document, waives the right to enforce, and agrees not to assert, Section 4d of CC-BY-SA to the fullest extent permitted by applicable law.

Red Hat, Red Hat Enterprise Linux, the Shadowman logo, JBoss, MetaMatrix, Fedora, the Infinity Logo, and RHCE are trademarks of Red Hat, Inc., registered in the United States and other countries.

Linux® is the registered trademark of Linus Torvalds in the United States and other countries.

Java® is a registered trademark of Oracle and/or its affiliates.

XFS® is a trademark of Silicon Graphics International Corp. or its subsidiaries in the United States and/or other countries.

MySQL® is a registered trademark of MySQL AB in the United States, the European Union and other countries.

All other trademarks are the property of their respective owners.

1801 Varsity Drive
Raleigh, NC 27606-2072 USA
Phone: +1 919 754 3700
Phone: 888 733 4281
Fax: +1 919 754 3701

Building a Linux Virtual Server (LVS) system offers highly-available and scalable solution for production services using specialized routing and load-balancing techniques configured through the PIRANHA. This book discusses the configuration of high-performance systems and services with Red Hat Enterprise Linux and LVS for Red Hat Enterprise Linux 5.

Introduction

This document provides information about installing, configuring, and managing Red Hat Virtual Linux Server (LVS) components. LVS provides load balancing through specialized routing techniques that dispatch traffic to a pool of servers. This document does not include information about installing, configuring, and managing Red Hat Cluster software. Information about that is in a separate document.

The audience of this document should have advanced working knowledge of Red Hat Enterprise Linux and understand the concepts of clusters, storage, and server computing.

This document is organized as follows:

For more information about Red Hat Enterprise Linux 5, refer to the following resources:

- Red Hat Enterprise Linux Installation Guide — Provides information regarding installation of Red Hat Enterprise Linux 5.
- Red Hat Enterprise Linux Deployment Guide — Provides information regarding the deployment, configuration and administration of Red Hat Enterprise Linux 5.

For more information about Red Hat Cluster Suite for Red Hat Enterprise Linux 5, refer to the following resources:

- Red Hat Cluster Suite Overview — Provides a high level overview of the Red Hat Cluster Suite.
- Configuring and Managing a Red Hat Cluster — Provides information about installing, configuring and managing Red Hat Cluster components.
- Logical Volume Manager Administration — Provides a description of the Logical Volume Manager (LVM), including information on running LVM in a clustered environment.
- Global File System: Configuration and Administration — Provides information about installing, configuring, and maintaining Red Hat GFS (Red Hat Global File System).
- Global File System 2: Configuration and Administration — Provides information about installing, configuring, and maintaining Red Hat GFS2 (Red Hat Global File System 2).
- Using Device-Mapper Multipath — Provides information about using the Device-Mapper Multipath feature of Red Hat Enterprise Linux 5.
- Using GNBD with Global File System — Provides an overview on using Global Network Block Device (GNBD) with Red Hat GFS.
- Red Hat Cluster Suite Release Notes — Provides information about the current release of Red Hat Cluster Suite.

Red Hat Cluster Suite documentation and other Red Hat documents are available in HTML, PDF, and RPM versions on the Red Hat Enterprise Linux Documentation CD and online at http://www.redhat.com/docs/.

1. 文档约定

本手册使用几个约定来突出某些用词和短语以及信息的某些片段。

在 PDF 版本以及纸版中，本手册使用在 Liberation 字体[1]套件中选出的字体。如果您在您的系统中安装了 Liberation 字体套件，它还可用于 HTML 版本。如果没有安装，则会显示可替换的类似字体。请注意：红帽企业 Linux 5 以及其后的版本默认包含 Liberation 字体套件。

1.1. 排版约定

我们使用四种排版约定突出特定用词和短语。这些约定及其使用环境如下。

单行粗体

用来突出系统输入，其中包括 shell 命令、文件名以及路径。还可用来突出按键以及组合键。例如：

> 要看到文件您当前工作目录中文件 my_next_bestselling_novel 的内容，请在 shell 提示符后输入 cat my_next_bestselling_novel 命令并按 Enter 键执行该命令。

以上内容包括一个文件名，一个 shell 命令以及一个按键，它们都以固定粗体形式出现，且全部与上下文有所区别。

组合键可通过使用连字符连接组合键的每个部分来与按键区别。例如：

> 按 Enter 执行该命令。
>
> 按 Ctrl+Alt+F2 切换到第一个虚拟终端。Ctrl+Alt+F1 返回您的 X-Windows 会话。

第一段突出的是要按的特定按键。第二段突出了两个按键组合（每个组合都要同时按）。下。

如果讨论的是源码、等级名称、方法、功能、变量名称以及在段落中提到的返回的数值，那么都会以上述形式出现，即固定粗体。例如：

> 与文件相关的等级包括用于文件系统的 filesystem、用于文件的 file 以及用于目录的 dir。每个等级都有其自身相关的权限。

比例粗体

这是指在系统中遇到的文字或者短语，其中包括应用程序名称、对话框文本、标记的按钮、复选框以及单选按钮标签、菜单标题以及子菜单标题。例如：

> 在主菜单条中选择「系统」→「首选项」→「鼠标」启动 鼠标首选项。在「按钮」标签中点击「惯用左手鼠标」 复选框并点击 关闭切换到主鼠标按钮从左向右（让鼠标适合左手使用）。
>
> 要在 gedit 文件中插入一个特殊字符，请在主菜单中选择「应用程序」→「附件」→「字符映射表」。下一步在 字符映射表菜单条中选择「搜索」→「查找」，在「搜索」字段输入字符名称并点击 下一个 按钮。您输入的字符会在「字符表」中突出出来。双击这个突出的字符将其放入「要复制的文本」字段，然后点击 复制 按钮。现在切换回您的文档并在 gedit 菜单条中选择「编辑」→「粘贴」。

[1] https://fedorahosted.org/liberation-fonts/

以上文本包括应用程序名称、系统范围菜单名称及项目、应用程序特定菜单名称以及按钮和 GUI 界面中的文本，所有都以比例粗体出现并与上下文区别。

固定粗斜体 或者 比例粗斜体

无论固定粗体或者比例粗体，附加的斜体表示是可替换或者变量文本。斜体表示那些不直接输入的文本或者那些根据环境改变的文本。例如：

> 要使用 ssh 连接到远程机器，请在 shell 提示符后输入 ssh username@domain.name。如果远程机器是 example.com 且您在该其机器中的用户名为 john，请输入 ssh john@example.com。
>
> mount -o remount file-system 命令会重新挂载命名的文件系统。例如：要重新挂载 /home 文件系统，则命令为 mount -o remount /home。
>
> 要查看目前安装的软件包版本，请使用 rpm -q package 命令。它会返回以下结果 ： package-version-release。

请注意以上文字中的粗斜体字 — username、domain.name、file-system、package、version 和 release。无论您输入文本或者运行一个命令，还是该系统显示的文本，每个字都是一个占位符。

不考虑工作中显示标题的标准用法，斜体表示第一次使用某个新且重要的用语。例如：

> Publican 是一个 DocBook 发布系统。

1.2. 抬升式引用约定

终端输出和源代码列表要与周围文本明显分开。

将发送到终端的输出设定为 Mono-spaced Roman 并显示为：

```
books        Desktop   documentation  drafts  mss    photos   stuff  svn
books_tests  Desktop1  downloads      images  notes  scripts  svgs
```

源码列表也设为 Mono-spaced Roman，但添加下面突出的语法：

```
package org.jboss.book.jca.ex1;

import javax.naming.InitialContext;

public class ExClient
{
   public static void main(String args[])
       throws Exception
   {
      InitialContext iniCtx = new InitialContext();
      Object         ref    = iniCtx.lookup("EchoBean");
      EchoHome       home   = (EchoHome) ref;
      Echo           echo   = home.create();

      System.out.println("Created Echo");

      System.out.println("Echo.echo('Hello') = " + echo.echo("Hello"));
   }
}
```

1.3. 备注及警告

最后，我们使用三种视觉形式来突出那些可能被忽视的信息。

备注

备注是对手头任务的提示、捷径或者备选的解决方法。忽略提示不会造成负面后果，但您可能会错过一个更省事的诀窍。

重要

重要框中的内容是那些容易错过的事情：配置更改只可用于当前会话，或者在应用更新前要重启的服务。忽略‘重要’框中的内容不会造成数据丢失但可能会让您抓狂。

警告

警告是不应被忽略的。忽略警告信息很可能导致数据丢失。

2. Feedback

If you spot a typo, or if you have thought of a way to make this manual better, we would love to hear from you. Please submit a report in Bugzilla (http://bugzilla.redhat.com/bugzilla/) against the component Documentation-cluster.

Be sure to mention the manual's identifier:

```
Virtual_Server_Administration(EN)-5 (2010-02-08T16:55)
```

By mentioning this manual's identifier, we know exactly which version of the guide you have.

If you have a suggestion for improving the documentation, try to be as specific as possible. If you have found an error, please include the section number and some of the surrounding text so we can find it easily.

Linux 虚拟服务器总览

Linux 虚拟服务器（LVS）是一组用来在真实服务器间平衡 IP 负载的整合软件组件。LVS 在一对配置相同的计算机中运行：一个是活跃 LVS 路由器，一个是备用 LVS 路由器。活跃 LVS 路由器有两个作用：

- 平衡真实服务器中的负载。
- 检查每个真实服务器中服务的完整性。

备用路由器的任务是监控活跃路由器并在活跃路由器出错的事件中扮演它的角色。

本章提供了 LVS 组件和功能的总览，它们由以下部分组成：

1.1. A Basic LVS Configuration

图 1.1 “A Basic LVS Configuration” shows a simple LVS configuration consisting of two layers. On the first layer are two LVS routers — one active and one backup. Each of the LVS routers has two network interfaces, one interface on the Internet and one on the private network, enabling them to regulate traffic between the two networks. For this example the active router is using Network Address Translation or NAT to direct traffic from the Internet to a variable number of real servers on the second layer, which in turn provide the necessary services. Therefore, the real servers in this example are connected to a dedicated private network segment and pass all public traffic back and forth through the active LVS router. To the outside world, the servers appears as one entity.

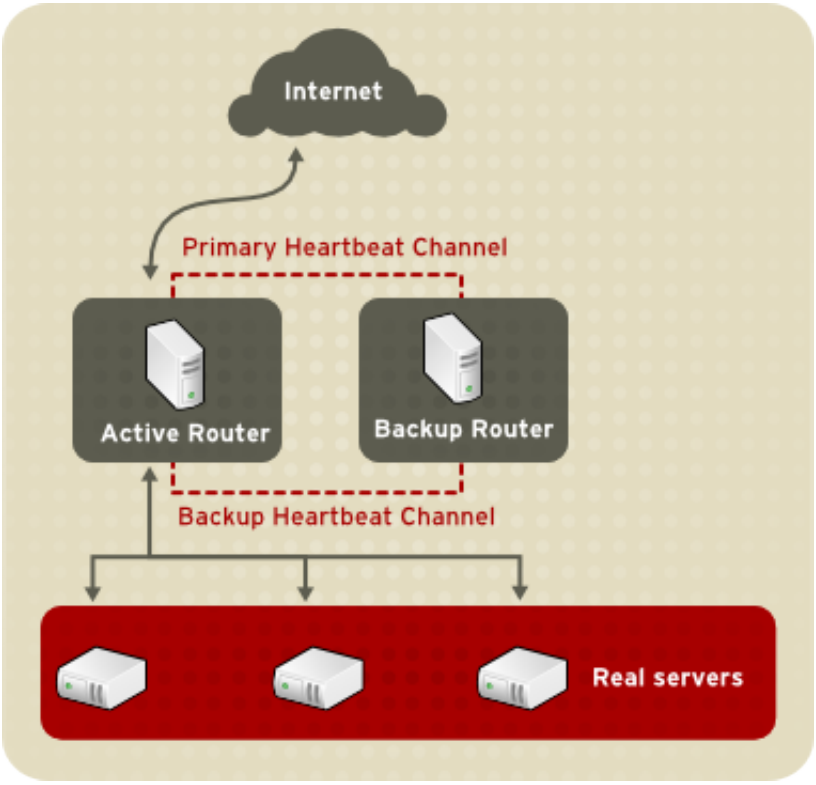

图 1.1. A Basic LVS Configuration

Service requests arriving at the LVS routers are addressed to a virtual IP address, or VIP. This is a publicly-routable address the administrator of the site associates with a fully-qualified domain name, such as www.example.com, and is assigned to one or more virtual servers. A virtual server is a service configured to listen on a specific virtual IP. Refer to 第 4.6 节 “VIRTUAL SERVERS” for more information on configuring a virtual server using the Piranha Configuration Tool. A VIP address migrates from one LVS router to the other during a failover, thus maintaining a presence at that IP address (also known as floating IP addresses).

VIP 地址还可以是同样将 LVS 路由器连接到互联网的设备的别名。例如：如果使用 eth0 连接到互联网，那么多个虚拟服务器就可以别名命名为 eth0:1。另外，每个虚拟服务器还可以根据服务关联到不同的设备。例如：HTTP 流量可由 eth0:1 处理，而 FTP 流量可由 eth0:2 处理。

Only one LVS router is active at a time. The role of the active router is to redirect service requests from virtual IP addresses to the real servers. The redirection is based on one of eight supported load-balancing algorithms described further in 第 1.3 节 “LVS 调度总览”.

活跃路由器还通过 send/expect 脚本动态监控真实服务器中特定服务的总体状态。侦测服务的状态需要动态数据，比如 HTTPS 或者 SSL。管理员还可以调用外部可执行程序。如果真实服务器中的某个服务失效，活跃路由器会停止向该服务器发送任务，直到它能够返回正常操作为止。

备用路由器是一个替补系统。LVS 路由器周期性地通过主要外部公共接口交换 heartbeat 信息，在失效切换的状态下，通过专用接口交换。备用节点应该无法在预期间隔之间接收 heartbeat 信息，它会启动一个失效切换，并假装执行活跃路由器的任务。在失效切换中，备用路由器接替了由出错的路由器提供的 VIP 地址，所用技术就是我们知道的 ARP 嗅探 — 在这里备用 LVS 路由器宣布它自己成为发往出错节点的 IP 数据包的目的地。当出错节点又可以提供服务时，备用节点由将自己设为随时可替换的角色。

The simple, two-layered configuration used in 图 1.1 “A Basic LVS Configuration” is best for serving data which does not change very frequently — such as static webpages — because the individual real servers do not automatically sync data between each node.

1.1.1. 在真实服务器之间的数据重复和数据共享

因为 LVS 中没有可用来在真实服务器之间共享相同数据的内置组件，所以管理员有两个基本选择：

- 在真实服务器池之间同步数据
- 为共享数据的访问在布局中添加第三层

对于不允许上传大量用户或者在真实服务器中进行数据修改的服务器来说，第一个选择是首选的。如果配置允许大量用户修改数据，比如电子商务网站，最好添加第三层。

1.1.1.1. 配置真实服务器来同步数据

管理员可用来同步真实服务器池数据的方法有很多。例如：可采用 shell 脚本，那么如果网页工程师更新了页面，就可同时将该页面发送到所有服务器中。还有，系统管理员可以使用类似 rsync 的程序来在设定的间隔期间重复所有节点中修改的数据。

但是，如果由于用户经常上传文件或者进行数据库传送造成配置超载，这种数据同步就不是最佳的同步方法。对于有高负载的配置，三层布局是最佳解决方案。

1.2. A Three-Tier LVS Configuration

图 1.2 “A Three-Tier LVS Configuration” shows a typical three-tier LVS topology. In this example, the active LVS router routes the requests from the Internet to the pool of real servers. Each of the real servers then accesses a shared data source over the network.

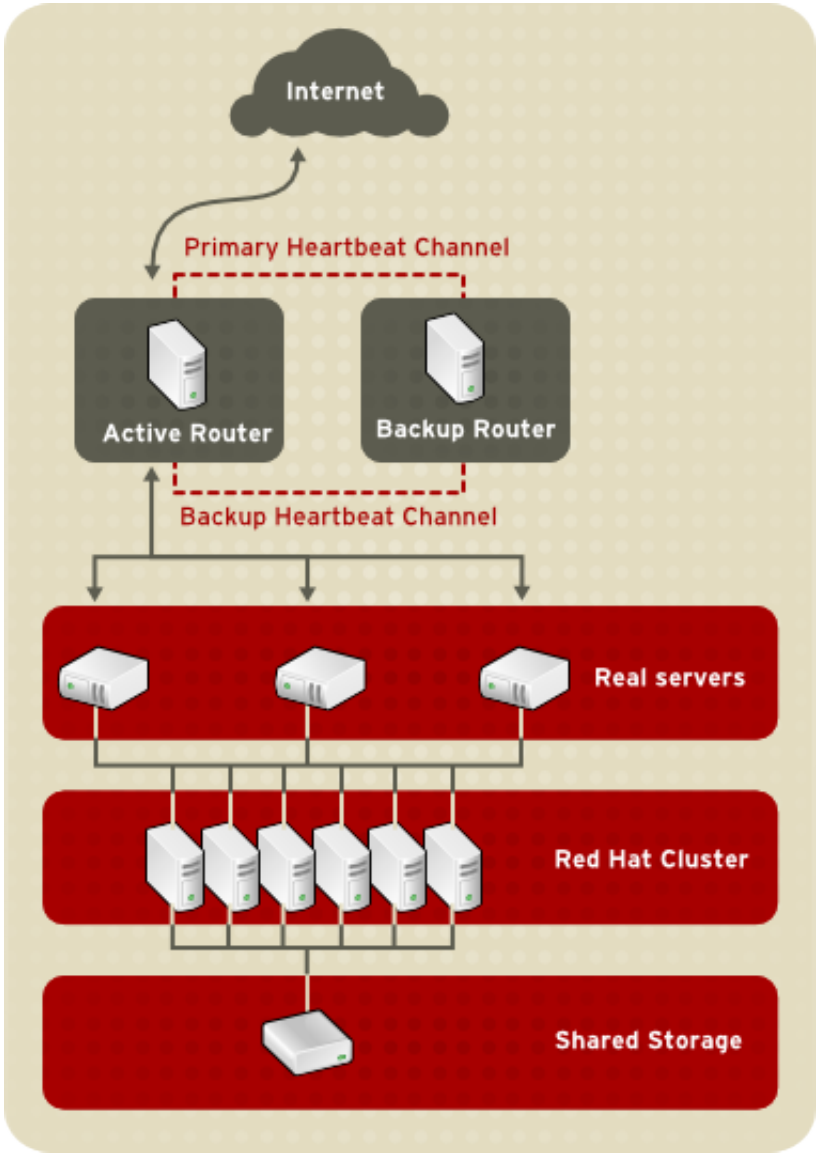

图 1.2. A Three-Tier LVS Configuration

此配置对于繁忙的 FTP 服务器最合适，服务器中可访问的数据都被保存在一个集中的高度可用的服务器中，且每个真实服务器都可通过一个导出的 NFS 目录或者 Samba 共享访问这些数据。我们还推荐在使用集中、高度可用数据库进行传送的网站使用此布局。另外，使用Red Hat Cluster Manager的 active-active 配置，管理员可配置一个高度可用的群集来同时扮演这两个角色。

上面示例中提到的第三层不一定要使用Red Hat Cluster Manager，但无法使用高度可用解决方案会导致严重的单点失败。

1.3. LVS 调度总览

使用 LVS 的优点之一就是它的灵活性，即可将 IP 级别负载均衡到真实服务器中。这种灵活性是因为配置 LVS 时管理员可以选择各种调度算法。LVS 负载平衡相对较少灵活性的方法来说更高级一些，比如轮叫 DNS，使用这种方法引起的 DNS 层级性和客户端机器缓存会导致负载失衡。另外，LVS 使用的底层过滤比应用程序层请求转发更有利，因为在网络数据包级别的平衡负载引起的计算超载最小，并可允许更大的可伸缩性。

Using scheduling, the active router can take into account the real servers' activity and, optionally, an administrator-assigned weight factor when routing service requests. Using

assigned weights gives arbitrary priorities to individual machines. Using this form of scheduling, it is possible to create a group of real servers using a variety of hardware and software combinations and the active router can evenly load each real server.

用于 LVS 的调度机制是由名为 IP 虚拟服务器或者 IPVS 模块的内核补丁集合提供的。这些模块启用了 layer 4 (L4) 传输层选项，该选项是设计用来在单一 IP 地址中更好地使用多个服务器。

要追踪数据包并将其有效路由到真实服务器中，IPVS 会在内核中建立一个 IPVS 表。活跃 LVS 路由器使用这个列表将来自虚拟服务器地址的请求重新路由并返回真实服务器池中的真实服务器中。ipvsadm 程序可随时更新 IPVS 列表 — 根据其可用性添加和删除群集成员。

1.3.1. 调度算法

The structure that the IPVS table takes depends on the scheduling algorithm that the administrator chooses for any given virtual server. To allow for maximum flexibility in the types of services you can cluster and how these services are scheduled, Red Hat Enterprise Linux provides the following scheduling algorithms listed below. For instructions on how to assign scheduling algorithms refer to 第 4.6.1 节 “「虚拟服务器」子界面”.

Round-Robin Scheduling

连续在真实服务器池中分配每个请求。使用此算法，所有真实服务器都会被同等对待，而不考虑其容量或者负载。这种调度模式延续了轮叫 DNS 但更加粗糙，因为它是基于网络连接而不是基于主机。LVS 轮叫调度不会陷入由 DNS 缓存查询造成的负载失衡状态。

Weighted Round-Robin Scheduling

Distributes each request sequentially around the pool of real servers but gives more jobs to servers with greater capacity. Capacity is indicated by a user-assigned weight factor, which is then adjusted upward or downward by dynamic load information. Refer to 第 1.3.2 节 “服务器加权和调度” for more on weighting real servers.

如果真实服务器池中的真实服务器之间有显著的差别，加权轮叫调度就是首选。但是，如果请求的负载有很大不同，那么加权强的服务器会回应更多的请求。

Least-Connection

为有较少活跃连接的服务器发送更多请求。因为它会通过 IPVS 列表追踪到真实服务器的活跃连接，最小连接是动态调度算法的一类，在请求负载差别很大时是上佳选择。它最适用于每个节点有类似容量的真实服务器池。如果一组服务器有不同的容量，加权最小连接调度则是更好的选择。

Weighted Least-Connections (default)

Distributes more requests to servers with fewer active connections relative to their capacities. Capacity is indicated by a user-assigned weight, which is then adjusted upward or downward by dynamic load information. The addition of weighting makes this algorithm ideal when the real server pool contains hardware of varying capacity. Refer to 第 1.3.2 节 “服务器加权和调度” for more on weighting real servers.

Locality-Based Least-Connection Scheduling

为与相对它们的目的 IP 有更少活跃连接的服务器分配更多的请求。这种算法是设计用于代理服务器缓存的服务器群集。它会为 IP 地址将数据包路由到服务器，除非该服务器已经超过了它的容量，并另有服务器只使用了容量的一半，在这种情况下，它会将 IP 地址分配给最小负载的真实服务器。

Locality-Based Least-Connection Scheduling with Replication Scheduling

为与相对它们的目的 IP 有更少活跃连接的服务器分配更多的请求。这种算法是设计用于代理服务器缓存的服务器群集。它和使用将目标 IP 与真实服务器节点的子网络进行映射的局部最小连接调度不同。请求会被路由到子网络中有最少连接的服务器中。如果目的 IP 的所有节点都超过了容量

，它会为那个目的 IP 复制一个新的服务器，这可通过将真实服务器池中那个有最小连接的真实服务器为目的地址 IP 添加到真实服务器的子网中实现。然后会从真实服务器子网中除去负载最大的节点以免过度重复。

Destination Hash Scheduling
: 通过在静态散列列表中查看目的 IP 来将请求分配到真实服务器池中。这个算法是设计用于代理服务器缓存的服务器群集。

Source Hash Scheduling
: 通过在静态散列列表中查看目的 IP 来将请求分配到真实服务器池中。这个算法是为带多个防火墙的 LVS 路由器设计的。

1.3.2. 服务器加权和调度

LVS 管理员可以为真实服务器池中的每一个节点分配一个加权。这个加权是一个整数值，它可成为考虑加权调度算法的一个因素（比如加权的最小连接），且可帮助 LVS 路由器为有不同容量的硬件更平均地分配负载。

加权充当服务器间比例的作用。例如：如果一个真实服务器的加权为 1，另一个的加权为 5，那么加权为 5 的服务器每有五个连接时，加权为 1 的服务器有一个连接。默认真实服务器加权值为 1。

尽管将加权添加到真实服务器池中的不同硬件配置可使群集的负载平衡更加有效，但它也会在将真实服务器池中添加一个真实服务器，或者在调度虚拟服务器使用加权的最小连接时造成暂时失衡。例如：假设在真实服务器池中有三个服务器，服务器 A 和 B 为加权 1 和 3，服务器 C 为加权 2。如果服务器 C 由于某种原因当机，服务器 A 和 B 就会平级分配被丢弃的负载。但服务器 C 重新上线后，LVS 路由器会视其为没有连接的服务器，并且将所有进入请求都一股脑发送到这台服务器中，直到和服务器 A 和 B 持平。

要防止此现象出现，管理员可将虚拟服务器设为 quiesce 服务器 — 无论何时当有新的服务器节点上线时，都将最小连接表重新设为 0，且 LVS 路由器象所有真实服务器都是刚刚添加到群集中一样路由请求。

1.4. 路由方法

Red Hat Enterprise Linux 在利用可用硬件或者将 LVS 整合到现有网络中时，使用可为管理员提供极大灵活性的网络地址转换或者 NAT 路由。

1.4.1. NAT 路由

图 1.3 “LVS Implemented with NAT Routing”，illustrates LVS utilizing NAT routing to move requests between the Internet and a private network.

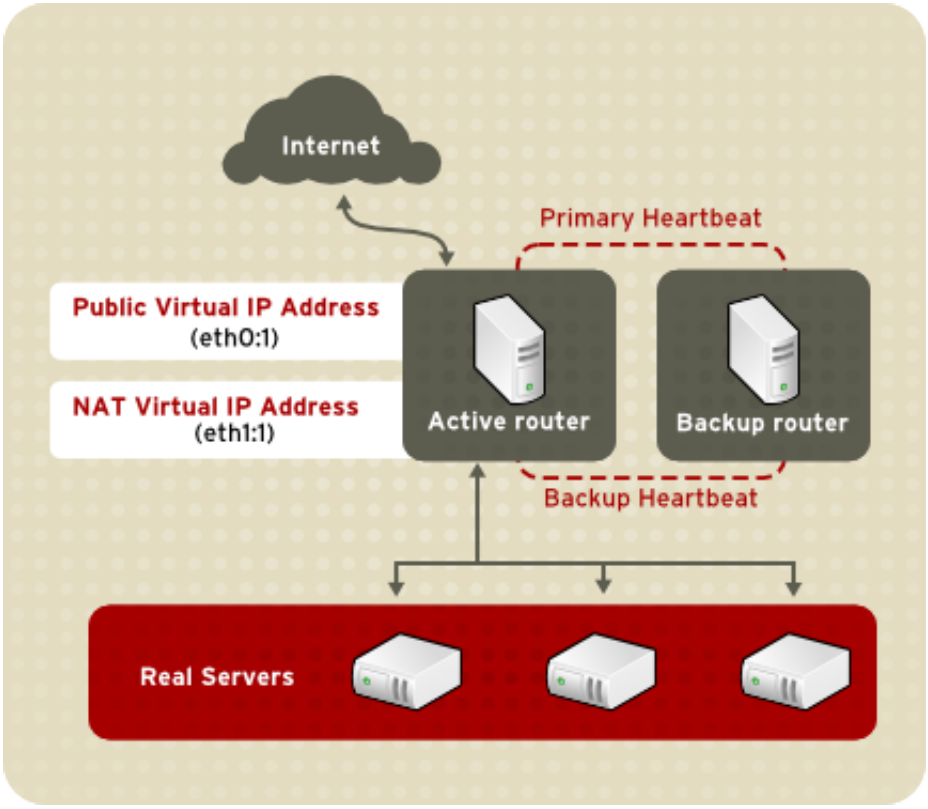

图 1.3. LVS Implemented with NAT Routing

在本示例中，活跃 LVS 路由器中有两个 NIC。用于互联网的 NIC 在 eth0 中有一个真实 IP 地址，并有一个别名为 eth0:1 的浮动 IP 地址。用于专用网络接口的 NIC 在 eth1 中有一个真实 IP 地址，并有一个别名为 eth1:1 的浮动 IP 地址。在发生失效切换时，面向互联网的虚拟接口和面向虚拟接口的专用接口同时由备用 LVS 路由器取代。所有位于专用网络中的真实服务器为 NAT 路由器使用浮动 IP 地址，因为它们默认路由是和活跃 LVS 路由器沟通，以便不会影响到对来自互联网请求的回应。

In this example, the LVS router's public LVS floating IP address and private NAT floating IP address are aliased to two physical NICs. While it is possible to associate each floating IP address to its own physical device on the LVS router nodes, having more than two NICs is not a requirement.

使用这种布局，活跃 LVS 路由器可接收请求并将其路由到适当的服务器。然后真实服务器处理该请求并将数据包返回到 LVS 路由器，该路由器使用网络地址转换将数据包中的真实服务器地址替换为 LVS 路由器公共 VIP 地址。这个过程被称为 IP 伪装，因为发出请求的客户端无法看到真实服务器的实际 IP 地址。

使用这种 NAT 路由，真实服务器可以是运行各种操作系统的机器。最大的缺点就是在较大群集部署中 LVS 路由器可能会成为瓶颈，因为它必须处理外发和进入请求。

1.4.2. 直接路由

建立使用直接路由的 LVS 设置和其它 LVS 联网布局相比有更好的性能。直接路由允许真实服务器将数据包直接处理并路由到发出请求的用户，而不是将所有外发的数据包通过 LVS 路由器发送给用户。直接路由通过将 LVS 路由器的任务变为仅仅处理进入的数据包，从而降低了出现网络性能问题的可能性。

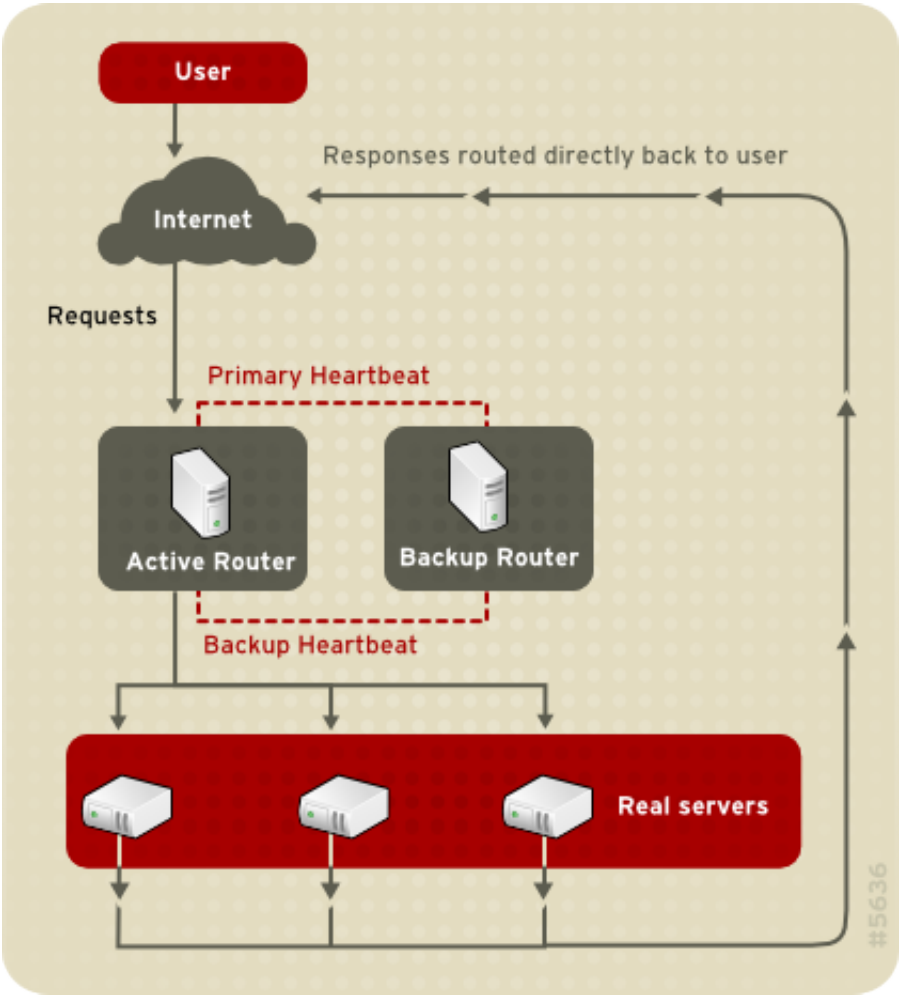

图 1.4. LVS Implemented with Direct Routing

在典型的直接路由设置中，LVS 路由器通过虚拟 IP（VIP）接收进入服务器的请求，使用调度算法将请求路由到真实服务器中。真实服务器会处理这些请求，并将回复绕过 LVS 路由器直接发送给客户端。这种路由方法允许在不增加 LVS 路由器将外发数据包从真实服务器路由到客户端负担的情况下添加真实服务器的能力，以免在网络负载较重的情况下形成瓶颈。

1.4.2.1. 直接路由和 ARP 限制

虽然在 LVS 中使用直接路由有很多优点，但也有一些局限。LVS 使用直接路由最常见的问题就出现在地址解析协议（ARP）。

In typical situations, a client on the Internet sends a request to an IP address. Network routers typically send requests to their destination by relating IP addresses to a machine's MAC address with ARP. ARP requests are broadcast to all connected machines on a network, and the machine with the correct IP/MAC address combination receives the packet. The IP/MAC associations are stored in an ARP cache, which is cleared periodically (usually every 15 minutes) and refilled with IP/MAC associations.

在直接路由 LVS 设置中出现 ARP 请求问题就是因为客户端对某个 IP 地址的请求必须与要处理请求的 MAC 地址关联，LVS系统的虚拟 IP 地址也必须与 MAC 关联。但由于 LVS 路由器和真实服务器有相同的 VIP，因此 ARP 请求会被广播到与该 VIP 关联的所机器。这会引发一些问题，比如完全绕过 LVS 路由器将 VIP 直接关联到真实服务器之一并直接处理请求，与设置 LVS 的初衷项背。

要解决这个问题，请确定总是将进入请求发送到 LVS 路由器，而不是真实服务器。使用 arptables_jf 或者 iptables 数据包过滤工具即可达到此目的，理由如下：

- arptables_jf 可防止 ARP 将 VIP 与真实服务器关联。

- iptables 方法完全避免了 ARP 问题，因为它从来没有在真实服务器中配置 VIP。

For more information on using arptables or iptables in a direct routing LVS environment, refer to 第 3.2.1 节 “直接路由及 arptables_jf” or 第 3.2.2 节 “直接路由及 iptables”.

1.5. 持久性和防火墙标记

在特殊情况下，可能会需要客户端重复地重新连接到同一个真实服务器，而不是让 LVS 负载平衡算法将请求发送到最可用的服务器。有关示例包括多屏幕网页表格、cookies、SSL 和 FTP 连接。在这些情况下，如果传送不是由同一个服务器处理来保持上下文环境，客户端可能无法正常工作。LVS 为处理这种情况提供了两个不同的特性：持久性和防火墙标记。

1.5.1. 持久性

启用后，持久性起到定时器的作用。当客户端连接一个服务时，LVS 会在指定的时间内记住最后的连接。如果具有相同 IP 地址的客户在这段时间内再次进行连接，它将被送往和上次连接相同的服务器 — 忽略负载平衡机制。而当连接在这段时间外发生，它会按照适当的调度规则进行处理。

持久性还允许管理员指定客户端 IP 地址测试使用的子网掩码，它可作为工具来控制什么地址可用有更高级别的持久性，从而将连接分组到那个子网中。

将目的地址为不同端口的连接分组对使用多个端口进行沟通的协议很重要，比如 FTP。但持久性在处理将目的地址为不同的端口的连接进行分组时并不是最有效的方法。在这种情况下，最佳方案是使用防火墙标记。

1.5.2. 防火墙标记

防火墙标记是为用于某个协议的端口进行分组或者为相关协议进行分组的简便、有效的方法。例如，如果将 LVS 部署到某个电子商务网站中，可使用防火墙标记将 HTTP 连接绑定到端口 80，将 HTTPS 连接固定到端口 443。通过为每个协议将相同的防火墙标价分配到虚拟服务器中，就可以保留传送的状态信息，因为 LVS 路由器会在打开某个连接后将所有请求都转发到同一个真实服务器中。

由于其高效、易用，LVS 管理员应该在任何可能需要对连接进行分组时使用防火墙标记而不是持久性。但是管理员应该仍然在虚拟服务器中添加持久性，使之与防火墙标记合并使用以确保客户端在一段特定时间内会重复连接到同一个服务器。

1.6. LVS — 框图

LVS routers use a collection of programs to monitor cluster members and cluster services. 图 1.5 “LVS Components” illustrates how these various programs on both the active and backup LVS routers work together to manage the cluster.

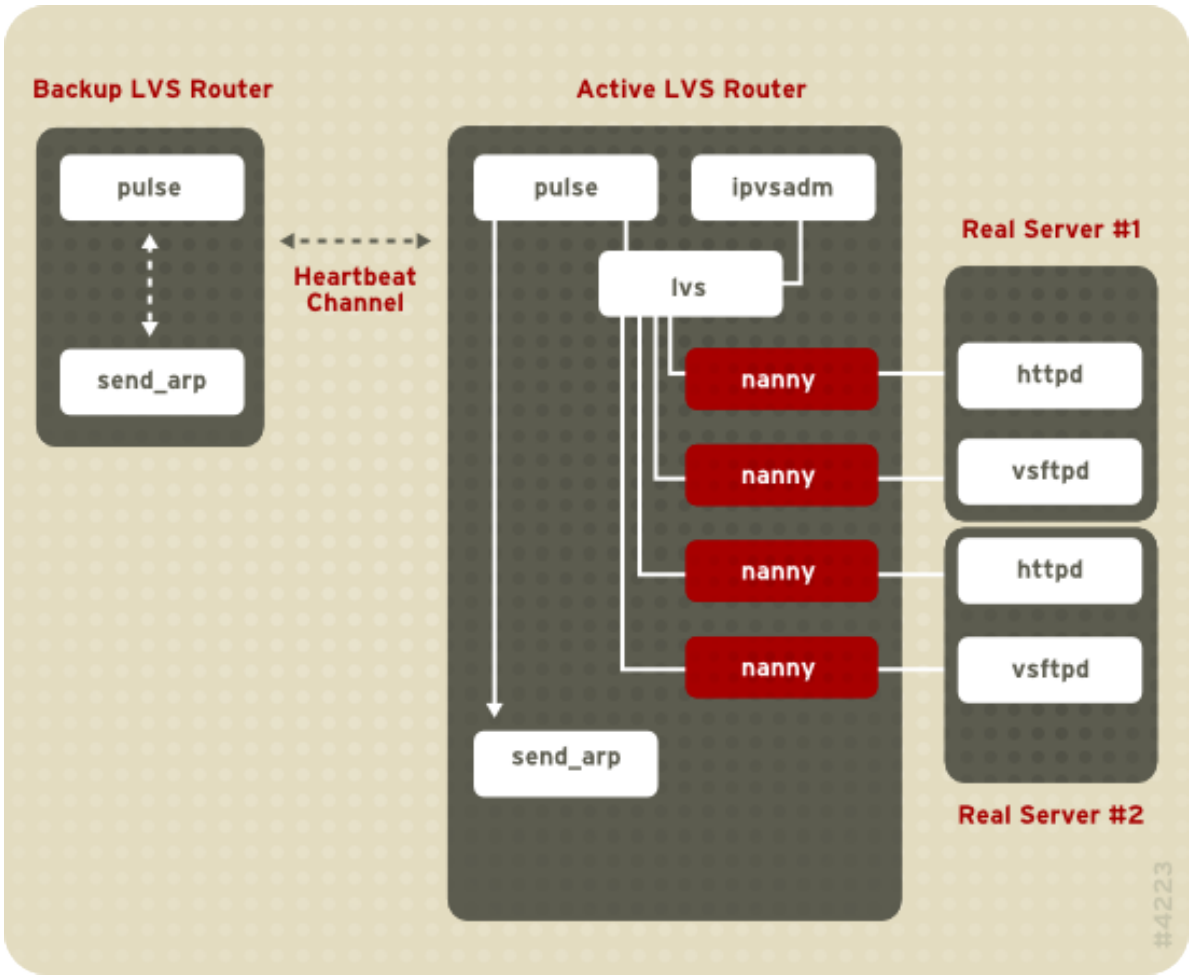

图 1.5. LVS Components

活跃和备用 LVS 路由器中都会运行 pulse 守护进程。在备用路由器中，pulse 向活跃服务器的公共接口发送一个 heartbeat 来确定活跃路由器仍可正常工作。在活跃服务器中，pulse 启动 lvs 守护进程，并回应来自备用 LVS 路由器的 heartbeat 查询。

启动后，lvs 守护进程调用 ipvsadm 程序来配置和维护内核中的 IPVS 路由表，并为每个真实服务器中配置的虚拟服务器启动 nanny 进程，同时告知 lvs 该真实服务器中的服务是否正常工作。如果发现故障，lvs 守护进程会向 ipvsadm 发出指令将那个真实服务器从 IPVS 路由表中删除。

如果备用路由器没有收到来自活跃路由器的响应，它会通过调用 send_arp 启动失效切换来将所有虚拟 IP 地址重新分配到备用节点的 NIC 硬件地址（MAC 地址），并通过公共和专用网络接口向活跃路由器发送一个命令来关闭活跃路由器中的 lvs 守护进程，启动备用节点中的 lvs 守护进程来为配置的虚拟服务器接收请求。

1.6.1. LVS Components

第 1.6.1.1 节 “pulse” shows a detailed list of each software component in an LVS router.

1.6.1.1. pulse

This is the controlling process which starts all other daemons related to LVS routers. At boot time, the daemon is started by the /etc/rc.d/init.d/pulse script. It then reads the configuration file /etc/sysconfig/ha/lvs.cf. On the active router, pulse starts the

LVS daemon. On the backup router, pulse determines the health of the active router by executing a simple heartbeat at a user-configurable interval. If the active router fails to respond after a user-configurable interval, it initiates failover. During failover, pulse on the backup router instructs the pulse daemon on the active router to shut down all LVS services, starts the send_arp program to reassign the floating IP addresses to the backup router's MAC address, and starts the lvs daemon.

1.6.1.2. lvs

一旦被 pulse 调用，lvs 守护进程就运行在活动 LVS 路由器中。它读取配置文件 /etc/sysconfig/ha/lvs.cf，调用 ipvsadm 工具来构建和维护 IPVS 路由表，并为每个配置的 LVS 服务分配 nanny 进程。如果 nanny 报告某个服务器关闭了，lvs 将指引 ipvsadm 工具从 IPVS 路由表中删除这个服务器。

1.6.1.3. ipvsadm

这个服务更新内核中的 IPVS 路由表。lvs 守护进程通过调用 ipvsadm 来添加、修改或者删除 IPVS 路由表中的条目来设置和管理 LVS。

1.6.1.4. nanny

nanny 监控运行在活跃 LVS 路由器中的守护进程。通过这个守护进程，活跃路由器可确定每个真实服务器的状态，有时还可以监控其工作负载。每个真实服务器中定义的每个服务都有一个独立进程为其运行。

1.6.1.5. /etc/sysconfig/ha/lvs.cf

这是 LVS 配置文件。所有守护进程都直接或者间接地从这个文件中获得它们的配置信息。

1.6.1.6. Piranha Configuration Tool

这是用来监控、配置和管理 LVS 的网页工具。它是用来维护 /etc/sysconfig/ha/lvs.cf LVS 配置文件的默认工具。

1.6.1.7. send_arp

在失效切换过程中，当浮动 IP 地址在节点间进行更改时，这个程序发送 ARP 广播。

第 2 章 初始 LVS 配置 reviews important post-installation configuration steps you should take before configuring Red Hat Enterprise Linux to be an LVS router.

初始 LVS 配置

安装Red Hat Enterprise Linux 后，您必须执行一些基本操作步骤来设置 LVS 路由器和真实服务器（real server）。本章对这些初始化步骤进行了详细的论述。

注记

当启动群集后，LVS 路由器节点就成为活跃节点，也叫主节点。在配置 LVS 时，请使用主节点中的 Piranha Configuration Tool。

2.1. 在 LVS 路由器中配置服务

Red Hat Enterprise Linux 安装程序会安装所有设置 LVS 所需要的组件，但必须在配置群集前激活正确的服务。请为两个 LVS 路由器在引导时启动正确的服务。Red Hat Enterprise Linux 中有三个主要的工具可用来将服务设置为在引导时激活，它们是命令行程序chkconfig、ncurses-based 程序 ntsysv 和图形界面程序 Services Configuration Tool。这些工具都要求有根访问才可以使用。

注记

要获得根访问权限，请在 shell 提示符后输入 su - 命令和根密码。例如：

```
$ su - root password
```

在 LVS 路由器中，需要将三个服务设置为在引导时激活：

- piranha-gui 服务（只用于主节点）
- pulse 服务
- sshd 服务

如果您正在群集多端口服务或者正在使用防火墙标记，您还必须启用 iptables 服务。

最好是将这些服务设置为在运行级别 3 和运行级别 5 都激活。要达到此目的，请使用 chkconfig，并为每个服务输入以下命令：

```
/sbin/chkconfig --level 35 daemon on
```

在上面的命令中，请使用您想要激活的服务名称替换 daemon。要获得系统中的服务及在什么运行级别将其设定为激活的列表，请使用以下命令：

```
/sbin/chkconfig --list
```

警告

Turning any of the above services on using chkconfig does not actually start the daemon. To do this use the /sbin/service command. See 第 2.3 节 “启动 Piranha Configuration Tool服务” for an example of how to use the /sbin/service command.

For more information on runlevels and configuring services with ntsysv and the Services Configuration Tool, refer to the chapter titled "Controlling Access to Services" in the Red Hat Enterprise Linux System Administration Guide.

2.2. 为 Piranha Configuration Tool设置密码

第一次在主 LVS 路由器中使用 Piranha Configuration Tool之前，您必须创建一个密码来限制对它的访问。创建密码时，请以根用户登录，并使用以下命令：

/usr/sbin/piranha-passwd

输入此命令后，根据提示创建管理密码。

警告

比较安全的密码不可以是专有名词、常用缩写或者可在任意语言字典中查到的单词。不要在系统中留下任何未加密的密码。

如果要在激活的 Piranha Configuration Tool会话中修改密码，系统会为管理员提示输入新密码。

2.3. 启动 Piranha Configuration Tool服务

在您为 Piranha Configuration Tool设定密码后，请启动或者重启位于 /etc/rc.d/init.d/piranha-gui 的 piranha-gui 服务。此时请以根用户身份输入以下命令：

/sbin/service piranha-gui start

or

/sbin/service piranha-gui restart

Issuing this command starts a private session of the Apache HTTP Server by calling the symbolic link /usr/sbin/piranha_gui -> /usr/sbin/httpd. For security reasons, the piranha-gui version of httpd runs as the piranha user in a separate process. The fact that piranha-gui leverages the httpd service means that:

1. 必须在系统中安装 Apache HTTP Server。

2. 通过用 service 命令终止 piranha-gui 服务来终止或者重启 Apache HTTP Server。

警告

如果是在 LVS 路由器中使用 /sbin/service httpd stop 或者 /sbin/service httpd restart 命令，您必须使用以下命令启动 piranha-gui 服务：

```
/sbin/service piranha-gui start
```

The piranha-gui service is all that is necessary to begin configuring LVS. However, if you are configuring LVS remotely, the sshd service is also required. You do not need to start the pulse service until configuration using the Piranha Configuration Tool is complete. See 第 4.8 节 “启动 LVS” for information on starting the pulse service.

2.3.1. 配置 Piranha Configuration Tool网页服务器端口

默认情况下，Piranha Configuration Tool在端口 3636 运行。要更改此端口号，请在 piranha-gui 网页服务器配置文件 /etc/sysconfig/ha/conf/httpd.conf 中修改第二部分的 Listen 3636 行。

要使用 Piranha Configuration Tool，您至少需要一个文本网页浏览器。如果您在主 LVS 路由器中启动了网页浏览器，请打开位置 http://localhost:3636。您可以用主 LVS 路由器的主机名或者 IP 地址替换 localhost，以便通过网页浏览器从任意位置进入 Piranha Configuration Tool。

当您的浏览器连接到 Piranha Configuration Tool时，您必须登录来访问配置服务。在「用户名」字段输入 piranha，在「密码」字段输入 piranha-passwd。

现在 Piranha Configuration Tool正在运行，您可能想要考虑要对网络中访问此工具的人员进行限制。下面的部分就是总结如何达到此目的。

2.4. 限制对 Piranha Configuration Tool的访问

Piranha Configuration Tool提示一个可用的用户名和密码组合。但所有传递给 Piranha Configuration Tool的数据都是明文的，因此建议您将对它的访问限制在可信的网络或者本地机器。

The easiest way to restrict access is to use the Apache HTTP Server's built in access control mechanisms by editing /etc/sysconfig/ha/web/secure/.htaccess. After altering the file you do not have to restart the piranha-gui service because the server checks the .htaccess file each time it accesses the directory.

默认情况下，对此目录的访问控制允许任何人浏览目录内容。以下就是默认访文件的形式：

```
Order deny,allow
Allow from all
```

要将对 Piranha Configuration Tool的访问限制在只允许本地主机访问，请将 .htaccess 文件修改为只允许来自回送设备（127.0.0.1）的访问。有关回送设备的详情请参考Red Hat Enterprise Linux Reference Guide的网络脚本一章。

```
Order deny,allow
Deny from all
```

```
Allow from 127.0.0.1
```

本示例中您还可以允许特定主机或者子网：

```
Order deny,allow
Deny from all
Allow from 192.168.1.100
Allow from 172.16.57
```

在本示例中，只有来自 IP 地址为 192.168.1.100 和 172.16.57/24 网络中的机器使用的网页浏览器可访问 Piranha Configuration Tool。

警告

编辑 Piranha Configuration Tool .htaccess 文件可限制对位于 /etc/sysconfig/ha/web/secure/ 目录中的配置页的访问，但不会限制登录和对位于 /etc/sysconfig/ha/web/ 目录的帮助页的访问。要限制对此目录的访问，请在 /etc/sysconfig/ha/web/ 目录中创建 .htaccess 文件，并使用和 /etc/sysconfig/ha/web/secure/.htaccess 文件一样的 order、allow 和 deny 行。

2.5. 启动数据包转发

要让 LVS 路由器将网络数据包正确转发给真实服务器，那么每个 LVS 路由器节点必须在内核中打开 IP 转发功能。以根用户登录，将 /etc/sysctl.conf 文件中的 net.ipv4.ip_forward = 0 改为：

```
net.ipv4.ip_forward = 1
```

重启系统即可使修改生效。

要检查是否打开了 IP 转发，请以根用户身份使用以下命令：

```
/sbin/sysctl net.ipv4.ip_forward
```

如果以上命令返回结果为 1，那么 IP 转发就启用了。如果返回的结果是 0，那么您就需要用以下命令手动打开此功能：

```
/sbin/sysctl -w net.ipv4.ip_forward=1
```

2.6. 在真实服务器中配置服务

如果真实服务器是Red Hat Enterprise Linux 系统，您可设置在引导时激活适当的服务器守护进程。这些守护进程包括网页服务的 httpd 或者 FTP 和 Telnet 服务的 xinetd。

还可以远程访问真实服务器，这就需要安装并运行 sshd 守护进程。

设置 LVS

LVS 群集包括两个基本群组：LVS 路由器和真实服务器。要防止单点失败，每个群组应该包含至少两个成员系统。

LVS 路由器群组应该包括两个相同或者非常类似的运行Red Hat Enterprise Linux 的系统。其中一个作为活跃 LVS 路由器使用，同时另一个处于热等待模式，因此它们需要有尽可能相似的容量。

在为真实服务器组群选择和配置硬件时，您必须决定使用三种 LVS 布局中的哪一种。

3.1. NAT LVS 网络

NAT 布局允许大限度利用现有硬件，但因为所有进出服务器池的数据包都经过 LVS 路由器，所以会限制其处理大负载的能力。

网络布局

使用 NAT 路由的 LVS 布局是根据网络方案透视进行配置的最简单的方法，因为只需要一个切入点访问公共网络。真实服务器会将所有请求返回到 LVS 路由器，这样就可以让它们在其专用网络中了。

硬件

从硬件考虑，NAT 布局是最灵活的布局，因为真实服务器不一定是 Linux 机器才能正常工作。在 NAT 布局中，每个真实服务器只需要一个 NIC，因为它只响应 LVS 路由器。另一方面，LVS 路由器需要两个 NIC 来在两个网络间路由流量。因为此布局在 LVS 路由器中产生了网络瓶颈，所以可以在每个 LVS 路由器中部署千兆以太网 NIC（gigabit Ethernet NIC）来提高 LVS 路由器可处理的带宽。如果在 LVS 路由器中使用了千兆以太网 NIC，每个连接真实服务器和 LVS 路由器的开关必须至少有两个千兆以太网端口来有效处理负载。

软件

因为 NAT 布局需要使用 iptables 进行某些配置，所以在 Piranha Configuration Tool之外还需要配置相当数量的软件。特别是在使用FTP 服务和防火墙标记时需要额外手动配置 LVS 路由器以便正确路由请求。

3.1.1. 为带 NAT 的 LVS 配置网络接口

To set up LVS with NAT, you must first configure the network interfaces for the public network and the private network on the LVS routers. In this example, the LVS routers' public interfaces (eth0) will be on the 192.168.26/24 network (I know, I know, this is not a routable IP, but let us pretend there is a firewall in front of the LVS router for good measure) and the private interfaces which link to the real servers (eth1) will be on the 10.11.12/24 network.

So on the active or primary LVS router node, the public interface's network script, /etc/sysconfig/network-scripts/ifcfg-eth0, could look something like this:

```
DEVICE=eth0
BOOTPROTO=static
ONBOOT=yes
IPADDR=192.168.26.9
NETMASK=255.255.255.0
GATEWAY=192.168.26.254
```

专用 LVS 路由器 NAT 接口的 /etc/sysconfig/network-scripts/ifcfg-eth1 应类似如下：

```
DEVICE=eth1
BOOTPROTO=static
ONBOOT=yes
IPADDR=10.11.12.9
NETMASK=255.255.255.0
```

In this example, the VIP for the LVS router's public interface will be 192.168.26.10 and the VIP for the NAT or private interface will be 10.11.12.10. So, it is essential that the real servers route requests back to the VIP for the NAT interface.

重点

The sample Ethernet interface configuration settings in this section are for the real IP addresses of an LVS router and not the floating IP addresses. To configure the public and private floating IP addresses the administrator should use the Piranha Configuration Tool, as shown in 第 4.4 节 “GLOBAL SETTINGS” and 第 4.6.1 节 “「虚拟服务器」子界面”.

After configuring the primary LVS router node's network interfaces, configure the backup LVS router's real network interfaces — taking care that none of the IP address conflict with any other IP addresses on the network.

重点

请确定每个位于备用节点接口的服务提供与主节点相同的接口。例如，如果在主节点中使用 eth0 连接到公共网络，那么也要使用它在备用节点连接公共网络。

3.1.2. 在真实服务器中路由

在配置 NAT 布局的真实服务器网络接口时，最重要的是要记住为 LVS 路由器的 NAT 浮动 IP 地址设定网关。在本示例中，该地址应该是 10.11.12.10。

注记

Once the network interfaces are up on the real servers, the machines will be unable to ping or connect in other ways to the public network. This is normal. You will, however, be able to ping the real IP for the LVS router's private interface, in this case 10.11.12.8.

So the real server's /etc/sysconfig/network-scripts/ifcfg-eth0 file could look similar to this:

```
```

```
DEVICE=eth0
ONBOOT=yes
BOOTPROTO=static
IPADDR=10.11.12.1
NETMASK=255.255.255.0
GATEWAY=10.11.12.10
```

警告

如果真实服务器有超过一个网络接口配置了 GATEWAY= 行，第一个出现的将是网关。因此，如果同时配置了 eth0 和 eth1，而且 eth1 用于 LVS，真实服务器可能无法正确路由请求。

最好是在他们位于 /etc/sysconfig/network-scripts/ 目录的网络脚本 ONBOOT=no 中设定关闭无关的网络接口，或者确定在第一个要出现的接口中正确设置了网关。

3.1.3. 启动 LVS 路由器中的 NAT 路由

In a simple NAT LVS configuration where each clustered service uses only one port, like HTTP on port 80, the administrator needs only to enable packet forwarding on the LVS routers for the requests to be properly routed between the outside world and the real servers. See 第 2.5 节 “启动数据包转发” for instructions on turning on packet forwarding. However, more configuration is necessary when the clustered services require more than one port to go to the same real server during a user session. For information on creating multi-port services using firewall marks, see 第 3.4 节 “多端口服务和 LVS ”.

Once forwarding is enabled on the LVS routers and the real servers are set up and have the clustered services running, use the Piranha Configuration Tool to configure LVS as shown in 第 4 章 用 Piranha Configuration Tool配置 LVS 路由器.

警告

Do not configure the floating IP for eth0:1 or eth1:1 by manually editing network scripts or using a network configuration tool. Instead, use the Piranha Configuration Tool as shown in 第 4.4 节 “GLOBAL SETTINGS” and 第 4.6.1 节 “「虚拟服务器」子界面 ”.

When finished, start the pulse service as shown in 第 4.8 节 “启动 LVS”. Once pulse is up and running, the active LVS router will begin routing requests to the pool of real servers.

3.2. 使用直接路由的 LVS

As mentioned in 第 1.4.2 节 “直接路由”, direct routing allows real servers to process and route packets directly to a requesting user rather than passing outgoing packets through the LVS router. Direct routing requires that the real servers be physically connected to a network segment with the LVS router and be able to process and direct outgoing packets as well.

网络布局

在直接路由 LVS 设置中，LVS 路由器需要接收进入请求，并将其路由到适当的真实服务器进行处理。接着真实服务器需要直接将响应路由给客户端。例如：如果客户端在互联网中并通过 LVS 路由器向真实服务器发送数据包，那么真实服务器必须可以通过互联网直接连接到客户端。这可通过为真实服务器配置网关来将数据包发送到互联网中。服务器池中的每个真实服务器可以有它们独立的网关（且每个网关都有其自身的互联网连接），这可允许最大限度的吞吐量和可伸缩性。但对于典型 LVS 设置，真实服务器可通过一个网关（也就是一个网络连接）进行沟通。

重点

我们不推荐您将 LVS 路由器作为真实服务器的网关使用，因为这样会带来不必要的对 LVS 路由器复杂设置和网络负载，这些内容我们将在 NAT 路由中存在的网络瓶颈中再次论述。

硬件

使用直接路由的 LVS 系统的硬件要求与其它 LVS 布局类似。当需要在Red Hat Enterprise Linux 中运行 LVS 路由器来处理进入请求并为真实服务器执行负载平衡时，真实服务器不一定是 Linux 机器才可正常工作。每个 LVS 路由器需要一个或者两个 NIC（要看是否有备用路由器）。您可以用两个 NIC 来缓解配置并完全分离流量 — 进入请求由一个 NIC 处理，用另一个 NIC 将数据包路由到真实服务器。

因为真实服务器会绕过 LVS 路由器并将外发的数据包直接发送给客户端，所以需要连接到互联网的网关。要获得最高性能和可用性，每个真实服务器应使用独立网关连接到互联网，这些独立网关有其专用的连接连接到客户端连接的载体网络（比如互联网或者内部网络）。

软件

There is some configuration outside of Piranha Configuration Tool that needs to be done, especially for administrators facing ARP issues when using LVS via direct routing. Refer to 第 3.2.1 节 “直接路由及 arptables_jf” or 第 3.2.2 节 “直接路由及 iptables” for more information.

3.2.1. 直接路由及 arptables_jf

In order to configure direct routing using arptables_jf, each real server must have their virtual IP address configured, so they can directly route packets. ARP requests for the VIP are ignored entirely by the real servers, and any ARP packets that might otherwise be sent containing the VIPs are mangled to contain the real server's IP instead of the VIPs.

用 arptables_jf 方法，可将应用程序绑定到每个 VIP 或者所有真实服务器服务的端口。例如：arptables_jf 方法允许 Apache HTTP Server 的多个事件明确绑定到系统中的不同 VIP 而运行。使用 arptables_jf 的 IPTables 选项还有非常优越的性能。

但使用 arptables_jf 方法，无法使用标准Red Hat Enterprise Linux 系统配置工具将 VIP 配置为在引导时启动。

要将每个真实服务器配置为忽略虚拟 IP 地址的 ARP 请求，请按以下步骤操作：

1. 为每个真实服务器的每个虚拟 IP 地址创建 ARP 表条目（均衡器使用 real_ip 作为联络真实服务器的 IP，通常此 IP 会绑定到 eth0）：

```
arptables -A IN -d <virtual_ip> -j DROP
```

```
arptables -A OUT -s <virtual_ip> -j mangle --mangle-ip-s <real_ip>
```

这会导致真实服务器忽略所有来自虚拟 IP 地址的 ARP 请求，并修改可能包含虚拟 IP 的外发 ARP 响应，以便其包含服务器的真正 IP。Piranha 唯一应该回应 ARP 请求的节点应该为目前活跃的 LVS 节点。

2. 为每个真实服务器完成此操作后，在每个真实服务器中输入以下命令保存 ARP 表条目：

   service arptables_jf save

   chkconfig --level 2345 arptables_jf on

   chkconfig 命令将导致系统在重新引导时重新载入 arptables 配置 — 在启动网络之前。

3. 在所有真实服务器中使用 ifconfig 命令配置虚拟 IP 地址来生成 IP 别名。例如：

   ```
   # ifconfig eth0:1 192.168.76.24 netmask 255.255.252.0 broadcast 192.168.79.255 up
   ```

   或者用利用 ip 命令的 iproute2，例如：

   ```
   # ip addr add 192.168.76.24 dev eth0
   ```

   如前所述，使用红帽系统配置工具无法将虚拟 IP 地址配置为在引导时启动。一种解决方法就是将这些命令放在 /etc/rc.d/rc.local 文件中。

4. Configure Piranha for Direct Routing. Refer to 第 4 章 用 Piranha Configuration Tool配置 LVS 路由器 for more information.

3.2.2. 直接路由及 iptables

您可能还会通过创建 iptables 防火墙规则使用直接路由方法处理 ARP 事件。要使用 iptables 配置直接路由，您必须添加可生成透明代理服务器的规则，以便真实服务器可在系统中并不存在的 VIP 地址的情况下还可将数据包发送到 VIP 地址。

iptables 方法是比 arptables_jf 更简单的配置方法。此方法还可完全绕过 LVS ARP 事件，因为虚拟 IP 地址只存在于活跃的 LVS 负载均衡器（LVS director）中。

但是与 arptables_jf 相比，使用 iptables 方法有一些性能上的问题，因为每次在转发/伪装数据包时都会超载。

您还无法重新利用使用 iptables 方法的端口。例如，无法将两个独立 Apache HTTP Server 服务绑定到端口 80，因为它们必须绑定到 INADDR_ANY 而不是虚拟 IP 地址。

要使用 iptables 方法配置直接路由，请执行以下步骤：

1. 在每个真实服务器中，为每个 VIP、端口和协议（TCP 或者 UDP）组合运行以下命令使其为真实服务器服务：

   iptables -t nat -A PREROUTING -p <tcp|udp> -d <vip> --dport <port> -j REDIRECT

   此命令可使真实服务器处理目的地址为 VIP 和给定端口的数据包。

2. 在每个真实服务器保存配置：

   ```
   ```

```
# service iptables save
# chkconfig --level 2345 iptables on
```

以上命令可使系统在引导时重新载入 iptables 配置 — 在启动网络前。

3.3. 将配置组合到一起

在决定使用以上哪种路由方法后，应该在网络中将硬件链接起来。

重点

必须将 LVS 路由器的适配器设备配置为可访问相同的网络。例如：如果 eth0 连接的是公共网络，eth1 连接的是专用网络，那么在备用 LVS 路由器中的相同的设备必须连接相同的网络。

还有，要将在引导时第一个出现的接口列出的网关添加到路由表中，之后忽略所有在其它接口中列出的网关。这在配置真实服务器时要特别重点考虑。

请在物理连接硬件后，配置主 LVS 路由器和备用 LVS 路由器中的网络接口。这可使用类似 system-config-network 的图形程序或者手动编辑网络脚本完成。有关使用 system-config-network 添加设备的详情请参考Red Hat Enterprise Linux 部署指南中网络配置一章。本章剩余的内容，即有关网络接口更换示例中所述的内容可通过手动编辑或者使用 Piranha Configuration Tool程序完成。

3.3.1. 通用 LVS 联网提示

在试图使用 Piranha Configuration Tool配置 LVS 前，请配置 LVS 路由器中公共和专用网络的真实 IP 地址。每个布局的这一部分都给出示例网络地址，但需要配置实际使用的网络地址。以下是一些使用网络接口或者检查其状态的命令。

使用真实联网接口

要使用网络接口，请以根用户身份使用以下命令，其中使用接口对应的数字替换 N（eth0 和 eth1）。

```
/sbin/ifup ethN
```

警告

Do not use the ifup scripts to bring up any floating IP addresses you may configure using Piranha Configuration Tool (eth0:1 or eth1:1). Use the service command to start pulse instead (see 第 4.8 节 “启动 LVS” for details).

关闭真实网络接口

要关闭某个真实网络接口，请以根用户身份使用以下命令，并用相关接口数替换 N（eth0 和 eth1）。

```
/sbin/ifdown ethN
```

查看网络接口状态

如果您需要在任意时间检查哪些接口是打开的，请输入以下命令：

/sbin/ifconfig

要浏览某台机器的路由表格，请使用以下命令：

/sbin/route

3.4. 多端口服务和 LVS

LVS routers under any topology require extra configuration when creating multi-port LVS services. Multi-port services can be created artificially by using firewall marks to bundle together different, but related protocols, such as HTTP (port 80) and HTTPS (port 443), or when LVS is used with true multi-port protocols, such as FTP. In either case, the LVS router uses firewall marks to recognize that packets destined for different ports, but bearing the same firewall mark, should be handled identically. Also, when combined with persistence, firewall marks ensure connections from the client machine are routed to the same host, as long as the connections occur within the length of time specified by the persistence parameter. For more on assigning persistence to a virtual server, see 第 4.6.1 节 “「虚拟服务器」子界面”.

遗憾的是，用来平衡真实服务器中负载的机制 — IPVS — 可以识别为数据包分配的防火墙标记，但无法自己分配防火墙标记。分配防火墙标记的工作必须由网络数据包过滤器 iptables 在 Piranha Configuration Tool之外执行。

3.4.1. 分配防火墙标记

要为目的地址为特定端口的数据包分配防火墙标记，管理员必须使用 iptables。

This section illustrates how to bundle HTTP and HTTPS as an example; however, FTP is another commonly clustered multi-port protocol. If an LVS is used for FTP services, refer to 第 3.5 节 “配置 FTP” for configuration details.

在使用防火墙标记时要记住的最基本规则就是在 Piranha Configuration Tool使用的每一个防火墙标记都必须有一个相应的 iptables 规则来将标记分配给网络数据包。

在创建网络数据包过滤器规则前，请确定没有规则在运行。要做到这一点，请在 shell 提示符后以根用户身份登录，并输入：

/sbin/service iptables status

如果没有运行 iptables，提示符会马上重新出现。

如果激活了 iptables，它会显示一组规则。如果显示了规则，请输入以下命令：

/sbin/service iptables stop

如果正在运行的规则很重要，请检查 /etc/sysconfig/iptables 中的内容并在操作前将有保留价值的规则复制到一个安全的地方。

以下是分配了相同防火墙标记 80 的规则，它们在端口 80 和 443 接收目的地址为浮动 IP 地址 n.n.n.n 的进入流量。

/sbin/modprobe ip_tables

/sbin/iptables -t mangle -A PREROUTING -p tcp -d n.n.n.n/32 --dport 80 -j MARK --set-mark 80

/sbin/iptables -t mangle-A PREROUTING -p tcp -d n.n.n.n/32 --dport 443 -j MARK --set-mark 80

For instructions on assigning the VIP to the public network interface, see 第 4.6.1 节 “「虚拟服务器」子界面”. Also note that you must log in as root and load the module for iptables before issuing rules for the first time.

在以上的 iptables 中，应该使用虚拟服务器的浮动 IP 替换您的 HTTP 和 HTTPS n.n.n.n。这些命令具有为在防火墙标记为 80 的适当端口将所有流量分配到 VIP 地址的网络效应，这些流量可依次由 IPVS 识别并进行正确转发。

警告

The commands above will take effect immediately, but do not persist through a reboot of the system. To ensure network packet filter settings are restored upon reboot, refer to 第 3.6 节 “保存网络数据包过滤设置”

3.5. 配置 FTP

文件传输协议（FTP）是一个古老而且复杂的多端口协议，会为 LVS 环境带来很多复杂的情况。要了解这些情况的实质，您必须首先了解有关 FTP 网络的一些关键问题。

3.5.1. FTP 是如何工作的?

和大多数其它服务器客户端关系一样，客户端会在某个特定端口打开一个到服务器的连接，然后服务器在那个端口对客户端进行响应。当 FTP 客户端连接到 FTP 服务器时，它会打开一个到控制端口 21 的连接。然后客户端会告知 FTP 服务器是建立主动连接还是被动连接。由客户端选择的连接类型决定服务器如何进行响应以及在什么端口进行传输。

数据连接有两种类型:

主动连接
: 当建立了主动连接，服务器会在客户端机器的端口 20 或者更高的端口打开一个到客户端的数据连接。所有来自服务器的数据都会通过此连接传输。

被动连接
: 当建立了被动连接时，客户端会要求服务器在高于 10,000 的端口中建立一个被动连接端口。接着服务器会为此次会话绑定到此高数值端口，并将此端口号转交会客户端。客户端的每个数据请求都会形成一个独立的数据连接。最先进的 FTP 客户端会在服务器发出数据请求时试图建立被动连接。

注记

客户端决定连接的类型，不是服务器。这就是说对于有效的群集 FTP 来说，您必须将 LVS 路由器配置为既可处理主动连接也可处理被动连接。

FTP 客户端/服务器关系有可能打开大量 Piranha Configuration Tool和 IPVS 都不了解的端口。

3.5.2. 这对 LVS 路由有什么影响?

IPVS 数据包转发只允许在识别其端口号或者防火墙标记的基础上接入或者接出。如果群集之外的客户端试图打开一个 IPVS 无法处理的端口，连接就会断开。同样，如果真实服务器试图在某个 IPVS 不了

解的端口打开返回互联网的连接，连接也会断开。这就是说所有来自互联网的 FTP 连接必须分配了同一个防火墙标记，而且所有来自 FTP 服务器的连接都必须使用网络数据包过滤规则进行了正确转发。

3.5.3. 创建网络数据包过滤规则

Before assigning any iptables rules for FTP service, review the information in 第 3.4.1 节 “分配防火墙标记” concerning multi-port services and techniques for checking the existing network packet filtering rules.

Below are rules which assign the same firewall mark, 21, to FTP traffic. For these rules to work properly, you must also use the VIRTUAL SERVER subsection of Piranha Configuration Tool to configure a virtual server for port 21 with a value of 21 in the Firewall Mark field. See 第 4.6.1 节 “「虚拟服务器」子界面” for details.

3.5.3.1. 主动连接规则

主动连接的规则告知内核接受并转发在端口 20（FTP 数据端口）中进入内部浮动 IP 地址的连接。

以下 iptables 命令允许 LVS 路由器接受 IPVS 不了解的真实服务器的外发连接。

```
/sbin/iptables -t nat -A POSTROUTING -p tcp -s n.n.n.0/24 --sport 20 -j MASQUERADE
```

In the iptables command, n.n.n should be replaced with the first three values for the floating IP for the NAT interface's internal network interface defined in the GLOBAL SETTINGS panel of Piranha Configuration Tool.

3.5.3.2. 被动连接的规则

被动连接规则为来自互联网到浮动 IP 的连接分配适当的防火墙标记， 这些标记为端口范围 — 10,000 到 20,000 的服务。

警告

如果您要为被动连接限制端口范围，您必须还要配置 VSFTP 服务器使用一个观察端口范围。在 /etc/vsftpd.conf 文件中添加以下行即可达到此目的：

```
pasv_min_port=10000

pasv_max_port=20000
```

您还必须控制服务器为被动 FTP 连接向客户端显示的地址。在使用 NAT 进行路由的 LVS 系统中，在 /etc/vsftpd.conf 文件中添加以下行来覆盖连接到 VIP 的真实服务器 IP 地址，该地址就是可用的可以在连接中看到的地址。例如：

```
pasv_address=n.n.n.n
```

使用 LVS 系统的 VIP 地址替换 n.n.n.n。

要配置其它 FTP 服务器，请参考有关文档。

范围的幅度应该适用与大多数情况，但您可修改命令中的 10000:20000，将其增加到包含所有可用的不安全端口，以下为 1024:65535。

下面的 iptables 命令有可将任意地址为浮动 IP 的流量分配给防火墙标记为 21 的适当端口的网络效应，然后这些地址可由 IPVS 识别，并正确转发：

```
/sbin/iptables -t mangle -A PREROUTING -p tcp -d n.n.n.n/32 --dport 21 -j MARK --set-mark 21
```

```
/sbin/iptables -t mangle -A PREROUTING -p tcp -d n.n.n.n/32 --dport 10000:20000 -j MARK --set-mark 21
```

在 iptables 命令中，可使用在 Piranha Configuration Tool的「虚拟服务器」子界面中定义的 FTP 虚拟服务器浮动 IP 地址替换 n.n.n.n。

警告

The commands above take effect immediately, but do not persist through a reboot of the system. To ensure network packet filter settings are restored after a reboot, see 第 3.6 节 “保存网络数据包过滤设置”

Finally, you need to be sure that the appropriate service is set to activate on the proper runlevels. For more on this, refer to 第 2.1 节 “在 LVS 路由器中配置服务”.

3.6. 保存网络数据包过滤设置

在为您的系统配置了正确的网络数据包过滤器之后，请保存设置以便在重启后可重新载入。对于 iptables，请输入以下命令：

```
/sbin/service iptables save
```

这可将这些设置保存到 /etc/sysconfig/iptables 文件中以便在引导时重新调用。

Once this file is written, you are able to use the /sbin/service command to start, stop, and check the status (using the status switch) of iptables. The /sbin/service will automatically load the appropriate module for you. For an example of how to use the /sbin/service command, see 第 2.3 节 “启动 Piranha Configuration Tool服务”.

Finally, you need to be sure the appropriate service is set to activate on the proper runlevels. For more on this, see 第 2.1 节 “在 LVS 路由器中配置服务”.

下面的一章将论述如何使用 Piranha Configuration Tool 配置 LVS 路由器，并描述了激活 LVS 路由器所需的步骤。

用 Piranha Configuration Tool配置 LVS 路由器

Piranha Configuration Tool提供结构方法来为 LVS 创建所需配置文件 — /etc/sysconfig/ha/lvs.cf。本章论述了 Piranha Configuration Tool的基本操作，以及如何在完成配置后激活群集。

重点

LVS 的配置文件有很严格的格式规则。使用 Piranha Configuration Tool是最好的预防在 lvs.cf 中出现语法错误的方法，并可因此防止软件失败。

4.1. 必需的软件

必须要在主 LVS 路由器中运行 piranha-gui 服务才可使用 Piranha Configuration Tool。要配置 LVS，您至少需要一个只显示文本的网页浏览器，比如 links。如果您从其它机器访问 LVS 路由器，您还需要作为根用户使用 ssh 连接到主 LVS 路由器。

当配置主 LVS 路由器时，最好在终端窗口中保持共存的 ssh 连接。该连接提供了一个重启 pulse 和其它服务、配置网络数据包过滤以及在故障排除时监控 /var/log/messages 文件的安全方法。

以下的四个部分将分别对 Piranha Configuration Tool的配置页进行说明，并给出使用此工具设置 LVS 的具体操作。

4.2. 登录到 Piranha Configuration Tool

When configuring LVS, you should always begin by configuring the primary router with the Piranha Configuration Tool. To do this,verify that the piranha-gui service is running and an administrative password has been set, as described in 第 2.2 节 “为 Piranha Configuration Tool设置密码”.

If you are accessing the machine locally, you can open http://localhost:3636 in a Web browser to access the Piranha Configuration Tool. Otherwise, type in the hostname or real IP address for the server followed by :3636. Once the browser connects, you will see the screen shown in 图 4.1 “The Welcome Panel”.

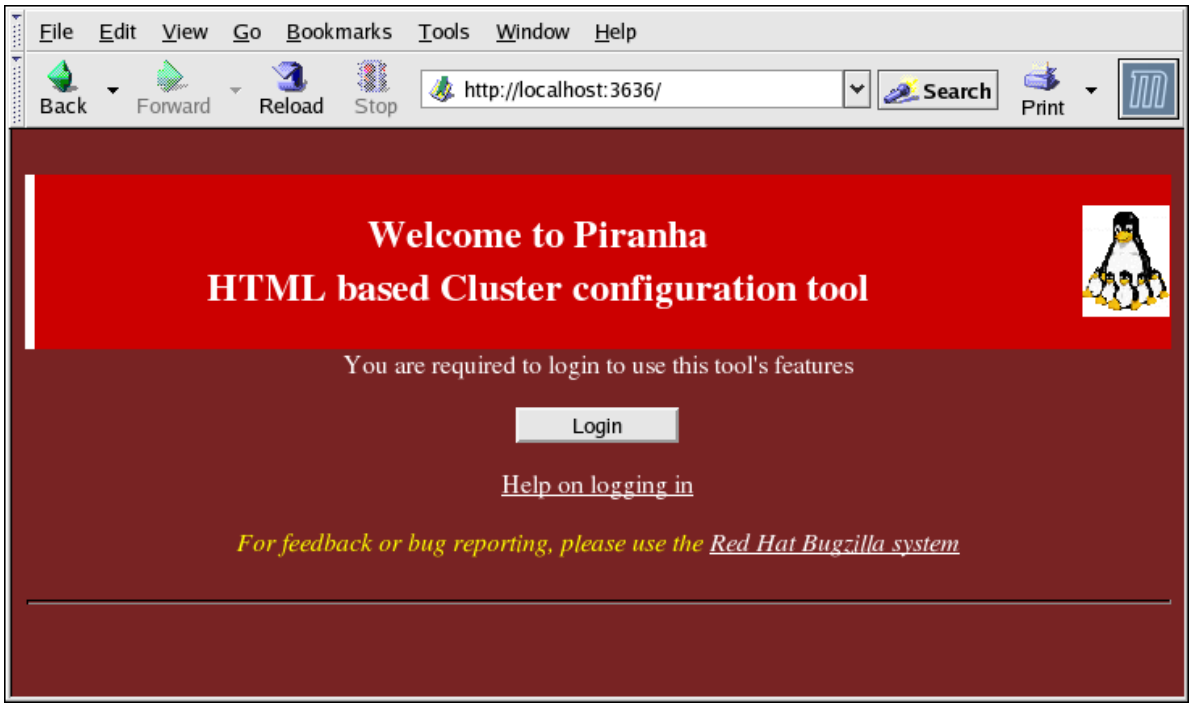

图 4.1. The Welcome Panel

点击「登录」按钮，并在「用户名」字段输入 piranha，在「密码」字段输入您生成的管理密码。

Piranha Configuration Tool由四个主要界面或者面板组成。另外，「虚拟服务器」面板包括四个子界面。「控制/监控（CONTROL/MONITORING」面板是登录屏幕之后出现的第一个界面。

4.3. CONTROL/MONITORING

「控制/监控」面板列出了 LVS 的受限的运行时间状态。它显示了 pulse 守护进程、LVS 路由表和 LVS 生成的 nanny 进程的状态。

注记

The fields for CURRENT LVS ROUTING TABLE and CURRENT LVS PROCESSES remain blank until you actually start LVS, as shown in 第 4.8 节 “启动 LVS”.

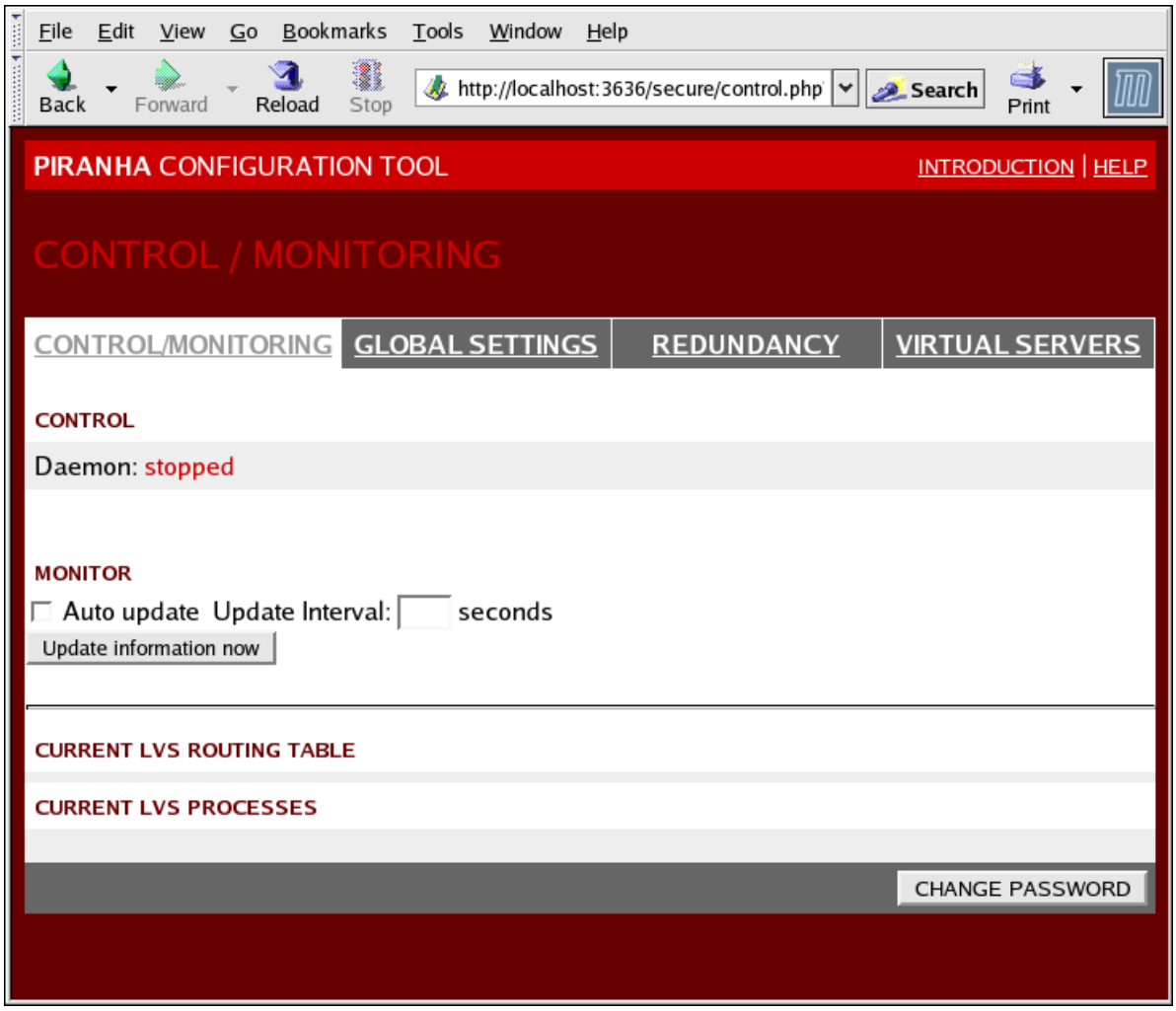

图 4.2. The CONTROL/MONITORING Panel

Auto update
本页面显示的状态可在用户可配置的界面中自动更新。要启用此性能，请点击「自动更新」复选框，并在「更新频率（Update frequency in seconds）」文本栏中设定所需更新频率（默认值为 10 秒）。

我们不建议将自动更新的时间间隔设置为小于 10 秒。这样使重新配置「自动更新」时间间隔变得困难，因为页面会过于频繁地更新。如果您遇到这个问题，只要点击另外一个面板即可返回「控制/监控」面板。

「自动更新」性能不适用于所有浏览器，比如 Mozilla。

Update information now
您可以点击此按钮手动更新状态信息。

CHANGE PASSWORD
点击这个按钮会使您进入一个帮助屏幕，上面有如何修改 Piranha Configuration Tool管理密码的信息。

4.4. GLOBAL SETTINGS

The GLOBAL SETTINGS panel is where the you define the networking details for the primary LVS router's public and private network interfaces.

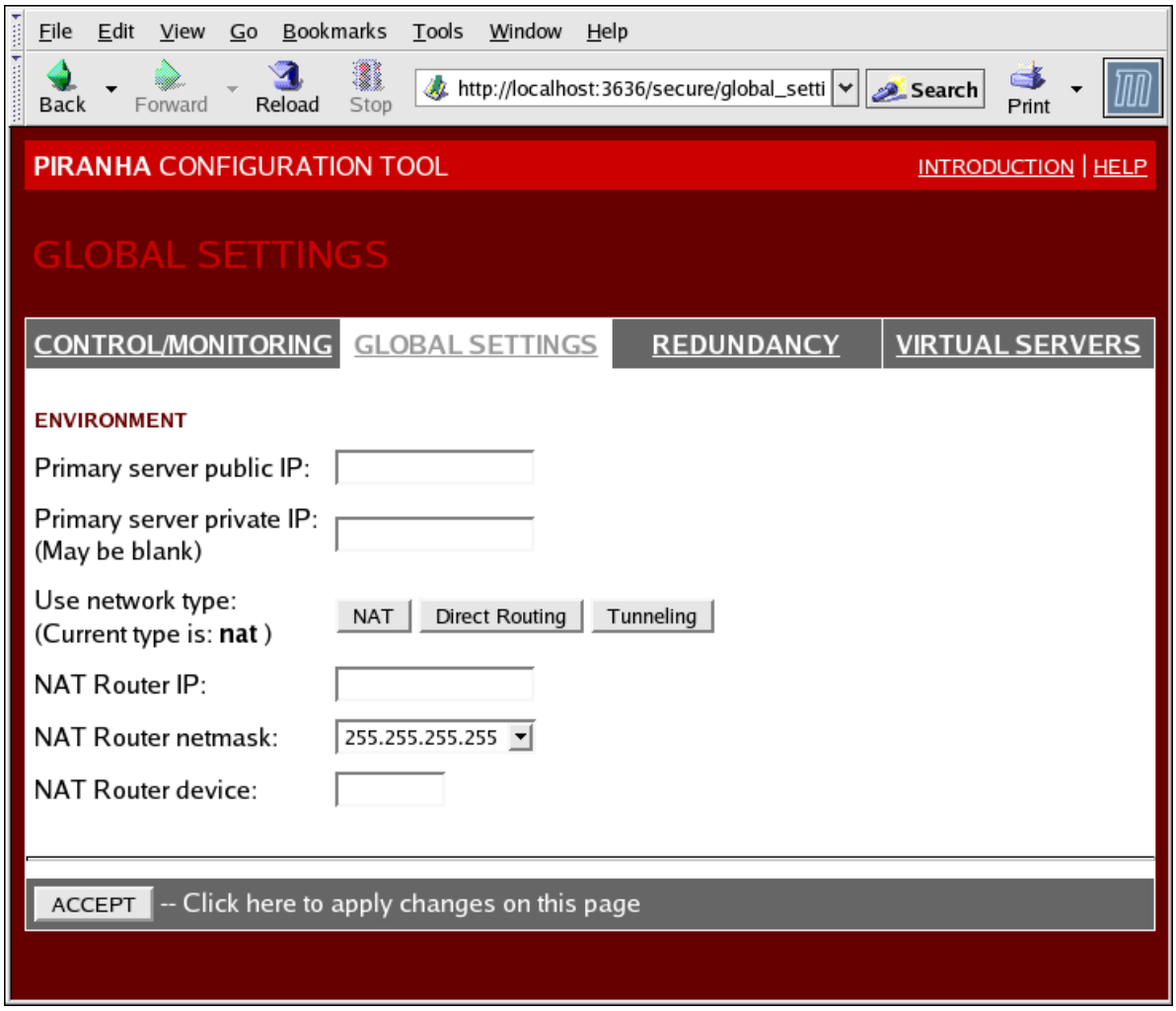

图 4.3. The GLOBAL SETTINGS Panel

The top half of this panel sets up the primary LVS router's public and private network interfaces. These are the interfaces already configured in 第 3.1.1 节 “为带 NAT 的 LVS 配置网络接口”.

Primary server public IP
在此字段，为主 LVS 节点输入可公开路由的真实 IP 地址。

Primary server private IP
Enter the real IP address for an alternative network interface on the primary LVS node. This address is used solely as an alternative heartbeat channel for the backup router and does not have to correlate to the real private IP address assigned in 第 3.1.1 节 “为带 NAT 的 LVS 配置网络接口”. You may leave this field blank, but doing so will mean there is no alternate heartbeat channel for the backup LVS router to use and therefore will create a single point of failure.

注记

「直接路由」配置不需要专用 IP 地址，因为所有真实服务器 以及 LVS 主控服务器共享相同的虚拟 IP 地址，并应该有相同的 IP 路由配置。

注记

The primary LVS router's private IP can be configured on any interface that accepts TCP/IP, whether it be an Ethernet adapter or a serial port.

Use network type
: 点击「NAT」按钮选择 NAT 路由。

  点击「直接路由」按钮选择直接路由。

The next three fields deal specifically with the NAT router's virtual network interface connecting the private network with the real servers. These fields do not apply to the direct routing network type.

NAT Router IP
: 在此文本字段输入专用浮动 IP，该浮动 IP 应该作为真实服务器的网关使用。

NAT Router netmask
: If the NAT router's floating IP needs a particular netmask, select it from drop-down list.

NAT Router device
: 使用此文本字段为浮动 IP 地址的网络接口定义设备名称，比如 eth1:1。

注记

您应该为连接到专用网络的以太网接口定义 NAT 浮动 IP 地址别名。在本示例中，专用网络为 eth1 接口，因此 eth1:1 就是浮动 IP 地址。

警告

完成此页面设置后，点击「接受」按钮确定您没有在选择新的面板时丢失任何修改的数据。

4.5. REDUNDANCY

「冗余」面板允许您配置备用 LVS 路由器节点并设置不同的 heartbeat 监控选项。

注记

The first time you visit this screen, it displays an "inactive" Backup status and an ENABLE button. To configure the backup LVS router, click on the ENABLE button so that the screen matches 图 4.4 “The REDUNDANCY Pane1”.

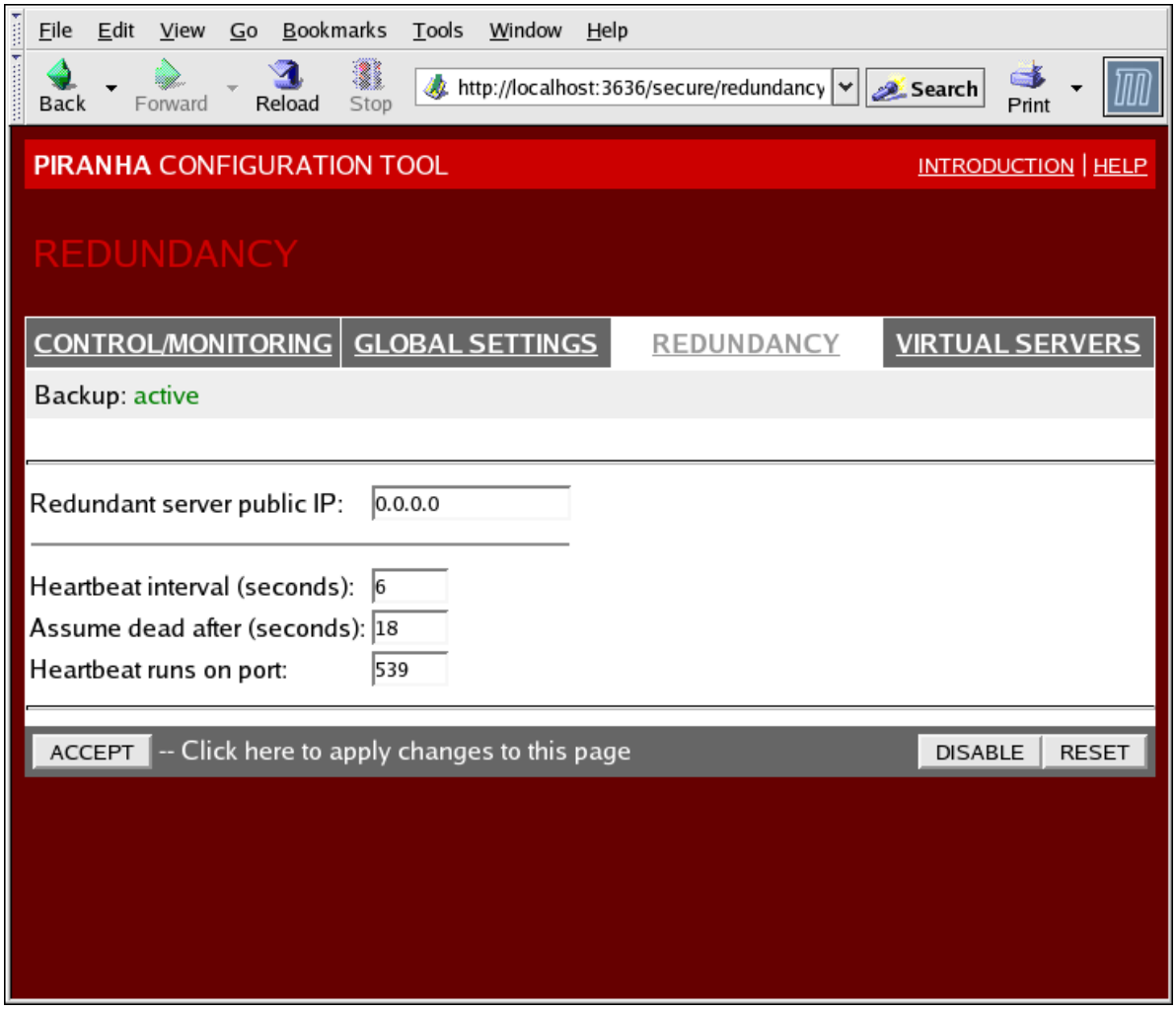

图 4.4. The REDUNDANCY Pane1

Redundant server public IP
为备用 LVS 路由器节点输入公共真实 IP 地址。

Redundant server private IP
Enter the backup node's private real IP address in this text field.

如果您没有看到「冗余服务器专用 IP」，请返回「全局设置」面板并输入「主服务器专用 IP」地址并点击「接受」。

面板的其余部分为配置 heartbeat 频道，该频道是用来让备用节点监控主节点的错误。

Heartbeat Interval (seconds)
此字段设置 heartbeat 间隔的时间（秒）— 即备用节点检查主 LVS 节点功能状态的间隔时间。

Assume dead after (seconds)
: 如果主 LVS 节点在这段时间内没有响应，备用 LVS 路由器节点将启动失效切换。

Heartbeat runs on port
: 此字段设定 heartbeat 和主 LVS 节点进行沟通所使用的端口。如果该字段是空白，即使用默认端口 539。

警告

请记住在修改此面板中的内容后点击「接受」按钮，以确定在选择新面板时没有丢失您所做的修改。

4.6. VIRTUAL SERVERS

「虚拟服务器」面板显示每个当前定义的虚拟服务器的信息。表格里的每个条目都显示了虚拟服务器的状态、服务器名称、分配的 IP 地址、虚拟 IP 的掩码、服务端口、使用的协议以及虚拟设备接口。

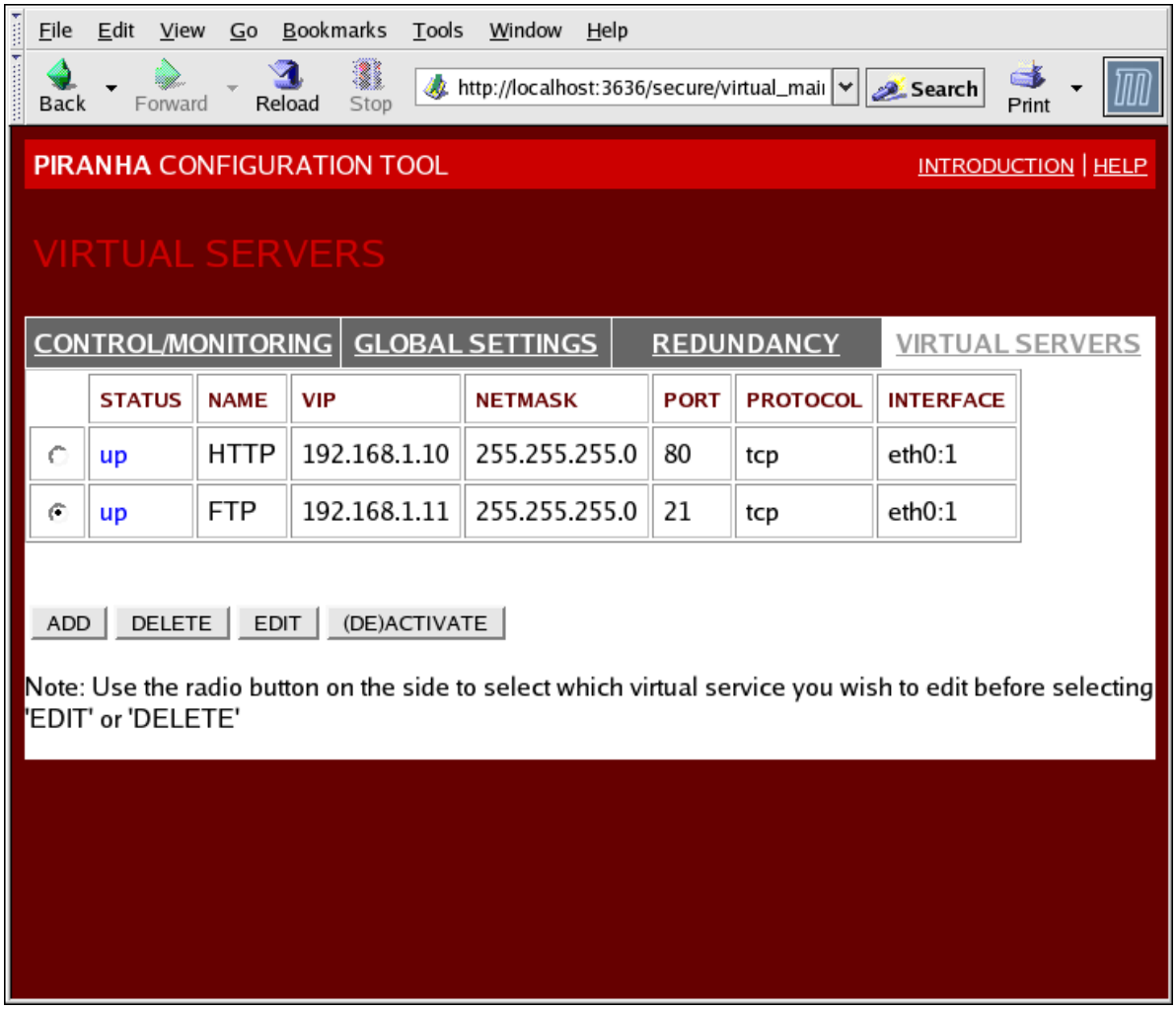

图 4.5. The VIRTUAL SERVERS Panel

「虚拟服务器」面板中显示的每个服务器都可以在接下来的屏幕或子界面（subsections）中进行配置。

点击「添加」按钮可以添加一个服务。要删除某个服务，可以选中虚拟服务器旁的单选按钮并按「删除」按钮。

要启用或禁用列表中的虚拟服务器，选中单选框并点击「激活/取消激活」按钮。

添加了虚拟服务器后，您可以选中其左边的单选按钮并点击「编辑」按钮进入「虚拟服务器」子界面来进行配置。

4.6.1. 「虚拟服务器」子界面

The VIRTUAL SERVER subsection panel shown in 图 4.6 “The VIRTUAL SERVERS Subsection” allows you to configure an individual virtual server. Links to subsections related specifically to this virtual server are located along the top of the page. But before configuring any of the subsections related to this virtual server, complete this page and click on the ACCEPT button.

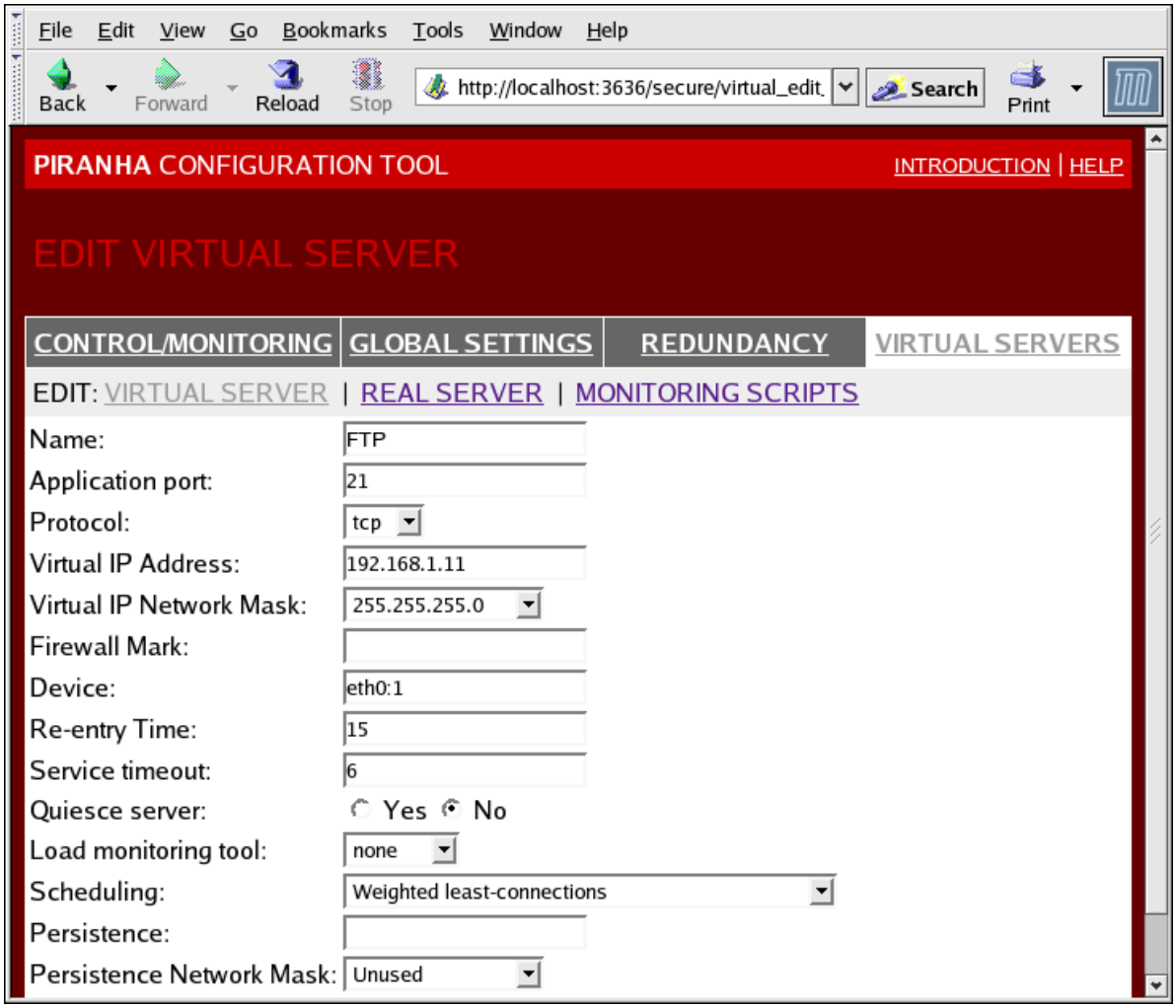

图 4.6. The VIRTUAL SERVERS Subsection

Name

输入描述名称来确定虚拟服务器。这个名称不是机器的主机名，因此可使用具有描述性并容易设别的名称。您甚至可以参考虚拟服务器使用的协议，比如 HTTP。

Application port

请输入用来侦听服务应用程序的端口号。因为本示例中要侦听的是 HTTP 服务，因此使用端口 80。

Protocol

在下拉菜单中选择 UDP 或者 TCP。网页服务器一般使用 TCP 协议进行沟通，如上例所示。

Virtual IP Address

Enter the virtual server's floating IP address in this text field.

Virtual IP Network Mask

使用下拉菜单为此虚拟服务器设定子网掩码。

Firewall Mark

不要在此字段输入防火墙标记整数值，除非您正在捆绑多个端口协议或者为独立但关联的协议创建多端口虚拟服务器。在本示例中，以上虚拟服务器的「防火墙标记」为 80，因为我们正在使用防火墙标记值 80 在端口 80 将连接捆绑至 HTTP，在端口 443 将连接捆绑至 HTTPS。当与持久性合并使用时，该技术可确保将访问不安全或者安全网页的用户路由到同一个真实服务器，并保持此状态。

警告

Entering a firewall mark in this field allows IPVS to recognize that packets bearing this firewall mark are treated the same, but you must perform further configuration outside of the Piranha Configuration Tool to actually assign the firewall marks. See 第 3.4 节 “多端口服务和 LVS ” for instructions on creating multi-port services and 第 3.5 节 “配置 FTP” for creating a highly available FTP virtual server.

Device

输入在「虚拟 IP 地址」字段定义您希望使用浮动 IP 地址的网络设备名称。

您应该将公共浮动 IP 地址命名为连接到公共网络的以太网接口的别名。在本示例中，公共网络位于 eth0 接口，因此设备名称应为 eth0:1。

Re-entry Time

输入一个整数值来定义时间的长度（以秒计），即在激活的 LVS 路由器试图在失败后将一个真实服务器带回服务器池的时间。

Service Timeout

输入一个整数值，用该数值定义在认为真实服务器死亡并将其从服务器池中删除的时间长度（以秒计）。

Quiesce server

当「静默服务器」单选按钮被选中后，每次有新的真实服务器节点上线时，最少连接表被重置为 0，这样活跃 LVS 路由器就会象所有的真实服务器都是刚加入群集一样路由请求。这个选项可以避免当大量的连接进入服务器池时，新加入的服务器超载。

Load monitoring tool

LVS 路由器可以用 rup 或 ruptime 监控不同服务器上的负载。如果您从下拉菜单里选择了 rup，那么每个服务器都必须运行 rstatd 服务。如果您选择了 ruptime，每个服务器都必须运行 rwhod 服务。

警告

负载监控和负载平衡不同，它可导致在与加权调度算法合并使用时难于预测预定的行为。同时，如果您使用负载监控，真实服务器必须是 Linux 机器。

Scheduling

Select your preferred scheduling algorithm from the drop-down menu. The default is Weighted least-connection. For more information on scheduling algorithms, see 第 1.3.1 节 “调度算法”.

Persistence

如果管理员在客户端传送的过程中需要持续连接到虚拟服务器，请在此文本字段输入在连接超时前允许的非激活状态的秒数。

重点

If you entered a value in the Firewall Mark field above, you should enter a value for persistence as well. Also, be sure that if you use firewall marks and persistence together, that the amount of persistence is the same for each virtual server with the firewall mark. For more on persistence and firewall marks, refer to 第 1.5 节 “持久性和防火墙标记”.

Persistence Network Mask

要把持久性限制到特定的子网，您可以从下拉菜单里选择合适的网络掩码。

注记

在出现防火墙标记之前，由子网限制的持久性是捆绑连接的原始方法。现在，最好在和防火墙标记的关联中使用持久性来达到同样的结果。

警告

请记住在此面板中进行任何修改后点击「接受」按钮以确定在选择一个新面板时没有丢失所做的修改。

4.6.2. 「真实服务器」子界面

点击面板顶部的「真实服务器」子界面链接将显示「编辑真实服务器」子界面。

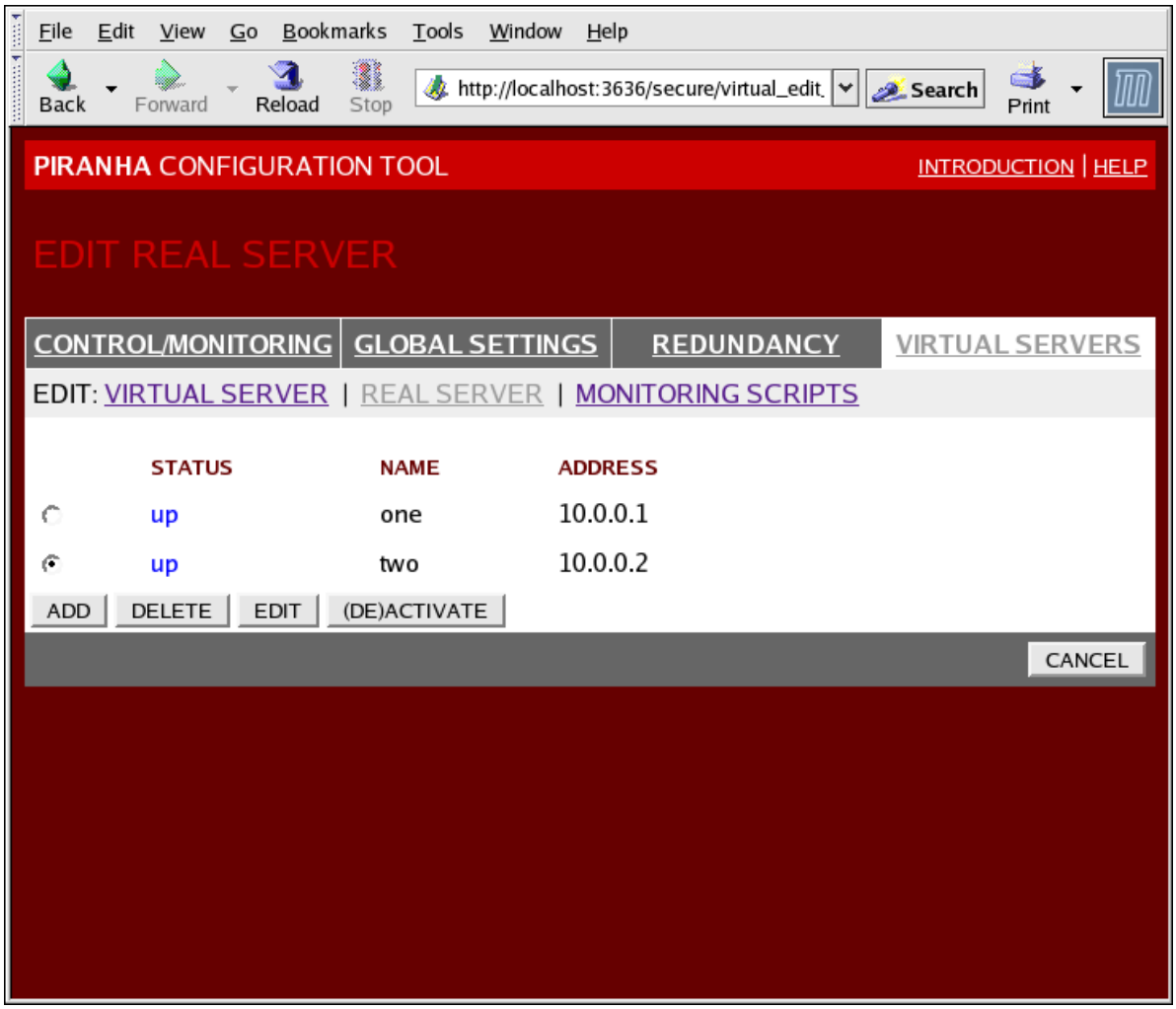

图 4.7. The REAL SERVER Subsection

Click the ADD button to add a new server. To delete an existing server, select the radio button beside it and click the DELETE button. Click the EDIT button to load the EDIT REAL SERVER panel, as seen in 图 4.8 “The REAL SERVER Configuration Panel” .

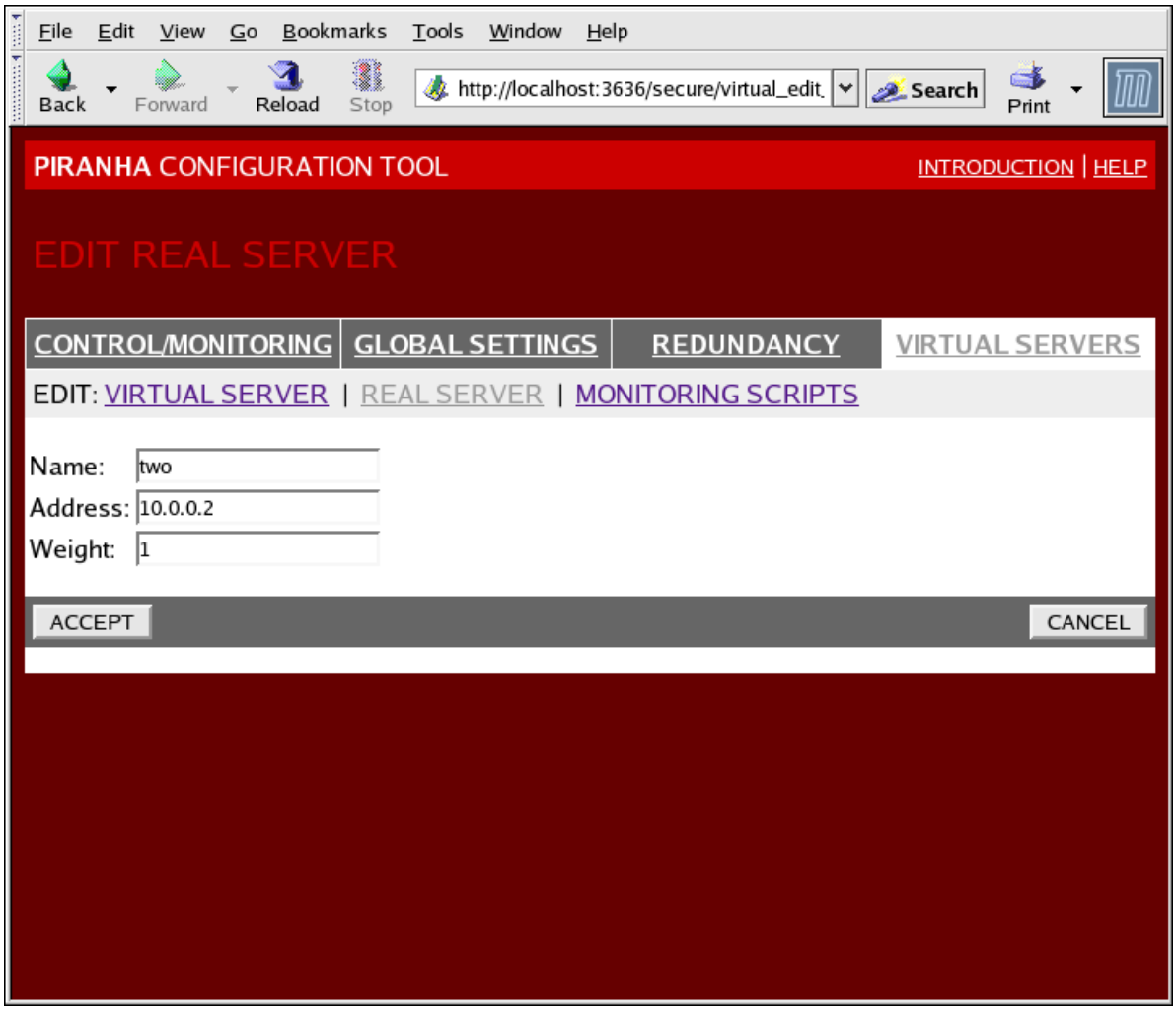

图 4.8. The REAL SERVER Configuration Panel

这个面板由 3 个字段组成:

Name

真实服务器的描述性名称。

注记

这个名称不是机器的主机名，它应该具有描述性且易于识别。

Address

The real server's IP address. Since the listening port is already specified for the associated virtual server, do not add a port number.

Weight

An integer value indicating this host's capacity relative to that of other hosts in the pool. The value can be arbitrary, but treat it as a ratio in relation to other real servers in the pool. For more on server weight, see 第 1.3.2 节 “服务器加权和调度”.

警告

请记住在修改此面板中的内容后点击「接受」按钮，以确定在选择新面板时没有丢失您所做的修改。

4.6.3. EDIT MONITORING SCRIPTS Subsection

点击页面顶部的「监控脚本」链接。「编辑监控脚本」子界面允许管理员指定一个 send/expect 字符串序列来验证虚拟服务器服务在每个真实服务器上是否正常运行。管理员也可以在这里定义检查需要动态更新数据服务的自定义脚本。

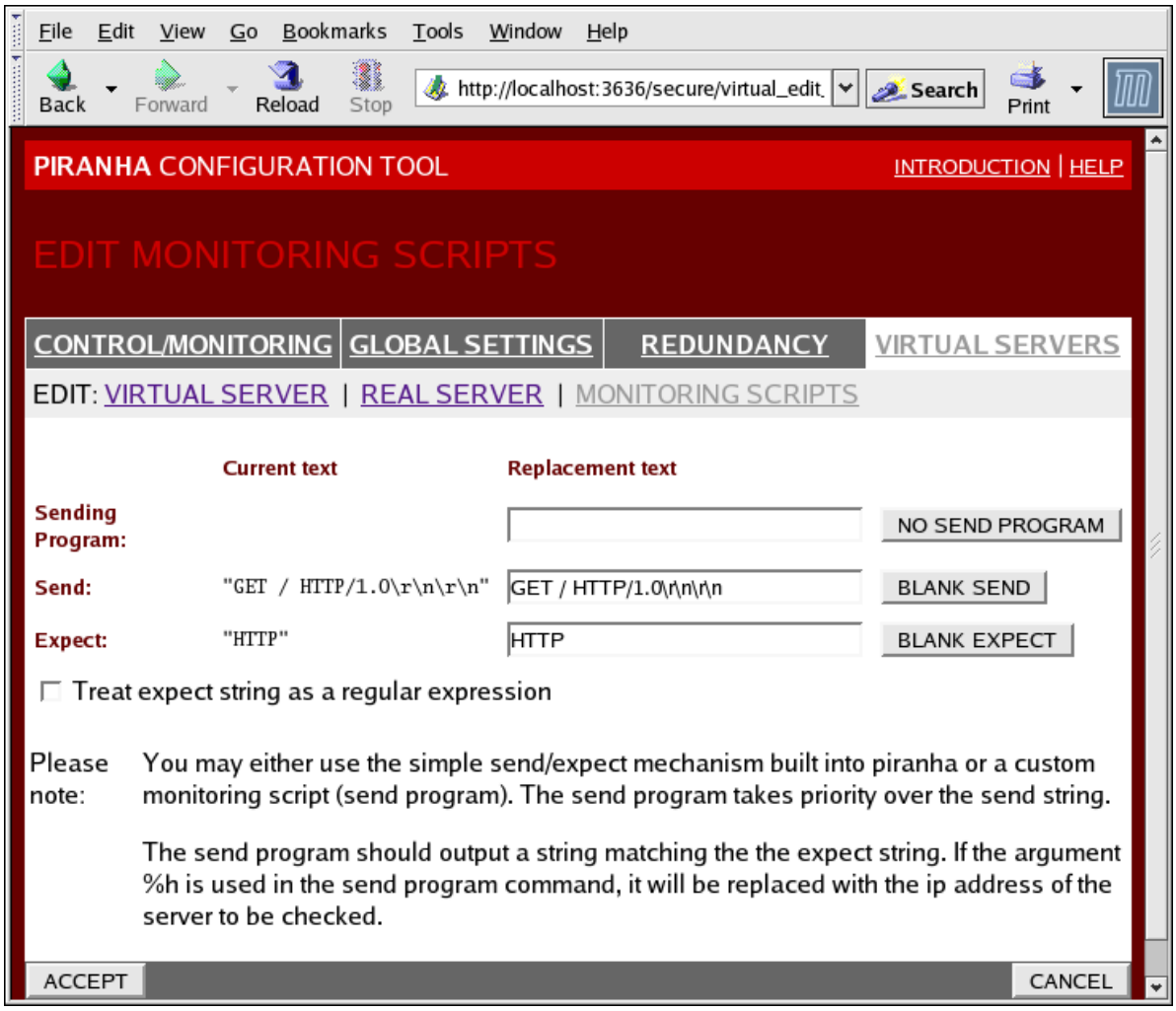

图 4.9. The EDIT MONITORING SCRIPTS Subsection

Sending Program

有关更多高级服务确认，您可以使用此字段来指定服务检查脚本的路径。此功能对那些需要动态修改数据的服务特别有帮助，比如 HTTPS 或者 SSL。

要使用此功能，您必须写出可返回文本响应的脚本，然后将其发送到可执行程序，并在「发送程序」字段输入该路径。

注记

To ensure that each server in the real server pool is checked, use the special token %h after the path to the script in the Sending Program field. This token is replaced with each real server's IP address as the script is called by the nanny daemon.

以下是在制作外置服务检查脚本时可作为指南使用的脚本样本：

```
#!/bin/sh

TEST=`dig -t soa example.com @$1 | grep -c dns.example.com

if [ $TEST != "1" ]; then
 echo "OK
else
 echo "FAIL"
fi
```

注记

如果在「发送程序」字段里输入了外部程序，「发送」字段将被忽略。

Send

在此字段为 nanny 守护进程输入字符串来将其发送到所有真实服务器。默认情况下此字段用于 HTTP。您可根据您的需要修改此字段值。如果您保留此字段空白，nanny 守护进程会试图打开端口并在端口成功开启时假设服务正在运行。

这个字段里只允许输入一个发送序列，且它只能包含可打印的、ASCII 字符串以及下面的转义符：

- \n 代表新一行。
- \r 代表回车。
- \t 代表制表符。
- \ 代表将下一个字符转义。

Expect

输入如果正常工作时服务器应该返回的文本响应。如果您自己写发送程序，请输入该程序成功时应该给出的响应。

注记

要确定为给定服务发送什么，您可以打开真实服务器中连接 telnet 的端口，并看返回了什么。例如，FTP 在端口 220 报告连接，那么应该在「Send」字段输入 quit，并在「Expect」字段输入 220。

警告

请记住在修改此面板中的内容后点击「接受」按钮，以确定在选择新面板时没有丢失您所做的修改。

Once you have configured virtual servers using the Piranha Configuration Tool, you must copy specific configuration files to the backup LVS router. See 第 4.7 节 “同步配置文件” for details.

4.7. 同步配置文件

配置完主 LVS 路由器之后，在启动 LVS 前必须将一些配置文件复制到备用 LVS 路由器中。

这些文件包括：

- /etc/sysconfig/ha/lvs.cf — LVS 路由器配置文件。
- /etc/sysctl — 在内核中打开数据包转发功能的配置文件。
- /etc/sysconfig/iptables — 如果您使用防火墙标记，您应该根据您使用的网络数据包过滤器同步以上这些文件之一。

重点

在您使用 Piranha Configuration Tool 配置 LVS 时， /etc/sysctl.conf 和 /etc/sysconfig/iptables 文件不会改变。

4.7.1. 同步 lvs.cf

无论何时在创建或者更新 LVS 配置文件 /etc/sysconfig/ha/lvs.cf 时，您必须将其复制到备用 LVS 路由器节点。

警告

活跃和备用 LVS 路由器节点必须有相同的 lvs.cf 文件。如果两节点间 LVS 配置文件不匹配会导致无法进行失效切换。

进行此操作的最好方法是使用 scp 命令。

重点

To use scp the sshd must be running on the backup router, see 第 2.1 节 “在 LVS 路由器中配置服务” for details on how to properly configure the necessary services on the LVS routers.

在主 LVS 路由器中以根用户身份使用以下命令来同步路由器节点间的 lvs.cf 文件。

```
scp /etc/sysconfig/ha/lvs.cf n.n.n.n:/etc/sysconfig/ha/lvs.cf
```

在命令中，使用备用 LVS 路由器的真正 IP 地址替换 n.n.n.n。

4.7.2. 同步 sysctl

sysctl 文件在大多数情况下只修改一次。该文件在引导时读取并告知内核打开数据包转发功能。

重点

If you are not sure whether or not packet forwarding is enabled in the kernel, see 第 2.5 节 “启动数据包转发” for instructions on how to check and, if necessary, enable this key functionality.

4.7.3. 同步网络数据包过滤规则

如果您使用 iptables，您将会需要同步备用 LVS 路由器中的适当配置文件。

如果您更换了任何网络数据包过滤规则，请在主 LVS 路由器中以根用户身份输入以下命令：

```
scp /etc/sysconfig/iptables n.n.n.n:/etc/sysconfig/
```

在命令中，使用备用 LVS 路由器的真正 IP 地址替换 n.n.n.n。

接下来，您可以打开一个到备用路由器的 ssh 会话，也可以根用户身份登录到机器并输入以下命令：

```
/sbin/service iptables restart
```

Once you have copied these files over to the backup router and started the appropriate services (see 第 2.1 节 “在 LVS 路由器中配置服务” for more on this topic) you are ready to start LVS.

4.8. 启动 LVS

要启动 LVS，最好同时打开两个根终端，或者以根用户同时打开两个连接到主 LVS 路由器的 ssh 会话。

在一个终端中，用以下命令观察内核日志信息：

```
tail -f /var/log/messages
```

然后在另一个终端中输入以下命令启动 LVS：

```
/sbin/service pulse start
```

Follow the progress of the pulse service's startup in the terminal with the kernel log messages. When you see the following output, the pulse daemon has started properly:

```
gratuitous lvs arps finished
```

要停止观察 /var/log/messages，请按 Ctrl+c 键。

从这里开始，主 LVS 路由器也就成为活跃 LVS 路由器了。尽管您可以在此向 LVS 发出请求，但您还是应该在使用 LVS 进行服务前启动备用 LVS 路由器。要做到这一点，只要在备用 LVS 路由器节点重复以上所述过程即可。

完成最后的步骤后，将开启并运行 LVS。

附录 A. 使用带 Red Hat 的 LVS 群集

您可以使用带 Red Hat LVS 路由器的群集来部署高度可用的商业网站，以提供负载平衡、数据完整性和源程序的可用性。

The configuration in 图 A.1 “LVS with a Red Hat Cluster” represents an e-commerce site used for online merchandise ordering through a URL. Client requests to the URL pass through the firewall to the active LVS load-balancing router, which then forwards the requests to one of the Web servers. The Red Hat Cluster nodes serve dynamic data to the Web servers, which forward the data to the requesting client.

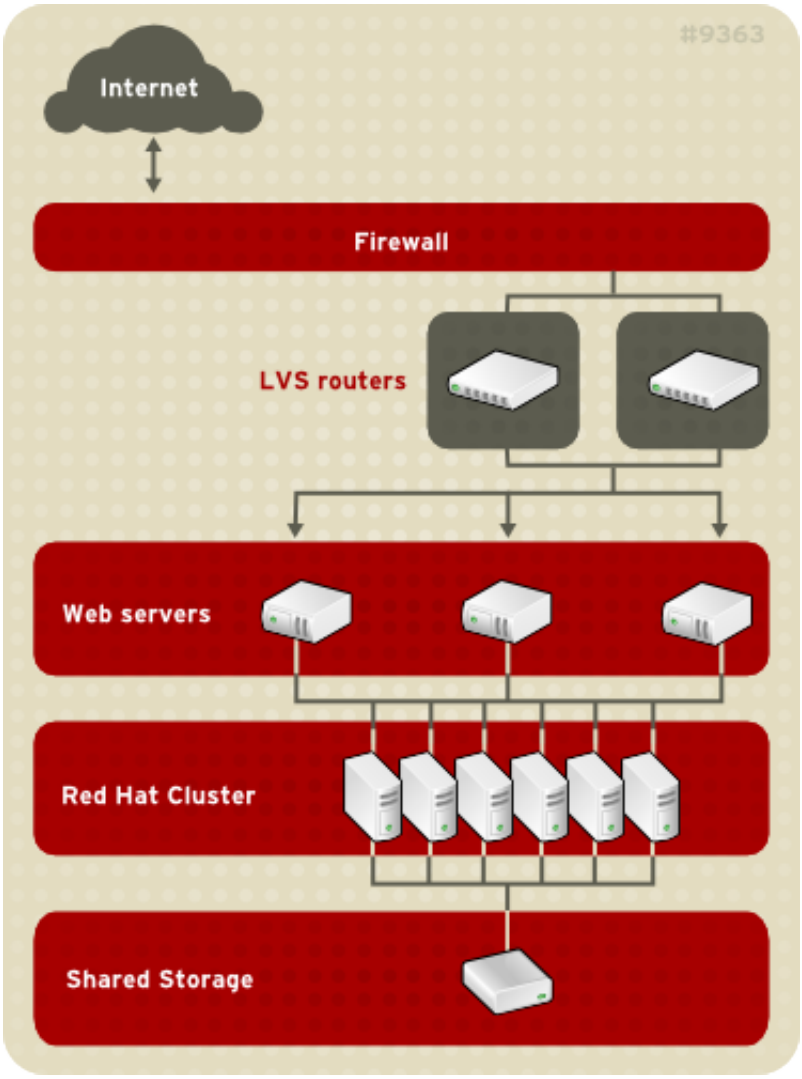

图 A.1. LVS with a Red Hat Cluster

Serving dynamic Web content with LVS requires a three-tier configuration (as shown in 图 A.1 “LVS with a Red Hat Cluster”). This combination of LVS and Red Hat Cluster allows for the configuration of a high-integrity, no-single-point-of-failure e-commerce site. The Red Hat Cluster can run a high-availability instance of a database or a set of databases that are network-accessible to the Web servers.

进行三层配置可提供动态内容。如果网页服务器只提供静态网页内容（包括少数不经常修改的数据），那么两层 LVS 配置就可以了，但如果需要提供动态网页内容，两层配置就不行了。动态内容可包括产品目录、购买订单或者客户数据库，这些内容必须在所有网页服务器中都是一致的，以确保客户可以访问最新的准确信息。

每层可提供以下功能:

- 第一层 — LVS 路由器执行负载平衡来分配网页请求。
- 第二层 — 为请求提供服务的一组网页服务器。
- 第三层 — 为网页服务器提供数据的 Red Hat 群集。

In an LVS configuration like the one in 图 A.1 “LVS with a Red Hat Cluster”, client systems issue requests on the World Wide Web. For security reasons, these requests enter a Web site through a firewall, which can be a Linux system serving in that capacity or a dedicated firewall device. For redundancy, you can configure firewall devices in a failover configuration. Behind the firewall are LVS load-balancing routers, which can be configured in an active-standby mode. The active load-balancing router forwards the requests to the set of Web servers.

Each Web server can independently process an HTTP request from a client and send the response back to the client. LVS enables you to expand a Web site's capacity by adding Web servers behind the LVS routers; the LVS routers perform load balancing across a wider set of Web servers. In addition, if a Web server fails, it can be removed; LVS continues to perform load balancing across a smaller set of Web servers.

附录 B. Revision History

修订 2.0-0 Mon Feb 08 2010 Paul Kennedy pkennedy@redhat.com

Resolves: 492000

Changes -d to -s in arptables "OUT" directive in "Direct Routing and arptables_jf" section.

修订 1.0-0 Tue Jan 20 2009 Paul Kennedy pkennedy@redhat.com

Consolidation of point releases

索引

符号

A

C

D

F

I

J

L

M

N

P

R

S

W