《运维社区 －LVS+Keepalived 集群》

UNIXHOT 运维社区

http://www.unixhot.com

使用说明：

1.为保证本文的完整性和可用性，本文遵循 GFDL 协议，转载请注明“运维社区”字样和链接。

2.可以在 http://www.unixhot.com/pdf/lvs.pdf 找到本文的最新版本。

3.本文仅供参考使用，不承担任何因文档错误而造成的任何损失。

4.有任何问题可以在“UnixHot 运维社区”讨论交流。

5.有相关问题或业务合作。请邮件至 admin@unixhot.com。

修订历史记录

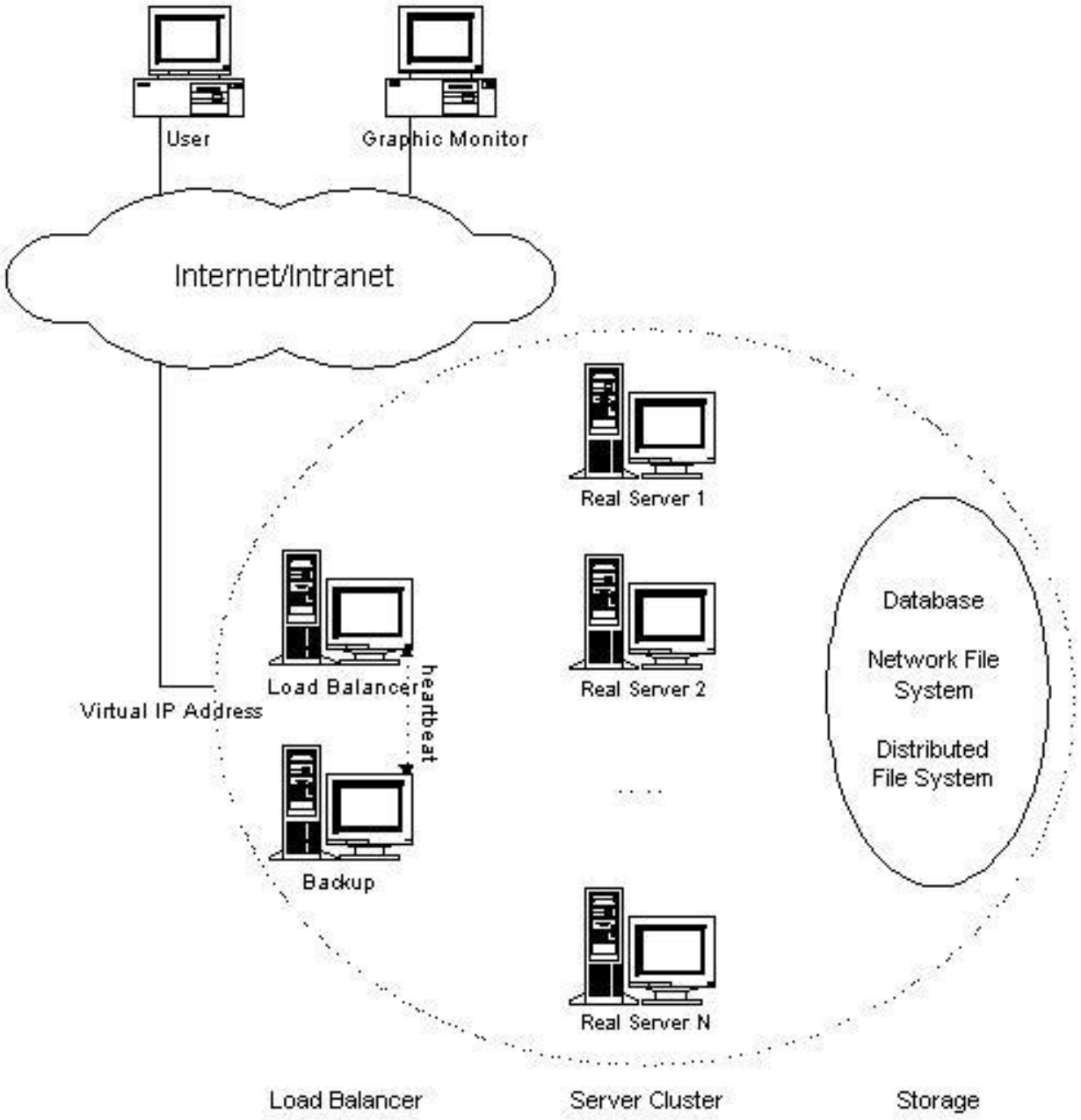
日　期	版　本	说　明
2010-01-01	V1.0	以 PDF 方式网络发布。
2011-11-01	V2.0	第一次完整修正后发布（该系列其它文档也在本月做修正）。

第 1 章 LVS 简介

LVS 是 Linux Virtual Server 的缩写，是由章文嵩博士开发的基于 Linux 内核的负载均衡技术。建议读者看本文档之前熟读 LVS 官方手册。

1.1 LVS 体系结构

LVS 建筑于实际的服务器集群之上，用户看不到提供服务的多台实际服务器，而只能看见一台作为负载平衡器的服务器。实际的服务器通过高速局域网或地理上分散的广域网连接。实际服务器的前端是一台负载均衡器，他将用户的请求调度到实际服务器上完成，这样看起来好像所有服务都是通过虚拟服务器来完成的。Linux 虚拟服务器能够提供良好的可升级性、可靠性和可用性。用户可以透明地增加或减少一个节点，可以对实际服务器进行监测，如果发现有节点失败就重新配置系统。



1.2 LVS 调度算法

LVS 提供了十种调度算法：

可以在这里查看：

```
[root@Web-node ~]# ls /lib/modules/2.6.18-164.el5/kernel/net/ipv4/ipvs/
ip_vs_dh.ko    ip_vs.ko        ip_vs_lblcr.ko  ip_vs_nq.ko  ip_vs_sed.ko  ip_vs_wlc.ko
```

ip_vs_ftp.ko ip_vs_lblc.ko ip_vs_lc.ko ip_vs_rr.ko ip_vs_sh.ko ip_vs_wrr.ko

1.轮叫（Round Robin RR）

调度器通过"轮叫"调度算法将外部请求按顺序轮流分配到集群中的真实服务器上，它均等地对待每一台服务器，而不管服务器上实际的连接数和系统负载。

2.加权轮叫（Weighted Round Robin WRR）

调度器通过"加权轮叫"调度算法根据真实服务器的不同处理能力来调度访问请求。这样可以保证处理能力强的服务器处理更多的访问流量。调度器可以自动问询真实服务器的负载情况，并动态地调整其权值。

3.最少链接（Least Connections LC）

调度器通过"最少连接"调度算法动态地将网络请求调度到已建立的链接数最少的服务器上。如果集群系统的真实服务器具有相近的系统性能，采用"最小连接"调度算法可以较好地均衡负载。

4.加权最少链接（Weighted Least Connections WLC）

在集群系统中的服务器性能差异较大的情况下，调度器采用"加权最少链接"调度算法优化负载均衡性能，具有较高权值的服务器将承受较大比例的活动连接负载。调度器可以自动问询真实服务器的负载情况，并动态地调整其权值。

5.基于局部性的最少链接（Locality-Based Least Connections LBLC）

"基于局部性的最少链接" 调度算法是针对目标 IP 地址的负载均衡，目前主要用于 Cache 集群系统。该算法根据请求的目标 IP 地址找出该目标 IP 地址最近使用的服务器，若该服务器是可用的且没有超载，将请求发送到该服务器；若服务器不存在，或者该服务器超载且有服务器处于一半的工作负载，则用"最少链接"的原则选出一个可用的服务器，将请求发送到该服务器。

6.带复制的基于局部性最少链接（Locality-Based Least Connections with Replication LBLCR

"带复制的基于局部性最少链接"调度算法也是针对目标 IP 地址的负载均衡，目前主要用于 Cache 集群系统。它与 LBLC 算法的不同之处是它要维护从一个目标 IP 地址到一组服务器的映射，而 LBLC 算法维护从一个目标 IP 地址到一台服务器的映射。该算法根据请求的目标 IP 地址找出该目标 IP 地址对应的服务器组，按"最小连接"原则从服务器组中选出一台服务器，若服务器没有超载，将请求发送到该服务器，若服务器超载；则按"最小连接"原则从这个集群中选出一台服务器，将该服务器加入到服务器组中，将请求发送到该服务器。同时，当该服务器组有一段时间没有被修改，将最忙的服务器从服务器组中删除，以降低复制的程度。

7.目标地址散列（Destination Hashing DH）

"目标地址散列"调度算法根据请求的目标 IP 地址，作为散列键（Hash Key）从静态分配的散列

表找出对应的服务器，若该服务器是可用的且未超载，将请求发送到该服务器，否则返回空。

8.源地址散列（Source Hashing SH）

"源地址散列"调度算法根据请求的源 IP 地址，作为散列键（Hash Key）从静态分配的散列表找出对应的服务器，若该服务器是可用的且未超载，将请求发送到该服务器，否则返回空。

9.最短期望延迟（Shortest Expected Delay Scheduling SED）

分配一个接踵而来的请求以最短的期望的延迟方式到服务器。

10.最小队列调度（Never Queue Scheduling NQ）

分配一个接踵而来的请求到一台空闲的服务器，此服务器不一定是最快的那台，如果所有服务器都是繁忙的，它采取最短的期望延迟分配请求。

1.3 LVS 负载均衡方法

LVS 提供了三种 IP 级的负载平衡方法：

Virtual Server via NAT 、Virtual Server via IP Tunneling、Virtual Server via Direct Routing。

Virtual Server via NAT 方法使用了报文双向重写的方法， Virtual Server via IP Tunneling 采用的是报文单向重写的策略， Virtual Server via Direct Routing 采用的是报文转发策略，这些策略将在以后的文章中详细描述。

1.4 常用术语介绍

DGW	公网 IP 地址的默认网关
VIP	客户端访问的公网 IP 地址，Director 的虚拟 IP 地址
DIP	Director 的真实 IP 地址
RIP	真实机的 IP 地址
CIP	客户端 IP 地址

1.5 相关资源

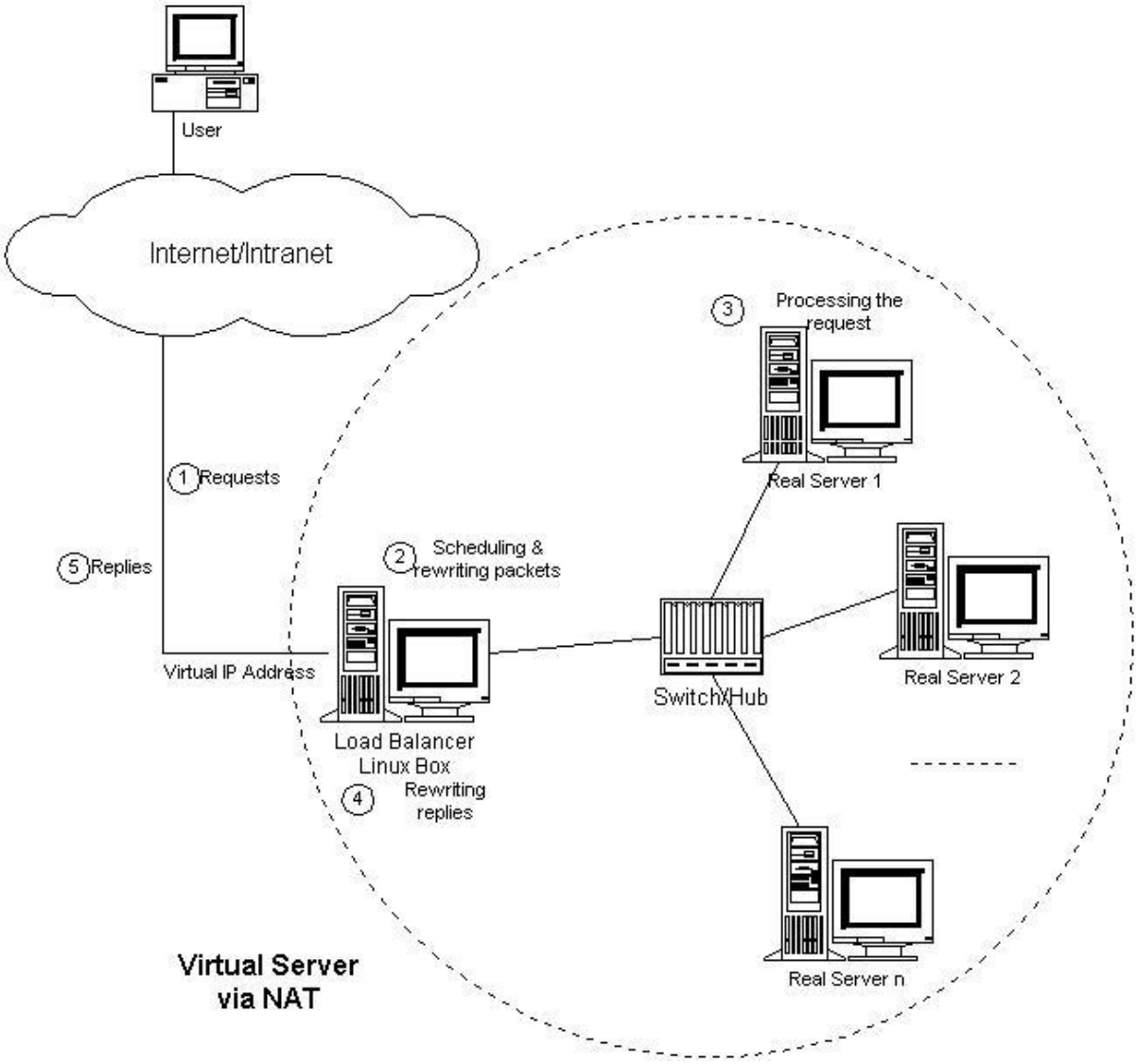
LVS 官方网站：http://www.linuxvirtualserver.org/

LVS 官方中文文档：http://www.linuxvirtualserver.org/zh/index.html

第 2 章 LVS-NAT 方式部署

2.1 LVS-NAT 方式体系结构

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。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。



2.2 部署前的准备工作

1> 服务器规划

IP 地址	主机名	描述
192.168.130.130	Web-Master	Director 分发器（VIP）
192.168.140.132	Web-Master	Director 分发器（DIP）
192.168.140.133	Web-node1	Real Server Web 节点 1
192.168.140.134	Web-node2	Real Server Web 节点 2

192.168.130.131	Web-Client	测试客户端

2 > 设置主机名解析和 ssh 验证

```
[root@Web-node ~]# vim /etc/hosts
# Do not remove the following line, or various programs
# that require network functionality will fail.
127.0.0.1       localhost.localdomain localhost
::1     localhost6.localdomain6 localhost6
192.168.140.132 Web-node
192.168.140.133 Web-node1
192.168.140.134 Web-node2
[root@Web-node ~]# ssh-keygen
[root@Web-node ~]# cp .ssh/id_rsa.pub .ssh/authorized_keys
[root@Web-node1 ~]# mkdir .ssh
[root@Web-node2 ~]# mkdir .ssh
[root@Web-node ~]# scp .ssh/authorized_keys Web-node1:/root/.ssh
[root@Web-node ~]# scp .ssh/authorized_keys Web-node2:/root/.ssh
[root@Web-node ~]# scp /etc/hosts Web-node1:/etc
[root@Web-node ~]# scp /etc/hosts Web-node2:/etc
```

2.3 在 Real Server 上的部署

1> Web-node1：

```
[root@Web-node1 ~]# mount /dev/cdrom /mnt
[root@Web-node1 ~]# rpm -ivh /mnt/Server/httpd-2.2.3-31.el5.i386.rpm
[root@Web-node1 ~]# echo Web-node1 > /var/www/html/index.html
[root@Web-node1 ~]# /etc/init.d/httpd start
[root@Web-node1 ~]# vim /etc/sysconfig/network-scripts/ifcfg-eth0
DEVICE=eth0
```

```
BOOTPROTO=static
IPADDR=192.168.140.133
NETMASK=255.255.255.0
GATEWAY=192.168.140.132 注意：默认网关设置为 DIP
ONBOOT=yes
HWADDR=00:0c:29:7c:cf:ba
[root@Web-node1 ~]# /etc/init.d/network restart
```

2> Web-node2

```
[root@Web-node2 ~]# mount /dev/cdrom /mnt
[root@Web-node2 ~]# rpm -ivh /mnt/Server/httpd-2.2.3-31.el5.i386.rpm
[root@Web-node2 ~]# echo Web-node2 > /var/www/html/index.html
[root@Web-node2 ~]# /etc/init.d/httpd start
[root@Web-node1 ~]# vim /etc/sysconfig/network-scripts/ifcfg-eth0
DEVICE=eth0
BOOTPROTO=static
IPADDR=192.168.140.134
NETMASK=255.255.255.0
GATEWAY=192.168.140.132 注意：默认网关设置为 DIP
ONBOOT=yes
HWADDR=00:0c:29:5d:2d:90
[root@Web-node2 ~]# /etc/init.d/network restart
```

请用浏览器访问两个 Web 几点，保证服务是正常运行。

2.4 在 Director 上的部署

1> 打开 IP_Forward

```
[root@Web-node ~]# vi /etc/sysctl.conf
net.ipv4.ip_forward = 1
[root@Web-node ~]# sysctl -p
```

2> 绑定 DIP 和 VIP

```
[root@Web-node ~]# vim /etc/sysconfig/network-scripts/ifcfg-eth0
DEVICE=eth0
BOOTPROTO=static
IPADDR=192.168.140.132
NETMASK=255.255.255.0
ONBOOT=yes
HWADDR=00:0c:29:17:39:9c
```

3> 安装 ipvsadm 软件包

```
[root@Web-node ~]# rpm -ivh /mnt/Cluster/ipvsadm-1.24-10.i386.rpm
```

4> 设置 ipvsadm

```
[root@Web-node ~]# modprobe iptable_nat
[root@Web-node ~]# ipvsadm -A -t 192.168.130.130:80 -s rr
[root@Web-node ~]# ipvsadm -a -t 192.168.130.130:80 -r 192.168.140.133 -m
[root@Web-node ~]# ipvsadm -a -t 192.168.130.130:80 -r 192.168.140.134 -m
[root@Web-node ~]# service ipvsadm save
Saving IPVS table to /etc/sysconfig/ipvsadm:                [  OK  ]
[root@Web-node ~]# chkconfig ipvsadm on
```

2.5 LVS -NAT 方式集群测试

1> 手动效果测试

[root@Web-Client ~]# elinks http://192.168.130.130 发现访问的是 Web-node1

[root@Web-Client ~]# elinks http://192.168.130.130 发现访问的是 Web-node2

2> 压力负载测试

在这里使用 Apache 自带的 ab 测试工具。

```
[root@Web-node ~]# ab -n 100 -c 100 http://192.168.140.133/
This is ApacheBench, Version 2.0.40-dev <$Revision: 1.146 $> apache-2.0
Copyright 1996 Adam Twiss, Zeus Technology Ltd, http://www.zeustech.net/
Copyright 2006 The Apache Software Foundation, http://www.apache.org/
```

Benchmarking 192.168.140.133 (be patient).....done

Server Software: Apache/2.2.3 #服务器平台和版本
Server Hostname: 192.168.140.133 #服务器主机名
Server Port: 80 #服务器端口号

Document Path: / #测试的页面
Document Length: 10 bytes #测试的页面大小

Concurrency Level: 100 #并发数
Time taken for tests: 0.91990 seconds #整个测试持续时间
Complete requests: 100 #完成的请求数
Failed requests: 0 #失败的请求数
Write errors: 0
Total transferred: 27200 bytes #整个测试场景的网络传输量
HTML transferred: 1000 bytes #整个测试场景的 HTML 内容传输量（10X100）
Requests per second: 1087.07 [#/sec] (mean) #平均每秒处理的事务数
Time per request: 91.990 [ms] (mean) #平均事物响应时间
Time per request: 0.920 [ms] (mean, across all concurrent requests) #每个请求运行的平均时间
Transfer rate: 282.64 [Kbytes/sec] received
#平均每秒网络上的流 量，可以帮助排除是否存在网络流量过大导致响应时间延长的问题
Connection Times (ms)

	min	mean[+/-sd]	median	max
Connect:	1	19 9.3	21	33
Processing:	9	32 14.2	32	65
Waiting:	8	31 14.1	31	56
Total:	10	52 22.9	54	90

```
Percentage of the requests served within a certain time (ms)
  50%     54
  66%     66
  75%     71
  80%     75
  90%     83
  95%     87
  98%     89
  99%     90
 100%     90 (longest request)
```

测试图表

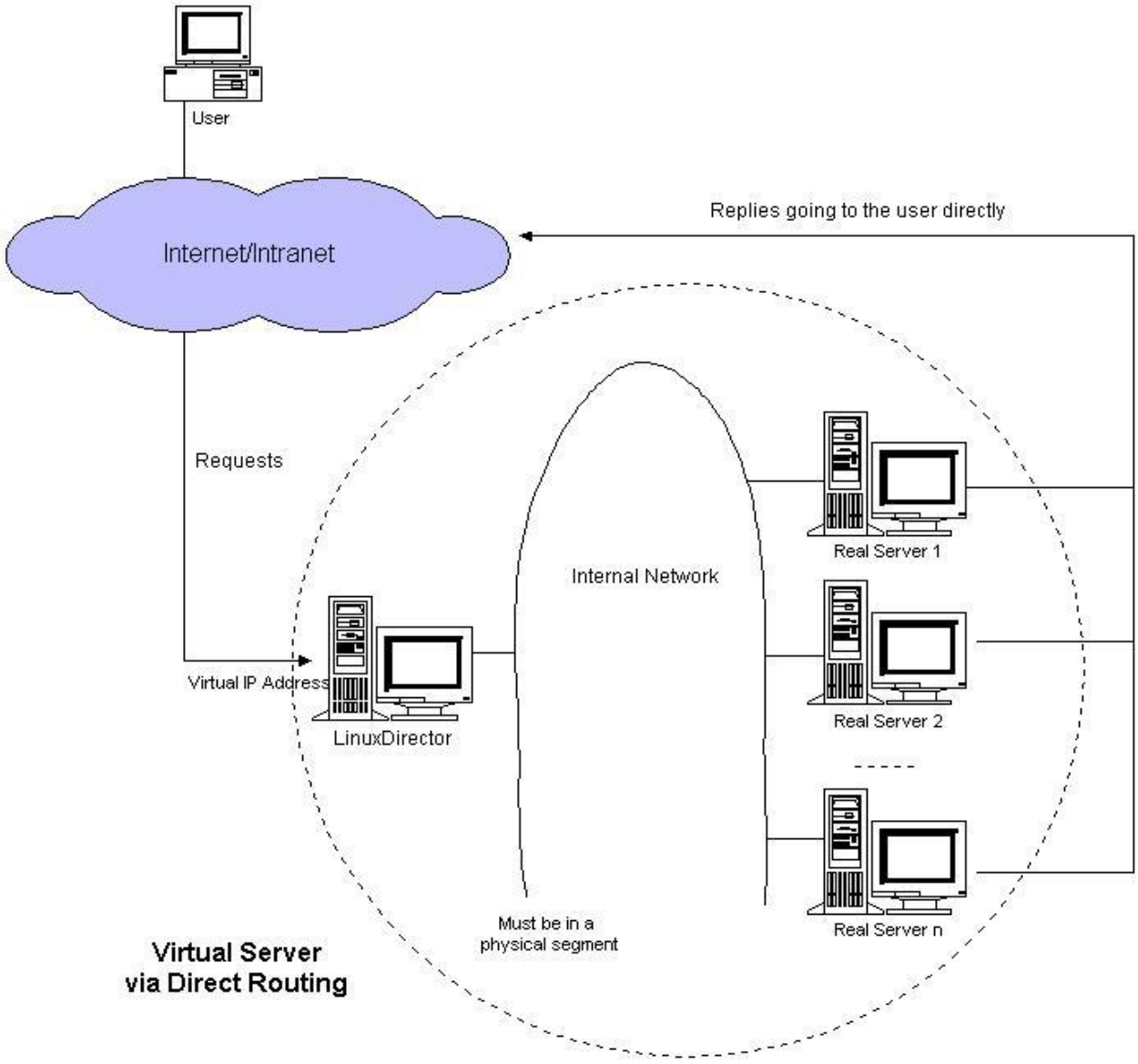
类　型	请求和并发数	每个请求平均处理时间
单个节点 Web-node1	100 个请求，每个请求 100 并发	0.961 [ms]
单个节点 Web-node2	100 个请求，每个请求 100 并发	0.950 [ms]
LVS 负载均衡	200 个请求，每个请求 200 并发	0.819

第 3 章 LVS-DR 方式部署

3.1 LVS-DR 方式体系结构

在 LVS-DR 中，调度器根据各个服务器的负载情况，动态地选择一台服务器，不修改也不封装 IP 报文，而是将数据帧的 MAC 地址改为选出服务器的 MAC 地址，再将修改后的数据帧在与服务器组的局域网上发送。因为数据帧的 MAC 地址是选出的服务器，所以服务器肯定可以收到这个数据帧，从中可以获得该 IP 报文。当服务器发现报文的目标地址 VIP 是在本地的网络设备上，服务器处理这个报文，然后根据路由表将响应报文直接返回给客户。

在 LVS-DR 中，根据缺省的 TCP/IP 协议栈处理，请求报文的目标地址为 VIP，响应报文的源地址肯定也为 VIP，所以响应报文不需要作任何修改，可以直接返回给客户，客户认为得到正常的服务，而不会知道是哪一台服务器处理的。



3.2 部署前的准备工作

1> 服务器规划

IP 地址	主机名	描述

192.168.140.128	Web-node	Director 分发器（VIP）
192.168.140.132	Web-node	Director 分发器（DIP）
192.168.140.133	Web-node1	Real Server Web 节点 1
192.168.140.134	Web-node2	Real Server Web 节点 2
192.168.140.130	Web-Client	测试客户端

在这里仅需要修改上个实验的 Direcoor 分发器的 IP 地址即可，使用 192.168.140.0 这个网段模拟公网地址。

2> 清除上个实验的配置

[root@Web-node ~]# echo "" > /etc/sysconfig/ipvsadm

[root@Web-node ~]# service ipvsadm restart

3.3 在 Real Server 上的部署

1> Web-node1

第一步：

[root@Web-node1 ~]# vim /etc/sysconfig/network-scripts/ifcfg-eth0

修改 GATEWAY 为 DGW：192.168.140.1

[root@Web-node1 ~]# /etc/init.d/network restart

第二步：

LVS-DR 方式中的 Real Server 需要忽略 ARP 解析，在这里用脚本实现

[root@Web-node1 ~]# vim /etc/sysconfig/network-scripts/arp.sh

```
#!/bin/bash
#===================================
# $Name:         RealServer.sh
# $Revision:     1.0
# $Function:     Config realserver lo and apply noarp
# $Author:       Shundong Zhao
# $organization: UnixHot
# $Create Date:  2010-08-10
```

```
#==================================

        WEB_VIP=192.168.140.140

        . /etc/rc.d/init.d/functions

        case "$1" in
        start)
                ifconfig lo:0 $SNS_VIP netmask 255.255.255.255 broadcast $WEB_VIP
                /sbin/route add -host $WEB_VIP dev lo:0
                echo "1" >/proc/sys/net/ipv4/conf/lo/arp_ignore
                echo "2" >/proc/sys/net/ipv4/conf/lo/arp_announce
                echo "1" >/proc/sys/net/ipv4/conf/all/arp_ignore
                echo "2" >/proc/sys/net/ipv4/conf/all/arp_announce
                sysctl -p >/dev/null 2>&1
                echo "RealServer Start OK"
                ;;
        stop)
                ifconfig lo:0 down
                route del $WEB_VIP >/dev/null 2>&1
                echo "0" >/proc/sys/net/ipv4/conf/lo/arp_ignore
                echo "0" >/proc/sys/net/ipv4/conf/lo/arp_announce
                echo "0" >/proc/sys/net/ipv4/conf/all/arp_ignore
                echo "0" >/proc/sys/net/ipv4/conf/all/arp_announce
                echo "RealServer Stoped"
                ;;
        *)
                echo "Usage: $0 {start|stop}"
                exit 1
        esac
```

```
        exit 0
[root@Web-node1 ~]# bash /etc/sysconfig/network-scripts/arp.sh
[root@Web-node1 ~]# ifconfig lo:0
lo:0      Link encap:Local Loopback
          inet addr:192.168.140.128  Mask:255.255.255.255
          UP LOOPBACK RUNNING  MTU:16436  Metric:1
```

2> Web-node2

第一步：

```
[root@Web-node2 ~]# vim /etc/sysconfig/network-scripts/ifcfg-eth0
```

修改 GATEWAY 为 DGW： 192.168.140.1

```
[root@Web-node2 ~]# /etc/init.d/network restart
```

第二步：

```
[root@Web-node1 ~]#
scp /etc/sysconfig/network-scripts/arp.sh Web-node2:/etc/sysconfig/network-scripts/
[root@Web-node2 ~]# bash /etc/sysconfig/network-scripts/arp.sh
[root@Web-node2 ~]# ifconfig lo:0
lo:0      Link encap:Local Loopback
          inet addr:192.168.140.128  Mask:255.255.255.255
          UP LOOPBACK RUNNING  MTU:16436  Metric:1
```

3.4 在 Director 上的部署

1> 绑定 VIP

```
[root@Web-node ~]# vim /etc/sysconfig/network-scripts/ifcfg-eth0:1
DEVICE=eth0:1
BOOTPROTO=static
IPADDR=192.168.140.128
NETMASK=255.255.255.0
ONBOOT=yes
```

HWADDR=00:0c:29:17:39:9c

2> 配置 ipvsadm

```
[root@Web-node ~]# ipvsadm -A -t 192.168.140.128:80 -s wrr -p 3600
[root@Web-node ~]# ipvsadm -a -t 192.168.140.128:80 -r 192.168.140.133:80 -g -w 10
[root@Web-node ~]# ipvsadm -a -t 192.168.140.128:80 -r 192.168.140.134:80 -g -w 20
[root@Web-node ~]# service ipvsadm save
```

3.5 LVS-DR 方式集群测试

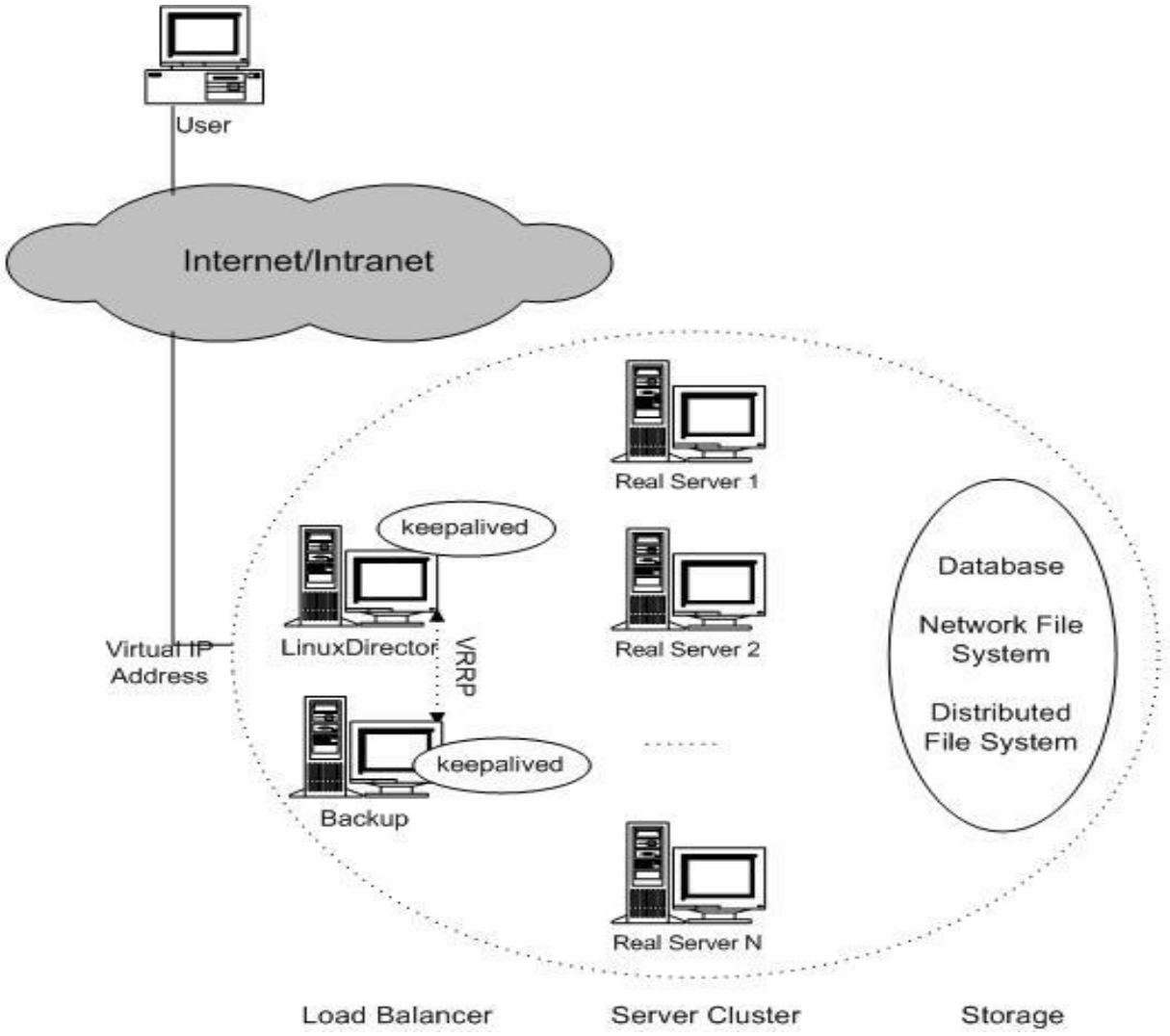
请读者用 ab 测试，模拟环境测试的性能指标可能区别不太明显（略）

第 4 章 LVS Keepalived 集群

4.1 LVS Keepalived 方案简介

LVS Keepalived 方案是和 LVS Heartbeat 一样的高可用加负载均衡方案，Keepalived 于 Heartbeat 相比在部署和管理方面都大大简化了。在多层负载均衡的架构中，它也是前段最好的选择，在下面会根据业务需求拓展架构，比如使用 Nginx 或者 Haproxy 做 Layer 7 的负载均衡，而前段仍使用 LVS+keepalived 来承载高并发的流量。

4.2 集群架构图



4.3 集群环境列表

IP 地址	主机名	说明
192.168.140.140		VIP
192.168.140.141	Web-Director	LVS Director
192.168.140.143	Web-Backup	Keepalived Backup
192.168.140.137	Web-node1	Web 节点
192.168.140.139	Web-node2	Web 节点

4.4 其它注意事项

1> LVS Keepalived 方案 ipvsadm 的管理有 keepalived 负责。

2> LVS 后端节点的监控检查也是 keepalived 进行。

3> 注意内核版本和 ipvsadm 版本的对应。

4> 注意 ip_vs 模块是否载入内核。

第 5 章 部署 LVS Keepalived

5.1 检查内核版本和模块

```
[root@LVS-node1 ~]# uname -r
2.6.18-194.el5
[root@LVS-node1 ~]# modprobe ip_vs
[root@LVS-node1 ~]# lsmod | grep ip_vs
ip_vs                 122113  0
```

5.2 下载软件包

```
[root@LVS-node1 ~]# ln -s /usr/src/kernels/2.6.18-194.el5-x86_64/ /usr/src/linux
[root@LVS-node1 ~]# cd /usr/local/src
[root@LVS-node1 src]# wget
http://www.linuxvirtualserver.org/software/kernel-2.6/ipvsadm-1.24.tar.gz
```

5.3 安装 ipvsadm

```
[root@LVS-node1 src]# tar zxvf ipvsadm-1.24.tar.gz
[root@LVS-node1 src]# cd ipvsadm-1.24
```

有关安装的相关信息，可以查看该目录下的 README 文件。

```
[root@LVS-node1 ipvsadm-1.24]# make && make install
```

安装完毕后会产生以下文件：

```
/sbin/ipvsadm
/sbin/ipvsadm-save
/sbin/ipvsadm-restore
/usr/man/man8/ipvsadm.8
/usr/man/man8/ipvsadm-save.8
/usr/man/man8/ipvsadm-restore.8
/etc/rc.d/init.d/ipvsadm
```

5.4 安装 Keepalived

```
[root@LVS-node1 src]# wget http://www.keepalived.org/software/keepalived-1.1.20.tar.gz
[root@LVS-node1 src]# tar zxf keepalived-1.1.20.tar.gz
[root@LVS-node1 src]# cd keepalived-1.1.20
[root@LVS-node1 keepalived-1.1.20]# ./configure --prefix=/usr/local/keepalived
Keepalived configuration
------------------------
Keepalived version       : 1.1.20
Compiler                 : gcc
Compiler flags           : -g -O2
Extra Lib                : -lpopt -lssl -lcrypto
Use IPVS Framework       : Yes
IPVS sync daemon support : Yes
Use VRRP Framework       : Yes
Use Debug flags          : No
```

如果出现以上输出，说明可以正常编译安装。

```
[root@LVS-node1 keepalived-1.1.20]# make && make install
[root@LVS-node1 keepalived-1.1.20]# cd /usr/local/keepalived/
[root@LVS-node1 keepalived]# cp etc/rc.d/init.d/keepalived /etc/rc.d/init.d/
```

```
[root@LVS-node1 keepalived]# cp etc/sysconfig/keepalived /etc/sysconfig/
[root@LVS-node1 keepalived]# /usr/local/keepalived/sbin/keepalived --help
```

如果你执行了 keepalived 命令，你会发现它默认去/etc/keepalived/keepalived.conf 找配置文件，所以我们要把配置文件创建到此处方便启动，当然，你也可以自己制定配置文件位置。

```
[root@LVS-node1 keepalived]# mkdir /etc/keepalived
[root@LVS-node1 keepalived]# cp etc/keepalived/keepalived.conf /etc/keepalived/
[root@LVS-node1 keepalived]# cp sbin/keepalived /usr/sbin/
```

第 6 章 LVS 配置

6.1 LVS Director 配置

LVS 配置是由 keepalived 依靠/etc/keepalived/keepalived.conf 来进行配置的，也就是说我们不需要手动配置 ipvsadm 程序，但是我们做为 LVS DR 模式运行，不能忘了对真实机的配置，对 RealServer 的配置参考“第 2 章 2.3.3 小节”。

6.2 RealServer 配置

```
[root@Web-node1 ~]# /sbin/realserver.sh start
RealServer Start OK
[root@Web-node2 ~]# /sbin/realserver.sh start
RealServer Start OK
```

在 RealServer 执行完 realserver.sh 后，一定要检查：

```
[root@Web-node1 ~]# ifconfig lo:0
lo:0      Link encap:Local Loopback
          inet addr:192.168.140.140  Mask:255.255.255.255
          UP LOOPBACK RUNNING  MTU:16436  Metric:1
[root@Web-node2 ~]# ifconfig lo:0
lo:0      Link encap:Local Loopback
```

```
inet addr:192.168.140.140  Mask:255.255.255.255
UP LOOPBACK RUNNING  MTU:16436  Metric:1
```

第 7 章 Keepalived 配置

这是 LVS Keepalived 集群方案的核心，主要集中在/etc/keepalived/keepalived 文件上。

Keepalived 有两种状态，MASTER 和 BACKUP，配置大致相似，但也有不同的地方，在 MASTER 的配置时，会用红色将，MASTER 和 BACKUP 不同的地方标出。

7.1 Web-Director 配置

下面是一个配置好的 keepalived.conf 文件，对于每行用注释说明。

```
[root@Web-Director ~]# cat /etc/keepalived/keepalived.conf
! Configuration File for keepalived
global_defs {                          #全局定义块
   notification_email {
    admin@unixhot.com                  #邮件通知模块，定义通知邮件地址。
   }
   notification_email_from localhost  #
   smtp_server 127.0.0.1          #定义 SMTP Server
   smtp_connect_timeout 30        #SMTP 链接超时时间
   router_id LVS_MASTER           #运行 keepalived 机器的一个标识，注意：Web-Backup 应该不同，修改为 LVS_BACKUP
}

vrrp_instance VI_1 {
    state MASTER                   #实例状态，注意：Web-Backup 应该修改为 BACKUP
    interface eth0                     #指定对外服务的网卡。
    virtual_router_id 51
    priority 101                       #优先级，注意：Web-Backup 应该修改比这个值小
```

```
    advert_int 1
    authentication {
        auth_type PASS
        auth_pass 1111
    }
    virtual_ipaddress {
    192.168.140.140
    }
}

virtual_server 192.168.140.140 80 {
    delay_loop 6                        #健康检查的时间间隔
    lb_algo wrr                         #负载均衡的调度算法
    lb_kind DR                         #负载均衡转发模式
    nat_mask 255.255.255.0
    persistence_timeout 50              #会话保持时间，针对动态网站
    protocol TCP                        #转发的协议

    real_server 192.168.140.137 80 {       #真实机的设置
        weight 1                       #真实机的权重，在带有加权调度的调度算法中有用
    TCP_CHECK {                       #TCP 健康检查
        connect_timeout 10
        nb_get_retry 3
        delay_before_retry 3
        connect_port 80
        }
     }

    real_server 192.168.140.139 80 {
        weight 1
```

```
    TCP_CHECK {
        connect_timeout 10
        nb_get_retry 3
        delay_before_retry 3
        connect_port 80
        }
        }
    }
```

7.2 BACKUP 端配置

BACKUP 端配置和 MASTER 几乎一样，可以直接用 scp 从 MASTER 端复制一份过来，做以下修改即可：

```
    router_id LVS_MASTER
    state MASTER
    priority 101
```

第 8 章 LVS Keepalived 方案测试

8.1 启动 keepalived 服务

```
[root@Web-Director ~]# /etc/init.d/keepalived start
[root@Web-Backup ~]# /etc/init.d/keepalived start
```

8.2 VIP 切换测试

```
[root@Web-Director ~]# ip ad li eth0
2: eth0: <BROADCAST,MULTICAST,UP,LOWER_UP> mtu 1500 qdisc pfifo_fast qlen 1000
    link/ether 00:0c:29:57:04:db brd ff:ff:ff:ff:ff:ff
    inet 192.168.140.141/24 brd 192.168.140.255 scope global eth0
```

```
    inet 192.168.140.140/32 scope global eth0
    inet6 fe80::20c:29ff:fe57:4db/64 scope link
       valid_lft forever preferred_lft forever
```

可以发现，默认的 VIP 是绑定在优先级比较高的这台 MASTER 上，也就是 Web-Director 服务器。

```
[root@Web-Backup ~]# ip ad li eth0
2: eth0: <BROADCAST,MULTICAST,UP,LOWER_UP> mtu 1500 qdisc pfifo_fast qlen 1000
    link/ether 00:0c:29:45:5e:0e brd ff:ff:ff:ff:ff:ff
    inet 192.168.140.143/24 brd 192.168.140.255 scope global eth0
    inet6 fe80::20c:29ff:fe45:5e0e/64 scope link
       valid_lft forever preferred_lft forever
```

你可以手动来停掉 Web-Director 上的 keepalived 服务，VIP 就会自动切到 Web-Backup 上。

附录：GFDL 协议

（注：可以参考本协议的中文翻译版本 http://www.thebigfly.com/gnu/FDLv1.3/）

GNU Free Documentation License

Version 1.3, 3 November 2008

Everyone is permitted to copy and distribute verbatim copies of this license document, but changing it is not allowed.

0. PREAMBLE

The purpose of this License is to make a manual, textbook, or other functional and useful document "free" in the sense of freedom: to assure everyone the effective freedom to copy and redistribute it, with or without modifying it, either commercially or noncommercially. Secondarily, this License preserves for the author and publisher a way to get credit for their work, while not being considered responsible for modifications made by others.
This License is a kind of "copyleft", which means that derivative works of the document must themselves be free in the same sense. It complements the GNU General Public License, which is a copyleft license designed for free software.
We have designed this License in order to use it for manuals for free software, because free software needs free documentation: a free program should come with manuals providing the same freedoms that the software does. But this License is not limited to software manuals; it can be used for any textual work, regardless of subject matter or whether it is published as a printed book. We recommend this License principally for works whose purpose is instruction or reference.

1. APPLICABILITY AND DEFINITIONS

This License applies to any manual or other work, in any medium, that contains a notice placed by the copyright holder saying it can be distributed under the terms of this License. Such a notice grants a world-wide, royalty-free license, unlimited in duration, to use that work under the conditions stated herein. The "Document", below, refers to any such manual or work. Any member of the public is a licensee, and is addressed as "you". You accept the license if you copy, modify or distribute the work in a way requiring permission under copyright law.
A "Modified Version" of the Document means any work containing the Document or a portion of it, either copied verbatim, or with modifications and/or translated into another language.
A "Secondary Section" is a named appendix or a front-matter section of the Document that deals exclusively with the relationship of the publishers or authors of the Document to the Document's overall subject (or to related matters) and contains nothing that could fall directly within that overall subject. (Thus, if the Document is in part a textbook of mathematics, a Secondary Section may not explain any mathematics.) The relationship could be a matter of historical connection with the subject or with related matters, or of legal, commercial, philosophical, ethical or political position regarding them.
The "Invariant Sections" are certain Secondary Sections whose titles are designated, as being those of Invariant Sections, in the notice that says that the Document is released under this License. If a section does not fit the above definition of Secondary then it

is not allowed to be designated as Invariant. The Document may contain zero Invariant Sections. If the Document does not identify any Invariant Sections then there are none.
The "Cover Texts" are certain short passages of text that are listed, as Front-Cover Texts or Back-Cover Texts, in the notice that says that the Document is released under this License. A Front-Cover Text may be at most 5 words, and a Back-Cover Text may be at most 25 words.
A "Transparent" copy of the Document means a machine-readable copy, represented in a format whose specification is available to the general public, that is suitable for revising the document straightforwardly with generic text editors or (for images composed of pixels) generic paint programs or (for drawings) some widely available drawing editor, and that is suitable for input to text formatters or for automatic translation to a variety of formats suitable for input to text formatters. A copy made in an otherwise Transparent file format whose markup, or absence of markup, has been arranged to thwart or discourage subsequent modification by readers is not Transparent. An image format is not Transparent if used for any substantial amount of text. A copy that is not "Transparent" is called "Opaque".
Examples of suitable formats for Transparent copies include plain ASCII without markup, Texinfo input format, LaTeX input format, SGML or XML using a publicly available DTD, and standard-conforming simple HTML, PostScript or PDF designed for human modification. Examples of transparent image formats include PNG, XCF and JPG. Opaque formats include proprietary formats that can be read and edited only by proprietary word processors, SGML or XML for which the DTD and/or processing tools are not generally available, and the machine-generated HTML, PostScript or PDF produced by some word processors for output purposes only.
The "Title Page" means, for a printed book, the title page itself, plus such following pages as are needed to hold, legibly, the material this License requires to appear in the title page. For works in formats which do not have any title page as such, "Title Page" means the text near the most prominent appearance of the work's title, preceding the beginning of the body of the text.
The "publisher" means any person or entity that distributes copies of the Document to the public.
A section "Entitled XYZ" means a named subunit of the Document whose title either is precisely XYZ or contains XYZ in parentheses following text that translates XYZ in another language. (Here XYZ stands for a specific section name mentioned below, such as "Acknowledgements", "Dedications", "Endorsements", or "History".) To "Preserve the Title" of such a section when you modify the Document means that it remains a section "Entitled XYZ" according to this definition.
The Document may include Warranty Disclaimers next to the notice which states that this License applies to the Document. These Warranty Disclaimers are considered to be included by reference in this License, but only as regards disclaiming warranties: any other implication that these Warranty Disclaimers may have is void and has no effect on the meaning of this License.

2. VERBATIM COPYING

You may copy and distribute the Document in any medium, either commercially or noncommercially, provided that this License, the copyright notices, and the license notice saying this License applies to the Document are reproduced in all copies, and that you add no other conditions whatsoever to those of this License. You may not use technical measures to obstruct or control the reading or further copying of the copies you make or distribute. However, you may accept compensation in exchange for copies. If you distribute a large enough number of copies you must also follow the conditions in section 3.
You may also lend copies, under the same conditions stated above, and you may publicly display copies.

3. COPYING IN QUANTITY

If you publish printed copies (or copies in media that commonly have printed covers) of the Document, numbering more than 100, and the Document's license notice requires Cover Texts, you must enclose the copies in covers that carry, clearly and legibly, all these Cover Texts: Front-Cover Texts on the front cover, and Back-Cover Texts on the back cover. Both covers must also clearly and legibly identify you as the publisher of these copies. The front cover must present the full title with all words of the title equally prominent and visible. You may add other material on the covers in addition. Copying with changes limited to the covers, as long as they preserve the title of the Document and satisfy these conditions, can be treated as verbatim copying in other respects.
If the required texts for either cover are too voluminous to fit legibly, you should put the first ones listed (as many as fit reasonably) on the actual cover, and continue the rest onto adjacent pages.
If you publish or distribute Opaque copies of the Document numbering more than 100, you must either include a machine-readable Transparent copy along with each Opaque copy, or state in or with each Opaque copy a computer-network location from which the general network-using public has access to download using public-standard network protocols a complete Transparent copy of the Document, free of added material. If you use the latter option, you must take reasonably prudent steps, when you begin distribution of Opaque copies in quantity, to ensure that this Transparent copy will remain thus accessible at the stated location until at least one year after the last time you distribute an Opaque copy (directly or through your agents or retailers) of that edition to the public.
It is requested, but not required, that you contact the authors of the Document well before redistributing any large number of copies, to give them a chance to provide you with an updated version of the Document.

4. MODIFICATIONS

You may copy and distribute a Modified Version of the Document under the conditions of sections 2 and 3 above, provided that you release the Modified Version under precisely this License, with the Modified Version filling the role of the Document, thus licensing

distribution and modification of the Modified Version to whoever possesses a copy of it. In addition, you must do these things in the Modified Version:

- A. Use in the Title Page (and on the covers, if any) a title distinct from that of the Document, and from those of previous versions (which should, if there were any, be listed in the History section of the Document). You may use the same title as a previous version if the original publisher of that version gives permission.
- B. List on the Title Page, as authors, one or more persons or entities responsible for authorship of the modifications in the Modified Version, together with at least five of the principal authors of the Document (all of its principal authors, if it has fewer than five), unless they release you from this requirement.
- C. State on the Title page the name of the publisher of the Modified Version, as the publisher.
- D. Preserve all the copyright notices of the Document.
- E. Add an appropriate copyright notice for your modifications adjacent to the other copyright notices.
- F. Include, immediately after the copyright notices, a license notice giving the public permission to use the Modified Version under the terms of this License, in the form shown in the Addendum below.
- G. Preserve in that license notice the full lists of Invariant Sections and required Cover Texts given in the Document's license notice.
- H. Include an unaltered copy of this License.
- I. Preserve the section Entitled "History", Preserve its Title, and add to it an item stating at least the title, year, new authors, and publisher of the Modified Version as given on the Title Page. If there is no section Entitled "History" in the Document, create one stating the title, year, authors, and publisher of the Document as given on its Title Page, then add an item describing the Modified Version as stated in the previous sentence.
- J. Preserve the network location, if any, given in the Document for public access to a Transparent copy of the Document, and likewise the network locations given in the Document for previous versions it was based on. These may be placed in the "History" section. You may omit a network location for a work that was published at least four years before the Document itself, or if the original publisher of the version it refers to gives permission.
- K. For any section Entitled "Acknowledgements" or "Dedications", Preserve the Title of the section, and preserve in the section all the substance and tone of each of the contributor acknowledgements and/or dedications given therein.
- L. Preserve all the Invariant Sections of the Document, unaltered in their text and in their titles. Section numbers or the equivalent are not considered part of the section titles.
- M. Delete any section Entitled "Endorsements". Such a section may not be included in the Modified Version.
- N. Do not retitle any existing section to be Entitled "Endorsements" or to conflict

in title with any Invariant Section.

- O. Preserve any Warranty Disclaimers.

If the Modified Version includes new front-matter sections or appendices that qualify as Secondary Sections and contain no material copied from the Document, you may at your option designate some or all of these sections as invariant. To do this, add their titles to the list of Invariant Sections in the Modified Version's license notice. These titles must be distinct from any other section titles.

You may add a section Entitled "Endorsements", provided it contains nothing but endorsements of your Modified Version by various parties—for example, statements of peer review or that the text has been approved by an organization as the authoritative definition of a standard.

You may add a passage of up to five words as a Front-Cover Text, and a passage of up to 25 words as a Back-Cover Text, to the end of the list of Cover Texts in the Modified Version. Only one passage of Front-Cover Text and one of Back-Cover Text may be added by (or through arrangements made by) any one entity. If the Document already includes a cover text for the same cover, previously added by you or by arrangement made by the same entity you are acting on behalf of, you may not add another; but you may replace the old one, on explicit permission from the previous publisher that added the old one.

The author(s) and publisher(s) of the Document do not by this License give permission to use their names for publicity for or to assert or imply endorsement of any Modified Version.

5. COMBINING DOCUMENTS

You may combine the Document with other documents released under this License, under the terms defined in section 4 above for modified versions, provided that you include in the combination all of the Invariant Sections of all of the original documents, unmodified, and list them all as Invariant Sections of your combined work in its license notice, and that you preserve all their Warranty Disclaimers.

The combined work need only contain one copy of this License, and multiple identical Invariant Sections may be replaced with a single copy. If there are multiple Invariant Sections with the same name but different contents, make the title of each such section unique by adding at the end of it, in parentheses, the name of the original author or publisher of that section if known, or else a unique number. Make the same adjustment to the section titles in the list of Invariant Sections in the license notice of the combined work.

In the combination, you must combine any sections Entitled "History" in the various original documents, forming one section Entitled "History"; likewise combine any sections Entitled "Acknowledgements", and any sections Entitled "Dedications". You must delete all sections Entitled "Endorsements".

6. COLLECTIONS OF DOCUMENTS

You may make a collection consisting of the Document and other documents released under

this License, and replace the individual copies of this License in the various documents with a single copy that is included in the collection, provided that you follow the rules of this License for verbatim copying of each of the documents in all other respects.
You may extract a single document from such a collection, and distribute it individually under this License, provided you insert a copy of this License into the extracted document, and follow this License in all other respects regarding verbatim copying of that document.

7. AGGREGATION WITH INDEPENDENT WORKS

A compilation of the Document or its derivatives with other separate and independent documents or works, in or on a volume of a storage or distribution medium, is called an "aggregate" if the copyright resulting from the compilation is not used to limit the legal rights of the compilation's users beyond what the individual works permit. When the Document is included in an aggregate, this License does not apply to the other works in the aggregate which are not themselves derivative works of the Document.
If the Cover Text requirement of section 3 is applicable to these copies of the Document, then if the Document is less than one half of the entire aggregate, the Document's Cover Texts may be placed on covers that bracket the Document within the aggregate, or the electronic equivalent of covers if the Document is in electronic form. Otherwise they must appear on printed covers that bracket the whole aggregate.

8. TRANSLATION

Translation is considered a kind of modification, so you may distribute translations of the Document under the terms of section 4. Replacing Invariant Sections with translations requires special permission from their copyright holders, but you may include translations of some or all Invariant Sections in addition to the original versions of these Invariant Sections. You may include a translation of this License, and all the license notices in the Document, and any Warranty Disclaimers, provided that you also include the original English version of this License and the original versions of those notices and disclaimers. In case of a disagreement between the translation and the original version of this License or a notice or disclaimer, the original version will prevail.
If a section in the Document is Entitled "Acknowledgements", "Dedications", or "History", the requirement (section 4) to Preserve its Title (section 1) will typically require changing the actual title.

9. TERMINATION

You may not copy, modify, sublicense, or distribute the Document except as expressly provided under this License. Any attempt otherwise to copy, modify, sublicense, or distribute it is void, and will automatically terminate your rights under this License. However, if you cease all violation of this License, then your license from a particular copyright holder is reinstated (a) provisionally, unless and until the copyright holder

explicitly and finally terminates your license, and (b) permanently, if the copyright holder fails to notify you of the violation by some reasonable means prior to 60 days after the cessation.
Moreover, your license from a particular copyright holder is reinstated permanently if the copyright holder notifies you of the violation by some reasonable means, this is the first time you have received notice of violation of this License (for any work) from that copyright holder, and you cure the violation prior to 30 days after your receipt of the notice.
Termination of your rights under this section does not terminate the licenses of parties who have received copies or rights from you under this License. If your rights have been terminated and not permanently reinstated, receipt of a copy of some or all of the same material does not give you any rights to use it.

10. FUTURE REVISIONS OF THIS LICENSE

The Free Software Foundation may publish new, revised versions of the GNU Free Documentation License from time to time. Such new versions will be similar in spirit to the present version, but may differ in detail to address new problems or concerns. See http://www.gnu.org/copyleft/.
Each version of the License is given a distinguishing version number. If the Document specifies that a particular numbered version of this License "or any later version" applies to it, you have the option of following the terms and conditions either of that specified version or of any later version that has been published (not as a draft) by the Free Software Foundation. If the Document does not specify a version number of this License, you may choose any version ever published (not as a draft) by the Free Software Foundation. If the Document specifies that a proxy can decide which future versions of this License can be used, that proxy's public statement of acceptance of a version permanently authorizes you to choose that version for the Document.

11. RELICENSING

"Massive Multiauthor Collaboration Site" (or "MMC Site") means any World Wide Web server that publishes copyrightable works and also provides prominent facilities for anybody to edit those works. A public wiki that anybody can edit is an example of such a server. A "Massive Multiauthor Collaboration" (or "MMC") contained in the site means any set of copyrightable works thus published on the MMC site.
"CC-BY-SA" means the Creative Commons Attribution-Share Alike 3.0 license published by Creative Commons Corporation, a not-for-profit corporation with a principal place of business in San Francisco, California, as well as future copyleft versions of that license published by that same organization.
"Incorporate" means to publish or republish a Document, in whole or in part, as part of another Document.
An MMC is "eligible for relicensing" if it is licensed under this License, and if all works

that were first published under this License somewhere other than this MMC, and subsequently incorporated in whole or in part into the MMC, (1) had no cover texts or invariant sections, and (2) were thus incorporated prior to November 1, 2008.
The operator of an MMC Site may republish an MMC contained in the site under CC-BY-SA on the same site at any time before August 1, 2009, provided the MMC is eligible for relicensing.

ADDENDUM: How to use this License for your documents

To use this License in a document you have written, include a copy of the License in the document and put the following copyright and license notices just after the title page:

> Copyright (C) YEAR YOUR NAME.
> Permission is granted to copy, distribute and/or modify this document
> under the terms of the GNU Free Documentation License, Version 1.3
> or any later version published by the Free Software Foundation;
> with no Invariant Sections, no Front-Cover Texts, and no Back-Cover Texts.
> A copy of the license is included in the section entitled "GNU
> Free Documentation License".

If you have Invariant Sections, Front-Cover Texts and Back-Cover Texts, replace the "with … Texts." line with this:

> with the Invariant Sections being LIST THEIR TITLES, with the
> Front-Cover Texts being LIST, and with the Back-Cover Texts being LIST.

If you have Invariant Sections without Cover Texts, or some other combination of the three, merge those two alternatives to suit the situation.
If your document contains nontrivial examples of program code, we recommend releasing these examples in parallel under your choice of free software license, such as the GNU General Public License, to permit their use in free software.