# Clustering und Failover mit Linux 2004


Markus Oswald <moswald@iirc.at>

# Worum geht es?

- Load-Balanced Cluster
- Failover Cluster
- Shared Storage (DRBD)

# Worum es nicht geht?

- Computational Cluster
- Beowulf
- Distributed Computing

# Inhalt

Was ist neu?

Theorie (I/II/III)
Anwendungen (I/II)
Load-Balancing Cluster
- Aufbau
- Software
- Konfiguration (I/II)
- High Availability

Failover Cluster
- Aufbau
- Software
- Funktion: heartbeat
- Funktion: keepalived
- Funktion: DRBD
- Konfiguration

Einsatzszenarien

# Was ist neu?

Was ist neu seit letztem Jahr:

- Keepalived mit VRRP
  - Redundante LVS Balancer / Router

- OpenBSD 3.5 mit CARP + pfsync
  - Ersatz für VRRP – nicht mit Patenten belastet
  - Stateful Failover !

- DRBD 0.7 "demnächst" fertig (?)
  - Kompatibel mit Linux 2.6
  - Kompatibel zu XFS
  - Activity Log (kein full-sync nach Ausfall der prim. Node)

# Theorie I

Warum Clustern?

- Ausfallsicherheit
  - Ein Server kann schnell ausfallen
  - Es geht nicht immer ohne rebooten (Hardware, Kernel)
- Lastverteilung
  - Ein Server schafft die Last oft einfach nicht mehr
- Skalierbarkeit
  - Ein Pool aus mehreren Servern lässt sich dynamisch den Anforderungen anpassen
- Kosten
  - Viele kleine Server sind oft billiger als ein großer

# Theorie II

Cluster Typen:

Load-Balanced Cluster

Mehrere Server teilen sich die Last

Technik: LVS (oder round-robin DNS, iptables, ...)
Application-Level (z.B. mod_backhand)

Failover Cluster

Backup Server übernimmt bei Ausfall die Dienste

Technik: heartbeat (Failsave, Kimberlite, ... )
VRRP (keepalived) / CARP (OpenBSD)

# Theorie III

Typischer Einsatz:

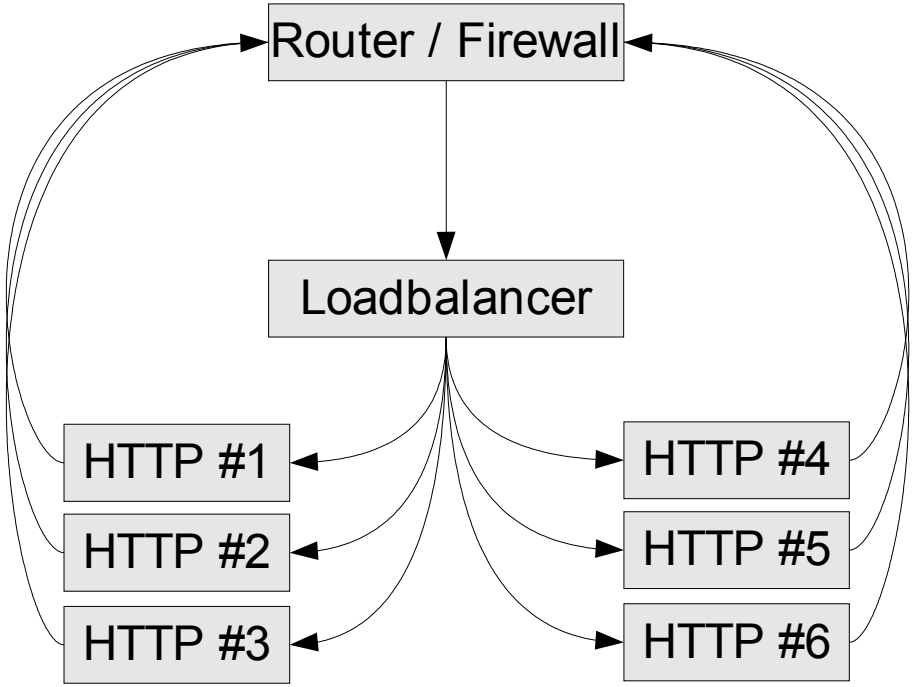
- Load-Balanced Cluster
  - Webserver
  - Mailserver
  - ...
- Failover Cluster
  - DB Server
  - Fileserver
  - Firewall / Router
  - ...


Markus Oswald, <moswald@iirc.at>

# Anwendungen I

Load Balanced:

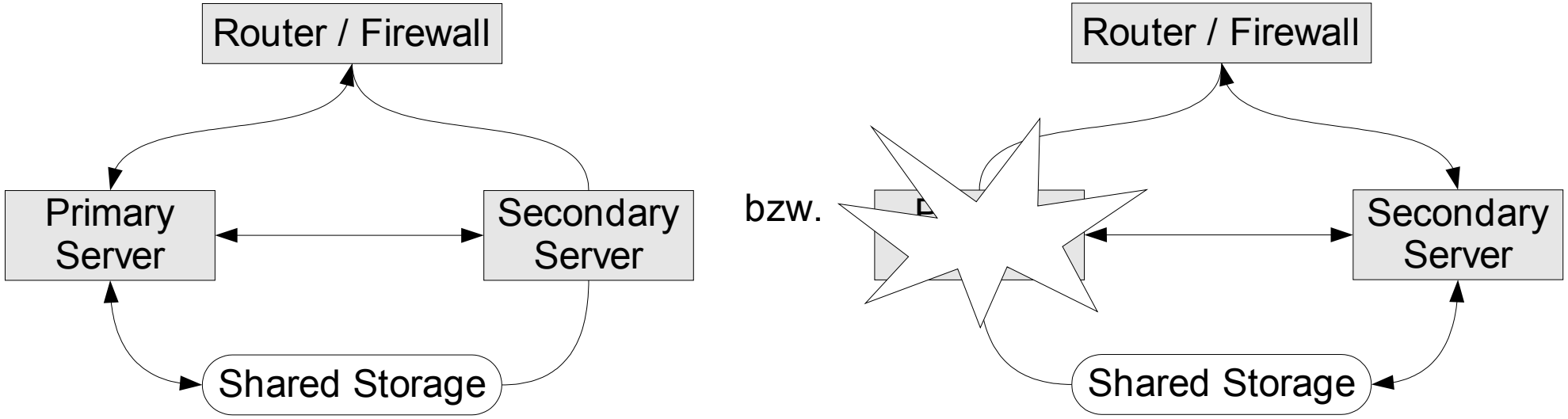
Webserver-Farm



# Anwendungen II

Failover:

DB Server





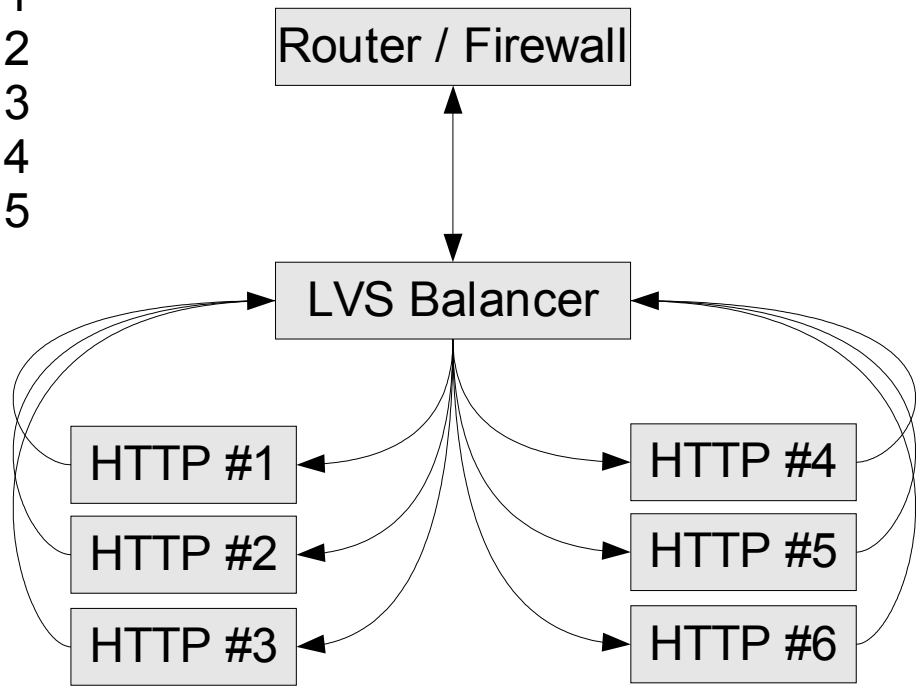
Markus Oswald, <moswald@iirc.at>

# Load Balancing – Aufbau

Bei NAT Forwarding:

| | | |
|---|---|---|
| LVS Balancer: | eth0: | 207.175.44.110 |
| | eth1: | 192.168.10.250 |
| HTTP #1: | eth0: | 192.168.10.1 |
| HTTP #2: | eth0: | 192.168.10.2 |
| HTTP #3: | eth0: | 192.168.10.3 |
| HTTP #4: | eth0: | 192.168.10.4 |
| HTTP #5: | eth0: | 192.168.10.5 |

....





Markus Oswald, <moswald@iirc.at>

# Load Balancing – Software

Kernel:

- Linux 2.2.x mit IPVS Patch
- Linux 2.4.x mit IPVS Patch oder Netfilter Modulen
- ab Linux 2.4.23 / 2.6.x direkt im Kernel

Konfiguration:

- ipvsadm

High-Availability:

- Keepalived

# Load Balancing – Konfiguration I

Konfiguration mittels ipvsadm ist recht einfach:

```
ipvsadm -A -t 207.175.44.110:80 -s rr
ipvsadm -a -t 207.175.44.110:80 -r 192.168.10.1:80 -m
ipvsadm -a -t 207.175.44.110:80 -r 192.168.10.2:80 -m
ipvsadm -a -t 207.175.44.110:80 -r 192.168.10.3:80 -m
ipvsadm -a -t 207.175.44.110:80 -r 192.168.10.4:80 -m
ipvsadm -a -t 207.175.44.110:80 -r 192.168.10.5:80 -m
```

*Verteilt alle Verbindungen auf 207.175.44.110:80 gleichmäßig über 192.168.10.1 bis 192.168.10.5*

```
bash$ ipvsadm -L -n
IP Virtual Server version 1.0.2 (size=4096)
Prot LocalAddress:Port Scheduler Flags
  -> RemoteAddress:Port           Forward Weight ActiveConn InActConn
TCP  207.175.44.110:80 rr
  -> 192.168.10.1:80              Masq    10     4          29
  -> 192.168.10.2:80              Masq    10     3          24
  -> 192.168.10.3:80              Masq    10     4          25
  -> 192.168.10.4:80              Masq    10     3          27
  -> 192.168.10.5:80              Masq    10     4          21
```

# Load Balancing – Konfiguration II

## Weitere Optionen:

`-f, --fwmark-service`
Firewall-Mark anstelle von Adresse, Port und Protokoll verwenden um virtuelle Server zu unterscheiden

`-s, --scheduler`
Algorithmus zur Verteilung der Verbindungen. Derzeit gibt es 7 Möglichkeiten:

- rr - Robin Robin
- wrr - Weighted Round Robin
- lc - Least-Connection
- wlc - Weighted Least-Connection
- lblcr - Locality-Based Least-Connection with Replication
- dh - Destination Hashing
- sh - Source Hashing

`-p, --persistent`
Mehrere Requests eines Clients kommen immer zum gleichen Server.
Sinnvoll bei FTP oder SSL Verbindungen

`-w, --weight`
Kapazität eines Servers – z.B. bei unterschiedlich starken Servern im Pool

# Load Balancing – High Availability

## Hochverfügbarkeit mittels Keepalived:

- Überwacht Services der Server
- Entfernt Server aus dem Pool oder fügt diese wieder hinzu
- Weiterleitung auf Sorry-Server bei Totalausfall
- Benachrichtigt den Administrator über Ausfäll via Email
- Handelt LVS Failover bei redundanten Setups mittels VRRP

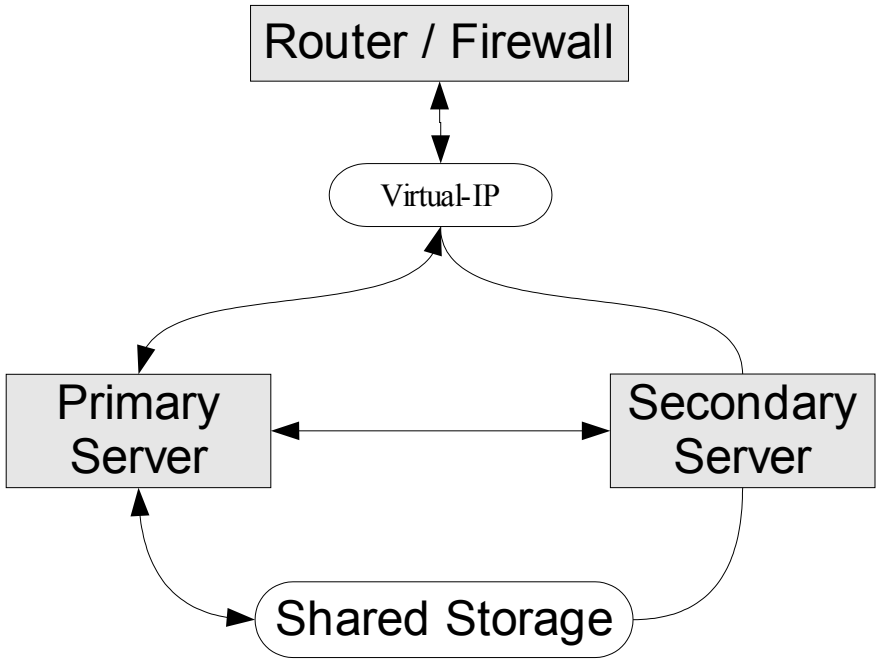
## Konfiguration einigermaßen einfach und gut dokumentiert:

```
virtual_server 207.175.44.110 80 {
delay_loop 10
lb_algo rr
lb_kind NAT
nat_mask 255.255.255.0
protocol TCP

  real_server 192.168.10.1 80 {
      weight 10
      HTTP_GET {
          url {
            path /status/status.php
            digest c4ca4238a0b923820dcc509a6f75849b
          }
          connect_timeout 10
          connect_port 80
          nb_get_retry 3
          delay_before_retry 3
      }
  }
}
```


Markus Oswald, <moswald@iirc.at>

# Failover – Aufbau




```
eth0      192.168.1.10                        eth0      192.168.1.11
eth0:0    192.168.1.15    - - - - - - - - ->
eth1      10.0.0.1                            eth1      10.0.0.2
```


Markus Oswald, <moswald@iirc.at>

# Failover – Software

Kernel:

Linux 2.2.x, Linux 2.4.x oder Linux 2.6.x

High-Availability:
- keepalived (VRRP)
- heartbeat

Shared-Storage:
- DRBD (Kernel-Patch)
- Hardware (externes Disk-Array, SAN, ...)

# Failover – Funktion: keepalived

Was macht keepalived?

- LVS Server-Pool Management
  - Konfiguration des Pools (Adressen, Ports, server weight, ...)
  - Überwachung der Server (tcp_connect/http_get/...)

- Failover Framework für Router, Firewalls und Balancer
  - Übernahme von VRRP Instanzen (mehrere IP-Adressen)
  - Media Link failure detection via LINKWATCH

Typischer Einsatz:
- Redundante Firewalls/Router/Balancer

# Failover – Funktion: heartbeat

Was macht heartbeat?

- Host monitoring via serial, UDP, PPP/UDP heartbeats
- IP address takeover (Übernahme der VIP)
- Ressourcen Management (Start/Stop von Services)
- Unterstützt STONITH (Shoot The Other Node In The Head)

Typischer Einsatz:

- Failover Cluster z.B. mit DRBD

# Failover – Funktion: DRBD

Was macht DRBD:

- Wird als Block-Device angesprochen, daher transparent
- Daten sind via Netzwerk gespiegelt
- Arbeitet "auf" Partitionen oder ganzen Platten
- Dateisystemunabhängig
- Zwischen 50 und 98% des theoretisch maximal möglichen Datendurchsatzes
- DRBD 0.7 "real soon now"

## "RAID 1 over Ethernet"

# Denkbarer Einsatz

LVS (keepalived):
- Webserver - HTTP / HTTPS
- Mailserver – SMTP / POP / IMAP
- Proxy-Server

VRRP (keepalived) / CARP (OpenBSD):
- Firewalls
- Loadbalancer
- Router

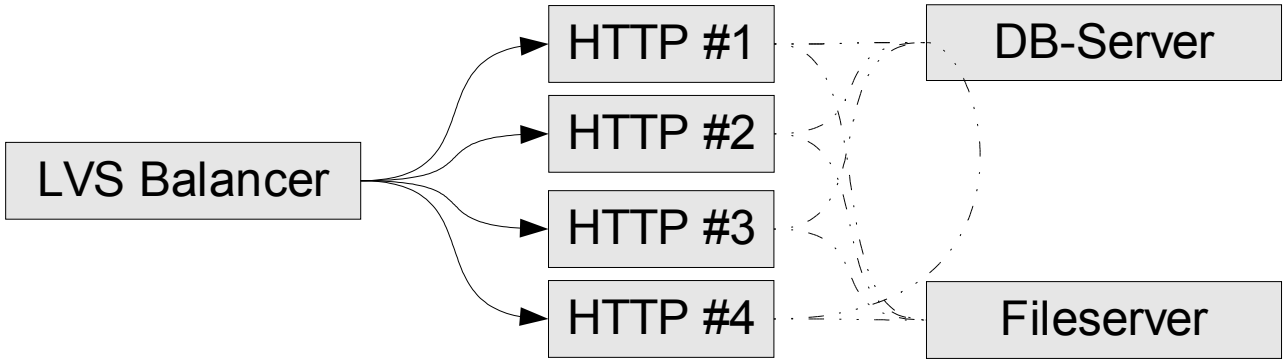
heartbeat + DRBD oder anderem shared Storage:
- Datenbanken – MySQL / PostgreSQL
- Webserver
- Mailserver
- Fileserver

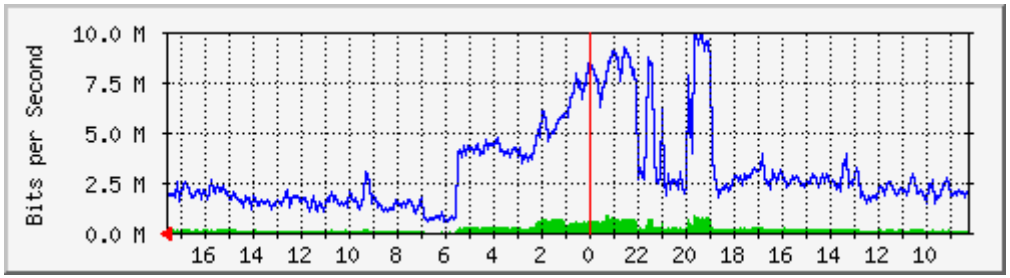
# Beispiel-Installationen I

## Typischer Webcluster

LVS Balancer: Gibraltar 1.2 , Kernel 2.4.23 P3 900 / 128MB CPU-Last: <10 %
Webserver: Debian 3.0 Woody

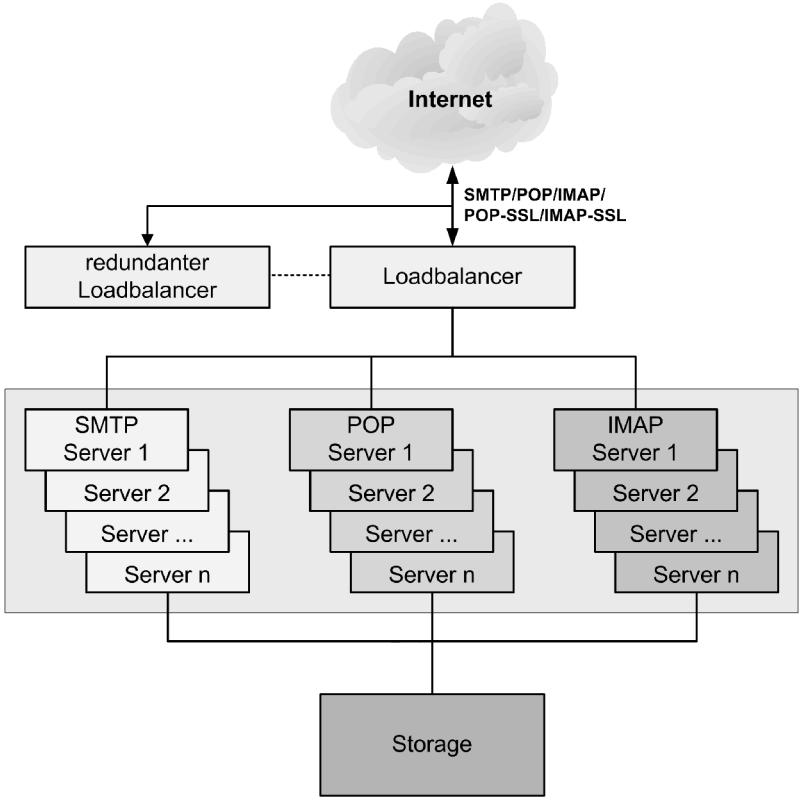
```
Prot LocalAddress:Port Scheduler Flags
  -> RemoteAddress:Port           Forward      Weight   ActiveConn   InActConn
TCP  XXX.XXX.XXX.XXX:80           rr  persistent  0
  -> 192.168.1.11:80              Masq         10          138          317
  -> 192.168.1.12:80              Masq         10          143          412
  -> 192.168.1. ...
```



# Beispiel-Installationen II

Einsatz als High-Performance Mailserver mit Postfix



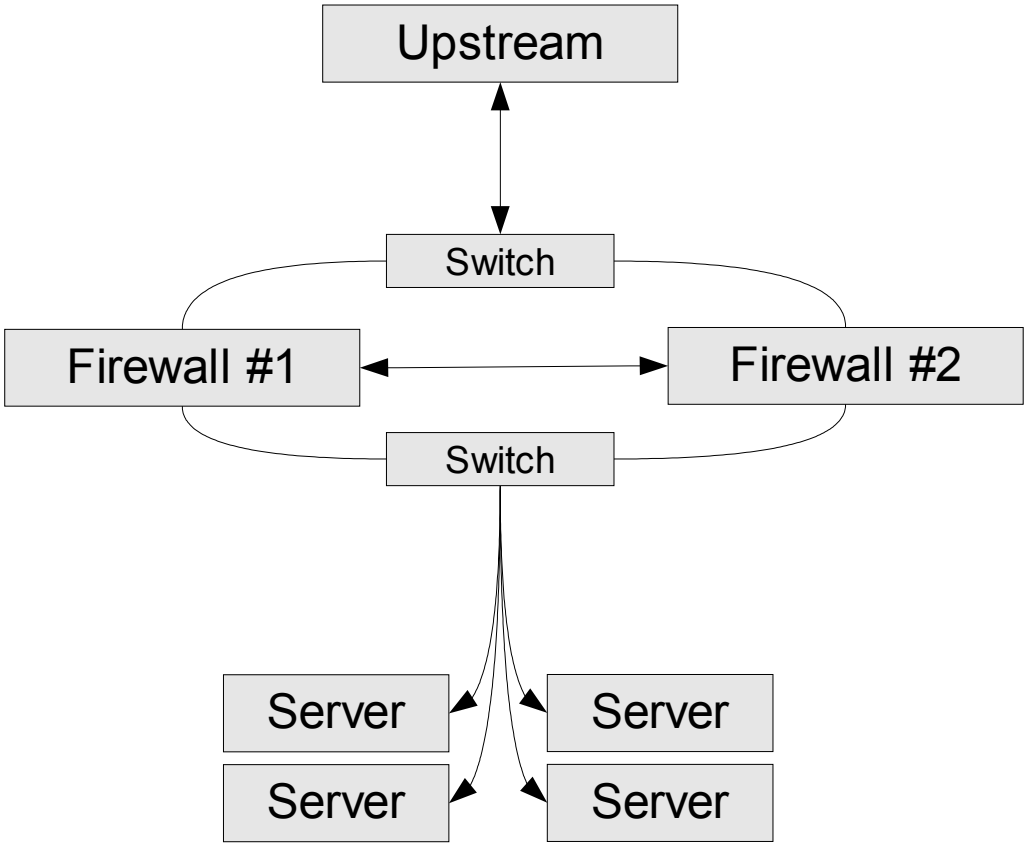

Vortrag: Postfix als High Performance Mailserver
Zeit: Samstag, 08. Mai, 16:30-17:30
Ort: HS1: Hörsaal 15 – Fachhochschule, Erdgeschoß

# Beispiel-Installationen III

Failover Firewall / Router

Gibraltar Linux
basierend auf Debian
Diskless Operation
(bootet von CD, Konfiguration auf Floppy/USB)
Failoverzeiten < 5 Sekunden

Hardware:
FSC Primergy L100 Rack-Server



# Fragen?

# Links

| | |
|---|---|
| LVS: | http://www.linuxvirtualserver.org/ |
| Keepalived: | http://www.keepalived.org |
| heartbeat: | http://www.linux-ha.org/heartbeat/ |
| DRBD: | http://www.drbd.org |
| Gibraltar: | http://www.gibraltar.at |
| Debian: | http://www.debian.org |

Vortrag: morgen, 14:00-16:00

E-Mail: Markus Oswald, <moswald@iirc.at>