Failover Procedures

A. Cavalli

A. Pagano

Information Society

Outline

- Actions
- Wiki
- Replication Status
- Linux vserver virtual machines
- Geo Failover (DNS)
- Next steps

Actions

1) Wiki: create sections for the replication details and status of each tool
2) Proceed with replication of tools
3) Test the DNS idea
4) Get a contact for Failover at CERN

Wiki

http://goc.grid.sinica.edu.tw/gocwiki/ --> OpDocs

http://goc.grid.sinica.edu.tw/gocwiki/Failover_mechanisms

- **Current COD tools replication status and Failover ideas wrote into the Wiki:**
  - Introduction
  - Operation toos:
    - CIC Portal*
    - GGUS
    - GOCDB
    - GSTAT
    - MAILING LISTS*
    - SFT/SAME*
    - SFT ADMIN
  - Geo Failover

***not yet**

Replication status

- **CIC Portal: TODO**
  - structure analysed in Lyon
  - Web+Lavoisier: find site+hosts and start
  - MySQL: wait for Oracle
- **GGUS: currently not considered – locally done**
- **GOCDB: MySQL done in TW, Web TODO**
- **GSTAT: DONE at CNAF and documented**
- **Mailing Lists: TODO**
- **SFT/SAME: TODO**
- **SFT ADMIN: foreseen into CIC Portal**

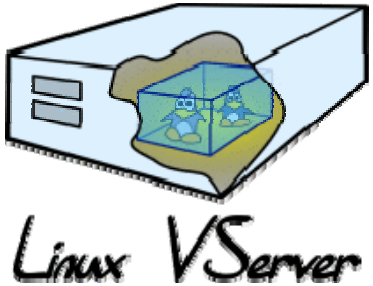
Virtual machines

A soft partitioning concept based on
Security Contexts which permits the creation of many independent
Virtual Private Servers (VPS) that run simultaneously on a single physical server at full speed, efficiently sharing:

- the same kernel
- the same hardware resources

Vserver: Main Points

- Impact ~0% - Overhead inside a Context less than 2%
- There are no special daemons running
- No pre-allocated disk space needed (40-100 megs)
- Independent updates
- 32-/64-bit independence



- Security: a cracked vserver can't reach the host server

Why do you need vservers ?

Fields of Application

- Many Tasks on the same Box
- Resource Independence: Moving vservers (and to optimise load)
- Experimenting and Upgrading
- Development
- Distribution Independence

- Administrative Separation
- Service Separation
- Enhancing Security
- Easy Maintenance
- Fail-over Scenarios

Step to step installation:

- Compile and patch the Kernel ...
- Reboot with the new kernel ...
- Create a new vm

```
[root@vm1 ~]# vserver vm6 build -m yum --hostname vm6.cnaf.infn.it –interfac domain=eth0:131.154.100.169/24 --
      initstyle sysv -- -d slc43

================================================================================

 Package            Arch      Version        Repository       Size

================================================================================

openssl             i686      0.9.7a-43.8    slc-base         1.1 M .....

     Install    276 Package(s)

     Update       0 Package(s)

     Remove       0 Package(s)

     Total download size: 146 M
```

1

2

```
[root@vm1 ~]# vserver vm6 start
Starting system logger:      [  OK  ]
Starting kernel logger:      [  OK  ]
```

3

```
[root@vm1 ~]# vserver vm6 enter
[root@vm6 /]# ps uax
USER       PID %CPU %MEM    VSZ  RSS TTY      STAT START   TIME COMMAND
root         1  1.6  0.0   1688  576 ?        S    11:37   0:01 init [3]
root     20952  0.0  0.0   1596  552 ?        Ss   11:37   0:00 syslogd -m 0
root     20973  3.0  0.0   4284 1376 pts/0    S    11:38   0:00 /bin/bash -login
root     21012  0.0  0.0   2372  760 pts/0    R+   11:38   0:00 ps uax
```

Our example:

```
[root@vm1 ~]# vserver-stat
CTX   PROC    VSZ    RSS  userTIME   sysTIME    UPTIME NAME
0       98 158.9M  51.7M  20m48s43  12m12s68   7d05h08 root server
49178    6  76.2M   10M   0m01s67   0m01s90   4d00h54 vm4
49182    5  71.8M   9.1M  0m05s71   0m03s76   2d22h31 vm2
49191    6   76M    10M   9m08s74   0m01s36   2d02h12 vm5
49193    6  75.1M   9.6M  0m00s23   0m00s14   30m44s74 vm3
```

- Vm1 physical machine
- Vm2 dns1
- Vm3 client dns
- Vm4 Root dns
- Vm5 dns2

References

- http://linux-vserver.org
- http://linux-vserver.org/CentOS_HowTo

The End (just my part!)

Questions

Geo Failover with DNS

- All-in-one-vserver-node testbed
- Complete DNS hierarchy reproduced (Bind 9.2.4)
- Tests with real IPv4 addresses of Operations tools
- nsupdate client to change mappings
- Resolver floods to test Bind under stress
- Simple php interface to pilot nsupdate (not yet ready to present)

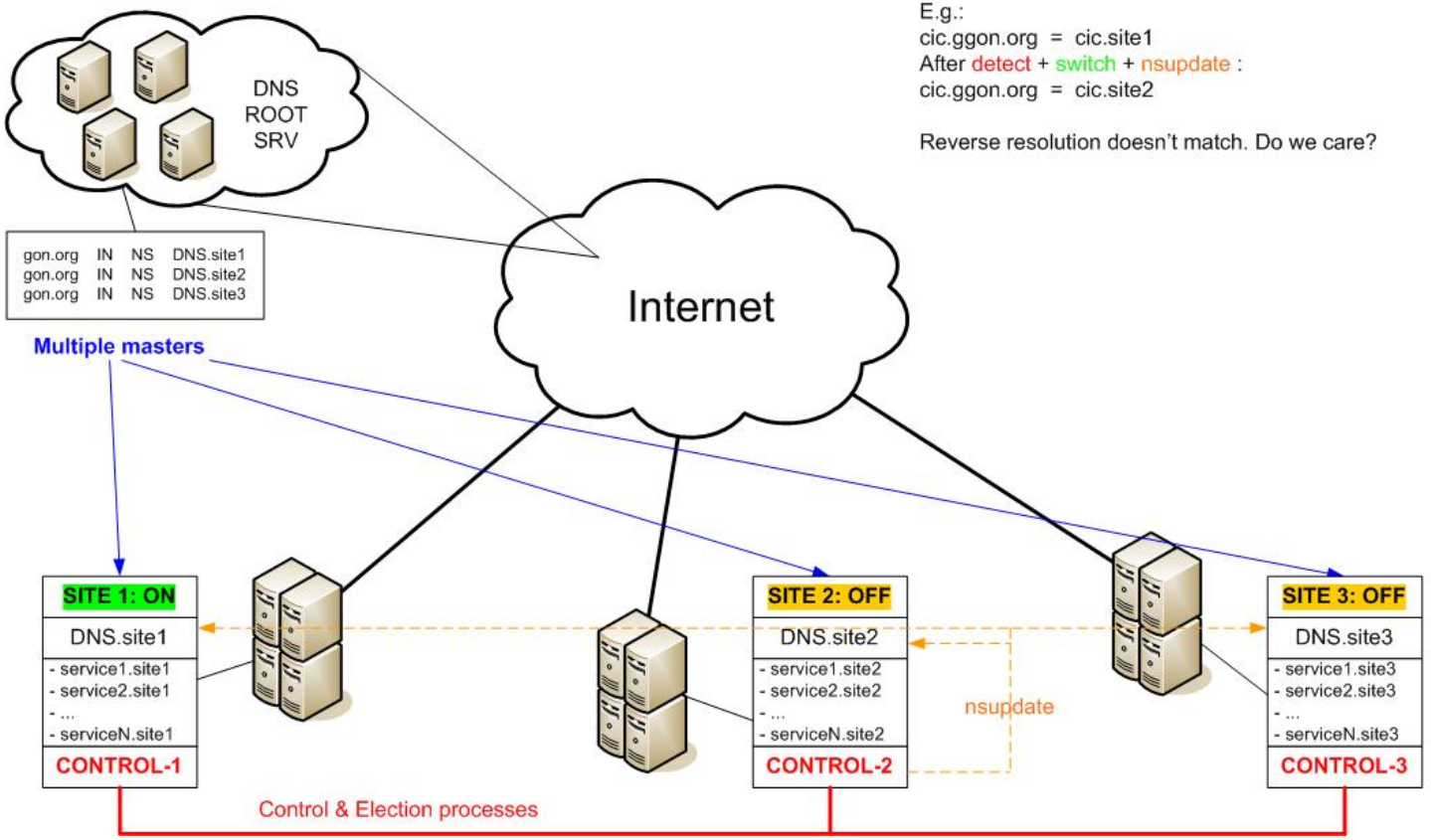
The general schema

ggon.org – Global Grid Operations Network

Control: if n-1 sites detect master down => switch
Election: round-robin?

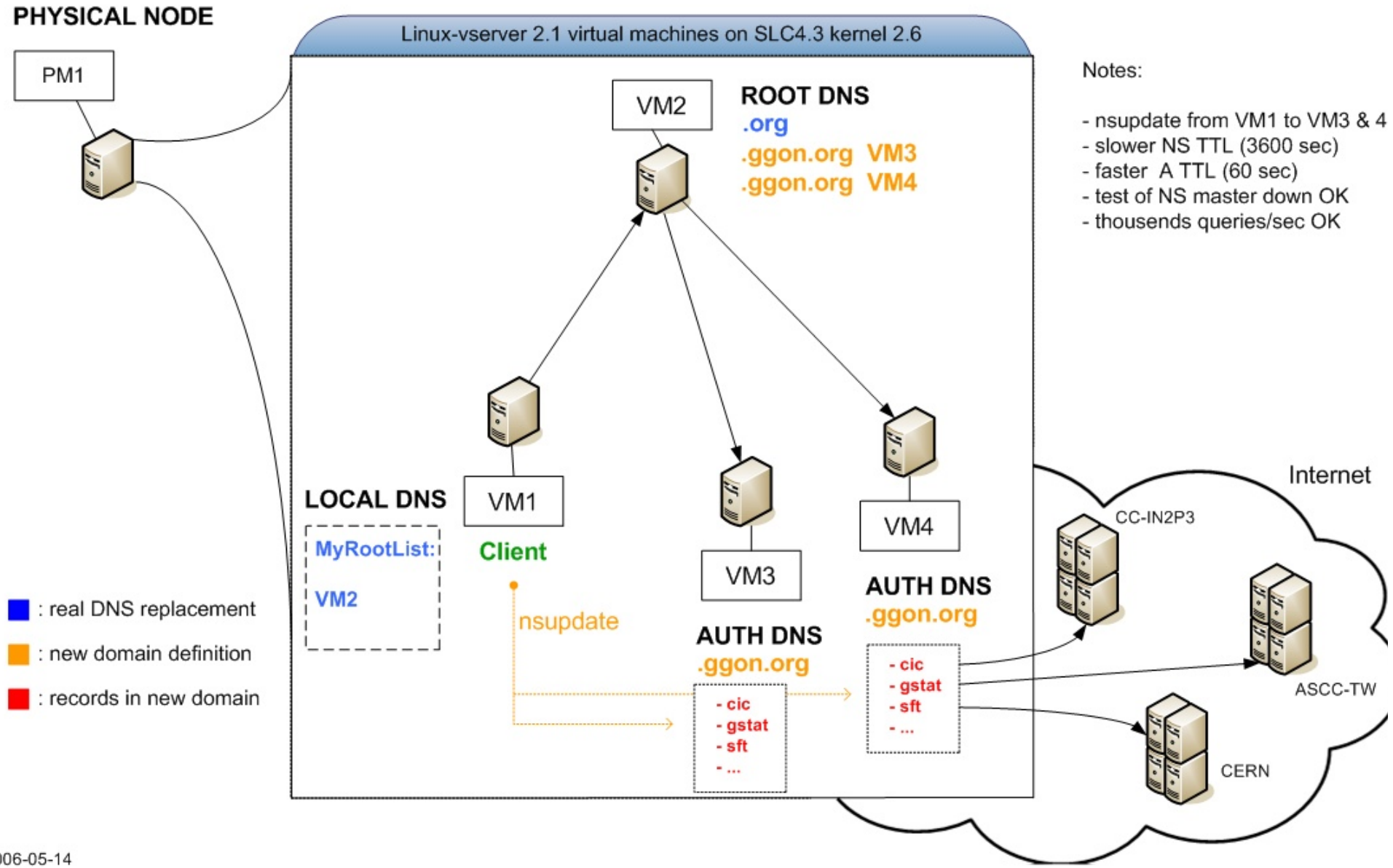
E.g.:
cic.ggon.org = cic.site1
After detect + switch + nsupdate :
cic.ggon.org = cic.site2

Reverse resolution doesn't match. Do we care?



ggon.org – Global Grid Operations Network

Tested in a closed vserver environment



Geo failover: 1st test results

- **Experience gathered:**
  - setting up a real DNS hierarchy
  - remote-controlling Bind daemons (rndc)
  - remote-updating DNS records (nsupdate)
  - managing linux-vserver VMs
- **Fast TTLs (60s) used for records: need to be fine tuned at the end, but**
- **Bind daemons have been stressed by thousends queries/sec without loadavg nor response time changes**
- **DNS multiple masters:**
  - no delay noticed when one of the masters turned off => DNS redundancy OK
- **Successfully tested remap of OPS tools under new names. E.g. try:**
  - `$ host cic.ggon.org vm3.cnaf.infn.it`
  - `$ dig @vm2.cnaf.infn.it ggon.org axfr`
  - or put 131.154.100.163 as nameserver in your resolv.conf
  - see next slide: how many links in the OPS tools are not ready to be remapped

Absolute links in COD tools

- **GSTAT:** all links ok
- **CIC**: almost everything works, including most important features
  - Certificate belongs to cic.in2p3.fr warning
  - All upper TAB links
  - Homepage: many links
  - Home/Portal features: all links inside
  - RC staff/Homepage: one link at bottom
  - CIC/ROC operation metrics (points to egee.in2p3.fr)
  - Dashboard:
    - "Back to CIC portal" link at top & bottom
    - Manage problems list
    - HandOver: several links and the "log" php
  - CIC/On duty procedures: one link
  - one link in the "Broadcast Info" page
- **GOCDB**: **everything ok**
  - Certificate belongs to goc.grid-support.ac.uk warning
- **LCG-SFT**:
  - Certificate belongs to lcg-sft.cern.ch warning
  - All status and single test links
- **GOC Wiki**:
  - all Wiki-managed links seem to be relative, except for some that could be human error

DNS Performance under stress

```
[root@vm queryperf-nominum-2.1]# ./queryperf -d file.list -s IP

Statistics:
  Parse input file:     once
  Ended due to:         reaching end of file

  Queries sent:         1000 queries
  Queries completed:    1000 queries
  Queries lost:         0 queries
  Percentage completed: 100.00%
  Percentage lost:      0.00%

  Ran for:              0.522707 seconds

  Queries per second:   1913.117674
```

CPU 2 x Xeon 2.80GHz
Mem: 2 gb ram
Hd raid 1 SATA 120 gb

DNS loadavg increase:

almost nothing,
0.01 0.00 0.00

After artificially loaded server:

response-time increase not noticeable

Next steps

Item	Start	Done
CIC portal: **find a place and replicate its components**	**COD-8**	**2006-09-01**
SFT/SAME: **start to gather components info**	**COD-8**	**COD-8**
SFT/SAME: **find place and start replication of its components**	**COD-9 ?**	**It depends on development status and Oracle "replica-bility"**
GOCDB: **collect all info about replica done in TW**	**after COD-8**	**COD-9**
Geo-DNS idea phase2: **complete php interface, write & test a prototype automatic control-switch logic**	**after COD-8**	**COD-10 ?**
Keep wiki **updated**	**-**	**-**

Acknowledgements

Thanks a lot to:

- **luca.dellagnello@cnaf.infn.it**
- **riccardo.veraldi@cnaf.infn.it**

for their help, answering questions as DNS experts.