OpenVZ management with Quattor

Luis Fernando Muñoz Mejías

Universidad Autónoma de Madrid

6th Quattor Workshop
Nikhef, 10/2008

Outline

Outline

Features

- Guests have no kernel
  - No worries about static resource assignment
    - But, of course, you can set limits if you need them
  - No worries about hardware
  - No worries about filesystems
- Virtual machines are isolated in namespaces inside the host kernel
- Very high performance
  - Over 90% of native performance, even on I/O operations
- Very high density of VMs
  - We have up to 10 VMs hosting production services on the same host
  - We have up to 18 VMs on the same host for testing instances

Limitations

- Only Linux inside Linux
  - But that's what I want
- VMs are not allowed to use kernel threads
  - VMs can't run NFS servers
- OpenVZ tools don't support all Linux capabilities
  - NSCD fails when trying to drop privileges on SL5 guests
- OpenVZ doesn't work with SELinux
- Tools for intrusion detection may return false positives

Outline

No kernel

- We use a single filesystem for all VEs
- No Anaconda during installation
  - We have to mimic a few things with existing NCM components
- Installations are based on images
- We use x86_64 SL5 hosts with SL4 and SL5 guests, i386 and x86_64
- In principle, SL4 hosts are possible too

Outline

Special kernel

- We need an OpenVZ kernel
- Vanilla kernels contain some namespacing patches, but they are not yet ready
- Add it to AII installation, too

```
"/system/aii/osinstall/ks/extra_packages" =
    list ("ovzkernel");
```

Special tools

- Only vzctl and friends

Outline

Enable network forwarding

- The host acts as a router for its guests
- Enable `net.ipv4.ip_forward` on sysctl
  - Use ncm-sysctl for that
- Adjust your iptables settings for the FORWARD chain
  - Either set `ACCEPT` default policy or carefully allow traffic from and to guests
- Set NAT rules, if needed

Guest declaration

- Mandatory information
  - Container ID
  - Network information
  - URLs for bootstrap image (template) and pseudo-ks
- Resource limits may be configured too
- The list is indexed by guest name

Guest declaration
Example

```
"/software/components/openvz/guests" = nlist (
    “foo.bar.baz”,
    nlist (“id”, 1000,
        “template_url”, “http://your/image/url”,
        “ksurl”, “http://your/ks/url”,
        “ip”, list (“ip1”, “ip2”, ...),
        “dns”. “dns.server.com”,
        “onboot”, true,
        “start_upon_creation”, false,
);
```

Other considerations

- Don't install all your VMs at the same time
  - You don't want 18 instances of SPMA at the same time ;)
  - Be careful with your first reboot
- Don't start your VMs during the host's first boot
  - The host will reboot and abort VMs installations

Outline

No kernel!

- No PXE boot

  ```
  "/system/aii/nbp" = null;
  ```

- No need for DHCP configuration

  ```
  "/system/aii/dhcp" = null;
  ```

- No Anaconda phase
- No partitions
- No filesystems
- No hardware

Filesystem template on an OpenVZ guest

```
template filesystems/openvz/guest;
“/system/blockdevices” = nlist();
“/system/filesystems” = list();
```

Hardware template on an OpenVZ guest

```
structure template hardware/machine/openvz/guest;
```

Hardware template on an OpenVZ guest

```
structure template hardware/machine/openvz/guest;

# That's it, really
```

Osinstall plug-in

- We only need the post_reboot script
- There is a plugin, aii-openvz for that
  - It has no schema, or fields

```
variable AII_OSINSTALL_GEN =
    'quattor/aii/openvz/config';
```

- Use AII as usual
- The component places the correct post_reboot script on /etc/init.d
- No NBP configuration or hooks
- Your usual post_reboot hooks will run

Outline

No hardware

- Disable ncm-network
- Disable ncm-lmsensors
- Disable ncm-modprobe

No Anaconda during installation

- Timezone is not set during installation
  - Use ncm-symlink for that

Outline

Virtual Ethernet

- Slower, less secure than OpenVZ's default network implementation
- But it can receive broadcasts
- The component doesn't handle such devices
- The veth device has to be integrated in a bridge on the host
  - ncm-network already handles this

Useful links

- UAM pre-built images for SL4 and SL5
  http://gridp03.ft.uam.es/ostemplates/
- How to build your own images http://openvz.org/pipermail/users/2007-November/001372.html