LPC 2012: libvirt-sandbox

Building application sandboxes on top of LXC and KVM with libvirt

Daniel P. Berrangé <berrange@redhat.com>

Talk Overview

- Introduction
- Architecture
- Examples

Introduction

Application Sandboxes

- Isolate general purpose applications
- Target specific use cases
- Variety of approaches
  - Seccomp – Linux syscall restriction
  - Java VM – bytecode verification
  - SELinux – MCS isolation
  - Virtualization – OS separation
- Multiple layers of defence

Virtualization

- Full machine or container virtualization
- The 3 common virt use cases
  - Cloud, server, desktop
- The 4th use case
  - Application development toolkit
  - Facility for application developers
  - Sysadmin application sandboxing

Sandbox Goals

- API for constructing sandboxes
- Agnostic to type of virtualization
- Isolated from hardware config
- No separate OS installation
  - Forget JEOS (Just Enough OS)
  - Try NAOS (No Additional OS)
  - Inherit host FS
  - Custom overrides sub-trees

Architecture

Low Level Infrastructure

- Kernel
  - Namespaces (pid,uts,fs,net,user,ipc)
  - CGroups
  - KVM
  - SELinux
- QEMU
  - Device model emulation (for KVM)

Low Level APIs

- libvirt: virtualization management API
- C library, lang bindings, object mappings
- KVM, LXC, VMware ESX, VirtualBox, Hyper-V, Parallels, etc
- Generic API & config model
- Hypervisor specific hardware

High Level APIs

- libvirt-sandbox C library
- Built on libvirt-gconfigibvirt-gobject
- Access from Python/Perl/JavaScript

High Level Tools

- virt-sandbox
  - For ad-hoc CLI application sandbox
  - Exposes most of C API via CLI args
- virt-sandbox-service
  - For persistent system services
  - Sets up custom environment
  - Create systemd unit files

Examples

virt-sandbox

- Simple invocation (R/O root, no network)
  - virt-sandbox [OPTIONS] BINARY [ARGS...]
- Choose virt driver using option
  - --connect LIBVRT-URI (or -c LIBVIRT-URI)
- Run 'date' inside LXC
  - virt-sandbox -c lxc:/// /bin/date
- Run 'cat /proc/cpuinfo' inside KVM
  - virt-sandbox -c qemu:///session /bin/cat /proc/cpuinfo

virt-sandbox

- Bind host files/dirs to guest R/W
  - --host-bind GUEST-PATH=HOST-PATH
- Create empty /home/fred with tmp dir
  - --host-bind /home/fred=
- Create /home/fred from /tmp/home
  - --host-bind /home/fred=/tmp/home
- Create /home/fred from /tmp/home.img
  - --host-image /home/freq=/tmp/home.img

virt-sandbox

- Bind /etc/krb5.conf from /tmp/krb5.conf
  - --guest-bind /etc/krb5.conf=/tmp/krb5.conf
- Copy /home/fred/.firefox into guest
  - --include /home/fred/.firefox
- sandbox.img w/ firefox prof & krb5 conf
  - --host-image /tmp=/home/fred/sandbox.img
  - --guest-bind /etc/krb5.conf=/tmp/krb5.conf
  - --guest-bind /home/fred=/tmp/home
  - --include /home/fred/.firefox

virt-sandbox

- Add DHCP configured NIC
  - --network dhcp
- Add static configured NIC
  - --network address=192.168.1.1/255.255.255.0
- SELinux dynamic config
  - --security label=svirt_sandbox_t,dynamic
- SELinux static config
  - --security label=svirt_sandbox_t:s0:c123,c123;static

Example: Server Virtual Hosting

- Goal:
  - Deploy multiple Apache virtual hosts
  - Strong isolation between virtual hosts
- Solution:
  - One apache instance per virtual host
  - Run apache inside a sandbox

virt-sandbox-service

- virt-sandbox-service create -C -u httpd.service apache1
  - Config /etc/libvirt-sandbox/service/apache1.sandbox
  - Multiple unit files allowed
  - SystemD unit file
    - /etc/systemd/system/httpd@apache1.service
  - Create state directories or image
    - /var/lib/libvirt/filesystem/apache1
      - Chroot type directory
    - Examines rpm payload
    - Clone - /var and /etc config
    - Share /usr
  - Allocate unique MCS security label

virt-sandbox-service

- virt-sandbox-service start apache1
  - Starts service from config
- virt-sandbox-service stop apache1
  - Stop service
- virt-sandbox-service connect apache1
  - Connect admin debug shell to container
- virt-sandbox-service execute -C ifconfig apache1
  - Execute command within container
  - virt-sandbox-service.logrotate
    - /usr/bin/virt-sandbox-service execute -C /etc/cron.daily/logrotate $i

systemd

- systemctl start httpd@apache1.service
  - Starts one sandboxed service
- systemctl start httpd@.service
  - Starts all sandboxed services
- systemctl reload httpd.service
  - Trigger reload of all httpd@XXXX.service

Other Use cases

- Run mock within a container
- Run customer services on gluster nodes
- Run mysql within a container
- OpenShift work loads

Example: Audio Transcode

- Obtained 'ogg' from untrusted source
- Decode to 'raw' format
- Prevent all filesytem & network access
- Only R/W on stdin/out
  - virt-sandbox -c lxc:/// -- /usr/bin/oggdec -o - - \
    < /path/to/untrusted.ogg \
    > /path/to/trusted.raw

Example: mock RPM Build

- setgid binary for users in 'mock' group
- Installs chroot with target distro RPMs
- Runs RPM as 'mock' user inside chroot
- Problem:
  - RPM chroot install runs as 'root'
  - RPM %post/%pre scripts run as 'root'
  - 'root' user can escape any chroot
    - => Malicious %post/%pre scripts can escape chroot

Example: mock RPM Build

- Solution:
  - Install chroot using 'rpm' in a sandbox
  - %pre/%post scripts run as 'root'
  - 'root' cannot escape from sandbox
    - => %pre/%post scripts cannot escape

LPC 2012: libvirt-sandbox

http://libvirt.org/