UTOSC 2012

Creating a Cloud-Like Infrastructure With FreeBSD

Shawn Webb
shwebb@wayfair.com

wayfair

Who Am I

- **Independent security researcher**
- **Tech blogger**
  - http://0xfeedface.org/
- **Software Engineer and Security Analyst for Wayfair LLC**
- **Disclaimer: any beliefs, opinions, etc. are mine and do not necessarily reflect those of my employer**

FreeBSD®

What's Covered

- **Definitions**
- **What is virtualization?**
- **Quick history**
  - Jails, virtual networking, and ZFS
- **Why jails?**
- **Setting up virtual networking**
- **Basic jailing**
- **Combining virtual network, jailing, and ZFS**
- **Future work**

Definitions

- **FreeBSD: An awesome POSIX-compliant 4.4BSD-Lite-based operating system**
- **Virtual network: a completely virtualized TCP/IP stack for use within the OS itself**
- **ZFS: The God of filesystems, developed by Sun. Oracle closed source after zpool v28.**
- **Jail: OS-based virtualization**
- **Drupal: A popular opensource CMS**

What is Virtualization?

- **The process by which one or more virtual machines run on a single physical server**
- **Hardware virtualization virtualizes everything, from hardware to OS**
  1. Xen, VirtualBox, VMWare, etc.
- **OS-level virtualization virtualizes only the OS**
  1. VMs share the same kernel
  2. VMs share the hardware

History of Jails

- **Introduced in FreeBSD 4.x by Poul-Henning Kamp**
- **Continuously being improved**
- **Secure replacement for chroot**
- **OS-based virtualization**
- **Inspired Solaris' Zones/Containers**

History of Virtual Networking

- **Called VIMAGE (or vnet)**
- **Work started in 7-CURRENT**
- **Official feature in 9-RELEASE/9-STABLE**
- **Analogous to Solaris Crossbow**
  1. Not as feature-complete as Crossbow
- **Reasons to use VIMAGE**
  1. Network security
  2. NATing jails

History of ZFS

- **The God of filesystems**
- **ZFS first integrated on 06 April 2007**
- **zpool v28 merged into 8-STABLE and in 9-RELEASE/9-STABLE**
- **Many wonderful features**
  1. New, powerful features coming from Delphix and Joyent

Reasons to Combine all Three

- **Basic cloud-like infrastructure**
  1. ZFS for instant snapshots and clones
  2. vnet for VLANs, dedicated network stack for each jail (routing tables)
  3. Jails for VMs
- **Why jails?**
  1. Much more CPU efficient than VMs
  2. Very little disk space consumption
- **Gotcha's**
  1. No pf or ipf support
  2. Must use IPFW

Setting up Virtual Networking

- Special kernel config
  1. Enable VIMAGE, IPFW
- Set up firewall
- Set up NAT
  1. Not required, but useful

```
# Kernel Config
options VIMAGE
options IPFIREWALL
options IPDIVERT

# rc.conf NAT
gateway_enable="YES"
firewall_enable="YES"
firewall_type="OPEN" # Change!
natd_enable="YES"
natd_interface="em0" # Change!
natd_flags=""
```

Setting up Virtual Networking

- Create bridge
- epair devices
  1. Pair of two ifconfig-able devices (epair[n]{a,b})
  2. Two ends of an ethernet cable
  3. Plug one end into bridge
  4. Plug other end into jail

```
ifconfig bridge0 create
ifconfig epair0 create
ifconfig bridge0 inet
192.168.2.1
ifconfig bride0 addm epair0a
ifconfig epair0a up

# Next commands assume jail is
# already booted up

ifconfig epair0b vnet [jail]
jexec [jail] ifconfig epair0b
inet 192.168.2.2
jexec [jail] route add default
192.168.2.1
```

Setting up Basic Jailing

- Use ZFS to create template jail dataset
  1. Create snapshot
  2. Attack of the clones
- Install world/distribution
- Install ports tree
- Install ports

```
# Initial installation
D=/jails/template
zfs create -omountpoint=/jails \
tank/jails
zfs create tank/jails/template
cd /usr/src
make installworld DESTDIR=$D
make distribution DESTDIR=$D
portsnap -p $D/usr/ports fetch \
extract

# Set DNS resolution
echo 'nameserver 8.8.8.8' > \
$D/etc/resolv.conf

# Set up temporary vnet
ifconfig bridge0 create
ifconfig bridge0 inet 192.168.2.1
ifconfig epair0 create
ifconfig bridge0 addm epair0a
ifconfig epair0a up
```

```
# Start the jail and set up networking in it
mount -t devfs devfs /jails/template/dev
mount -t nullfs /usr/ports /jails/template/usr/ports
jail -c vnet host.hostname=template name=template path=/jails/template \
persist
ifconfig epair0b vnet template
jexec template ifconfig epair0b inet 192.168.2.2
jexec template route add default 192.168.21.1

# Install ports
jexec template sh
*** NOW IN JAIL ***
cd /usr/ports/security/sudo
make install clean distclean

*** EXIT JAIL ***
# Snapshot for clones
zfs snapshot tank/jails/template@date

# New jail:
zfs clone tank/jails/template@date tank/jails/newjail
```

Not So Dumb After All

- **Dummynet: ipfw module to shape traffic**
- **Combine vnet with dummynet**
  1. Bandwidth limitation
  2. Queuing
  3. Simulated packet loss
  4. Packet delays
- **Disclaimer: dummynet can be buggy when mixed with vnet**
  1. Use only one dummynet config per pipe

So Many Commands!

- **A lot of initial work**
- **Takes a lot of time**
- **Problems:**
  1. FreeBSD's rc.d does not support vnet jails
  2. People reporting kernel panics destroying epair devices
     - I have had one or two kernel panics
  3. XENHVM + VIMAGE doesn't work

Making it Easy

- **I've written a Drupal 7 module to admin vnet jails**
- **Plans:**
  1. epair ifconfig aliases
  2. Reporting
  3. Privilege separation
  4. External API
  5. Make vnet optional
  6. Support IPv6
  7. Manage multiple jail hosts
- **Will not support non-ZFS setups**
- **https://github.com/lattera/drupal-jailadmin**
- **In Ports: http://www.freshports.org/www/drupal7-jailadmin**
- **Version 0.3 released**
- **C version soon**
  1. I wrote it for my GUI-less netbook

Real-World Examples

- **Wayfair will be switching to using vnet jails for development environments**
  1. Currently using hundreds of FreeBSD VMs in a XenCenter environment
  2. Will restructure to:
     - Single FreeBSD XC VM
     - Run Drupal 7 + www/drupal7-jailadmin ports
     - Hundreds of vnet jails

Future Work

- **XENHVM kernel and VIMAGE don’t mix well**
- **Obvious rc.d support**
- **Dtrace support**
  1. Like Solaris Zones
  2. Metrics
  3. Debugging
- **Complete virtualization**
  1. Certain resources still shared (i.e. 127.0.0.1)
- **Puppet classes**
- **Live migration**
- **KVM in a jail? (Need KVM first)**

Demo

- **Demo creating a template jail from scratch**
- **Demo using jailadmin Drupal module**
- **Demo using the ncurses-based command-line utility**

- **Comments/questions**
- **exit(0);**