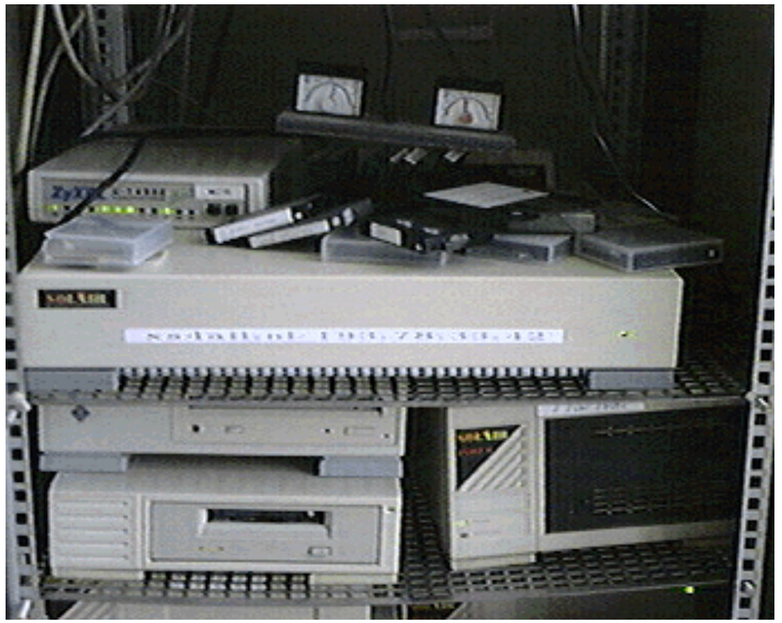
The OpenBSD hypervisor in the wild, a short story

Who dis

- Began at XS4ALL (ISP) in 1995
- Working for $vendor since 1998
- Started with FreeBSD in 1998
- Hosting / Co-Location since 1999



What about you?

- who is using OpenBSD?
- who is using vmm(4)/vmd(8)?
- who is on OpenBSD Amsterdam? ;)

How it all began

Always on the lookout for *easy* segmentation and virtualisation

- Started with and still using jails(8)
- Used bhyve(8)
- Using vmm(4)/vmd(8)

How it all began

- Spare rackspace
- Spare hardware
- Spare IP space
- Domain with something BSD
- Contributing back to the community
- How far can we take this
- Let's go!

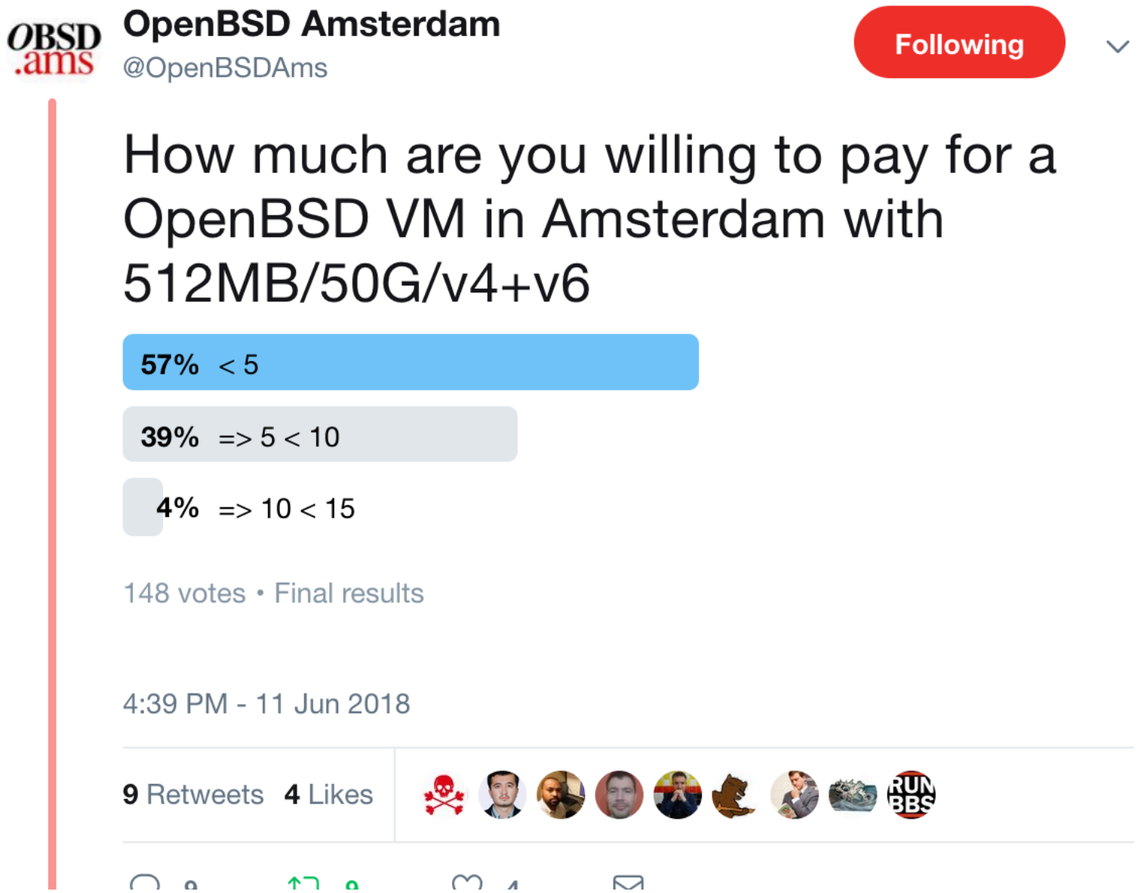
Where is it?

- Amsterdam!
- XS4ALL (KPN) Datacenter
- Dell R610 -> Foundry FLS448 -> Foundry MLX-4

Started on Twitter

[s1] Xeon(R) CPU E3-1220 V2 @ 3.10GHz w/ 8G RAM

What are people willing to pay



Proper machine online

[s2] Intel(R) Xeon(R) CPU E5-2620 v3 @ 2.40GHz w/ 32G RAM

First donation

OpenBSD Amsterdam
@OpenBSDAms

Following

Donated to the OpenBSD Foundation! Thank you all for making this possible!
#OpenBSD #RUNBSD

8:46 PM - 27 Jul 2018

5 Retweets **31** Likes

Statistics

- Latest donation €395
- 2018 €1850 (6 months)
- 2019 €3505
- 2020 €395 (YTD)
- Total: €5750
- Active Hosts: 10
- Active VMs: 330

OpenBSD Amsterdam
@OpenBSDAms

In the first month of the new year, and decade, we donated €395 to
the #OpenBSD Foundation, totalling €5750.

In January 17 new VMs were added and 15 VMs were renewed.

#OpenBSD #RUNBSD

10:18 AM · Jan 31, 2020 · TweetDeck

View Tweet activity

14 Retweets **34** Likes

What do you get?

- Opinionated VM

What do you get?

- Opinionated VM
- 512M RAM
- 50G Disk
- IPv4 assigned via DHCP
- IPv6 statically assigned (/56 is assigned to a host)
  - Host is gateway for each VM

Setup

BASE

Everything we use is in base

BASE

Everything we use is in base
- perl(1)
- vmm(4)/vmd(8)
- dhcpd(8)
- autoinstall(8)
- siteXX.tgz
- httpd(8)
- sensorsd(8)
- vi(1)

perl(1)

- /etc/vm.conf
- /etc/dhcpd.conf
- /var/www/htdocs/install/<MAC>-install.conf
- /etc/doas.conf
- user creation
- vm image creation

vm.conf(5)

```
socket owner :_vmdusers

switch "uplink_vlan931" {
    interface bridge931
}

vm "vm13" {
    disable
    owner alice
    disk "/var/vmm/vm13.qcow2"
    interface tap {
        switch "uplink_vlan931"
        lladdr fe:e1:bb:f1:c8:01
    }
}
```

dhcpd.conf(5)

```
option domain-name "openbsd.amsterdam";
option domain-name-servers 46.23.80.26;

subnet 46.23.93.0 netmask 255.255.255.0 {
    option routers 46.23.93.1;
    server-name "server8.openbsd.amsterdam";

    host vm13 {
        hardware ethernet fe:e1:bb:f1:c8:13;
        fixed-address 46.23.93.13;
        filename "auto_install";
        option host-name "puffy.openbsd.amsterdam";
    }
}
```

autoinstall(8)

/var/www/htdocs/autoinstall/fe:e1:bb:f1:c8:13-install.conf

```
# vm13-install.conf
System hostname = puffy.openbsd.amsterdam
Password for root = [password]
Which speed should com0 = 115200
Network interfaces = vio0
IPv4 address for vio0 = dhcp
IPv6 address for vio0 = 2a03:6000:6f64:613::13
IPv6 default router = 2a03:6000:6f64:613::1
Setup a user = alice
Password for user = [password]
Public ssh key for user = ssh-ed25519 AAAAC3N...U7KKt alice@domain.tld [password]
Which disk is the root disk = sd0
What timezone are you in = Europe/Amsterdam
Location of sets = http
Server = server8.openbsd.amsterdam
Set name(s) = -x* +xb* +xf* +site*
Continue anyway = yes
Continue without verification = yes
```

siteXX.tgz

- installurl(5):

```
https://cdn.openbsd.org/pub/OpenBSD
```

- rc.conf.local(8):

```
ntpd_flags="-s"
sndiod_flags=NO
```

- sysmerge(8).ignore:

```
/etc/ttys
```

siteXX.tgz

- install.site:

```
echo "/usr/sbin/syspatch && touch /etc/rc.local.forcereboot" >> /etc/rc.firsttime
```

- rc.local(8):

```
if [ -r /etc/rc.local.forcereboot ]; then
    rm -f /etc/rc.local.forcereboot
    printf '\n*** Reboot after CPU microcode/OS updates\n\n'
    sleep 2
    reboot
fi
```

httpd(8)

/etc/httpd.conf

```
server "default" {
    listen on * port 80
    root "/htdocs/autoinstall"
    location "/pub/OpenBSD/6.6/amd64/*" {
        root "/htdocs/6.6"
        request strip 4
        directory { auto index }
    }
}
```

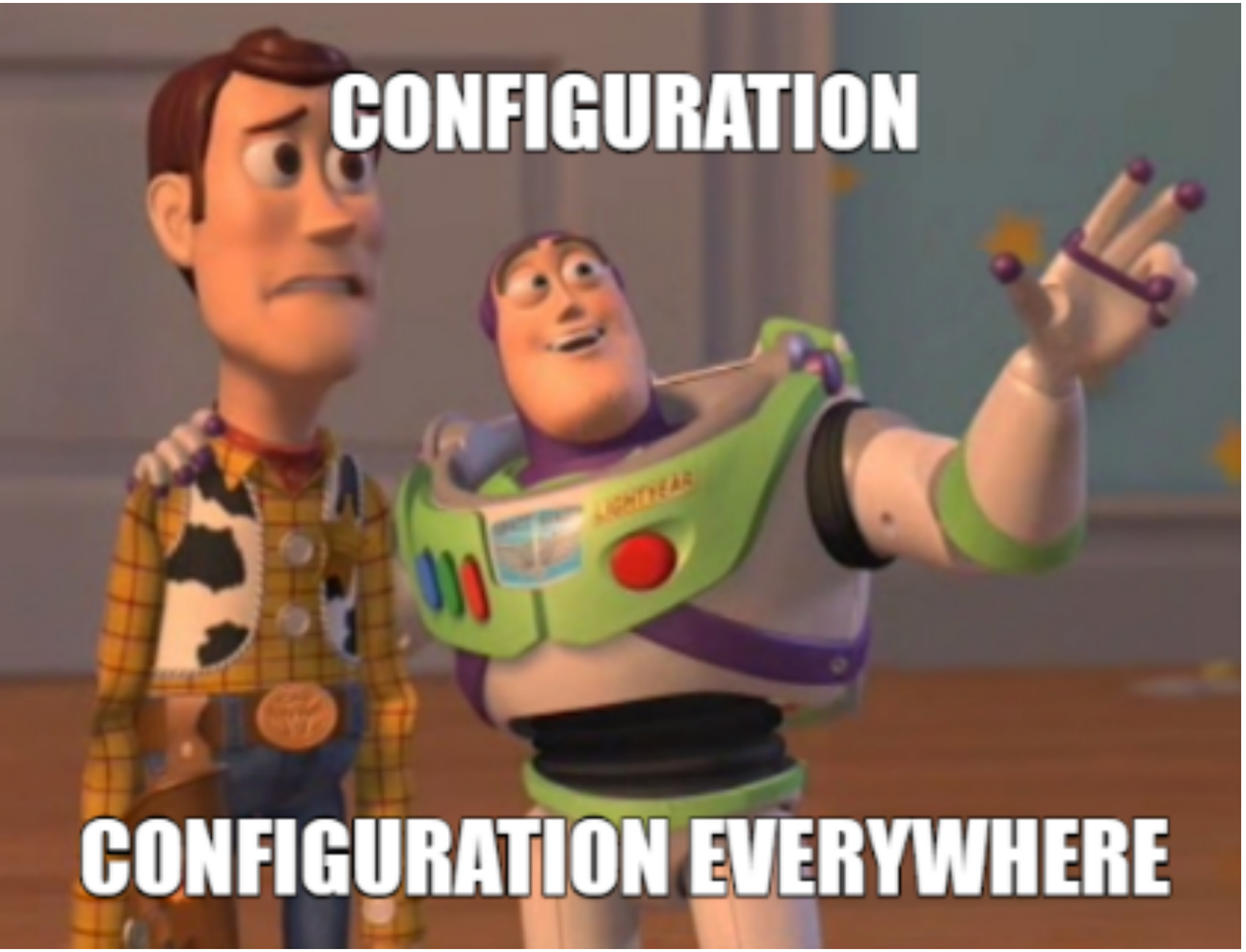
sensorsd(8)

/etc/sensorsd.conf

`drive:command=/etc/sensorsd/drive %t %n %2 %s`

```
#!/bin/sh
#
#       %t      The type of sensor.
#       %n      The sensor number.
#       %2      The sensor's current value.
#       %s      The sensor status.
#
#drive:command=/etc/sensorsd/drive %t %n %2 %s
#Subject: Sensor drive0 changed
#Raid state: drive0 online OK
echo "Current raid state: ${1}${2} ${3} ${4}" | mail -s "$(hostname) ${1}${2} ${4}" -r noreply@domain.tld mischa@domain.tld
```

Deploying!




```
server8:~ # cat _deploy.conf
# Server config for <MAC>-install.conf
SERVER="server8"
DOMAIN="openbsd.amsterdam"
# IP / MAC config
IP_PREFIX="46.23.93"
IP_START=100
IPV6_PREFIX="2a03:6000:6f64"
IPV6_START=600
MAC_PREFIX="fe:e1:bb:f1:c8"
# .conf locations
VMS="/home/mischa/vms"
ETC="/etc"
IMAGES="/var/vmm"
HTDOCS="/var/www/htdocs/default"
# vm.conf
MEMORY="512M"
DISKSIZE="50G"
FORMAT="qcow2"
VMDUSERS="_vmdusers"
SWITCH="uplink_vlan931"
INTERFACE="bridge931"
# dhcpd.conf
ROUTER="46.23.93.1"
DNS="46.23.80.26"
SUBNET="46.23.93.0"
NETMASK="255.255.255.0"
```

Deploy-flow

- form > email > file
- run deploy.pl on the host
- restart dhcpd
- reload vmd
- start vm
- run installer - Hit (A)

Form

Type-in your name *

Alice

email *

alice@example.com

and your SSH public key *

ssh-ed25519 FRhkxldn1...sDZUdP

hostname *

example

username *

alice

RAM

Standard 512M

HDD

Standard 50G

disk format

qcow2

referral code

OBSD-XXXX-XXXX

note?

I like VMs

Book it

Email > ~/vms/vm13.txt

```
date="2020/02/16"
payment=""
donated=""
name="Alice"
email="alice@domain.tld"
sshkey="ssh-ed25519 AAAAC3N...U7KKt alice@domain.tld"
hostname="puffy"
username="alice"
note=""
memory="512M"
disk2=""
format="qcow2"
referral=""
```

deploy.pl

```
server10:~ # deploy.pl
autoinstall(8) files:
            vm13 /var/www/htdocs/default/fe:e1:bb:f1:c8:13-install.conf created
useradd(8) creation:
            alice
vmm(4)/vmd(8) files:
            vm13 /var/vmm/vm13.qcow2 created (size 50G)
```

```
server10:~ # vmctl reload
server10:~ # rcctl restart dhcpd
dhcpd(ok)
dhcpd(ok)
server10:~ # vmctl start -c vm13
Connected to /dev/ttypk (speed 115200)
Copyright (c) 1982, 1986, 1989, 1991, 1993
        The Regents of the University of California.  All rights reserved.
Copyright (c) 1995-2019 OpenBSD. All rights reserved.  https://www.OpenBSD.org

OpenBSD 6.6 (RAMDISK_CD) #349: Sat Oct 12 11:03:52 MDT 2019
    deraadt@amd64.openbsd.org:/usr/src/sys/arch/amd64/compile/RAMDISK_CD
real mem = 520093696 (496MB)
avail mem = 500412416 (477MB)
mainbus0 at root
bios0 at mainbus0
acpi at bios0 not configured
cpu0 at mainbus0: (uniprocessor)
cpu0: Intel(R) Xeon(R) CPU X5690 @ 3.47GHz, 3459.93 MHz, 06-2c-02
...
sd0 at scsibus0 targ 0 lun 0: <VirtIO, Block Device, > SCSI3 0/direct fixed
sd0: 51200MB, 512 bytes/sector, 104857600 sectors
...
root on rd0a swap on rd0b dump on rd0b
erase ^?, werase ^W, kill ^U, intr ^C, status ^T
Welcome to the OpenBSD/amd64 6.6 installation program.
(I)nstall, (U)pgrade, (A)utoinstall or (S)hell? a
```

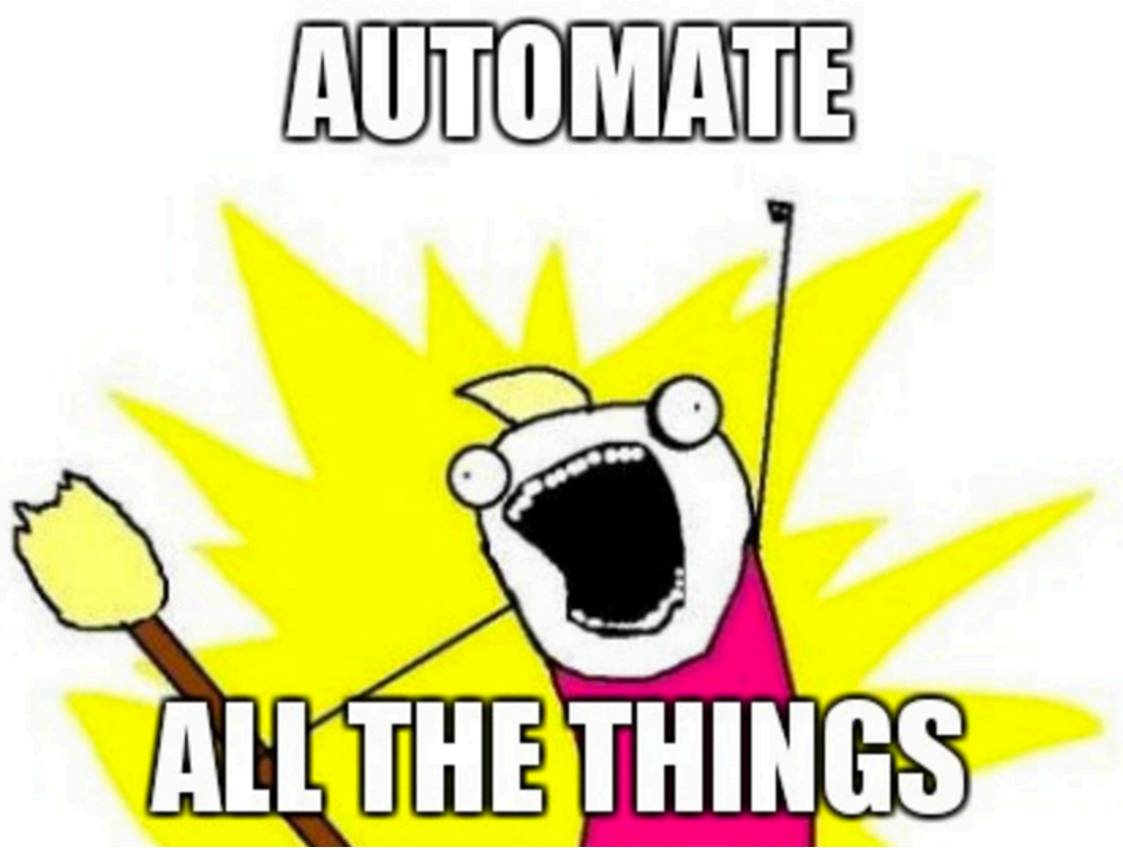
```
server10:~ # cat wrapper.sh
#!/bin/sh

deploy.pl
rcctl restart dhcpd
vmctl reload
sleep 10

./auto-deploy.exp $1
if [ $? ]; then
        vmctl stop -f $1
fi

deploy.pl
rcctl restart dhcpd
vmctl reload
sleep 10

./auto-start.exp $1
if [ $? ]; then
        echo
        echo "SUCCES!"
fi
```



expect TCL

```
server10:~ # cat auto-deploy.exp
#!/usr/local/bin/expect -f

set vmid [lindex $argv 0]
set timeout -1

spawn vmctl start -c $vmid
expect "(I)nstall, (U)pgrade, (A)utoinstall or (S)hell? "
sleep 1
send "a\r"
expect "(I)nstall, (U)pgrade, (A)utoinstall or (S)hell? "
```

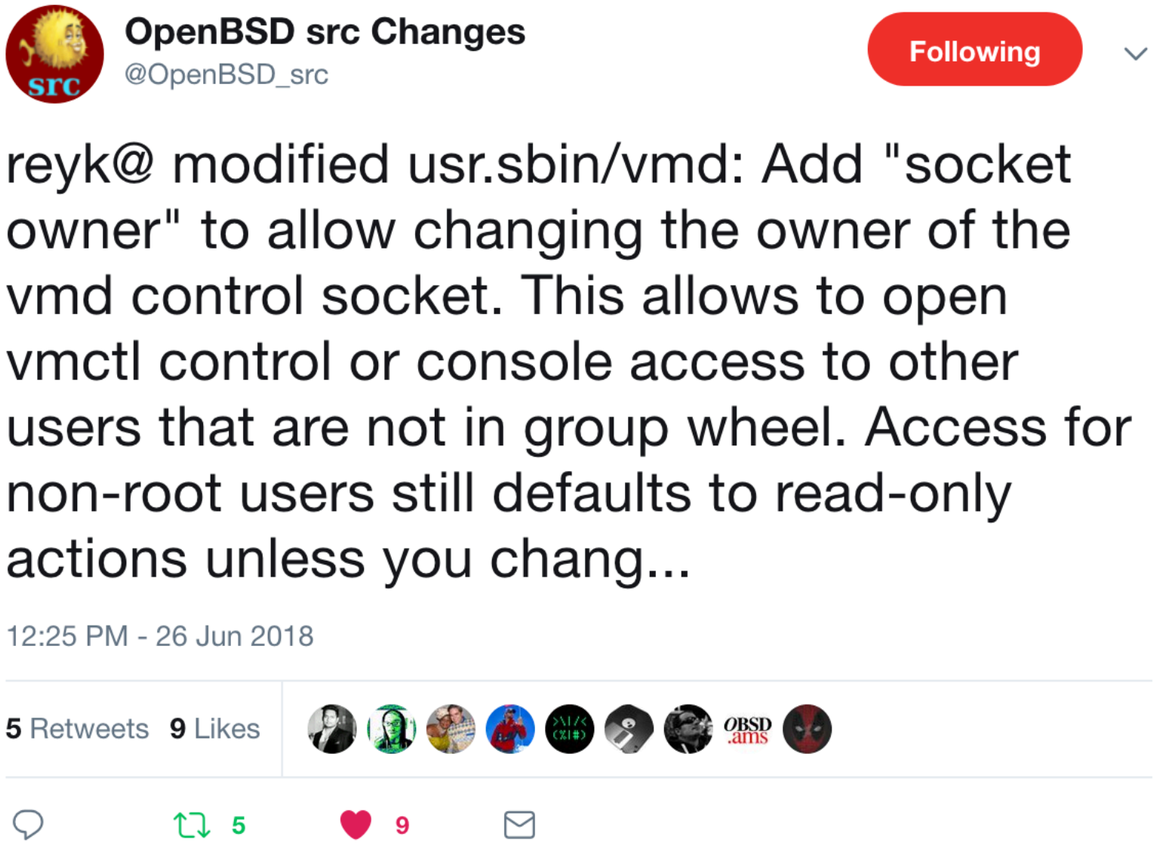
```
server10:~ # cat auto-start.exp
#!/usr/local/bin/expect -f

set vmid [lindex $argv 0]
set timeout -1

spawn vmctl start -c $vmid
expect "login:"
```

What did we find

socket owner




```
socket owner :group
    Set the control socket owner to the specified group.
```

socket https://marc.info/?l=openbsd-cvs&m=153003284400760&w=2

tap(4) interfaces

```
$ cd /dev
$ ls -al tap*
crw-------  1 root  wheel   93,   0 Apr 25 09:28 tap0
crw-------  1 root  wheel   93,   1 Apr 25 09:28 tap1
crw-------  1 root  wheel   93,   2 Apr 25 09:28 tap2
crw-------  1 root  wheel   93,   3 Apr 25 09:28 tap3

$ for i in $(jot 50 4 50); do doas sh MAKEDEV tap$i; done
```

share password??

```
jot -rcs '' 20 33 126
```

```
-r          Generate random data.  By default, jot generates sequential
            data
-c          This is an abbreviation for -w %c.
-w word     Print word with the generated data appended to it.  Octal,
            hexadecimal, exponential, ASCII, zero-padded, and right-
            adjusted representations are possible by using the
            appropriate printf(3) conversion specification inside word,
            in which case the data is inserted rather than appended.
-s string   Print data separated by string.  Normally, newlines
            separate data.
```

added to ~/.ssh/authorized_keys

[ascii] http://www.asciitable.com/

stopping VMs

Used to do this with:

```
$ vmctl show | for i in $(awk '!/ID| - / {print $1}'); do doas vmctl stop $i; sleep 30; done
```

Now there is:

```
$ doas vmctl stop -aw
```

[-a] https://marc.info/?l=openbsd-cvs&m=153806854327569&w=2

starting VMs

```
$ vmctl show | for i in $(awk '!/ID/ {print $1}'); do doas vmctl start $i; sleep 30; done
```

Or

```
$ vmctl show | for i in $(awk '!/ID/ {print $1}'); do doas vmctl start $i; sleep 90; done
```

arpq

```
$ sysctl net.inet.ip.arpq.drops
net.inet.ip.arpq.drops=524
$ sysctl net.inet.ip.arpq.maxlen
net.inet.ip.arpq.maxlen=50

$ doas sysctl net.inet.ip.arpq.maxlen=1024
```

What users experience

Clock drift

Clock drifts, sometimes more severe.

```
# Sync clock every 15 minutes
*/15 * * * * /usr/sbin/rdate -s pool.ntp.org
```

clock https://openbsd.amsterdam/clock.html

High CPU interrupts

VMs have a constant high intr CPU state:

```
CPU states: 0.0% user, 0.0% nice, 0.1% sys, 0.0% spin, 98.0% intr, 1.9% idle
```

---

intr https://marc.info/?l=openbsd-misc&m=154834783313341&w=2

Connectivity drops

Cron + tmux

```
@reboot /usr/bin/tmux new -d 'while true; do ping -i5 <gateway>; done' \;
```

Unresponsive VM

When `vmctl stop -f <vm-name>` doesn't work. [6.5]

/etc/doas.conf

```
permit nopass <vm-owner> as root cmd pkill args -9 -f <vm-name>
```

User runs:

```
$ doas pkill -9 -f vm13
```

[6.5] vmctl stop <vm-name> -f (https://marc.info/?l=openbsd-cvs&m=155916557307145&w=2)

Redundancy

Layer2 - carp(4)

```
vm1:~ # cat /etc/hostname.carp188
inet 46.xx.xx.51 255.255.255.0 NONE vhid 188 pass cisco carpdev vio0 advskew 100 carppeer 46.xx.xx.50

vm2:~ # cat /etc/hostname.carp188
inet 46.xx.xx.51 255.255.255.0 NONE vhid 188 pass cisco carpdev vio0 advskew 110 carppeer 46.xx.xx.49
```

Layer3 - relayd(8) + bgpd(8)

```
server10:~ # cat /etc/relayd.conf
vms = "46.23.94.51"

table <vms> { $vms ip ttl 1 }
router "vmd" {
        route 46.23.94.68/32
        forward to <vms> check script "/home/mischa/hc.sh"
}
```

```
server10:~ # cat hc.sh
#!/bin/sh
nc -zw3 $1 22 > /dev/null 2>&1
if [ $? -eq 0 ]; then
        exit 1
else
        exit 0
fi
```

```
server10:~ # cat /etc/bgpd.conf
AS 65065
router-id 65523
fib-update no
prefix-set mynetworks {
        46.23.94.68/32
}
network inet static
group "OpenBSDAms" {
        remote-as 65512
        neighbor ...
}
allow to group "OpenBSDAms"
```

Wishlist / Future

- iPXE
- no/less clock drift
- tackle connectivity drops

Couldn't be possible without!

Mike Larkin (@mlarkin2012)
Reyk Flöter (@reykfloeter)
Carlos Cardenas (@cobracmder)
Stefan Kempf
Claudio Jeker
Jasper Lievisse Adriaanse (@jasper_la)
Ori Bernstein (@oribernstein)

Roman Zolotarev (@romanzolotarev)

47

Thank you!

More information https://openbsd.amsterdam
Deploy script https://git.high5.nl/deploy.pl
Twitter https://twitter.com/OpenBSDAms
Mastodon https://bsd.network/@OpenBSDAms

Just URLs

More information https://openbsd.amsterdam
Deploy script https://git.high5.nl/deploy.pl
Twitter https://twitter.com/OpenBSDAms
Mastodon https://bsd.network/@OpenBSDAms