The OpenBSD hypervisor in the wild, a short story... but getting longer

Who dis

- Began at XS4ALL 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 it all began

- Amsterdam!

- XS4ALL (KPN) Datacenter

- Dell R610 -> Foundry FLS448 -> Foundry MLX-4

Started on Twitter

OpenBSD Amsterdam
@OpenBSDAms

Following

#OpenBSD VPS #6 on #OpenBSD #VMM #RUNBSD

2:13 PM - 10 Jun 2018

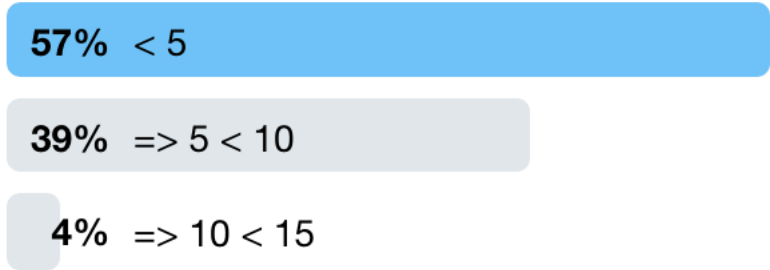[s1] Xeon(R) CPU E3-1220 V2 @ 3.10GHz w/ 8G RAM on OpenBSD 6.3

What are people willing to pay

OpenBSD Amsterdam
@OpenBSDAms

Following

How much are you willing to pay for a OpenBSD VM in Amsterdam with 512MB/50G/v4+v6

57% < 5

39% => 5 < 10

4% => 10 < 15

148 votes • Final results

4:39 PM - 11 Jun 2018

9 Retweets **4 Likes**

Proper machine online

OpenBSD Amsterdam
@OpenBSDAms

Following

For all the people wh voted, server #2 is ready!! #announcement #OpenBSD #RUNBSD

openbsd.amsterdam/server2.html

Start contributing to #OpenBSD while running a VPS!

> **OpenBSD Amsterdam** @OpenBSDAms
> How much are you willing to pay for a OpenBSD VM in Amsterdam with 512MB/50G/v4+v6
> Show this thread

7:59 AM - 1 Jul 2018

3 Retweets **4** Likes

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

First donation

Statistics

- Latest donation €610

- 2018 €1850 (6 months)

- 2019 €3505

- 2020 €5825

- 2021 €4965 (YTD)

- Total: €16.145

- Active Hosts: 13

- Active VMs: 495

OpenBSD Amsterdam
@OpenBSDAms

Will November be the month? 500 VMs? :)

19 new VMs were added and 28 VMs were renewed.

We donated €610 to the #OpenBSD Foundation, €16145 since we started.

Thank you, our users, and OpenBSD developers for an awesome OS!

Stay safe & healthy!

#RUNBSD in 2021

Thanks for donating 610,00 EUR to The OpenBSD Foundation

Share and inspire others to make an impact

7:41 AM · Nov 1, 2021 · TweetDeck

View Tweet activity

16 Retweets   **50** Likes

What do you get?

- Opinionated VM

What do you get?

- Opinionated VM

- ~~512M RAM~~ 1G 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)
- doas(1)
- vi(1)

CATS: ALL YOUR BASE ARE BELONG TO US.

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_veb931" {
    interface veb931
}


vm "vm13" {
    disable
    owner alice
    memory 1G
    disk "/var/vmm/vm13.qcow2" format qcow2
    interface tap {
        switch "uplink_veb931"
        lladdr fe:e1:bb:f1:c8:01
    }
}
```

dhcpd.conf(5)

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

subnet 46.23.90.0 netmask 255.255.255.0 {
    option routers 46.23.90.193;
    server-name "server8.openbsd.amsterdam";

    host vm13 {
        hardware ethernet fe:e1:bb:f1:c8:13;
        fixed-address 46.23.90.216;
        #fixed-address-ipv6 2a03:6000:6f64:613::216;
        filename "auto_install";
        next-server server8.openbsd.amsterdam;
        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 = autoconf
IPv6 address for vio0 = 2a03:6000:6f64:613::216
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) = +site*
Continue anyway = yes
Continue without verification = yes
```

siteXX.tgz

- **installurl(5)**:

```
https://mirror.openbsd.amsterdam/pub/OpenBSD
```

- **rc.conf.local(8)**:

```
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 "mirror.openbsd.amsterdam" {
    listen on * port 80
    tcp { nodelay, sack }
    log style forwarded
    root "/htdocs/autoinstall"
    location "/pub/OpenBSD/7.0/amd64/*" {
        root "/htdocs/7.0"
        request strip 4
        directory { auto index }
    }
}
```
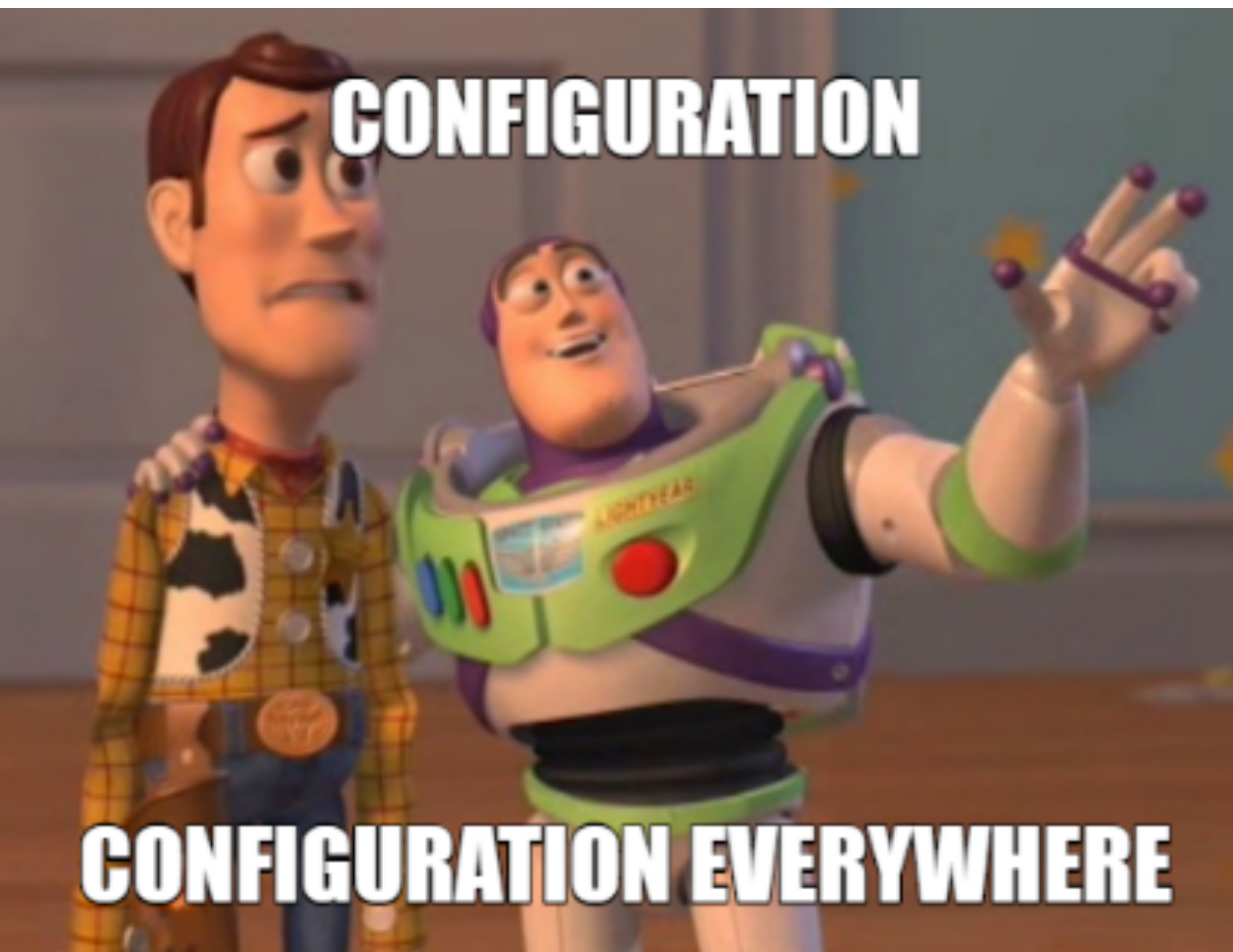
sensorsd(8)

/etc/sensorsd.conf

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

```sh
#!/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!

CONFIGURATION

CONFIGURATION EVERYWHERE

```
root@server8:~ # cat _deploy.conf
#
SERVER="server8"
DOMAIN="openbsd.amsterdam"
# IP / MAC config
IP_PREFIX="46.23.90"
IP_START=203
IPV6_PREFIX="2a03:6000:6f64"
IPV6_START=600
MAC_PREFIX="fe:e1:bb:6f:64"
# .conf locations
VMS="/home/mischa/vms"
ETC="/etc"
IMAGES="/var/vmm"
HTDOCS="/var/www/htdocs/default"
TEMPLATES="/home/mischa/templates"
# vm.conf
MEMORY="1G"
DISKSIZE="50G"
FORMAT="qcow2"
VMDUSERS="_vmdusers"
SWITCH="uplink_veb903"
INTERFACE="veb903"
# dhcpd.conf
ROUTER="46.23.90.193"
DNS="46.23.91.18"
SUBNET="46.23.90.192"
NETMASK="255.255.255.192"
```

Deploy-flow

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

Form

Book your VM

The service is provided "as is" and the operator disclaims all warranties with regard to this service including all implied warranties of merchantability and fitness. In no event shall the operator be liable for any special, direct, indirect, or consequential damages or any damages whatsoever resulting from loss of use, data or profits, whether in an action of contract, negligence or other tortious action, arising out of or in connection with the use or performance of this service.

Keep in mind that vmm(4)/vmd(8) is in **development**, things break!

Type-in your name *

| Alice |

RAM

| Standard 1G |

email *

| alice@example.com |

HDD

| Standard 50G |

disk format

| qcow2 |

and your SSH public key *

| ssh-ed25519 FRhkxldn1...sDZUdP |

referral code

| OBSD-XXXX-XXXX |

hostname *

| example |

note?

| I like VMs |

username (max 12 chars) *

| alice |

enter the following *, 3 + 2 =

| |

VAT # (optional)

| NL123456789B01 |

Book it

Email > ~/vms/vm13.txt

```
date="2020/02/16"
payment=""
subscription=""
donated=""
name="Alice"
email="alice@domain.tld"
sshkey="ssh-ed25519 AAAAC3N...U7KKt alice@domain.tld"
hostname="puffy"
username="alice"
note=""
memory="1G"
disk2=""
format="qcow2"
referral=""
vat=""
```

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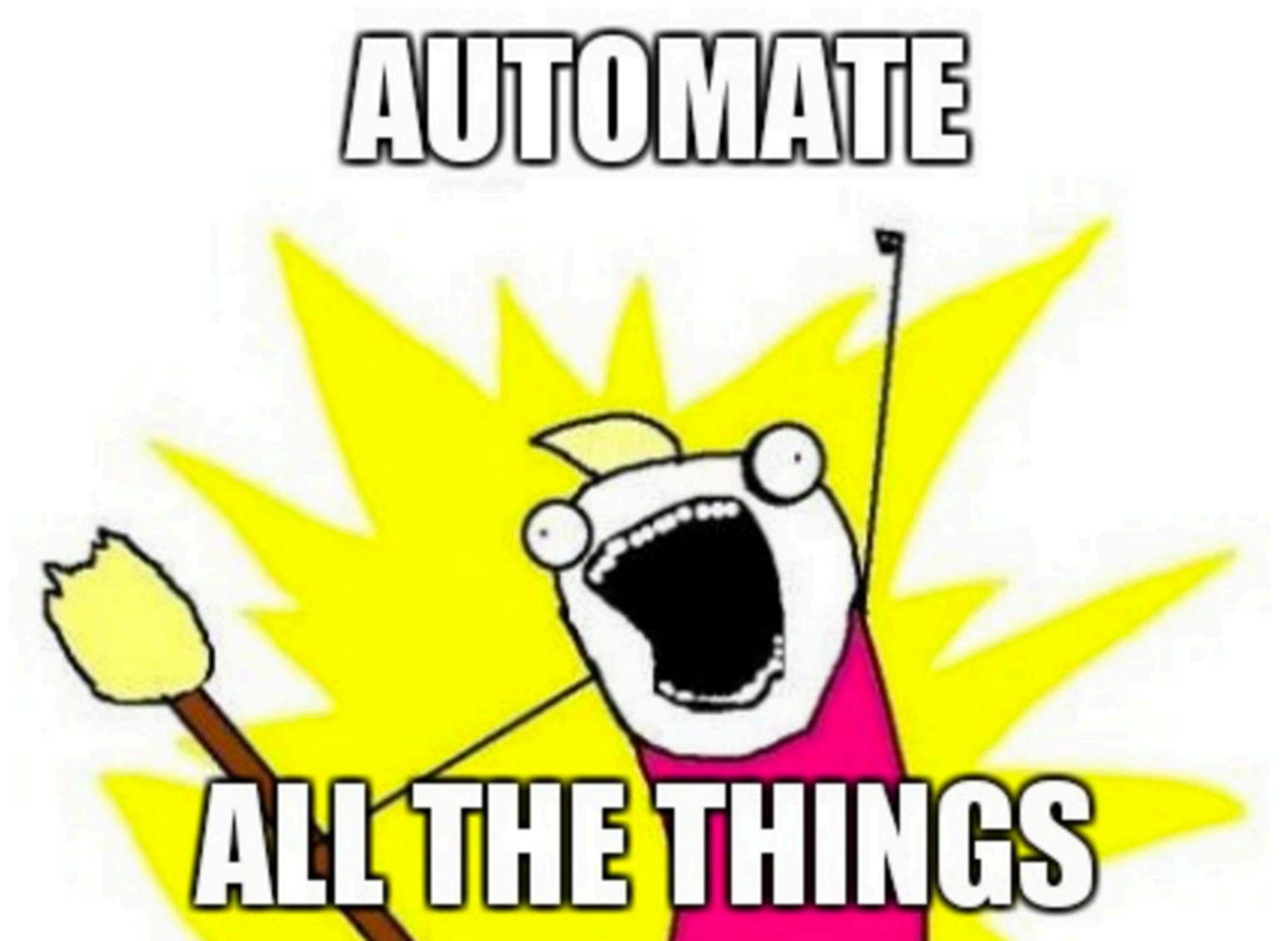
```
server10:~ # cat bin/launch.sh
#!/bin/sh

deploy.pl
rcctl restart dhcpd
vmctl reload
sleep 10


./auto-deploy.exp ${1}
if [ $? ]; then
        vmctl stop -f ${1}
        pushover.pl -m "$(hostname -s) ${1} failed deploy" -p 1
fi

deploy.pl
rcctl restart dhcpd
vmctl reload
sleep 10


./auto-start.exp ${1}
if [ $? ]; then
        echo
        echo "SUCCES!"
        notify.pl deployed ${1}
        pushover.pl -m "$(hostname -s) ${1} deployed"
fi
```


AUTOMATE ALL THE THINGS

expect TCL

```
server10:~ # cat bin/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 3
send "a\r"
expect {
        -re "Response file location?.*" {
                exit 1
        }
        "Performing non-interactive install..." {
                exp_continue
        }
        "(I)nstall, (U)pgrade, (A)utoinstall or (S)hell? " {
                exit 0
        }
}
```

```
server10:~ # cat bin/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

OpenBSD src Changes
@OpenBSD_src

Following ∨

reyk@ modified usr.sbin/vmd: Add "socket owner" to allow changing the owner of the vmd control socket. This allows to open vmctl control or console access to other users that are not in group wheel. Access for non-root users still defaults to read-only actions unless you chang...

12:25 PM - 26 Jun 2018

5 Retweets 9 Likes

💬   🔁 5   ❤️ 9   ✉️

```
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

Reyk Flöter
@reykfloeter

Replying to @blakkheim @NicoSchottelius and @datacenterlight

Better? Login is puffy@, cloud-agent can now generate a random password and write it as a comment into .ssh/authorized_keys - I shamelessly stole the idea from @OpenBSDAms.

```
$ ssh puffy@2a0a:e5c0:2:2:0:c8ff:fe68:bf16
Last login: Wed Jun  5 22:05:52 2019 from 2001:8e0:2002:8913:2eaa:c7ca:253c:d589
OpenBSD 6.5 (GENERIC.MP) #0: Wed Apr 24 23:38:54 CEST 2019

Welcome to OpenBSD: The proactively secure Unix-like operating system.

Please use the sendbug(1) utility to report bugs in the system.
Before reporting a bug, please try to reproduce it with the latest
version of the code.  With bug reports, please try to ensure that
enough information to reproduce the problem is enclosed, and if a
known fix for it exists, include that as well.

vm0200c868bf16$ head -1 .ssh/authorized_keys
# XXXXXXXXXXXXXXXX
vm0200c868bf16$ doas -s
doas (puffy@vm0200c868bf16) password: XXXXXXXXXXXXXXXX
vm0200c868bf16#
```

10:39 PM · Jun 5, 2019 · Twitter for iPhone

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

stopping VMs

```sh
#!/bin/sh
SLEEP=30
CPU=$(($(sysctl -n hw.ncpuonline)-2))

COUNTER=0
for i in $(vmctl show | sort | awk '/ running / {print $9}' | xargs); do
        VMS[${COUNTER}]=${i}
        COUNTER=$((${COUNTER}+1))
done
...
CYCLES=$((${#VMS[*]}/${CPU}+1))
echo "Stopping ${#VMS[*]} VMs on ${CPU} CPUs in ${CYCLES} cycle(s), waiting ${SLEEP} seconds after each cycle."

COUNTER=0
for i in ${VMS[*]}; do
        COUNTER=$((${COUNTER}+1))
        vmctl stop ${i}
        if [ $COUNTER -eq $CPU ]; then
                sleep ${SLEEP}
                COUNTER=0
        fi
done
sleep 30
```

starting VMs

```
#!/bin/sh
SLEEP=240
CPU=$(($(sysctl -n hw.ncpuonline)-2))

COUNTER=0
for i in $(vmctl show | sort | awk '/ - / {print $9}' | xargs); do
        VMS[${COUNTER}]=${i}
        COUNTER=$((${COUNTER}+1))
done

CYCLES=$((${#VMS[*]}/${CPU}+1))
echo "Starting ${#VMS[*]} VMs on ${CPU} CPUs in ${CYCLES} cycle(s), waiting ${SLEEP} seconds after each cycle."

COUNTER=0
for i in ${VMS[*]}; do
        COUNTER=$((${COUNTER}+1))
        vmctl start ${i}
        if [ $COUNTER -eq $CPU ]; then
                sleep ${SLEEP}
                COUNTER=0
        fi
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=512
```

What users experience

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

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 -xf "vmd: <vm-name>"`
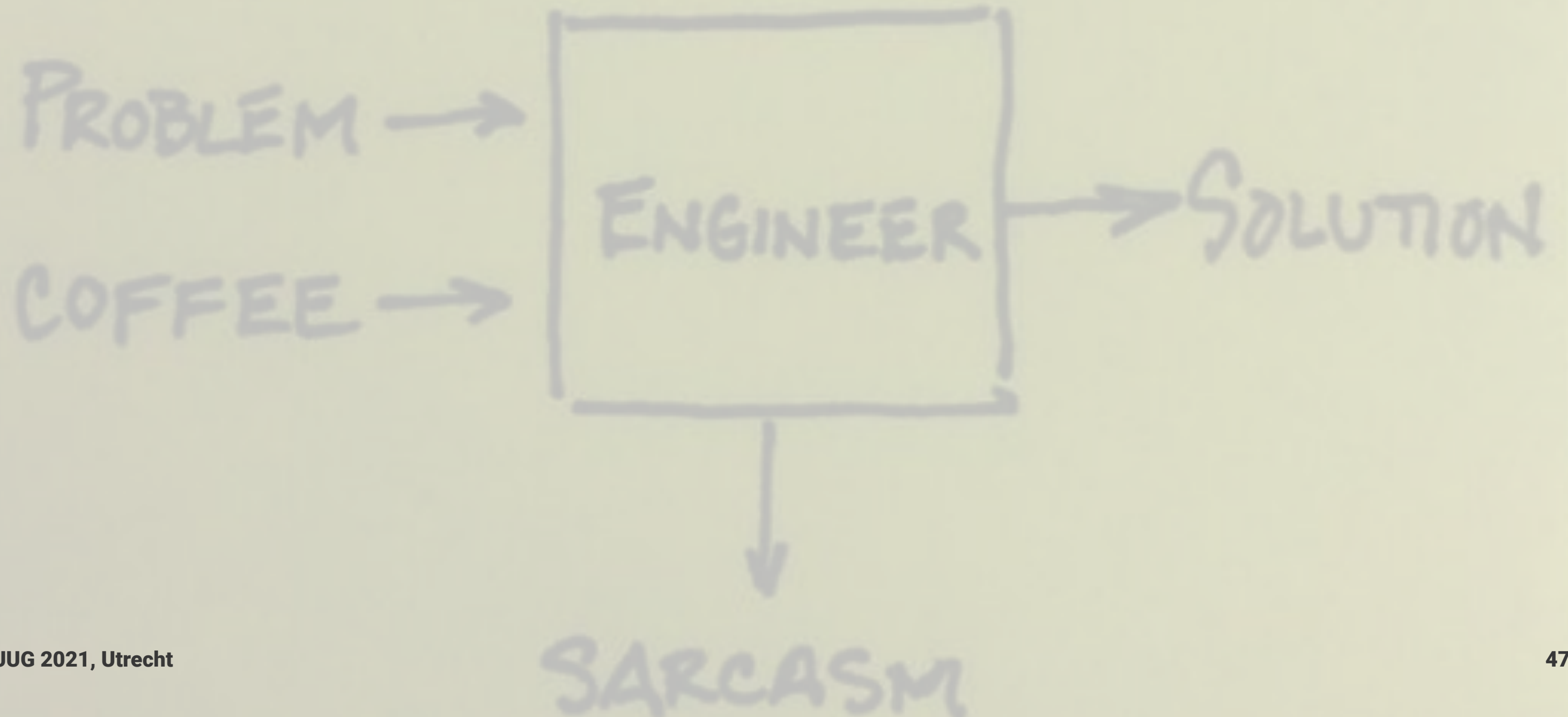
User runs:

`$ doas pkill -9 -xf "vmd: vm13"`

[6.5] https://marc.info/?l=openbsd-cvs&m=155916557307145&w=2

Good things

Good things

- Every release vmm(4)/vmd(8) gets better

- Clock sync issues, gone

- VM random hangs, gone

- Connectivity issues, almost gone

Excellent!

What else did we do?

PROBLEM →

COFFEE →

ENGINEER → SOLUTION

↓

SARCASM

Moving!

Moved all VMs from Dell R610 to Dell R620

```
Dell PowerEdge R610 w/ 2 x Intel(R) Xeon(R) CPU X5670 @ 2.93GHz
Dell PowerEdge R610 w/ 2 x Intel(R) Xeon(R) CPU X5675 @ 3.07GHz
Dell PowerEdge R610 w/ 2 x Intel(R) Xeon(R) CPU X5690 @ 3.47GHz
```
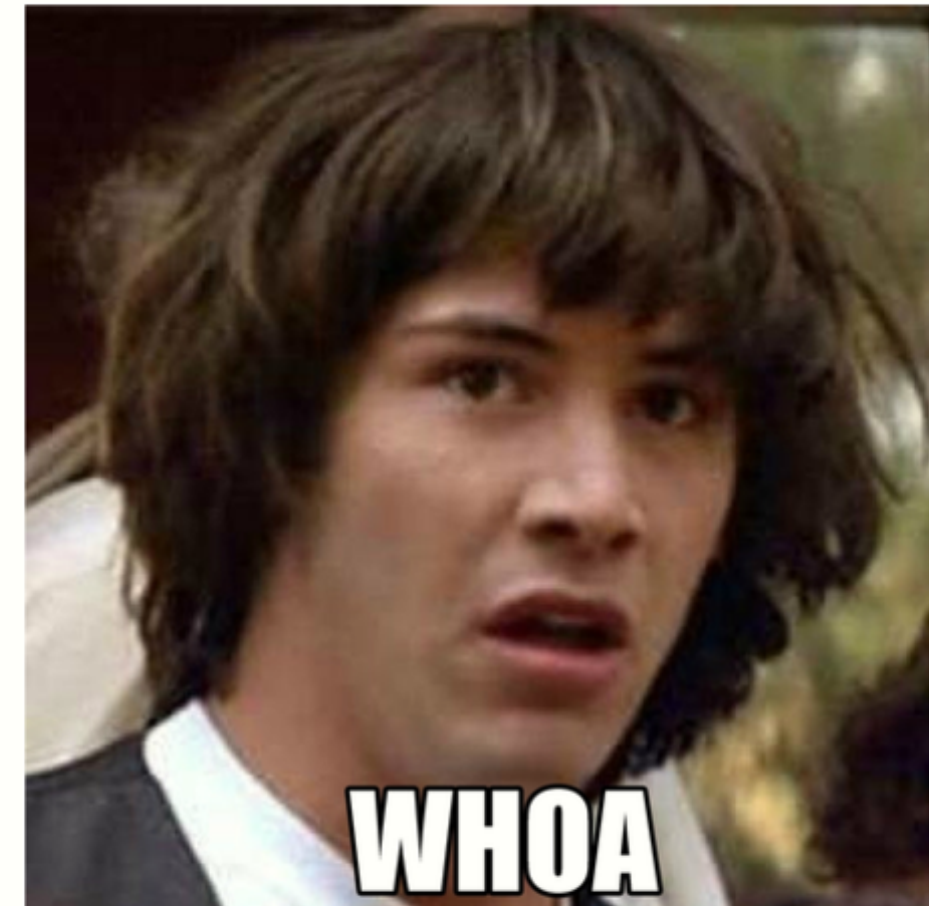
To

```
Dell PowerEdge R620 w/ 2 x Intel(R) Xeon(R) CPU E5-2630 0 @ 2.30GHz
- 128GB, PERC H710 Mini, 2 x 146GB 15K SAS, 2 x 900GB 10K SAS
```

R620 Extremely generous donations by CloudVPS

Moving!

- Started Saturday 2021-04-10

- Last VM moved Tuesday 2021-06-01

- Roughly 400 VMs moved

  - Converted from raw to qcow2

  - IPv4 and IPv6 changed

- Switched to routed setup [6.9]

[6.9] On Sunday 2021-05-16 from bridge(4)/vether(4) to veb(4)/vport(4)

Moving!

```
# mvvm.sh
Usage: mvvm.sh <source vm> <dest host> <dest vm>
```

Using rsync, ssh, doas and pushover

```
...
    rsync -ve 'ssh -i ~mischa/.ssh/id_ed25519' --rsync-path='doas rsync' \
    /var/vmm/${1}.qcow2 mischa@${2}.obsda.ms:/var/vmm/${3}.qcow2 && \
        pushover.pl -m "rsync ${3} to ${2}
done"
```

Moving! For reals this time.

Started shutting down VMs 2021-07-28 at 19:00 UTC

The last VM was started again around 20:30 UTC.

Left the DC around 20:45 UTC

Where is it now?

- Amsterdam! Of course!

- Digital Realty, Amsterdam Data Tower

- Dell R620 -> Arista DCS-7050T-64 -> Arista DCS-7050QX-32S

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)
David Gwynne (@dlgwynne)
Dave Voutila(@voutilad)
Roman Zolotarev (@romanzolotarev)

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