SmartOS Homerouter

fedi: @wiedi@fru.bar mail: sw@core.io

aka
abusing enterprise systems at home

aka
move slow and fix things

Agenda

- What are we building? Why do this?
- History (20+ years in the making)
- Side quests
  - APU Bring up
  - Freifunk & porting OLSRd
- SmartOS networking basics
- Current config walk through
- What’s next?

Goals

- Routing
- NAT
- Firewall
- DHCP / RA
- DNS resolver
- low power usage
- stability
- some nerdy bonus features

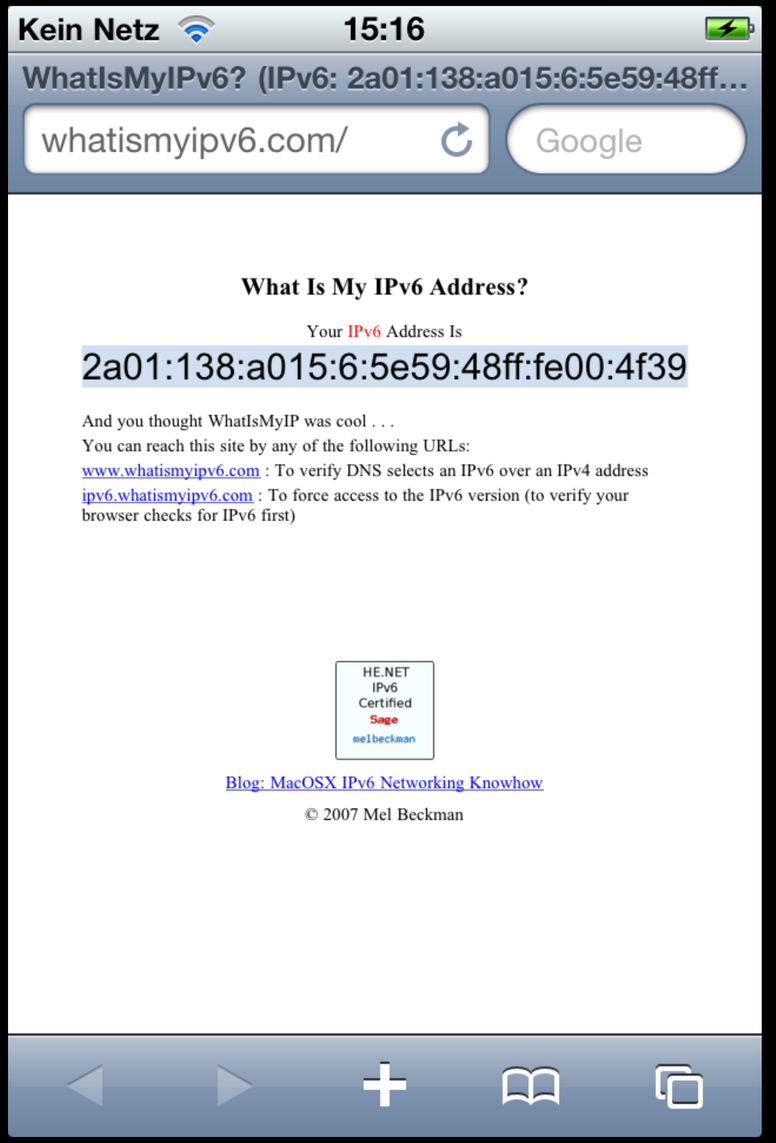
SmartOS because crossbow & zones

History
~2000

- Software
  - fli4l
  - boots from floppy into ro linux system
- Hardware
  - uplink: DSL (analog phone line)
  - lan: 100 mbit/s
  - Old 486 system

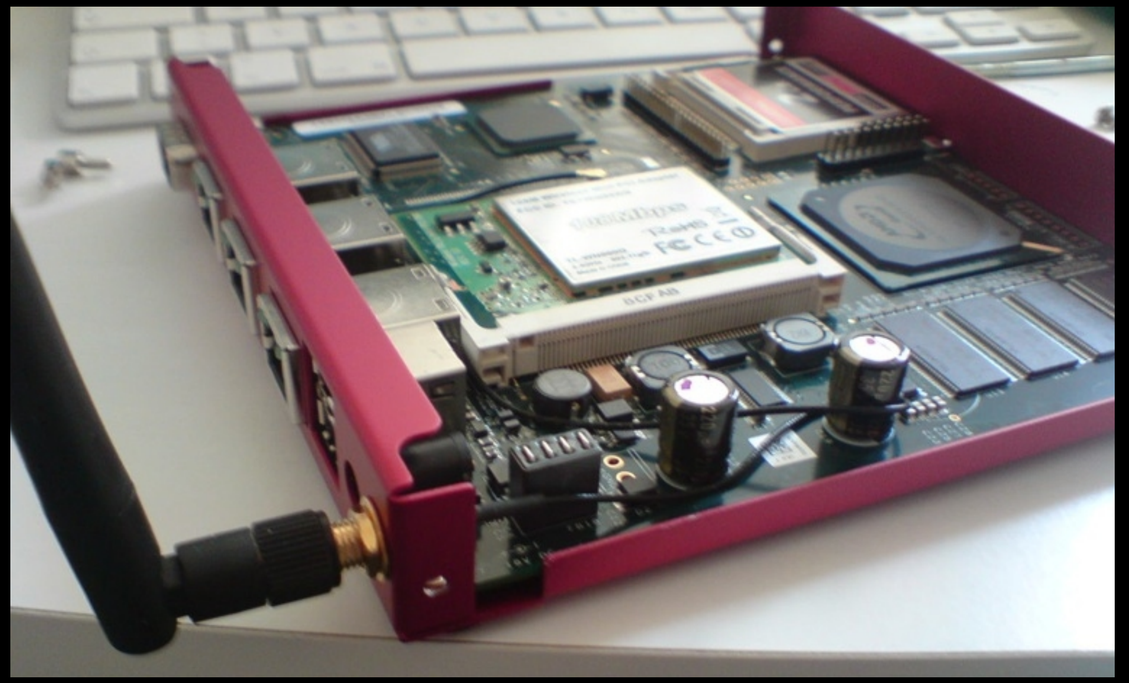
History
2007

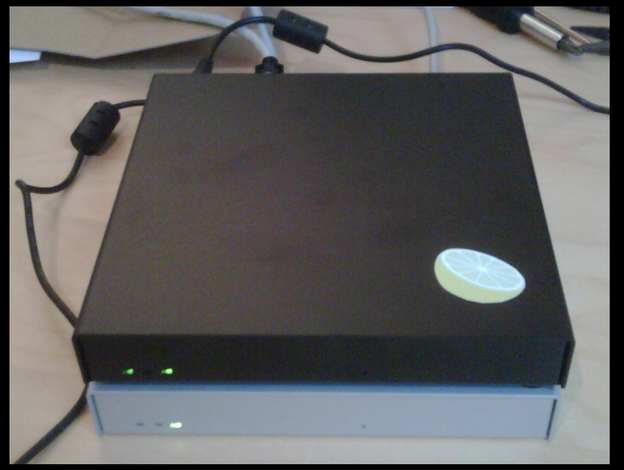
- OpenVPN between friends
- IPv6
  - first ::/48 delegation
- Quagga
  - OSPF for IPv4
  - RIPng for IPv6



History
2009

- Hardware
  - PC Engines ALIX
  - x86 embedded board
  - 3x 100mbit/s nics
  - CompactFlash
  - no VGA - but serial
- Software
  - voyage linux -> http://linux.voyage.hk/features





History
2014

- PC Engines APU 1C beta prototype board
  - new embedded AMD CPU
  - 3x 1000 mbit/s
  - SD Card
  - mSATA / miniPCI
  - coreboot opensource firmware
- Idea: use SmartOS as a routing os
- Problem: doesn’t boot

```
PC Engines board rev ID = 0/0
SPD has a invalid or zero-valued CRC
coreboot-SageBios_PCEngines_APU-8 Fri Feb 21 17:07:53 MST 2014 booting...
CBFS: ERROR: No file header found at 0x1ffb80 - try next aligned address: 0x1ffbc0.
CBFS: Could not find file 'pci1002,4390.rom'.
CBFS: ERROR: No file header found at 0x1ffb80 - try next aligned address: 0x1ffbc0.
CBFS: Could not find file 'pci1002,4385.rom'.
RTC: 02/21/2014 day of week=6
CBFS: ERROR: No file header found at 0x1ffb80 - try next aligned address: 0x1ffbc0.
CBFS: Could not find file 'pci1022,1700.rom'.
CBFS: ERROR: No file header found at 0x1ffb80 - try next aligned address: 0x1ffbc0.
CBFS: Could not find file 'pci1022,1701.rom'.
CBFS: ERROR: No file header found at 0x1ffb80 - try next aligned address: 0x1ffbc0.
CBFS: Could not find file 'pci1022,1702.rom'.
CBFS: ERROR: No file header found at 0x1ffb80 - try next aligned address: 0x1ffbc0.
CBFS: Could not find file 'pci1022,1703.rom'.
CBFS: ERROR: No file header found at 0x1ffb80 - try next aligned address: 0x1ffbc0.
CBFS: Could not find file 'pci1022,1704.rom'.
CBFS: ERROR: No file header found at 0x1ffb80 - try next aligned address: 0x1ffbc0.
CBFS: Could not find file 'pci1022,1718.rom'.
CBFS: ERROR: No file header found at 0x1ffb80 - try next aligned address: 0x1ffbc0.
CBFS: Could not find file 'pci1022,1716.rom'.
CBFS: ERROR: No file header found at 0x1ffb80 - try next aligned address: 0x1ffbc0.
CBFS: Could not find file 'pci1022,1719.rom'.
CBFS: ERROR: No file header found at 0x1ffb80 - try next aligned address: 0x1ffbc0.
CBFS: Could not find file 'pci10ec,8168.rom'.
CBFS: ERROR: No file header found at 0x1ffb80 - try next aligned address: 0x1ffbc0.
CBFS: Could not find file 'pci10ec,8168.rom'.
CBFS: ERROR: No file header found at 0x1ffb80 - try next aligned address: 0x1ffbc0.
CBFS: Could not find file 'pci10ec,8168.rom'.
CBFS: ERROR: No file header found at 0x1ffb80 - try next aligned address: 0x1ffbc0.
CBFS: Could not find file 'pci168c,002a.rom'.
PC-Engines/APU: adding bootorder to CBMEM
Reading data from file [bootorder]
bootorder file size 1000h
bootorder total size 1018h
bootorder struct in CBMEM
```

```
Displaying memory from 7efe7000 to 7efe8018
7efe7000: 54 4f 4f 42 00 10 00 00 62 6f 6f 74 6f 72 64 65  TOOB....bootorde
7efe7010: 72 00 28 00 a2 99 24 00 2f 70 63 69 40 69 30 63  r.(...$./pci@i0c
7efe7020: 66 38 2f 75 73 62 40 31 36 2c 32 2f 75 73 62 2d  f8/usb@16,2/usb-
7efe7030: 2a 40 31 0d 0a 2f 70 63 69 40 69 30 63 66 38 2f  *@1../pci@i0cf8/
7efe7040: 75 73 62 40 31 32 2c 32 2f 75 73 62 2d 2a 40 31  usb@12,2/usb-*@1
7efe7050: 0d 0a 2f 70 63 69 40 69 30 63 66 38 2f 75 73 62  ../pci@i0cf8/usb
7efe7060: 40 31 32 2c 32 2f 75 73 62 2d 2a 40 35 0d 0a 2f  @12,2/usb-*@5../
7efe7070: 70 63 69 40 69 30 63 66 38 2f 2a 40 31 31 2f 64  pci@i0cf8/*@11/d
7efe7080: 72 69 76 65 40 30 2f 64 69 73 6b 40 30 0d 0a 2f  rive@0/disk@0../
7efe7090: 70 63 69 40 69 30 63 66 38 2f 2a 40 31 31 2f 64  pci@i0cf8/*@11/d
7efe70a0: 72 69 76 65 40 31 2f 64 69 73 6b 40 30 0d 0a 2f  rive@1/disk@0../
7efe70b0: 70 63 69 40 69 30 63 66 38 2f 2a 40 31 31 2f 64  pci@i0cf8/*@11/d
7efe70c0: 72 69 76 65 40 32 2f 64 69 73 6b 40 30 0d 0a 2f  rive@2/disk@0../
7efe70d0: 72 6f 6d 40 67 65 6e 72 6f 6d 73 2f 70 78 65 62  rom@genroms/pxeb
7efe70e0: 6f 6f 74 2e 72 6f 6d 0d 0a 70 78 65 6e 31 0d 0a  oot.rom..pxen1..
7efe70f0: 73 63 6f 6e 31 0d 0a 2f 72 6f 6d 40 69 6d 67 2f  scon1../rom@img/
7efe7100: 73 65 74 75 70 0a 00 00 00 00 00 00 00 00 00 00  setup...........
7efe7110: 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00  ................
          00 ...
7efe8010: 00 00 00 00 00 00 00 00                          ........
SeaBIOS (version ?-20140221_170714-ubuntubuilderx64)
SeaBIOS (version ?-20140221_170714-ubuntubuilderx64)
Found coreboot cbmem console @ 7e150400
Found mainboard PC Engines APU
Relocating init from 0x000e8e61 to 0x7e1065e0 (size 39259)
Found CBFS header at 0xfffffb90
found file "bootorder" in cbmem
CPU Mhz=1000
Found 32 PCI devices (max PCI bus is 06)
Copying PIR from 0x7e160400 to 0x000f2790
Copying MPTABLE from 0x7e161400/7e161410 to 0x000f25a0 with length 1e4
Copying ACPI RSDP from 0x7e162400 to 0x000f2580
Copying SMBIOS entry point from 0x7e16d800 to 0x000f2560
Using pmtimer, ioport 0x808
Scan for VGA option rom
EHCI init on dev 00:12.2 (regs=0xf7f08420)
Found 1 lpt ports
Found 2 serial ports
AHCI controller at 11.0, iobase f7f08000, irq 11
EHCI init on dev 00:13.2 (regs=0xf7f08520)
EHCI init on dev 00:16.2 (regs=0xf7f08620)
Searching bootorder for: /pci@i0cf8/*@11/drive@0/disk@0
AHCI/0: registering: "AHCI/0: SuperSSpeed S238 16GB ATA-7 Hard-Disk (15258 MiBytes)"
Searching bootorder for: /rom@img/setup
Searching bootorder for: /rom@img/memtest
OHCI init on dev 00:12.0 (regs=0xf7f04000)
OHCI init on dev 00:13.0 (regs=0xf7f05000)
OHCI init on dev 00:14.5 (regs=0xf7f06000)
```

```
[0]> :c
SunOS Release 5.11 Version joyent_20140221T181910Z 64-bit
Copyright (c) 2010-2014, Joyent Inc. All rights reserved.
…
panic[cpu0]/thread=ffffffffbc3f4a0: assertion failed: pci_bus_res[secbus].dip
== NULL, file: ../../intel/io/pci/pci_boot.c, line: 2806

Warning - stack not written to the dump buffer
ffffffffbc81800 genunix:process_type+162bf0 ()
ffffffffbc818b0 pci_autoconfig:add_ppb_props+101 ()
ffffffffbc81a00 pci_autoconfig:process_devfunc+561 ()
ffffffffbc81a60 pci_autoconfig:enumerate_bus_devs+111 ()
ffffffffbc81a80 pci_autoconfig:pci_setup_tree+80 ()
ffffffffbc81ab0 pci_autoconfig:pci_enumerate+20 ()
ffffffffbc81ad0 unix:impl_bus_initialprobe+65 ()
ffffffffbc81b00 unix:impl_setup_ddi+dc ()
ffffffffbc81b20 genunix:create_devinfo_tree+cc ()
ffffffffbc81b30 genunix:setup_ddi+13 ()
ffffffffbc81b60 unix:startup_modules+f6 ()
ffffffffbc81b70 unix:startup+4a ()
ffffffffbc81bb0 genunix:main+9b ()
ffffffffbc81bc0 unix:_locore_start+90 ()
```

History
2014

- PC Engines quickly provides a new BIOS with potential fix
- `flashrom -p internal -c MX25L1605A/MX25L1606E -w BIOS.rom`
- triggers bug in flashrom

```
flashrom -w apu_bios_20140225.rom
flashrom v0.9.5.2-r1546 on Linux 3.2.0-4-amd64 (x86_64)
flashrom is free software, get the source code at http://www.flashrom.org

Calibrating delay loop... OK.
coreboot table found at 0x7e17cc00.
Found chipset "AMD SB7x0/SB8x0/SB9x0". Enabling flash write... OK.
Found Macronix flash chip "MX25L1605" (2048 kB, SPI) at physical address 0xffe00000.
Reading old flash chip contents... done.
Erasing and writing flash chip... Erase/write done.
Verifying flash... VERIFY FAILED at 0x00000000! Expected=0x4c, Read=0xcc, failed byte count from 0x00000
000-0x001fffff: 0xe1e3
Your flash chip is in an unknown state.
Get help on IRC at chat.freenode.net (channel #flashrom) or
mail flashrom@flashrom.org with the subject "FAILED: <your board name>"!
-------------------------------------------------------------------------------
DO NOT REBOOT OR POWEROFF!
```

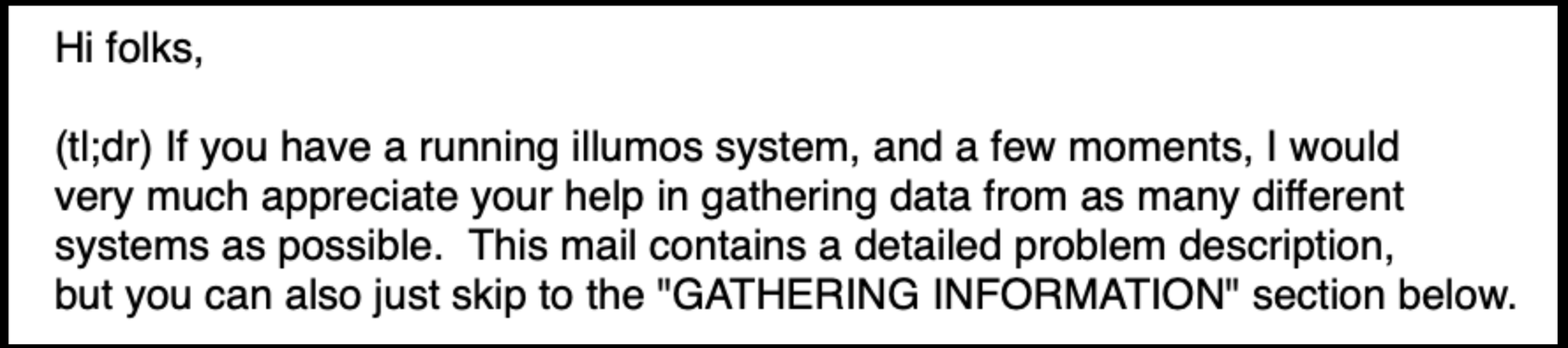
History
2014

- devs in #flashrom help us debug this on the spot
- fix commited to repo the same day

Amazing work and thanks all around!

History
2015 / 2016

- FreeBSD fixed but SmartOS still doesn’t boot…
- 2015 June - „ACPI MCFG Tables and You“ collecting data on the mailinglist
- 2015 Nov - jclulow has a patch - with that we can successfully boot our APU
- 2016 May - 6859 missing MCFG table should lead to I/O PCIe config access

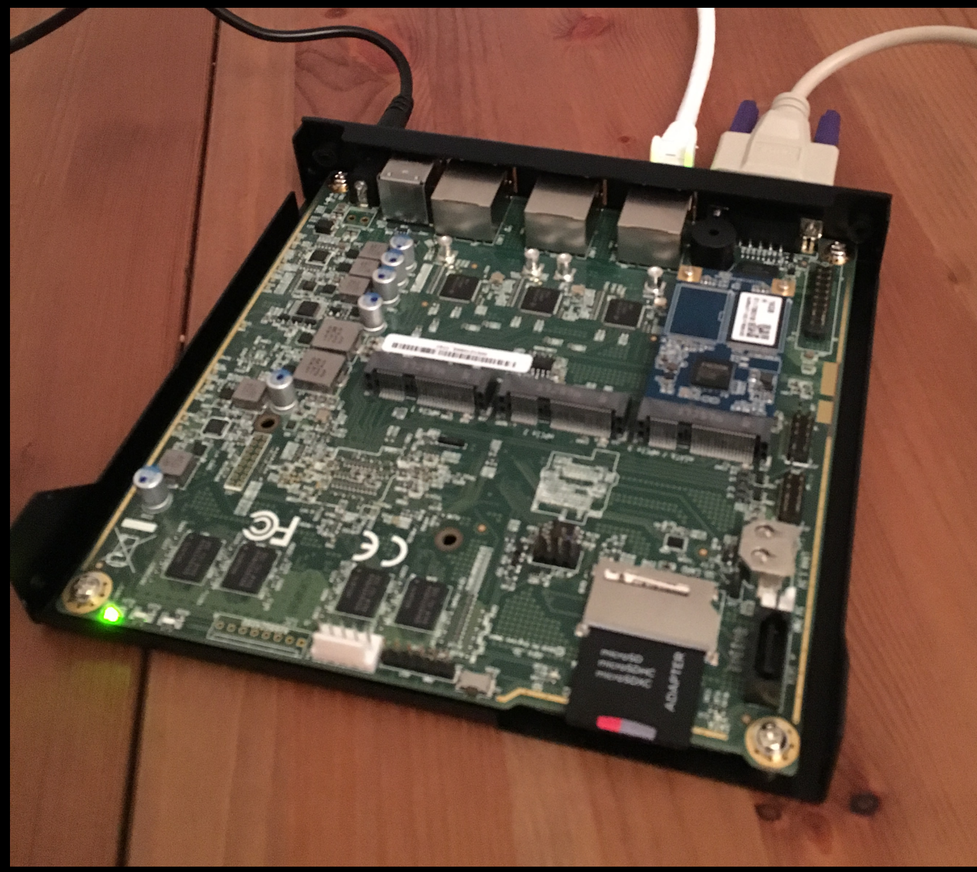


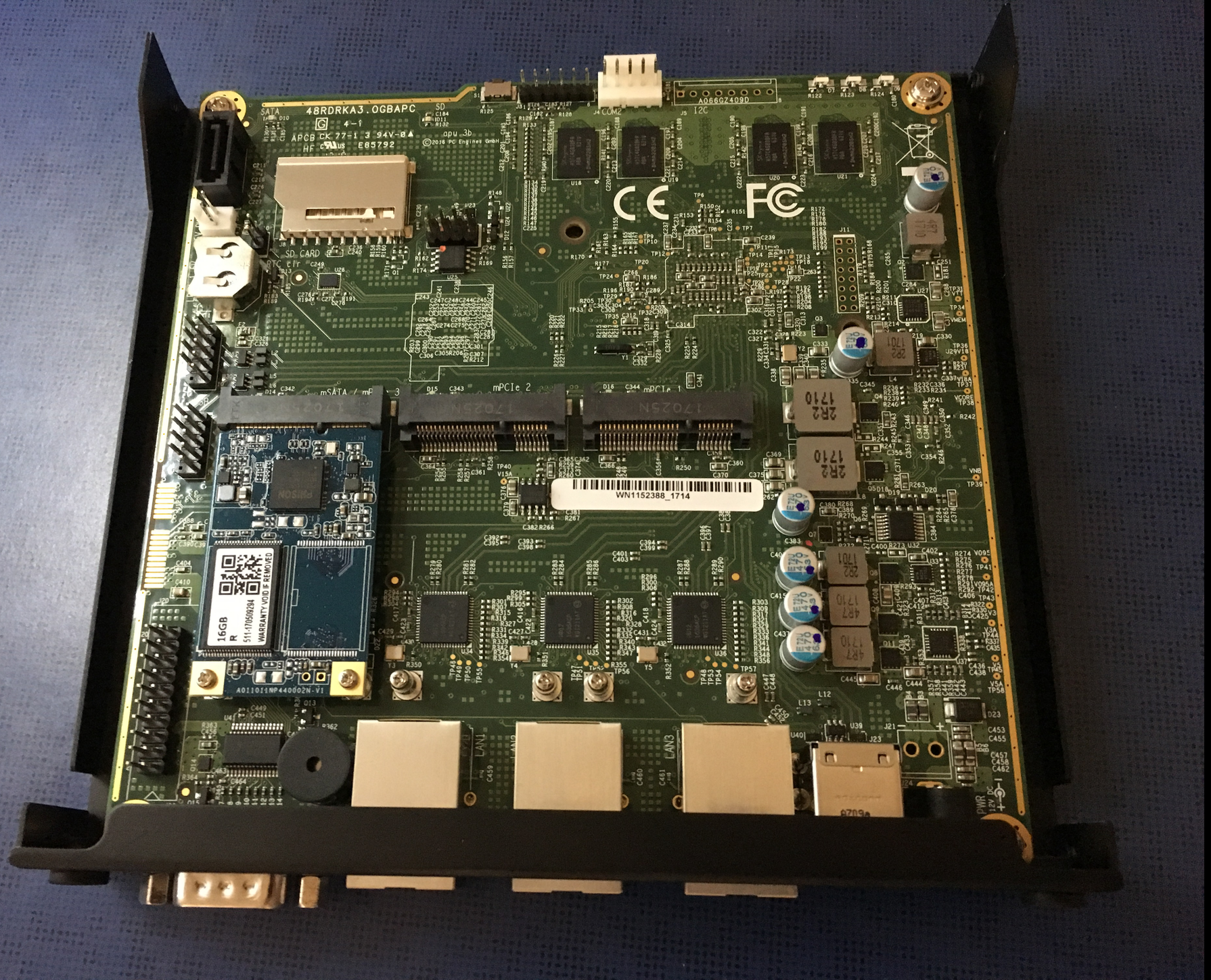
Hi folks,

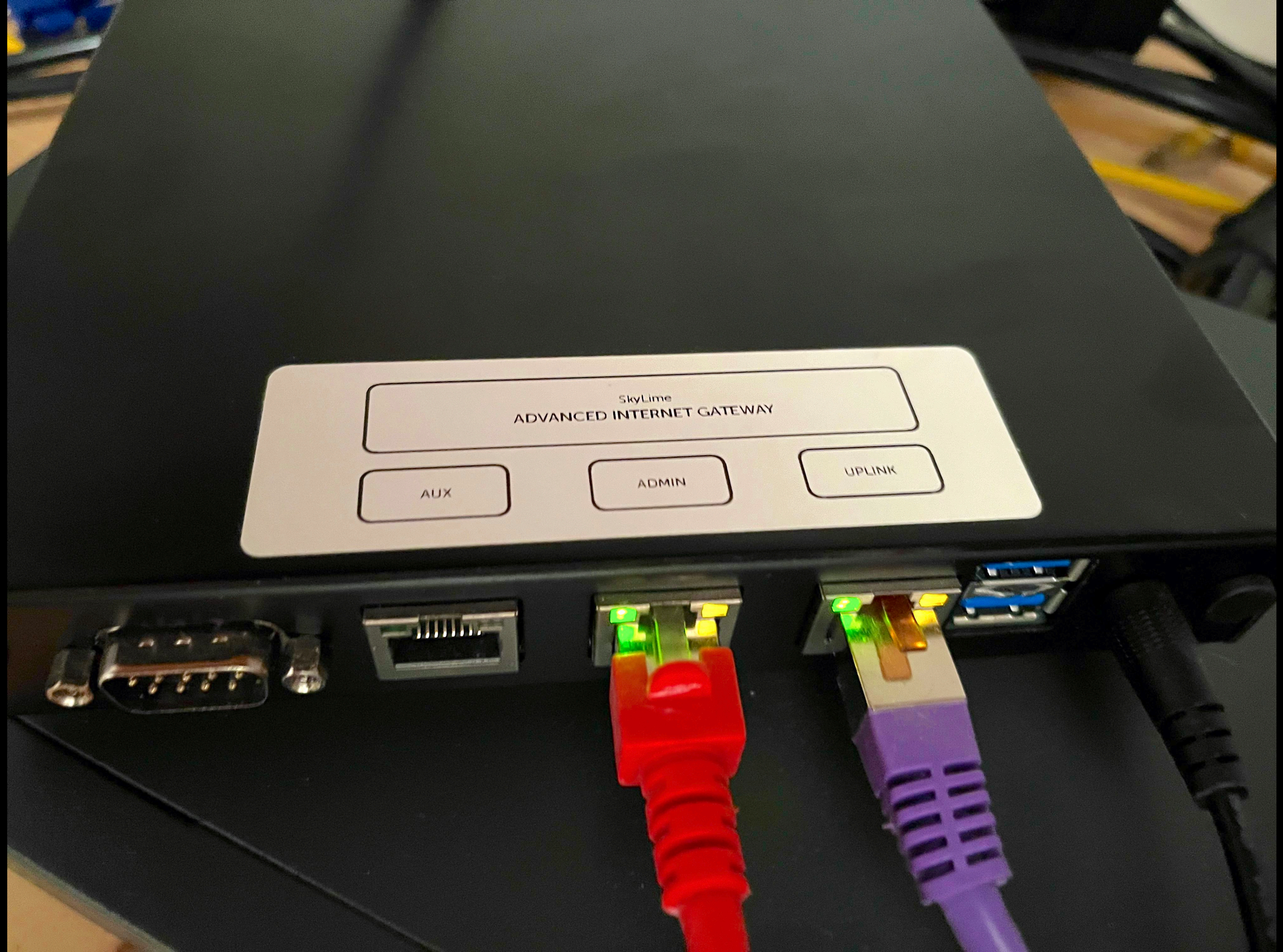
(tl;dr) If you have a running illumos system, and a few moments, I would very much appreciate your help in gathering data from as many different systems as possible. This mail contains a detailed problem description, but you can also just skip to the "GATHERING INFORMATION" section below.

History
2017

- Hardware & Software ready
- APU3B4 Systemboard
  - 1 GHz
  - 4 GB DDR3 RAM
  - 3x 1000 mbit/s
  - 12V DC, about 6 to 12W depending on CPU load
- 8GB Transcend SD Card
- 16GB mSATA SSD
- ~200 euro (incl. case, shipping)




48RDRKA3.OGBAPC
SATA
SD
SD CARD
RTC clr
mSATA / mP
mPCIe 2
mPCIe 1
I2C
COM2
LAN1
LAN3
PWR 12V DC
WN1152388_1714
16GB


SkyLime
ADVANCED INTERNET GATEWAY
AUX
ADMIN
UPLINK

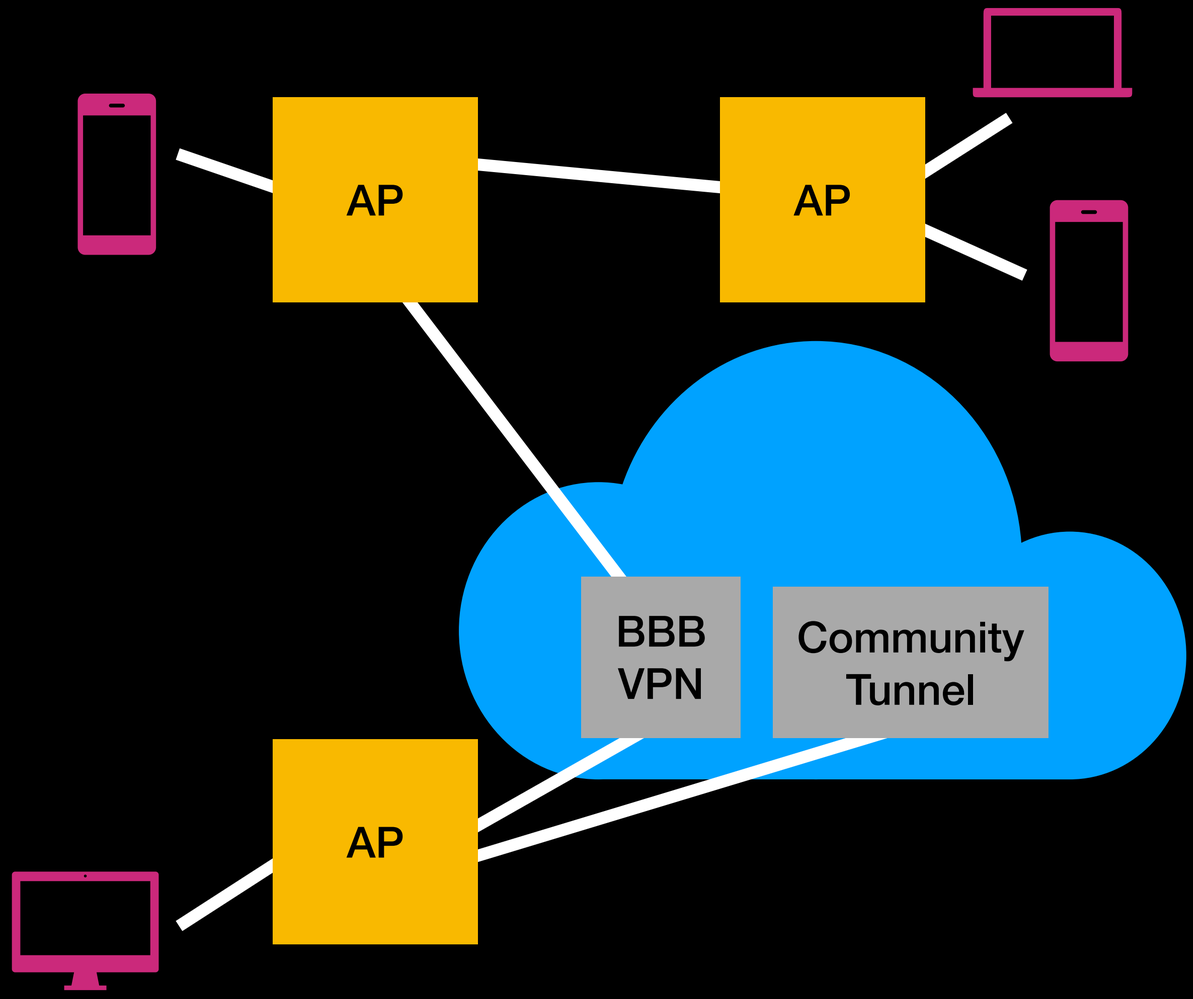
Freifunk

- wireless community network
- mesh backbone
- regional communities
- off the shelf wifi router + custom openwrt firmware


f15-cafe
f15-sw3
Berhala
Sarumpaet
weissenseerspitze
stadtoase-balkon
FreifunkFriesicke
Lehder11
behaim4-tunnel
druschba.ff.lunatiki.de
Lychener-Str-76-HH
2.-Lychenerstr-76-HH
kiezhaus-eg
hdj-wedding
kommune65
GruentalerStr_11_13357
Badstraße
Volkspark Rehberge
Kamerguinea
FrischAuf-Core2
FrischAuf-Suedwest-2GHz
funkpilz
Bruesseler_10
scherer8-abb
hoch13_icebar
fkkwedding-0
funky_time
torf29-core
buch8
nokl-core
k11-ff-garten
Lino5
k12-core
s--35
BuL_WinsStr34
Chor46
segen-core
PiratenBerlinHof
PiratenBerlinVorne
kls0e-l26
perleberger59
perleberger36-turm-2ghz-so
Somraw
ffunk39
habersaath-core
BuL_Hufelandstr9
lilo2_wdr3600
storkower139c
Borsigstr-11-VH
linie206-core
notunterkunft-tieck17
stadalbert-core
saarbruecker-gw
Turmstraße
kamatz
Buchhandlung
gerudo
OpenWrt
strom-gw
dreispitz.freifunk.net
hds-ffraum
hds-containers
NeuePromenade6
ang-pos-3-core
commander-root
fkkbersarin
Bartningallee16
Otto-Suhr-Allee
Kohlrsch5
Großer Tiergarten
Fra29
agym-core
dr-ogen
Sesamstrasse
10-36-79-193
Annenstrasse12
blackbox
rt-tunnel
rt-park
xdorf-core
kko-rudolf-lehmbruck
fkkostkreuz
BerlinerTeesalon
JugendKulturZentrum_PUMPE
GalerieGustavvonHirschheydt
l105-gw
halle-core
tommyhausff3
DanneckerStr24
freifunk-fhain
HausMallorca
Lietzenburger Straße
mgh-core
suedblock-core
Schlesische14QG
Kerngehaeuse4
BuL_Cuvrystr22
Humpty-VH2
dtmb-ladestr-bbb-nord
HKPA1
PallasT
nhu-emma
ohlauer-gw
bundes-0
rhxb-core
Urbanstraße
manstein10-core
grossgor28
Tatch
BuL_ZossenerStr34
NeueRepublikFunk

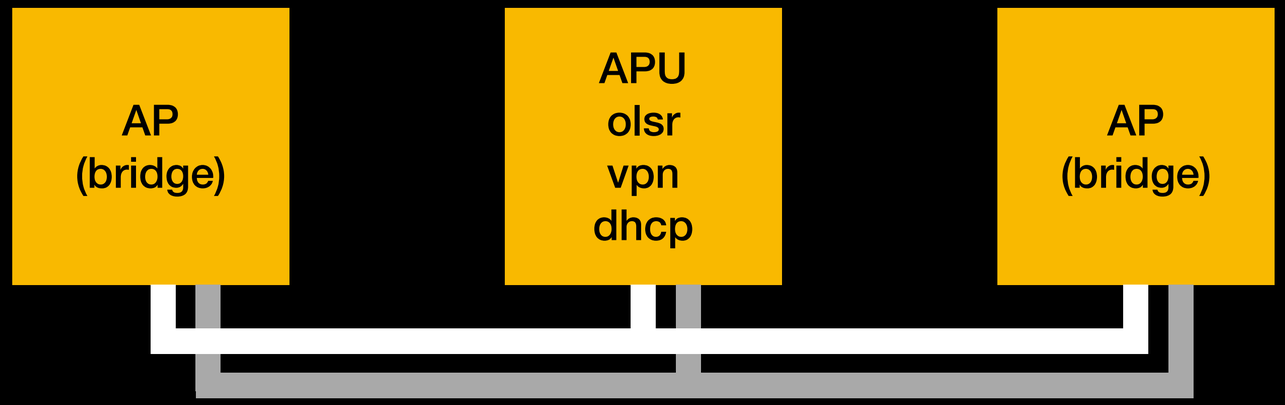
berlin.freifunk.net

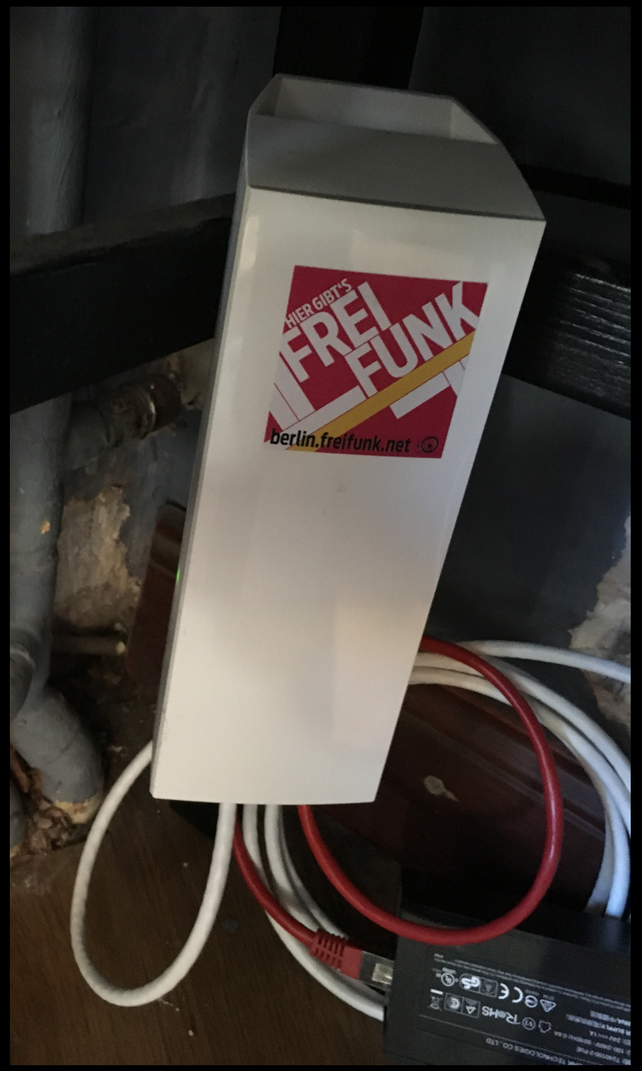
- OLSRv1 (RFC3626)
- 2 VPNs for backbone and legal protection
  - OpenVPN at the time
  - now wireguard „tunneldigger“
  - backbone + community tunnel
- separate SSIDs
  - clients (DHCP)
    - berlin.freifunk.net
  - mesh / point-to-point
    - intern-chXX.freifunk.net
    - …



freifunk offloaded

- Idea: run VPNs + olsrd on APU
- TP-Link CPE210 / CPE510 APs
  - bridge mode
- roaming!





olsrd

- mostly portable with some OS specific parts (adding / removing routes)
- started porting to illumos in 2016
- timing bug fixed & upstreamed
- needs `getifaddrs()` to lookup MAC addresses (AF_LINK)
- illumos version only returns AF_INET & AF_INET6

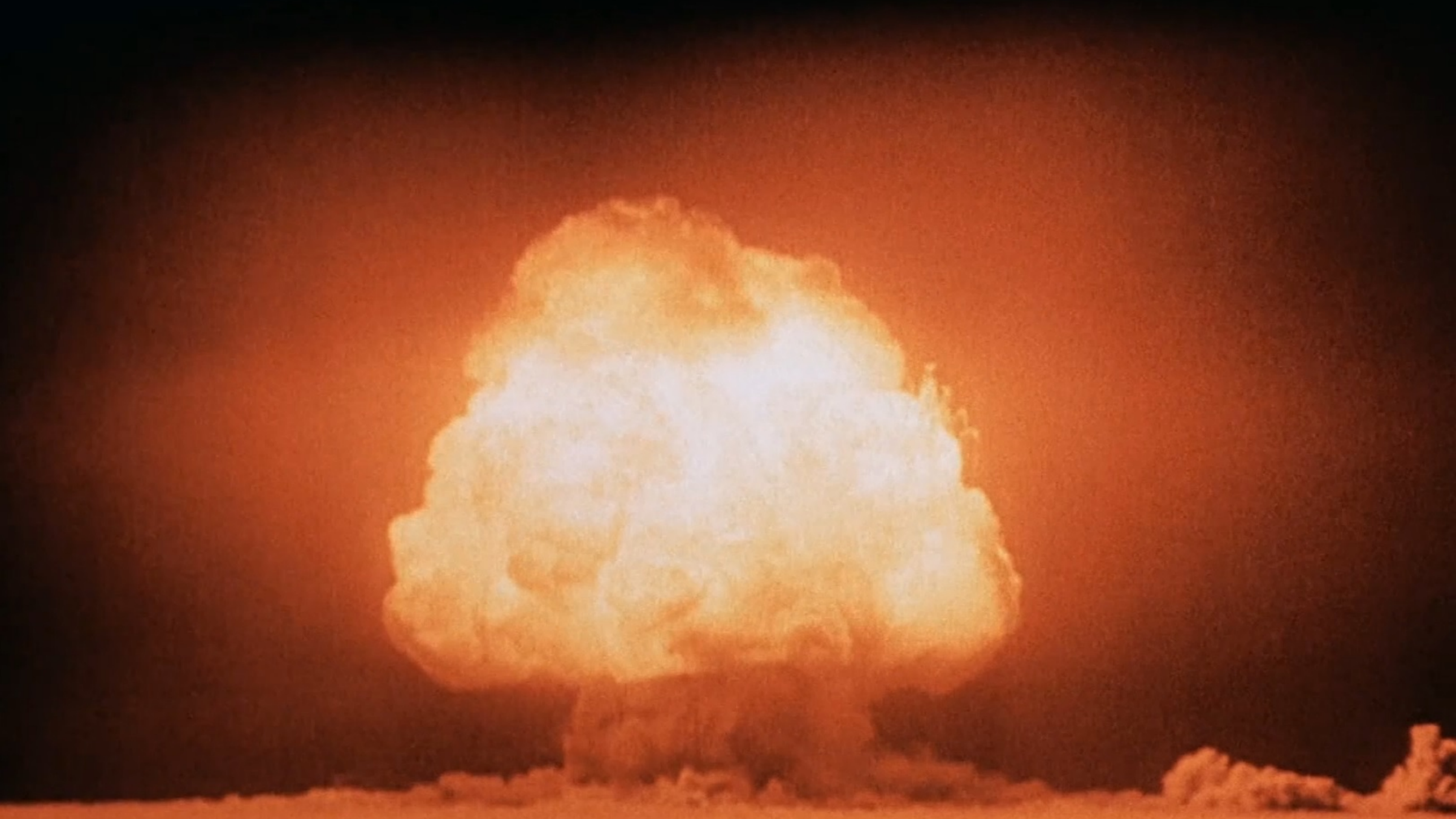
getifaddrs(3socket)

int getifaddrs(struct ifaddrs **ifap)

```
struct ifaddrs {
        struct ifaddrs      *ifa_next;
        char                *ifa_name;
        uint64_t            ifa_flags;
        struct sockaddr     *ifa_addr;
        struct sockaddr     *ifa_netmask;
        union {
            struct sockaddr *ifu_broadaddr;
            struct sockaddr *ifu_dstaddr;
        } ifa_ifu;
        void                *ifa_data;
};


struct sockaddr {
        sa_family_t     sa_family;      /* address family */
        char            sa_data[14];    /* up to 14 bytes of direct address */
};
```

getifaddrs

- 3729 - getifaddrs must learn to stop worrying and love the other address families
  - opened 2013
- jan 2017 - posted first version for review
  - only worked for links with ip interfaces on top
- may 2017 - version that does door calls to dlmgmtd to enumerate links
  - lot’s of help from rmustacc
- june 2017 - RTI & merged into illumos-gate



OpenIndiana nwam-manager

```
if (getifaddrs(&ifap) == 0) {
        for (idx = ifap; idx; idx = idx->ifa_next) {
                ...
                if (idx->ifa_addr->sa_family == AF_INET) {
                        /* some ipv4 things */
                } else {
                        /* some ipv6 things */
                }
        }
        freeifaddrs(ifap);
}
```

getifaddrs
2017 June

- nwam-manager fixed
  - OI rolling release means users get the fixed version immediately
- then reports come in that node.js crashes
  - SmartOS uses node.js (a lot)
  - Bug is actually in libuv
  - fixed & upstreamed quickly
  - but: all existing node.js binaries have this bug compiled in
  - will be years before people stop using those
- backout 3729
- reviewed all getifaddrs() callers in pkgsrc via opengrok for similar bugs (turned up a more libuv consumers)

getifaddrs
2022

- `#pragma redefine_extname getifaddrs __getifaddrs`
- New compiled software will get AF_LINK entries
- Old binaries will work as before
- Finally integrated for good

almost there

SmartOS networking basics

nictags

- names for interfaces (idependent of MAC address or driver)
- physical nics or local-only etherstubs
- managed using `nictagadm`
- settings stored in /usbkey/config

```
[root@gw (de-bln-f15) ~]# nictagadm list
NAME            MACADDRESS          LINK            TYPE
uplink          00:0d:b9:46:57:12   igb2            normal
admin           00:0d:b9:46:57:11   igb1            normal
aux             00:0d:b9:46:57:10   igb0            normal
routing0        -                   -               etherstub
```

vmadm network options

- supports multiple nics, one primary
- „ips“ array lists addresses in CIDR notation
  - „dhcp“ for DHCPv4
  - „addrconf“ for SLAAC / DHCPv6
- protection settings:
  - allow_ip_spoofing
  - allow_mac_spoofing
  - allow_dhcp_spoofing
  - allow_restricted_traffic
- more options in vmadm(8)

```
"nics": [
  {
    "nic_tag": „example0",
    "ips": ["192.0.2.1/24", "2001:db8::1/64"],
    "allow_ip_spoofing": true,
    "vlan_id": 42
  }, {
    "nic_tag": "uplink",
    "ips": ["dhcp", "addrconf"],
    "primary": true
  }
]
```

routeadm

- forwarding != routing
- `routeadm -ue ipv4-forwarding`
- `routeadm -ue ipv6-forwarding`
- ipv4-routing/ipv6-routing enable routing daemon
  - rip/ripng by default
- OSPF/IS-IS/BGP/… available via Quagga/FRR

DIY Config management

- git repo
- easy to bootstrap
  - expect little working infrastructure
  - script to generate vmadm json definitions with templates
  - minimal software dependencies (shell + jq)
  - copy & paste commands into serial console
- easy to recreate

vmadm templates

```
(sed -e "s;__BASE64__;$BASE64;g" \
        -e "s;__ALIAS__;$alias;g" \
        -e "s;__DOMAIN__;$DOMAIN;g" \
        -e "s;__DNS1__;$DNS1;g" \
        -e "s;__DNS2__;$DNS2;g" \
        -e "s;__RAM__;$RAM;g" \
        -e "s;__ROOT_AUTHORIZED_KEYS__;$SSH_KEY;g" \
        < base.json; cat vms/$alias/template.json) | jq -s add
```

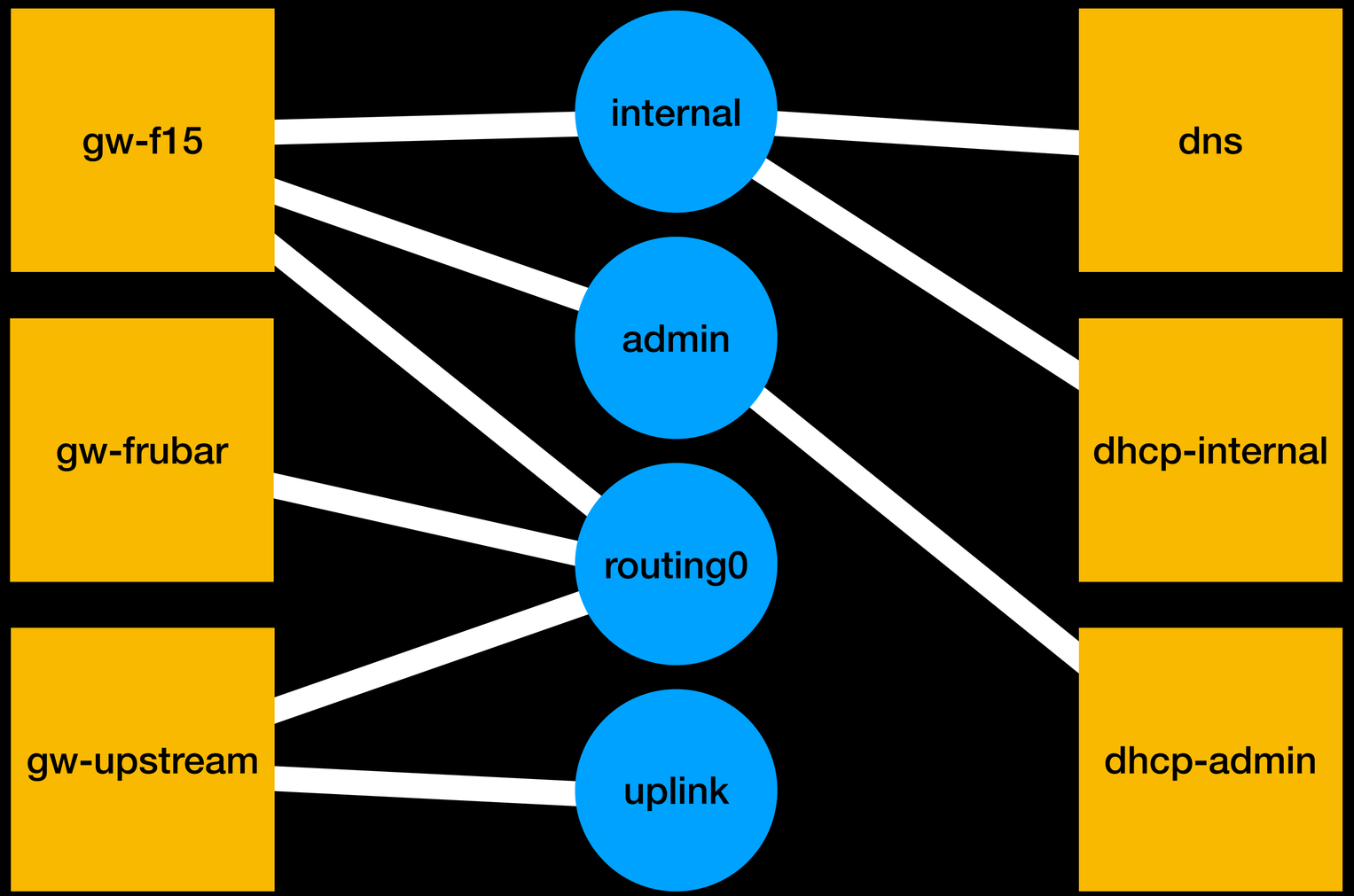
vmadm templates

```
# base.json
{
  "brand": "joyent",
  "image_uuid": "__BASE64__",
  "alias":    "__ALIAS__",
  "hostname": "__ALIAS__.__DOMAIN__",
  "resolvers": [
    "__DNS1__",
    "__DNS2__"
  ],
  "max_physical_memory": __RAM__,
  "max_swap":            __RAM__,
}

# gw-upstream/template.json
{
  "nics": [
    { "nic_tag": "routing0", "ips": ["192.168.35.1/24", "2a01:138:a015:35::1/64"], "allow_ip_spoofing" : true },
    { "nic_tag": "uplink",   "ips": ["dhcp"], "primary": true }
  ],
  "routes": {
    "192.168.15.0/24":      "192.168.35.2",
    "192.168.10.0/24":      "192.168.35.2",
    "192.168.0.0/16":       "192.168.35.3",
 }
}
```

Networks

name	nictag	vlan	ipv4	ipv6
admin	admin		192.168.15.0/24	2a01:138:a015:15::/64
internal	admin	2	192.168.10.0/24	2a01:138:a015:10::/64
routing0	routing0		192.168.35.0/24	2a01:138:a015:35::/64

Zones


gw-f15
gw-frubar
gw-upstream
internal
admin
routing0
uplink
dns
dhcp-internal
dhcp-admin

ipf nat

```
echo 'map net1 192.168.0.0/16 -> 0/32 portmap tcp/udp 1025:65534' >  /etc/ipf/ipnat.conf
echo 'map net1 192.168.0.0/16 -> 0/32'                            >> /etc/ipf/ipnat.conf

svcadm enable ipfilter
```

ndpd

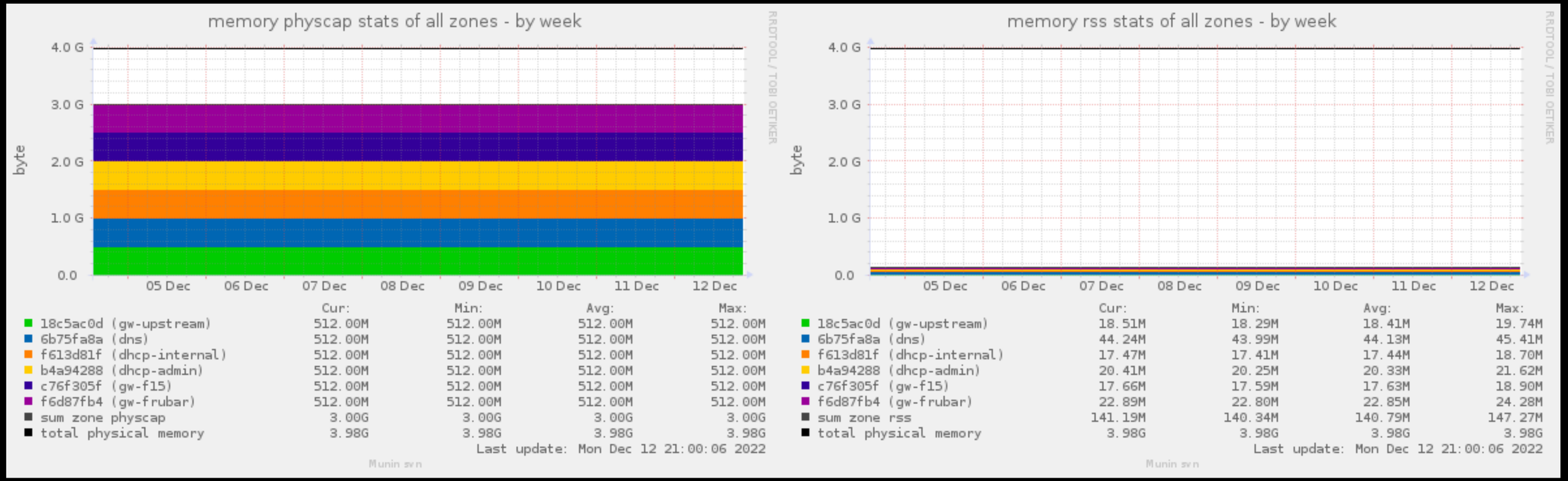
```
cat > /etc/inet/ndpd.conf << EOF
ifdefault AdvSendAdvertisements 0
ifdefault StatelessAddrConf 0
ifdefault StatefulAddrConf 0
prefixdefault AdvOnLinkFlag on AdvAutonomousFlag on

if net1 AdvSendAdvertisements 1
prefix 2a01:138:a015:15::/64 net1
if net2 AdvSendAdvertisements 1
prefix 2a01:138:a015:10::/64 net2
EOF
```

unbound

```
pkgin -y install unbound

cat > /opt/local/etc/unbound/unbound.conf << EOF
server:
  verbosity: 0
  interface: 0.0.0.0
  interface: ::0
  interface-automatic: yes
  access-control: ::0/0 allow
  access-control: 0.0.0.0/0 allow
remote-control:
  control-enable: yes
EOF

unbound-control-setup

svcadm enable svc:/pkgsrc/unbound:default
```

policy via firewall
otherwise adjust!

RAM usage


memory physcap stats of all zones - by week
byte
4.0 G
3.0 G
2.0 G
1.0 G
0.0
05 Dec
06 Dec
07 Dec
08 Dec
09 Dec
10 Dec
11 Dec
12 Dec
RRDTOOL / TOBI OETIKER
Cur: Min: Avg: Max:
18c5ac0d (gw-upstream) 512.00M 512.00M 512.00M 512.00M
6b75fa8a (dns) 512.00M 512.00M 512.00M 512.00M
f613d81f (dhcp-internal) 512.00M 512.00M 512.00M 512.00M
b4a94288 (dhcp-admin) 512.00M 512.00M 512.00M 512.00M
c76f305f (gw-f15) 512.00M 512.00M 512.00M 512.00M
f6d87fb4 (gw-frubar) 512.00M 512.00M 512.00M 512.00M
sum zone physcap 3.00G 3.00G 3.00G 3.00G
total physical memory 3.98G 3.98G 3.98G 3.98G
Last update: Mon Dec 12 21:00:06 2022
Munin svn
memory rss stats of all zones - by week
byte
4.0 G
3.0 G
2.0 G
1.0 G
0.0
05 Dec
06 Dec
07 Dec
08 Dec
09 Dec
10 Dec
11 Dec
12 Dec
RRDTOOL / TOBI OETIKER
Cur: Min: Avg: Max:
18c5ac0d (gw-upstream) 18.51M 18.29M 18.41M 19.74M
6b75fa8a (dns) 44.24M 43.99M 44.13M 45.41M
f613d81f (dhcp-internal) 17.47M 17.41M 17.44M 18.70M
b4a94288 (dhcp-admin) 20.41M 20.25M 20.33M 21.62M
c76f305f (gw-f15) 17.66M 17.59M 17.63M 18.90M
f6d87fb4 (gw-frubar) 22.89M 22.80M 22.85M 24.28M
sum zone rss 141.19M 140.34M 140.79M 147.27M
total physical memory 3.98G 3.98G 3.98G 3.98G
Last update: Mon Dec 12 21:00:06 2022
Munin svn

demo repo

https://github.com/wiedi/homerouter-demo

What’s next

- illumos#11916 - Support for RFC8106 in in.ndpd (DNS RA options)
- illumos#4033 - DHCPv6 prefix delegation
- USB NCM support (for newer LTE modems)
- investigate+fix unbound anycast issue
- native wireguard?
- more modern firewall?
- updated wifi stack?

THANK YOU

and the many people who helped make this possible