SUNET

UPDATES AND A SHORT INTRODUCTION TO LOGICAL NETWORK DESIGN.

Alright, time for some updates.

We have installation-teams driving around all of Sweden and most of the core-network is starting to be built out, we had a few good weeks with good pace so at this very moment we have installed 82% of the DWDM-nodes and 64% of the core-routers. In the same time we roll out DWDM and Routers to core-locations we also install and commision the 4G Out-of-Band console servers so we can reach and start to configure all equipment.

Both ROADMs and routers is shipped without any configuration at all and we do not require the field-technicians that install the equipment to do anything else than to connect the console-server to the 4G-antennas (they are preconfigured as mentioned in a earlier post) and we should be able to access all equipment at the sites. This means we can commision links, wavelengths and optical spans before we have light on the fiber.

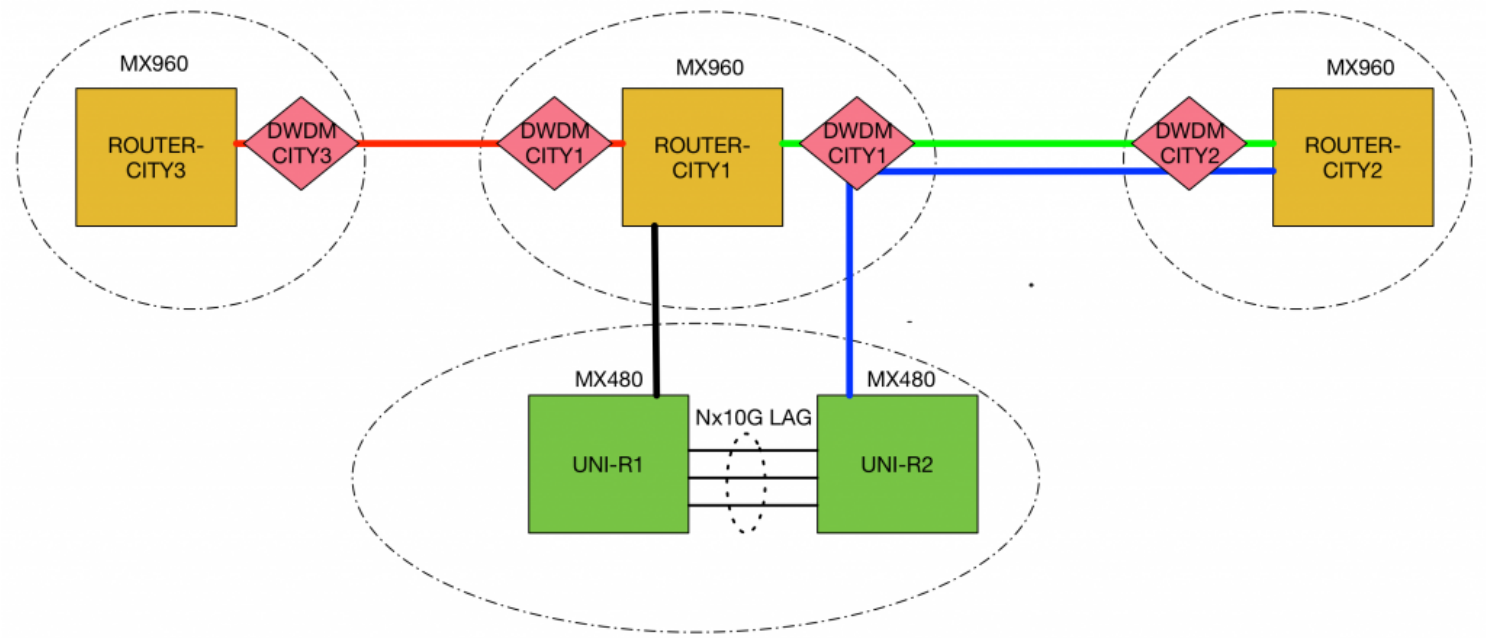
There is still work going on throughout the network to fix and repair all reported problems with the dark fiber we have encountered and to not push the time-schedule further we install all equipment and connect the fiber nevertheless if the fiber-span is okay or not, and hopefully when the fiber is up to standard we can go ahead and light spans immediately.

We have been running a few successful tests on the completed optical-spans and so far we have not encountered any problems to light up long (or short) links as long as the amplifiers is performing as expected and that the fiber is decent, the longest working link e we have today with the Juniper 100G Coherent interfaces is Uppsala to Luleå, that is about 900km of fibre. Hopefully we can get enough links up and running so we can test Malmö-Kirtuna in a not so distant future (2000km fibre).

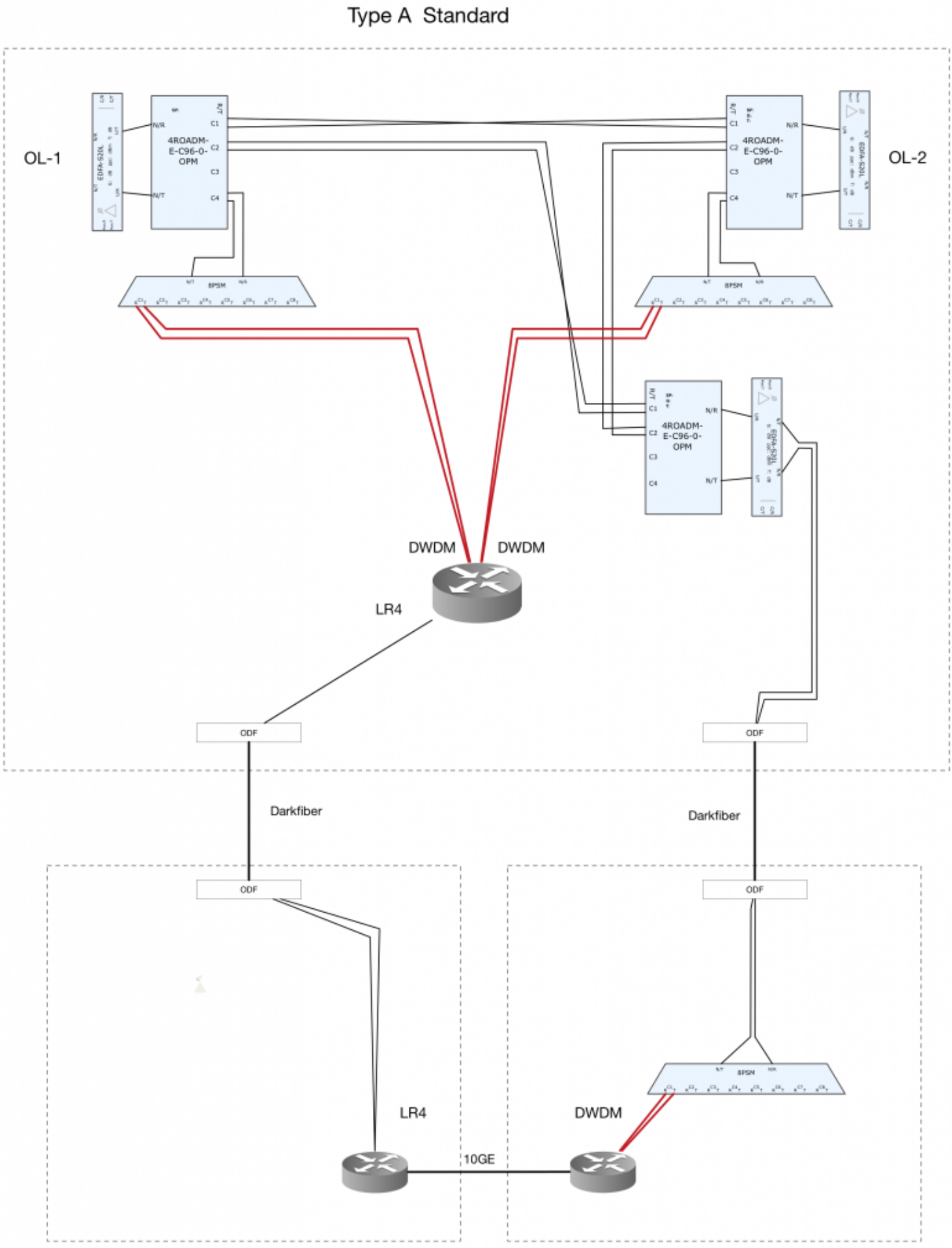
The idea to get the maximum amount of performance out from a RAMAN-amplifier is a project that will continue long in 2017 and maybe even in 2018 as well. The main reason we use these high grade hybrid RAMAN-amplifiers is not to be able to light up the current network as it is, but to prepare for the future and make sure we don't need to truckroll and change all amplifiers when and if we upgrade to the next technology, which may be 400G, or it may even be 1000G.

During may and all during summer our sub-contractor for fiber will start to measure and prepare all access-fibers. The fiber going from the core/POP site out to the customer in the other end. The requirements for these fibers is not as harsh as the long-distance links where we more or less expect them to be perfect, but for access-fibre perfect will not exist. There is local broadband and power companys in almost all cities we are active in and the quality of installations and what you can do is very different. Most of them is also in a state of monopoly and does not really need to deliver something out of the ordinary. What we have is that we require one of the access-fibres to be within specifications that a regular 100g-LR4 can work on it, the other access-fibre can be worse since we will run the Coherent optics on top of that and have much more room for high attenuation and reflection.

Alright, enough of that. Let's talk some logical design.

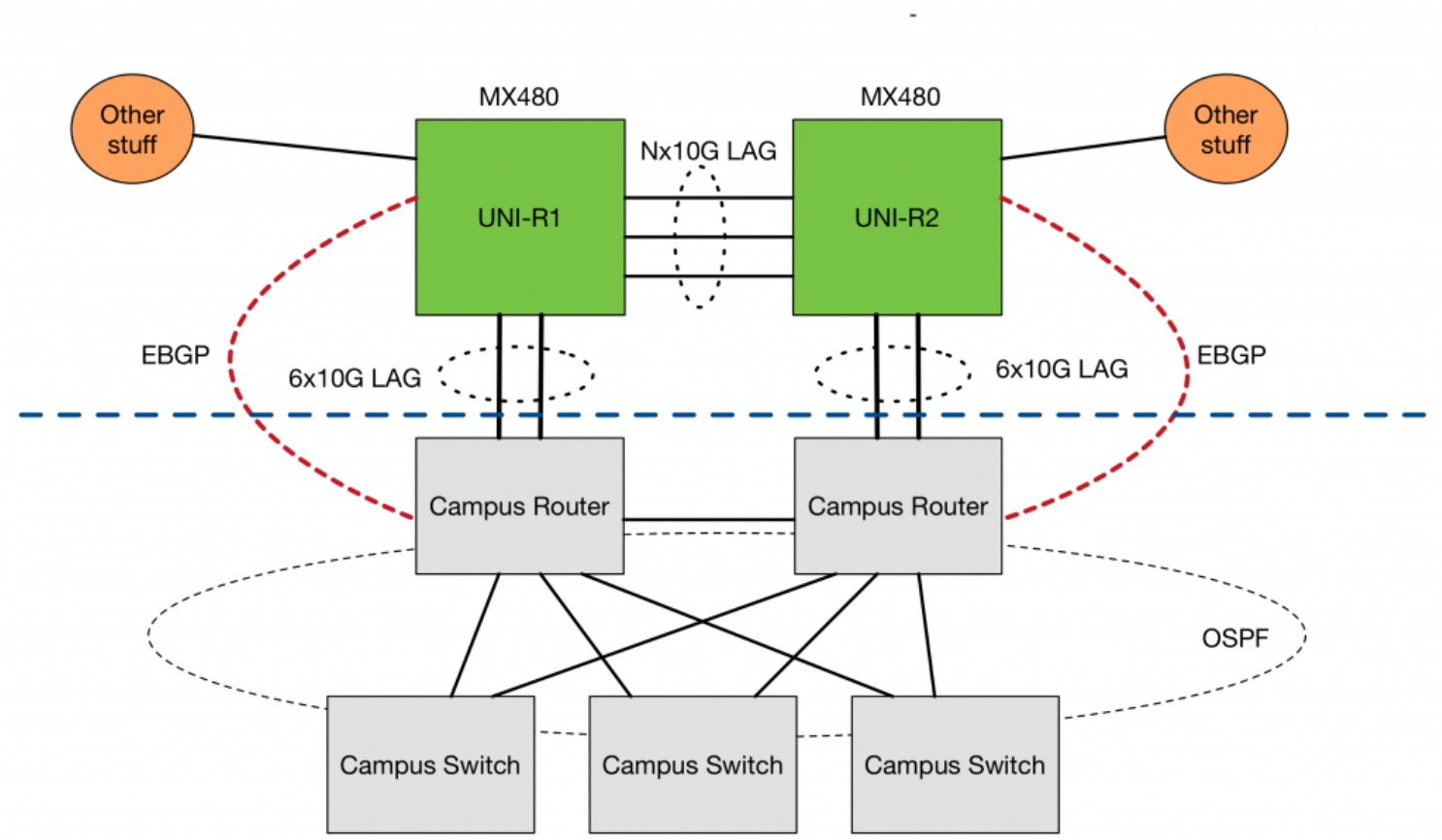


This is a very simple overview on the physical design of university-access. There is two Juniper MX480 located on a diverse location at campus, these are tied together with fiber in between consisting of Nx10G. The uplinks is one 100G LR4 locally in the city, and a 100G Coherent to the neighbouring city. And then the core-router in the city is also tied together to two neighbouring cities using 100G Coherent interfaces.


Type A Standard
OL-1
OL-2
4ROADM-E-C96-0-OPM
8PSM
DWDM
DWDM
LR4
ODF
ODF
Darkfiber
Darkfiber
ODF
ODF
LR4
DWDM
10GE

This is a more block-diagram how it will look on how it will be connected if its a Standard type of handoff. Click it for the full resolution to see how the idea is to connect things. The 8PSM ”filter” is a passive splitter/combiner and is gridless.

Time to look at logical things...

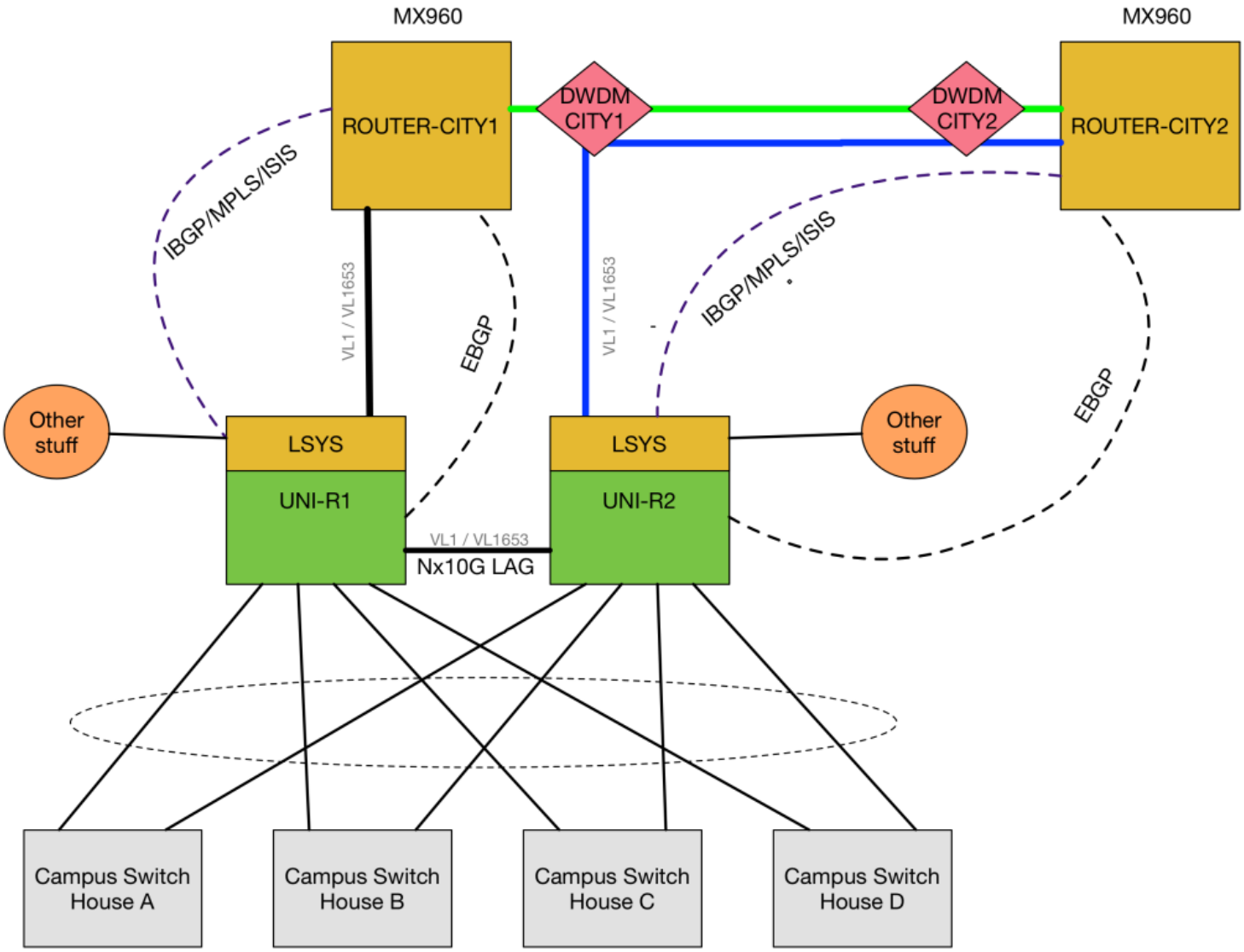


So this is the simplest way of doing a handoff to the university. A classical ”ISP-handoff” with EBGP. Both Juniper MX480 is SUNET-controlled and runs as part of the core-network. ISIS, MPLS and IBGP is terminated in MX480 and EBGP is used down to the customers AS-number and we can preferably send a full table down to the university to take advantage of all diverse paths. Day 1 we have allocated 6x10G interfaces for downstream purposes which is free for the university to use as they like, either to different equipment or to build a Nx10G LAG. SUNET also has the possibility to deliver any type of services through auxiliary interface such as MPLS-tunnels, interconnect-interfaces, other type of customers that connects through campus (dormitories, museums, science projects, etc), on the diagram shown as ”other stuff”.

This requires the university to a have a device that can speak BGP accordingly, and most likely that it can also do a compatible LAG/LACP to the Juniper devices to aggregate the links.

The positive things with this solution is that it is very simple, for both the customer and the operator. Minimal configurations and clear boundaries. The negative thing is that this requires the university to run and take care off BGP themselves in their own equipment. Which sets quite high demands on customers equipment.

This is when things get interesting, what if the the SUNET could co-exist with the Campus network in the same equipment? seeing as the investment is already done for the Juniper-router.



What we have done for many years in SUNET and continue with in the new network is that we run logical systems on all of our routers. This cuts the physical router into logical systems (virtualization), and you assign resources to the logical system as you like. In the old network we provided a logical system (guest) to the university and had SUNET in the main-instance (the host), this was mostly to be able to provide a BGP-to-OSPF-translator seeing as no university was interested back then to house their own BGP-capable devices.

What we will do in the new network is to flip the steak and put SUNET in a logical system, and put the campus in the main-instance. This is because there is a few features thats not supported on a logical system, such as collecting netflow and using mlag. SUNET can and will collect netflow in the core-routers instead and the university could then extract netflow from their own MX-routers instead. The main-instance and the SUNET logical system will be using different VLANS on the 100G-ports up to core. The university will peer with EBGP with the core-routers and not the LSYS on the same box, this is mainly because the logical-tunnel interfaces used to connect logical systems is limited to half the ASIC-speed. The LSYS which will house the SUNET-instance will terminate MPLS, IBGP and ISIS (which is the IGP of choice for SUNET.) This logical system will be used to manage the box itself but also to be able to terminate non-university customers at the campus-sites and the ability to provide virtual services to them one needing them.

The positive thing with this type of setup is that we leverage the very potent routing and protocol-functionality from the MX from both the ISP-side and down to the campus-side as well. There is plenty of resources for everyone to do whats needed and with the broad featureset of a service-router type of equipment the university can enjoy technologies which is probably not available when using a more classic campus-core type of device. For the ones that decides to utilize the SUNET MX-router as their L3-Core will instantly get the benefit of less devices in the network and no need to re-invest in core-equipment for the foreseeable future.

Logical system will of course be provided free-of-charge since it does not cost SUNET anything to produce (no licensing-fees or such).

BONUS

Juniper has recently changed vital infrastructure on how the RE is built out, some interesting tidbits while we have (tried to) upgrade releases and get things to work.

```
Running /initial_setup.sh
Type ^C to abort
Running initial setup scripts
-> running /etc/initial_setup/00read_only_root.sh
Check ...
/dev/sda4 on / type ext4 (ro,relatime,seclabel,data=ordered)
sysfs on /sys type sysfs (rw,relatime,seclabel)
selinuxfs on /sys/fs/selinux type selinuxfs (rw,relatime)
proc on /proc type proc (rw,relatime)
Check ...
Starting udev
  4 logical volume(s) in volume group "jvg_S" now active
  4 logical volume(s) in volume group "jvg_P" now active
mount: proc is already mounted or /proc busy
       proc is already mounted on /proc
Starting Bootlog daemon: bootlogd.
ALSA: Restoring mixer settings...
/usr/sbin/alsactl: load_state:1729: No soundcards found...
net.ipv4.conf.default.rp_filter = 1
net.ipv4.conf.all.rp_filter = 1
vm.nr_hugepages = 64
kernel.core_pattern = |/bin/bash /usr/bin/coredump %h %e %t

INIT: Entering runlevel: 3
Starting system message bus: dbus.
Mounting cgroups...Done
Starting OpenBSD Secure Shell server: sshd
done.
Configuring network interfaces... done.
Starting rpcbind daemon...done.
creating NFS state directory: done
starting statd: done
Starting Advanced Configuration and Pow
acpid: starting up

acpid: 1 rule loaded
```

Anyone got a juniper-soundcard?

```
Main Menu

1.  Boot [J]unos volume
2.  Boot Junos volume in [S]afe mode

3.  [R]eboot

4.  [B]oot menu
5.  [M]ore options




Choice:
Booting from Junos volume ...
-
/packages/sets/pending/boot/os-kernel/kernel text=0x3a59c8 data=0x7a500+0xdc810 syms=[0x8+0x8da08+0x8+0xa539f]
/packages/sets/pending/boot/netstack/netstack.ko size 0xec23a0 at 0xa30000
loading required module 'crypto'
/packages/sets/pending/boot/os-crypto/crypto.ko size 0x2e9c0 at 0x18f3000
/packages/sets/pending/boot/os-kernel/miibus.ko size 0x411e0 at 0x1922000
/packages/sets/pending/boot/os-kernel/if_em.ko size 0x7d6c0 at 0x1964000
/packages/sets/pending/boot/os-kernel/if_fxp.ko size 0x13d60 at 0x19e2000
/
```

booting

```
Starting domain name service: named.
Creating the macvlan0 interface...
Cannot change scatter-gather
Cannot change tcp-segmentation-offload
Starting glusterd:
exportfs: can't open /etc/exports for reading
starting 8 nfsd kernel threads: done
starting mountd: done
Starting ntpd: done
starting rsyslogd ... done
Starting wdmd:  Success
Starting internet superserver: xinetd.
 * Starting Avahi mDNS/DNS-SD Daemon: avahi-daemon
Creating VFs...
Starting sanlock:  Success
Starting Postfix...postsuper: fatal: scan_dir_push: open directory defer: Permission denied
postfix/postfix-script: fatal: Postfix integrity check failed!
 Failed
modprobe: FATAL: Module openvswitch not found.
Inserting openvswitch module ... failed!
Module has probably not been built for this kernel.
Install the openvswitch-datapath-source package, then read
/usr/share/doc/openvswitch-datapath-source/README.Debian
Starting ovsdb-server.
Configuring Open vSwitch system IDs.
modprobe: FATAL: Module openvswitch not found.
Inserting openvswitch module ... failed!
Enabling remote OVSDB managers.
Enabling gre with iptables.
 * Starting virtualization library daemon: libvirtd
no /usr/bin/dnsmasq found; none killed
 * Starting ovs-controller  ovs-controller
/etc/openvswitch-controller/cacert.pem: CA certificate missing
Starting crond: OK
starting mcelog... done
```

Postfix....what

```
Booting SSD1: EFI Hard Drive (StorFly VSF202CI050G-JUN)...

Decrementing retry count for SSD1
Welcome to GRUB!

                             GNU GRUB  version 2.00

 /----------------------------------------------------------------------------\
 |Serial console efi_P                                                        |
 |Serial console jspare_P                                                     |
 |                                                                            |
 |                                                                            |
 |                                                                            |
 |                                                                            |
 |                                                                            |
 |                                                                            |
 |                                                                            |
 |                                                                            |
 |                                                                            |
 |                                                                            |
 |                                                                            |
 \----------------------------------------------------------------------------/

    Use the ^ and v keys to select which entry is highlighted.
    Press enter to boot the selected OS, `e' to edit the commands
    before booting or `c' for a command-line. ESC to return
  The highlighted entry will be executed automatically in 0s.
```

grub

Skriven av

FREDRIK “HUGGE” KORSBÄCK

Network architect and chaosmonkey for AS1653 and AS2603. Fluent in BGP hugge@nordu.net