Bifrost och 10Gbit routing

Software Freedom Day
2010-09-18/Stockholm

Robert Olsson
Uppsala Universitet och KTH

What is bifrost?

Small Linux suited for USB
- For infrastructure
- For research & education

Name of human network and collaboration

Why a new Linux distro? We've been around
For 15 years soon.

What about Red Hat/Debian Ubuntu etc?

For who?

Networking people
Unix/Linux
Needing a small flexible distro
Research & Education

Objectives infrastucture?

Hardware selection – Crucial
Software selection - Crucial
Testing - Crucial
Bug Fixes/Support Crucial
Development - Crucial

Basic functions?

Routing
Firewalling
Login services
Traffic logging
Gateways etc
Virtualization and/or Namespaces (New)
Ipv4, ipv6

Network testing pktgen/netperf etc

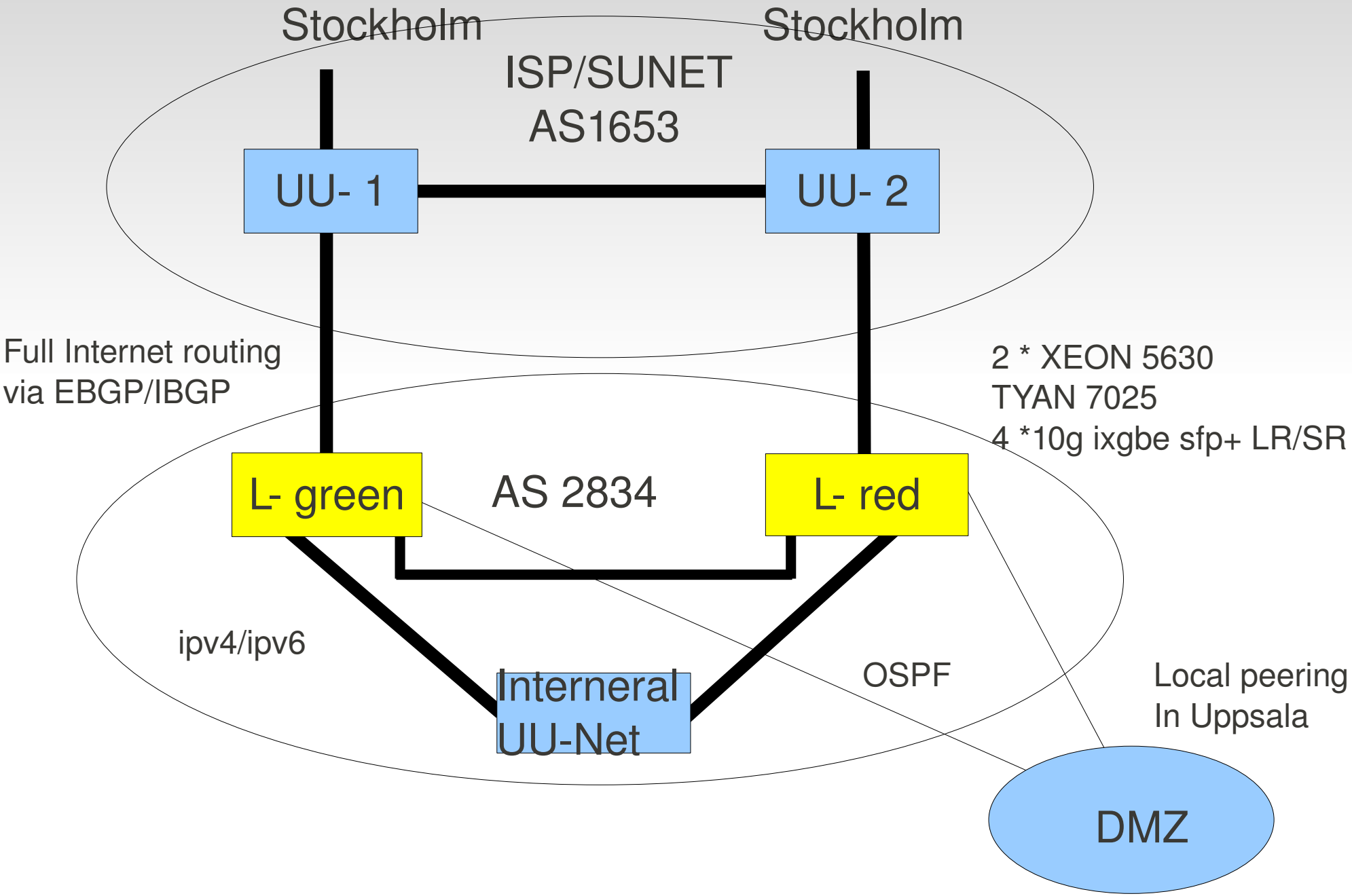
Routing deamons

Packet forwarding is done in Linux kernel
But routing protocols is run by userpace Deamons

Currently tested versions of quagga
Bgp, OSPF both Ipv4, ipv6
Cisco API

Of course other software can be used
Bird?

More than 10 year in production at Uppsala University



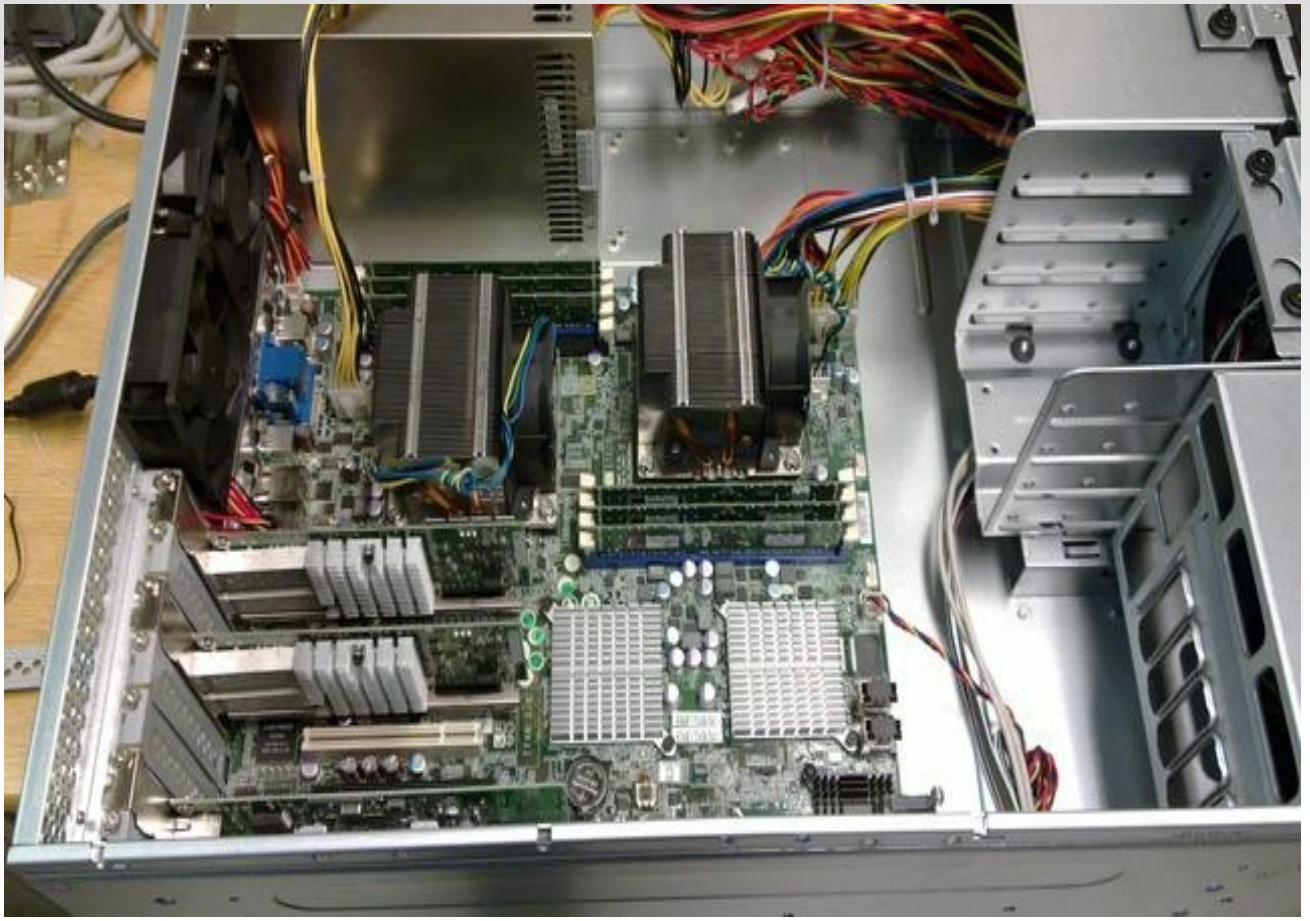
R & D related to bifrost

Close collaboration with Linux networking Developers and industry

NAPI (3 years) now in most network drivers
Pktgen testing in linux,
fib_trie, (routing algo)
routing stats to monitor and understand network

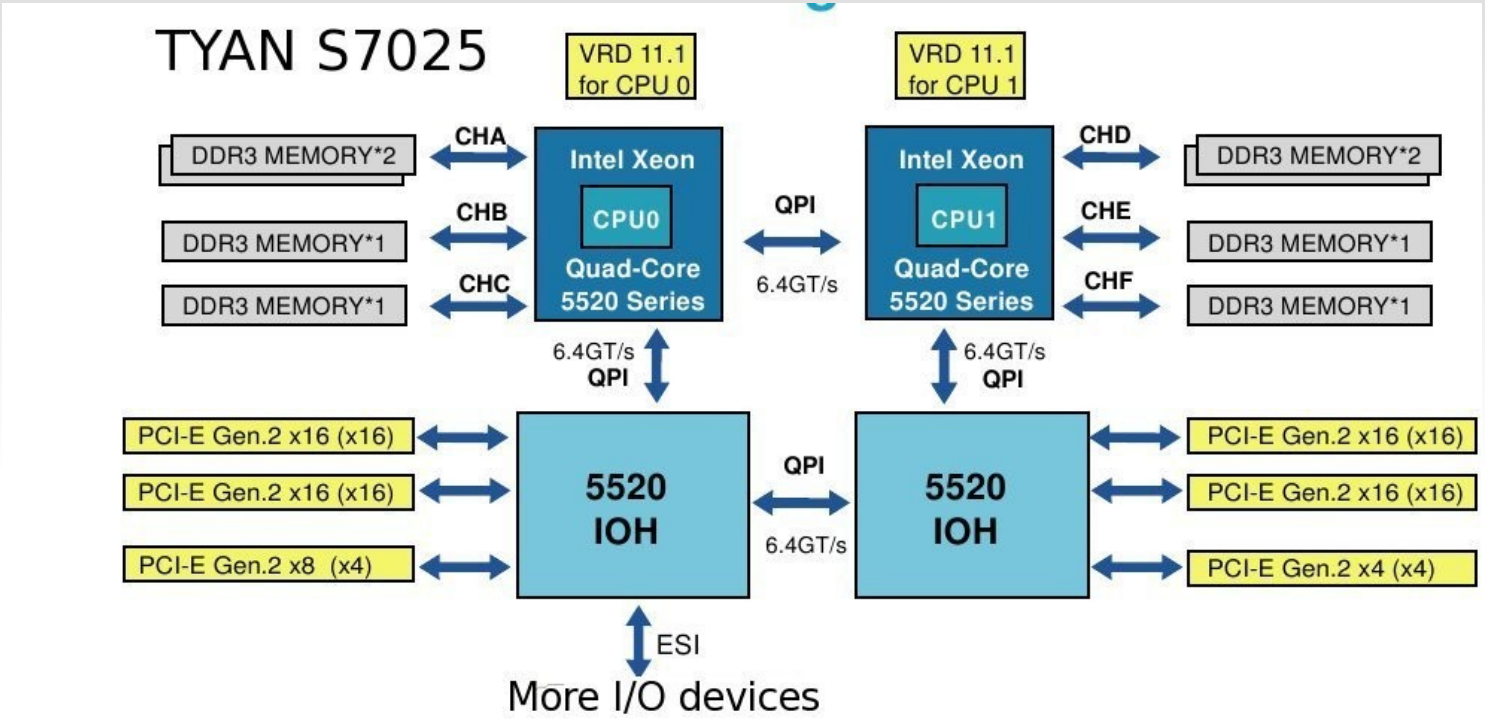
Etc, etc

Hi-End Hardware



XEON 2 x E5630
TYAN S7025 Motherboard
Intel 82599

Block hw structure



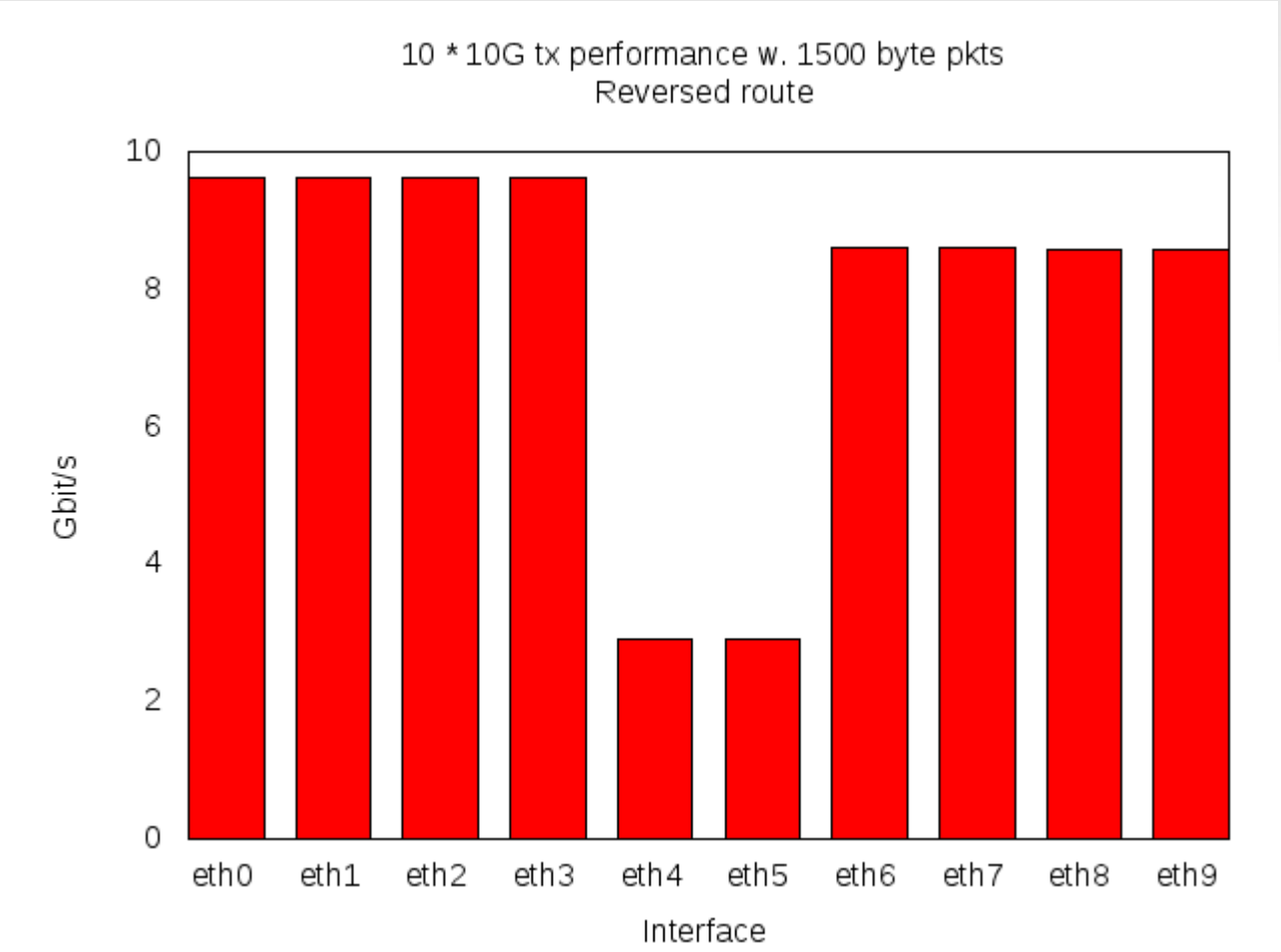
Recent project IIS

- DMA 93 Gbit/s TX
- Separation based HW classifier. BGP and ssh

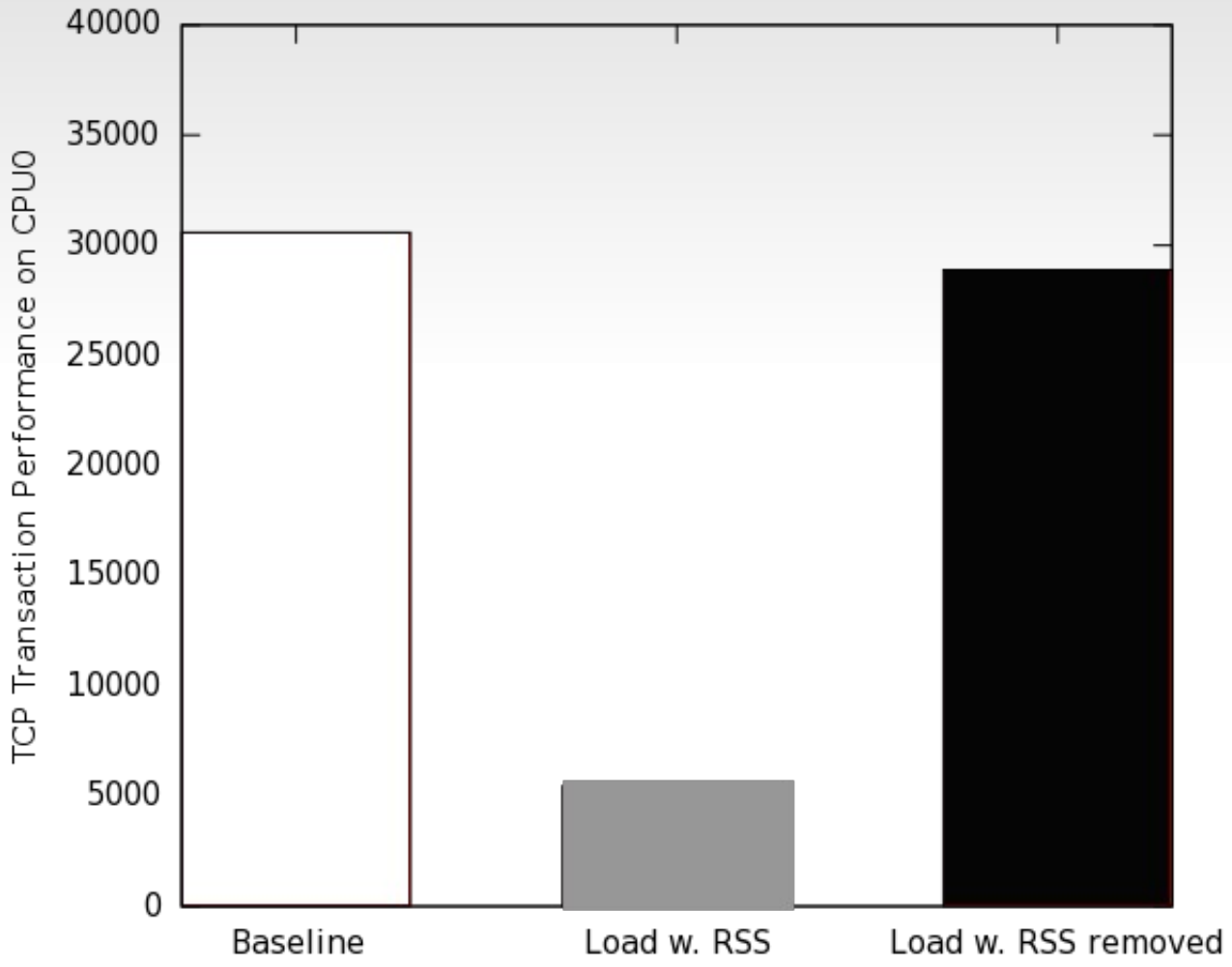
  In control plane (CPU0)

TX w. 10 * 10g ports 93Gb/s “Optimal”

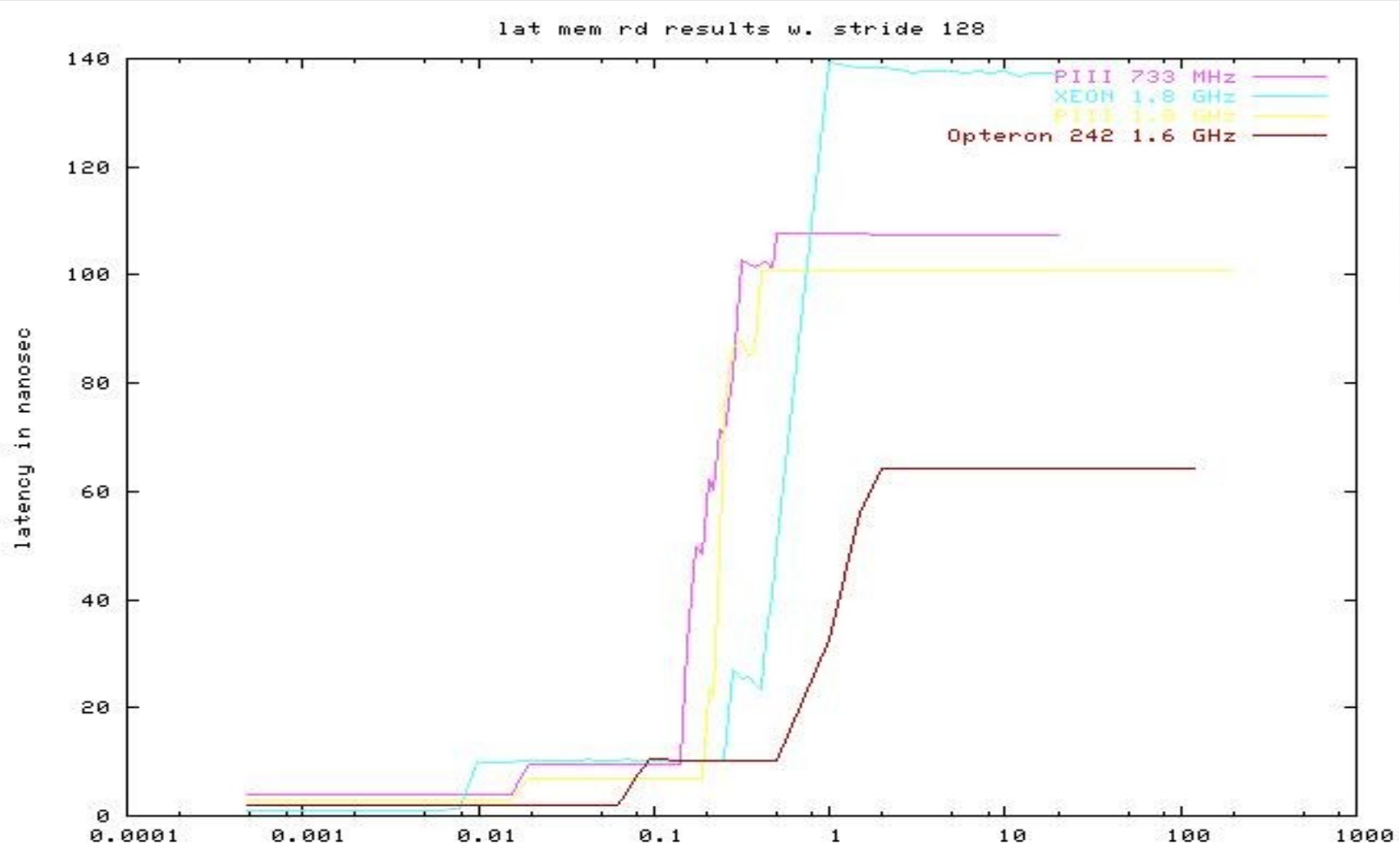
TX w. 10 * 10g ports 93Gb/s “unoptimal”



Network flow separation for bgp, ssh

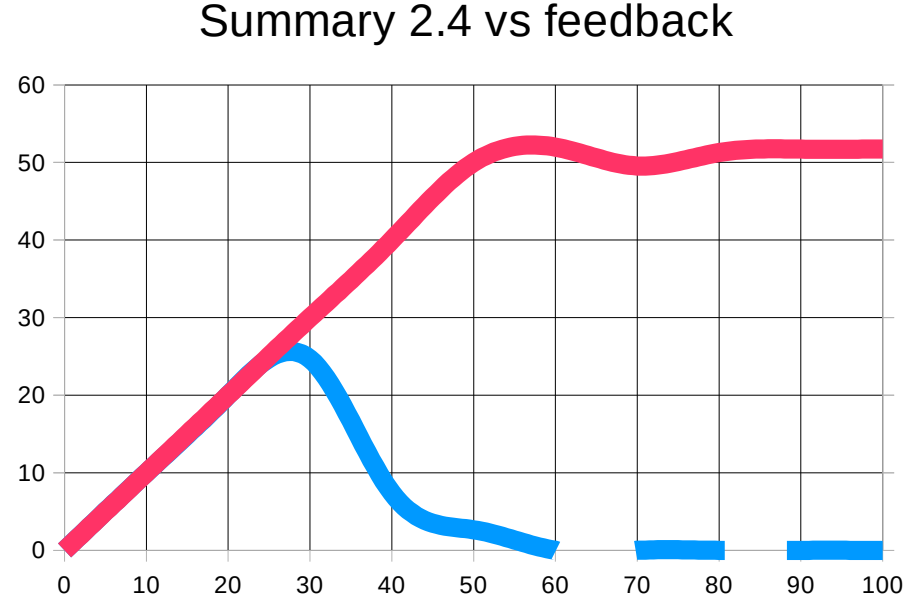


Cache effect/Performance using lmbench



NAPI Effect

- Inelegant handling of heavy net loads
  - System collapse
- Scalabiity affected
  - System and number of NICS
    - A single hogger netdev can bring the system to its knees and deny service to others



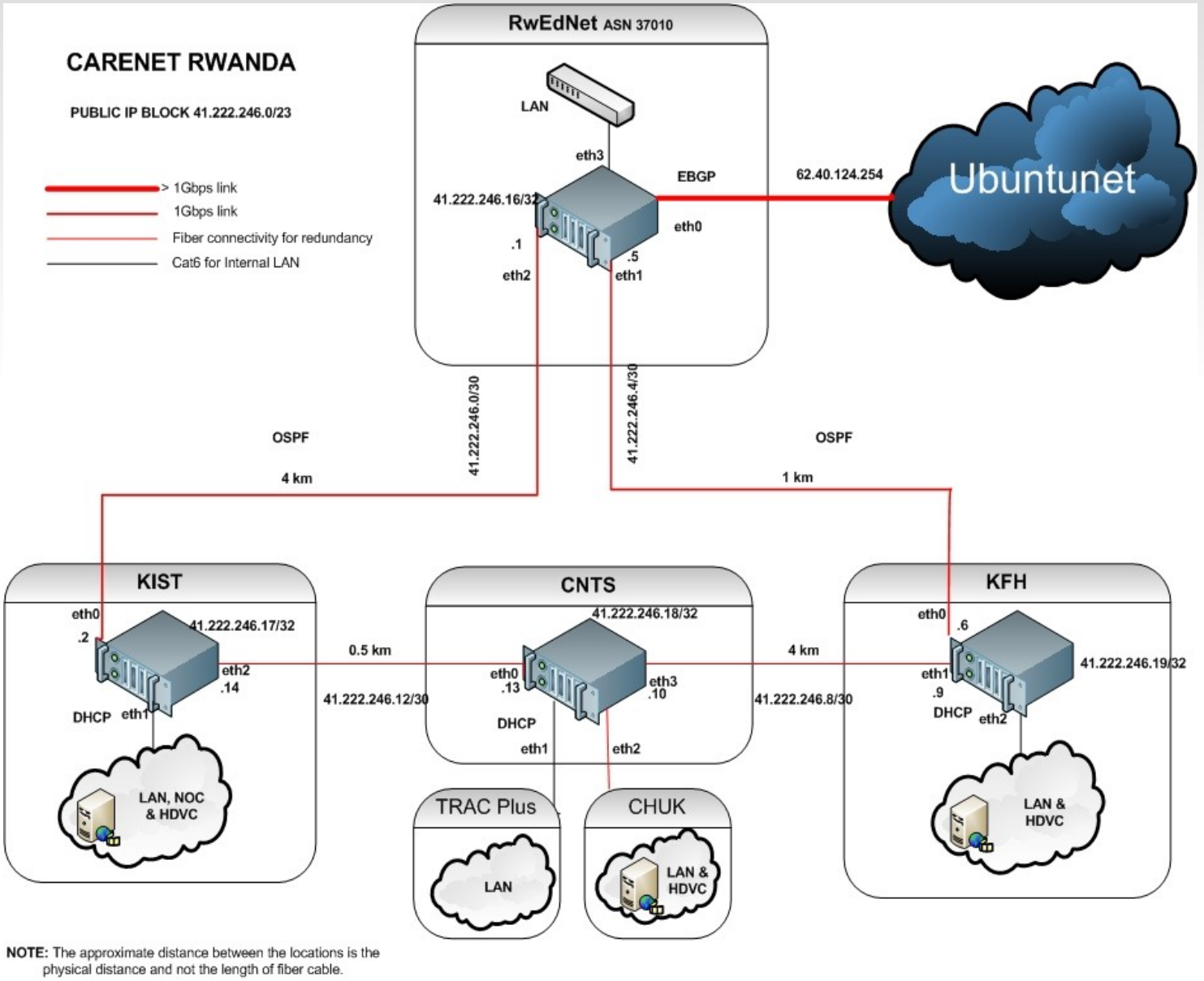
March 15 report on lkml

Thread: "How to optimize routing perfomance"

reported by Marten.Wikstron@framsfab.se

- Linux 2.4 peaks at 27Kpps
- Pentium Pro 200, 64MB RAM

Rwanda example



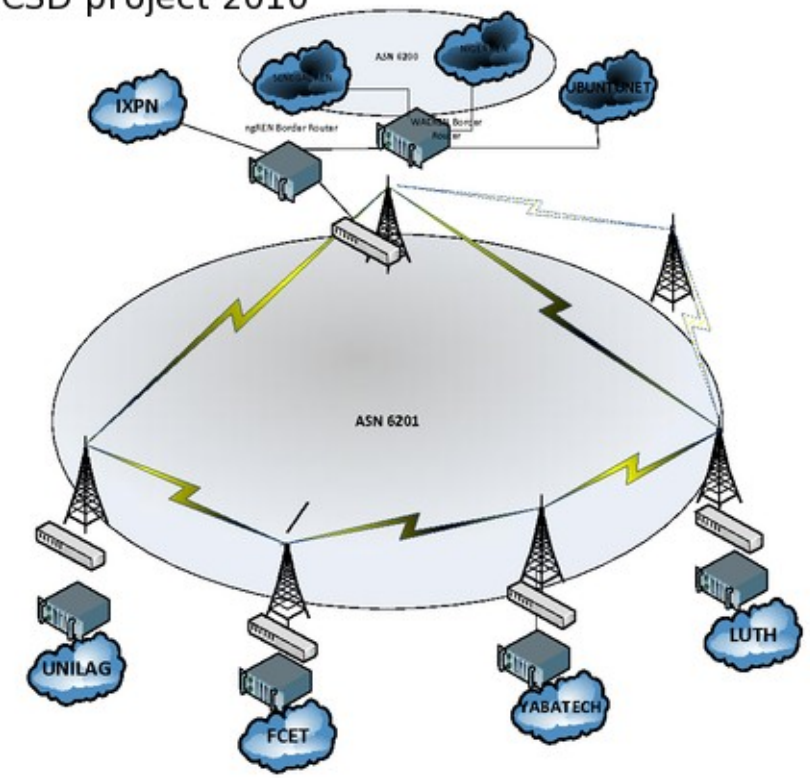
Lagos next

Carenet WACREN

CSD summer 2010

This document represents the topology of pilot phase of EKO-connect project based on wireless links

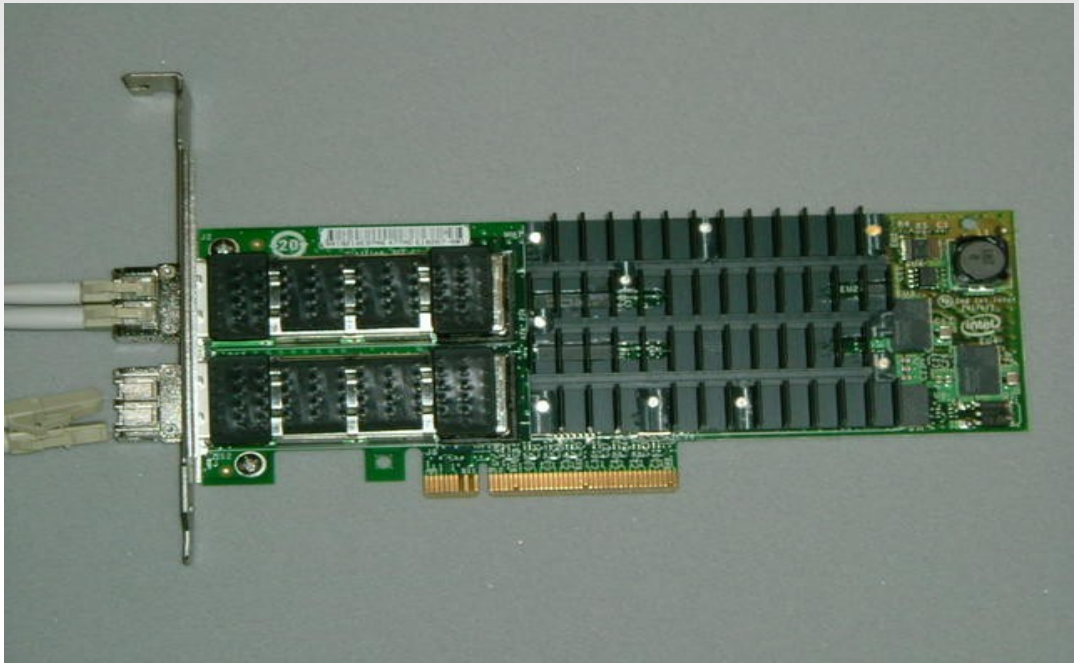
KTH/CSD project 2010



Lab testing HW and SW

Not all were selected...

Hardware - NIC



Intel 10g board Chipset 82599

Open chip specs. Thanks Intel!

KTH/CSD

Current focus

Optical to Open Source Router
Low-Power & Renewable Energy
Virtualization/Namespaces
Performance
Cost

Low-Power Development

Number of packet per second per Watt
but
Still addressing infrastructure.

So need for SFP fiber modules, multiqueue etc

Low-Power Development
Some ideas

Power consumption
SuperMicro X7SPA @ 16.5 Volt with picoPSU

Watt	Test
1.98	Power-Off
13.53	Idle
14.35	1 core
15.51	2 Core
15.84	3 Core
16.50	4 Core

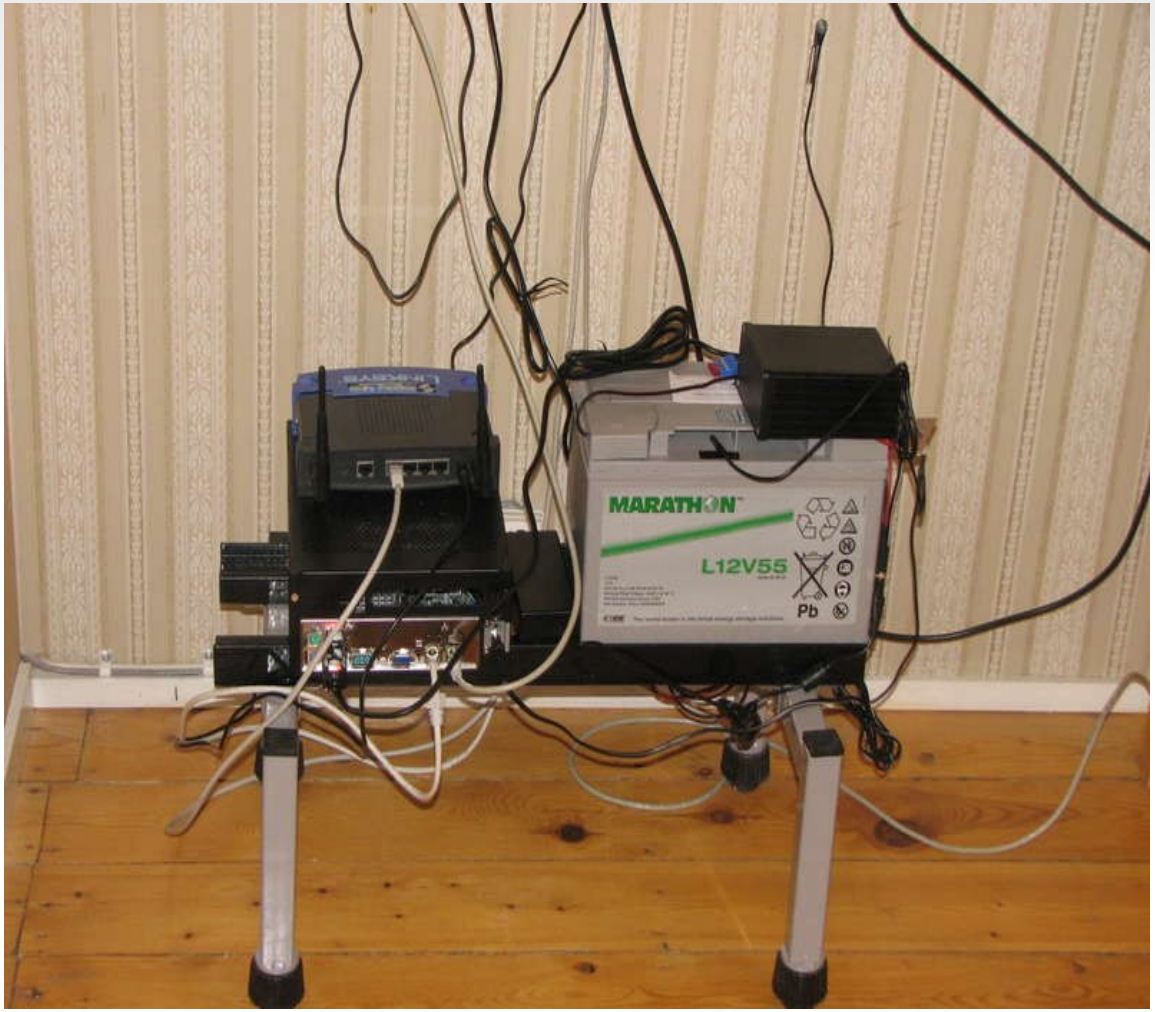
Routing Performance about 500.000 packet/sec
in optimal setup.

Low Power NIC

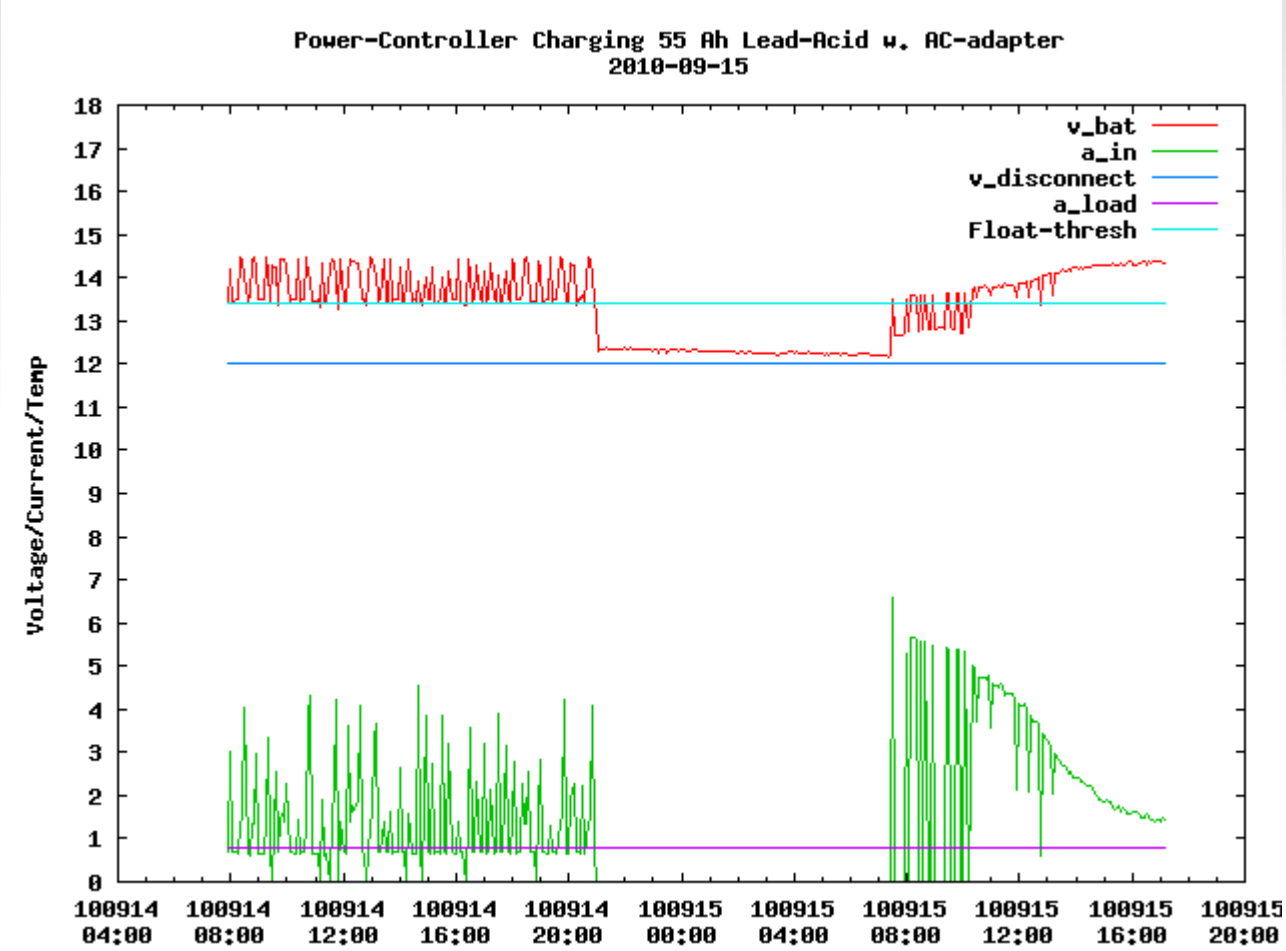
Remember we talk server chips

We're waiting for Intel 82580 for GE

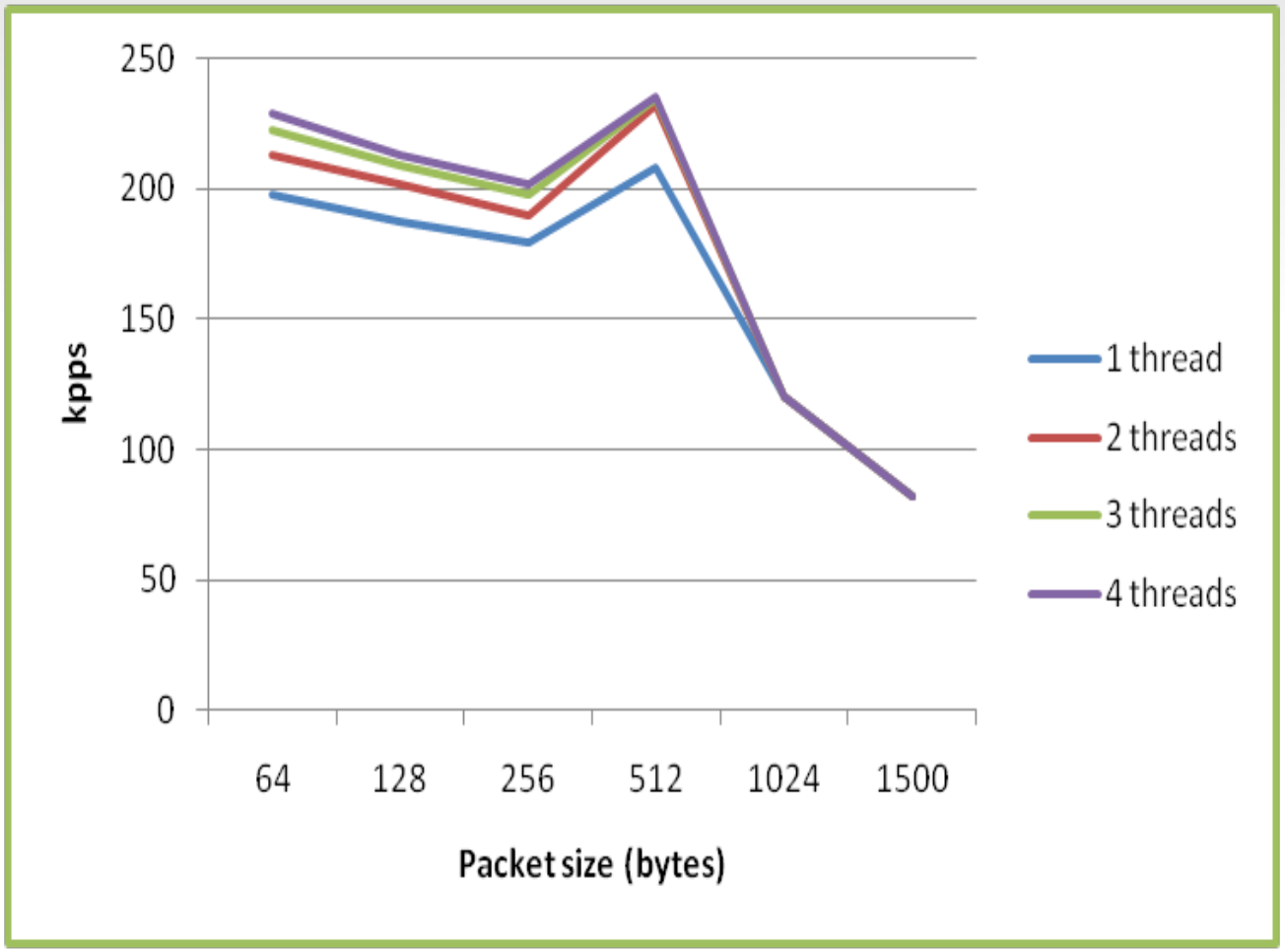
Example herjulf.se
14 Watt by 55Ah battery
bifrost/USB + lowpower disk



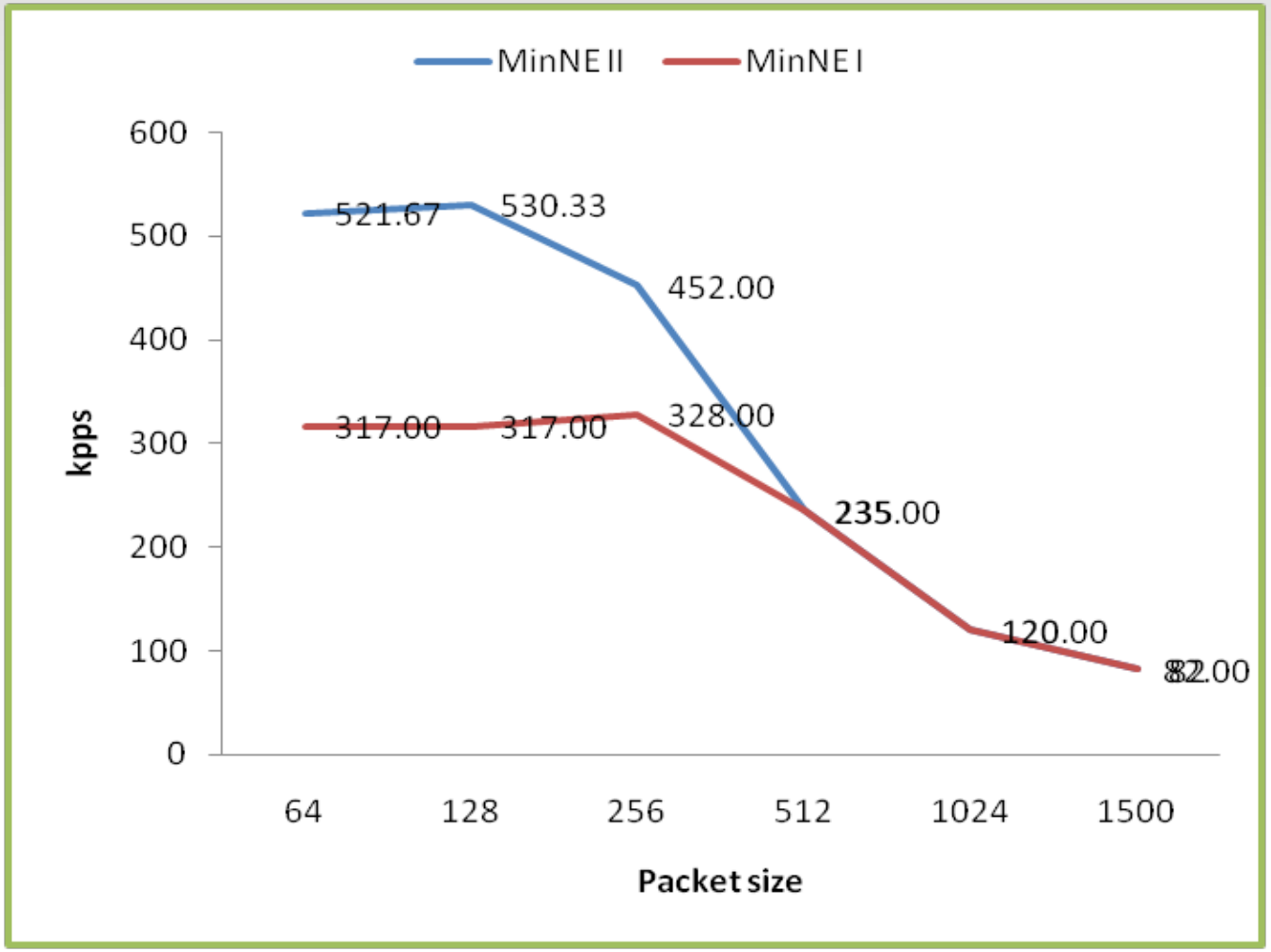
Running on a 55 Ah Battery



PPS per thread



Packet Performance



Optical modules

Optical sender and receiver in one module

SFP 1G
XFP 10G
SFP+ 10G

DOM - Optical Monitoring

Optical modules can support optical link monitoring
RX, TX power, temperatuers, alarms etc

Newly added support to Bifrost/Linux

DOM

ethtool -D eth3

Int-Calbr: Avr RX-Power: RATE_SELECT: Wavelength: 1310 nm

Temp: 25.5 C

Vcc: 3.28 V

Tx-Bias: 20.5 mA

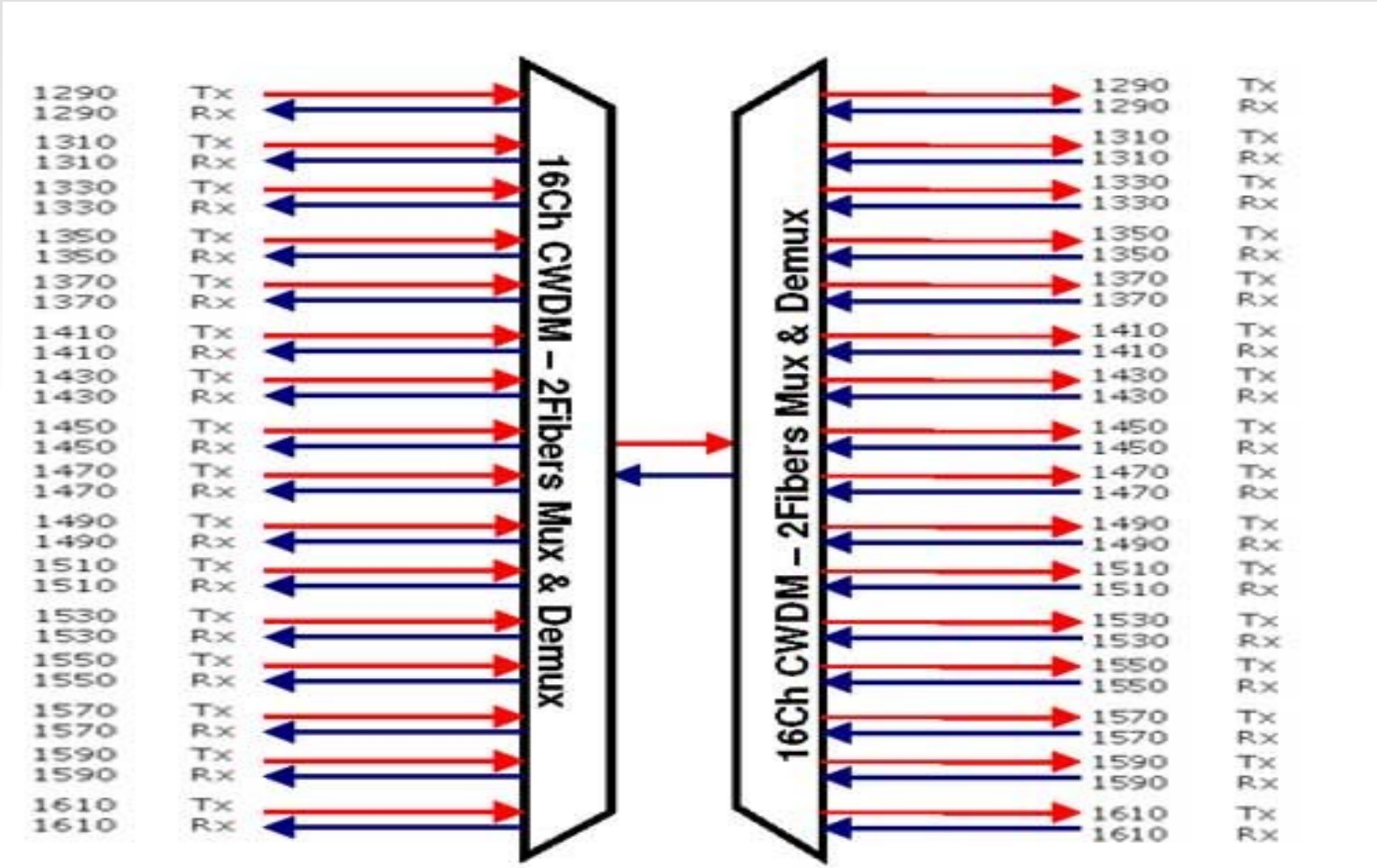
TX-pwr: -3.4 dBm (0.46 mW)

RX-pwr: -15.9 dBm (0.03 mW)

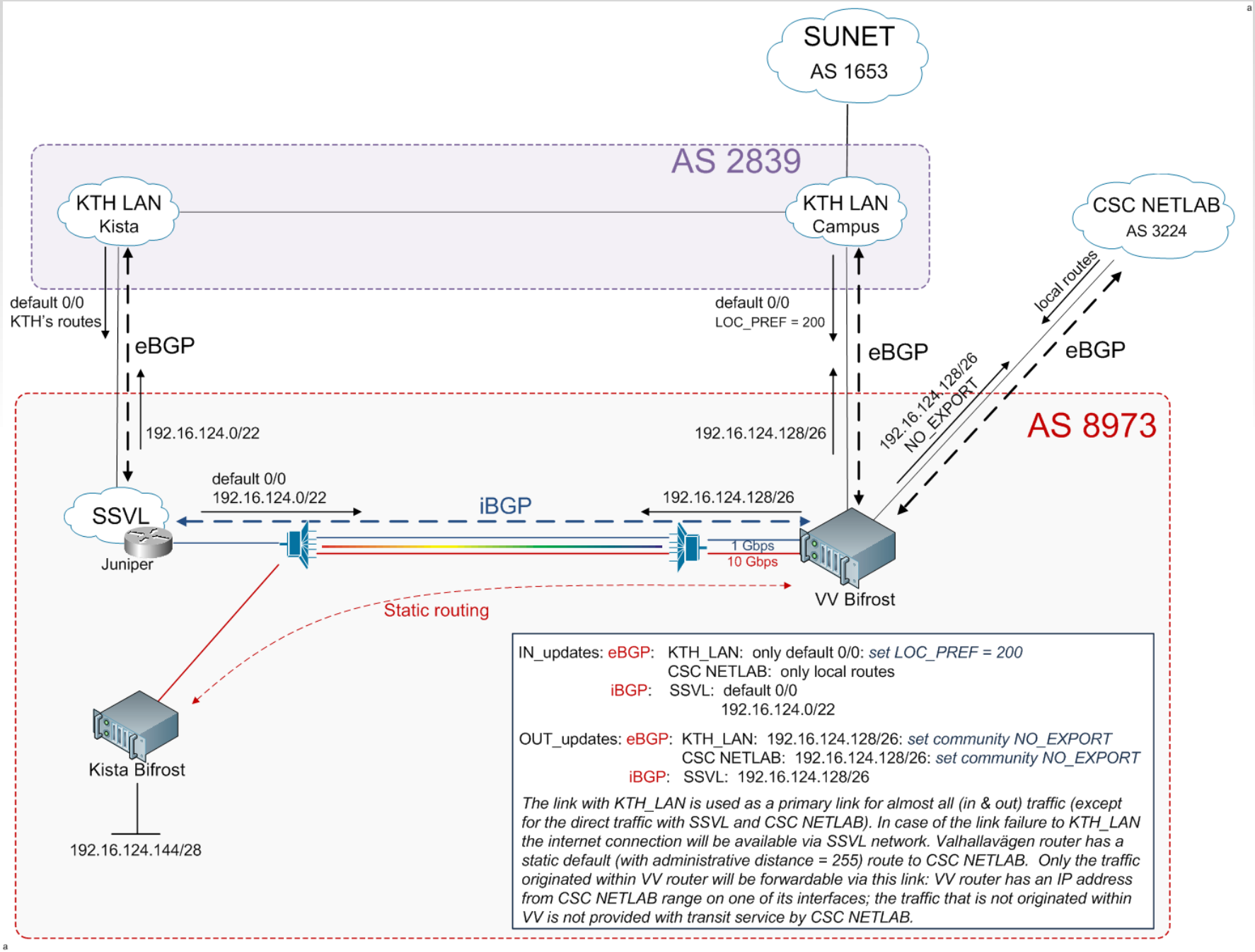
CWDM MUX/DEMUX 4 Ports

Price idea 600 Euro

CWDM MUX/DEMUX 16 Ports



Usage


SUNET
AS 1653
AS 2839
KTH LAN
Kista
KTH LAN
Campus
CSC NETLAB
AS 3224
default 0/0
KTH's routes
eBGP
default 0/0
LOC_PREF = 200
eBGP
local routes
eBGP
192.16.124.128/26
NO_EXPORT
AS 8973
192.16.124.0/22
192.16.124.128/26
default 0/0
192.16.124.0/22
iBGP
192.16.124.128/26
SSVL
Juniper
1 Gbps
10 Gbps
VV Bifrost
Static routing
Kista Bifrost
192.16.124.144/28
IN_updates: eBGP: KTH_LAN: only default 0/0: set LOC_PREF = 200
CSC NETLAB: only local routes
iBGP: SSVL: default 0/0
192.16.124.0/22
OUT_updates: eBGP: KTH_LAN: 192.16.124.128/26: set community NO_EXPORT
CSC NETLAB: 192.16.124.128/26: set community NO_EXPORT
iBGP: SSVL: 192.16.124.128/26
The link with KTH_LAN is used as a primary link for almost all (in & out) traffic (except for the direct traffic with SSVL and CSC NETLAB). In case of the link failure to KTH_LAN the internet connection will be available via SSVL network. Valhallavägen router has a static default (with administrative distance = 255) route to CSC NETLAB. Only the traffic originated within VV router will be forwardable via this link: VV router has an IP address from CSC NETLAB range on one of its interfaces; the traffic that is not originated within VV is not provided with transit service by CSC NETLAB.

That's all

Questions?

Project's ideas

DOM 4-port GIGE card/Linux igb driver
GateWay Optical/WiFi
Energy – Router Power Reduction/Linux
Energy – Ultracaps/Solar or UPS