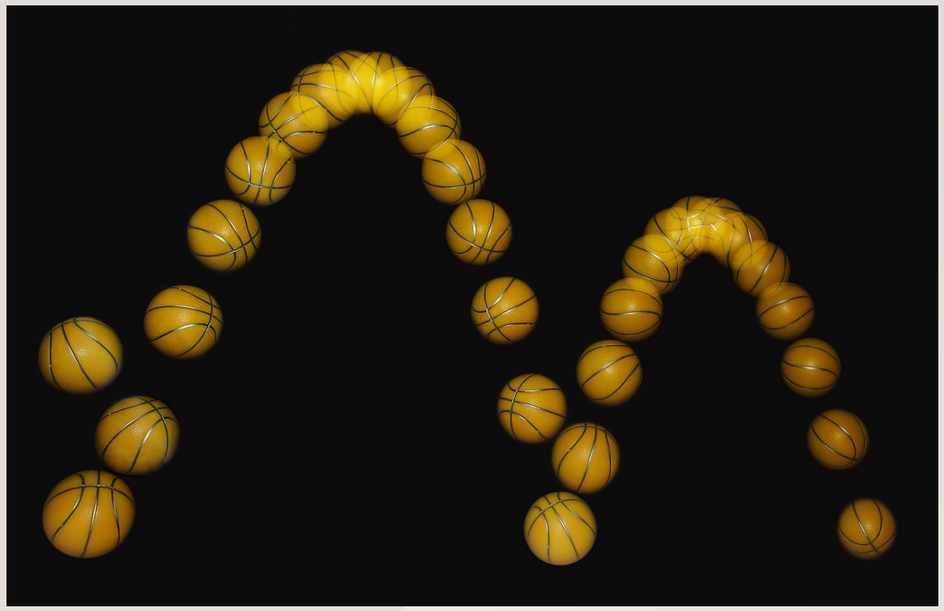
Stroboscoping The Internet

Olav Kvittem
UNINETT
03.11.16

Picture from wikipedia.org

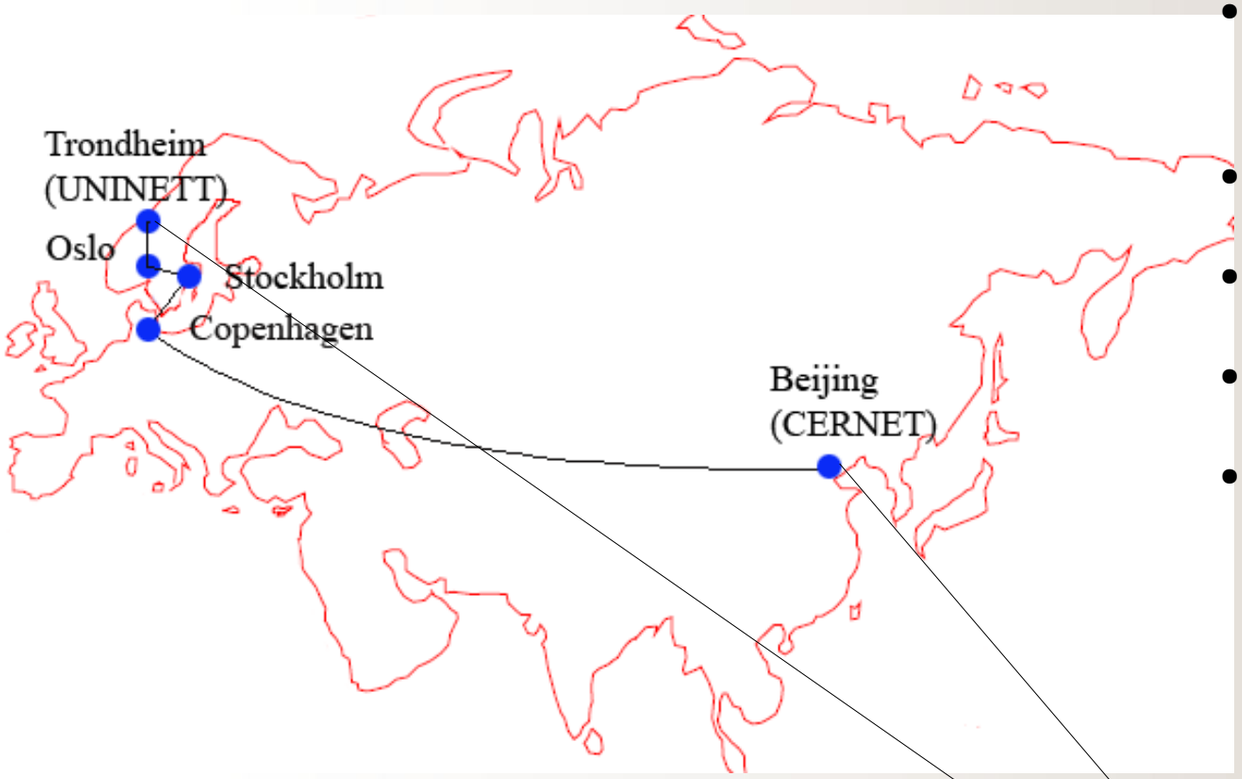


Overview

- Measureing global end to end quality
- Focus on routing convergence time
- Tools
- Metods
- Results – routing sucks
  - A tool to improve routing

UNINETT

Dragonlab



- Dragon-lab project at Tsinguha Univ.
- Dragonlab confederation is affiliated
- Traffic measurements
- with CERNET, U. Auckland, NTE (commercial internet operator)
- Micro dependability – rerouting effects
- Commercial and academic path
- Cernet is out
- Switch, RNP, Nordunet is in

UNINETT

Micro dependability

- Global routing is complex –
  - how well does The Internet perform on small time scales ?
  - We have triple redundancy but does it work
- Packet loss incurs on
  - Congestion
  - Fault detection
  - Route computation
  - Route change propagation
    - Micro loops
- Tuning
  - Signal loss detection time
  - Routing updates – keepalive
  - BGP – active/passive mode
  - IGP – Fast hello
  - IP Fast Reroute

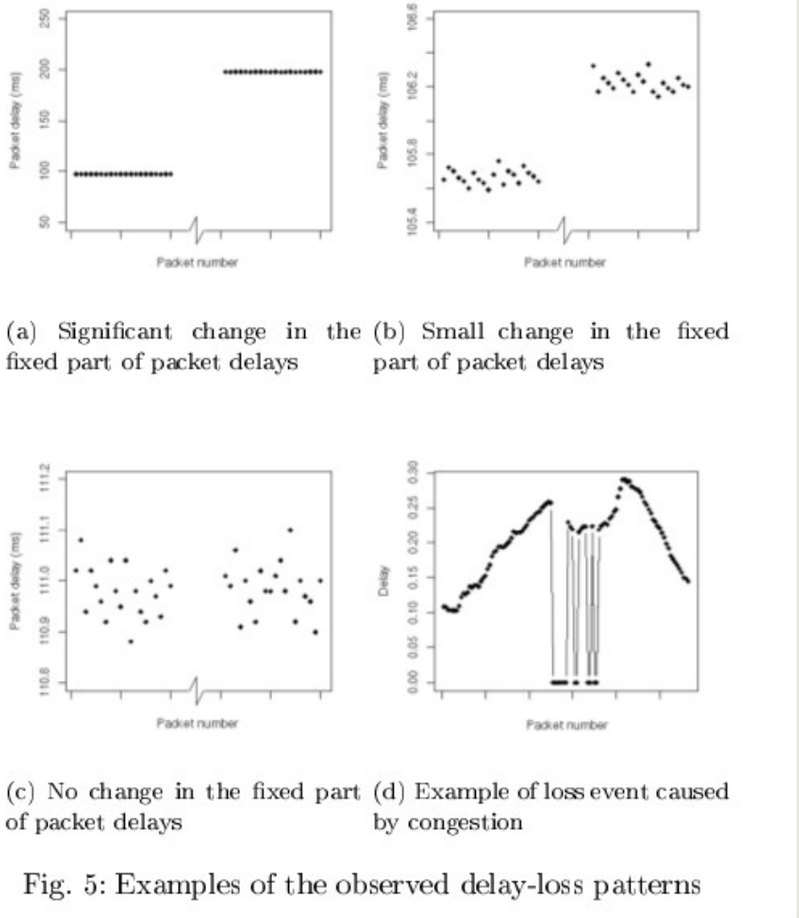
UNINETT

Measurement setup

- Linux systems
- Rude/crude, traceroute
- Fine grained probing – 100(0) pps
  - Cover SDH 50ms protection time
- Problems :
  - precision on timestamps in user space o (ms)
    - Linux kernel changed to immediate receiver sceduling 4+ years ago
    - Buffering in nic's - o(100us)
  - NTP o(ms) – absolute delay difficult to measure
    - Down to sub ms by tuning ntpd
    - Packet arrival time differences - OK
  - Delay vary with path
    - checking remaining TTL in packets to indicate route change

UNINETT

Loss analysis



(a) Significant change in the fixed part of packet delays

(b) Small change in the fixed part of packet delays

(c) No change in the fixed part of packet delays

(d) Example of loss event caused by congestion

Fig. 5: Examples of the observed delay-loss patterns

- How to separate congenstion from failure automatically ?
- Increasing + decreasing delay => congestion ?
- Flat delay => failure ?
- Changed ttl => rerouting

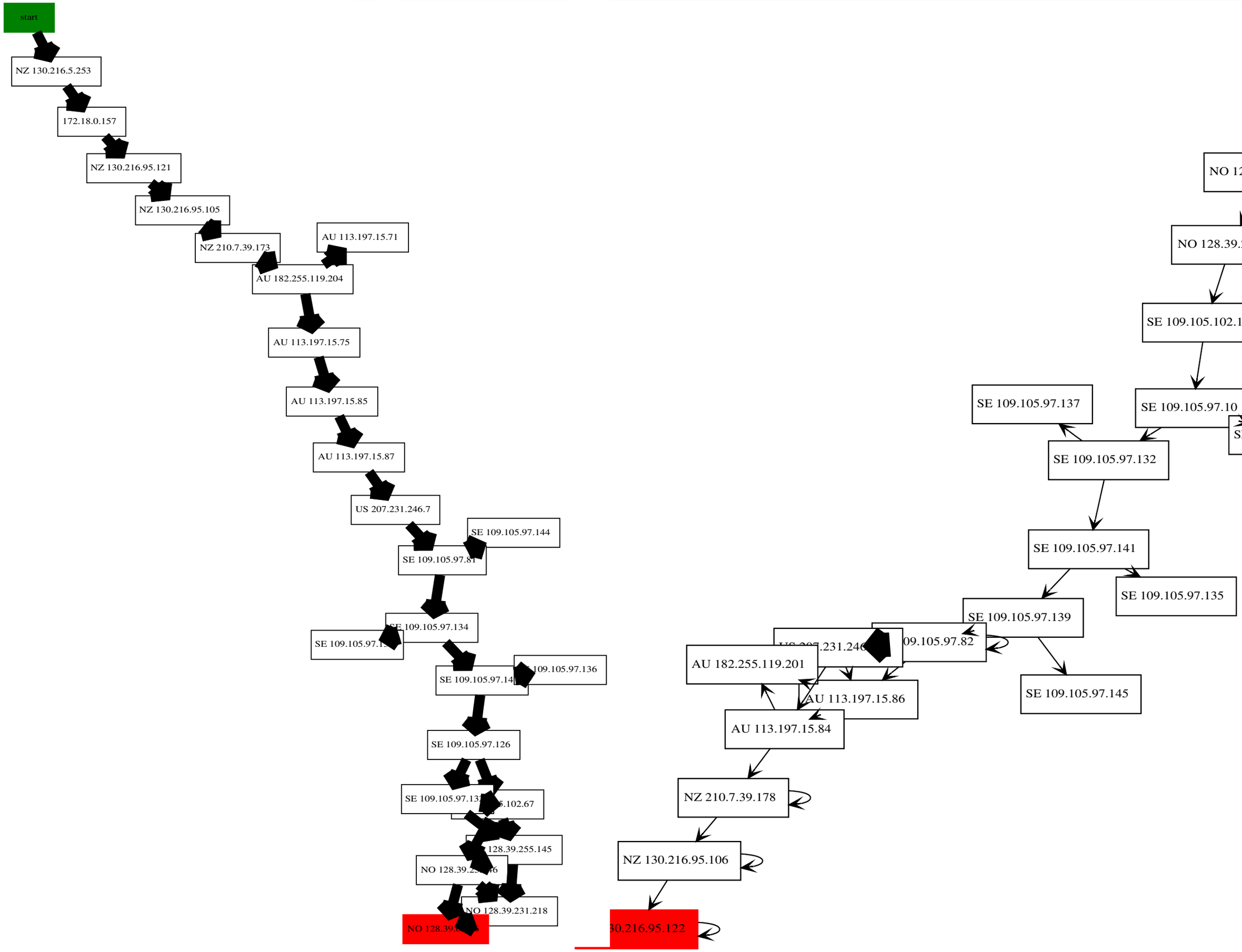
UNINETT

Traceroute patterns

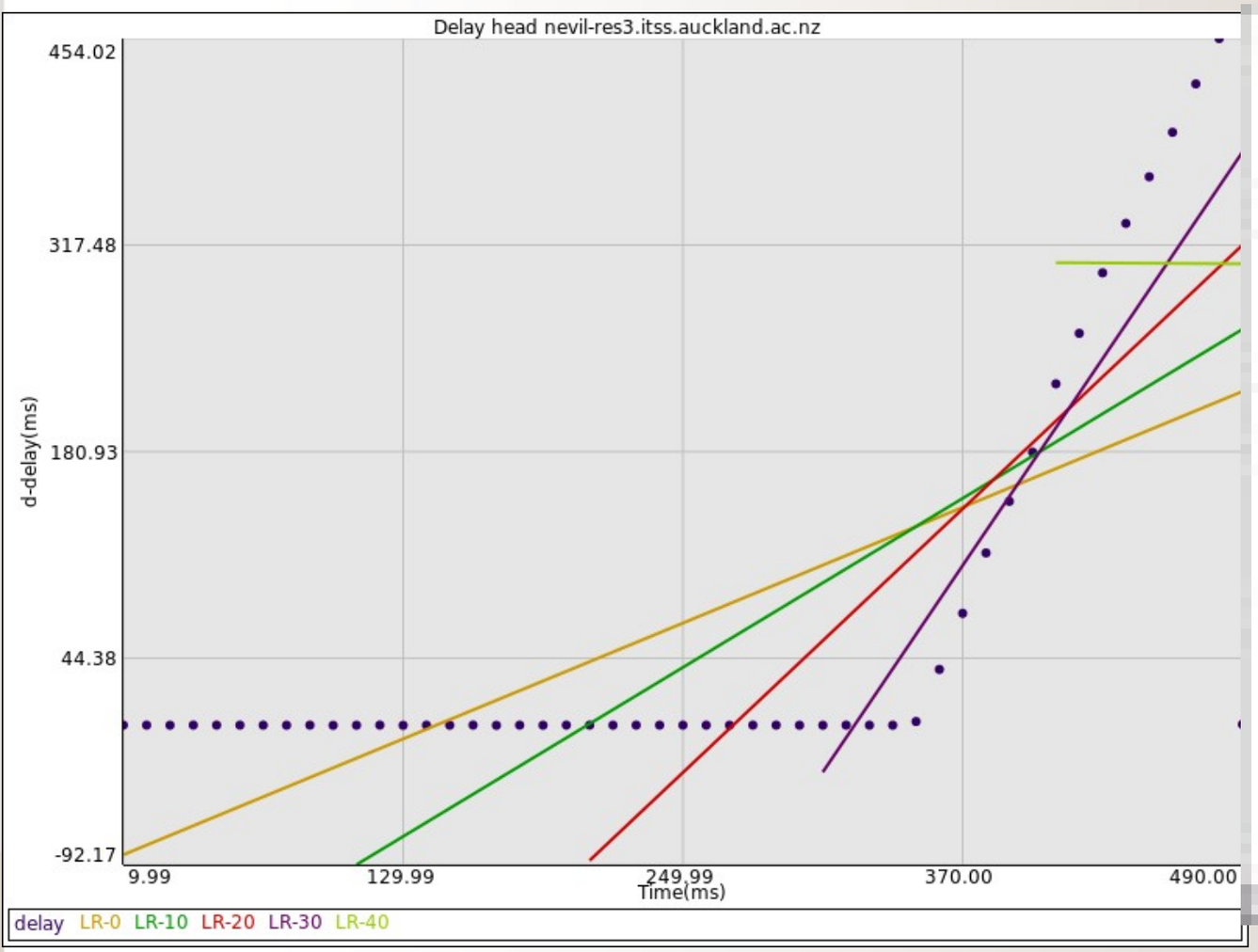
- Traceroute packets frequently dropped by routers ?
  - Normally not
- Traceroute method
  - Repeats 6 times
  - Changes ports every packet – applies to different flows
  - We see sammle of hops for different flows – not The Route
- Graphing routes –
  - add nodes for ip-addresses
  - Add edges for consequent nodes
  - Auto layout
- Geoip – location of address owner
- Shows equal cost path sharing paths
  - Alternate routes visible if longer that traceroute interval (1min)

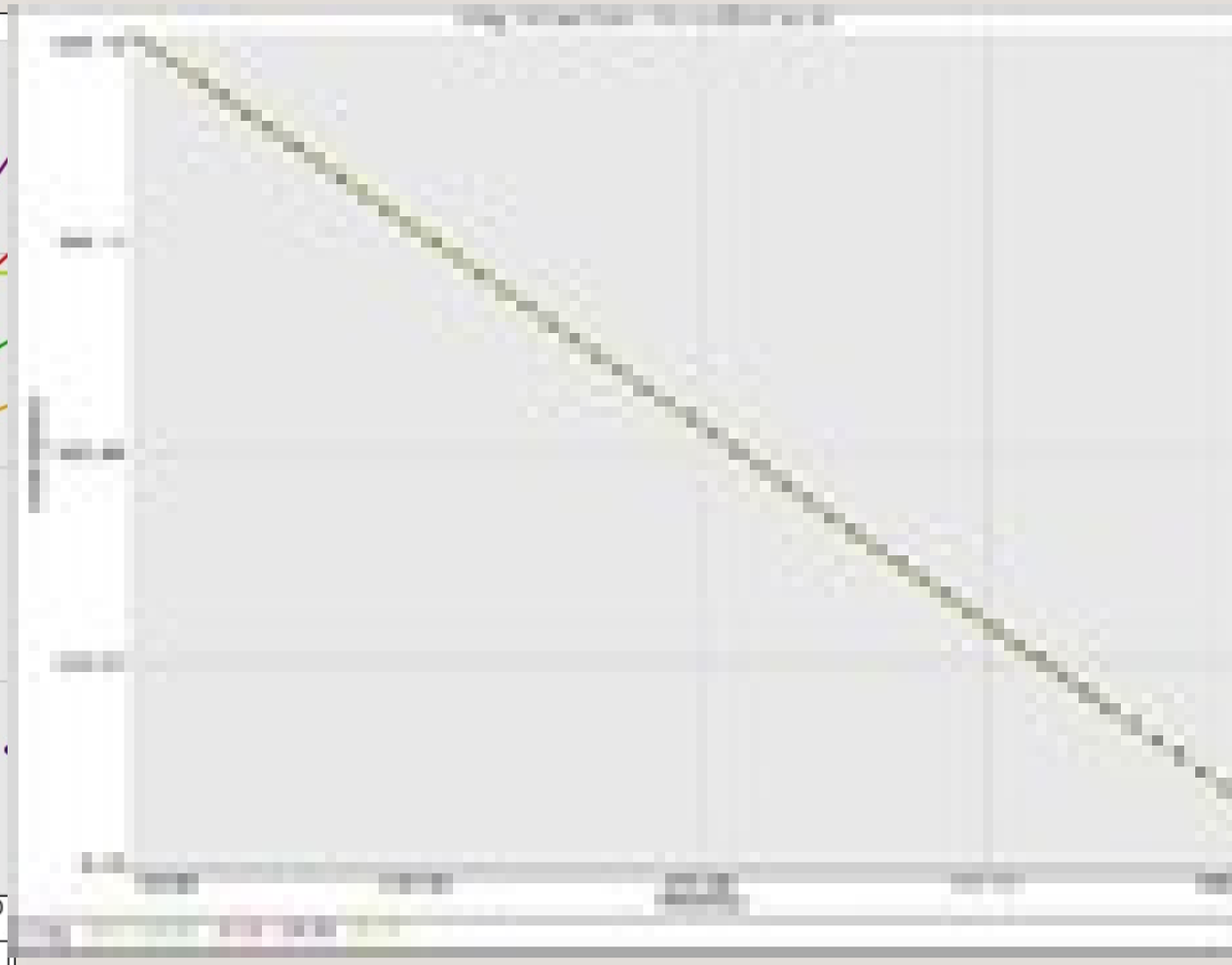
UNINETT

Uninett to NZ



Delay patterns around loss





graph/m-trondheim-mp.png

01:51:49 head 1 1	01:51:49 tail 1 2	04:18:00 head 1 3	04:18:00 tail 1 4	04:18:01 head 1 5	04:18:01 tail 1 6	06:15:06 head 1 7	06:15:06 tail 1 8	06:15:07 head 1 9
06:15:07 head 2 1	06:15:07 tail 2 2	06:15:07 head 2 3	06:15:07 tail 2 4	06:15:09 head 2 5	06:15:09 tail 2 6	06:15:09 head 2 7	06:15:09 tail 2 8	06:15:09 head 2 9
06:15:10 head 3 1	06:15:10 tail 3 2	06:15:10 head 3 3	06:15:10 tail 3 4	06:15:10 head 3 5	06:15:10 tail 3 6	06:15:11 head 3 7	06:15:11 tail 3 8	06:15:11 head 3 9
06:15:11 head 4 1	06:15:11 tail 4 2	06:15:12 head 4 3	06:15:12 tail 4 4	06:15:12 head 4 5	06:15:12 tail 4 6	06:15:13 head 4 7	06:15:13 tail 4 8	06:15:14 head 4 9
06:15:14 head 5 1	06:15:14 tail 5 2	06:15:14 head 5 3	06:15:14 tail 5 4	06:15:15 head 5 5	06:15:15 tail 5 6	08:04:23 head 5 7	08:04:23 tail 5 8	11:06:26 head 5 9
11:09:30 head 6 1	11:09:30 tail 6 2	11:10:04 head 6 3	11:10:04 tail 6 4	11:11:42 head 6 5	11:11:42 tail 6 6	11:12:54 head 6 7	11:12:54 tail 6 8	15:14:34 head 6 9

Dependability measurements

- Assumptions
  - NRENs don't have congestion
  - Longer gaps are dure to outage
  - Rerouteing may take unneccessary time
  - Outages are unwanted and uneccessary
- Investigating longer gaps
  - Locate router with traceroute logs
  - See alternate pats by looking at use counts
  - Estimate congestion by looking at jitter in rude and traceroute delay
  - Check the NOC logs to find matching incidents
-

Measures

- Stream of UDP packets with
  - sequence numbers
    - Length of outage
    - Reordering is not frequent but happens on bad days
  - time stamps
    - Last 50 packets before and after outage
    - Minimum delay in latest 1000 pakcets
    - Relative delay rise – not sensitive to clock skew
    - Regression coeffcient to see rise of delay
    - Jitter

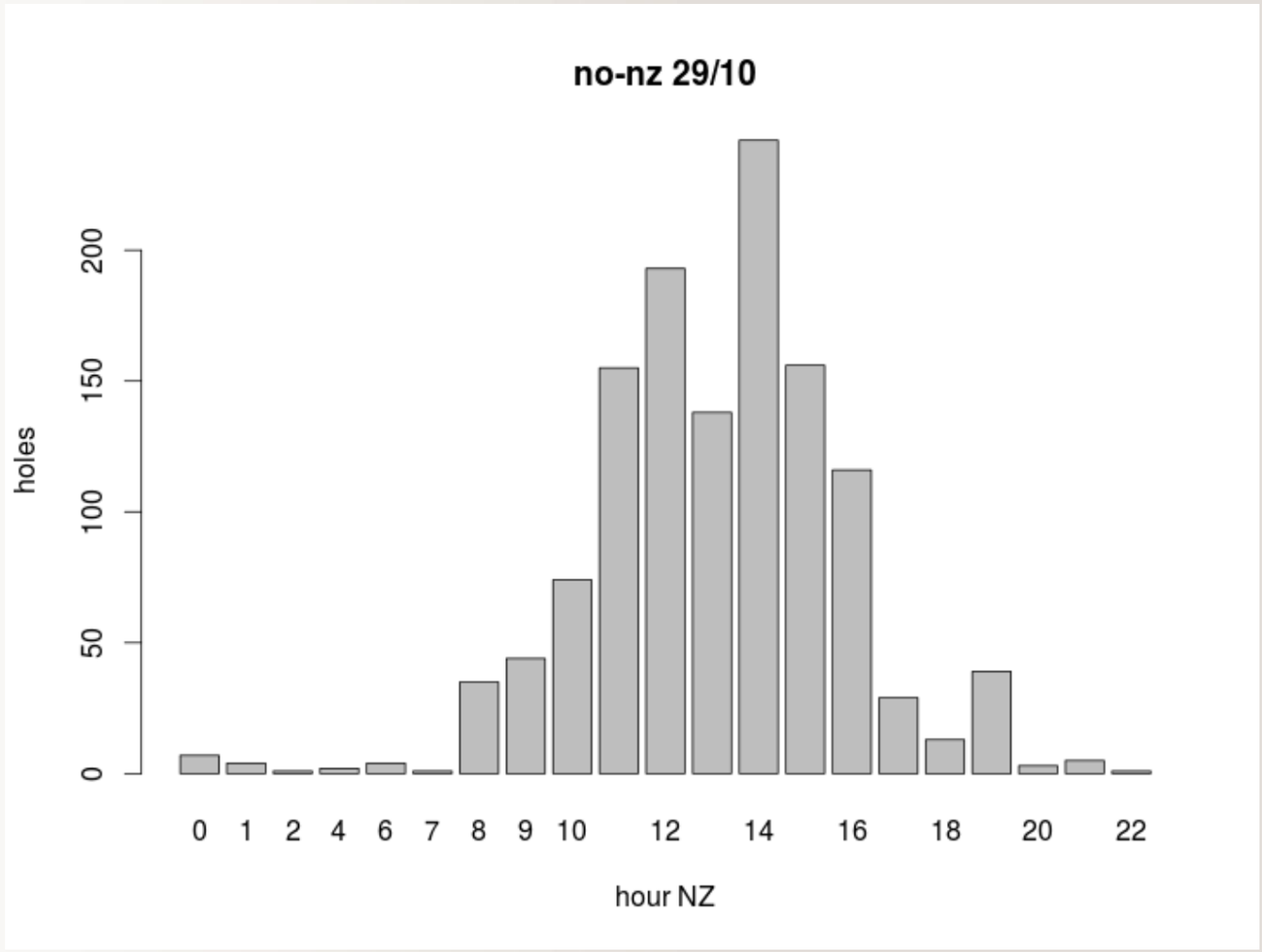
UNINETT

Cases

- Saw 130seconds outage when rebooting border router
  - Passive BGP alternate path had to be reannounced
  - IP Fast Reroute brought it down to a few seconds
- Unstable fibre caused 10 times 1 min outage
  - Take down path manually
  - Update of full forwarding table takes 76 seconds
  - partial rewrite feature in newer routers

UNINETT

Loss day profile



Loss summary UNINETT router upgrade

eduroam info | UNINET × | GÉANT × | Aktivitet | Eureka × | iou2.uninett.no/mp-rud × |

iou2.uninett.no/mp-rude-sum/gap-sum-2016-02-12.txt

bergen-mp.uib.no	alesund-mp.hials.no	2016-02-12 20:45:52	1455309952.920081	1560
ntnu-mp.ntnu.no	alesund-mp.hials.no	2016-02-12 20:45:52	1455309952.926841	1120
bergen-mp.uib.no	drammen-mp.hibu.no	2016-02-12 21:25:53	1455312353.090096	1230
bergen-mp.uib.no	forde-mp.hisf.no	2016-02-12 21:25:53	1455312353.090100	1230
bergen-mp.uib.no	gjovik-mp.hig.no	2016-02-12 21:25:53	1455312353.090108	1230
bergen-mp.uib.no	ntnu-mp.ntnu.no	2016-02-12 21:25:53	1455312353.090141	1230
bergen-mp.uib.no	pil32-mp.hioa.no	2016-02-12 21:25:53	1455312353.090149	1230
bergen-mp.uib.no	steinkjer-mp.hint.no	2016-02-12 21:25:53	1455312353.090165	1230
bergen-mp.uib.no	tromso-mp.uit.no	2016-02-12 21:25:53	1455312353.090176	1230
tromso-mp.uit.no	bergen-mp.uib.no	2016-02-12 21:25:53	1455312353.752486	129230
time.hint.no	bergen-mp.uib.no	2016-02-12 21:25:53	1455312353.758595	129190
porsgrunn-mp.hit.no	bergen-mp.uib.no	2016-02-12 21:25:53	1455312353.759205	129180
drammen-mp.hibu.no	bergen-mp.uib.no	2016-02-12 21:25:53	1455312353.760234	129180
oslo-mp.uio.no	bergen-mp.uib.no	2016-02-12 21:25:53	1455312353.760423	129180
fredrikstad-mp.hiof.no	bergen-mp.uib.no	2016-02-12 21:25:53	1455312353.760810	129180
grimstad-mp.uia.no	bergen-mp.uib.no	2016-02-12 21:25:53	1455312353.761274	129190
kristiansand-mp.uia.no	bergen-mp.uib.no	2016-02-12 21:25:53	1455312353.761684	129190
halden-mp.hiof.no	bergen-mp.uib.no	2016-02-12 21:25:53	1455312353.762032	129160
alesund-mp.hials.no	bergen-mp.uib.no	2016-02-12 21:25:53	1455312353.763051	129160
forde-mp.hisf.no	bergen-mp.uib.no	2016-02-12 21:25:53	1455312353.763168	129170
kunnskapsv-mp.hioa.no	bergen-mp.uib.no	2016-02-12 21:25:53	1455312353.764316	129180
gjovik-mp.hig.no	bergen-mp.uib.no	2016-02-12 21:25:53	1455312353.764670	129160
pil32-mp.hioa.no	bergen-mp.uib.no	2016-02-12 21:25:53	1455312353.764761	129160
ntnu-mp.ntnu.no	bergen-mp.uib.no	2016-02-12 21:25:53	1455312353.766872	129190
molde-mp.hiMolde.no	bergen-mp.uib.no	2016-02-12 21:25:53	1455312353.767300	129170
stavanger-mp.uis.no	bergen-mp.uib.no	2016-02-12 21:25:53	1455312353.767370	129160
stord-mp.hsh.no	bergen-mp.uib.no	2016-02-12 21:25:53	1455312353.767847	129130
sogndal-mp.hisf.no	bergen-mp.uib.no	2016-02-12 21:25:53	1455312353.768175	129180
haugesund-mp.hsh.no	bergen-mp.uib.no	2016-02-12 21:25:53	1455312353.768636	129170
bergen-mp.uib.no	alesund-mp.hials.no	2016-02-12 21:25:53	1455312353.770081	550
bergen-mp.uib.no	fredrikstad-mp.hiof.no	2016-02-12 21:25:53	1455312353.770102	550
bergen-mp.uib.no	grimstad-mp.uia.no	2016-02-12 21:25:53	1455312353.770111	550
bergen-mp.uib.no	halden-mp.hiof.no	2016-02-12 21:25:53	1455312353.770115	550
bergen-mp.uib.no	kristiansand-mp.uia.no	2016-02-12 21:25:53	1455312353.770127	550
bergen-mp.uib.no	kunnskapsv-mp.hioa.no	2016-02-12 21:25:53	1455312353.770131	550
bergen-mp.uib.no	molde-mp.himolde.no	2016-02-12 21:25:53	1455312353.770135	550
bergen-mp.uib.no	oslo-mp.uio.no	2016-02-12 21:25:53	1455312353.770144	550
bergen-mp.uib.no	porsgrunn-mp.hit.no	2016-02-12 21:25:53	1455312353.770151	550
bergen-mp.uib.no	stavanger-mp.uis.no	2016-02-12 21:25:53	1455312353.770160	550
fredrikstad-mp.hiof.no	bergen-mp.uib.no	2016-02-12 21:37:58	1455313078.350810	310
grimstad-mp.uia.no	bergen-mp.uib.no	2016-02-12 21:37:58	1455313078.351269	310
kristiansand-mp.uia.no	bergen-mp.uib.no	2016-02-12 21:37:58	1455313078.351684	310
halden-mp.hiof.no	bergen-mp.uib.no	2016-02-12 21:37:58	1455313078.352032	310
kunnskapsv-mp.hioa.no	bergen-mp.uib.no	2016-02-12 21:37:58	1455313078.354324	310
time.hint.no	bergen-mp.uib.no	2016-02-12 21:37:58	1455313078.358595	420
haugesund-mp.hsh.no	bergen-mp.uib.no	2016-02-12 21:37:58	1455313078.358636	300
drammen-mp.hibu.no	bergen-mp.uib.no	2016-02-12 21:37:58	1455313078.360234	820
tromso-mp.uit.no	bergen-mp.uib.no	2016-02-12 21:37:58	1455313078.362488	410
alesund-mp.hials.no	bergen-mp.uib.no	2016-02-12 21:37:58	1455313078.363059	420
forde-mp.hisf.no	bergen-mp.uib.no	2016-02-12 21:37:58	1455313078.363182	420
gjovik-mp.hig.no	bergen-mp.uib.no	2016-02-12 21:37:58	1455313078.364670	410
pil32-mp.hioa.no	bergen-mp.uib.no	2016-02-12 21:37:58	1455313078.364762	410

TTL change

- TTL may change when routing changes
- Crude instrumented to log remaining TTL
- Most days have no change in end TTL
- Seen changes 10 times a day +-1

UNINETT

Presentation

- Log incidents/gaps to Elasticsearch
- Graph sum and count of gaps
  - Variation over time
  - If total indendable or not
  - Independability in PPM – because networks are very stable
- Search for combination of parameters

UNINETT

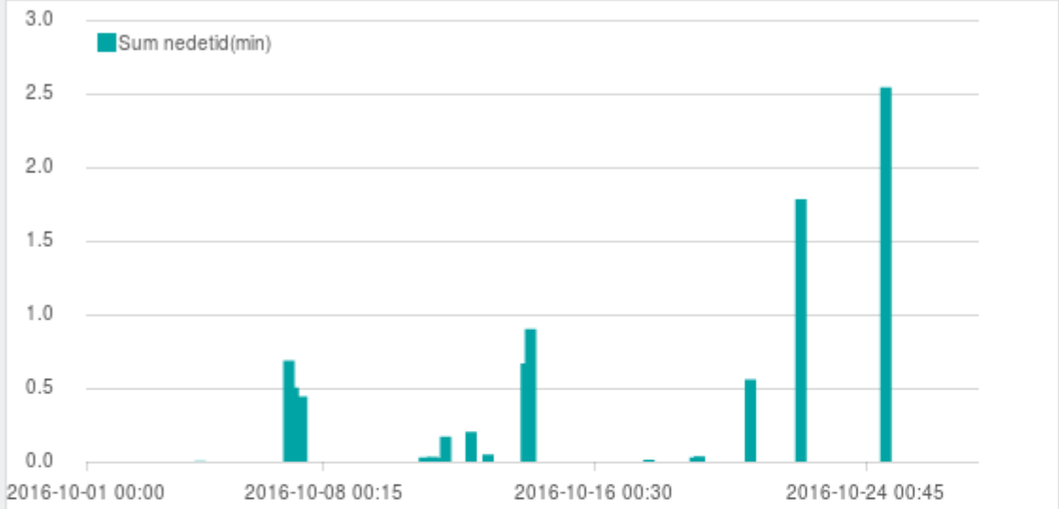
kibana

Discover | Visualize | Dashboard | Settings | Month to date

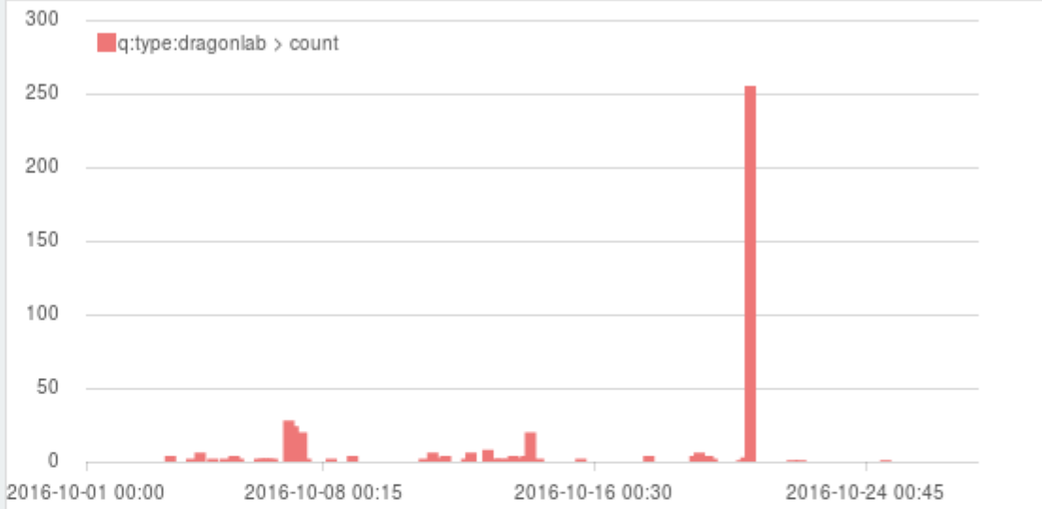
dragonlab nedetid

copenhagen AND zurich

dragonlab nedetid



Dragonlab antall gap per dag



dragonlab loss 300

1 2 3 4 5 ...10 »

Time	tloss	h_jit	h_delay	from	to
October 24th 2016, 16:56:27.000	152,480	0	19.832	copenhagen-mp	zurich-mp
October 22nd 2016, 04:32:07.000	106,830	0.003	19.886	copenhagen-mp	zurich-mp
October 14th 2016, 02:38:46.000	10,019.9	-0.002	19.731	copenhagen-mp	zurich-mp
October 14th 2016, 02:38:46.000	10,019.9	-0.002	19.731	copenhagen-mp	zurich-mp
October 14th 2016, 02:38:46.000	9,960	0	19.301	zurich-mp	copenhagen-mp

UNINETT

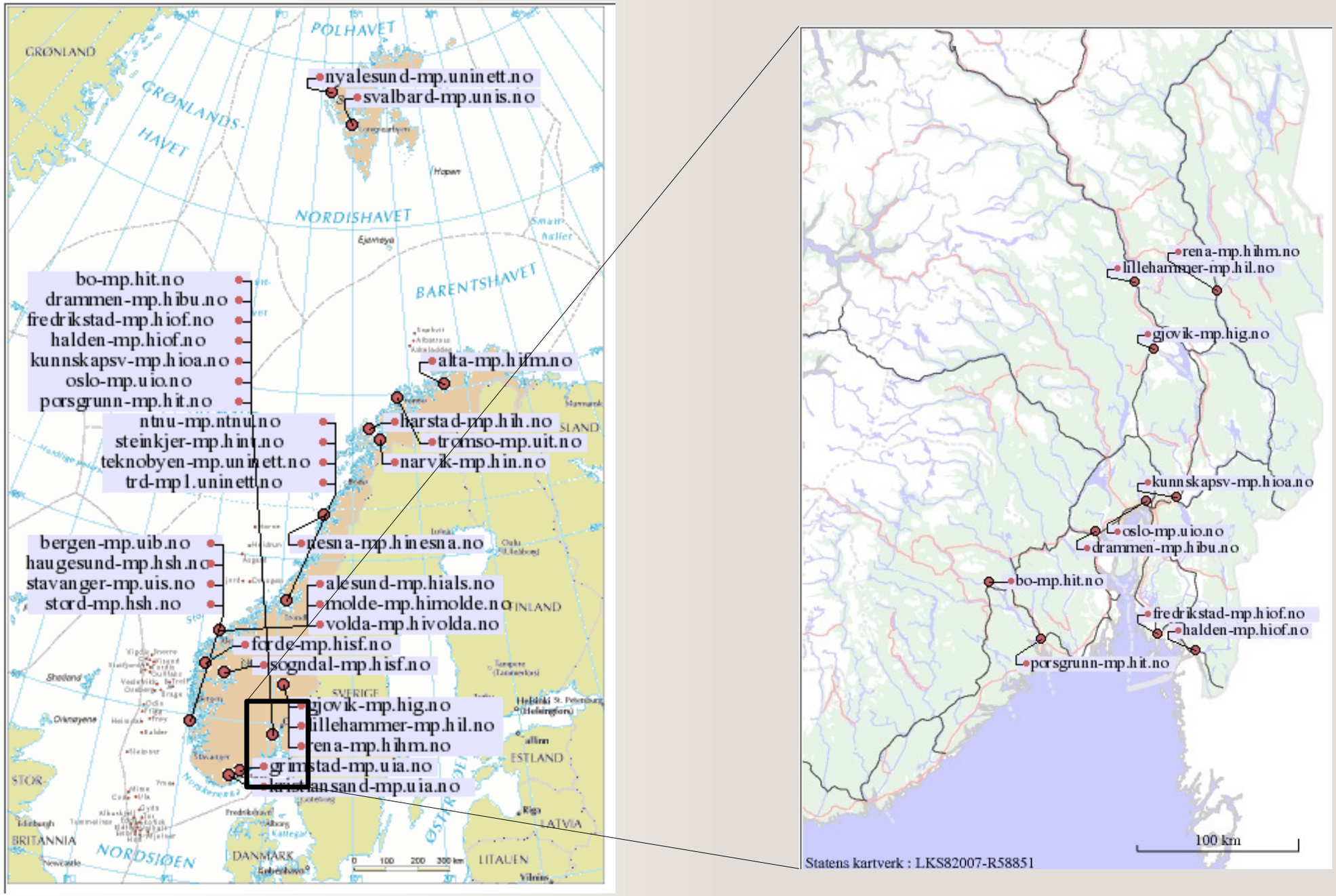
Status

- Contribution to 2 PHD's
- Separate congestion from other failures
- Find a indicative measure for reliability ?
-
- UNINETT operational service
- International
  - Brasil, Auckland, Zurich, Copenhagen, Stockholm, Trondheim
  - Measure Geant ?
  - More partners – anyone ?

Thanks

- Don't ask for software - yet
- Overall http://drift.uninett.no/dragonlab/
- Publications : https://openwiki.uninett.no/dragonlab:publ

UNINETT monitoring infrastructure



Calibrating linux timestamping

- Crude **reads** udp measurement packets and logs to **disk**
- Problem : input udp packet loss
- Extending IO-buffer did not help
- **Discovered udp buffer overflow**
  - Kernel write blocked for .5 seconds every ~ 30 seconds
  - to flush to disk ?
- Increase input udp buffers and monitor to catch overruns
  - Induced jitter still problem
- Buffer output = fewer writes
- Forking into read and write process and non-blocking write on pipe
- Kernel timestamp per packet
- Pcap – pick up packet in IO-buffer

UNINETT

Jitter in Linux networking with crude

Kernels	Lowlatency kernel laptop				Normal kernel fast machine			
Feature/params	Jitter(us)	Jitter sdv	Jitter max	Loss %	Jitter	Jitter sdv	Jitter max	Loss%
Original 0.62	3.6	120.0	253356.0	0.12	3.4	8.6	4363	0
Orig with -P	3.6	15.0	32906.3	0.06				
Buffered -f	50.0	39.4		2.1	50.0	39.6	4066	0
Buffered -P	50.0	39.7		0.6				
Buffered -P -k	6.3	5.6		2				
Buffered -k	5.9	6.3	1024	0	2.8	8.1	4076	0
Buffered -f -k	7.1	7.4	3462	410	3.1	6.3	1029.0	30
Buffered -f -P	50.0	39.5	4039.0	0				
Buffered pcap -C -P	3.9	5.9	922.0	1.2				
Buffered Pcap -C	3.2	5.8	672.1	0				
Buffered -k ram iou2					0.0	0.0	0.7	0
Buffered -k disk iou2					0.0	0.0	1.8	0

UNINETT

Jitter in linux timestamps

- Memory buffering helps jitter
- Forking and non-blocking write to pipe has high jitter due to overhead in extra process switches
- Kernel udp timestamp or pcap timestamp has lowest jitter
- Kernel timestamp is an extra kernel call
- Increase driver buffer and udp-buffer prevent packet loss
- Low latency kernel helps a bit on jitter variation

UNINETT