Karlsruher Institut für Technologie

P2P – pilot – dynamic circuit

Bruno Hoeft / KIT

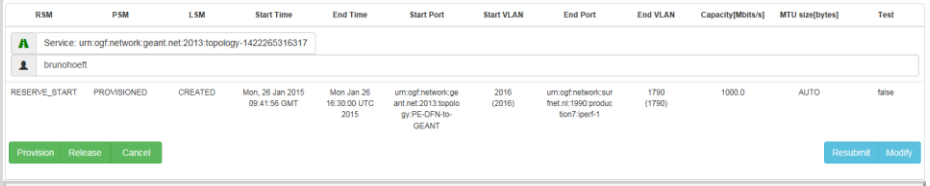
STEINBUCH CENTRE FOR COMPUTING - SCC

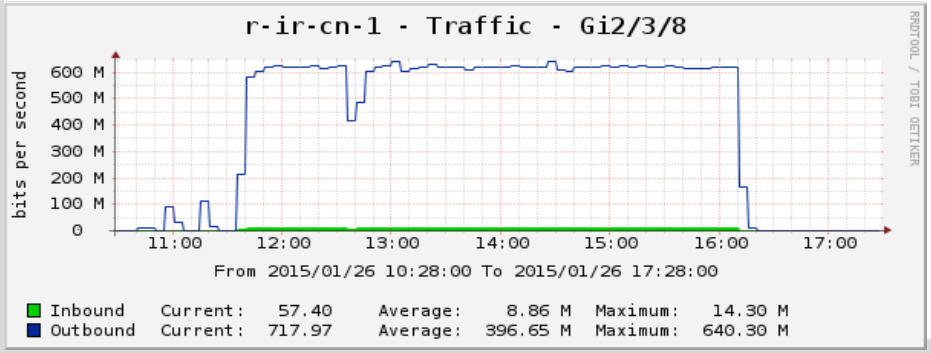
Deployment of dc at KIT

- 1GE Interface => NL-T1 reachable with 10GE bandwidth
  - Upgrade to 10GE interface to DC (still pending)
- Setting up DC
  - several prior Interventions were required
    - DC interface of Surfnet for NL-T1
      - urn:ogf:network:surfnet.nl:1990:production7:46825 setup
      - needed to be entered at other endpoint and a vlan specified (1790)
    - Interface had to be made available in Géant BoD portal under surfnet 7 production7
      - Involved Autobahn developer/operation : (Asd001A_F48S01 Te-1-0-04 iperf1 (eth3))
    - A token had to be created
      - Since KIT has no relation to (SURF[net/sara/…]) NL-T1 and it was not possible to generate via onegini
      - A token was created and has to be entered manualy in the reservation portal
  - One issue apeared at KIT
    - Took the historic DC interface at KIT it was a lightpath between KIT and the DFN/Géant peering in Frankfurt, but not anymore lighted up
    - A MPLS link between the DFN XR-router at Karlsruhe and the DFN CR-Router in Frankfurt on the path to the DFN/Géant peering in Frankfurt is active and after switching to this interface the DC could get brought up and first packets (ping) exchanged
  - No activation of DC between christmas and mid of Jan.
    - Mid of Jan. Géant DC operation had to intervene and sort out pending DC requests before a new reservation was valid and accepted
  - only end of Jan. iperf sessions at both ends could be started and achive 600Mbps over the 1GE reservation

TO-DOs

- Activate dynamic circuit
  - Manually activation only (switchport operator)
  - Dynamic circuit activation integrated in Grid middleware required
  - Some hurdles during manual activation:
    - Token is requested at SARA
      - No ability to generate the token by the remote site
      - → certificate based authorization?
    - Automate representation of actualization and including additional circuits in “Control Pane” → currently still manual intervention necessary
    - Blockade/issues during dc establishing process → see next slide






```
[root@AutoKNF-H1 ~]# ping 10.250.90.24
PING 10.250.90.24 (10.250.90.24) 56(84) bytes of data.
  64 bytes from 10.250.90.24: icmp_seq=1 ttl=64 time=20.5 ms
  64 bytes from 10.250.90.24: icmp_seq=2 ttl=64 time=9.49 ms
  64 bytes from 10.250.90.24: icmp_seq=3 ttl=64 time=9.47 ms
 --- 10.250.90.24 ping statistics ---
3 packets transmitted, 3 received, 0% packet loss, time 2001ms
rtt min/avg/max/mdev = 9.479/13.160/20.506/5.194 ms

[root@AutoKNF-H1 ~]# iperf -c 10.250.90.24
------------------------------------------------------------
Client connecting to 10.250.90.24, TCP port 5001
TCP window size: 16.0 MByte (default)
------------------------------------------------------------
[  3] local 10.250.90.26 port 47564 connected with 10.250.90.24 port 5001
  [ ID] Interval       Transfer     Bandwidth
  [  3]  0.0-10.0 sec    769 MBytes    645 Mbits/sec
```

Fault finding in error situation

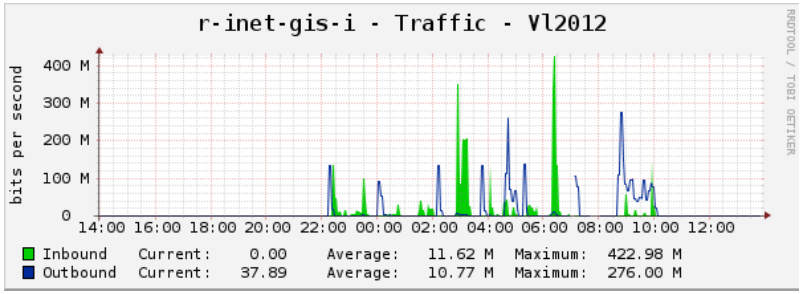
- Endsite :
  - Only able to identify packets leaving the edge router (/interface)
- Call your attached NREN
- Benificial could be:
  - To see if there are known issues along the path
  - Ping and traceroute is over long destination without any shown hop only suboptimal
  - See packet flow (like looking glass) along the dc

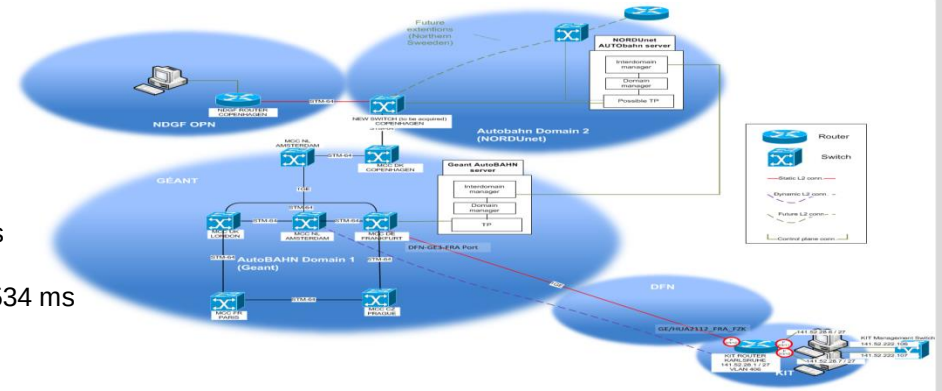
BGP
“lossless” Packet routing change

```
[root@f01-070-101-e ~]# date
Wed Oct  3 21:11:20 CEST 2012
[root@f01-070-101-e ~]# traceroute 109.105.124.141
traceroute to 109.105.124.141 (109.105.124.141), 30 hops max, 40 byte packets
 1  def-gw-v461.gridka.de (192.108.46.1)  0.682 ms  0.762 ms  0.915 ms
 2  tn1481-vrf-gtn41-ospf47.gridka.de (192.108.46.217)  0.360 ms  0.450 ms  0.499 ms
 3  tn1475-rig1-ospf1.gridka.de (192.108.46.181)  0.442 ms  0.489 ms  0.625 ms
 4  l513-e-rbrml-1-be9.cern.ch (192.16.166.33)  11.043 ms  11.074 ms  11.096 ms
 5  dk-ore.nordu.net (192.16.166.50)  33.146 ms  33.149 ms  33.171 ms
 6  bio-vobox.ndgf.org (109.105.124.141)  33.704 ms  33.390 ms  33.313 ms
```

RTT - change 33ms to 23 ms

```
[root@f01-070-101-e ~]# date
Wed Oct  3 21:20:01 CEST 2012
[root@f01-070-101-e ~]# traceroute 109.105.124.141
traceroute to 109.105.124.141 (109.105.124.141), 30 hops max, 40 byte packets
 1  def-gw-v461.gridka.de (192.108.46.1)  0.488 ms  0.655 ms  0.764 ms
 2  tn1481-vrf-gtn41-ospf47.gridka.de (192.108.46.217)  0.398 ms  0.471 ms  0.534 ms
 3  tn1475-rig1-ospf1.gridka.de (192.108.46.181)  0.402 ms  0.486 ms  0.569 ms
 4  dk-ndgf.nordu.net (109.105.124.25)  23.362 ms  23.395 ms  23.407 ms
 5  bio-vobox.ndgf.org (109.105.124.141)  23.804 ms  23.808 ms  23.803 ms
```





activation of dynamic circuit

PREV. NEXT

Page 1 of 5

Last update: Thu Jun 13 12:38:37 UTC 2013

State	Start Time	End Time	Start Port	Start VLAN	End Port	End VLAN	Capacity [Mbits/s]	MTU size [bytes]
Service: GEANT@1371126933989			User: brunohoeft				Justification: AutoKNF - DE-KIT - NDGF - 20130613	
ACTIVE (10)	Thu Jun 13 12:35:33 UTC 2013	Thu Jun 13 13:10:10 UTC 2013	Link to NORDUnet (GEANT.pc.43838714)	2012 (2012)	Link to DFN for BoD (GEANT.pc.bddcfc22)	2012 (2012)	1000.0	0

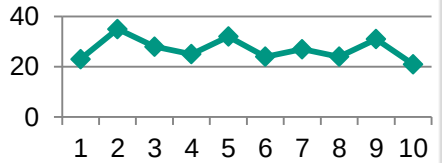
Cancel

AutoKNF move to **Production Env.**

- Circuit deployed between DE-KIT -- NDGF
  - deploy bgp instance @ border router of DE-KIT and NORDUnet
  - **activate dynamic circuit**
    - BGP instances exchanging routing prefix
    - routing table update
  - **decommission dynamic circuit**
    - BGP instances does not reach each other → timeout (? sec.)
    - Routing table update
      neighbor X.X.X.X advertisment-interval XX →default value 30 seconds (eBGP)
    - ssh session survive outage

No outage
0 packet lost

sec.



- LHC project requirements
  - no time constraints
  - every thing within the TCP/IP protocol tolerance

Stable + Robust

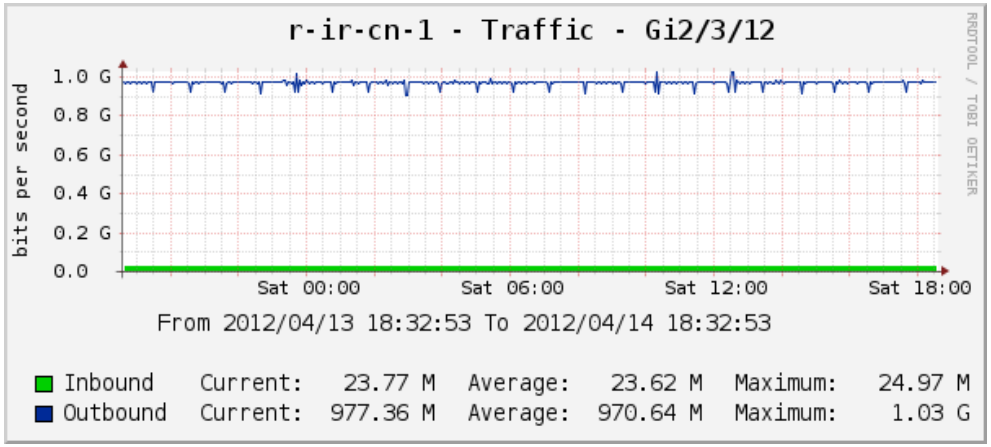
2009 Oct Nov Dec … 2010 Jun Jul Aug Sep Oct Nov Dec Jan Feb Mar Apr May Jun Jul Aug Sep Oct Nov Dec Jan Feb Mar Apr May Jun Jul Aug Sep Oct … 2012

Home Domain	GEANT
User	brunohoeft
Justification	AutoKNF LongTermTest

State	Start time	End Time	Start port	Start mode	Start VLAN	End port	End mode	End VLAN	Capacity [Mbits/s]	Mtu size [bytes]
ACTIVE (10)	Fri Apr 13 11:43:27 EEST 2012	Mon Apr 16 13:59:00 EEST 2012	GEANT Frankfurt port 12/01 (GEANT:pc:2681a0c3)	VLAN	2011	GEANT Connection to Nordunet at Amsterdam port 12/03 (GEANT:pc:a6098c56)	VLAN	2011	1000.0	0

- 86,4 terabit in 24h, $86{,}4 * 10^{12}$

• CRC error	0	✓
• input/output error	0	✓
• packet drop	0	✓
• constant jitter	0	✓



→ extreme low error ratio > 10^{-14}

Comments to LHC P2P architectural draft

- BGP setup for each possible peer
  - Will only be possible for a “small” and static community
  - Several bgp session will be idle (→ P2P peering down)
  - Even equal CIDR to LHC[OPN/ONE] announcement should be possible, the site has to steer the routing (e.g. local prefixes)

- Route server? →
  - Not sure if it will reduce the BGP complexity level
  - Add/remove “automated” bgp peerings