Measuring Delay and Packet Loss at an IXP

Theory and Practice

RIPE 70

Christoph Dietzel

Junior Researcher / PhD Student

Agenda

- Agreed Service Levels
- History
- Challenges
- Implementation
- Questions & Answers

Agreed Service Levels

- Requirements:
  - One way delay: < 500 µs for up to 97.5% of the packets
  - Jitter: < 100 µs for 97.5% of the packets
  - Packet loss: < 0.05% on a daily average (24 hours)
  - All physical links must be covered

- Graphs on the customer portal

SLA History

- RIPE-TTM
  - Discontinued service in 2014
- Accedian MetroNODE / MetroNID
  - Limitations regarding path selection (Y.1731 protocol)
  - Pricy for our use case
- Custom implementation
  - Measure RTT
  - Delay := RTT / 2
  - Jitter := Avg. deviation of the mean latency [1]
  - Packet loss & all links covered: No loss at a representative number of packets over all links

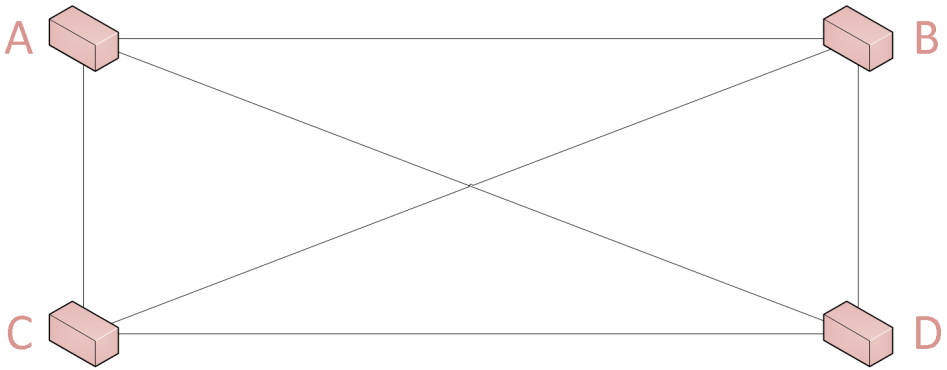
[1] RFC 3393, http://tools.ietf.org/html/rfc3393

Challenges of Latency Measurement

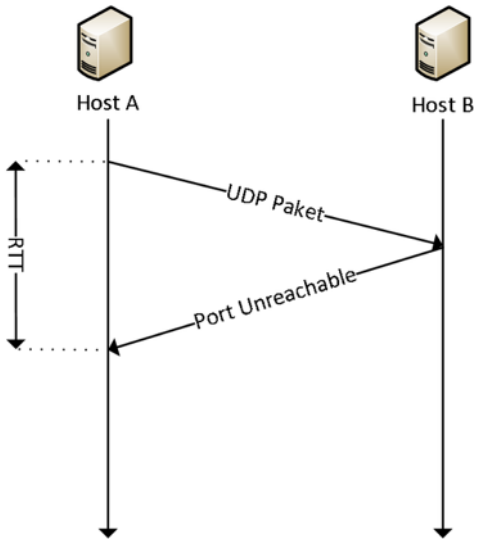
- Multiple paths from A to B on platform (how many?)
- Limited control over LAG (Link Aggregation Group) and LAG member choice
- Be nice, not too much bandwidth consumption
- Platform and OS limitations (protocol stack delay)

Measurement Tools

- 4 edge switches, 4 probing systems
- Using UDP & ICMP (nping)
- Unidirectional $i = n(n-1)$
  - 12 sending instances (AB, AC, AD, BA, BC, BD, CA, CB, CD, DA, DB, DC)
  - How to verify success?

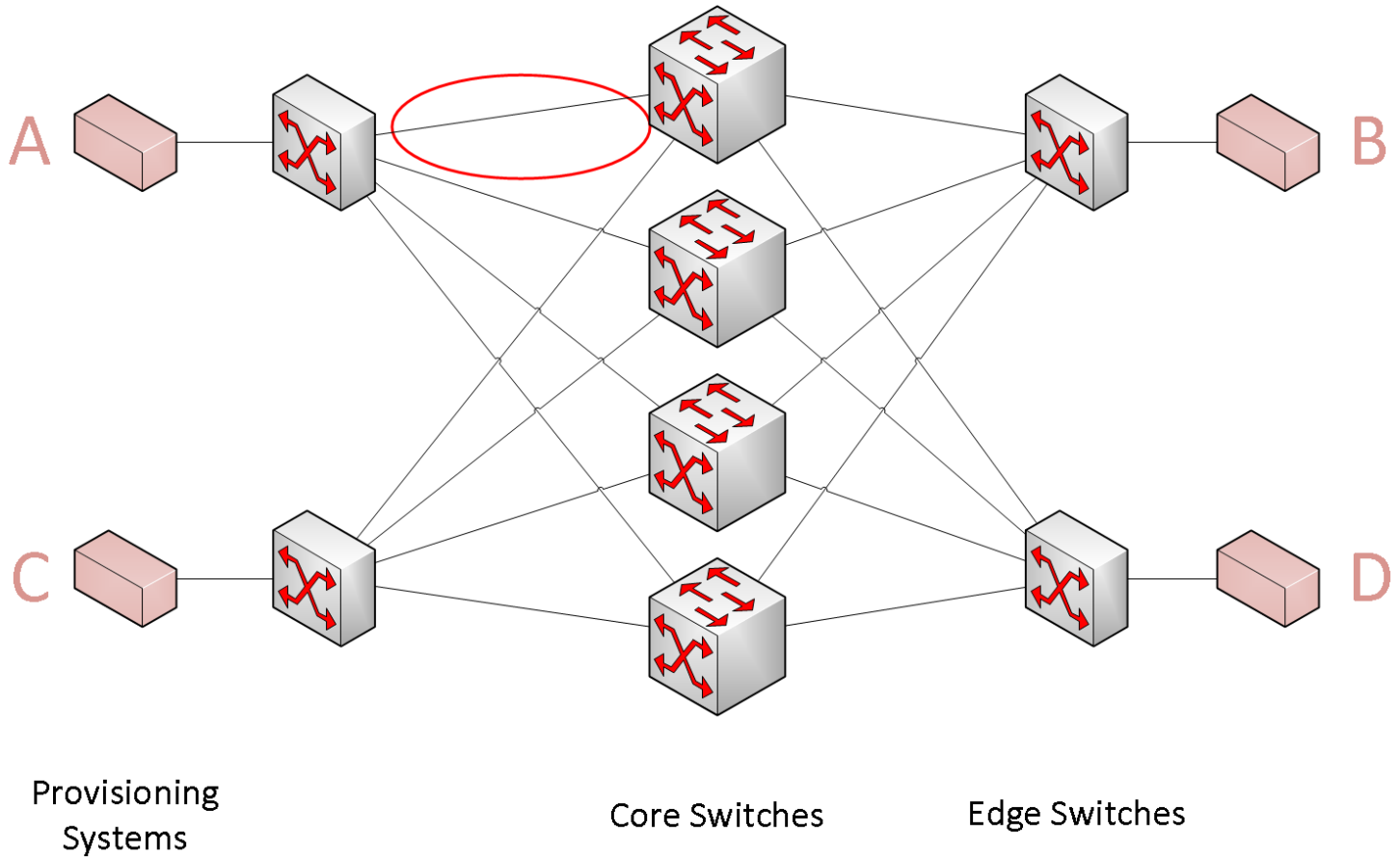


Provisioning Systems



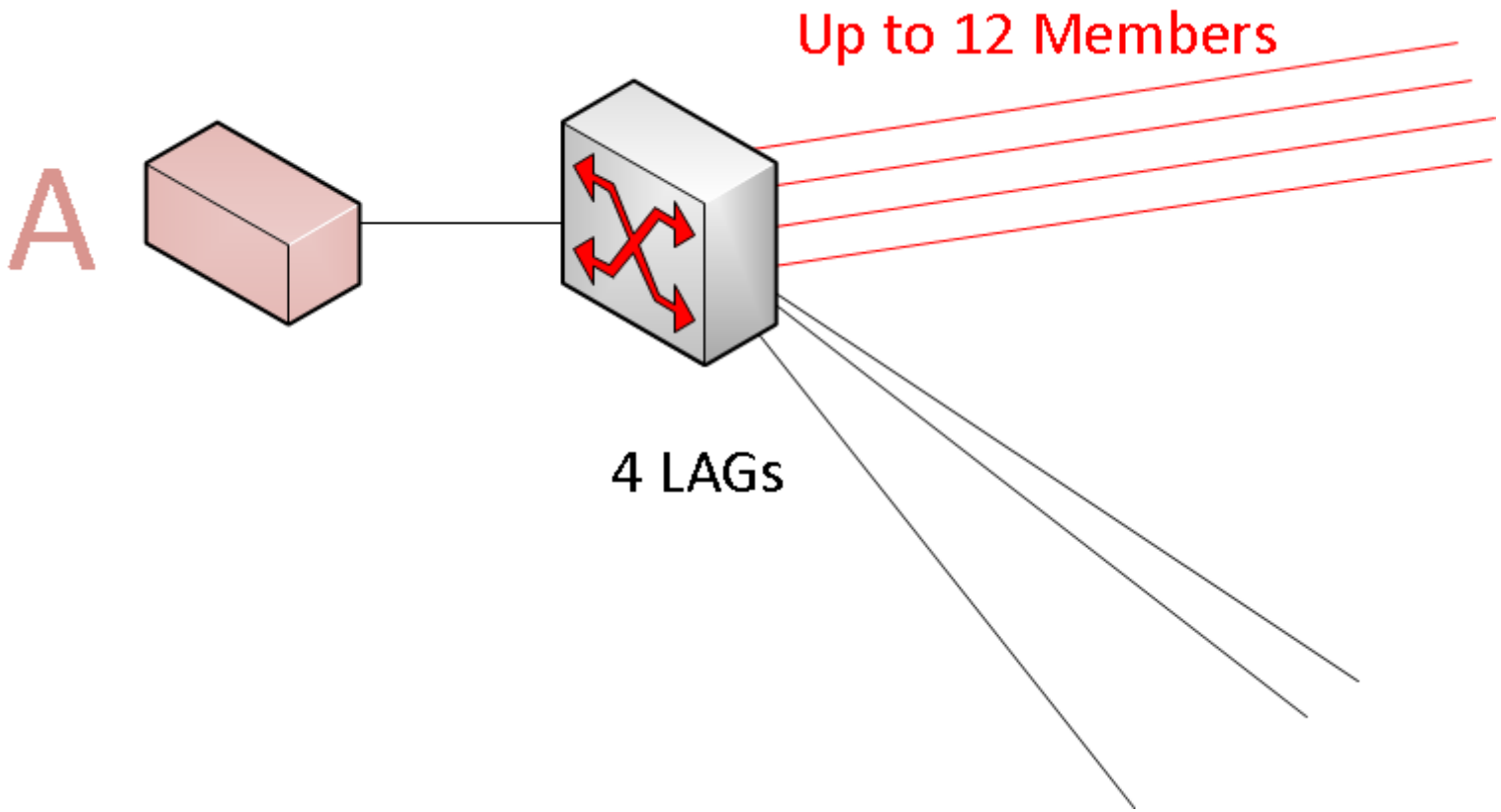
Number of Different Paths

» DE-CIX LAG members



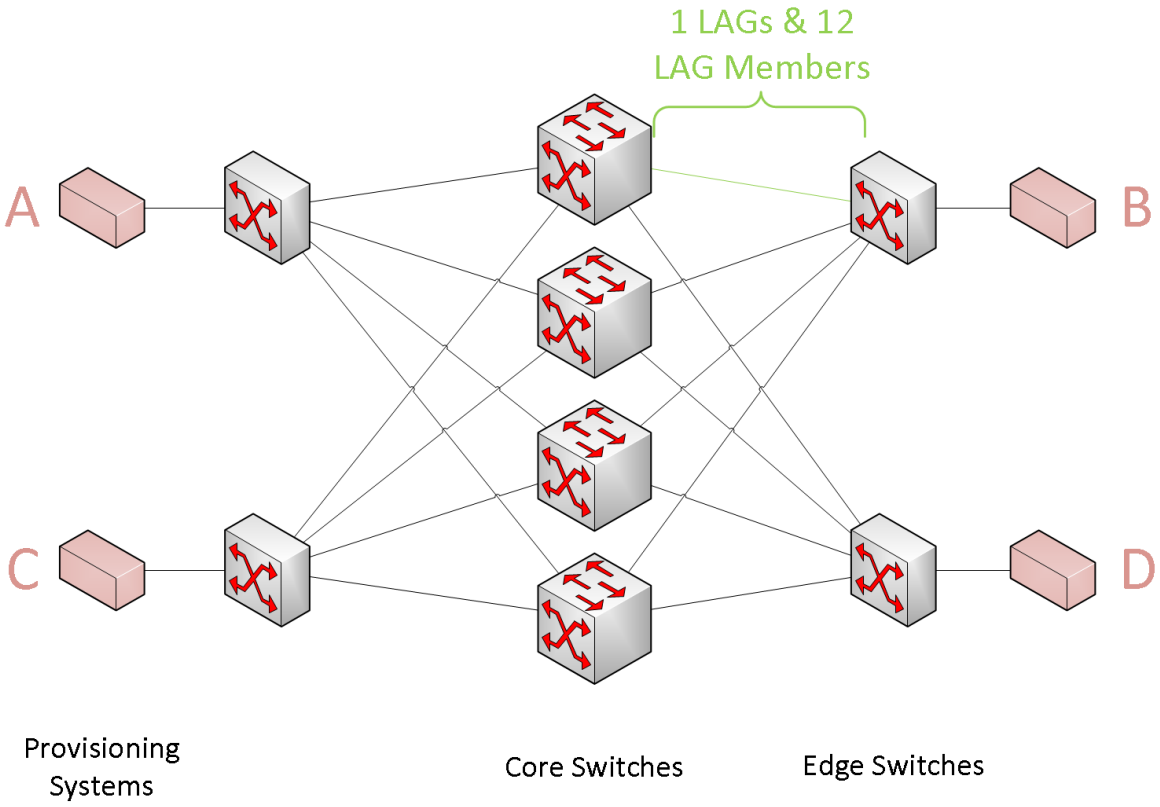
Number of Different Paths

» DE-CIX LAG members



Number of Different Paths

- Probe from A to B (unidirectional)
  - 4 LAGs from edge to core, 12 members per LAG
  - 1 LAG from core to edge, 12 members per LAG



Number of Different Paths

- Probe from A to B (unidirectional)
  - 4 LAGs from edge to core, 12 members per LAG
  - 1 LAG from core to edge, 12 members per LAG

$$(4 \cdot 12) \cdot (1 \cdot 12) = 576$$

  - Probe from A to B, response from B to A not considered

Path Selection: LAG and LAG Member Choice

- LAG choice
  - ECMP (MPLS/VPLS)
  - Assumption: equal chance for each LAG
- LAG member choice
  - Hash space divided by LAG members
  - Hash {src mac, dest mac, src ip, dest ip, src port, dest port} -> deterministic path
    - mac, ip have to be immutable for a path A to B
    - port the only source of entropy ☹
  - Assumption: hash space is equally distributed over all LAG members
    - even though only the port is dynamic

Number of Probes to Test All Paths with 95% Certainty

- Mathematical foundations: coupon collector's problem
  - How many pictures must be bought to have the full set with a chance of 50%
  - How many pings must be send … maps to collectors problem
- Certainty k = 0.95
- Number of paths n = 576 (unidirectional)
- Number of probes x = ?
- Limit theorem [2]: $\mathrm{P}(T < n \log n + cn) \to e^{-e^{-c}}, \quad \text{as } n \to \infty.$

For $n = 576$ paths with a probability of $k = 0.95$ one needs X_n probes.

$X_n = n\ log\ n + nc$ where $e^{-e^{-c}} = k$

For $e^{-e^{-c}} = 0.95,\ c \approx 2.97$
such that
$576\ log(576)\ +\ 576 \cdot 2.97 \approx 5371.84$

[2] Erdős, Paul; Rényi, Alfréd, On a classical problem of probability theory

Implementation

» 3 instances of nping at all 4 provisioning systems

```
nping --privileged -v0 -c 5372 --rate 1000 --udp -p $rand 192.168.1.2
```

» Chose random port for each probe

```
$ports->{40000 + int(rand(25536))}
```

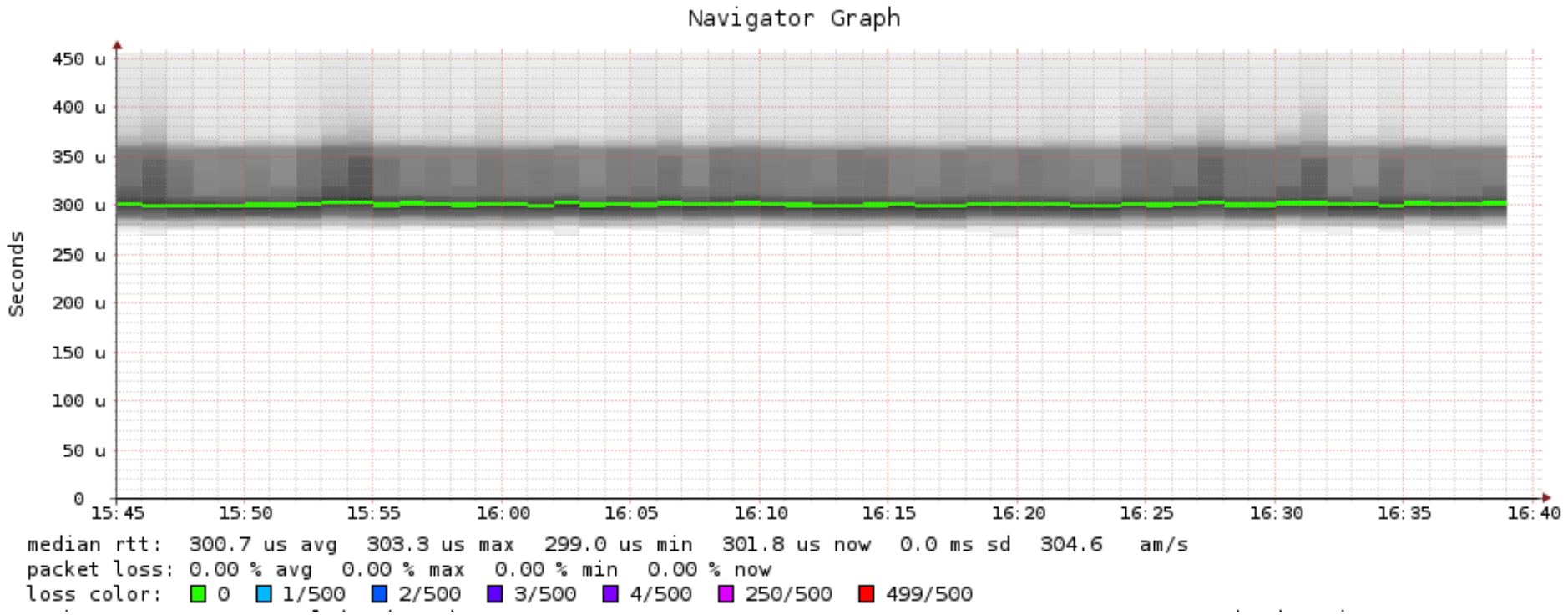
» IP tables rule reduce protocol stack caused delay

```
iptables -I INPUT --proto udp --dport 40000:65536 -j REJECT
```

» Disable ICMP kernel rate limit

```
echo 'net.ipv4.icmp_ratemask=6160' >> /etc/sysctl.conf
```

Customer View



DE·CIX

By joining DE-CIX, you become part of a universe of networks. Everywhere.

DE-CIX. Where networks meet.

DE-CIX Management GmbH
Lindleystr. 12
60314 Frankfurt
Germany
Phone +49 69 1730 902 0

rnd@de-cix.net

www.de-cix.net