perfSONAR Microdep – an add-on for event-based real-time analysis and presentation

4th European perfSONAR Users Workshop
May 14-16th 2024, Trondheim, Norway

Otto J Wittner, Sikt

History

- Microdep ~15 years old
- 15 years ago:
  - perfSONAR's focus: performance tests + scheduled measurements
  - Microdep's focus: small time scale outages + 24/7 measurements
  - perfSONAR impractical for Microdep => "home grown" system
  - ~10 years later: Microdep adopts Elastic search (open distro) as database, scale of production system increases and map-based web-GUI development begins.
- 2024:
  - perfSONAR 5 supports latencybg test (since long) and adopts Opensearch as database.
  - … time is due to merge Microdep into perfSONAR.

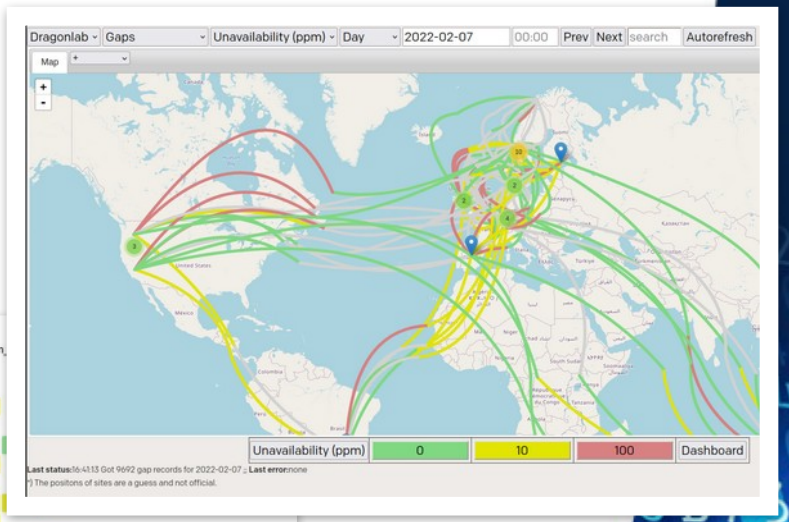
Microdep's objectives

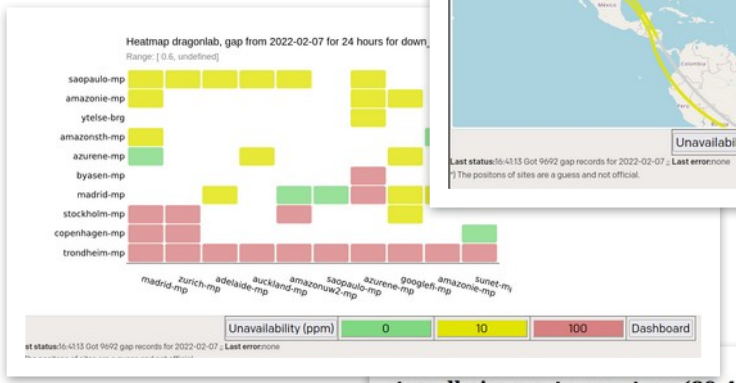
- Monitor Internet w.r.t. telecom standards for depenability
  - five 9s (0.99999 availability)
  - > 50ms considered an outages (increasing relevance for real-time AV)
- Amend NOCs **end-to-end "blindness"**
  - Per-network unit and single-hop monitoring is well covered
  - End-to-end QoS and QoE awareness limited
  - **Early awareness for QoE drop** - "… before unhappy customer calls."
- Monitor routing dependability - detect
  - Sub-optimal configurations
  - Level of outage during planned maintenance (should be zero)
- Queuing and congestion detection – **Interdomain debugging**

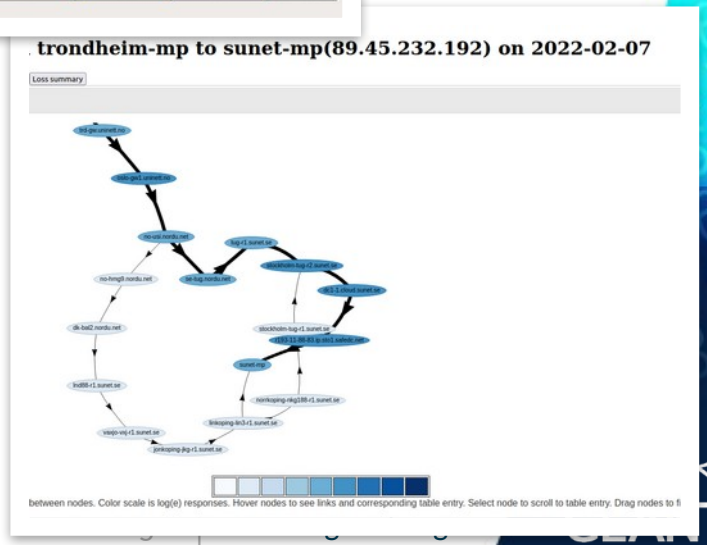
Microdep in production

https://microdep.uninett.no

- End-to-end 24/7 latency measurements
  - UDP probes: 100 pkt/s
  - traceroutes: 10 s intervals
  - ICMP “back scatter” monitoring
- ~240 end-to-end flows in global topology
  - 8 DC nodes (amazon, azure, google)
- **Real-time event analysis**
  - **Gaps (>5 packets lost, > 50ms)**
  - **Queues (diff in latency from min latency)**
  - **Route changes and failures (traceroute)**
  - **Event correlations (gap + routechange)**

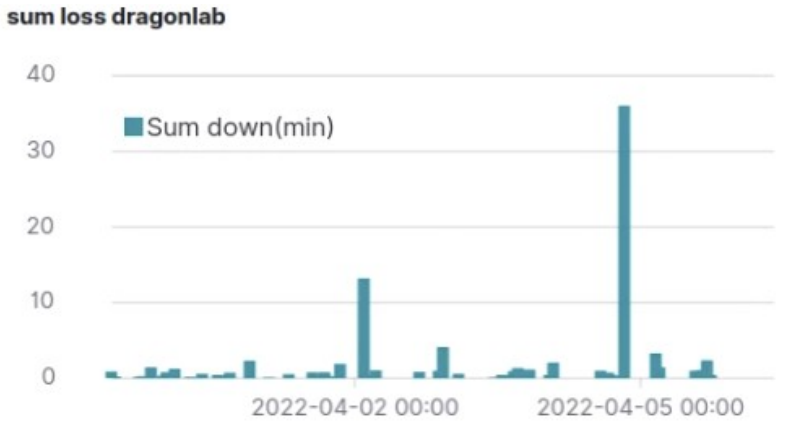


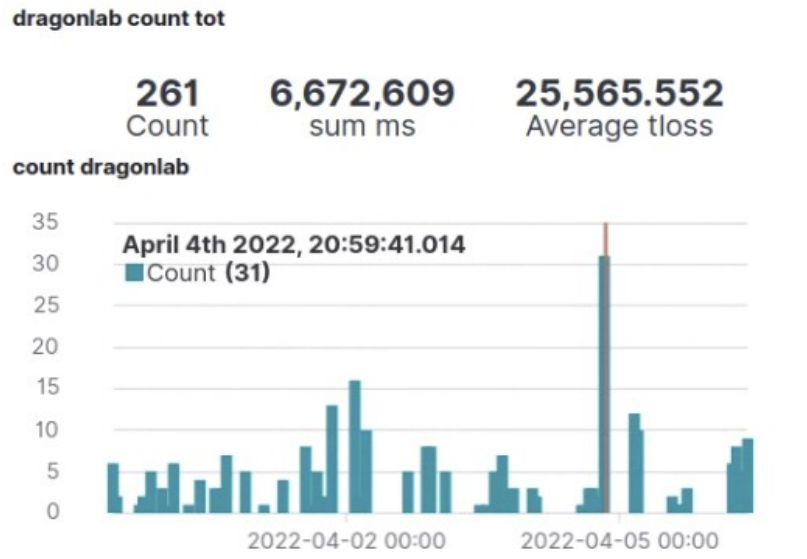




Gaps / packet loss events

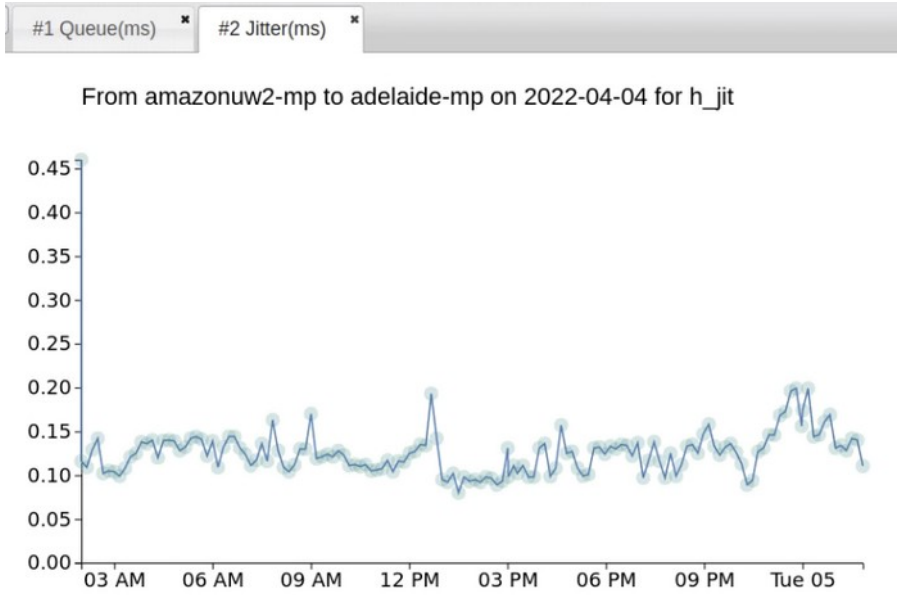
- Windows of 2000 pkts -> min one-way delay
- Gap event = 5 or more pkts lost, i.e. 50 ms downtime
  - 5 successfull pkts ends gap
- Stats on head and tail of gaps (50 pkts)
- Smaller gaps + other stats in daily summary reports

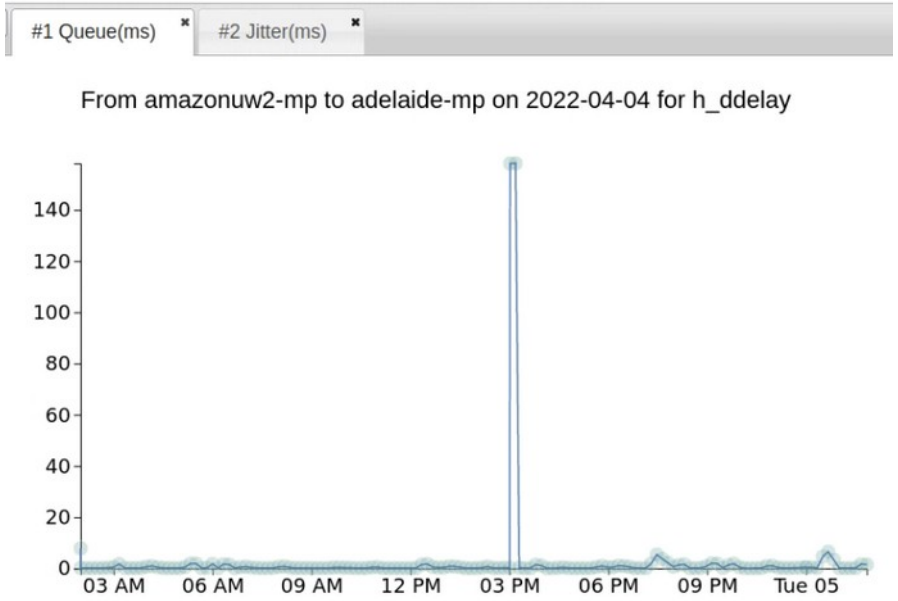




Queues / Jitter events

- Jitter definition from RTCP (rfc3550)
  - ... but show only minor variances
  - Order of few ms
- Queue-buildup events by change in differential one-way delay
  - (delayB – delayA) – mindelay
  - Order of 10-100 ms





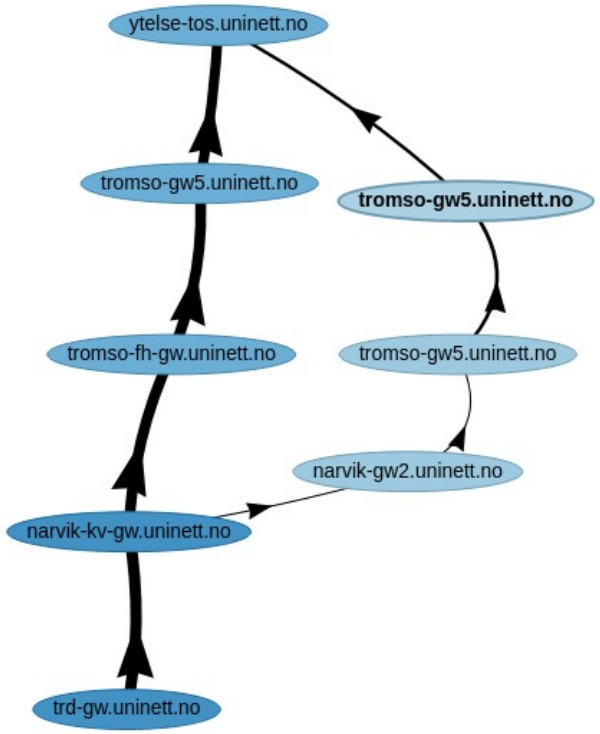
Route failure events

- Route failure = «never ending» traceroute
- Detect periodes with route failures
  - Find «* * * * * *» at max-hops
- Report ICMP errors
  - Network unreachable (N!)
  - ...

```
traceroute to 109.105.116.52 (mp-cph.nordu.net) 30  hops max, 60  byte packets
 1  100.64.102.1 (100.64.102.1) 0.578 ms 0.715 ms 0.815 ms 100.64.102.2 (100.64.10
 2  195.178.64.232 (195.178.64.232) 0.844 ms 100.64.0.1 (100.64.0.1) 1.032 ms 195.
 3  195.113.235.89 (195.113.235.89) 0.777 ms 0.753 ms 0.750 ms 195.178.64.232 (195
 4  195.113.235.89 (195.113.235.89) 4.105 ms 62.40.124.29 (cesnet.mx1.pra.cz.geant
 5  62.40.124.29 (cesnet.mx1.pra.cz.geant.net) 0.550 ms 0.526 ms 0.525 ms 0.572 ms
 6  62.40.98.69 (ae0.mx1.ham.de.geant.net) 15.379 ms 62.40.98.192 (ae8.mx1.fra.de.
 7  62.40.125.206 (nordunet-bckp2-gw.mx1.ham.de.geant.net) 15.350 ms 15.468 ms 62.
 8  62.40.125.206 (nordunet-bckp2-gw.mx1.ham.de.geant.net) 15.409 ms 109.105.97.56
 9  109.105.97.197 (dk-ore-sw-a01.nordu.net) 20.597 ms 109.105.97.207 (dk-ore-sw-a
10  109.105.99.180 (dk-ore-fw.nordu.net) 20.117 ms 20.079 ms 20.237 ms 109.105.97.
11  109.105.116.52 (mp-cph.nordu.net) 20.780 ms 20.973 ms 109.105.99.180 (dk-ore-f
1649029226 starttime 01:40:26
traceroute to 109.105.116.52 (mp-cph.nordu.net) 30  hops max, 60  byte packets
 1  100.64.102.1 (100.64.102.1) 0.424 ms 100.64.102.2 (100.64.102.2) 0.584 ms 100.
 2  100.64.0.1 (100.64.0.1) 0.718 ms 195.178.64.232 (195.178.64.232) 2.856 ms 2.86
 3  195.113.235.89 (195.113.235.89) 3.886 ms 3.861 ms 195.178.64.232 (195.178.64.2
 4  62.40.124.29 (cesnet.mx1.pra.cz.geant.net) 0.403 ms 195.113.235.89 (195.113.23
 5  62.40.124.29 (cesnet.mx1.pra.cz.geant.net) 0.595 ms 0.487 ms 0.681 ms 0.613 ms
 6  62.40.98.69 (ae0.mx1.ham.de.geant.net) 15.240 ms 62.40.98.192 (ae8.mx1.fra.de.
 7  62.40.125.206 (nordunet-bckp2-gw.mx1.ham.de.geant.net) 15.527 ms 15.486 ms 62.
 8  62.40.125.206 (nordunet-bckp2-gw.mx1.ham.de.geant.net) 15.366 ms 109.105.97.56
 9  109.105.97.56 (dk-ore.nordu.net) 20.216 ms 25.275 ms 20.303 ms 109.105.97.197
10  109.105.99.180 (dk-ore-fw.nordu.net) 20.115 ms 109.105.97.207 (dk-ore-sw-a01.n
11  109.105.99.180 (dk-ore-fw.nordu.net) 20.509 ms 20.161 ms 20.542 ms 20.113 ms 2
12  * * * * * *
13  * * * * * *
14  * * * * * *
15  * * * * * *
16  * * * * * *
17  * * * * * *
18  * * * * * *
19  * * * * * *
20  * * * * * *
21  * * * * * *
22  * * * * * *
23  * * * * * *
24  * * * * * *
25  * * * * * *
26  * * * * * *
27  * * * * * *
28  * * * * * *
29  * * * * * *
30  * * * * * *
1649029288 starttime 01:41:28
traceroute to 109.105.116.52 (mp-cph.nordu.net) 30  hops max, 60  byte packets
 1  100.64.102.2 (100.64.102.2) 0.531 ms 100.64.102.1 (100.64.102.1) 0.725 ms 0.86
 2  100.64.0.1 (100.64.0.1) 1.300 ms 1.437 ms 1.576 ms 1.913 ms 2.076 ms 195.178.6
 3  195.113.235.89 (195.113.235.89) 1.297 ms 195.178.64.232 (195.178.64.232) 6.357
 4  62.40.124.29 (cesnet.mx1.pra.cz.geant.net) 0.429 ms 195.113.235.89 (195.113.23
 5  62.40.124.29 (cesnet.mx1.pra.cz.geant.net) 0.520 ms * 0.574 ms 0.641 ms * 0.55
 6  62.40.98.69 (ae0.mx1.ham.de.geant.net) 15.302 ms 62.40.98.192 (ae8.mx1.fra.de.
 7  62.40.98.69 (ae0.mx1.ham.de.geant.net) 15.241 ms 62.40.125.206 (nordunet-bckp2
 8  62.40.125.206 (nordunet-bckp2-gw.mx1.ham.de.geant.net) 15.463 ms 15.439 ms 109
```

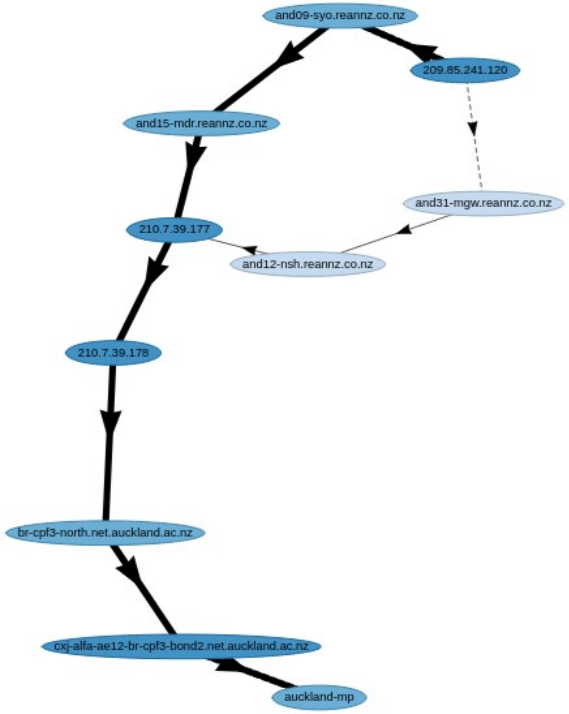
Route change events

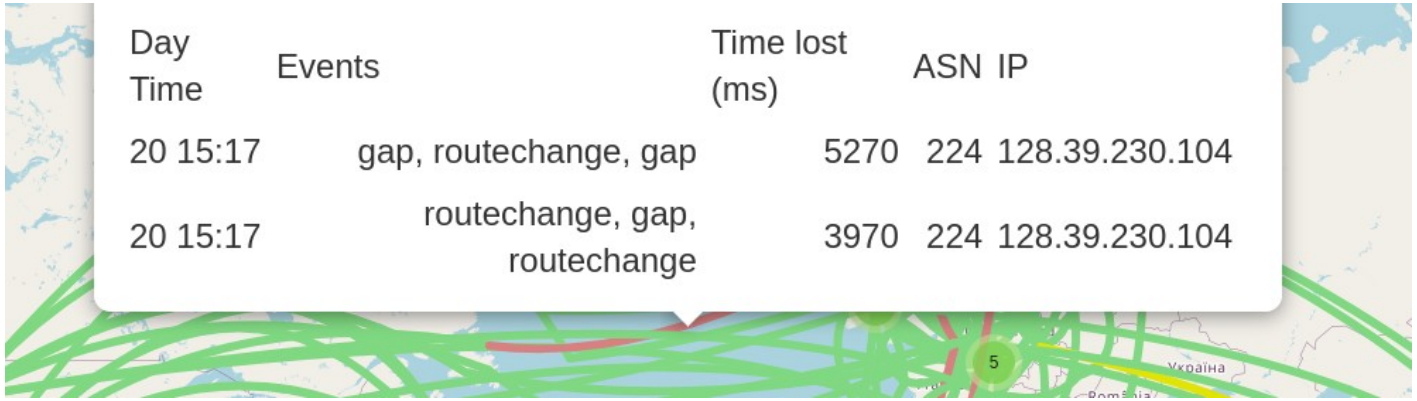
- Route change = **significant** new route
- Detects **change in distribution of seen ip** addresses for each traceroute hop
  - Differential cross entropy
- «Learns» which route changes are normal



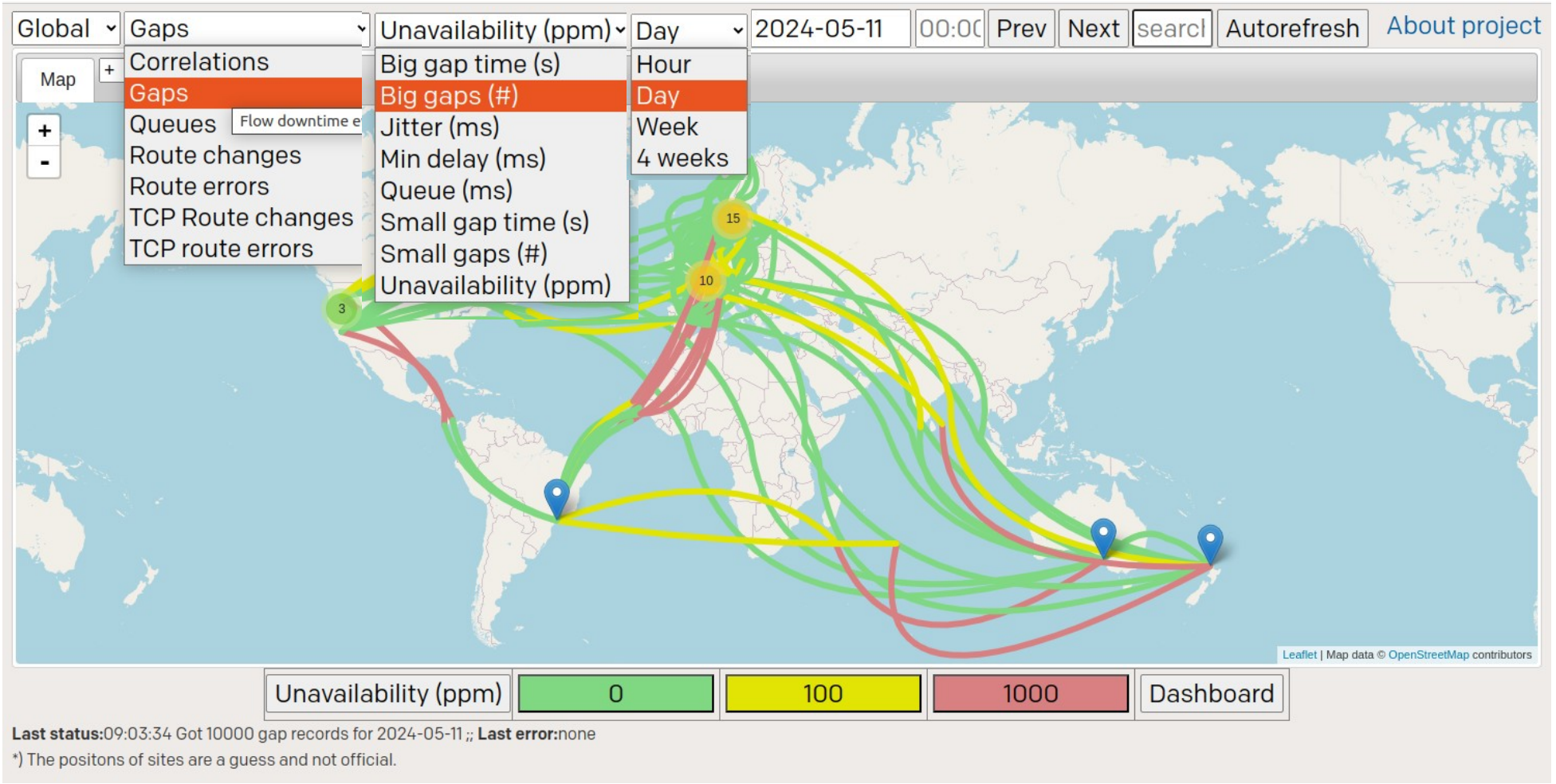
Correlated events

- Gap and routechange in same time window
- Downtime + path anomality
- Identity and ASN of responsible router

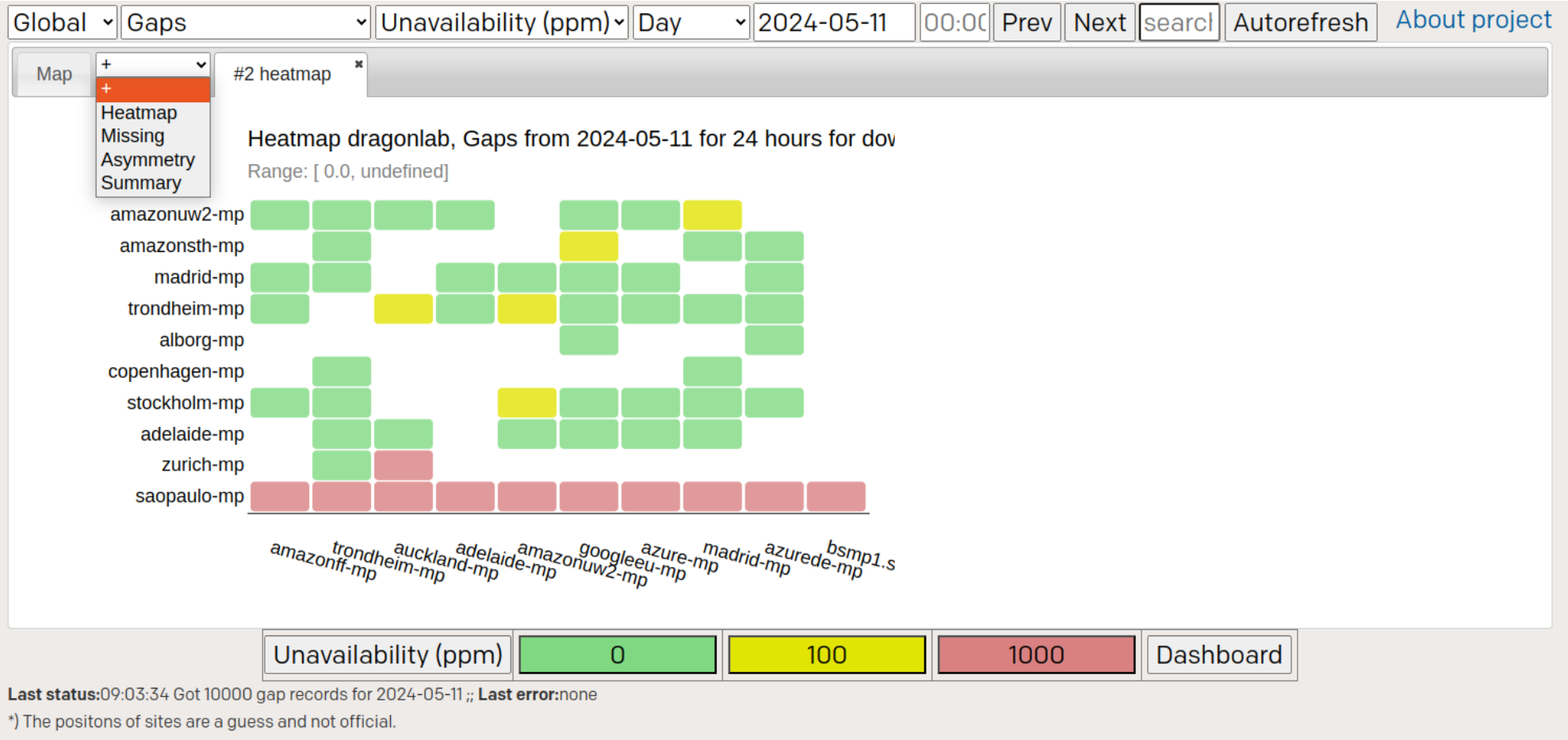




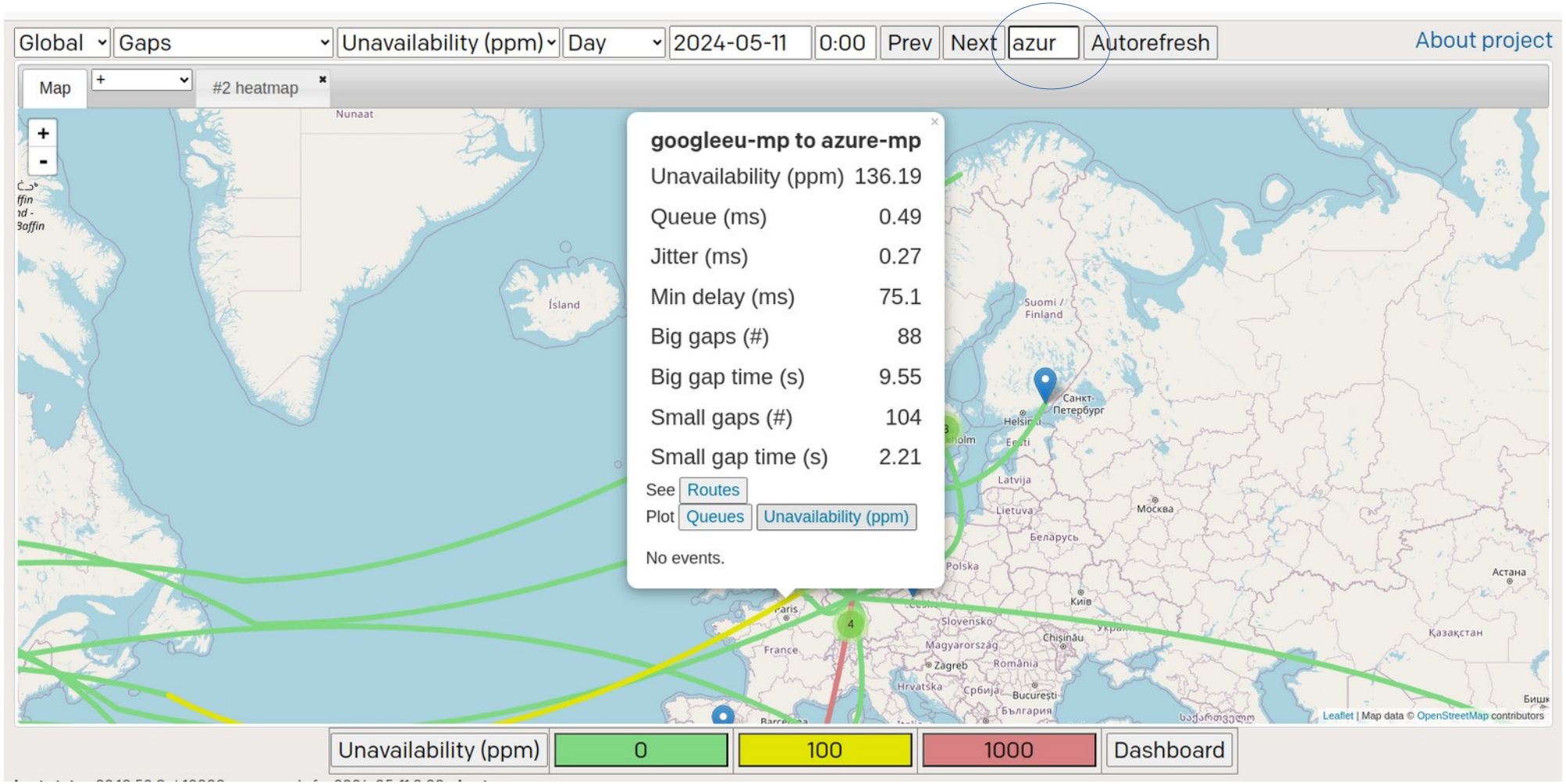
Microdep Web GUI - map



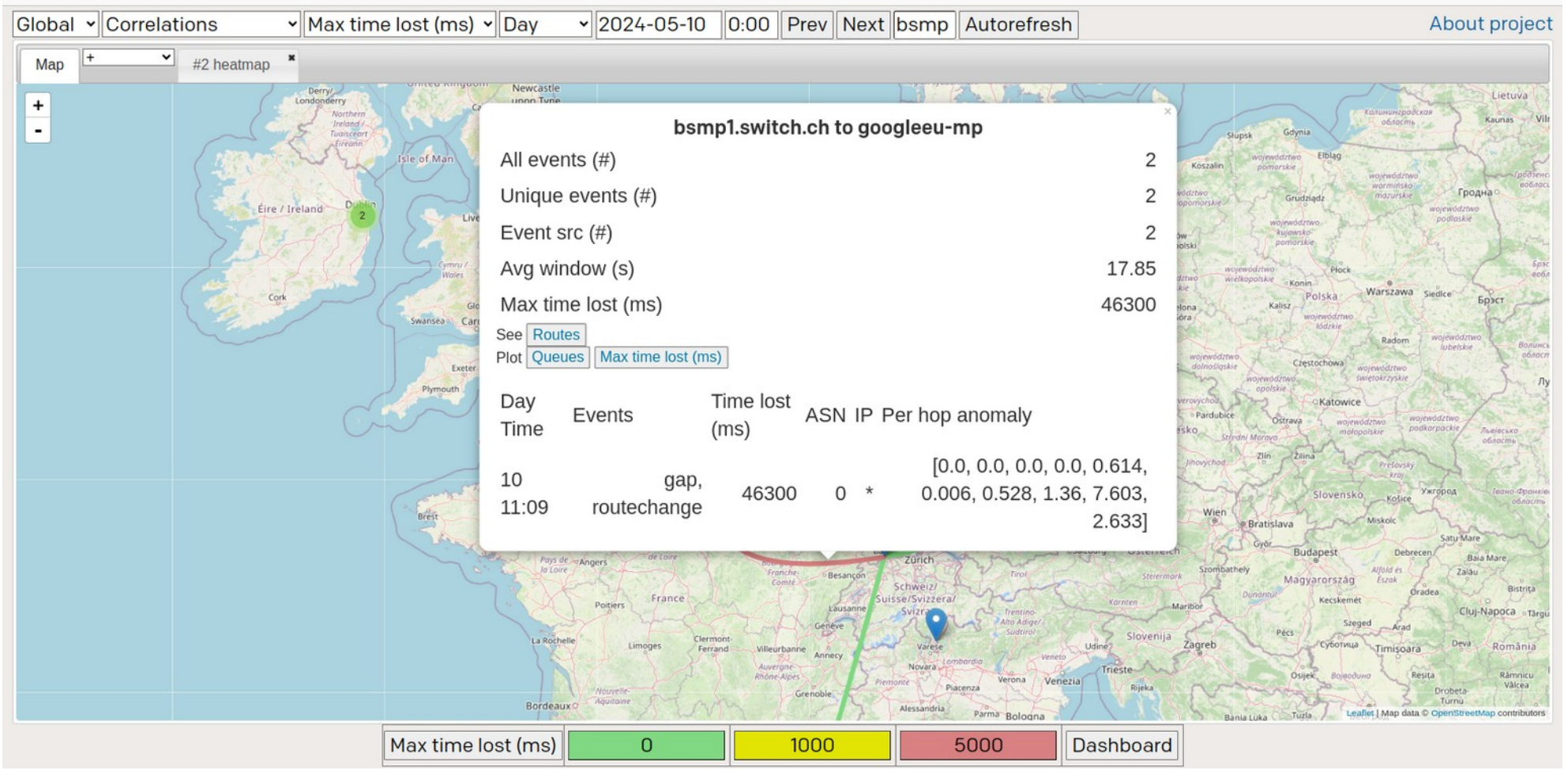
Microdep Web GUI - reports



Microdep Web GUI – popups and filters

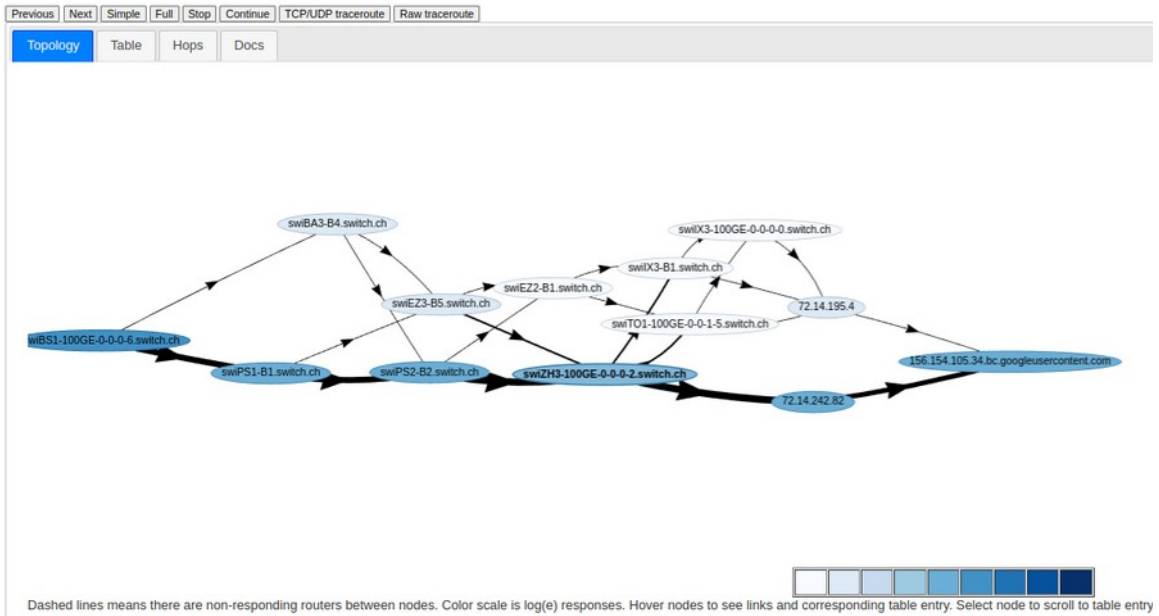


Microdep Web GUI – Correlations



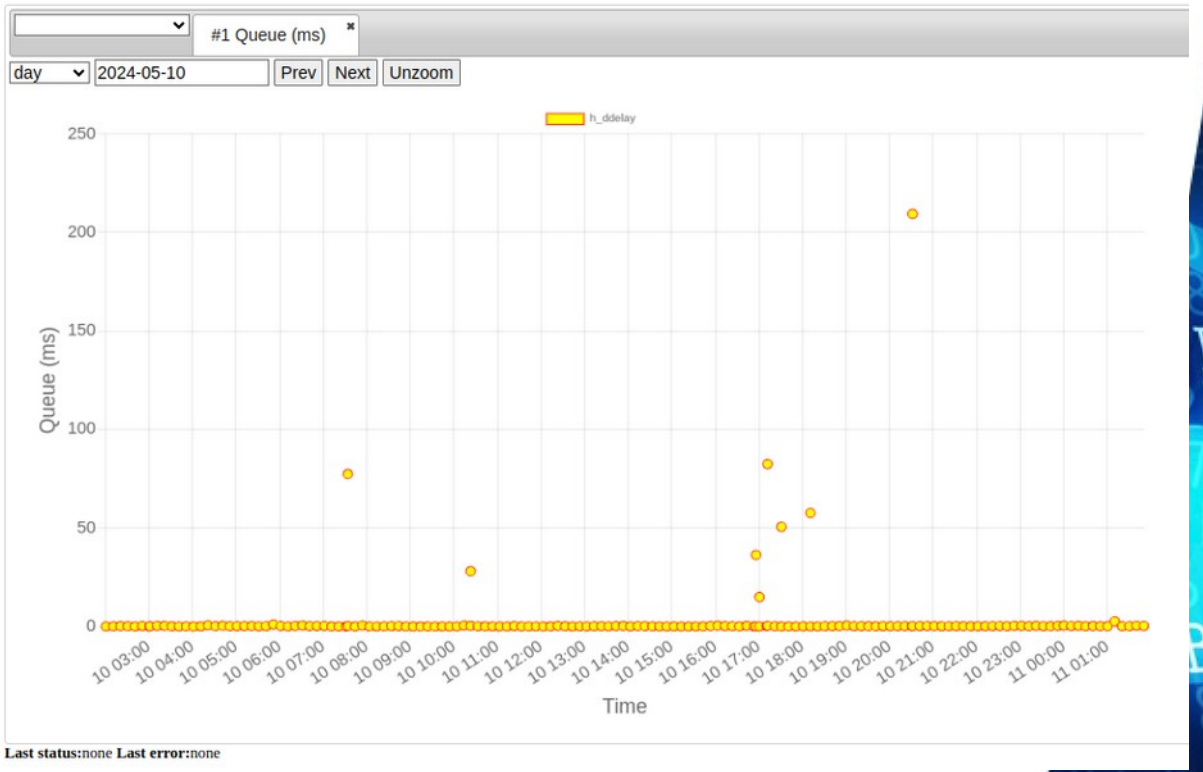
Microdep Web GUI – pstracetree and curve chart

Summary for traceroute charts from bsmp1.switch.ch to googleeu-mp(34.105.154.156) on 2024-05-10



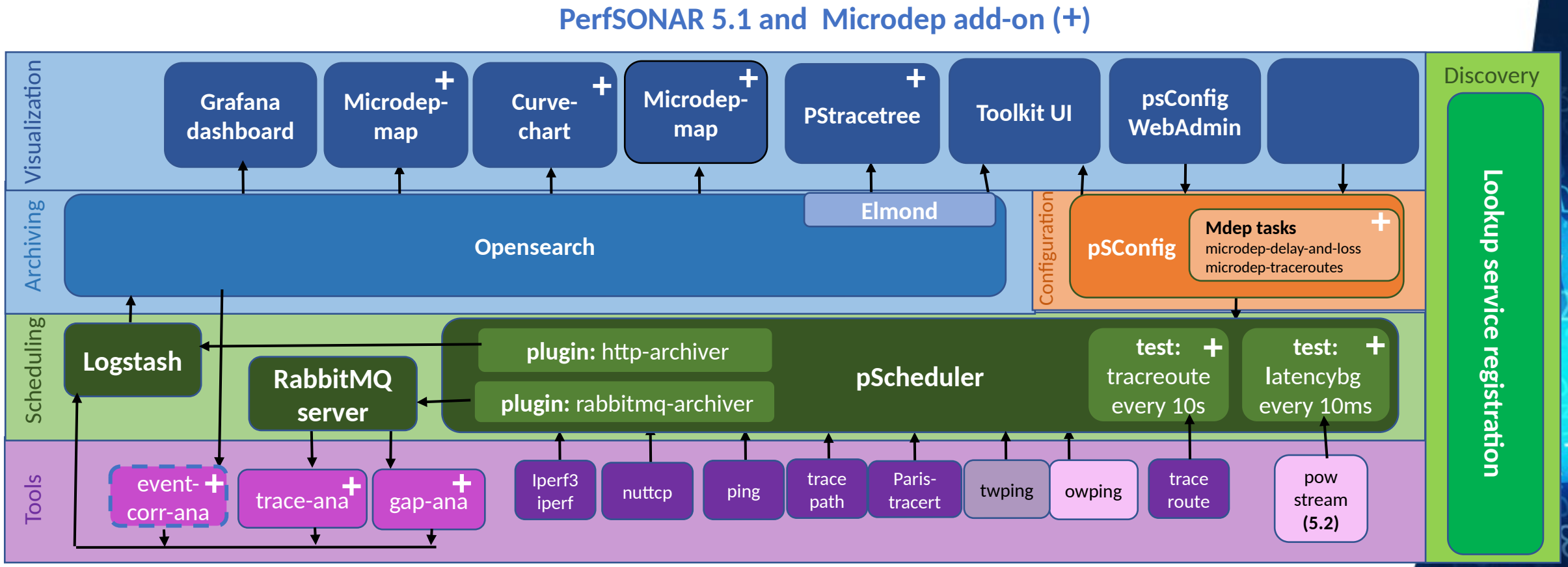
Hop	Router	Avg ms	Min	Max	Sdv	Loss%	Seen	Address	Start	End	Error
1	swiBS1-100GE-0-0-0-6.switch.ch	0.9	0.5	1.4	0.1	0.00%	8712	130.59.35.149	10 00:00:06	10 23:58:46	
2	swiBA3-B4.switch.ch	1.0	0.8	1.1	0.1	0.00%	4	130.59.37.32	10 11:11:22	10 11:11:22	
	swiPS1-B1.switch.ch	2.3	1.5	4.2	0.3	0.00%	8708	130.59.37.189	10 00:00:06	10 23:58:46	
3	swiPS2-B2.switch.ch	2.2	1.4	4.6	0.2	0.00%	8708	130.59.36.179	10 00:00:06	10 23:58:46	
	swiEZ3-B5.switch.ch	1.7	1.6	1.7	0.0	0.00%	4	130.59.37.6	10 11:11:22	10 11:11:22	
4	swiEZ2-B1.switch.ch	5.6	5.6	5.7	0.0	0.00%	3	130.59.36.125	10 11:11:22	10 11:11:22	
	swiZH3-100GE-0-0-0-2.switch.ch	2.0	1.8	5.5	0.1	0.00%	8709	130.59.36.170	10 00:00:06	10 23:58:46	

From azurede-mp to bsmp1.switch.ch



Last status:none **Last error:**none

Microdep in perfSONAR



Microdep installation and configuration

- Meta package “**perfsonar-microdep**” in perfsonar 5.2 repo with
  - *microdep-map* (GUI)
  - *microdep-ana* (event analysis)
  - *pstracetree* (traceroute viewer)
- Installation
  - Add perfsonar repo (see https://docs.perfsonar.net/index.html#installation)
  - rpm/deb: `dnf/apt install perfsonar-toolkit perfsonar-microdep`
- Configuration
  - Adapt **`/etc/perfsonar/psconfig/pscheduler.d/microdep-tests.json`**
  - ...

microdep-tests.json : Topology

```
{
   "addresses" : {
      "toolkit_1" : { "address" : "172.150.1.2"},
      "testpoint_1" : { "address" : "172.150.2.2", "no-agent" : true },
      ...
   }
   "groups" : {
      "Star-topology" : {
         "a-addresses" : [{ "name" : "toolkit_1"}],
         "b-addresses" : [{ "name" : "testpoint_1"}, ... ],
         "type" : "disjoint"}
      }
    },
   "schedules" : {
      "Every-60s" : {
         "repeat" : "PT60S",
         "slip" : "PT60S",
         "sliprand" : true
      }
   },
   ...
```

microdep-tests.json : Tests

```
 ...
"tests" : {
    "Delay-and-loss-ipv4-100pps" : {
       "spec" : {
          "dest" : "{% address[1] %}",
          "dest-node" : "{% pscheduler_address[1] %}",
          "flip" : "{% flip %}",
          "ip-version" : 4,
          "source" : "{% address[0] %}",
          "source-node" : "{% pscheduler_address[0] %}",
          "output-raw" : true,
          "packet-count" : 360000,
          "packet-subcount" : 200,
          "packet-interval": 0.01
       },
       "type" : "latencybg"
    },
    "Traceroute-ipv4" : {
       "spec" : {
          "dest" : "{% address[1] %}",
          "ip-version" : 4,
          "source" : "{% address[0] %}",
          "source-node" : "{% pscheduler_address[0] %}",
          "probe-type" : "tcp"
       },
       "type" : "trace"
    }
 },
 ...
```

powstream patch

microdep-tests.json : Archives

```
...
"archives": {
    "logstash": { ... },
    "gap-ana-rmq": {
        "archiver": "rabbitmq",
        "data": {
            "schema": 2,
            "_url": "amqp://guest:guest@localhost//",
            "exchange": "gap-ana",
            "connection-expires": "PT60S",
            "retry-policy": [ { "attempts": 5,  "wait": "PT5S" } ]
        },
        "ttl": "PT1H"
    },
    "trace-ana-rmq": {
        "archiver": "rabbitmq",
        "data": {
            "schema": 2,
            "_url": "amqp://guest:guest@localhost//",
            "exchange": "trace-ana",
            "connection-expires": "PT120S",
            "retry-policy": [ { "attempts": 5,  "wait": "PT5S" } ]
        },
        "ttl": "PT1H"
    }
},

...
```

microdep-tests.json : Tasks

Applied to select topology (5.1 beta)

```
    ...
    "tasks" : {
        "microdep-delay-and-loss" : {
            "_meta" : {
                "display-name" : "Microdep delay and loss measurements"
            },
            "group" : "Star-topology",
            "test" : "Delay-and-loss-ipv4-100pps",
            "archives": ["logstash","gap-ana-rmq"]
        },
        "microdep-traceroutes" : {
            "_meta" : {
                "display-name" : "Microdep traceroutes every minute"
            },
            "group" : "Star-topology",
            "schedule" : "Every-60s",
            "test" : "Traceroute-ipv4",
            "archives": ["logstash","trace-ana-rmq"]
        }
    }
}
```

To do...

- Before release
  - powstream patch
    - Enable raw packet subsession reporting (not only aggregations)
  - Support for config of multiple measurement networks
  - Debian packages
- Future release
  - Support IPv6
  - Move and merge web GUI info Grafana framework
  - Enable event analysis on testnodes
  - Improve correlation analysis
  - Improve traceroute based pinpointing of source-of-error
  - ...

Let’s explore a production version…

- Visit https://microdep.sikt.no

GÉANT
Networks • Services • People

Thank you

www.geant.org

The research leading to these results has received funding from the European Union's Horizon 2020 research and innovation programme under Grant Agreement No. 731122 (GN4-2).