IETF Hackathon

Application Layer Traffic Optimization (ALTO) WG

Optimizing Large Scale Global Data Transfers for Science Networks

Presenter: Mahdi Soleimani on behalf of the ALTO WG

IETF 114
23-24 July 2022
Philadelphia, PA

Thank you to all participants

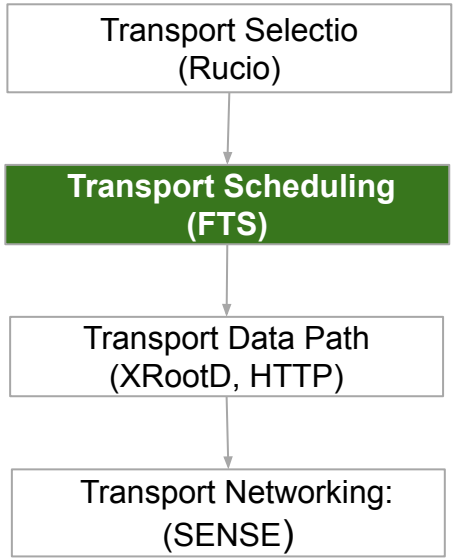
Jensen Zhang, Mahdi Soleimani, Jian Luo, Jie Chen, Ryan Yang, Jacob Dunefski, Kai Gao, Jordi Ros-Giralt, Y. Richard Yang, John Graham, Mario Lassnig, Harvey Newman, Jacob Dunefsky.

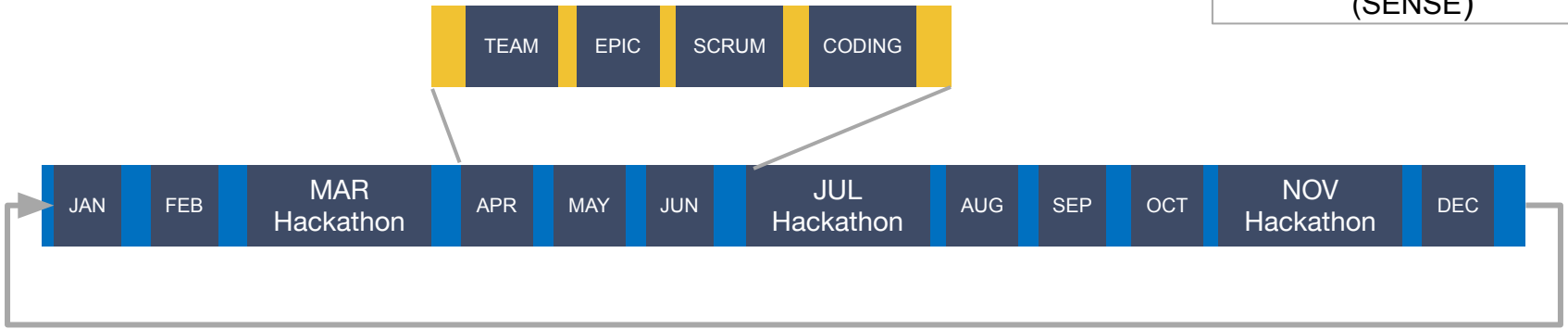
And all the members from the ALTO WG, Yale, Tongji and Sichuan Universities, FTS and Rucio teams, the Pacific Research Platform in California, and companies involved in the WG activities.

Working endless hours managing 3 time zones! (US, EU, China) for the Hackathon.

OpenALTO: Continuous Integration with Hackathon Checkpoints

- In IETF 113 Hackathon, ALTO WG integrated the *network map and cost map (RFC 7285)* into *Rucio* (the data management system of CERN) to optimize *selection of data sources*
- In IETF 114 Hackathon, ALTO WG integrates the *path vector (RFC-to-be)* into *FTS (the data transfer service of CERN)* to achieve *rate (network resource) control and optimization.*



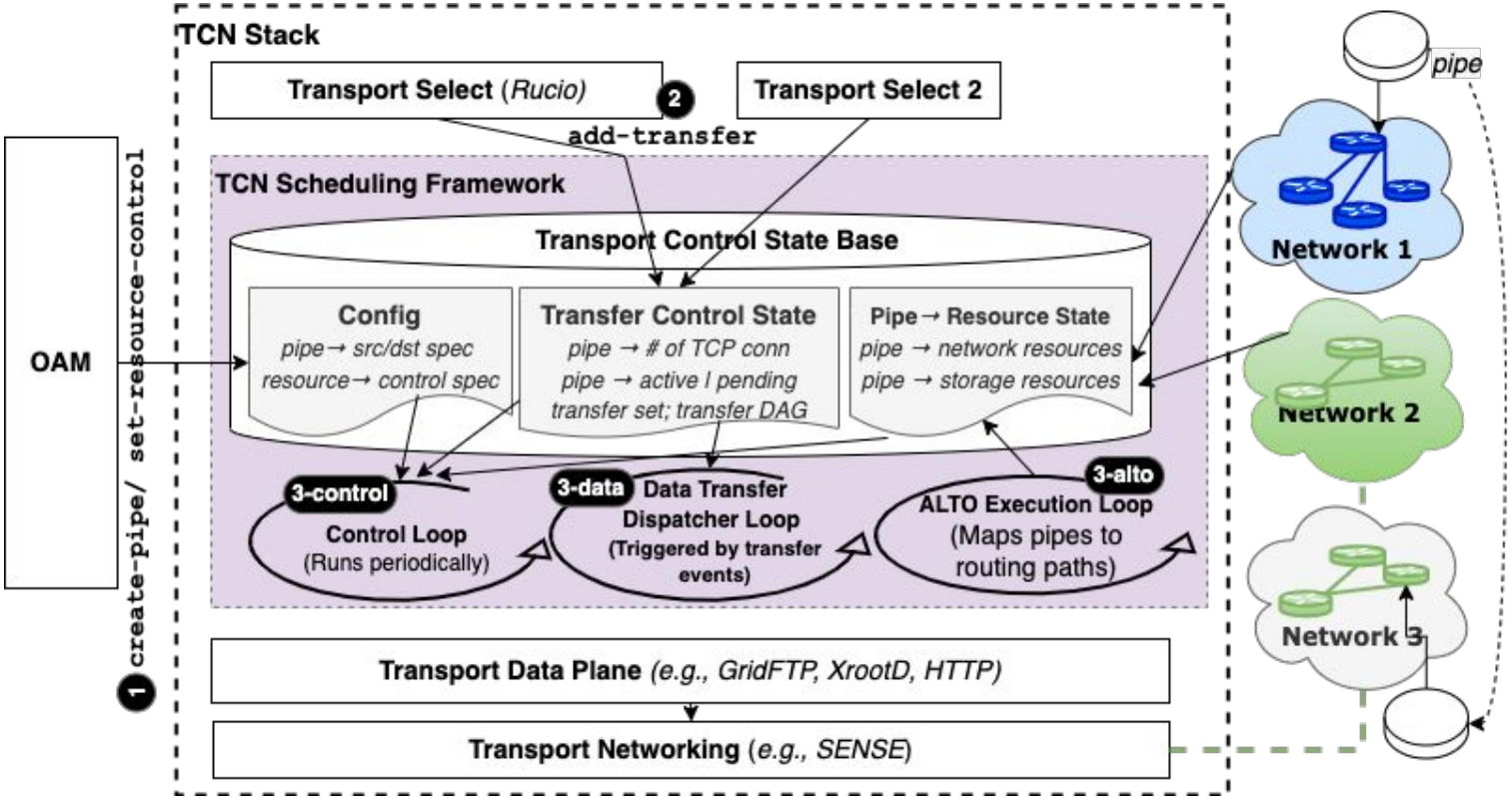


FTS Control Architecture

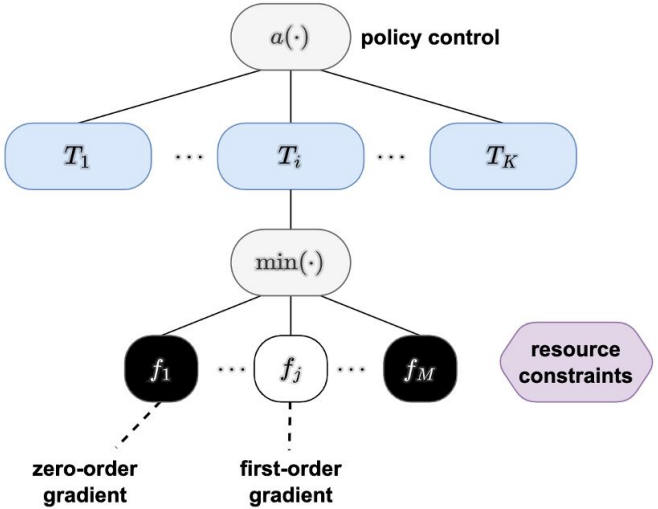
- Software architecture
  - Transport control using widely available, end-to-end control knob (i.e., #tcp connections)
  - Transport control using light-weight control knob (i.e., #tcp connections, not packets)
  - FTS providing a single point in control flow graph

- Performance properties
  - Efficiency (optimal usage of resources)
  - Resource control

TCN (FTS) Components



Overall Scheduling Framework



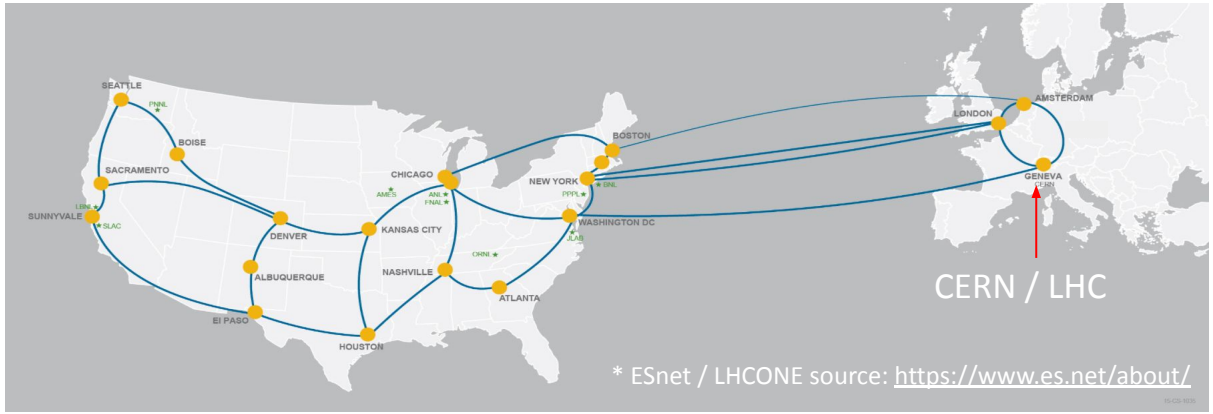
otocol 1 FTS Model Analyzed (Called for High Success Rate)

```
: Define RL(x)=round(log_B(x))
: procedure OPTIMIZEGOODSUCCESSRATE(state)
:     if cur.ema < prev.ema then
:         if RL(cur.filesizeAvg)) < RL(prev.filesizeAvg) then
:             decision = prevValue + increaseStepSize
:         else if RL(cur.ema)< RL(prev.ema) then
:             decision = prevValue - decreaseStepSize
:         else
:             decision = prevValue
:         end if
:     else if cur.ema > prev.ema then
:         decision = prevValue + increaseStepSize
:     else
:         decision = prevValue + increaseStepSize
:     end if
: end procedure
```

Implemented control algorithm

Goals in this Hackathon

- Use ALTO provided information to optimize transfers for FTS, the main data transfer scheduler tool for LHC and other large scientific projects.
- To integrate the Zero-order algorithm and the ALTO client with the fts controller. (Prototype)

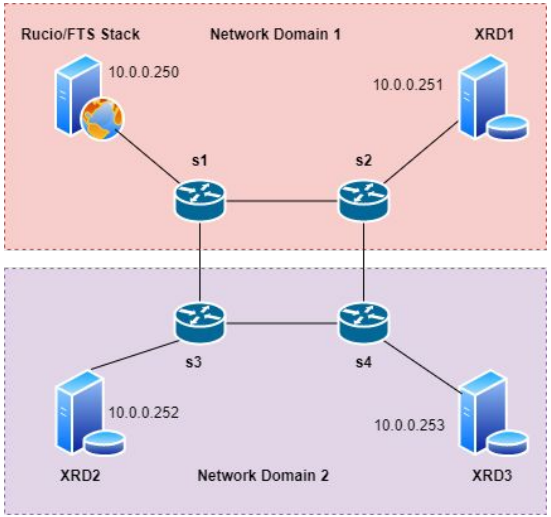


RFCs Involved in this Hackathon

- **RFC-to-be: An ALTO Extension:** Path Vector
  - https://datatracker.ietf.org/doc/draft-ietf-alto-path-vector/
- RFC 9240: An Extension for Application-Layer Traffic Optimization (ALTO): Entity Property Maps
  - https://datatracker.ietf.org/doc/rfc9240/
- I-Draft ALTO Extension: Flow-based Cost Query
  - https://datatracker.ietf.org/doc/draft-gao-alto-fcs/

Demo 1: ALTO Path Vector

Reference:
https://datatracker.ietf.org/doc/draft-ietf-alto-path-vector/




```
fno in ietf-hackathon at fno-anywhere on ᚹ fts [?]
→ docker compose -p static ps
NAME                COMMAND                  SERVICE             STATUS
rucio               "httpd -D FOREGROUND"    rucio               running
1:8443->443/tcp, 127.0.0.1:3306->3306/tcp, 127.0.0.1:5432->5432/tcp, 127.
/tcp, 127.0.0.1:61613->61613/tcp
static-activemq-1   "/app/run.sh"           activemq            running
static-fts-1        "/docker-entrypoint.…"   fts                 running
static-ftsdb-1      "docker-entrypoint.s…"   ftsdb               running
static-mininet-1    "/bin/bash -c 'servi…"   mininet             running
static-ruciodb-1    "docker-entrypoint.s…"   ruciodb             running
```

```
fno in ietf-hackathon at fno-anywhere on ᚹ fts [?]
→ docker compose -p dynamic ps
NAME                COMMAND                  SERVICE             STATUS
dynamic-odl1-1      "/bin/bash -c '/opt/…"   odl1                running
dynamic-odl2-1      "/bin/bash -c '/opt/…"   odl2                running
xrd1                "/docker-entrypoint.…"   xrd1                running
xrd2                "/docker-entrypoint.…"   xrd2                running
xrd3                "/docker-entrypoint.…"   xrd3                running
```

ALTO Endpoint Cost Map using Path Vector

```
Sending Request: {
  "ane-property-names" :
  [
    "bandwidth"
  ],
  "cost-type" :
  {
    "cost-metric" : "ane-path",
    "cost-mode" : "array"
  },
  "endpoint-flows" :
  [
    {
      "dsts" :
      [
        "ipv4:10.0.0.252",
        "ipv4:10.0.0.253"
      ],
      "srcs" :
      [
        "ipv4:10.0.0.251"
      ]
    },
    {
      "dsts" :
      [
        "ipv4:10.0.0.253"
      ],
      "srcs" :
      [
        "ipv4:10.0.0.252"
      ]
    }
  ]
}
```

ALTO Path Vector Response

```
[root@84cbb0cfa7f2 build]# src/alto/alto_client http://192.168.208.1:9090
/pathvector/pv 'xrd1->xrd2,xrd3;xrd2->xrd3' bandwidth | tail -n+36
}
--62f59eb2ec394a36bec324f743f64db2
Content-Type: application/alto-endpointcost+json
Content-ID: <ecs@192.168.208.1:9090>

{"meta": {"vtag": {"resource-id": "pv.ecs", "tag": "fa7be4f0bc714a0587b63
c426861f456"}, "cost-type": {"cost-metric": "ane-path", "cost-mode": "arr
ay"}}, "endponit-cost-map": {"10.0.0.251": {"10.0.0.253": ["L1", "L2", "L
3"], "10.0.0.252": ["L1", "L4", "L5", "L6"]}, "10.0.0.252": {"10.0.0.253"
: ["L6", "L5", "L4", "L2", "L3"]}}}
--62f59eb2ec394a36bec324f743f64db2
Content-Type: application/alto-propmap+json
Content-ID: <propmap@192.168.208.1:9090>

{"meta": {"dependent-vtags": [{"resource-id": "pv.ecs", "tag": "fa7be4f0b
c714a0587b63c426861f456"}]}, "property-map": {".ane:L1": {"bandwidth": 10
}, ".ane:L2": {"bandwidth": null}, ".ane:L3": {"bandwidth": 10}, ".ane:L4
": {"bandwidth": null}, ".ane:L5": {"bandwidth": null}, ".ane:L6": {"band
width": 10}}}
--62f59eb2ec394a36bec324f743f64db2--
```

ALTO Property Map for Abstract Network Element (ANE)

Reference:
https://datatracker.ietf.org/doc/rfc9240/

ALTO Path Vector Request with Flow Extension

Reference: https://datatracker.ietf.org/doc/draft-gao-alto-fcs/

Demo 2: ALTO-Enabled Transport Scheduler for FTS

Modified FTS Database:

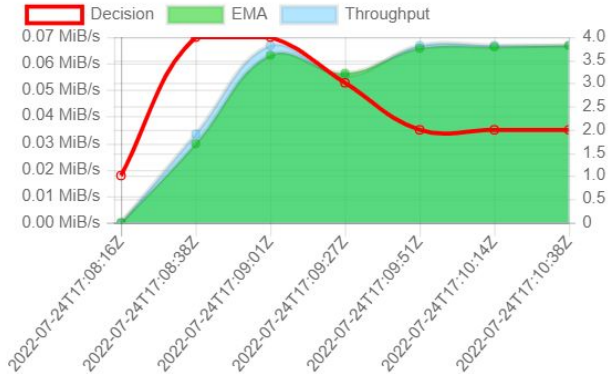
```
mysql> select * from t_server_config;
+-------+----------------+------------+---------------+---------+--------------+--------------+-----------------------------------------+
| retry | max_time_queue | sec_per_mb | global_timeout | vo_name | no_streaming | show_user_dn | alto_pv_uri                             |
+-------+----------------+------------+---------------+---------+--------------+--------------+-----------------------------------------+
|     0 |              0 |          0 |             0 | *       | NULL         | NULL         | http://mininet:9090/pathvector/pv |
+-------+----------------+------------+---------------+---------+--------------+--------------+-----------------------------------------+
1 row in set (0.00 sec)

mysql> select source_se, dest_se, min_active, max_active, optimizer_mode, tcn_abs_limit, tcn_rel_weight from t_link_config;
+------------+------------+------------+------------+----------------+---------------+---------------+
| source_se  | dest_se    | min_active | max_active | optimizer_mode | tcn_abs_limit | tcn_rel_weight |
+------------+------------+------------+------------+----------------+---------------+---------------+
| *          | *          |          2 |         10 |              2 |          NULL |          NULL |
| root://xrd1 | *          |          2 |          8 |              4 |          NULL |            10 |
| root://xrd2 | root://xrd3 |          2 |          8 |              4 |        200000 |          NULL |
+------------+------------+------------+------------+----------------+---------------+---------------+
3 rows in set (0.00 sec)
```

URI to access ALTO PV resource

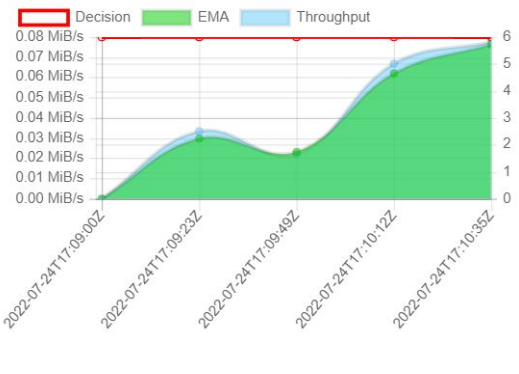
TCN resource control specification

TCN Optimizer
(zero order gradient algorithm with ALTO PV constraints)



Throughput	EMA	Diff	Explanation
68.27 KiB/s	68.26 KiB/s	0	Calculated by zero order gradient optimizer.
68.27 KiB/s	68.16 KiB/s	0	Calculated by zero order gradient optimizer.
68.27 KiB/s	67.20 KiB/s	-1	Calculated by zero order gradient optimizer.
56.89 KiB/s	57.65 KiB/s	-1	Calculated by zero order gradient optimizer.
68.27 KiB/s	64.51 KiB/s	0	Calculated by zero order gradient optimizer.
34.13 KiB/s	30.72 KiB/s	3	Calculated by zero order gradient optimizer.
0 bytes/s	0 bytes/s	-3	Calculated by zero order gradient optimizer.

FTS Optimizer



Wrap Up and Looking Forward

- **ALTO WG Contact:**
  - IETF ALTO WG: https://datatracker.ietf.org/wg/alto/about/
- **ALTO Code Base Project:**
  - Repo: https://github.com/openalto/
  - IETF Hackathon 114 ALTO Scrum Dashboard: https://github.com/orgs/openalto/projects/1/views/1
- **Next steps:**
  - Production evaluations and deployment at CERN
- **Want to contribute to OpenALTO as a developer?** Reach us out: **jros at qti.qualcomm.com**

Backup

What Got Done

- Application side:
  - C++ ALTO client integrated with FTS.
    - [TBD] Pull request to FTS
  - An ALTO-enabled TCN optimizer which improves aggregated throughput.
  - https://github.com/fno2010/fts3/tree/zero-order-grad
- Server side:
  - Path vector service is added to the OpenALTO project, providing bandwidth and routing information.
  - Network information collected from Mininet/OpenDaylight
  - https://github.com/openalto/alto/tree/ietf114
- Infrastructure:
  - *Customizable* containerized environment for FTS over ContainterNet.
  - https://github.com/openalto/ietf-hackathon/tree/fts