Steering Traffic with the ALTO Protocol to Optimize Network-Aware Application Performance

OPSAWG Meeting 7/29/2022

on behalf of IETF ALTO WG (Jordi Ros Giralt, Y. Richard Yang
Jacob, Jensen, Kai, Alex, Mahdi, Roland, Sabine …)

Outline

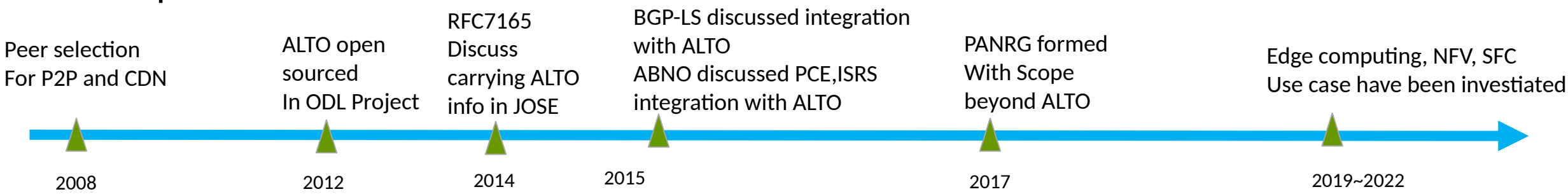
- What is the ALTO?
  - ALTO protocol overview
  - Basic ALTO Network Abstraction
  - Basic ALTO Transport Framework
  - ALTO Usage Examples
- ALTO WG Documents Status Update
- ALTO Implementation and Application Integration
- Summary and Takeaway

What is the ALTO?

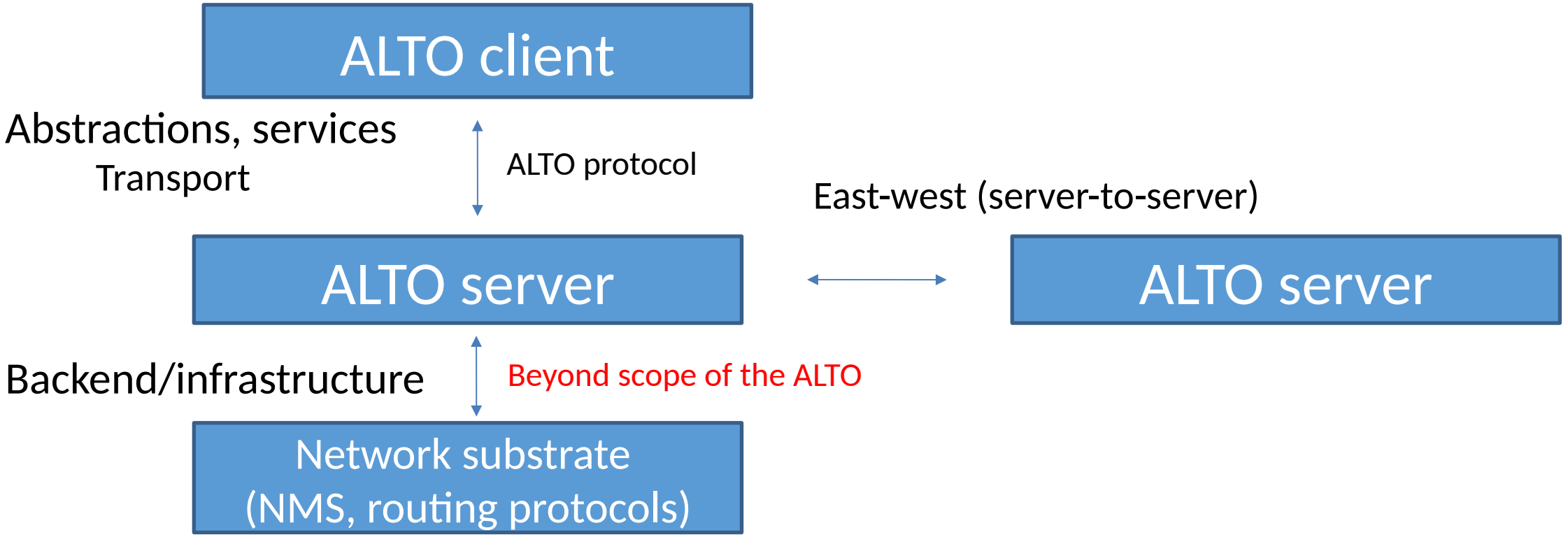
- ALTO is an effort of the ALTO Working Group in the Transport Area of Internet Engineering Task Force (IETF)

- **ALTO high-level goal**: provide a standard for applications and networks to work together to optimize traffic patterns to improve both network and application performance
  - Initial for data distribution by peer-to-peer applications (e.g., BitTorrent)
  - Evolved to include content distribution networks (CDN)

ALTO includes two components: (1) abstractions of network states/services, and (2) transport of network information



General ALTO Protocol Overview



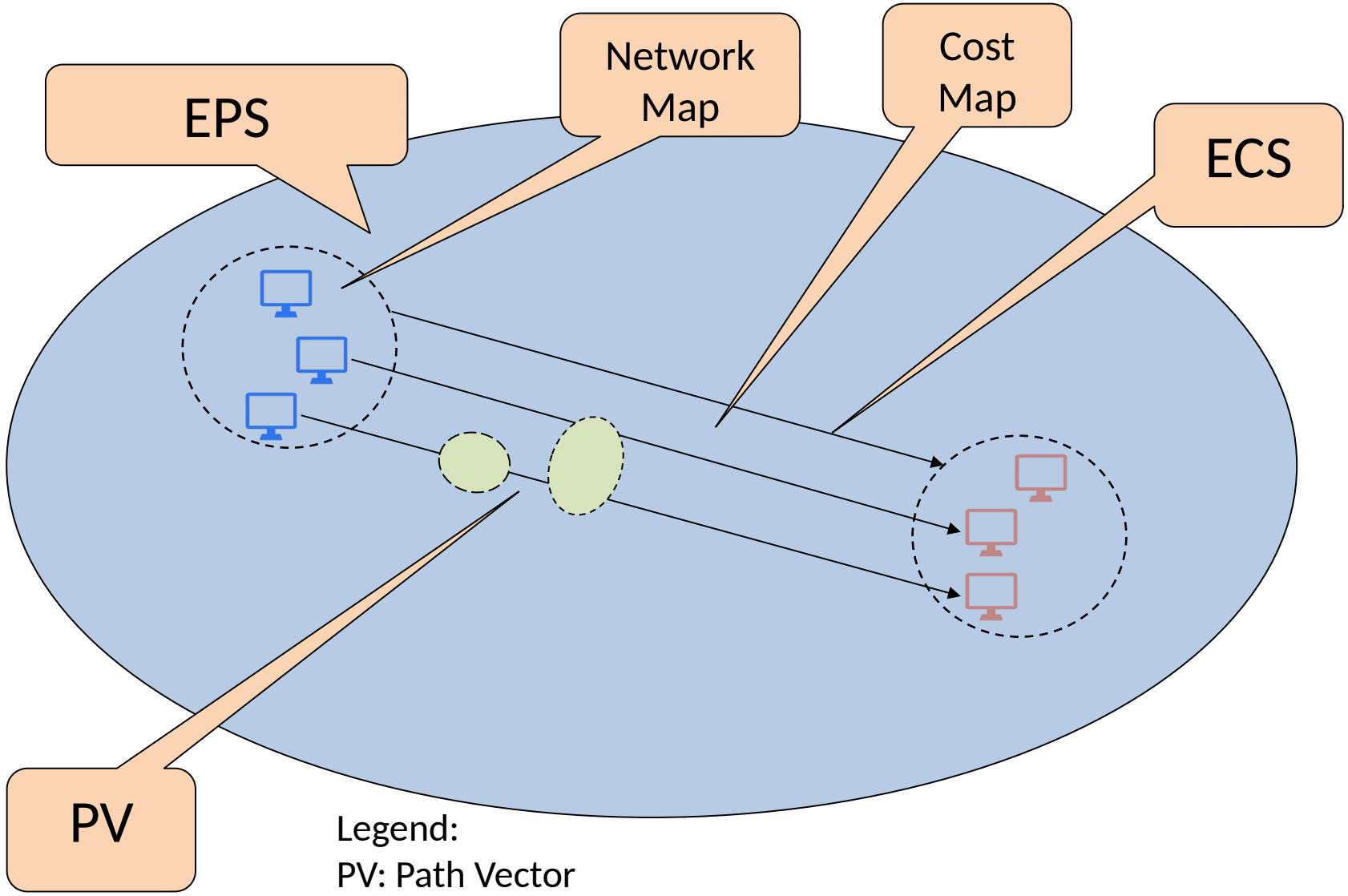
Basic ALTO Network Abstractions

A network consists of:

- A set of entities, where each entity can be an
  - Endpoint
  - Aggregations of endpoints (PID)
  - Abstract network element
- Entities have properties that can be inherited
  - Entities can have capabilities

- A set of paths, where each path is from a src entity to a dst entity
- A path has path properties:
  - Cost metrics (originally cost was called distance)
  - Complex properties: multicost, calendar, path vector
- Path Vector:A set of src-dst pairs can form a co-flow, with shared abstract network elements

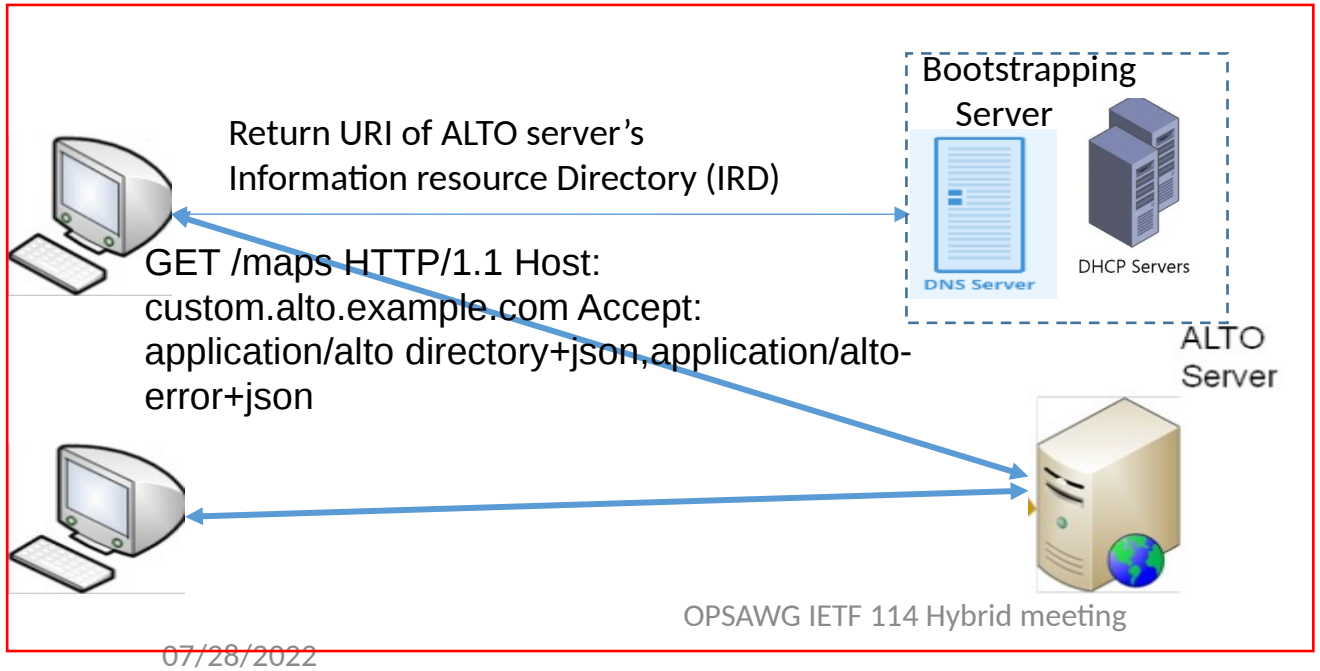
ALTO Network Abstractions and Services



Legend:
PV: Path Vector
ECS: Endpoint Cost Service
EPS: Endpoint Property Service

Basic ALTO Transport Framework

- Bootstrap server provided by server discovery
  - Server discovery (RFC7286), xdom discovery (RFC8686)
- Network information divided into (network) information resources
- List of available information resources provided by Information Resource Directory (IRD)
- Each individual information resource is provided as a RESTful service
- Information resources distributed using either pull (RFC 7285) or incremental-push (RFC8895)



Transport Protocol: HTTP 1.1
Encoding: JSON Plaint Text for Debug
Message Format: Independent of HTTP protocol

Example: Endpoint Cost and HTTP Transport

11.5.1.7. Example

```
POST /endpointcost/lookup HTTP/1.1
Host: alto.example.com
Content-Length: 248
Content-Type: application/alto-endpointcostparams+json
Accept: application/alto-endpointcost+json,application/alto-error+json

{
  "cost-type": {"cost-mode" : "ordinal",
                "cost-metric" : "routingcost"},
  "endpoints" : {
    "srcs": [ "ipv4:192.0.2.2" ],
    "dsts": [
      "ipv4:192.0.2.89",
      "ipv4:198.51.100.34",
      "ipv4:203.0.113.45"
    ]
  }
}
```

RFC7285

```
HTTP/1.1 200 OK
Content-Length: 274
Content-Type: application/alto-endpointcost+json

{
  "meta" : {
    "cost-type": {"cost-mode" : "ordinal",
                  "cost-metric" : "routingcost"
    }
  },
  "endpoint-cost-map" : {
    "ipv4:192.0.2.2": {
      "ipv4:192.0.2.89"    : 1,
      "ipv4:198.51.100.34" : 2,
      "ipv4:203.0.113.45"  : 3
    }
  }
}
```

- Many performance metrics: one-way delay (delay-ow), round-trip delay (delay-rt), delay variation (delay-var), loss rate (lossrate), residual bw (bw-residual), available bw (bw-available), TCP throughput (**tput**), hop count

Example: Entity Properties and SSE Incremental Push Update

```
POST /updates/properties HTTP/1.1
Host: alto.example.com
Accept: text/event-stream
Content-Type: application/alto-updatestreamparams+json
Content-Length: 511

{ "add": {
    "props-1": {
      "resource-id": "my-props",
      "input": {
        "properties" : [ "priv:ietf-bandwidth" ],
        "endpoints" : [
          "ipv4:198.51.100.1",
          "ipv4:198.51.100.2",
          "ipv4:198.51.100.3"
        ]
      }
    },
    "props-2": {
      "resource-id": "my-props",
      "input": {
        "properties" : [ "priv:ietf-load" ],
        "endpoints" : [
          "ipv6:2001:db8:100::1",
          "ipv6:2001:db8:100::2",
          "ipv6:2001:db8:100::3"
        ]
      }
    }
  }
}
```

```
HTTP/1.1 200 OK
Connection: keep-alive
Content-Type: text/event-stream

event: application/alto-updatestreamcontrol+json
data: {"control-uri":
data: "https://alto.example.com/updates/streams/1414213562373"}

event: application/alto-endpointprops+json,props-1
data: { "endpoint-properties": {
data:     "ipv4:198.51.100.1" : { "priv:ietf-bandwidth": "13" },
data:     "ipv4:198.51.100.2" : { "priv:ietf-bandwidth": "42" },
data:     "ipv4:198.51.100.3" : { "priv:ietf-bandwidth": "27" }
data:  } }

event: application/alto-endpointprops+json,props-2
data: { "endpoint-properties": {
data:     "ipv6:2001:db8:100::1" : { "priv:ietf-load": "8" },
data:     "ipv6:2001:db8:100::2" : { "priv:ietf-load": "2" },
data:     "ipv6:2001:db8:100::3" : { "priv:ietf-load": "9" }
data:  } }

   (pause)

event: application/merge-patch+json,props-1
data: { "endpoint-properties":
data:   {"ipv4:198.51.100.1" : {"priv:ietf-bandwidth": "3"}}
data: }

   (pause)

event: application/merge-patch+json,props-2
data: { "endpoint-properties":
data:   {"ipv6:2001:db8:100::3" : {"priv:ietf-load": "7"}}
data: }
```

```
POST /updates/streams/1414213562373" HTTP/1.1
Host: alto.example.com
Accept: text/plain,application/alto-error+json
Content-Type: application/alto-updatestreamparams+json
Content-Length: 448

{ "add": {
    "props-3": {
      "resource-id": "my-props",
      "input": {
        "properties" : [ "priv:ietf-bandwidth" ],
        "endpoints" : [
          "ipv4:198.51.100.4",
          "ipv4:198.51.100.5"
        ]
      }
    },
    "props-4": {
      "resource-id": "my-props",
      "input": {
        "properties" : [ "priv:ietf-load" ],
        "endpoints" : [
          "ipv6:2001:db8:100::4",
          "ipv6:2001:db8:100::5"
        ]
      }
    }
  }
}
```

- More details see RFC8895

Example: Path Vector Supporting Co-Flow

```
POST /endpointcost/pv HTTP/1.1
Host: alto.example.com
Accept: multipart/related;
        type=application/alto-endpointcost+json,
        application/alto-error+json
Content-Length: 362
Content-Type: application/alto-endpointcostparams+jso

{
  "cost-type": {
    "cost-mode": "array",
    "cost-metric": "ane-path"
  },
  "endpoints": {
    "srcs": [
      "ipv4:192.0.2.34",
      "ipv6:2001:db8::3:1"
    ],
    "dsts": [
      "ipv4:192.0.2.2",
      "ipv4:192.0.2.50",
      "ipv6:2001:db8::4:1"
    ]
  },
  "ane-property-names": [
    "max-reservable-bandwidth",
    "persistent-entity-id"
  ]
}
```

```
HTTP/1.1 200 OK
Content-Length: 1433
Content-Type: multipart/related; boundary=example-2;
              type=application/alto-endpointcost+json

--example-2
Content-ID: <ecs@alto.example.com>
Content-Type: application/alto-endpointcost+json

{
  "meta": {
    "vtags": {
      "resource-id": "endpoint-cost-pv.ecs",
      "tag": "bb6bb72eafe8f9bdc4f335c7ed3b10822a391cef"
    },
    "cost-type": {
      "cost-mode": "array",
      "cost-metric": "ane-path"

    }
  },
  "endpoint-cost-map": {
    "ipv4:192.0.2.34": {
      "ipv4:192.0.2.2":   [ "NET3", "L1", "NET1" ],
      "ipv4:192.0.2.50":   [ "NET3", "L2", "NET2" ]
    },
    "ipv6:2001:db8::3:1": {
      "ipv6:2001:db8::4:1": [ "NET3", "L2", "NET2" ]
    }
  }
}
```

```
--example-2
Content-ID: <propmap@alto.example.com>
Content-Type: application/alto-propmap+json

{
  "meta": {
    "dependent-vtags": [
      {
        "resource-id": "endpoint-cost-pv.ecs",
        "tag": "bb6bb72eafe8f9bdc4f335c7ed3b10822a391cef"
      },
      {
        "resource-id": "ane-props",
        "tag": "bf3c8c1819d2421c9a95a9d02af557a3"
      }
    ]
  },
  "property-map": {
    ".ane:NET1": {
      "max-reservable-bandwidth": 50000000000,
      "persistent-entity-id": "ane-props.ane:MEC1"
    },
    ".ane:NET2": {
      "max-reservable-bandwidth": 50000000000,
      "persistent-entity-id": "ane-props.ane:MEC2"
    },
    ".ane:NET3": {
      "max-reservable-bandwidth": 50000000000
    },
    ".ane:L1": {
      "max-reservable-bandwidth": 10000000000
      },
    ".ane:L2": {
      "max-reservable-bandwidth": 15000000000
      }
  }
```

- More details see https://datatracker.ietf.org/doc/html/draft-ietf-alto-path-vector-21#section-8.1

IETF ALTO RFCs/Ongoing/Drafts

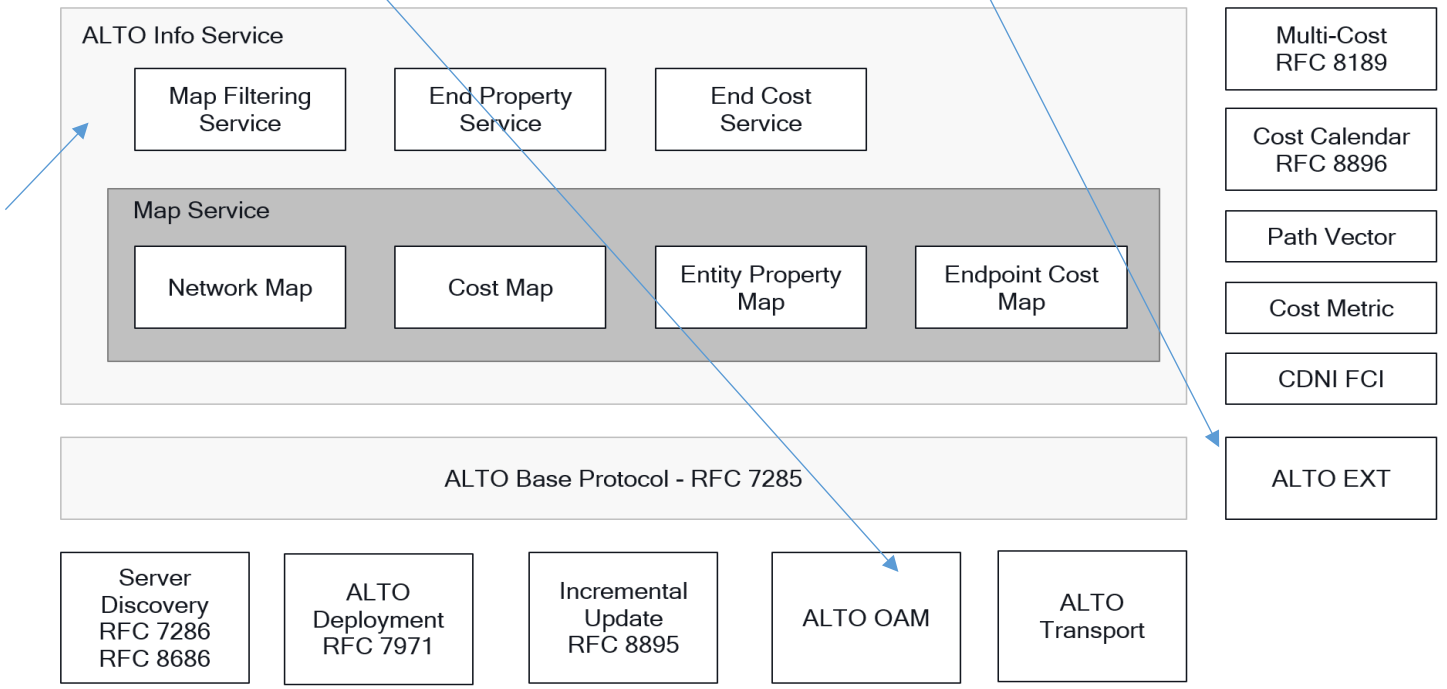
- Xdom Service Discovery (RFC8686)
- Multicost (RFC8189)
- Deployment (RFC7971)
- Server Discovery (RFC7286)
- Base Protocol (RFC7285)
- Requirements (RFC6708)
- Problem Statement (RFC5693)
- SSE/Incr Update (RFC8895)
- Cost Calendar (RFC8896)
- CDNi (RFC9241)
- Path Vector (RFC editor)
- Unified Properties (RFC9240)
- Cost Metrics (RFC editor)
- Cost Mode (RFC9274)

ALTO OAM

ALTO New Transport

MEC drafts (service edge, Service function handling, NEF, Compute Aware Networking)

Cellular/mobile networks drafts (MOWIE, cellular addresses, mobility network models)



ALTO Deployment Update

- Current implementations/deployments:
- Wiki list of implementations: https://trac.ietf.org/trac/alto/wiki/Impl
  - Examples:
    - Comcast P4P: RFC 5632: https://datatracker.ietf.org/doc/html/rfc5632
    - Benocs: https://people.csail.mit.edu/gsmaragd/publications/CoNEXT2019/CoNEXT2019.pdf
    - Telefonica: https://dl.ifip.org/db/conf/im/im2021mini/212012.pdf

- Forthcoming new deployments (work in progress):
  - Pacific Research Platform
  - CERN Rucio
  - UCSD 5G
  - NY City Cosmos 5G
  - ESnet

Example ALTO Deployment Setting: Flow Director

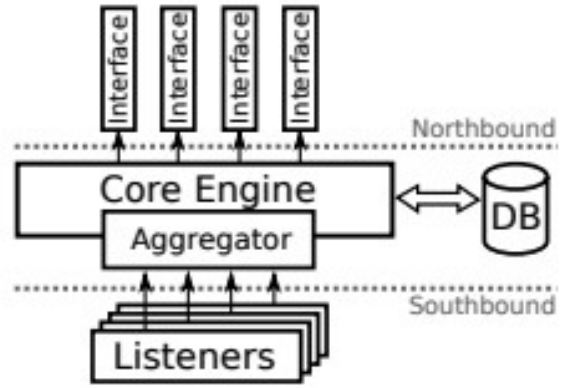


Figure 9: *Flow Director*: High-level system architecture.

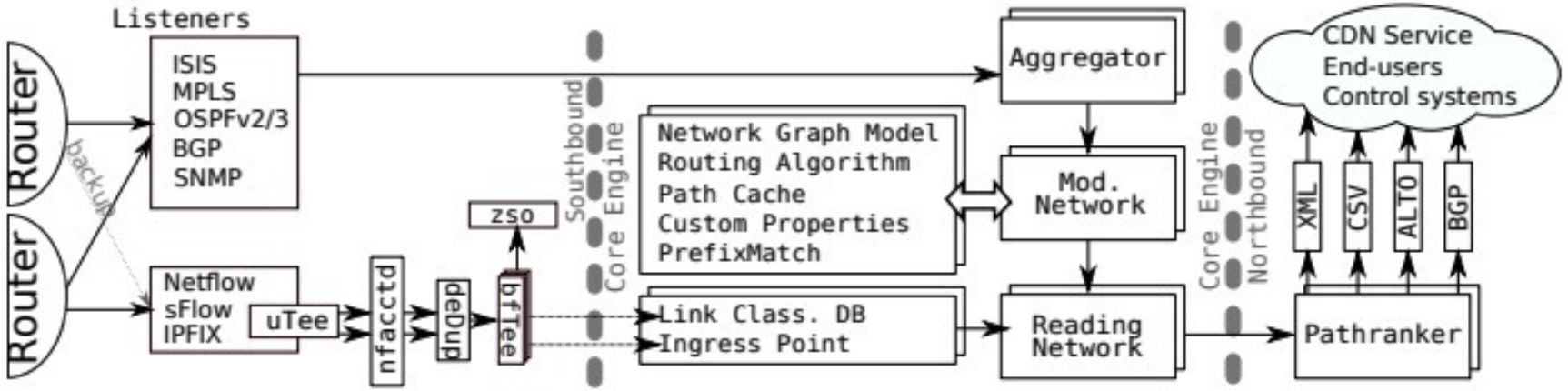


Figure 10: *Flow Director* processing pipeline.

source: http://people.csail.mit.edu/gsmaragd/publications/CoNEXT2019/CoNEXT2019.pdf
CoNEXT 2019 Best Paper Award; IETF/IRTF 2020 Applied Networking Research Prize

Example ALTO Deployment Setting: Telifonica CDN deployment roadmap

Technology lab tests

- Initial tests with ALTO module of ODL
- Integration with ODL BGP (originally LLDP)
- Monovendor router scenario
- Virtualized routers
- Virtualized ALTO
- Simple IP network based on OSPF as IGP
- Single AS
- Simple metrics (= hopcount)
- Some of the routers acting as RR

Pre-production network tests

- Migration to exaBGP
- Fixing of issues in exaBGP (3 ticketsraised and solved) mainly related to BGP-LS[*]
- Multivendor router scenario
- Physical routers
- Dedicated ALTO server
- Complex MPLS network combining OSPF and IS-IS
- Multiple private ASs
- More sophisticated metrics in IGP
- Dedicated RR, separated for BGP and BGP-LS

Integration in production network

- Adaptation to production processes and rules
- Hardening of all the environment to prevent security issues (HW, SW, …)
- Limited activation of BGP-LS by now
- Coexistence with many other services in the network
- Complete deployment expected for Q3’22

[*] https://github.com/Exa-Networks/exabgp/issues/1071
https://github.com/Exa-Networks/exabgp/pull/1075
https://github.com/Exa-Networks/exabgp/issues/1077

ALTO Hackathon Project Update (IETF 113,114)

- Implementation of an ALTO Client in Python (RFC 7285)
- Integration with CERN Rucio replica download
  - Submitted pull request to Rucio Project: – https://github.com/rucio/rucio/pull/5364
  - 3 Demos [https://github.com/openalto/ietf-hackathon/issues/8]
    - [D1] Single-flow replica node selection using ALTO BW Cost Map
    - [D2] ALTO Estimator: Multi-flow BW prediction
    - [D3] ALTO Scheduler: SLA-constrained multi-flow node selection
  - Southbound ALTO integration with SDN:
    - Mininet/Pox, OpenDaylight
    - Scrum dashboard: https://github.com/orgs/openalto/projects/1/views/1

Summary and Take away

- ALTO is originally designed for P2P application, has evolved to support CDN application
  - ALTO further development can be driven by real application such as CAN (Compute aware Networking)
    - ALTO provide centralized approach to support service placement
    - Both network metric and compute metric can be exposed to the client
- ALTO enable network and application integration and provide a good Network visibility
  - Intelligence: AI, ML
  - Control: Service Invocation, Peer Selection, Path Selection, Resource Schedule Control, Congestion Control, etc

- Further work plan (https://github.com/ietf-wg-alto/wg-materials):
  - Telefonica TCDN deployment is coming in the 2nd half of this year
  - Integration w/ Rucio (automatic workflow)
    - Manual workflow (client download)
    - Southbound: Provide automatic, network-aware, dynamic distance []
    - Northbound: Automatic replication QoS aware scheduling