ALTO Multi-Domain Use Cases and Services

Mario Lassnig, Ingmar Poese,
Jordi Ros Giralt, Y. Richard Yang,
Danny Lachos, Chin Guok

March 27, 2023

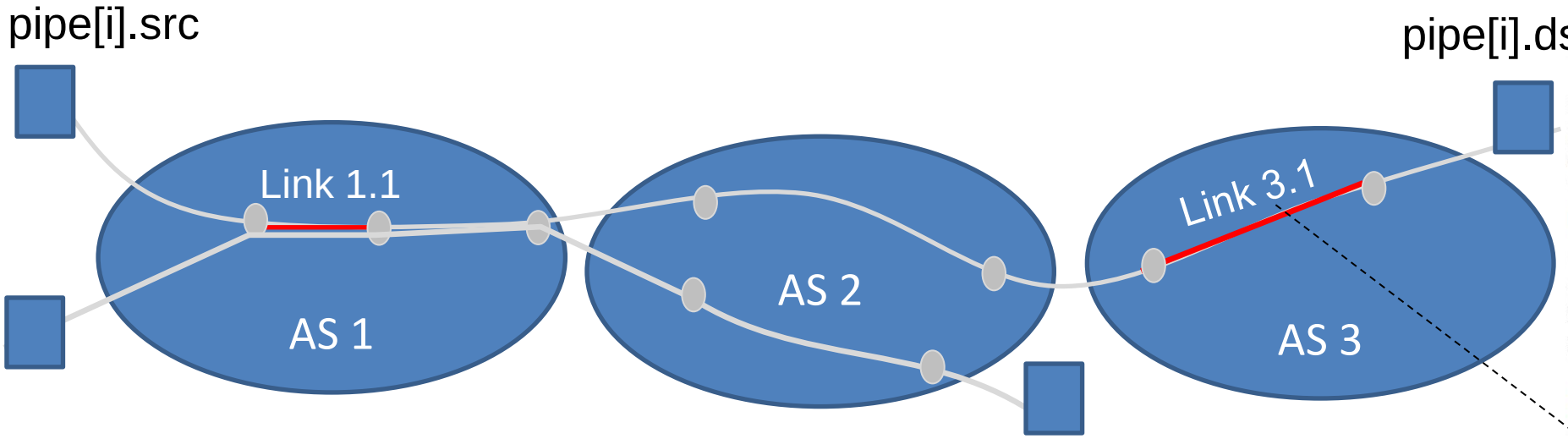
IETF 116

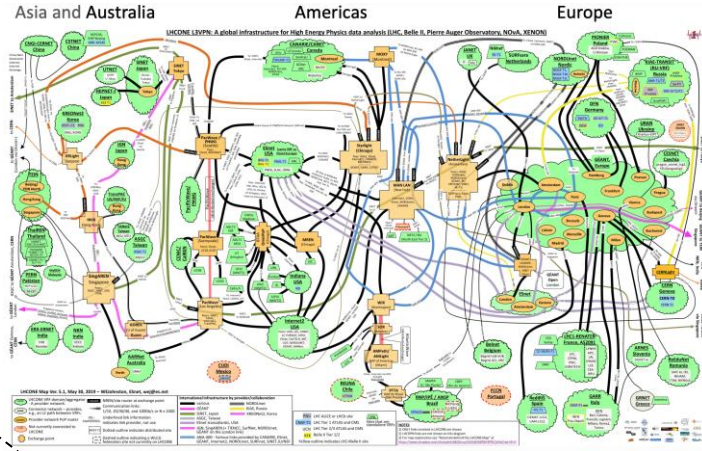
Problem (Relevance)

- RFC 7971: "The ALTO protocol is designed for use cases where the ALTO server and client can be located in different organizations or trust domains. ALTO is inherently designed for use in multi-domain environments. Most importantly, ALTO is designed to enable deployments in which the ALTO server and the ALTO client are not located within the same administrative domain. ”

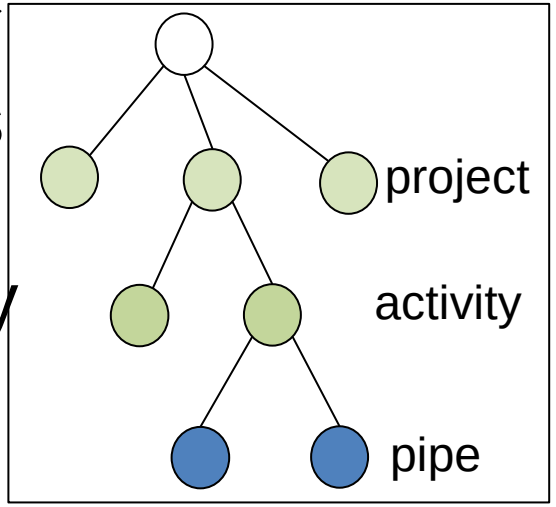
- However, existing core ALTO services including Endpoint Cost Service (ECS) and Cost Map Service query a single ALTO server for the ALTO properties (e.g., routing cost, latency, …) of the whole network path, but the path may span multiple networks.

Use Case Driven by Deployment: Multi-Domain Path->Link Usage (Example: CERN FTS Scheduling Integration)

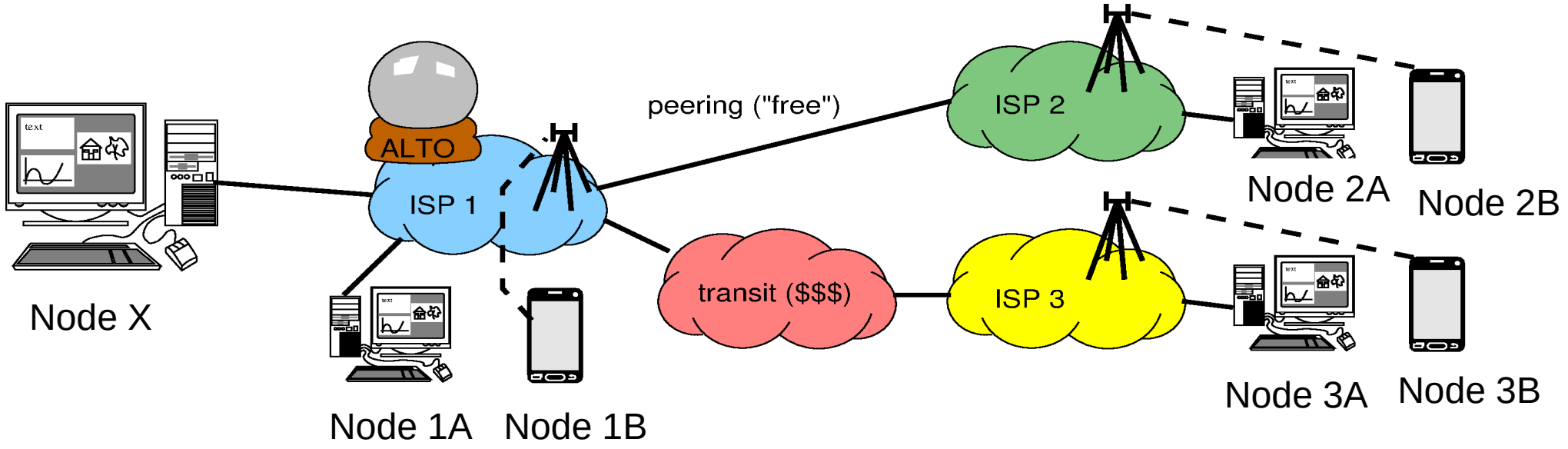




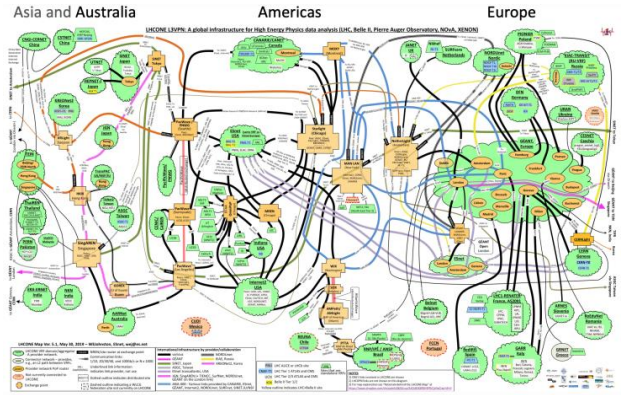
- Multi-domain applications
- App controls transfer pipe traversing a set of resources
- Each resource (link) has resource allocation model
- App supporting app-defined-networking need the ability to map pipe to the set of resources
- More detail see CERN ALTO/FTS integration.



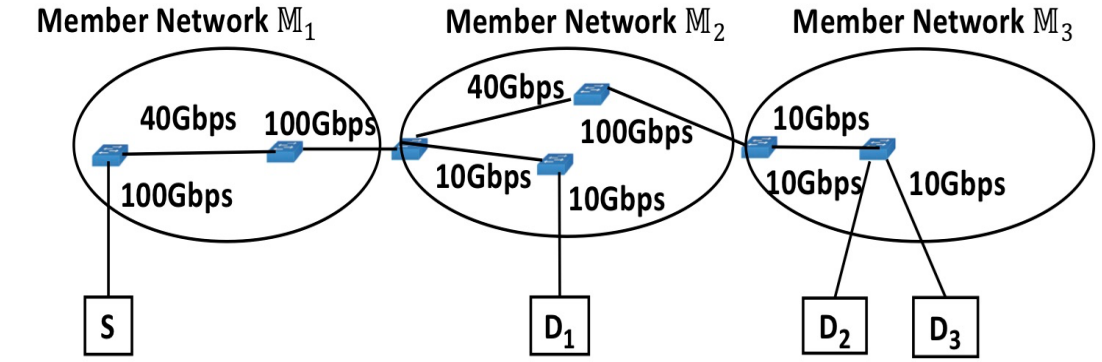
Use Case Driven by Deployment: Multi-domain Path Distance/Ranking (Example: Rucio Distance/Flow Director)



- Node X has 6 potential sources, Node [1-3]A, Node [1-3]B
- Sources span multiple domains
- How to compute distance/ranking for Node X?



Use Case: Multi-domain Co-Flow Resource Discovery (Example: AutoGOLE/SENSE)



$\mathbb{M}_1$:

$x_1 + x_2 + x_3 \leq 100,$

$x_1 + x_2 + x_3 \leq 40,$

$x_1 + x_2 + x_3 \leq 100,$

$\mathbb{M}_2$:

$x_2 + x_3 \leq 40, \quad x_1 \leq 10,$

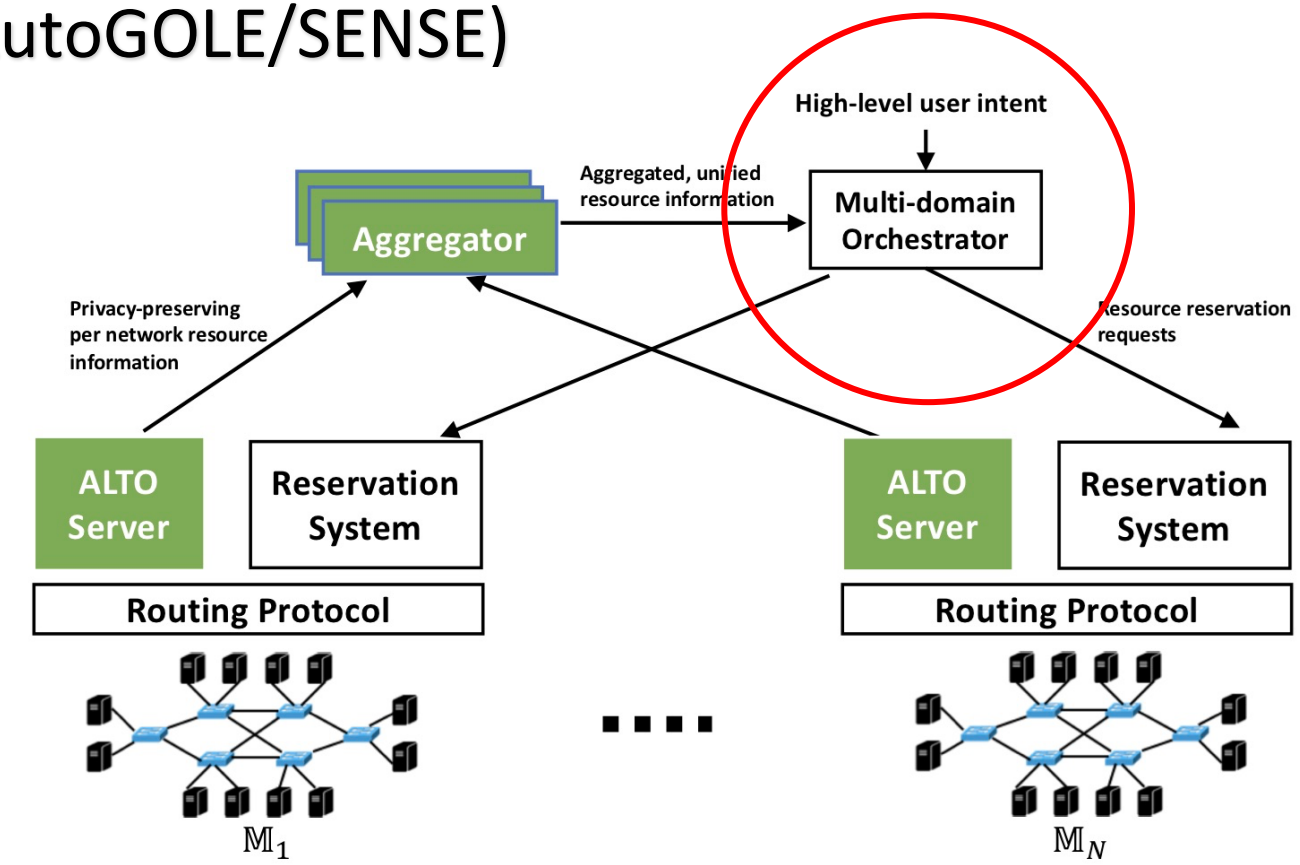
$x_2 + x_3 \leq 100, \quad x_1 \leq 10,$

$\mathbb{M}_3$:

$x_2 + x_3 \leq 10,$

$x_2 \leq 10,$

$x_3 \leq 10,$



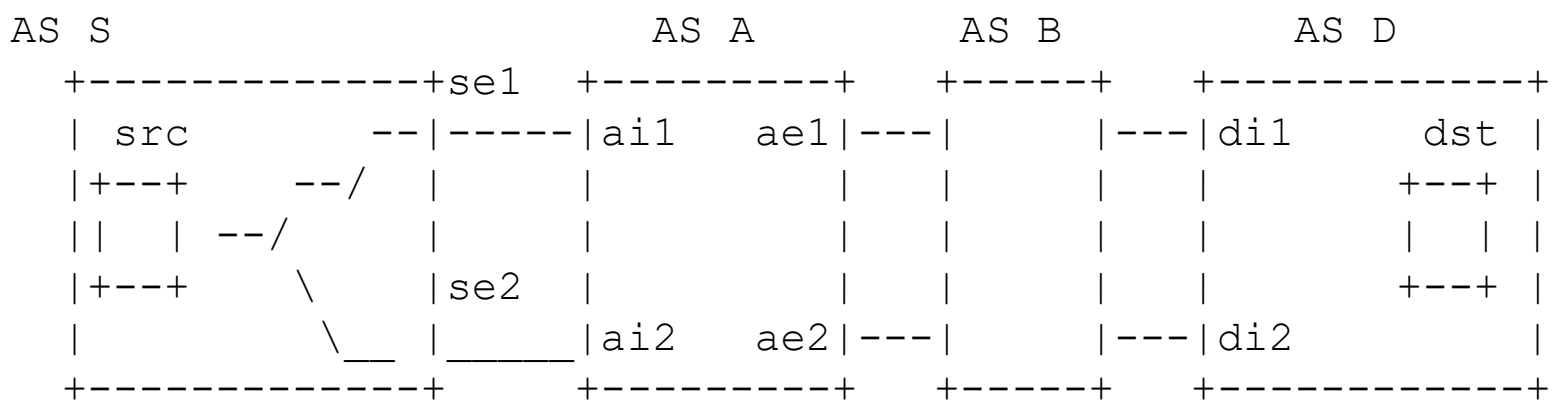
- Large-scale data analytics jobs span multiple networks
- Compute QoS (possible bandwidth) to optimize co-flow finishing time

"Fine-grained, multi-domain network resource abstraction as a fundamental primitive to enable high-performance, collaborative data sciences." ACM/IEEE Supercomputing 2018.

Additional Use Case

- Multi-domain bottleneck structure
- Details see https://datatracker.ietf.org/doc/draft-giraltyellamraju-alto-bsg-multidomain/

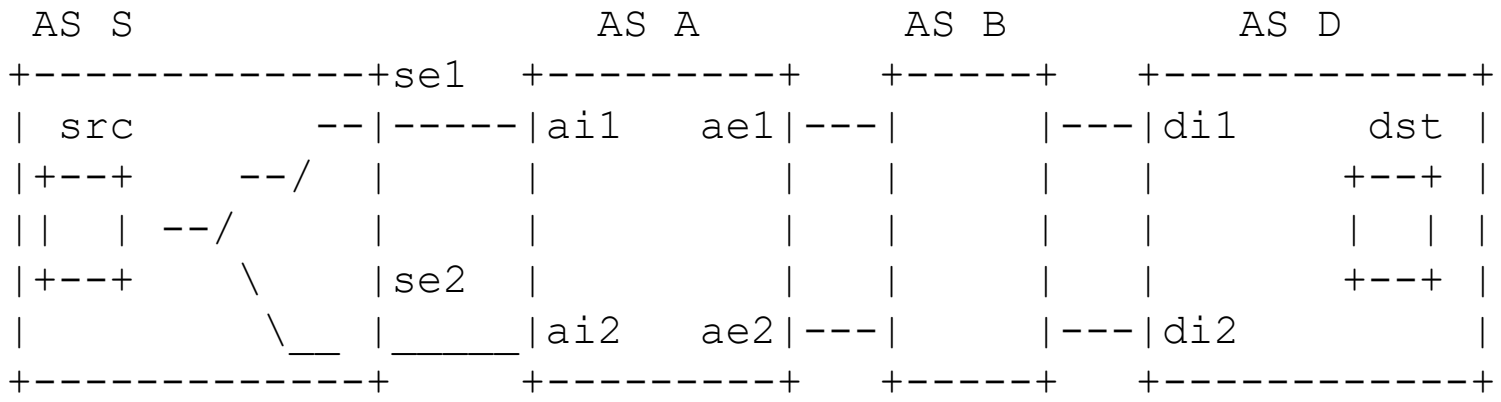
Gap in Current ALTO/Routing Systems



- For the same (src-dst) flow path,
  - Information propagation is upstream
    - AS S can see the whole AS path S A B D; AS A sees only A B D
  - upstream does not notify downstream choice (egress, corresponding ingress at downstream)
    - AS A does not know (by protocol) AS S chooses se1->ai1 or se2->ai2
    - BGP does not have a CHOSEN message; ALTO has no resource to provide the info

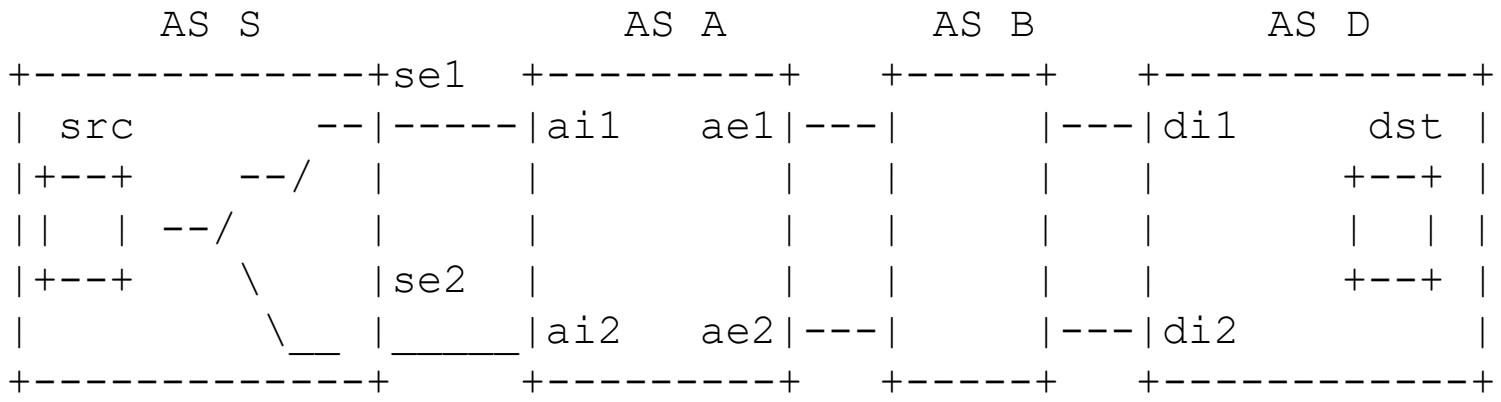
Basic ALTO Extension

- New ALTO service providing egress-notification resource (EN)
  - <flow-info, [ingress]>
    ->
    <egress, domid:next-ingress; [Sebastian proposal: next-alto-server-uri; handle blackhole…]>
- Useful beyond ALTO (egress/ingress verification)
- An east-west interface between ALTO/AS

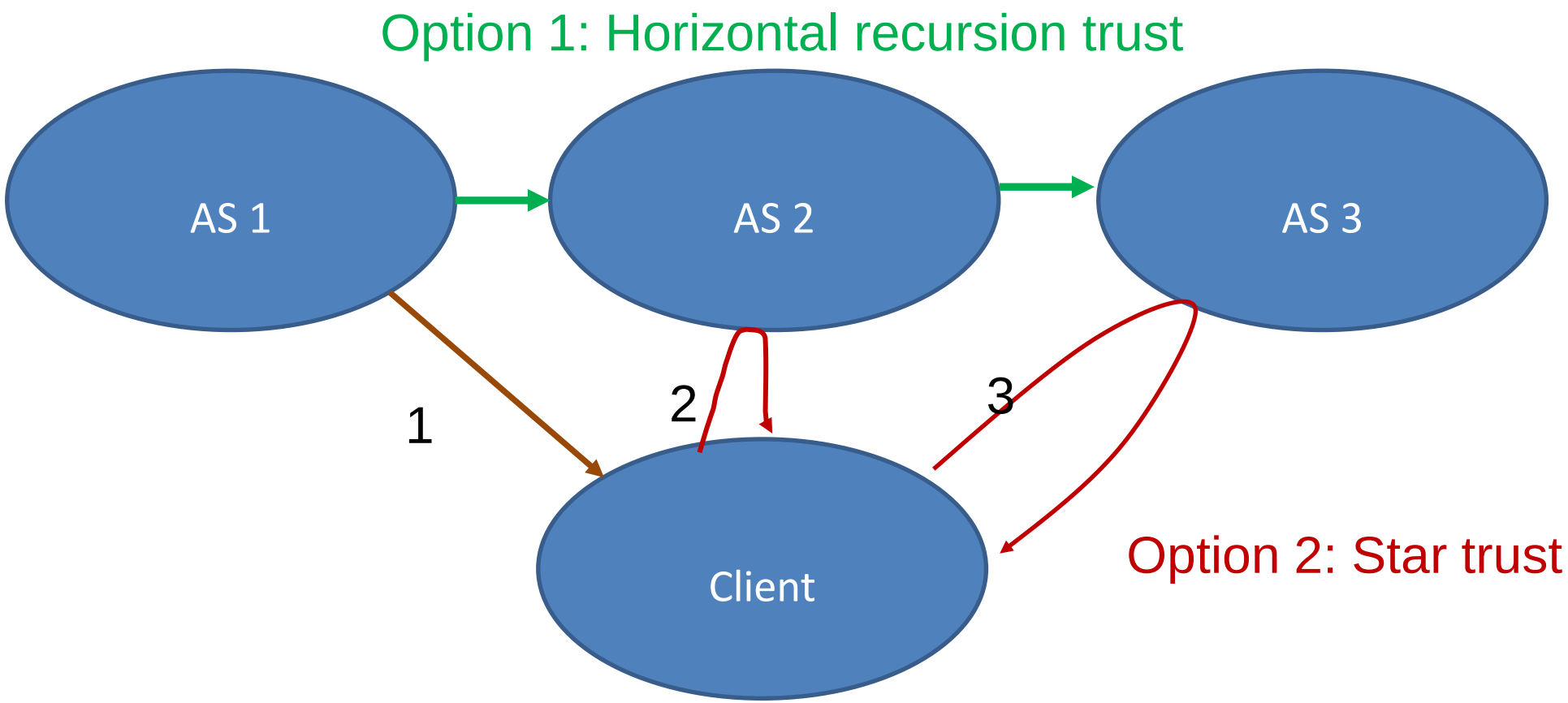


Multidomain Endpoint Cost Service using EN

- Option 1 (Horizontal)
  - ALTO server coordination: Downstream queries upstream for ingress point (can detect anyway; but protocol convey intent, not error, before traffic)
- Option 2 (Vertical)
  - ALTO client goes from upstream to downstream, collecting and informing info along the way

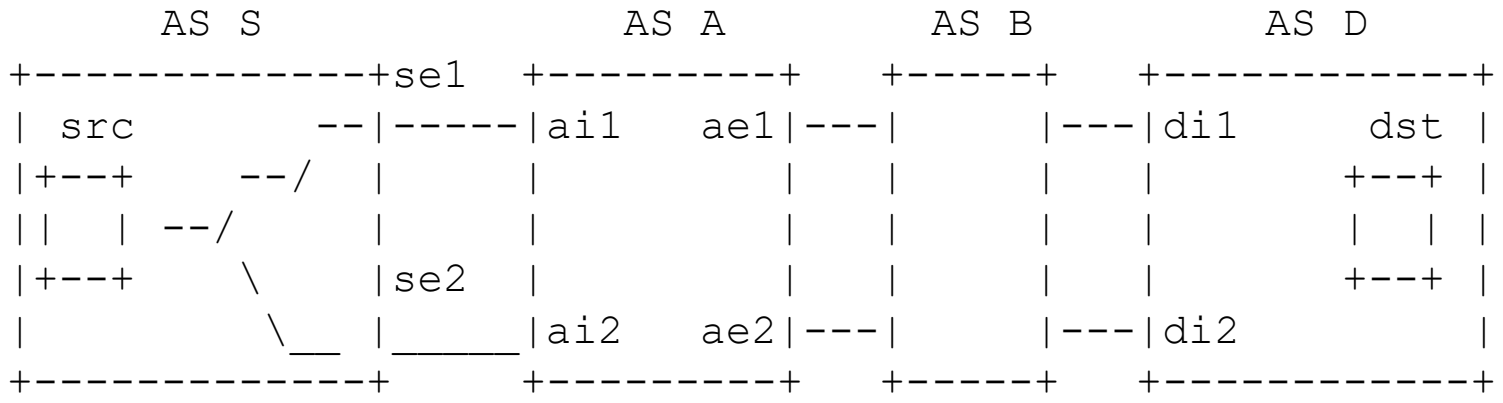


Important Technical Detail: Query and Trust Model



Important Technical Detail: Incremental Deployment

- Incremental deployment: the chaining of domains may be broken due to incremental deployment (e.g., domain sequence is S -> A -> B -> C -> D, but A does not provide EN)
  - Provide **guidance** on
    - how to detect ingress point at B, and
    - how to respond if B cannot detect ingress (multi-answers)
      - See general path abstraction discussion on the mailing list



Related References on Multidomain

- https://datatracker.ietf.org/doc/draft-lachos-alto-multi-domain-use-cases/
- https://datatracker.ietf.org/doc/draft-lachos-sfc-multi-domain-alto/
- https://datatracker.ietf.org/doc/draft-lachosrothenberg-alto-brokermdo/
- https://datatracker.ietf.org/doc/draft-lachosrothenberg-alto-md-e2e-ns/
- https://datatracker.ietf.org/doc/draft-giraltyellamraju-alto-bsg-multidomain/
- Old CERN use case
  - https://ieeexplore.ieee.org/abstract/document/8756056
  - https://www.sciencedirect.com/science/article/abs/pii/S0167739X18302413
- Inter-ALTO communication protocol
  - https://datatracker.ietf.org/doc/draft-dulinski-alto-inter-alto-protocol/
- ALTO network-server, server-server API
  - https://datatracker.ietf.org/doc/draft-medved-alto-svr-apis/

Next Steps

- Organizing interim meetings before May 2023
  - Discuss details of current design implementations
  - Involve operators

Backup Slides

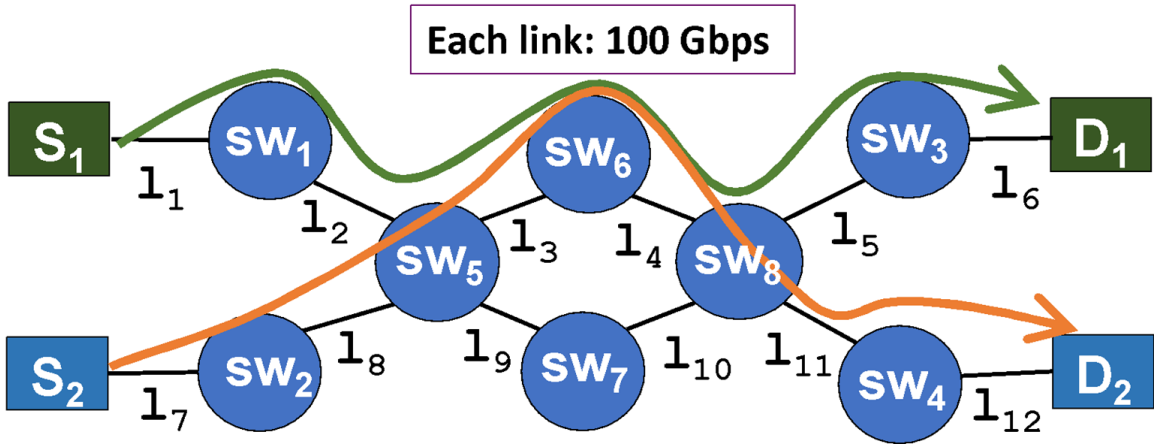
Additional Questions

- The “routing cost” metric makes it difficult to aggregate different point of views
  - See also RFC 8686, Appendix C
- The “ALTO advice” runs in the opposite direction of the money
  - will it always stop at the peering points / Tier-1 carriers?
  - what if the advice given by ISP1’s ALTO server impairs ISP2’s traffic engineering?
  - will ISP1 be legally liable? Thus, will ISP1 refuse to give details wrt. ISP2 even if they knew?

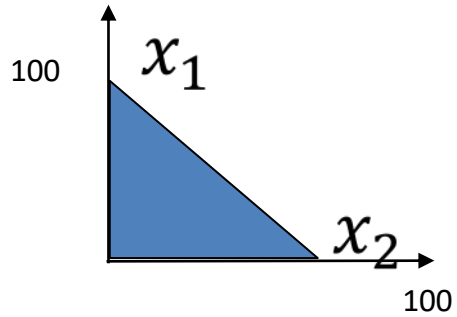
[Sebastian 9/30/2020]

(R)PV: Mathematical Programming as Abstraction Representation to Support Third Use Case

- **GOAL:** Use mathematical programming constraints to provide a compact representation of the **available bandwidth** of flows through **a network**.

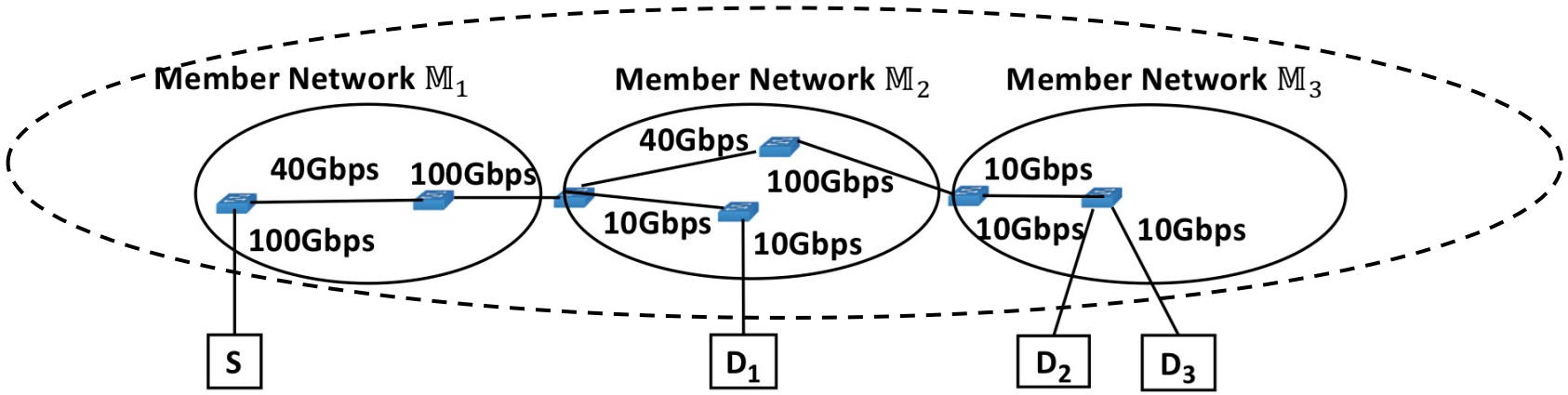


$$x_1 \leq 100 \quad \forall l_u \in \{l_1, l_2, l_5, l_6\},$$
$$x_2 \leq 100 \quad \forall l_u \in \{l_7, l_8, l_{11}, l_{12}\},$$
$$x_1 + x_2 \leq 100 \quad \forall l_u \in \{l_3, l_4\},$$



- **Redundant inequalities can be removed** via a polynomial-time, optimal algorithm.
- Remaining bottlenecks represented as abstract network elements (ANE).

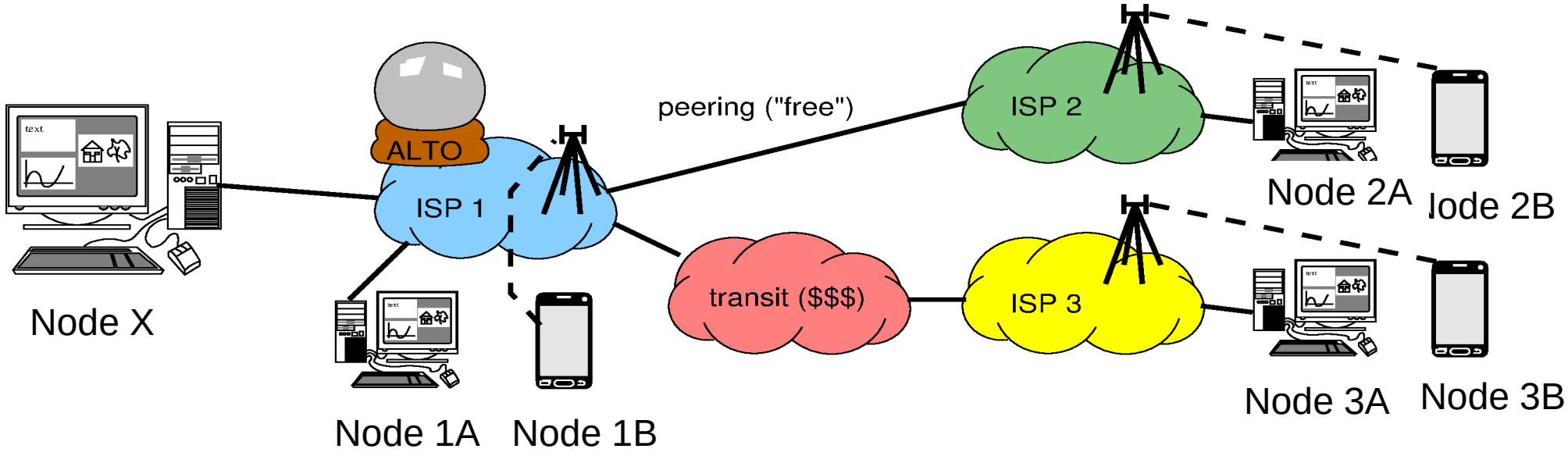
The Reverse View: Mathematical Constraints as Virtual Network Representation



Aggregate the abstraction in multiple networks into a **unified, single, virtual** representation:

$$x_1 \leq 10, \qquad x_2 + x_3 \leq 10$$

Use Case: Multi-domain Path Distance/Ranking (Cost Map/Flow Director/Rucio Distance)



Which distance/ranking should Node X receive?

1. Node1A
2. Node 1 B
3. Node 2A, 2B (*)
4. Node 3A, 3B (*)

1. Node 1A
2. Node 2A, 2B (*)
3. Node1B
4. Node3A, 3B (*)

Is "all within my domain" or "not in my wireless network" more preferable?

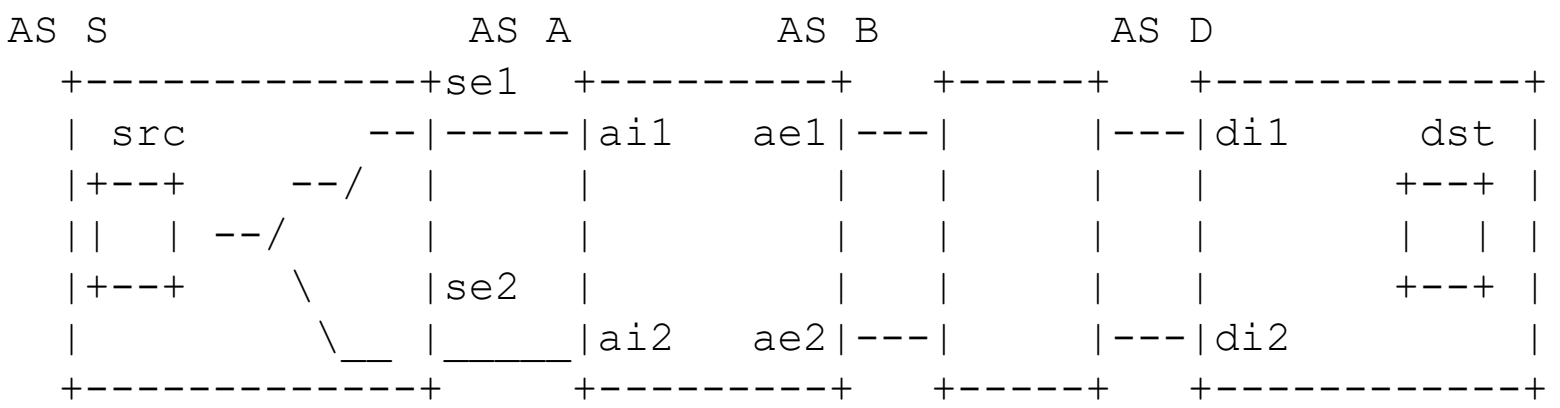
(*) = ?A and ?B are on the same level of preference, because ISP1 might not know that they are wireline vs. wireless, doesn't care (monetary cost is the same for ISP1), and/or wouldn't dare to tell even if they knew.

Feasibility: Simple ALTO Multi-Domain Abstraction

- Starts with a ***simple*** architecture called ALTO Multi-Domain Abstractions (AMDA)
  - The path of a flow from a src to a dst consists of a sequence (vector) of domain segments
    - Domain[0]:{src -> net_0-e} -> Domain[i]{net_i-i -> … -> net_i-e -> net_{i+1}-i -> … -> net_n-e -> dst

    - Domain segments obtained from BGP at source – bootstrapping starts at source

Gap in Current ALTO/Routing Systems



- Missing standard protocol to stitch information across domains
  - Take computing cost/distance src->dst as an example
    - AS S alone has complete path property, but only for BGP path
    - AS S and AS D together can provide only distance from endpoints (e.g., GeoIP)
    - AS A/B in the middle can provide path segments, if it can detect/determine ingress point (not common for downstream to know)
      - Gap: provide an ability to provide ingress point from upstream

Important Technical Detail: Multi-Domain Path Ranking

- The multi-domain path of a flow from a src to a dst consists of those of a sequence of domain segments
  - domain[0]:{src -> dom_0-egress-> dom_1-ingress}
    domain[i]{dom_i-ingress -> dom_i-egress -> dom_{i+1}-egress}
    - src = dom_0-ingress, dst = dom_n-egress

  - List of domains obtained from BGP at source by default => bootstrapping starts at source
- A vector of path cost may no longer define a total order; candidate designs MUST discuss clear guidelines to applications on how to utilize partial ordering, and the consequences (i.e., operations considerations)
  - Leverage SIGCOMM'20 multi-criteria routing design

Additional Issues

- Operation models for extensions [mechanisms, not policies]
  - iterative (client aggregation)
  - recursive (network helped aggregation)
  - Hybrid

- How to handle cost map, not only ECS
- How to handle bandwidth use case