ALTO Multi-Domain Use Cases and Services

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

March 27, 2023

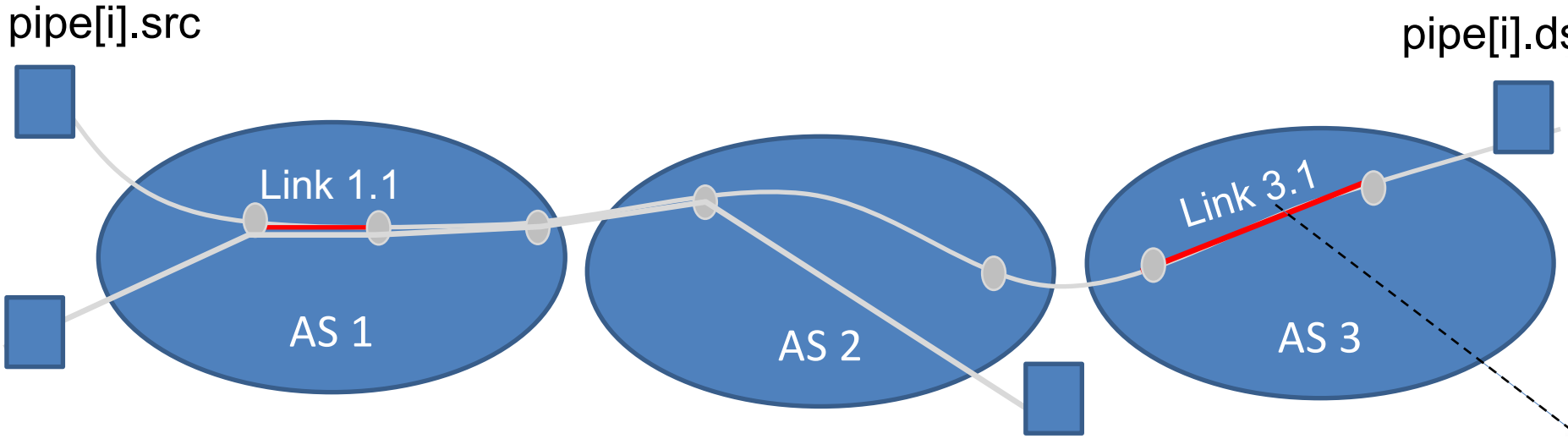
IETF 116

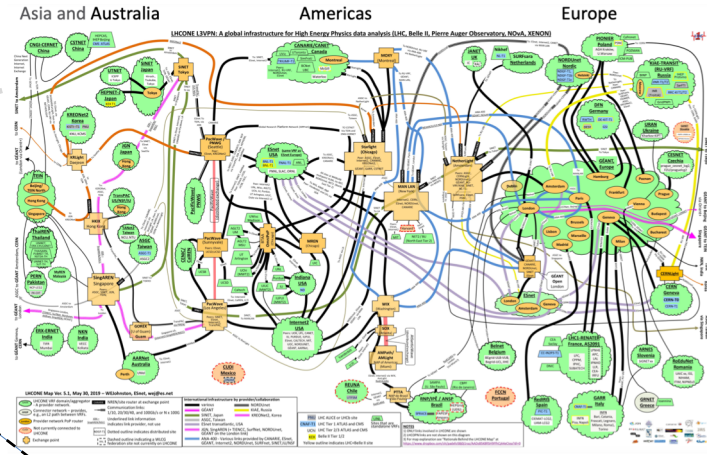
Problem (Relevance)

- 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. ”

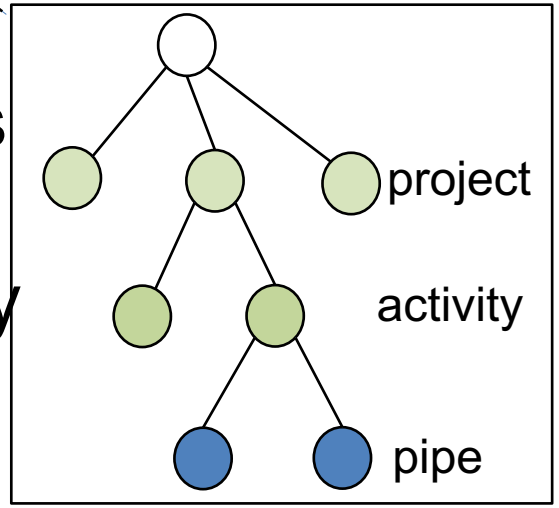
- However, existing core ALTO services including Endpoint Cost Service (ECS) and Cost Map Service query a single ALTO server for the ALTO properties (e.g., routing cost, latency, …) of the whole network path, but the path may span multiple networks.

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

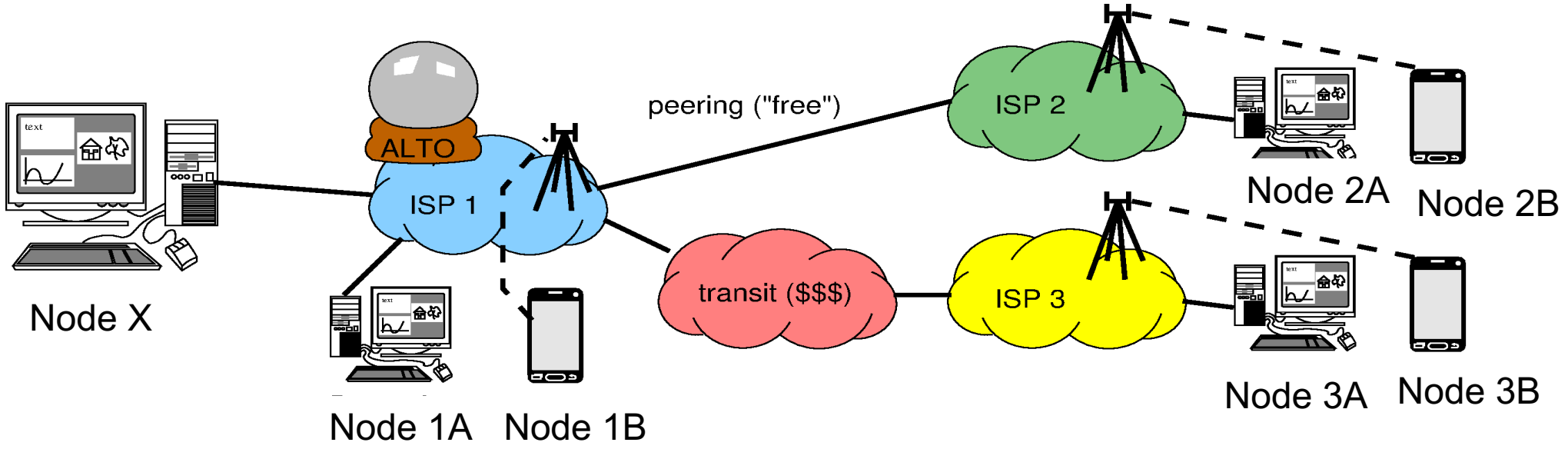




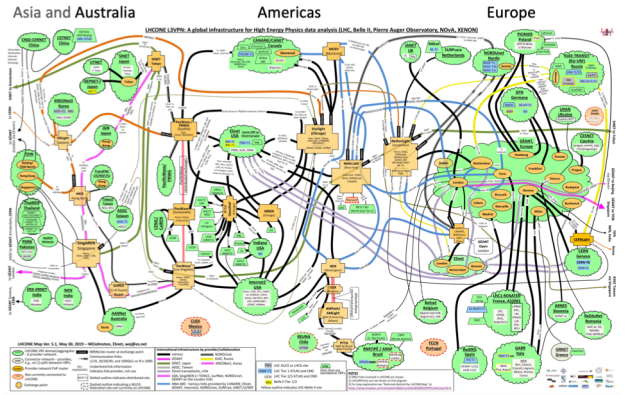
- Multi-domain applications
- App controls transfer pipe traversing a set of resources
- Each resource (link) has resource allocation model
- App supporting app-defined-networking need the ability to map pipe to the set of resources
- More detail see CERN ALTO/FTS integration.



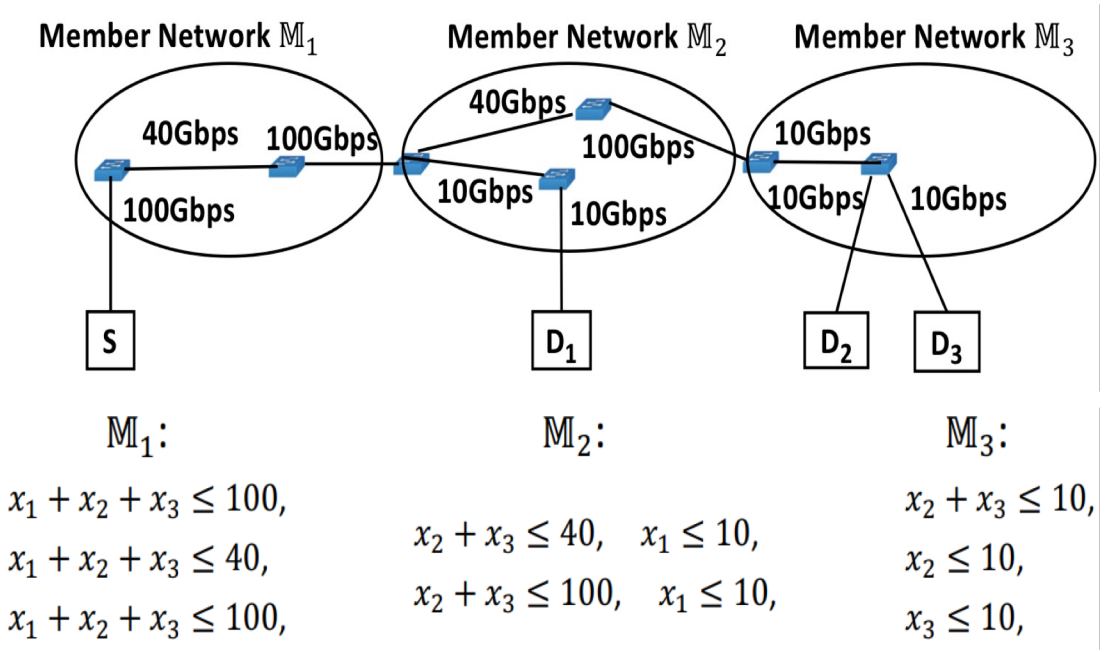
Use Case Driven by Deployment: Multi-domain Path Distance/Ranking (Example: Flow Director/Rucio Distance)

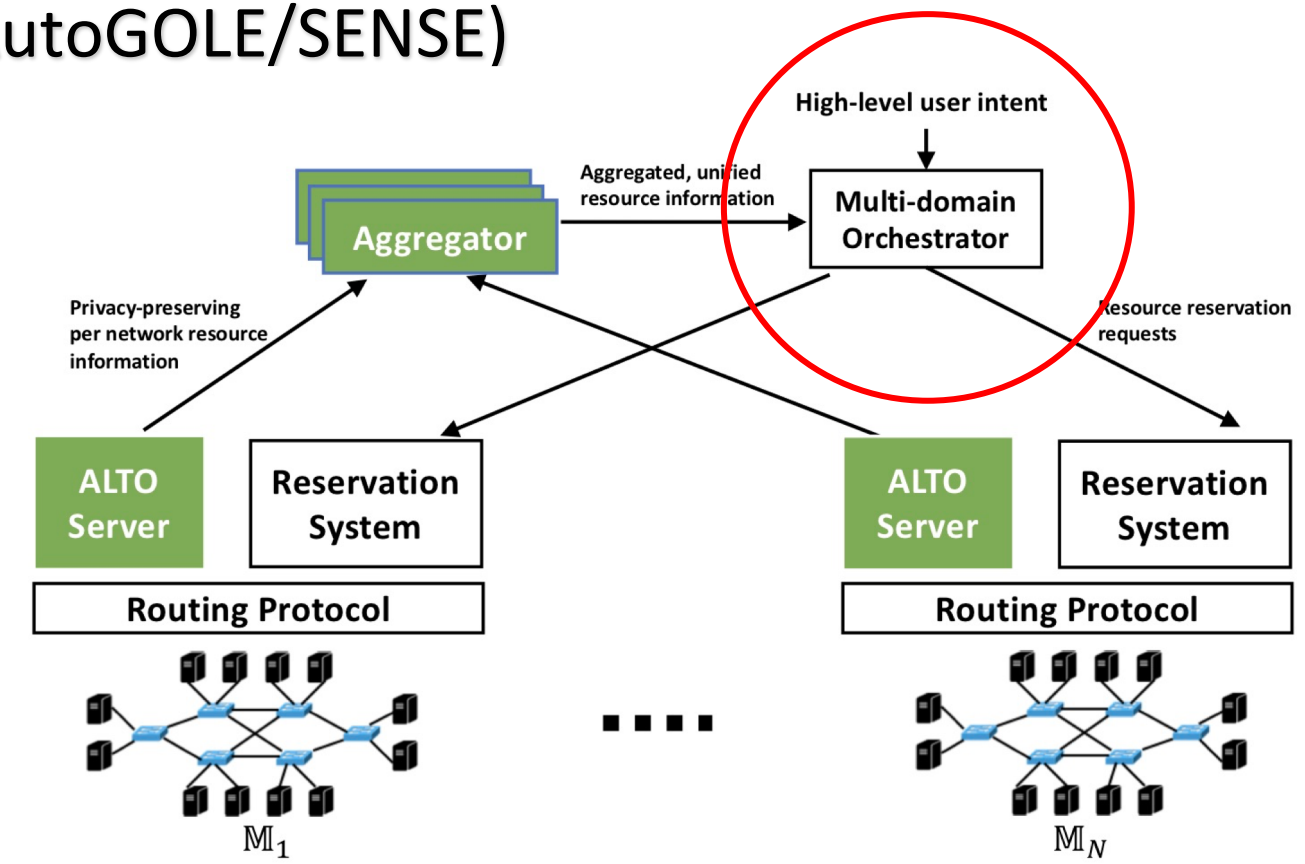


- Node X has 6 potential sources, Node [1-3]A, Node [1-3]B
- Sources span multiple domains
- Which distance/ranking should Node X receive?



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





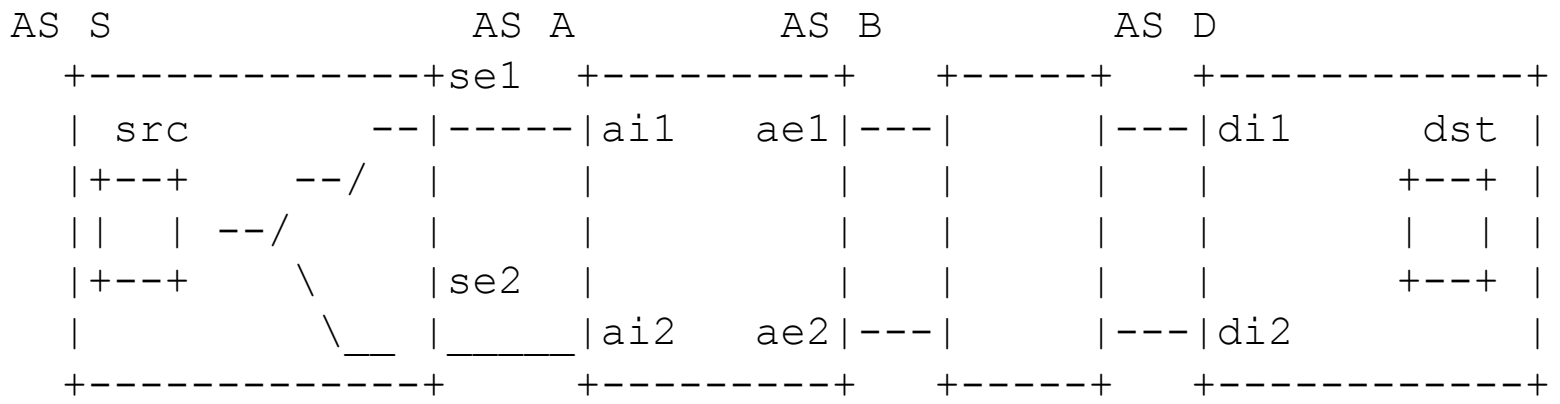
- Large-scale data analytics jobs span multiple networks
- Compute QoS (possible bandwidth) to optimize co-flow finishing time

"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 Jordi Ros Giralt work

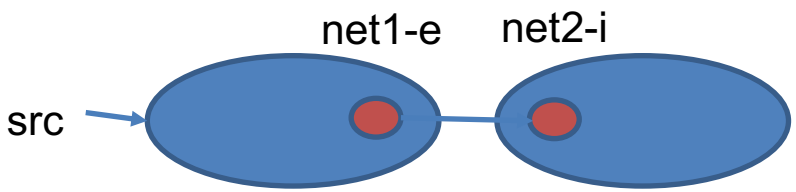
Gap in Current ALTO

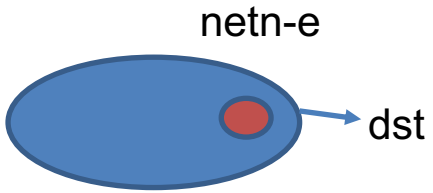


- Missing standard protocol to stitch information across domains
  - Take computing cost/distance src->dst as an example
    - AS S has complete path property, but only for BGP path
    - AS S and AS D can provide only GeoIP
    - AS A/B in the middle can provide only path segments
      - Assume it can detect ingress point

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
    - src -> net_1-e -> net_2-i -> … -> net_i-e -> net_{i+1}-i -> … -> net_n-e -> dst
      - 0-domain-hop: segment-src-dst
      - 1-domain-hop: segment-src -> segment-dst





    - Domain segments obtained from BGP at source => bootstrapping starts at source

ALTO Extensions to Realize AMDA

- Ext 1: Segment discovery [implemented]
  - <flow, netid:ingress>
    ->
    <netid:egress, netid:next-ingress; [Sebastian proposal: next-alto-server-uri; handle blackhole…]>

- Ext 2: Whole path cost compute
  - Indicate <flow, netid:ingress> to segment ALTO server
  - Operation models for extensions [mechanisms, not policies]
    - iterative (client aggregation)
    - recursive (network helped aggregation)
    - hybrid

Important Technical Detail: Incremental Deployment

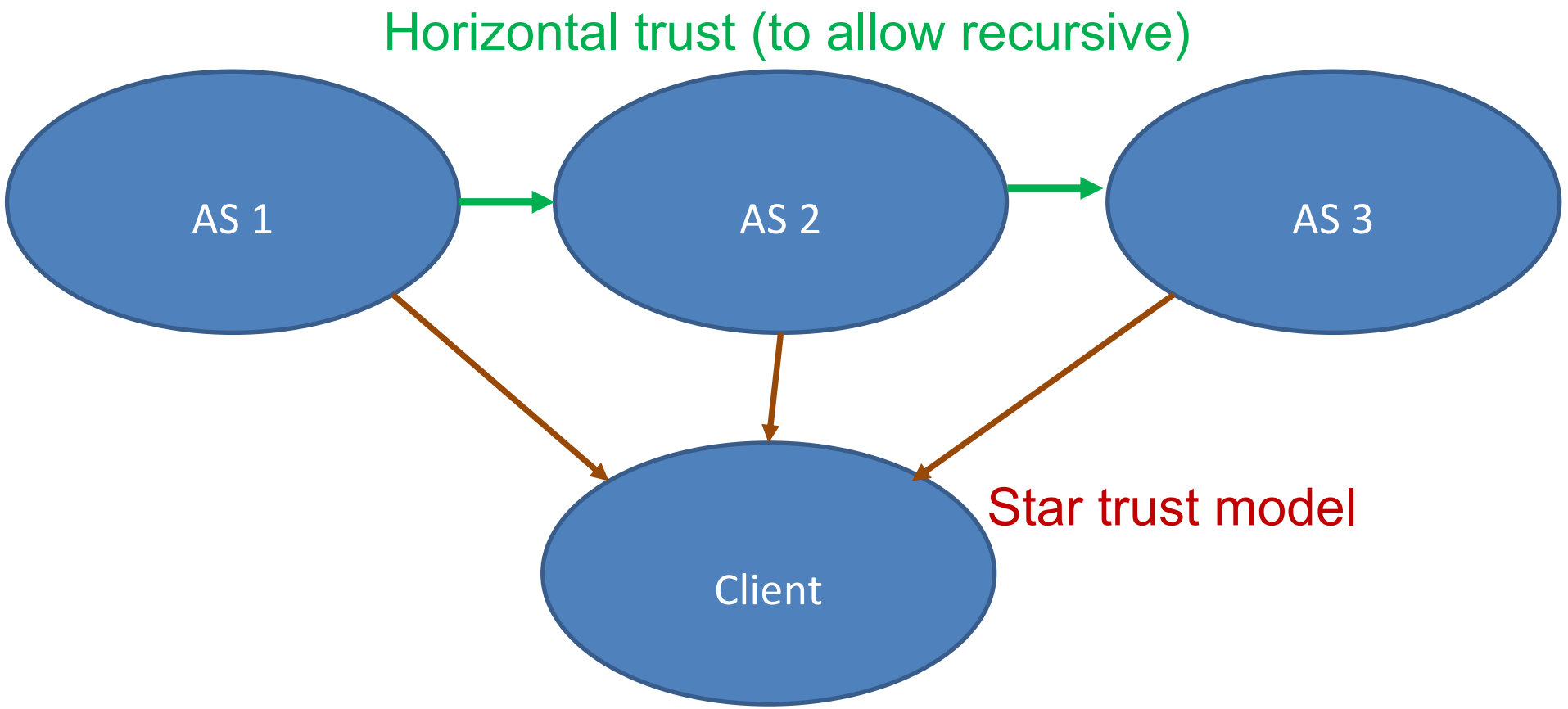
- Incremental deployment: the chaining of domains may be broken due to incremental deployment (e.g., domain sequence is S -> A -> B -> C -> D, but A does not provide AMDA)
  - Provide guidance on
    - how to detect ingress point at B, and
    - how to respond if B cannot detect ingress (multi-answers)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

Important Technical Detail: ADMA Ranking

- A vector of path cost may no longer defines 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

Important Technical Detail: Discuss Trust Model

- Discuss implications of revealing egress points
- Discuss star vs chaining trust models



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/
- 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

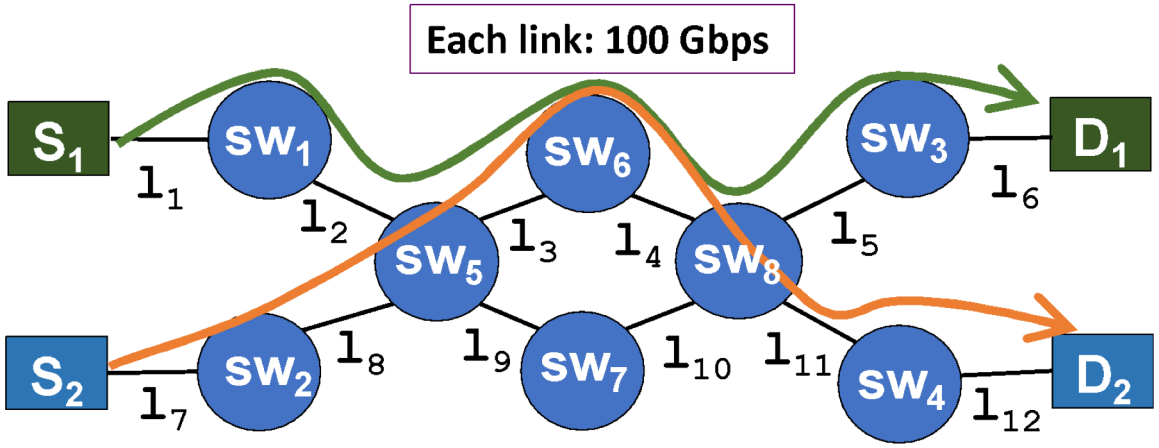
Additional Questions

- 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?

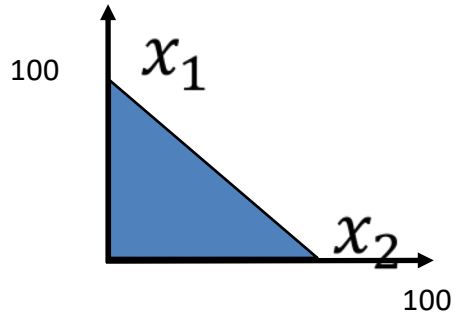
[Sebastian 9/30/2020]

(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**.

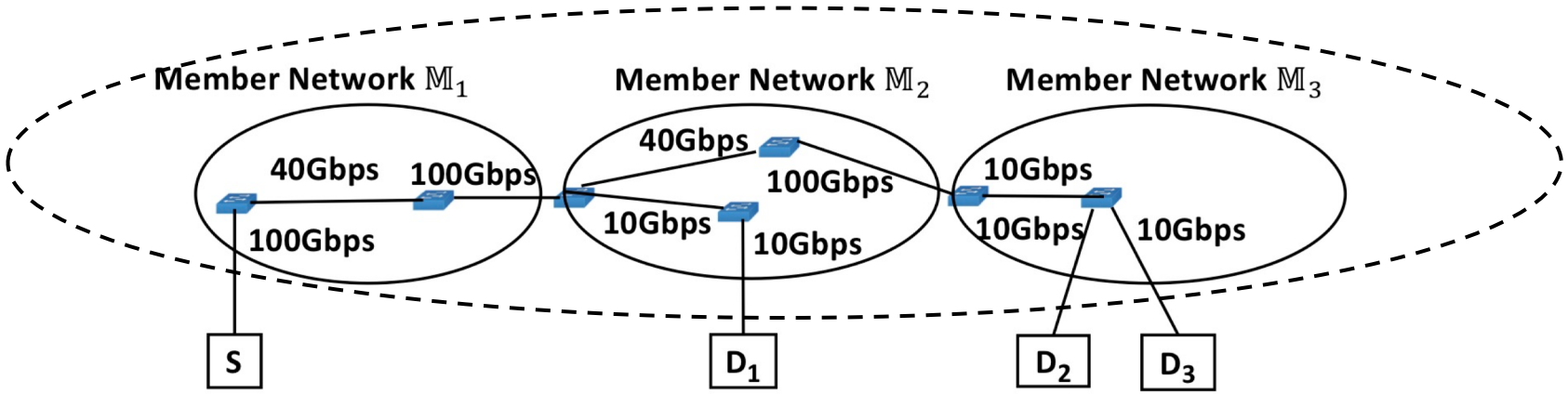


$$x_1 \leq 100 \quad \forall l_u \in \{l_1, l_2, l_5, l_6\},$$
$$x_2 \leq 100 \quad \forall l_u \in \{l_7, l_8, l_{11}, l_{12}\},$$
$$x_1 + x_2 \leq 100 \quad \forall l_u \in \{l_3, l_4\},$$



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

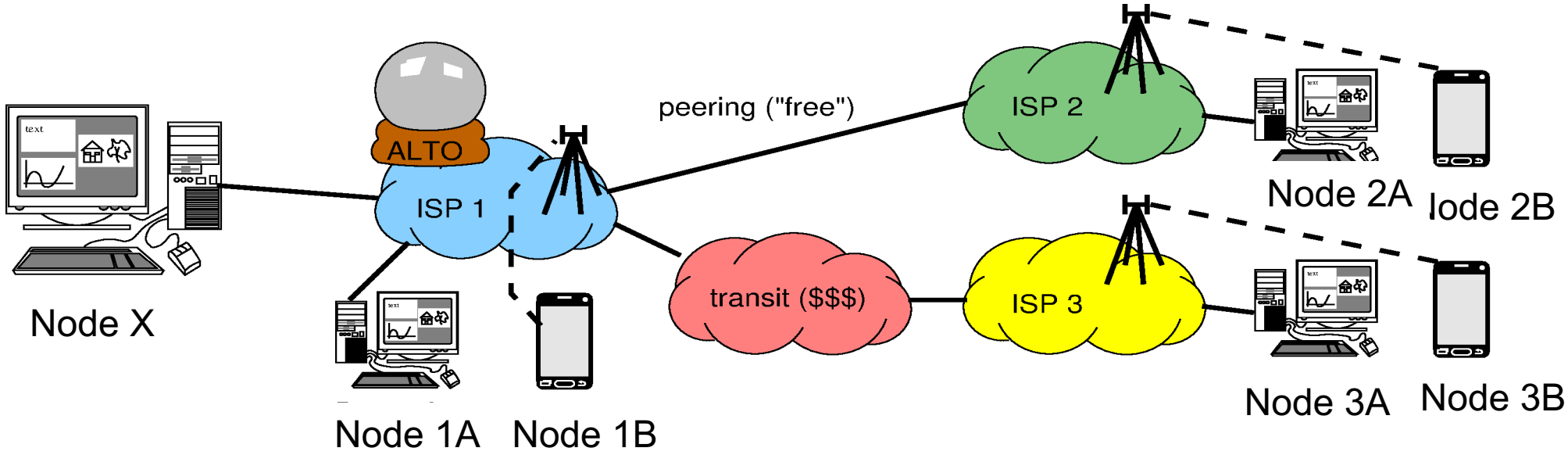
The Reverse View: Mathematical Constraints as Virtual Network Representation



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

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

(*) = ?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.

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