Multi-Cost ALTO

Updates in

draft-randriamasy-alto-multi-cost-03

S. Randriamasy, N. Schwan

Former discussions & proposals

- IETF'80 at Praha: Should the ALTO protocol allow multiple costs in responses?
  - Motivation for Multi-Cost ALTO transactions
    - Gain time, save resources
  - Previous drafts also proposed:
    - additional Cost Types, EP properties, attributes for time-sensitive costs

- ALTO WG discussions:
  ➔ no objections for multi-cost transactions. Protocol extensions need to be specified

- This draft initiates the next steps to this work
  - ALTO protocol extensions to support Multi-Cost transactions
  - Uses cases motivating additional EP properties and Cost Types

Impacted ALTO features

- Some new features
  - Introduction of arrays of N supported Cost Type IDs
  - Multi-Cost service on EPs + PIDs
  - EP cover: Peer, CDN storage location, party in grid computing or on-line gaming or other resources sharing applications.
  - Command in request: GET ➔ POST
  - Assumption: EP properties have constant values, EP costs MAY vary
  - Costs can be time sensitive and need appropriate attributes
- Rule:
  - when multiple cost types are requested, then the requested Cost Mode SHOULD be numerical
  - Reason: requests for multiple ordinal or numerical costs too complex to handle and btw nonsense

Impacted ALTO services

- First sketch of Multi-Cost transaction
  - EP Cost Lookup service
- Next updates
  - Information Resources Directory
    - Explicit or implicit limit on N in requests?
  - Cost Map Service
  - Cost Map Filtering Service
  - EP Cost Lookup service
- Updates will consider discussions on mailing list

Multi-Cost ALTO scenario and use cases

- Delay sensitive overlay applications
  - Real-time sensitive applications such as gaming, interactive video conferencing, medical services
    - need overlay topology with minimal delay
    - Accessible by e2e measurements but cumbersome

    ➔ALTO could provide statistical guidance on path w.r.t. delay, possibly combined with ‘routingcost’ or other Cost Type

Multi-Cost ALTO scenario and use cases

- CDN surrogate/cache selection
  - See draft «Use Cases for ALTO within CDNs»
  - To which cache to send a content request?
    - Request routing strategies prefer caches closer to the Client.
    - Cache nodes tend to be placed deeper in the network
  - Surrogate selection would benefit from criteria such as ‘routingcost’ PLUS
    - Criteria that help for load balancing
      - Statistical CPU utilization, EP memory usage, path bandwidth…

Thank you

- Next slides
  - based on former draft iterations

Properties & Costs information needed for IANA registry

- ID
- Intended semantics
  - Units ∈ {‘units’, ‘msecs’, ‘Mbytes’, ‘%’, … }
    - The unit ‘unit’ applies to ordinal values or generic values as for ‘routingcost’
  - Mode ∈ {‘numerical’, ‘ordinal’, other?}
    - Should we have an attribute indicating the presence of time-related attributes?
  - Time attributes ∈ {timeframe, lifetime, expiration time}
  - Optimal value ∈ {MIN, MAX}
- Security considerations
  - A property or cost is either ‘public’ or ‘provider confidential’
    - ‘routingcost’ MUST be public
    - Other P&C MAY be tagged as ‘provider confidential’ by the acting ALTO service management.

More on proposed P&C attributes

- Time scope attributes of ALTO P&C
  - Purpose: to reflect moderate variations and/or periodicity
  - Timeframe
    - Indicates on which duration statistics are made
    - DEFAULT: infinite
  - Lifetime: validity interval of information
    - E.g. Hourly statistics valid between 8 am and 7 pm
  - Validity period:
    - E.g. « working day » hourly stats expire on Friday at 7 pm
  - Age: date of last information update

- Example:
  - 'endpointloadcost'
    - Mode = stat:median ➔ whether 'statistic' is a mode needs to be discussed
    - Timeframe = 60 min
    - Lifetime = [8 am – 7 pm]
    - Age = YYYY/MM/DD HH/MN/??
    - Validity period: until Friday at 7 pm

Interaction Multi-Cost ALTO & EP selection functions

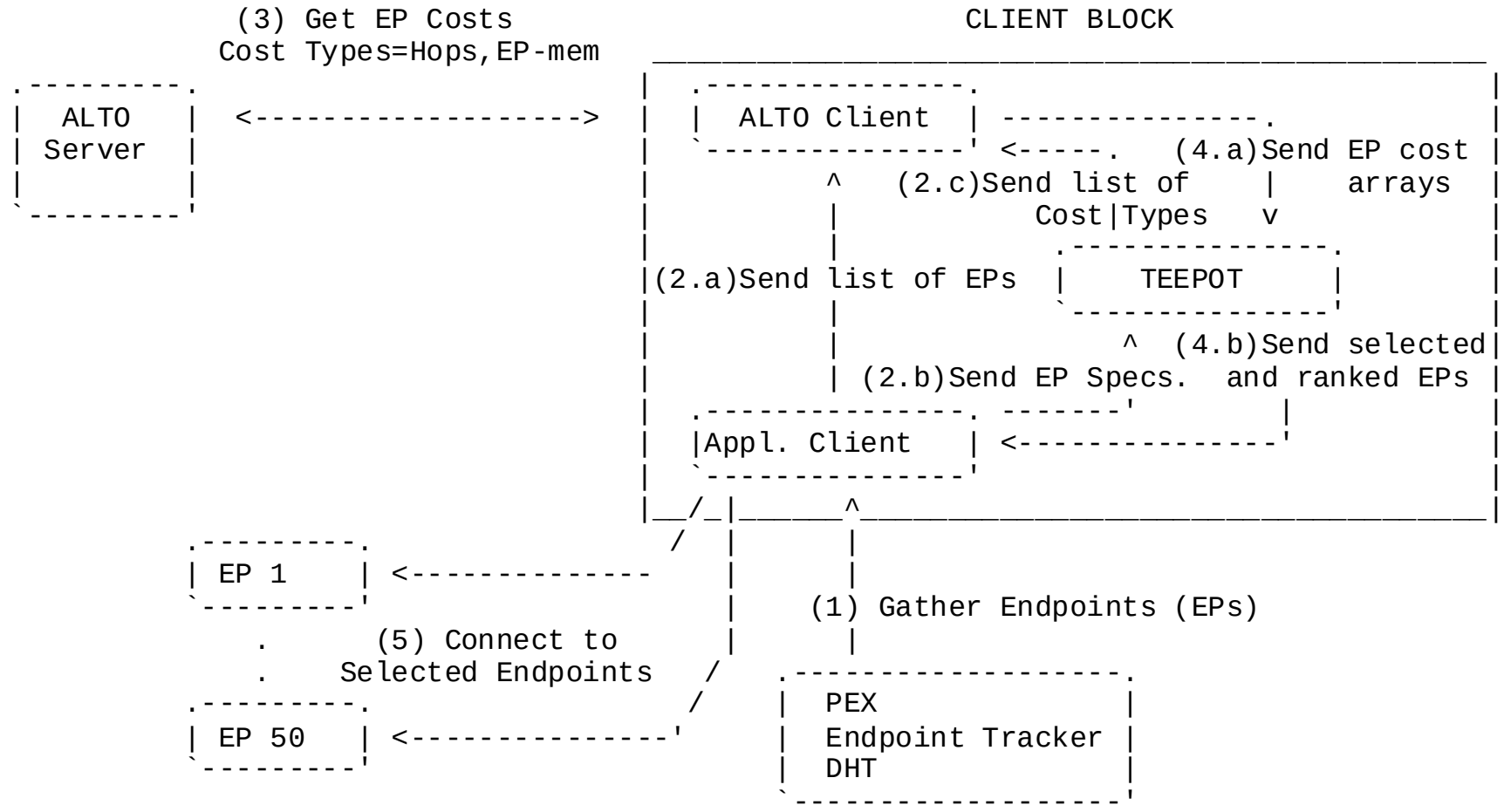

Figure 2: Multi-Cost ALTO scenario and transactions in a client block