Multi-Cost ALTO

draft-randriamasy-alto-multi-cost-02

Updates

S. Randriamasy

Multi-Cost ALTO proposal updates

- IETF79 draft stresses the need to
  - Save transaction time
  - Choose Endpoints with application-specific characteristics
  - According to traffic conditions with a moderate time variation
    - Not available otherwise
- This presentation focuses on
  - Draft updates according to discussions with ALTO list members
  - Additional considerations
- ISP preferences are reflected in ALTO by a ‘routingcost’
  - The Aplication Client (AC) is not supposed to know its nature
    - In particular which ISP constraints and metrics it involves
  - Other proposed metrics in ALTO tend to be static
- Discussion items
  - Should the ALTO protocol allow multiple costs in responses?
  - Proposed extension for multiple costs and proprieties?

Multi-Cost ALTO context

- The Application Client (AC) knows what it needs for the application, e.g.
  - Enough EP memory capacity
  - Particular access or path capacity characteristics
    - E.g. long delay & no losses, or short delay with moderate losses…
  - Several paths ➔ several EPs from which to download
- The ISP network knows about
  - The connection type of EPs (wireless/fixed)
  - The rush period for paths and endpoints
  - The proximity of EPs to the AC home network
  - The topology and size of its network
    - Important for small networks experiencing flash crowds

Multi-Cost specific ALTO extensions

- Diffs since IETF 79
- All Costs Types in one single ALTO transaction
- Extension of attribute Cost Type
  - From a single value to a vector/array of N >= 1 values
- Definition of ID for each Cost Type
  - Automatical mapping of Cost Type ID to Cost Vector components
- Update of the ALTO Server Capability Service
  - To integrate the ID of supported costs
- **Rule**:
  - when multiple cost types are requested, then the requested Cost Mode MUST be numerical

Impacted ALTO services and features

- Assuming that
  - EP properties have constant values and EP costs MAY vary

- EP Properties
  += memory capacity, nominal bandwidth, access technology
- Endpoint (EP) Cost
  - Cost attribute Type
    += EP load, EP path load, EP availability
  - Cost attribute Mode
    += statistical:
    - statistics need to be provided with time scope information that are proposed in next slides
- ALTO service description
  += supported Cost Type ID vector
- Cost Map between Network Locations
- Cost Map filtering

Proposed Time Scope attributes

- Time scope attributes of ALTO P&C
  - Purpose: to reflect predictable/periodical variations (rush hour, event)
  - 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
    - DEFAULT: permanent
  - Age: time elapsed since last update
  - Are a complement to 'Expiration Time' proposed in § 8.1.2
    - E.g. « working day » hourly stats expire on Friday at 7 pm
- Example: 'endpointloadcost' (provided by content network provider)
    - Mode = stat:median ➔ whether 'statistic' is a mode needs to be discussed
    - Timeframe = 60 min
    - Lifetime = [8 am – 7 pm]
    - Age = time elapsed – e.g. 3 weeks
    - Expiration time: Friday at 7 pm
- Time scope attributes should be applicable to all Costs Types

Proposed Time Scope attributes

- Allows ALTO Client to schedule its requests for updated ALTO information
- Allows Application Client to schedule its downloads
- Given that not all Cost Types have the same time scope
- If provided by ALTO services, saves the time for AC to do the measurements
- Use case:
  - An AC wants to download
    - N files with no losses and whatever delay
    - P files with minimal latency and moderate losses
  - AC home network has poor / intermittent infrastructure
  - AC has limited / intermittent access
  - AC has no time/means to measure EP performances
  - The ISP of such networks has interests in maximising the number of sucessful and resources efficient downloads

Properties & Costs attributes for IANA registry

- ID
- Intended semantics
  - Units: in {‘units’, ‘msecs’, ‘Mbytes’, ‘%’, ‘strings’… }
    - The unit ‘unit’ applies to ordinal values or generic values as for ‘routingcost’
  - Mode: in {‘numerical’, ‘ordinal’, ‘statistical’}
  - Time attributes: in {timeframe, lifetime, expiration time}
  - Optimum: in {MIN, MAX} i.e. best value(s) equal/closer to optimum
- Security considerations
  - A property or cost is either ‘public’ or ‘provider confidential’ (see REQs)
    - ‘routingcost’ MUST be public
    - Other P&C MAY be tagged as ‘provider confidential’ by the acting ALTO service management.

Multi-Cost ALTO scenario and applications

- ALTO protocol options to keep considering ISP preferences: examples
  - Request a significant weight for the ISP established ‘routingcost’
  - Among several EPs with equivalent ‘routingcost’
    - Allow to choose EPs with particular values on other costs
  - If ‘routingcost’ already involves queried metrics then
    - The ISP may simply not offer the service on involved cost types

- The Application Client (AC)
  - Collects Endpoint information via the ALTO Services or other techniques
  - Decides how to weight the decision metrics and organize the gathered information to choose the EPs
- Multi-Cost ALTO is applicable to
  - CDN
  - Trackerless P2P
  - Tracker-based P2P,
    - that then should « fairly » integrate the EP costs with the contribution level of peers.

Illustrative ALTO scenario

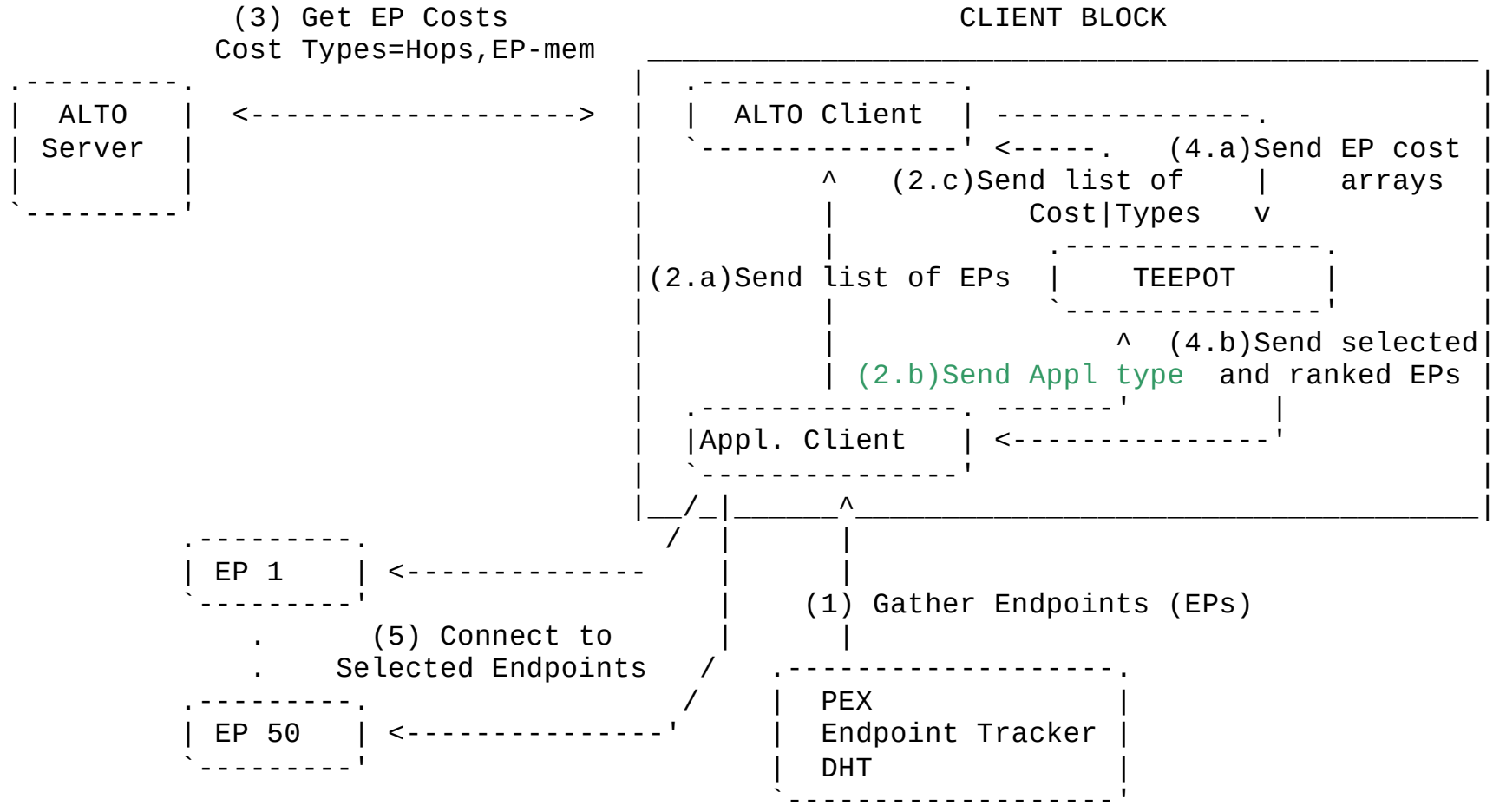

Figure 2: Multi-Cost ALTO scenario and transactions in a client block showing how multiple costs are requested and processed

Main discussion points

- Should the ALTO protocol allow multiple costs in responses?
- Proposed extension for multiple costs and proprieties
  - New Cost Types
  - Time scope attributes
  - Should ‘statistical’ be a mode or should the Cost timeframe hint this?
- Other reported suggestions during discussions
  - Modes: += ‘string’

Multi-Cost ALTO updates

Thank you

REQS and Multi-Cost ALTO

- REQ. ARv07-14: « The ALTO client protocol MUST support the usage of several different rating criteria types ».
- REQ ARv07-17: « The ALTO Client protocol specification MUST define an appropriate procedure for adding new rating criteria types e.g. by esablishing an IANA registry »
- REQ. ARv07-29: « The ALTO Client protocol SHOULD support lifetime attributes to enable the caching of recommendations at ALTO Clients»
- REQ. ARv07-30: « The ALTO Client protocol SHOULD specify an aging mechanism which allows to give newer recommendations precedence over older ones»
  - « Long » (TBD) term statistics or empirical ratings on performance oriented information may still be useful for a reliable choice of candidate endpoints.
  - Specific «short term» ALTO services can be specified for mobile core networks, which have a smaller scale and can afford and take advantage of using network information at a smaller time-scale

Other consideration

- Vector costs naturally provide several efficient solutions
  - Which are numerically robust and consistent
  - A set of efficient EPs provides a valuable input to Sefficient multi-path connections