Network Working Group
Internet-Draft
Intended status: Standards Track
Expires: August 17, 2017

S. Randriamasy
Nokia Bell Labs
R. Yang
Yale University
Q. Wu
Huawei
L. Deng
China Mobile
N. Schwan
Thales Deutschland

February 13, 2017

ALTO Cost Calendar
draft-ietf-alto-cost-calendar-01

Abstract

The goal of Application-Layer Traffic Optimization (ALTO) is to bridge the gap between network and applications by provisioning network related information in order to allow applications to make network informed decisions. The present draft extends the ALTO cost information so as to broaden the decision possibilities of applications to not only decide 'where' to connect to, but also 'when'. This is useful to applications that need to schedule their data transfers and connections and have a degree of freedom to do so. ALTO guidance to schedule application traffic can also efficiently help for load balancing and resources efficiency. Besides, the ALTO Cost Calendar allows to schedule the ALTO requests themselves and thus to save a number of ALTO transactions.

This draft proposes new capabilities and attributes on filtered cost maps and endpoint costs enabling an ALTO Server to provide "Cost Calendars". These capabilities are applicable to time-sensitive ALTO metrics. With ALTO Cost Calendars, an ALTO Server exposes ALTO Cost Values in JSON arrays where each value corresponds to a given time interval. The time intervals as well as other Calendar attributes are specified in the IRD and ALTO Server responses.

Requirements Language

The key words "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT", "SHOULD", "SHOULD NOT", "RECOMMENDED", "MAY", and "OPTIONAL" in this document are to be interpreted as described in RFC 2119 [RFC2119].

Status of This Memo

Copyright Notice

Table of Contents

1. Introduction

IETF is currently standardizing the ALTO protocol which aims at providing guidance to overlay applications needing to select one or several hosts from a set of candidates able to provide a desired resource. This guidance is based on parameters that affect performance and efficiency of the data transmission between the hosts such as the topological distance. The goal of ALTO is to improve the Quality of Experience (QoE) in the application while optimizing resource usage in the underlying network infrastructure.

The ALTO protocol in [RFC7285] specifies a Network Map which defines groupings of endpoints in provider-defined network regions (called PIDs). The Cost Map Service, Endpoint Cost Service (ECS) and Endpoint Ranking Service then provide ISP-defined costs and rankings for connections among the specified endpoints and PIDs and thus incentives for application clients to connect to ISP preferred locations, e.g. to reduce their costs. ALTO intentionally avoids provisioning realtime information as explained in the ALTO Problem Statement [RFC5693] and ALTO Requirements [RFC5693].Thus the current

Cost Map and Endpoint Cost Service are providing, for a given Cost Type, exactly one path cost value. Applications have to query one of these two services to retrieve the currently valid cost values. They therefore need to plan their ALTO information requests according to their own estimation of the frequency of cost value change.

With [RFC7285], an ALTO client should interpret the returned costs as those at the query moment. However, Network costs can fluctuate, e.g. due to diurnal patterns of traffic demand or planned events such as network maintenance, holidays or highly publicized events. Providing network costs for only the current time thus may not be sufficient, in particular for applications that can schedule their traffic in a span of time, for example by deferring backup to night during traffic trough.

In case the ALTO Cost value changes are predicable over a certain period of time and the application does not require immediate data transfer, it can save time to get the whole set of cost values over this period in one ALTO response. Using them to schedule data transfers allows optimising the network resources usage and QoE. ALTO Clients and Servers can also minimize their workload by accordingly scheduling their data exchanges.

This document extends RFC7285 to allow an ALTO server to provide network costs for a given duration of time. A sequence of network costs across a time span for a given pair of network locations is named an "ALTO Cost Calendar". The Filtered Cost Map Service and Endpoint Cost Service are extended to provide Cost Calendars. In addition to this functional ALTO enhancement, we expect to further gain on storage and on the wire data exchange by gathering multiple Cost Values for one Cost Type into one single ALTO Server response.

In this draft an "ALTO Cost Calendar" is specified by information resources capabilities that are applicable to time-sensitive ALTO metrics. An ALTO Cost Calendar exposes ALTO Cost Values in JSON arrays where each value corresponds to a given time interval. The time intervals as well as other Calendar attributes are specified in the IRD and in the Server response to allow the ALTO Client to interpret the received ALTO values. Last, the proposed extensions for ALTO Calendars are applicable to any Cost Mode and they ensure backwards compatibility with legacy ALTO clients.

In the rest of this document, Section 2 provides the design characteristics. Sections 3 and 4 define the formal specifications for the IRD and the information resources. Section 5 provides non-normative use cases to illustrate the usage of cost calendars. IANA considerations and security considerations will be completed in further versions.

2. Overview of ALTO Cost Calendars

An ALTO Cost calendar provided by the ALTO Server provides 2 information items:

- an array of values for a given metric, where each value corresponds to a time interval, where the value array can sometimes be a cyclic pattern that repeats a certain number of times.

- attributes describing the time scope of the calendar such as the size and number of the intervals and the date of the starting point of the calendar, allowing an ALTO Client to properly interpret the values.

An ALTO Cost Calendar can be used like a "time table" to figure out the best time to schedule data transfers and also to proactively manage application traffic given predictable events such as flash crowds, traffic intensive holidays and network maintenance. It may be viewed as a synthetic abstraction of real measurements that can be historic or be a prediction for upcoming time periods.

Most likely, the ALTO Cost Calendar would be used for the Endpoint Cost Service, assuming that a limited set of feasible Endpoints for a non-real time application is already identified, that they do not need to be accessed immediately and that their access can be scheduled within a given time period. The Filtered Cost Map service is also applicable as long as the size of the Map allows it.

2.1. ALTO Cost Calendar information features

The Calendar attributes are provided in the IRD and in ALTO Server responses. The IRD announces attributes with dateless values in its information resources capabilities, where as attributes with time dependent values are provided in the "meta" of Server responses. The ALTO Cost Calendar attributes provide the following information:

- attributes to interpret the time scope of the Calendar value array:

  * generic time zone,

  * applicable time interval for each calendar value: combining numbers and time units to reflect for example: 1 hour, 2 minutes, 10 seconds, 1 week, 1 month,

  * duration of the Calendar: e.g. the number of intervals provided in the calendar.

- "calendar-start-date": specifying when the calendar starts, that is to which date the first value of the cost calendar is applicable.

- "repeated": an optional attribute indicating for how many iterations the provided calendar will have the same values. The server may use it to allow the client to schedule its next request and thus save its own workload by avoiding to process useless requests.

2.2. ALTO Calendar design characteristics

The protocol extension placeholders for an ALTO Calendar are: the IRD, the ALTO requests and responses for Cost calendars.

Extensions are designed to be light and ensure backwards compatibility with base protocol ALTO Clients and with other extensions. It uses section 8.3.7 "Parsing of Unknown Fields" of RFC7285 that writes: "Extensions may include additional fields within JSON objects defined in this document. ALTO implementations MUST ignore unknown fields when processing ALTO messages."

The calendar-specific capabilities are integrated in the information resources of the IRD and in the "meta" member of ALTO responses to Cost Calendars requests. A calendar and its capabilities are associated to a given information resource and within this information resource to a given cost type. This design has several advantages:

- it does not introduce a new mode,

- it does not introduce new media types,

- it allows an ALTO Server to offer calendar capabilities on a cost type, with attributes values adapted to each information resource.

The Applicable Calendared information resources are:

- the Filtered Cost Map,

- the Endpoint Cost Map.

The ALTO Server can choose in which frequency it provides cost Calendars to ALTO Clients. It may either provide calendar updates starting at the request date, or carefully schedule its updates so as to take profit from a potential repetition/periodicity of calendar values.

2.2.1. ALTO Cost Calendar for all cost modes

Calendars are well-suited for values encoded in the 'numerical' mode. However, Calendars can also represent any metric considered as time-sensitive by an ALTO Server. For example, types of Cost values such as JSONBool can also be expressed as calendars, as states may be "true" or "false" depending on given time periods or likewise, values represented by strings, such as "medium", "high", "low", "blue", "open" .

Note also that a Calendar is applicable as well to time-sensitive metrics provided in the 'ordinal' mode, if these values are time-sentitive and their update is carefully managed by the ALTO Server.

2.2.2. Compatibility with legacy ALTO Clients

The ALTO protocol extensions for Cost Calendars have been defined so as to ensure that Calendar capable ALTO Servers can provide legacy ALTO Clients with legacy information resources as well. That is a legacy ALTO Client can request resources and receive responses as specified in RFC7285.

For compatibility with legacy ALTO Clients specified in RFC7285, calendared information resources are not applicable for full Cost Maps for the following reason: a legacy ALTO client would receive a Calendared Cost Map via an HTTP 'GET' command. As specified in section 8.3.7 of RFC7285, it will ignore the Calendar Attributes indicated in the "meta" of the responses. Therefore, lacking information on calendar attributes, it will not be able to correctly interpret and process the values of the received array of calendar cost values.

3. ALTO Calendar specification: IRD extensions

The Calendar attributes in the IRD information resources capabilities carry constant dateless values. A calendar is associated to an information resource rather than a cost type. For example, a Server can provide a "routingcost" calendar for the Filtered Cost Map Service at a granularity of one day and a "routingcost" calendar for the Endpoint Cost service at a finer granularity but for a limited number of endpoints.

3.1. Calendar attributes in the IRD resources capabilities

When for an applicable resource , an ALTO Server provides a Cost Calendar for a given Cost Type, it MUST indicate this in the IRD capabilities of this resource, by an object of type

'CalendarAttributes', associated to this Cost Type and specified below.

The capabilities of a Calendar aware information resource entry have a member named "calendar-attributes" which is an array of objects of type CalendarAttributes. It is necessary to use an array because of resources such as Filtered Cost Map and Endpoint Cost Map, for which the member "cost-type-names" is an array of 1 or more values.

RULE: a member "calendar-attributes" MUST appear only once for each applicable cost type name of a resource entry. If "calendar-attributes" are specified several times for a same "cost-type-name" in the capabilities of a resource entry, the ALTO client SHOULD ignore any calendar capabilities on this "cost-type-name" for this ressource entry.

```
CalendarAttributes calendar-attributes <1..*>;

object{
  [JSONString   cost-type-name;]
  JSONString    time-interval-size;
  JSONNumber    number-of-intervals;
} CalendarAttributes;
```

o "cost-type-name":

  * an optional member indicating the cost-type-name in the IRD entry to which the capabilities apply. If this not present, it MUST be assumed to correspond to its index in the "cost-type-names" list of the IRD resource entry.

o "time-interval-size":

  * is the duration of an ALTO calendar time interval, expressed as a time unit appended to the number of these units. The time unit, ranges from "second" to "year". The number is encoded with an integer. Example values are: "5 minute" , "2 hour", meaning that each calendar value applies on a time interval that lasts respectively 5 minutes and 2 hours.

o "number-of-intervals":

  * the integer number of values of the cost calendar array, at least equal to 1.

- Attribute "cost-type-name" , if used, provides a better readability to the calendar attributes specified in the IRD and avoids confusion with calendar attributes of other cost-types.

- Multiplying Attributes 'time-interval-size' and 'number-of-intervals' provides the duration of the provided calendar. For example an ALTO Server may provide a calendar for ALTO values changing every 'time-interval-size' equal to 5 minutes. If 'number-of-intervals' has the value 12, then the duration of the provided calendar is "1 hour".

3.2. Calendars in a delegate IRD

One option to clarify IRD resources is that a "root" ALTO Server implementing base protocol resources delegates "specialized" information resources such as the ones providing Cost Calendars to another ALTO Server running in a subdomain specified with its URI in the "root" ALTO Server. This option is described in Section 9.2.4 "Delegation using IRDs" of RFC7285.

This document provides an example, where a "root" ALTO Server runs in a domain called "alto.example.com". It delegates the announcement of Calendars capabilities to an ALTO Server running in a subdomain called "custom.alto.example.com". The location of the "delegate Calendar IRD" is assumed to be indicated in the "root" IRD by the resource entry: "custom-calendared-resources".

Another advantage is that some Cost Types for some resources may be more advantageous as Cost Calendars and it makes few sense to get them as a single value. For example, Cost Types with predictable and frequently changing values, calendared in short time intervals such as a minute.

3.3. Example IRD with ALTO Cost Calendars

The cost types in this example are either specified in the base ALTO protocol or may be proposed in other drafts see [draft-ietf-alto-performance-metrics]. In this example, the available cost metrics are indicated in the "meta" field by cost type names "num-routingcost", "num-latency", "num-pathbandwidth" and "string-quality-status". Metrics "routingcost" , 'latency' and 'Availbandwidth' are available in the "numerical" Cost Mode. Metric "quality-status" is available in the "string" Cost Mode.

The example IRD includes 2 particular URIs providing calendars:

- "http://custom.alto.example.com/calendar/costmap/filtered": a filtered cost map in which calendar capabilities are indicated for cost type names: "num-routingcost", "num-pathbandwidth" and "string-service-status",

- "http://custom.alto.example.com/calendar/endpointcost/lookup": an endpoint cost map in which in which calendar capabilities are indicated for cost type names: "num-routingcost", "num-latency", "num-pathbandwidth", "string-service-status".

The design of the Calendar capabilities allows that some calendars on a cost type name are available in several information resources with different Calendar Attributes. This is the case for calendars on "num-routingcost", "num-pathbandwidth" and "string-service-status" , available in both the Filtered Cost map and Endpoint Cost map service, but with different time interval sizes for "num-pathbandwidth" and "string-service-status".

```
GET /calendars-directory HTTP/1.1
Host: custom.alto.example.com
Accept: application/alto-directory+json,application/alto-error+json
---------------

HTTP/1.1 200 OK
Content-Length: [TODO]
Content-Type: application/alto-directory+json

   {
     "meta" : {
        "cost-types": {
           "num-routingcost": {
              "cost-mode" : "numerical",
              "cost-metric" : "routingcost"
           },
           "num-latency": {
              "cost-mode"  : "numerical",
              "cost-metric": "latency"
           },
           "num-pathbandwidth": {
              "cost-mode"  : "numerical",
              "cost-metric": "Availbandwidth",
           },
           "string-qual-status": {
              "cost-mode"  : "string",
              "cost-metric": "quality-status",
           }
         ... other meta ...
       },

    "resources" : {
       "filtered-cost-map-calendar" : {
           "uri" : "http://custom.alto.example.com/calendar/costmap/filtered",
           "media-type" : "application/alto-costmap+json",
```

```
           "accepts" : "application/alto-costmapfilter+json",
           "capabilities" : {
              "cost-constraints" : true,
              "cost-type-names"  : [ "num-routingcost", "num-pathbandwidth",
                                     "string-service-status" ],
              "calendar-attributes" : [
                 {"cost-type-names" : [ "num-routingcost", "num-
pathbandwidth" ],
                  "time-interval-size" : "1 hour",
                  "number-of-intervals" : 24
                 },
                 {"cost-type-names" : "string-service-status",
                  "time-interval-size" : "30 minute",
                  "number-of-intervals" : 48
                 }
               ] // end calendar-attributes
              "uses": [ "my-default-network-map" ]
              }
           },

       "endpoint-cost-calendar-map" : {
         "uri" : "http://custom.alto.example.com/calendar/endpointcost/lookup",
         "media-types" : [ "application/alto-endpointcost+json" ],
         "accepts" : [ "application/alto-endpointcostparams+json" ],
         "capabilities" : {
           "cost-constraints" : true,
           "cost-type-names" : [ "num-routingcost", "num-latency",
                                 "num-pathbandwidth", "string-service-status" ],
           "calendar-attributes" : [
              {"cost-type-names" : "num-routingcost",
               "time-interval-size" : "1 hour",
               "number-of-intervals" : 24
              },
              {"cost-type-names" : "latency",
               "time-interval-size" : "5 minute",
               "number-of-intervals" : 12
              },
              {"cost-type-names" : "num-pathbandwidth",
               "time-interval-size" : "1 minute",
               "number-of-intervals" : 60
              },
              {"cost-type-names" : "string-service-status",
               "time-interval-size" : "2 minute",
               "number-of-intervals" : 30
              }
             ]
           "uses": [ "my-default-network-map" ]
         } // ECM capab
```

```
      } //info resource N
```

```
} // ressources
```

In this example IRD, for the filtered cost map service:

- the Calendar for 'num-routingcost' and 'num-pathbandwidth' is an array of 24 values each provided on a time interval lasting 1 hour.

- the Calendar for "string-service-status": "is an array of 48 values each provided on a time interval lasting 30 minutes.

For the endpoint cost map service:

- the Calendar for 'num-routingcost': is an array of 24 values each provided on a time interval lasting 1 hour.

- the Calendar for 'latency': is an array of 12 values each provided on a time interval lasting 5 minutes.

- the Calendar for 'num-pathbandwidth': is an array of 60 values each provided on a time interval lasting 1 minute.

- the Calendar for "string-service-status": "is an array of 30 values each provided on a time interval lasting 2 minutes.

4. ALTO Calendar specification: Service Information Resources

This section documents the individual information resources defined to provide the Calendared information services defined in this document.

The reference time zone for the provided time values is GMT because the option chosen to express the time format is the HTTP header fields format:

```
Date: Tue, 15 Nov 2014 08:12:31 GMT
```

4.1. Calendar extensions for Filtered Cost Maps

A legacy ALTO client requests and gets filtered cost map responses as specified in RFC7285.

4.1.1. Calendar extensions in Filtered cost map requests

The input parameters of a "legacy" request for a filtered cost map, defined by object ReqFilteredCostMap in section 11.3.2 of RFC7285, are augmented with one additional member.

A Calendar-aware ALTO client requesting a Calendar on a given Cost Type for a Filtered Cost Map resource having Calendar capabilities MUST add the following field to its input parameters:

```
JSONBoolean    calendared<1..*>;
```

This field is an array of 1 to N boolean values, where N is the number of requested metrics. Each boolean value indicates whether or not the ALTO Server should provide the values for this Cost Type as a calendar.

This field MUST NOT be specified if member "calendar-attributes" is not present for this information resource.

A Calendar-aware ALTO client supporting single cost type values, as specified in RFC7285, MUST provide an array of 1 element:

```
"calendared" : [true];
```

A Calendar-aware ALTO client that is also Multi-Cost aware MUST provide an array of N values set to "true" or "false", depending whether it wants the applicable Cost Type values as a single or calendared value.

If this field is not present, it MUST be assumed to have only values equal to "false".

4.1.2. Calendar extensions in Filtered Cost map responses

The calendared costs are JSONArrays instead of JSONNumbers for the legacy ALTO implementation. All arrays have a number of values equal to 'number-of-intervals'.

The "meta" field of a Calendared Filtered Cost map response MUST include at least:

- if the ALTO Client supports cost values for one Cost Type at a time only: the "meta" fields specified in RFC 7285 for these information service responses:
  - "dependent-vtags ",

* "cost-type" field.

- if the ALTO Client supports cost values for several Cost Types at a time, as specified in [draft-ietf-alto-multi-cost] : the "meta" fields specified in [draft-ietf-alto-multi-cost] for these information service responses:

  * "dependent-vtags ",

  * "cost-type" field with value set to '{}', for backwards compatibility with RFC7285.

  * "multi-cost-types" field.

In addition, the "meta" field of a Calendared Filtered Cost map response MUST include the member "calendar-response-attributes" for the requested information resource, together with the values provided by the ALTO Server for these attributes. This member is an array of objects of type "CalendarResponseAttributes", defined as follows:

```
CalendarResponseAttributes calendar-response-attributes <1..*>;

object{
  JSONString    calendar-start-time;
  JSONString    time-interval-size;
  JSONNumber    number-of-intervals;
  [JSONNumber   repeated;]            [OPTIONAL]
} CalendarResponseAttributes;
```

- "calendar-start-time": indicates the date at which the first value of the calendar applies. By default, the value provided for the "calendar-start-time" attribute SHOULD be no later than the request date.

- "time-interval-size": as specified in section "Calendar attributes in the IRD resources capabilities",

- "number-of-intervals": as specified in section "Calendar attributes in the IRD resources capabilities",

- "repeated": is an optional field provided for Calendars. It is an integer N greater or equal to '1' that indicates how many iterations of the calendar value array starting at the date indicated by "calendar-start-time" have the same values. The number N includes the provided iteration.

Using the member "repeated" helps minimizing on the wire data exchange: by providing it, an ALTO Server will avoid unecessary

processing of requests for Calendars with unchanged values while it allows ALTO Clients to save their resources as well.

For example: if the "calendar-start-time" member has value "Mon, 30 Jun 2014 at 00:00:00 GMT" and if the value of member "repeated" is equal to 4, it means that the calendar values are the same values on Monday, Tuesday, Wednesday and Thursday. The ALTO Client thus may use the same calendar for the next 4 duration periods following "calendar-start-time".

4.1.3. Use case and example for a FCM with a bandwidth Calendar

An example of non-real time information that can be provisioned in a 'calendar' is the expected path bandwidth. While the transmission rate can be measured in real time by end systems, the operator of a data center is in the position of formulating preferences for given paths, at given time periods for example to avoid traffic peaks due to diurnal usage patterns. In this example, we assume that an ALTO Client requests a bandwidth calendar as specified in the IRD to shedule its bulk data transfers as described in the use cases.

In the example IRD, calendars for cost type name "num-pathbandwidth" are available for the information resources: "filtered-cost-calendar-map" and "endpoint-cost-calendar-map". The ALTO Client requests a calendar for "num-pathbandwidth" via a POST request for a filtered cost map.

We suppose in this example that the ALTO Client sends its request on Tuesday July 1st 2014 at 13:15

```
  POST /calendar/costmap/filtered HTTP/1.1
  Host: alto.example.com
  Content-Length: [TODO]
  Content-Type: application/alto-costmapfilter+json
  Accept: application/alto-costmap+json,application/alto-error+json

  {
    "cost-type" : {"cost-mode" : "numerical", "cost-metric" : "Availbandwidth"},
    "calendared" : [true],

    "pids" : {
       "srcs" : [ "PID1", "PID2" ],
       "dsts" : [ "PID1", "PID2", "PID3" ]
    }
  }


  HTTP/1.1 200 OK
  Content-Length: [TODO]
  Content-Type: application/alto-costmap+json

  {
    "meta" : {
      "dependent-vtags" : [...],
      "cost-type" : {"cost-mode" : "numerical", "cost-metric" :
"Availbandwidth"},
      "calendar-response-attributes" : [
         "calendar-start-time" : Tue, 1 Jul 2014 13:00:00 GMT,
         "time-interval-size" : "2 hour",
         "numb-intervals" : 12
       ]
     },

    "cost-map" : {
       "PID1": { "PID1": [v1,v2, ... v12],
                 "PID2": [v1,v2, ... v12],
                 "PID3": [v1,v2, ... v12] },
       "PID2": { "PID1": [v1,v2, ... v12],
                 "PID2": [v1,v2, ... v12],
                 "PID3": [v1,v2, ... v12] }
    }
  }
```

4.2. Calendar extensions in the Endpoint Cost Map Service

This document extends the Endpoint Cost Service, as defined in {11.5.1} of [RFC7285], by adding new input parameters and capabilities, and by returning JSONArrays instead of JSONNumbers as the cost values. The media type {11.5.1.1} and HTTP method {11.5.1.2} are unchanged.

4.2.1. Calendar specific input in Endpoint cost map requests

The extensions to the requests for calendared Endpoint Cost Maps are the same as for the Filtered Cost Map Service, specified in section XXXX of this draft.

The ReqEndpointCostMap object for a Calendared ECM request will have the following format:

```
object {
  CostType       cost-type;
  [JSONBoolean    calendared<1..*>;]
  EndpointFilter endpoints;
} ReqEndpointCostMap;

object {
  [TypedEndpointAddr srcs<0..*>;]
  [TypedEndpointAddr dsts<0..*>;]
} EndpointFilter;
```

4.2.2. Calendar attributes in the Endpoint Cost Map response

The "meta" field of a Calendared Endpoint Cost map response MUST include at least:

- if the ALTO Client supports cost values for one Cost Type at a time only: the "meta" fields specified in {11.5.1.6} of RFC 7285 for the Endpoint Cost response:
  - "cost-type" field.
- if the ALTO Client supports cost values for several Cost Types at a time, as specified in [draft-ietf-alto-multi-cost] : the "meta" fields specified in [draft-ietf-alto-multi-cost] for the the Endpoint Cost response:
  - "cost-type" field with value set to '{}', for backwards compatibility with RFC7285.

* "multi-cost-types" field.

If the client request does not provide member "calendared" or if it provides it with a value equal to 'false', then the ALTO Server response is exactly as specified in the above cited references.

If the ALTO client provides member "calendared" with a value equal to 'true' in the input parameters, the "meta" member of a Calendared Endpoint Cost Map response MUST include the same addifional member "calendar-response-attributes" as specified for the Filtered Cost Map Service. The Server response is thus changed as follows:

- the "meta" member has one additional field "CalendarResponseAttributes", as specified for the Filtered Cost Map Service,

- the calendared costs are JSONArrays instead of JSONNumbers for the legacy ALTO implementation. All arrays have a number of values equal to 'number-of-intervals'.

4.2.3. Use case and example for the ECS with a routingcost Calendar

Let us assume an Application Client is located in an end sytem with limited resources and having an access to the network that is either intermittent or provides an acceptable quality in limited but predictable time periods. Therefore, it needs to both schedule its resources greedy networking activities and its ALTO transactions.

The Application Client has the choice to trade content or resources with a set of Endpoints and needs to decide with which one it will connect and at what time. For instance, the Endpoints are spread in different time-zones, or have intermittent access. In this example, the 'routingcost' is assumed to be time sentitive with values provided as ALTO Calendars.

The ALTO Client associated to the Application Client queries an ALTO Calendar on 'routingcost' and will get the Calendar covering the 24 hours time period "containing" the date and time of the ALTO client request.

For Cost Type 'num-routingcost', the sollicited ALTO Server has defined 3 different daily patterns each represented by a Calendar, to cover the week of Monday June 30th at 00:00 to Sunday July 6th 23:59:

- C1 for Monday, Tuesday, Wednesday, Thursday, (week days)

- C2 for Saturday, Sunday, (week end)

- C3 for Friday (maintenance outage on July 4, 2014 from 02:00:00 GMT to 04:00:00 GMT, or big holiday such as New Year evening).

In the following example, the ALTO Client sends its request on Tuesday July 1st 2014 at 13:15.

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

{
  "cost-type" : {"cost-mode" : "numerical", "cost-metric" : "routingcost"},
  "calendared" : [true],
  "endpoints" : {
    "srcs": [ "ipv4:192.0.2.2" ],
    "dsts": [
      "ipv4:192.0.2.89",
      "ipv4:198.51.100.34",
      "ipv4:203.0.113.45",
      "ipv6:2000::1:2345:6789:abcd"
    ]
  }
}


HTTP/1.1 200 OK
Content-Length: [TODO]
Content-Type: application/alto-endpointcost+json

{
  "meta" : {
    "cost-type" : {"cost-mode" : "numerical", "cost-metric" : "routingcost"},
    "calendar-response-attributes" : [
       { "calendar-start-time" : Mon, 30 Jun 2014 00:00:00 GMT,
         "time-interval-size" : "1 hour",
         "numb-intervals" : 24,
         "repeated": 4 }
      ],
    } // end meta

    "endpoint-cost-map" : {
      "ipv4:192.0.2.2": {
        "ipv4:192.0.2.89"    : [v1, v2, ... v24],
        "ipv4:198.51.100.34" : [v1, v2, ... v24],
        "ipv4:203.0.113.45"  : [v1, v2, ... v24],
        "ipv6:2000::1:2345:6789:abcd" : [v1, v2, ... v24]
      }
    }
}
```

When the Client gets the Calendar for "routingcost", it sees that the "calendar-start-time" is Monday at 00h00 GMT and member "repeated" is equal to '4'. It understands that the provided values are valid until Thursday included and will only need to get a Calendar update on Friday.

4.2.4. use case and example for the ECS with a multi-cost calendar for routingcost and latency

In this example, it is assumed that the ALTO Server implements multi-cost capabilities, as specified in [draft-ietf-alto-multi-cost] . That is, an ALTO client can request and receive values for several cost types in one single transaction. An illustrating use case is a path selection done on the basis of 2 metrics: routing cost and latency.

As in the previous example, the IRD indicates that the ALTO Server provides "routingcost" Calendars in terms of 24 time intervals of 1 hour each.

For metric "latency", the IRD indicates that the ALTO Server provides Calendars in terms of 12 time intervals values lasting each 5 minutes.

In the following example transaction, the ALTO Client sends its request on Tuesday July 1st 2014 at 13:15.

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

  {
    "cost-type" : {},
    "multi-cost-types" : [
        {"cost-mode" : "numerical", "cost-metric" : "routingcost"},
        {"cost-mode" : "numerical", "cost-metric" : "latency"}
    ],
    "calendared" : [true, true],
    "endpoints" : {
      "srcs": [ "ipv4:192.0.2.2" ],
      "dsts": [
        "ipv4:192.0.2.89",
        "ipv4:198.51.100.34",
        "ipv4:203.0.113.45",
        "ipv6:2000::1:2345:6789:abcd"
      ]
```

```
    }
  }


  HTTP/1.1 200 OK
  Content-Length: [TODO]
  Content-Type: application/alto-endpointcost+json

  {
    "meta" : {
      "multi-cost-types" : [
        {"cost-mode" : "numerical", "cost-metric" : "routingcost"},
        {"cost-mode" : "numerical", "cost-metric" : "latency"}
       ],
      "calendar-response-attributes" : [
         { "cost-type-name : num-routingcost"
           "calendar-start-time" : Mon, 30 Jun 2014 00:00:00 GMT,
           "time-interval-size" : "1 hour",
           "numb-intervals" : 24,
           "repeated": 4 },
         { "cost-type-name : num-latency"
           "calendar-start-time" : Tue, 1 Jul 2014 13:00:00 GMT,
           "time-interval-size" : "5 minute",
           "numb-intervals" : 12}
       ],
      } // end meta

      "endpoint-cost-map" : {
        "ipv4:192.0.2.2": {
          "ipv4:192.0.2.89"    : [[r1, r2, ... r24], [l1, l2, ... l12]],
          "ipv4:198.51.100.34" : [[r1, r2, ... r24], [l1, l2, ... l12]],
          "ipv4:203.0.113.45"  : [[r1, r2, ... r24], [l1, l2, ... l12]],
          "ipv6:2000::1:2345:6789:abcd" : [[r1, r2, ... r24], [l1, l2, ... l12]]
        }
      }
  }
```

When receiving the response, the client sees that the calendar values for 'routing cost' are repeated for 4 iterations. Therefore, in its next requests until the routing cost calendar is expected to change, the client will only need to request a calendar for "latency".

Without the ALTO Calendar extensions, the ALTO client would have no clue on the dynamicity of the metric value change and would spend needless time requesting values at an inappropriate pace. In addition, without the Multi-Cost ALTO capabilities, the ALTO client

would duplicate this waste of time as it would need to send one request per cost metric.

4.3. Recap of rules related to ALTO Cost Calendars

XXXXX TO BE COMPLETED + MOVED AT THE END OF THE SPECS

A Calendar-aware ALTO Server MUST implement the base protocol specified in RFC7285.

When a metric is available as a calendar, it MUST be available as a single value as well. An ALTO Server aquiring cost values in limited time intervals only can construct a single value from the value array.

Calendared information resources MUST be requested via a POST method.

5. Use cases for ALTO Cost Schedule

[THIS SECTION NEEDS TO BE SHORTENED]

This section presents use cases showing the benefits of ALTO Cost calendars for applications needing to decide both "where" to connect and "when".

5.1. Bulk Data Transfer scheduling upon bandwidth calendars

Large Internet Content Providers (ICPs) like Facebook or YouTube, as well as CDNs rely on data replication across multiple sites and time zones to offload the core site and increase user experience through shorter latency from a local site. Typically the usage pattern of these data centers or caches follows a location dependent diurnal demand pattern. In these examples, data replication across the various locations of an ICP, leads to bulk data transfers between datacenters on a diurnal pattern.

In the meantime, there is a degree of freedom on when the content is transmitted from the origin server to the caching node, or from the core site to a local site. However, scheduling these data transfers is a non-trivial task as they should not infer with the user peak demand to avoid degradation of user experience and to decrease billing costs for the datacenter operator by leveraging off-peak hours for the transfer.

As a result, these ICPs need to have a good knowledge on the link utilization patterns between the different datacenters before making an efficient scheduling decision. While usage data today is already gathered and used to schedule data transfers, provisioning these data

gets increasingly complex with the number of CDN nodes and datacenter operators that are involved. In particular, privacy concerns prevent that this kind of data is shared across administrative domains. The ALTO Cost Calendar avoids these problems by presenting an abstracted view of time sensitive utilization maps through a dedicated ALTO service to allow ICPs a coherent scheduling of data transfers across administrative domains and time zones.

Likewise, bandwidth Calendaring allows network operators to reserve resources in advance according to agreements with their customers, enabling them to transmit data with specified starting time and duration, for example, for a scheduled bulk data replication between data centers. Traditionally, this can be supported by a Network Management System operation such as path pre-establishment and activation on the agreed starting time. However, this does not provide efficient network usage since the established paths exclude the possibility of being used by other services even when they are not used for undertaking any service.

An ALTO Cost calendar for TE metrics on transfer paths can support the scheduled bulk data replication with better efficiency since it can alleviate the processing burden on network elements.

Cost calendars for these time-sensitive ALTO TE metrics need to consider the network topology and the dynamicity of the traffic. For example, a small topology with low density and low capacity that carries inpredictable, heavy and bursty traffic has few chances to exhibit stationary TE metric value patterns over large periods and would benefit to use the ALTO Calendar over smaller time slots. Some ALTO TE metric values, even aggregated over time may need to be updated at a frequency that would require doing ALTO requests at a pace that would be overload both the ALTO Client and the Server. Large high capacity topologies would benefit from Cost Calendars with a coarse time granularity for the filtered cost map service where as Calendars of finer time granularity for the Endpoint Cost Service would be better suited for small low density and capacity topologies.

5.1.1. Applicable example transaction

Assuming a Large high capacity topology, an applicable example transaction for this us case is provided by section 4.1.3. "Example transaction for a FCM with a "request-date" bandwidth Calendar".

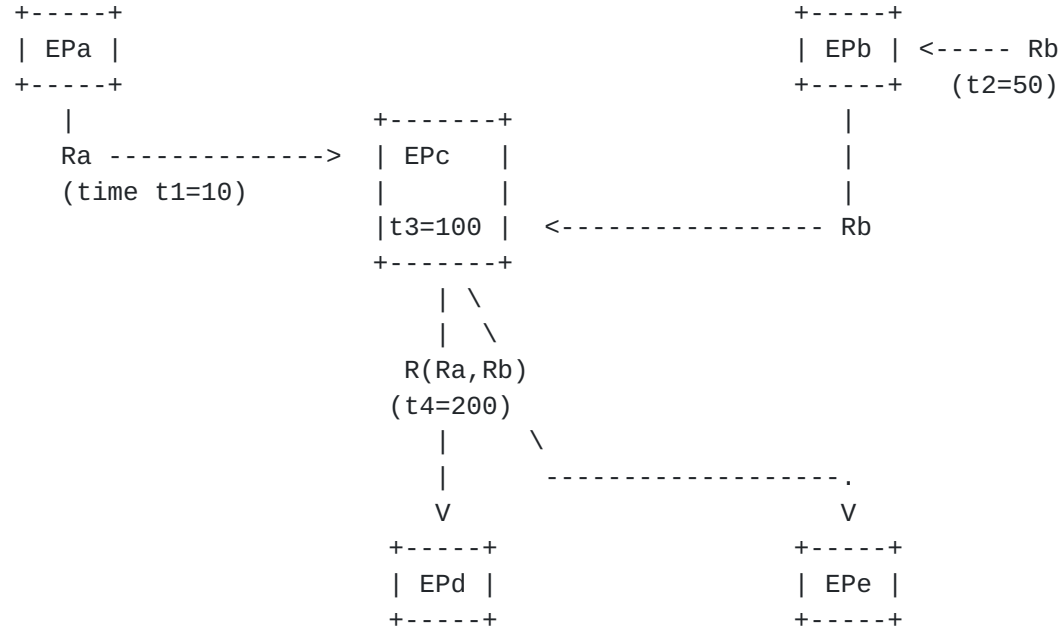
5.2. Applications with limited connectivity or access to datacenters

Some applications are limited in their connectivity either in time or resources or both. For example applications running on devices in remote locations or in developing countries that need to synchronize

their state with a data center periodically, in particular if sometimes there is no connection at all. Example applications are enterprise database update, remote learning, remote computation distributed on several data center endpoints.

Wireless connections have a variable quality and may even be intermittent. On the other hand, the wireless network conditions have a rapid impact on applications while they can sometimes be predicted over a span of time. Non real time applications and time-insensitive data transfers such as client patching, archive syncing, etc. can benefit from careful scheduling. It is thus desirable to provide ALTO clients with routing costs to connection nodes (i.e. Application Endpoints) over different time periods. This would allow end systems using ALTO aware application clients to schedule their connections to application endpoints.

Another challenge arises with applications using data and physical resources scattered around the world. For non-real time applications, the interaction with Endpoints can be orchestratrated and scheduled at the time slots corresponding to the best possible network conditions. For instance, resource Ra downloaded from Endpoint EPa at time t1, Resource Rb uploaded to EPb at time t2, some batch computation involving Ra and Rb done on EPc at time t3 and results R(A,B) downloaded to EPd and EPe at time t4.



5.2.1. Applicable example transaction

An applicable example transaction for this use case is provided by section 4.2.3. "Example transaction for the ECS with a "periodic" routingcost Calendar".

5.3. SDN Controller guided traffic scheduling with Calendars

An ALTO Server can assist an SDN Controller by hosting abstracted network information that can be provided to SDN aware applications via an ALTO Client.

Via the Northbound interface (NBI), applications may get QoE impacting information such as network provider preferences w.r.t. delay and bandwidth on the network paths. Such information may be provided via the ALTO Service.

One key objective of an SDN controller is the ability to balance the application traffic whenever possible. Resources availability may often be predicted and strong incentives for applications to time shift their traffic may be given by network operators appropriately setting routing cost values at different time values, according to their policy on network utilization over time.

To achieve this objective, the SDN controller can:

1. get the network state information from its controlled network elements through its southbound API and derive an estimation of these values over given time frames

2. abstract the network topology and end to end path costs and store them in an ALTO Server as Network Maps and Cost Calendars

3. deliver these values to ALTO Clients linked to SDN applications, through the NBI.

This way:

- On one hand, the applications get the best possible QoE, as they can pick the best time for them to access one or more Endpoints or PIDs,

- One the other hand, the SDN controller achieves load balancing and optimizes application traffic as it may guide the application traffic so as to better distribute the traffic over time.

5.3.1. Applicable example transaction

An applicable example transaction for this use case is provided by section 4.2.4. "Example transaction for the ECS with a calendar on both routingcost and latency".

6. IANA Considerations

Information for the ALTO Endpoint property registry maintained by the IANA and related to the new Endpoints supported by the acting ALTO server. These definitions will be formulated according to the syntax defined in Section on "ALTO Endpoint Property Registry" of [RFC7285] ,

Information for the ALTO Cost Type Registry maintained by the IANA and related to the new Cost Types supported by the acting ALTO server. These definitions will be formulated according to the syntax defined in Section on "ALTO Cost Type Registry" of [RFC7285],

6.1. Information for IANA on proposed Cost Types

When a new ALTO Cost Type is defined, accepted by the ALTO working group and requests for IANA registration MUST include the following information, detailed in Section 11.2: Identifier, Intended Semantics, Security Considerations.

6.2. Information for IANA on proposed Endpoint Propeeries

Likewise, an ALTO Endpoint Property Registry could serve the same purposes as the ALTO Cost Type registry. Application to IANA registration for Endpoint Properties would follow a similar process.

7. Acknowledgements

Thank you to Diego Lopez, He Peng and Haibin Song and the ALTO WG for fruitful discussions.

8. References

8.1. Normative References

[RFC2119] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", BCP 14, RFC 2119, DOI 10.17487/RFC2119, March 1997, <http://www.rfc-editor.org/info/rfc2119>.

[RFC5693] Seedorf, J. and E. Burger, "Application-Layer Traffic Optimization (ALTO) Problem Statement", RFC 5693, DOI 10.17487/RFC5693, October 2009, <http://www.rfc-editor.org/info/rfc5693>.

8.2. Informative References

[draft-ietf-alto-multi-cost]
S. Randriamasy, W. Roome, N. Schwan, , "Multi-Cost ALTO (work in progress), draft-ietf-alto-multi-cost", September 2016.

[draft-ietf-alto-performance-metrics]
Q. Wu, Y. Yang, Y. Lee, D. Dhody, S. Randriamasy, , "ALTO Performance Cost Metrics (work in progress)", September 2016.

[draft-yang-alto-topology]
Y. Yang, , "ALTO Topology Considerations (work in progress)", July 2013.

[ID-alto-protocol]
R.Alimi, R. Penno, Y. Yang, Eds., "ALTO Protocol, RFC 7285", September 2014.

[RFC7285] R. Alimi, R. Yang, R. Penno, Eds., "ALTO Protocol", September 2014.

[sdnrg] "Software Defined Network Research Group, http://trac.tools.ietf.org/group/irtf/trac/wiki/sdnrg".

[slides-88-alto-5-topology]
G. Bernstein, Y. Lee, Y. Yang, , , "ALTO Topology Service: Use Cases, Requirements and Framework (presentation slides IETF88 ALTO WG session), http://tools.ietf.org/agenda/88/slides/slides-88-alto-5.pdf", November 2013.

Authors' Addresses

Sabine Randriamasy
Nokia Bell Labs
Route de Villejust
NOZAY 91460
FRANCE

Email: Sabine.Randriamasy@nokia-bell-labs.com

Richard Yang
Yale University
51 Prospect st
New Haven, CT 06520
USA

Email: yry@cs.yale.edu

Qin Wu
Huawei
101 Software Avenue, Yuhua District
Nanjing, Jiangsu 210012
China

Email: sunseawq@huawei.com

Lingli Deng
China Mobile
China

Email: denglingli@chinamobile.com

Nico Schwan
Thales Deutschland

Email: nico.schwan@thalesgroup.com